Plant Omics Journal POJ

POJ 7(5):291-297 (2014) ISSN:1836-3644

Genome-wide analysis of the heat shock transcription factor family in *Triticum urartu* and *Aegilops tauschii*

Wenlong Yang[1, §], Juan Li[2, §], Dongcheng Liu[1], Jiazhu Sun[1], Lixiong He[2], Aimin Zhang[1,*]

[1]Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, 100101, Beijing, China
[2]College of Life Sciences, Hunan Agricultural University, 410128, Changsha, China
[§] These authors contributed equally to this work.

[*]Corresponding author: amzhang@genetics.ac.cn

Abstract

Heat shock proteins (Hsps) are believed to play essential roles in developmental processes and in responses to heat stress. Heat shock transcription factors (Hsfs) are important Hsp regulators, but their functions are poorly understood, especially in wheat. In this study, a comprehensive bioinformatics analysis was conducted in wheat A and D genome donors, *Triticum urartu* and *Aegilops tauschii*, the genomic sequences of which were published recently. The results showed that 13 Hsf proteins were identified in both *T. urartu* and *Ae. tauschii*, and they could be classified into three groups according to structure; seven Hsfs belonged to group A, two to group B, and three to group C. Expression analyses of these Hsf genes in different tissues of *T. urartu* and in the response to heat stress were conducted using quantitative RT-PCR. Several Hsf genes in group A (*Tuhsf03*, *Tuhsf05*, *Tuhsf06*, *Tuhsf10*) had 19–292-fold increases in transcript levels versus the control in different tissues of *T. urartu* and could be induced by heat stress, while the transcripts of group B and group C Hsf genes could hardly be detected. These results provide important information for cloning, expression, and functional studies of Hsfs in wheat.

Keywords: *Aegilops tauschii*, expression analysis, heat shock transcription factor, phylogenetic analysis, *Triticum urartu*.
Abbreviations: AHA, acidic surrounding motif, DBD, DNA-binding domain, HSE, heat stress element, Hsf, heat shock transcription factor, Hsp, heat shock protein, NES, leucine-rich export signal, NLS, nuclear localization signal, OD, oligomerization domain, qPCR, quantitative real-time polymerase chain reaction, RT-PCR, reverse transcription polymerase chain reaction.

Introduction

Wheat is one of the most important crops globally; it can be processed to bread, noodles, and biscuits, and feeds over 30% of the global population. However, global warming is bringing about serious effects on the productivity and quality of wheat, especially during the grain-filling stage (Ferris et al., 1998; Perrotta et al., 1998). Some studies have revealed that wheat productivity could be reduced by 7-8% with an increase of 1°C in the overnight temperature in the range of 21-34°C (Niu et al., 2008). Thus, it is of great significance to improve the heat tolerance of wheat. Plants possess adaptability to environmental stresses, gained during the evolutionary process, such as rapid responses to low or high temperatures, high salt stress, and water deficit. Heat stress often induces plants to rapidly accumulate heat shock proteins (Hsps), which play a central role in protection against stress damage, and are also involved in the folding, intracellular distribution, and degradation of proteins. The expression of Hsps—and other heat stress-inducible genes—is regulated by a group of heat stress transcription factors (Hsfs) that bind to a conserved binding motif (the heat stress element, HSE) found in the promoters of *Hsps* and other genes (Bienz and Pelham, 1987; Czarnecka-verner et al., 2000; Zhang, 2002). Hsfs can form trimers in the nucleus and function as sequence-specific DNA-binding proteins. Several studies have revealed that at least three repeated HSEs are required for Hsf binding to activate the expression of Hsp genes (Drees et al., 1997; Baniwal et al., 2004; Chen et al., 2006). Plant Hsfs genes have been isolated from many species since the first was identified in tomato (Scharf et al., 1990). Mishra et al. (2002) found that HsfA1 from tomato was a major regulator in the heat response that regulated the expression of Hsps and other Hsfs. Both HsfA2 and HsfB1 play important roles in the response to high-temperature stress in tomato (Howarth et al., 1993). The HsfA4a identified in wheat was found to be involved in the response to heavy metal stress. Overexpression of HsfA4a in rice significantly increased the resistance to heavy metal stress (Shim et al., 2009). To date, 21 Hsfs in *Arabidopsis*, 18 in tobacco, 16 in tomato, 25 in rice, and 34 in *Glycine max* have been identified (Nover et al., 2001; Kotak et al., 2004; Guo et al., 2008). Similar to other transcription factors, the Hsf family is well conserved in structure and function throughout eukaryotes. The DNA-binding domain (DBD), close to the N-terminus, is the most conserved part of Hsfs; it can form a helix-turn-helix structure for specific recognition of the conserved HSE motif (Harrison et al., 1994; Schultheiss et al., 1996; Cicero et al., 2001). An adjacent oligomerization domain (HR-A/B region) forms a helical coiled-coil structure, responsible for the trimerization of Hsfs (Peteranderl et al., 1999). Hsfs can be classified into three major classes (A, B, C) based on differences in the flexible linkers between the A and B parts of the HR-A/B region. The HR-A/B regions of class B Hsfs have no inserted sequences, while all class A and class C Hsfs have an extended HR-A/B region because of the insertion of 21 and 7 amino acid residues, respectively (Novel et al., 2001). Other motifs have also been found in Hsf proteins, such as a nuclear localization signal (NLS) rich in arginine and

lysine residues, which is responsible for the nuclear localization of Hsf proteins, a leucine-rich export signal (NES), and aromatic and large hydrophobic amino acid residues embedded in an acidic surrounding (AHA) motif in the C-terminus of some plant Hsfs (Lyck et al. 1997). The balance between NLS and NES determines the actual nucleocytoplasmic distribution of the Hsf protein, which is important in many signaling pathways involving transcription factors (Heerklotz et al., 2001). The AHA motif is specific for class A Hsfs and activates the transcription of Hsps by binding some basic transcription protein complexes (Morimoto, 1998; Doring et al., 2000). With increasing numbers of genomic sequences now available, phylogenetic analyses of a set of homologous sequences or gene families in different species have become a useful approach to deducing functional diversity and evolutionary relationships (Li and Yang, 2003; Feng et al., 2004). In this study, we identified a non-redundant set of Hsf genes from wheat A and D genome donors, *Triticum urartu* and *Aegilops tauschii*, predicted their structures, and analyzed their expression patterns. These results will help in understanding the evolutionary history and functions of Hsfs, and in improving the heat resistance of wheat.

Results

Identification of Hsf proteins from T. urartu and Ae. Tauschii

To identify Hsf proteins in *T. urartu* and *Ae. tauschii*, the amino acid sequence of the Hsf-type DBD domain (Pfam: PF00447) was used as a query in BLASTP searches for possible homologs in the *T. urartu* and *Ae. tauschii* genomes. As a result, 17 Hsf protein candidates were identified in each of the *T. urartu* and *Ae. tauschii* genomes. After checking using the SMART program, four sequences from each species were removed due to the absence of a coiled-coil structure (HR-A/B region). Consequently, 13 non-redundant Hsfs were identified in each of *T. urartu* and *Ae. tauschii* (Table 1).

Predicted primary structures and properties of T. urartu and Ae. tauschii Hsf proteins

Based on the identified functional domains of Hsfs from *Arabidopsis*, rice, and other species, we determined a set of putative functional domains or motifs of Hsfs from *T. urartu* and *Ae. tauschii* (Tables 2, 3). Five conserved domains were observed in most of the Hsf proteins. Close to the N-terminus, multiple sequence alignment results clearly showed highly conserved DBD domains among *T. urartu* and *Ae. tauschii* proteins. The coiled-coil structure, characteristic of HR-A/B regions in the protein sequences, was also observed. The nuclear localization signal (NLS) was seen in all the Hsf proteins from *T. urartu* and *Ae. tauschii*. Most of the Hsfs also had the NES motif. The conserved AHA motif existed in some of the group A Hsfs from *T. urartu* and *Ae. tauschii*.

Phylogenetic analysis of the Hsf gene family in T. urartu and Ae. Tauschii

To analyze the phylogenetic organization of the Hsf families, a phylogenetic analysis of 13 *T. urartu* Hsfs, 13 *Ae. tauschii* Hsfs, 25 rice Hsfs (OsHsfs), and 21 *Arabidopsis* Hsfs (AtHsfs) was performed, generating a phylogenetic tree (Fig. 1). Most of the Hsfs fell into three major classes: A, B, and C, except AetHsf01 and TuHsf12. There were seven Hsfs each from *T. urartu* and *Ae. tauschii* classified into group A, two in group B, and three in group C. Group A can be divided into A1, A2, A3, and A4 subgroups. TuHsf10 is a member of subgroup A1, TuHsf03, 05, 06, and AetHsf07, 08, 09, and 12 are members of subgroup A2, AetHsf04 belongs to subgroup A3, and TuHsf01, 04, 09, and AetHsf02 and 11 belong to subgroup A4. The Hsfs from *T. urartu* did not completely correspond to those from *Ae. tauschii*, indicating differences in gene structures between *T. urartu* and *Ae. tauschii*. Group B includes TuHsf02, 08 from *T. urartu* and AetHsf06, 10 from *Ae. tauschii*, group C has TuHsf07, 11 and 13 from *T. urartu* and AetHsf03, 05 and 13 from *Ae. tauschii.*

Expression analysis of the Hsf gene family in T. Urartu

The expression patterns of Hsf genes were analyzed in the root, stem, leaf, and spikelets of *T. urartu* using quantitative real-time PCR (qPCR). While transcripts of *hsf01*, *hsf02*, *hsf04*, *hsf07*, *hsf08*, *hsf09*, *hsf11*, *hsf12*, and *hsf13* could hardly be detected in the tissues, *hsf03*, *hsf05*, *hsf06*, and *hsf10* showed high levels of transcription in all the tissues evaluated, especially the leaves. They had 19–292-fold increases in transcript levels versus the control in different tissues of *T. urartu*. Figure 2 indicates that the expression levels of *hsf05* in roots and stems were higher than those of *hsf06*, while in leaves, the expression levels of *hsf06* were higher than those of *hsf05*. These results indicate the expression of the *hsf* genes showed a spatial pattern in *T. urartu*. To assess whether the *hsf* genes played a role in *T. urartu* under heat stress, *T. urartu* seedlings were treated at 42°C for 0, 1, 3, 5, 7, 12, and 24 h, and the transcripts of the *hsf* genes were assayed by qPCR. While transcripts of *hsf01*, *hsf02*, *hsf04*, *hsf07*, *hsf08*, *hsf09*, *hsf11*, and *hsf12* could hardly be detected, other *hsf* genes were induced by heat stress, they showed 0.04–102-fold increases in transcript levels versus the control of *T. urartu*. Transcripts of *hsf06* accumulated gradually during heat stress, and peaked at 7 h after heat treatment, while transcripts of *hsf03* and *hsf05* rapidly reached peak levels at 1 h after heat treatment, then decreased gradually. Expression of *hsf10* and *hsf13* remained essentially stable during heat stress. These results indicated that the *hsf* genes identified showed differing expression patterns during heat stress, and they may play unique roles in response to heat stress.

Discussion

The *hsf* gene family is quite large in plants. Hsf proteins can be divided into three groups based on their structures. In this study, we identified 13 Hsfs from *T. urartu* and 13 from *Ae. Tauschii*; seven of each belonged to group A, two of each to group B, and three of each were classified into group C. Among the group A Hsfs, which can be divided into four subgroups, A1, A2, A3, and A4, some members from *T. urartu* differed from those of *Ae. tauschii*. In subgroup A1, TuHsf10 of *T. urartu* had no apparent homolog in *Ae. tauschii*. AetHsf08 of subgroup A2 and AetHsf04 of subgroup A3 had no apparent homologs in *T. urartu*; TuHsf01 of subgroup A4 and TuHsf13 of group C also had no apparent homologs in *Ae. tauschii*, but TuHsf11 had two homologs, AetHsf03 and AetHsf13, in *Ae. tauschii*. *T. urartu* and *Ae. tauschii* are considered to be the donors of wheat genomes A and D, respectively; the differences in Hsfs between *T. urartu* and *Ae. tauschii* indicate heterogeneity between wheat genomes A and D. Functional analyses of Hsfs revealed that group A Hsfs are involved primarily in the regulation of Hsp genes. Group B proteins, which lacks the AHA motif, can act as transcriptional activators or repressors in tomatoes and *Arabidopsis* (Czarnecka-Verner et al., 2000; Kotak et al., 2004), indicating that group B *hsf* genes may be involved in transcription regulation in collaboration with group

Table 1. Heat shock transcription factor protein information for *Triticum urartu* and *Aegilops tauschii*.

Number	Protein names	NCBI accession no.	Size (aa)	MW (Da)	pI
1	TuHsf01	TRIUR3_30232	414	46616.49	5.58
2	TuHsf02	TRIUR3_31165	365	40285.49	7.87
3	TuHsf03	TRIUR3_27052	413	47307.83	4.59
4	TuHsf04	TRIUR3_18146	401	44588.08	4.81
5	TuHsf05	TRIUR3_19594	567	62355.27	5.00
6	TuHsf06	TRIUR3_02308	460	51171.46	5.80
7	TuHsf07	TRIUR3_19198	411	45344.00	5.98
8	TuHsf08	TRIUR3_26103	322	34877.71	9.00
9	TuHsf09	TRIUR3_11025	432	47408.55	5.10
10	TuHsf10	TRIUR3_08676	533	59897.62	4.99
11	TuHsf11	TRIUR3_34250	402	44582.84	8.98
12	TuHsf12	TRIUR3_33712	379	42443.55	5.26
13	TuHsf13	TRIUR3_12841	266	28058.51	6.77
14	AetHsf01	AEGTA00412	848	96636.92	7.73
15	AetHsf02	AEGTA00937	458	49700.94	5.06
16	AetHsf03	AEGTA02744	321	35380.37	6.43
17	AetHsf04	AEGTA04438	479	52813.48	5.87
18	AetHsf05	AEGTA08948	251	28429.47	9.76
19	AetHsf06	AEGTA11681	298	32057.19	8.95
20	AetHsf07	AEGTA13936	437	49082.78	4.90
21	AetHsf08	AEGTA19036	320	36277.62	5.00
22	AetHsf09	AEGTA19394	442	48832.49	4.86
23	AetHsf10	AEGTA19430	375	40627.69	5.55
24	AetHsf11	AEGTA21225	517	58100.95	5.16
25	AetHsf12	AEGTA27629	480	52874.86	4.96
26	AetHsf13	AEGTA33173	291	31959.24	5.06

Table 2. Functional domains and motifs of TuHsfs and AetHsfs.

Protein name	Group	DBD[a]	OD[a]	NLS[b]	NES[c]	AHA motif[c]
TuHsf01	A4	1-83	98-153	GKKRR	114LKCDNASLKL	350NDGFWQQFLT
TuHsf02	B	36-137	150-182	GRKRM		
TuHsf03	A2	178-277	251-303	KRRR	94LAALTI	382TDDFWEELLS
TuHsf04	A4	21-114	129-162	RRK	223LALVSM	343NDLFWERFLT
TuHsf05	A2	63-156	179-231	KKRRRR		393HDNFWEELLN
TuHsf06	A2	101-194	223-274	KKRRR		416HDDFWEELMS
TuHsf07	C	190-283	293-326	RKKKR		
TuHsf08	B	23-116	172-203	RRRK	238LHIGS	
TuHsf09	A4	1-84	99-130	KKRR	312LSLTL	387NDNFWEQFLT
TuHsf10	A1	1-139	179-217	KRR5KKRR		475IDSFWEQFLC
TuHsf11	C	98-191	227-263	RRKKR	83TELALGLI	
TuHsf12	A1	33-126	150-221	KRKKR	165SLLMQQL	
TuHsf13	C	11-104	127-162	KRK	171LLTVV	
AetHsf01	A1	1-94	126-203	KRKKR	141MKQLVDLRL	
AetHsf02	A4	16-110	125-156	KKRR	340LTL	DNFWEQFLTE
AetHsf03	C	21-114	146-186	RRKKR	10LGLI	
AetHsf04	A3	57-150	182-230	RVKRK	7LEPKEM 458LHL	
AetHsf05	C	15-108	123-162	RKKKR		
AetHsf06	B	26-119	174-203	RRRK	213VRQLDLRL	
AetHsf07	A2	117-210	234-285	KKRRR	242LKRDKNLLI	
AetHsf08	A2	8-101	125-175	KRRRR	131LKRDKQLLM	DDFWEDLLHE
AetHsf09	A2	35-128	216-250	KKRRR		DDFWEELMSR
AetHsf10	B	28-129	196-228	GRKR		
AetHsf11	A4	94-187	202-235	KKRR	343LALVSM	DLFWERFLTD
AetHsf12	A2	39-200	223-263	KKRRRR	150LSISEL	DNFWEELLNK
AetHsf13	C	17-110	147-193	RRKKKR	181LQQAAEKKLQRMQL	

a. Numbers indicate the position of amino acid residues in the protein sequence. b. Amino acid residues indicate the NLS. c. Number indicates the position of amino acid residues in the protein sequence. DBD, DNA-binding domain, OD, oligomerization domain, NLS, nuclear localization signal, NES, leucine-rich export signal, AHA, acidic surrounding motif.

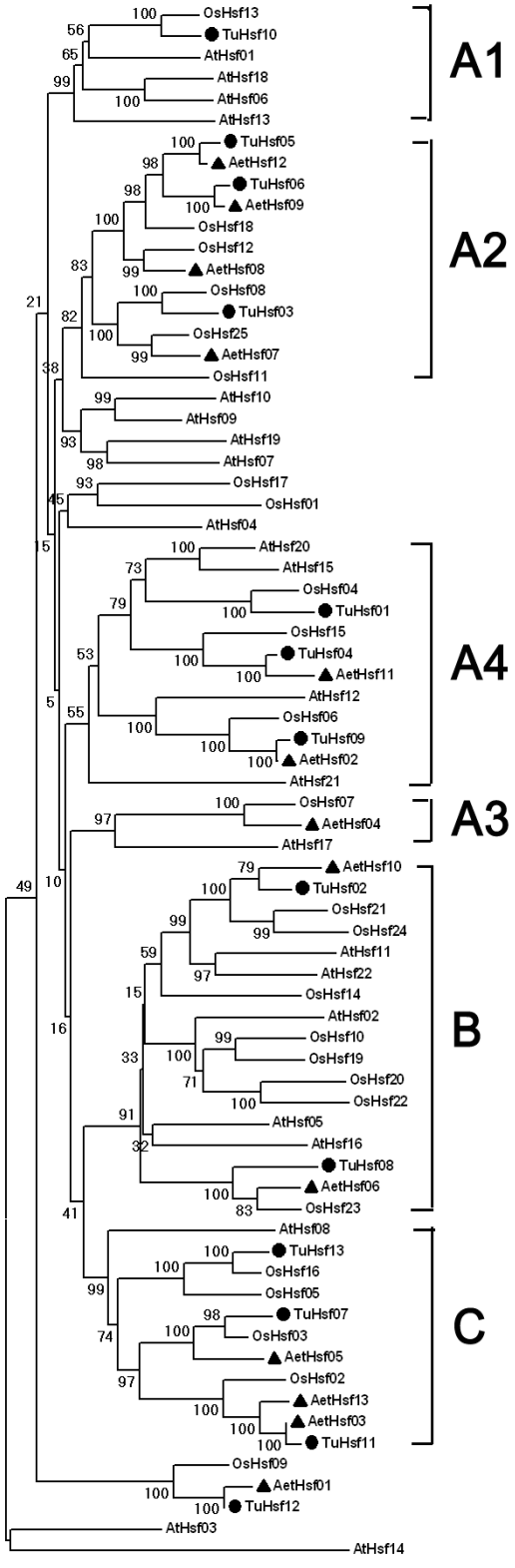

Fig 1. Phylogenetic tree constructed from *T. urartu*, *Ae. tauschii*, *Arabidopsis thaliana*, and *Oryza sativa* Hsfs using full-length amino acid sequences. The Hsfs can be classified into groups A (divided into subgroups A1, A2, A3, and A4), B, and C with high predictive values (bootstrap support of 50 or greater).

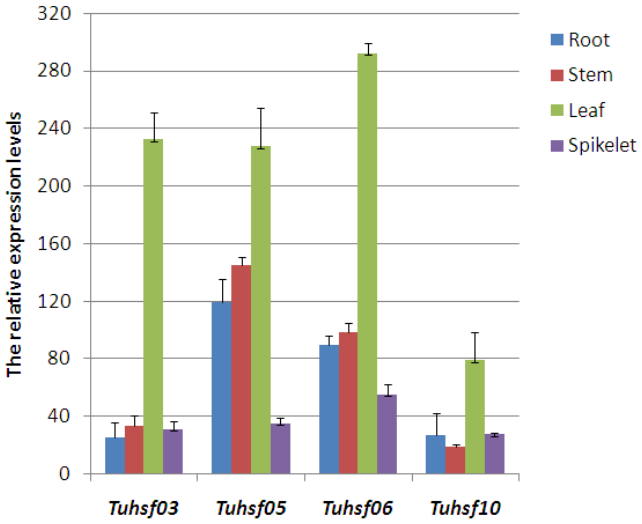

Fig 2. Expression patterns of *Tuhsfs* in root, stem, leaf and spikelets of *T. urartu*, detected by quantitative real-time PCR. Error bars indicate three biological replicates of each sample, together with two technical replicates. *Ta4050* was used as the reference gene.

A. The functions of group C *hsf* genes remain unclear because of limited information. In the present study, group A *hsf* genes (*Tuhsf03*, *05*, *06*, *10*) showed quite high transcription levels in the tissues tested and under heat stress, while the transcripts of group C *hsf* genes (*Tuhsf07*, *11*, *13*) could barely be detected, and those of the group B *hsf* genes could not be detected. These results are consistent with those of previous studies and indicate that group A *hsf* genes play important roles in developmental processes and in response to heat stress in plants, and differential accumulation of *hsf* transcripts in tissues tested, and in response to heat stress, suggest that different *hsf* genes have unique functions. The functions of group B and group C *hsf* genes need further investigation. In summary, the information obtained in this study of the Hsf gene family from *T. urartu* and *Ae. tauschii* will be useful for understanding the evolutionary history of Hsfs. These results will also facilitate the cloning, expression, and functional studies of Hsfs in wheat.

Materials and Methods

Plant materials and stress treatment

Seeds of *T. urartu* accession G1812 were germinated in Petri dishes with two sheets of wet filter paper, and planted in 12×12-cm trays in a greenhouse. At the heading stage, roots, stems, leaves, and spikelets were sampled and snap-frozen in liquid nitrogen, and then stored at -80°C until RNA was isolated for reverse transcription polymerase chain reaction (RT-PCR) analyses. Plants were grown in a greenhouse using liquid culture in Hoagland solution (1950). After growing for 2 weeks, the seedlings were transferred to a chamber at 42°C, and harvested after 0, 1, 3, 5, 7, 12, or 24 h of heat treatment, snap-frozen in liquid nitrogen, and stored at -80°C until RNA isolation.

Table 3. Motif sequences identified by MEME tools.

Motif	Multilevel consensus sequence
1	NNSFVVWDPHAFATVLLPRHFKHSNFSSFVRQLNTYGFRKV
2	DPDRWEFANEGFLRGQRHLLKNIRRRKPPAHTASNQQSLG
3	LGDAGPTPFLAKTYDMVDDPATDAVVSWT
4	QAMEKRLQGTEQKQQQMMSFLARVMQNP
5	GHFGYDAEIDRLKRDKQLLMAEVVKLRQE
6	FLKQLIAKNGMRKELHDAISKKRRRRIDG
7	LGTCEAQQNRAPGLFHDNFWEELLNKGLS
8	GEVSVELLADGVPPELESSVALLADGIPPDLEGAAELLVDV
9	SGSGMGITDGGTAVETPFPFCLLGQCFF
10	IFADMPALPDFEDMHLWFSEDGEPTLTIQDYDEFPQSGQDCQMEAQHNYN
11	NATSLHLLEQAAEKKRQRMQCPSRDFTSFPVALPLHPAPSP
12	EDSHESKDGVKMGLDCYWFGHRNNVDQITEQMGHLASAQKT
13	QSSHEDNGSPHGRHPPVHDGMGTGCLPLVPQIMELSDTGTSICPSKSSFF
14	VEATETASFCDDHSATSRQEMGNLLNQHFSDKLKLGLCSAATESNLVTLS
15	EEEDGHGTGDSPGATAAPRP
16	LALGDAAMGGVRVWQWAEPMPL
17	PAVNDEGLSPCHLSLTLASCSMDVDRGQASNADGGTTGDEGSDNPPEATA
18	PTAADLGAEDEEEKMSARLFGVCIGRKRMRHDGEDLTSRGA
19	SSSLEQELPVVFDSHGSVELL
20	DDNPVNVDGMDVMSEKTDHLVPNSPTRAT

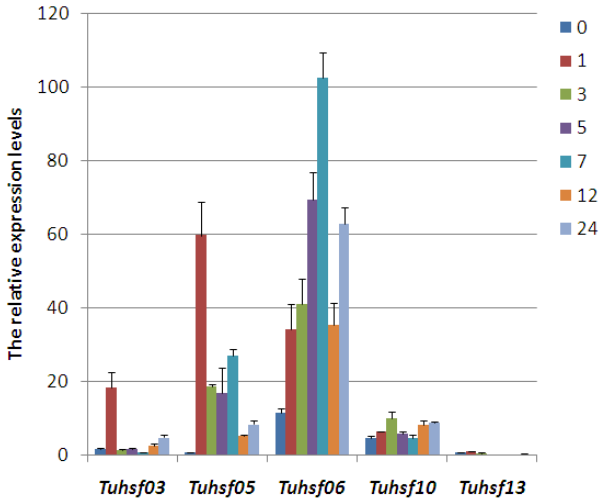

Fig 3. Expression profiles of *Tuhsfs* under heat stress, as determined by quantitative real-time PCR. At 2 weeks, growing seedlings were treated at 42°C for 0, 1, 3, 5, 7, 12, or 24 h. Error bars indicate three biological replicates of each sample, together with two technical replicates. *Ta4050* was used as the reference gene.

Identification of Hsf proteins in T. urartu and Ae. Tauschii

The *T. urartu* and *Ae. tauschii* genome sequences and filtered proteins and cds sequences are now available (their NCBI accession numbers are listed in Table 1; Ling et al., 2013; Jia et al., 2013). We first searched the protein family (Pfam) database with six known Hsf protein sequences (rice Hsf sequences LOC_Os10g28340 (class A), LOC_Os04g48030 (class B), and LOC_Os01g43590 (class C), and *Arabidopsis* Hsf sequences At4g17750 (class A), At4g36990 (class B), and At5g62020 (class C)) and obtained the conserved Hsf-type DBD domain sequences. Next, we generated local databases from the *T. urartu* and *Ae. tauschii* complete genome nucleotide sequences and protein sequences, and used the Hsf domain (PF00447) obtained from the Pfam database as a standard sequence to screen all possible homologs in *T. urartu* and *Ae. tauschii* by BLASTP searches (P = 0.0001). All candidate sequences were checked once more in the Pfam database to remove any sequences not containing the Hsf-type DBD domain, and were confirmed using the SMART program to possess the coiled-coil structure, the core of the HR-A/B region. Sequences lacking the coiled-coil structure were eliminated. Finally, a distinctive name was provided for each Hsf identified in *T. urartu* and *Ae. tauschii*.

Multiple sequence alignment and domain prediction

ClustalW was used to align amino acid sequences of Hsf proteins and GeneDoc was used to manually edit the results. Several web-based bioinformatics tools were used to predict the existence and location of potential domains within the Hsf protein sequences: MARCOIL, PredictNLS, and NetNES 1.1 Server were used to predict the HR-A/B domain, NLS, and NES, respectively. Conserved motif analyses within the determined Hsf groups were performed using MEME.

Phylogenetic analysis of Hsf proteins

A phylogenetic tree was constructed with the neighbor-joining (NJ) method in MEGA (ver. 5.0) using full-length amino acid

Table 4. Primers for qPCR of *hsf* genes in *Triticum urartu*.

Gene	Forward primer (5' → 3')	Reverse primer (5' → 3')
Tuhsf01	GAAGAGGAGACTGCCTAAACCAAT	TGACCTGGGACATCACCATC
Tuhsf02	GACGAGGAGGAGAAGATGAGC	TCCCGCTGTAGCAGTTTGAC
Tuhsf03	CGACGAAGAGGAGGAGGC	TCACCACAGCCCTGATTCC
Tuhsf04	GTCCAATGCTACTATTCCCAGAAAT	CCATCTTCACACCGTCCTTG
Tuhsf05	GCTCCTTCCTCTTATCTTGCTTCT	GGTTGGTGAGTTTGGATTTGAG
Tuhsf06	GAGCAAGGGTCACAGGTAATGTT	CAGCAGGTCCTCCCAGAAAT
Tuhsf07	TGACGCTCAGTGGAACGAA	GGCAACTATGGATGAACGAAAT
Tuhsf08	TCTTTGCGAGACACACACTACAC	TGGCTCAACACTTATGACACG
Tuhsf09	GAGAGCAACCTCGTCACCCT	CAGCCCTTCATCGTTCACAG
Tuhsf10	CTTCTGGGAGCAGTTCCTTTGT	CTATGGTTGGATGACAGCAAGC
Tuhsf11	TGGAGTCGTCGTCCTCATC	GCTGGAGAAGTCGGAATGC
Tuhsf12	CTGATTACGGGACTTCTGGC	CCCTGATGAACCTTGTGTGTT
Tuhsf13	GAGGTGGTGAGGCTGAAGAAG	GAAGTCGCCGTCAAGCAG

sequences from *Arabidopsis*, rice, *T. urartu* and *Ae. tauschii*. NJ analysis was performed with the pairwise deletion option and the Poisson correction. For statistical reliability, bootstrap analysis was conducted with 1000 replicates to assess statistical support for each node. The tree file was visualized using Treeview.

RNA extraction and cDNA synthesis

Total RNAs of *T. urartu* samples were extracted using the TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) and purified with RNase-free DNAse I treatment (Promega, Madison, WI, USA). The concentration of RNA was determined using a UV spectrometer. The first-strand cDNA was synthesized using MMLV (Promega) and an oligo(dT) primer according to the manufacturer's protocol. cDNA quality was assessed by PCR amplification of the tubulin gene of *T. urartu* (TRIUR3_09129).

Quantitative real-time PCR analysis

Total RNA was extracted from various organs of the *T. urartu* accession G1812 as mentioned above, and residual genomic DNA was removed by DNAse I (Promega) digestion prior to reverse transcription (RT). Equivalent amounts of total RNA (2 μg) from each tissue sample were reverse transcribed into cDNA according to the manufacturer's protocol. Quantitative real-time PCR (qPCR) was performed on a Roche LightCycler 480 system with the LightCycler 480 SYBR Green 1 Master Kit (Roche Diagnostics), according to Huang et al. (2012): a 10-μL reaction containing 5-μL SYBR green mix, 2-μL 15-fold-diluted cDNA template, and 3-μL 1 μM forward and reverse primers was used (Table 4). Three biological replicates of each sample, together with two technical replicates, were performed for all *hsf* genes. *Ta4050* (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi) was used as the reference gene.

Conclusions

Thirteen Hsf proteins in each of *Triticum urartu* and *Aegilops tauschii* were identified by searching the two genomic databases, and could be classified into three groups according to their structures. Among them, seven Hsfs belonged to group A, two to group B, and three to group C. Expression analysis of *Tuhsfs* showed that Hsf genes had different expression patterns in different tissues of *T. urartu* and under heat stress. These results provide important information for cloning, expression, and functional studies of Hsfs in wheat.

Acknowledgements

This work was supported by the 863 program (no. 2011AA100104) and the International Cooperation Project (no. 2010DFB33540).

References

Baniwal SK, Bharti K, Chan KY, Fauth M, Ganguli A, Kotak S, Mishra SK, Nover L, Port M, Scharf KD, Tripp J, Weber C, Zielinski D, von Koskull-Döring P. (2004) Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. J Biosci. 29: 471-487.

Bienz M, and Pelham HR (1987) Mechanisms of heat-shock gene activation in higher eukaryotes. Adv Genet. 24: 31−72.

Chen XJ, Ye CJ, Lu HY (2006) Cloning of GmHSFA1 gene and its overexpression leading to enhancement of heat tolerance in transgenic soybean. Acta Genet Sin. 28: 1411−1420.

Cicero MP, Hubl ST, Harrison CJ, Littlefield O, Hardy JA, Nelson HC. (2001) The wing in yeast heat shock transcription factor (HSF) DNA-binding domain is required for full activity. Nucleic Acids Res. 29: 1715−1723.

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

Doring P, Treuter E, Kistner C, Lyck R, Chen A, Nover L (2000) The role of AHA motifs in the activator function of tomato heat stress transcription factors HsfA1 and HsfA2. Plant Cell. 12: 265−278.

Drees BL, Grotkopp EK, Nelson HC (1997) The GCN4 leucine zipper can functionally substitute for the heat shock transcription factor's trimerization domain. J Mol Biol. 273: 61–74

Feng Y, Liu QP, Xue QZ (2004) Phylogenetic analysis of PHD-finger protein in Rice and *Arabidopsis thaliana*. Acta Genet Sin. 31: 1284−1293.

Ferris R, Ellis RH, Wheeler TR, Hadley P (1998) Effect of high temperature stress at anthesis on grain yield and biomass of field grown crops of wheat. Plant Cell Environ. 34: 67-78.

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

Harrison CJ, Bohm AA, Nelson HC (1994) Crystal structure of the DNA binding domain of the heat shock transcription factor. Science. 263: 224–227.

Heerklotz D, Doring P, Bonzelius F, Winkelhaus S, Nove L (2001) The balance of nuclear import and export determines the intracellular distribution and function of tomato heat stress transcription factor HsfA2. Mol Cell Biol. 21: 1759–1768.

Hoagland DR, and Arnon DI (1950) The water-culture method for growing plants without soil. Circ. 347. Berkeley, CA: Agric Exp Stn, Univ of Calif.

Howarth CJ, and Ougham HJ (1993) Gene expression under temperature stress. New Phytol. 125: 1-2.

Huang Y, Yang W, Pei Z, Guo X, Liu D, Sun J, Zhang A (2012) The genes for gibberellin biosynthesis in wheat. Funct Integr Genomics. 12: 199–206.

Jia JZ, Zhao SC, Kong XY, Li YG, Zhao GY, He WM, Appels R, Pfeifer M, Tao Y, Zhang XY, Jing RL, Zhang C, Ma YZ, Gao LF, Gao C, Spannagl M, Mayer KFX, Li D, Pan SK, Zheng FY, Hu Q, Xia XC, Li JW, Liang QS, Chen J, Wicker T, Gou CY, Kuang HH, He GY, Luo YD, Keller B, Xia QJ, Lu P, Wang JY, Zou HF, Zhang RZ, Xu JY, Gao JL, Middleton C, Quan ZW, Liu GG, Wang J, IWGSC, Yang HM, Liu X, He ZH, Mao L, Wang J. (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature. 496: 91-95.

Kotak S, Port M, Ganguli A, Bicker F, von Koskull-Döring P (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.

Li CX, and Yang Q (2003) Phylogenetic relationships among the genera of Taxodiaceae and Cupressaceae from 28S rDNA sequences. Acta Genet Sin. 25: 177−180.

Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C, Wu H, Li Y, Cui Y, Guo X, Zheng S, Wang B, Yu K, Liang Q, Yang W, Lou X, Chen J, Feng M, Jian J, Zhang X, Luo G, Jiang Y, Liu J, Wang Z, Sha Y, Zhang B, Wu H, Tang D, Shen Q, Xue P, Zou S, Wang X, Liu X, Wang F, Yang Y, An X, Dong Z, Zhang K, Zhang X, Luo MC, Dvorak J, Tong Y, Wang J, Yang H, Li Z, Wang D, Zhang A, Wang J. (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature. 496: 87-90.

Lyck R, Harmening U, Hohfeld I, Treuter E, Scharf KD, Nover L (1997) Intracellular distribution and identification of the nuclear localization signals of two plant heat-stress transcription factors. Planta. 202: 117–125.

Mishra S K, Tripp J, Winkelhaus S, Tschiersch B, Theres K, Nover L, Scharf KD (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.

Morimoto RI (1998) Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes Dev. 12: 3788−3796.

Niu H, Chen X, Bai R, Deng D, Li Q, Ren J, Li Y, Yin J (2008) Cloning and expression of TaHSP16.9-1 cDNA in wheat. J Triticeae Crops. 28: 719-723.

Nover L, Kapil B, Doring P, Mishra SK, Arnab G, Scharf KD (2001) *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperones. 6: 177-189.

Perrotta C, Treglia AS, Mita G, Giangrande E, Rampino P, Ronga G, Spano G, Marmiroli N (1998) Analysis of mRNAs from ripening wheat seeds: the effect of high temperature. J Cereal Sci. 27: 127-132.

Peteranderl R, Rabenstein M, Shin YK, Liu CW, Wemmer DE, King DS, Nelson HC (1999) Biochemical and biophysical characterization of the trimerization domain from the heat shock transcription factor. Biochemistry. 38: 3559−3569.

Qin D, Peng H, Ni Z, Xu P, Dong J, Zhang J, Sun Q, Li M (2012) Advances in relationship between heat shock transcription factors and thermo-tolerance of plants. J Triticeae Crops. 32: 178-183.

Qin D, Wu H, Peng H, Yao Y, Ni Z, Li Z, Zhou C, Sun Q. (2008) Heat stress-responsive transcriptome analysis in heat susceptible and tolerant wheat (*Triticum aestivum* L.) by using wheat genome array. BMC Genomics. 9: 432.

Scharf KD, Rose S, Zott W, Schoffl F, Novel L (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.

Schultheiss J, Kunert O, Gase U, Scharf KD, Nover L, Ruterjans H (1996) Solution structure of the DNA-binding domain of the tomato heat-stress transcription factor HSF24. Eur J Biochem. 236: 911−921.

Shim D, Hwang JU, Lee J, Lee S, Choi Y, An G, Martinoia E, Lee Y (2009) Orthologs of the class A4 hest shock transcription factor HsfA4a confer cadmium tolerance in wheat and rice. Plant Cell. 21: 4031-4043.

Xin H, Chen L, Li X, He X, Yi M (2010) Cloning and Expression Analysis of hsf from *Lilium longiforum*. Acta Horticult Sin. 37: 435-440.

Zha Z, and Wan B (2009) Advance in the Research of Heat Stress Transcription Factors in Plants. Hubei Agricult Sci. 48: 2263-2265.

Zhang W, and Wang D (2002) Regulation of heat shock gene transcription: the structure and function of heat shock transcription factor. Acta Academiae Medicinae Militaris Tertiae. 22: 198-200.