



Article

# Genome-Wide Investigation of Heat Shock Transcription Factor Family in Wheat (*Triticum aestivum* L.) and Possible Roles in Anther Development

Jiali Ye, Xuetong Yang, Gan Hu, Qi Liu, Wei Li, Lingli Zhang and Xiyue Song \*

College of Agronomy, Northwest A&F University, Yangling 712100, Shaanxi, China; yejiali@nwafu.edu.cn (J.Y.); yangxuotong@nwafu.edu.cn (X.Y.); huyunuo@163.com (G.H.); liuqi@nwafu.edu.cn (Q.L.); smallliwei@nwafu.edu.cn (W.L.); zhanglingli@126.com (L.Z.)

\* Correspondence: songxiyue@126.com; Tel.: +86-29-8708-2845

Received: 22 November 2019; Accepted: 15 January 2020; Published: 17 January 2020



**Abstract:** Heat shock transcription factors (HSFs) play crucial roles in resisting heat stress and regulating plant development. Recently, HSFs have been shown to play roles in anther development. Thus, investigating the HSF family members and identifying their protective roles in anthers are essential for the further development of male sterile wheat breeding. In the present study, 61 wheat HSF genes (*TaHsfs*) were identified in the whole wheat genome and they are unequally distributed on 21 chromosomes. According to gene structure and phylogenetic analyses, the 61 *TaHsfs* were classified into three categories and 12 subclasses. Genome-wide duplication was identified as the main source of the expansion of the wheat HSF gene family based on 14 pairs of homeologous triplets, whereas only a very small number of *TaHsfs* were derived by segmental duplication and tandem duplication. Heat shock protein 90 (HSP90), HSP70, and another class of chaperone protein called htpG were identified as proteins that interact with wheat HSFs. RNA-seq analysis indicated that *TaHsfs* have obvious period- and tissue-specific expression patterns, and the *TaHsfs* in classes A and B respond to heat shock, whereas the C class *TaHsfs* are involved in drought regulation. qRT-PCR identified three *TaHsfA2bs* with differential expression in sterile and fertile anthers, and they may be candidate genes involved in anther development. This comprehensive analysis provides novel insights into *TaHsfs*, and it will be useful for understanding the mechanism of plant fertility conversion.

**Keywords:** evolution; genome-wide; heat shock transcription factor; thermo-sensitive male sterility; wheat

## 1. Introduction

The heat shock response is common in plants where a series of stress responses are generated under heat stress [1]. As a consequence of heat stress, plant heat shock elements (HSEs) present in the promoter regions upstream of the heat shock protein (HSP) genes are recognized by the activated heat shock transcription factors (HSFs), and they induce the transcription of *Hsp* genes as molecular chaperones to help refold, assemble, distribute, and degrade related proteins, as well as repairing damaged proteins and maintaining cell survival [2–4]. In particular, regulation of the activity of HSFs is the key mechanism responsible for the transcription and expression of *Hsps*.

A typical HSF has three components comprising the N-terminal DNA-binding domain (DBD), oligomeric domain (OD), and nuclear localization signal (NLS). Some HSFs also have a C-terminal activation domain (CTAD) and nuclear export signal (NES) [5]. DBDs are located at the N-terminus of HSFs and they are highly conserved regions with three helical structures (H1, H2, and H3) and

four inverted parallel  $\beta$ -sheets ( $\beta 1$ ,  $\beta 2$ ,  $\beta 3$ , and  $\beta 4$ ) [6]. The helix-turn-helix (H2-T-H3) structure at the hydrophobic center of the DBD allows the precise localization and recognition of HSE sequences (5'-AGAAAnnTTCT-3') [7]. The OD region is also located at the N-terminus of HSFs and it comprises two hydrophobic heptad repeat regions: A and B (HR-A/B). HR-A contains 5–6 sets of hydrophobic heptapeptide repeats, whereas HR-B has two overlapping heptapeptide repeats, which form a helical coiled-coil structure in space [8]. According to the characteristics of the HR-A/B region, plant HSFs can be classified into class A, class B, and class C. Class A and class C HSFs have 21 and 7 aa inserts between regions HR-A and HR-B, respectively, whereas class B HSFs are relatively simple with no amino acid insertions [9]. The synergy between the NLS and NES maintains the cell balance, thereby allowing plant HSFs to be freely distributed in the cytoplasm and nucleus [10]. Together with the NES, the CTAD acidic amino acid (AHA) motifs serve as a type-specific region in the C-terminus of class A HSFs in plants, which is characterized by aromatic and large hydrophobic amino acids [11]. By contrast, the B and C class HSFs have no activation functions due to the lack of an AHA domain [5].

The first HSF gene found in plants was cloned in tomato (*Solanum lycopersicum*) [12] and the HSF gene family have now been characterized in *Arabidopsis* (*Arabidopsis thaliana*), rice (*Oryza sativa*) [13], maize (*Zea mays*) [14], soybean (*Glycine max*) [15], chickpea (*Cicer arietinum*) [16], and sorghum (*Sorghum bicolor*) [17], as well as in vegetables including Chinese cabbage (*Brassica rapa*) [18] and pepper (*Capsicum annuum*) [19], and fruits such as apple (*Malus domestica*) [20] and pear (*Pyrus bretschneideri*) [21]. In addition to the *Hsfs* involved in the regulation of heat-resistance mechanism studied in *Arabidopsis* [22,23], many *Hsfs* have been found to participate in other responses in various plant species in recent years. The overexpression of *AtHsfA2* improves the tolerance of heat in *Arabidopsis* but it also enhances the salt tolerance [24], hypoxia resistance [25], and resistance to oxidative stress [26]. Under high temperature stress, the overexpression of *LeHsfA2* increases the fruit set rate in tomato and enhances fruit development [27]. Shim et al. found that the overexpression of *HsfA4a* in rice increased the tolerance to cadmium [28]. Almoguera et al. suggested that the overexpression of *HaHsfA9* alone or *HaHsfA9* together with *HaDreb2* can promote the synthesis and accumulation of HSPs in tobacco seeds and increase the seed life [29]. The heat-induced tomato *LeHsfB1* is a coactivator of class A *Hsfs* (e.g., *HsfA1a*) and it has an auxiliary activation function by enhancing the expression of target genes [30]. However, studies have rarely shown that class C *Hsfs* can be induced by various stresses [18].

Higher temperature directly leads to a decreased number of spikes per ear and spikelets per spike, and a lower grain weight in wheat [31]. There is no doubt that improving the heat tolerance of wheat and ensuring adequate food production are vital for humans. Members of the HSF family have been identified in wheat based on the old reference genome [32], but a lack of sufficient information has hindered in-depth studies of the stress resistance mechanisms related to *TaHsfs*. A limited number of *TaHsfs* have been functionally analyzed, and studies have demonstrated that *TaHsfA4a*, *TaHsfA6e*, *TaHsfB2d*, and *TaHsfC2a* function as wheat HSF genes to enhance the tolerance of heat, confer strong cadmium tolerance, and participate in protecting the grains [28,33–35]. Therefore, a more comprehensive and systematic analysis is required of *TaHsfs* based on the newly published wheat reference genome in order to provide a reliable basis for elucidating their functions.

The importance of heterosis utilization in increasing plant yield is self-evident. The successful use of two-line hybrid rice has effectively proved the great advantages of photo/ thermo-sensitive male sterility in practical production. The thermo-sensitive male-sterile wheat KT3315A obtained by our research team can be used for selfing breeding under high temperature conditions and for cross breeding under low temperature conditions, which greatly facilitates wheat breeding procedures and hybrid production. The fertility conversion of thermo-sensitive male sterile wheat is controlled by temperature; therefore, understanding its temperature response mechanism is the first task to understand the fertility conversion mechanism. Significant differences in microspore development and anther appearance in sterile and fertile KTM3315A have been verified [36]. In addition to genes that synthesize nutrients necessary for anther and microspore development, we found that *Hsfs* and *Hsps* were differentially expressed under different fertility conditions (different environmental

temperatures during the wheat growing season), and thus we speculated that they may be related to anther development [37]. The *Hsfs* on the sex chromosomes of humans and animals have been shown to affect spermatogenesis [38,39]. Coincidentally, in recent years, the HSFs involved in male reproductive cell protection have also attracted attention in plants. In the rice spike, eight HSF genes are up-regulated in anther development [40]. In addition, the up-regulated expression levels of *HsfA2* and *HsfA3* in tomato pollen grains demonstrate their involvement in anther development [27,41]. However, the HSF gene family members involved in anther protection have not yet been reported in wheat. Therefore, it is necessary to comprehensively investigate the HSF gene family in wheat and explore their possible protective roles in anther development.

In this study, we comprehensively analyzed the wheat HSF family based on the latest release of the complete wheat reference genome. We identified the members of the HSF gene family in wheat as well as investigated their evolutionary relationships, gene structures, and expression levels. Our findings may facilitate subsequent studies of the stress resistance mechanism in wheat, as well as provide a reference for understanding molecular temperature sensing and regulation mechanisms in plants with temperature-sensitive biological characteristics.

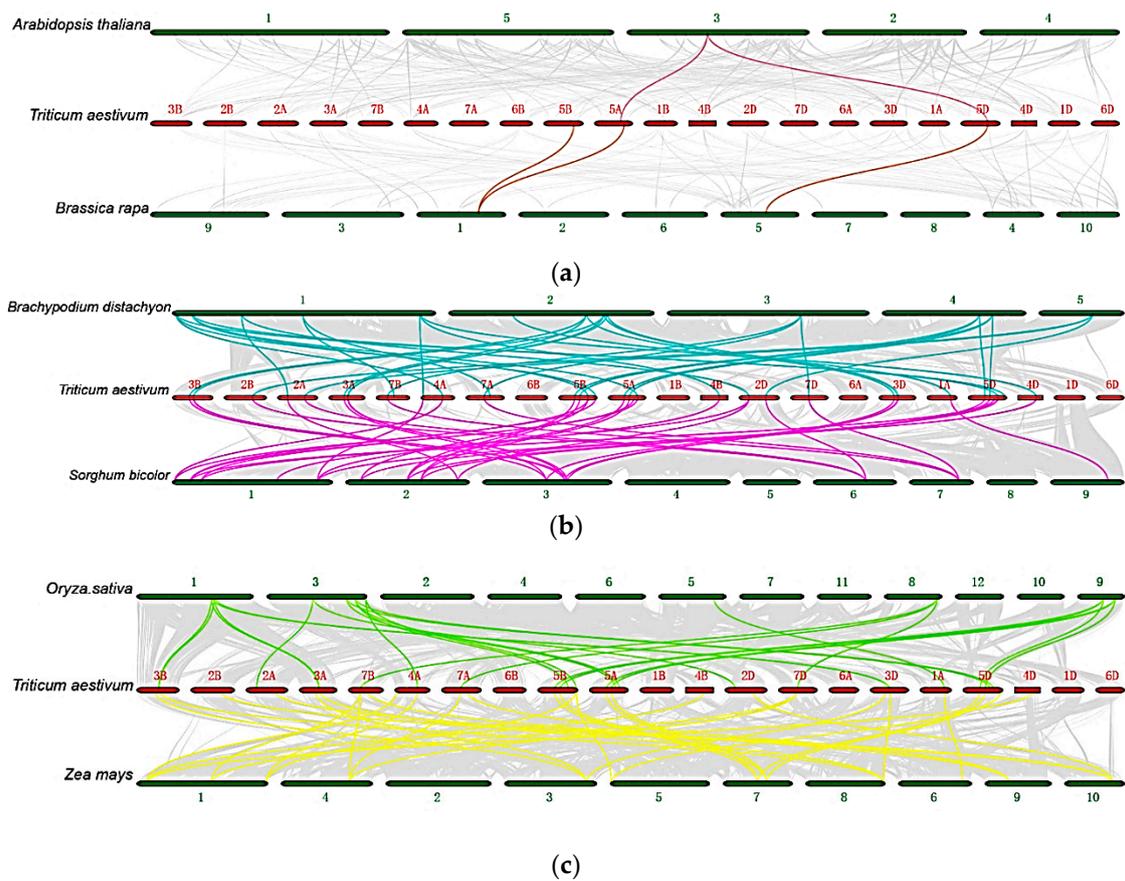
## 2. Results

### 2.1. Identification and Characteristics of *TaHsfs*

The newly constructed wheat-specific hidden Markov model (HMM) file was used to search the whole wheat protein sequences and 94 candidate *Hsfs* were obtained, which were consistent with the wheat candidate *Hsfs* obtained by aligning with *Arabidopsis* and rice *Hsfs*. After excluding candidate genes without a DBD domain and coiled-coil structure, 61 non-redundant genes were finally identified as wheat HSF family members (Table S1, Table S2). These 61 *TaHsfs* were found to be unequally distributed on 21 chromosomes (Figure S1), where the 5A chromosome contained the most *TaHsfs* (6) and one to five *TaHsfs* were unevenly distributed on the other chromosomes. These HSF genes identified in wheat encode proteins ranging from 227 to 569 amino acids (aa) with an average of 360 aa. The molecular weights of the *TaHsfs* vary from 24.69 (*TaHsfC1b-3*) to 59.76 (*TaHsfB2c-1*), and the predicted isoelectric points range from 4.85 (*TaHsfA2b-2*) to 9.52 (*TaHsfB1a-3*) (Table S3).

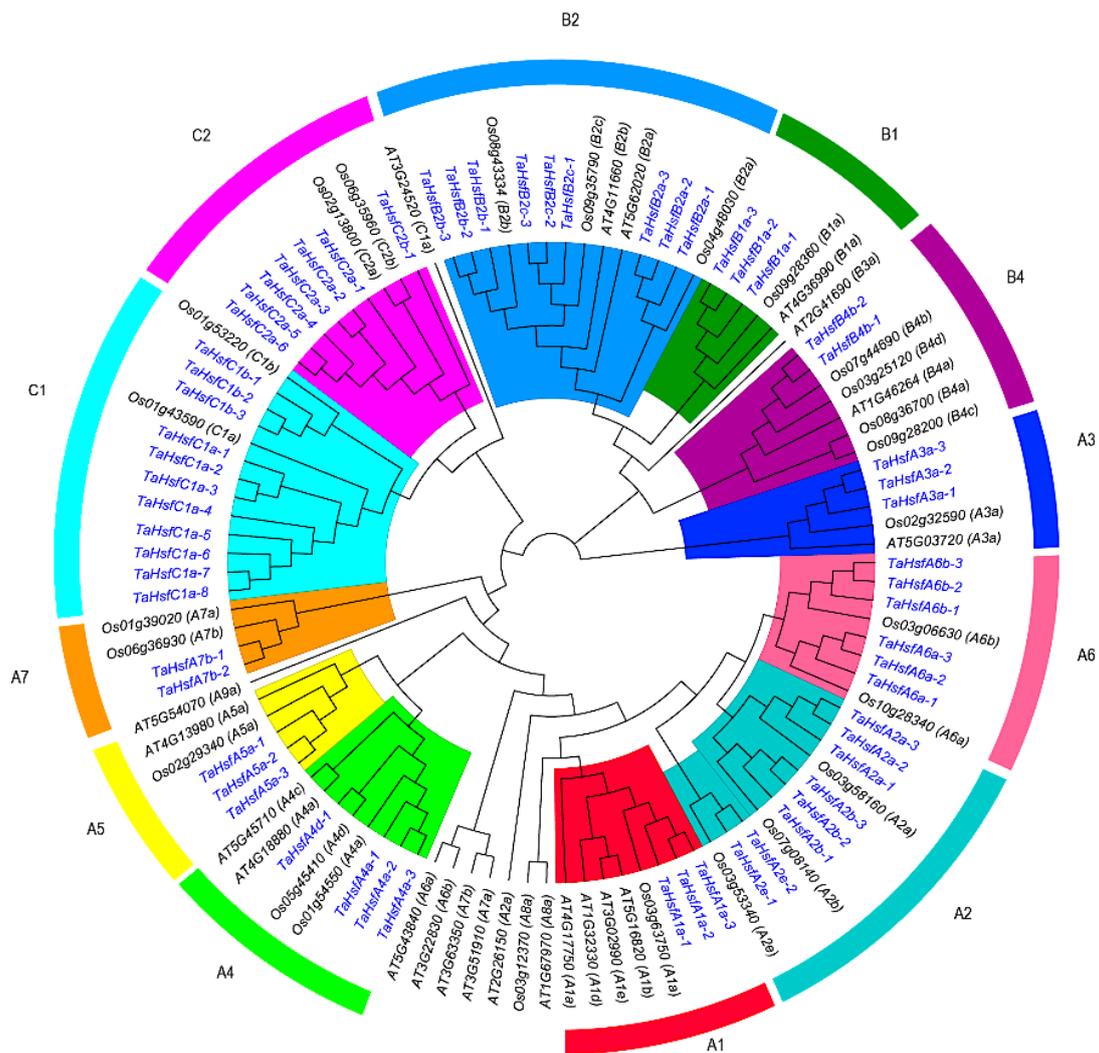
### 2.2. Evolution and Phylogenetic Analysis of *TaHsfs*

In two different species, the conservation of the position and order of homologous genes at a locus is called synteny of the genome [42]. If the species are closer, the genome coverage of the synteny fragment is the larger and it will contain more genes. To determine the evolutionary trends and genetic relationships in the wheat HSF gene family, we performed synteny analysis between wheat and *Arabidopsis*, Chinese cabbage, rice, maize, sorghum, and *Brachypodium* (*Brachypodium distachyon*). The results showed that *TaHsfs* are distantly related to the *Hsfs* in *Arabidopsis* and Chinese cabbage, with only two and three HSF genes within the synteny segment, respectively (Figure 1A). By contrast, the number of HSF genes in synteny were higher between wheat and rice (25), sorghum (39), maize (30), and *Brachypodium* (34) (Figure 1B,C), which all belong to the Gramineae. *Arabidopsis* and Chinese cabbage are both cruciferous, so 21 HSF genes were within the synteny segment (Figure S2). After the differentiation of gramineous species, synteny at the entire chromosome level was found [43]. In this study, the HSF gene family in wheat and other gramineous plants also exhibited closer kinship. Clearly, our results indicate that the HSF gene family remains stable and the evolution of *Hsfs* is consistent with that of the whole genome.



**Figure 1.** Synteny relationships of *Hsfs* in wheat and six representative species. (a) Synteny relationships of *Hsfs* between wheat with *Arabidopsis* (*Arabidopsis thaliana*) and Chinese cabbage (*Brassica rapa*). (b) Synteny relationships of *Hsfs* between wheat with rice (*Oryza sativa*) and maize (*Zea mays*). (c) Synteny relationships of *Hsfs* between wheat with sorghum (*Sorghum bicolor*) and *Brachypodium* (*Brachypodium distachyon*). Gray lines in the background indicate the synteny blocks for wheat and other plant genomes, and the lines in other colors highlight the synteny of *Hsf* pairs.

In order to clarify the classification and evolutionary status of the members of the wheat HSF gene family, 108 full-length amino acid sequences HSF in the related species rice (25), and model plants *Arabidopsis thaliana* (22) and wheat (61) were used to construct a phylogenetic tree (Figure 2). Based on the clustering results, 61 *TaHsfs* were classified into three major categories: A, B, and C. The *TaHsfs* in class A belonged to seven subclasses comprising A1–A7; the class B *TaHsfs* were divided into three subclasses comprising B1, B2, and B4; and the class C *TaHsfs* were divided into two subclasses comprising C1 and C2. Compared with the rice and *Arabidopsis* HSF gene families, the wheat HSF gene family lacks members belonging to A7, A8, A9, and B3. Few orthologous genes were detected in the HSF family members between *Arabidopsis* and wheat. However, in every subclass, at least one rice HSF was found to be highly homologous with the wheat HSFs, and almost every rice HSF was shown to be orthologous with three wheat HSFs. Therefore, the number of wheat HSFs is close to or equal to three times that in rice HSF in terms of the total number and the number in each subclass.



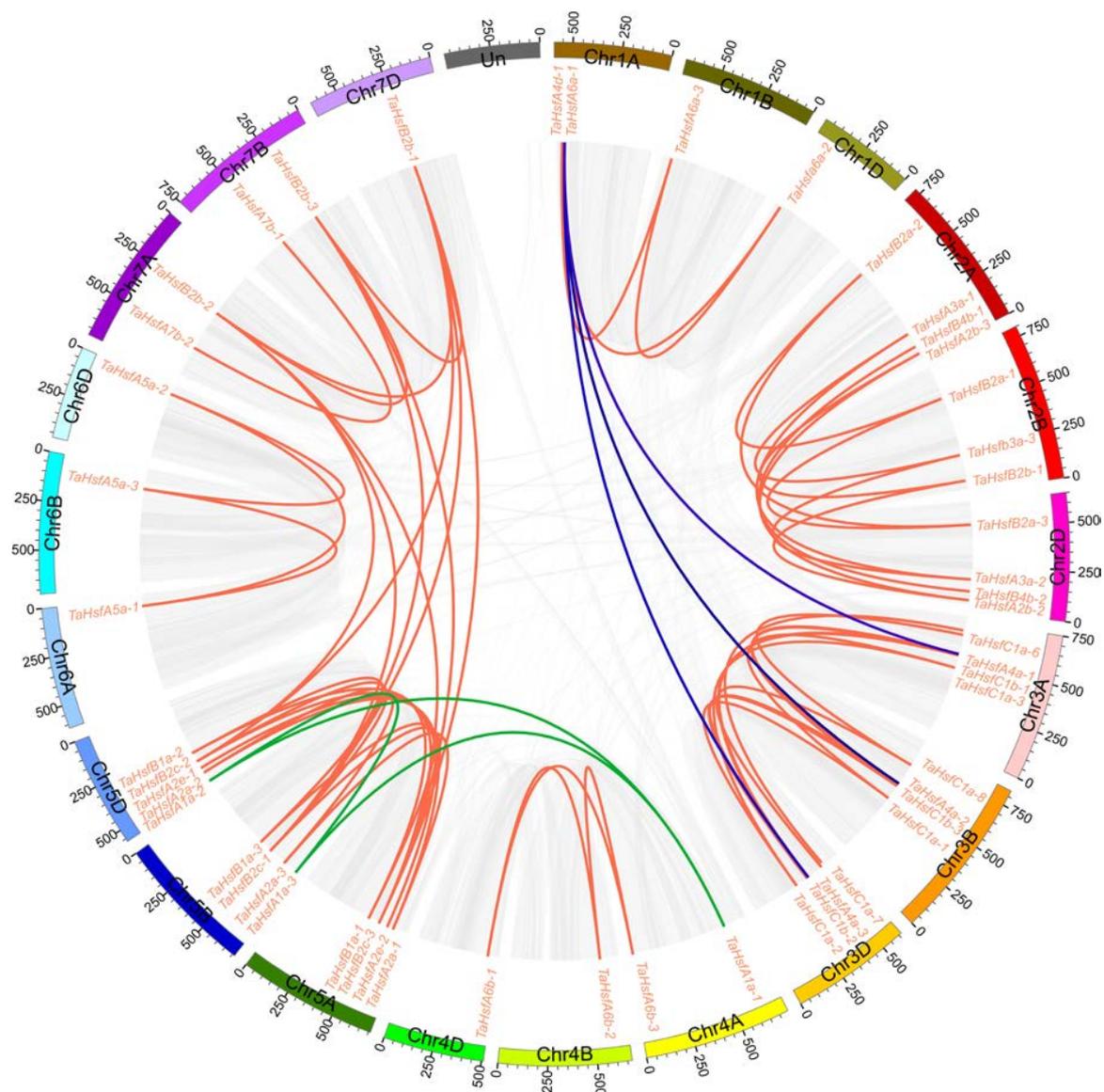
**Figure 2.** Phylogenetic tree obtained for the heat shock transcription factor (HSF) family members in wheat, rice, and Arabidopsis. Different colors represent different sub-classes in the HSF gene family.

### 2.3. Duplication Events in Wheat HSF Family

To understand the source of the larger number of *TaHsfs*, we analyzed the causes of the expansion of the wheat HSF family. We performed homeologous group analysis for the wheat HSF family based on the homologous alignment within the wheat HSF family, the distribution of each gene on the chromosomes, and clustering in the phylogenetic tree. As expected, 68.8% of the wheat HSFs exhibited homology of 1:1:1, where three *TaHsfs* localized on the three A, B, and D sub-genomes shared high homology, which we refer to as triplets (Table 1, Table S4). The proportion of homeologous triplets in the wheat HSF family was close to twice the proportion of homeologous triplets in the entire wheat genome (35.8%) [44]. This high proportion indicates that wheat polyploidization was the main cause of the expansion of the wheat HSF family, and the triplets remained stable during the long evolutionary process. However, gene losses (1:0:1/1:1:0) occurred in a very low number of triplets, where the proportion of losses was close to that in the whole wheat genome, and such losses are usually induced by function redundancy. Further investigation is needed to understand more complex phenomena (orphans/singletons and others), and the expansions in C1 and C2 classes indicated that some other gene duplication events occurred in the wheat HSF gene family.

We performed synteny analysis to further investigate the sources of other *TaHsfs*, i.e., the causes of wheat HSF gene family expansion in addition to genome-wide duplication. Figure 3 shows that all

*TaHsfs* derived from genome-wide duplication were involved in segmental duplication events (Table S5), which probably led to the high homology of the A, B, and D sub-genomes in wheat. The paralogues *TaHsfA1a-1*, -2, and -3 located on the 4A, 5B, and 5D chromosomes, respectively, were derived from segmental duplication events. Interestingly, the orphan gene *TaHsfA4d-1* (Chr1A) identified above was a segmental duplicate of *TaHsfA4a-1*, -2, and -3 (Chr3A, 3B, and 3D). In addition, two pairs of tandem duplications comprising *TaHsfC1a-6/TaHsfC1a-4* and *TaHsfC1a-7/TaHsfC1a-5* occurred in the wheat HSF gene family on the 3A and 3D chromosomes, respectively (Figure S1), and their KA/KS values were 0.367731 and 0.388683. The analysis of these two duplication events complemented the sources of the four genes described above (*A1a-a*, *A4d-1*, *C1a-4*, *C1a-5*), but the seven *TaHsfs* in the C2 subclass were unexpectedly not present in either case. The unusual features of the C2 sub-class members indicated that further analysis of their structure is required.



**Figure 3.** Genomic locations of *Hsfs* and segmental duplicated gene pairs in the wheat genome. Gray lines in the background indicate the syntenic blocks within the whole wheat genome, and red lines denote the segmental duplication HSF gene pairs. The blue lines show genes that have undergone segmental duplication with *TaHsfA4d-1*, while the green lines show the genes that have undergone segmental duplication with *TaHsfA1a-1*.

**Table 1.** Homeologous groups in wheat HSF gene family.

Homoeologous Group (A:B:D)	All Wheat Genes	Classes			Number of Groups	Number of Genes	% of Total TaHSFs
		A	B	C			
1:1:1	35.8%	7	4	3	14	42	68.8
1:1:0/1:0:1	13.2%	2	1	1	4	8	13.1
Orphans/singletons	37.1%	1		1	2	2	3.3
Others	-	1		1	2	9	14.8
Total	-	11	5	6	20	61	100

All wheat genes are distribution among homeologous groups in the whole wheat genome according to IWGSC. A, B, and D represent the three wheat sub-genomes. "Others" denotes circumstances other than those in the table above (see Table S4 for details).

#### 2.4. Gene Structure and Motif Composition in *TaHsfs*

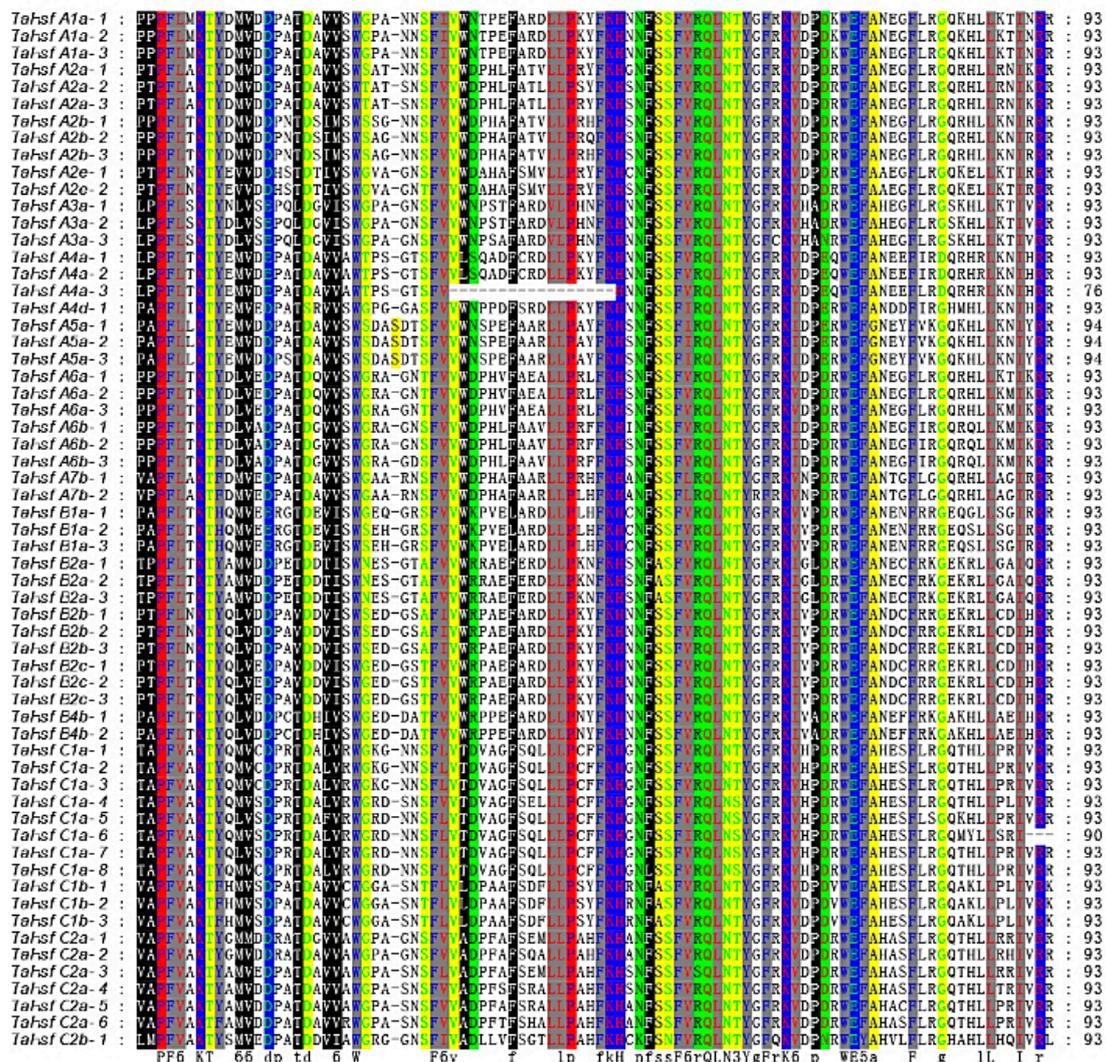
Multiple sequence alignments were performed to further analyze the degree of conservation in the DBD and OD domains in the *TaHsfs*. The results showed that the DBD amino acid lengths in the *TaHsfs* were 76–94 aa (Figure 4A). In particular, *TaHsfA4a-3* (76 aa) had a 17-aa deletion compared with most of the *TaHsfs* (93aa). According to the OD structure, the classification of the gene family was consistent with the results obtained by phylogenetic analysis, with 21 and 7 aa insertions between HR-A and HR-B in the class A and class C *TaHsfs*, respectively, and no insertions between HR-A and HR-B in the class B *TaHsfs* (Figure 4B). The positional information for the conserved DBD, OD, NLS, NES, and AHA domains showed that the same conserved domains in the same sub-class are almost identical, e.g., each conserved domain has exactly the same position and sequence among the three members of the A5 subclass (DBD: 17–110, OD: 138–188, NLS: 210-DALHKKRRLSGLDY, and AHA: 414-DNFWEQFLTE) (Table S6). Figure 5 shows that the number of introns is small in the *TaHsfs*, ranging from 0 to 4, but the lengths of the introns vary greatly. Motif prediction identified 15 conserved motifs with high confidence ranging between sizes of 11–50 aa in the *TaHsf* family, where motif 1 was the longest (50 aa) and motif 6 was the shortest (Table S7). Motifs 1, 2, 5, 6, 12, and 14 correspond to the DBD domain, and their highly conserved structures make them highly compatible with HSEs to ensure their accurate localization and the precise regulation of downstream gene expression [7]. Motifs 3 and 4 correspond to the OD domain, and motif 10 corresponds to the AHA motif. Structural analysis demonstrates that the *TaHsfs* in the C2 subclass belong to the wheat HSF gene family because they have the unique sequence characteristics of the C class but also the same conserved motifs as other members.

#### 2.5. Interaction Network Analysis

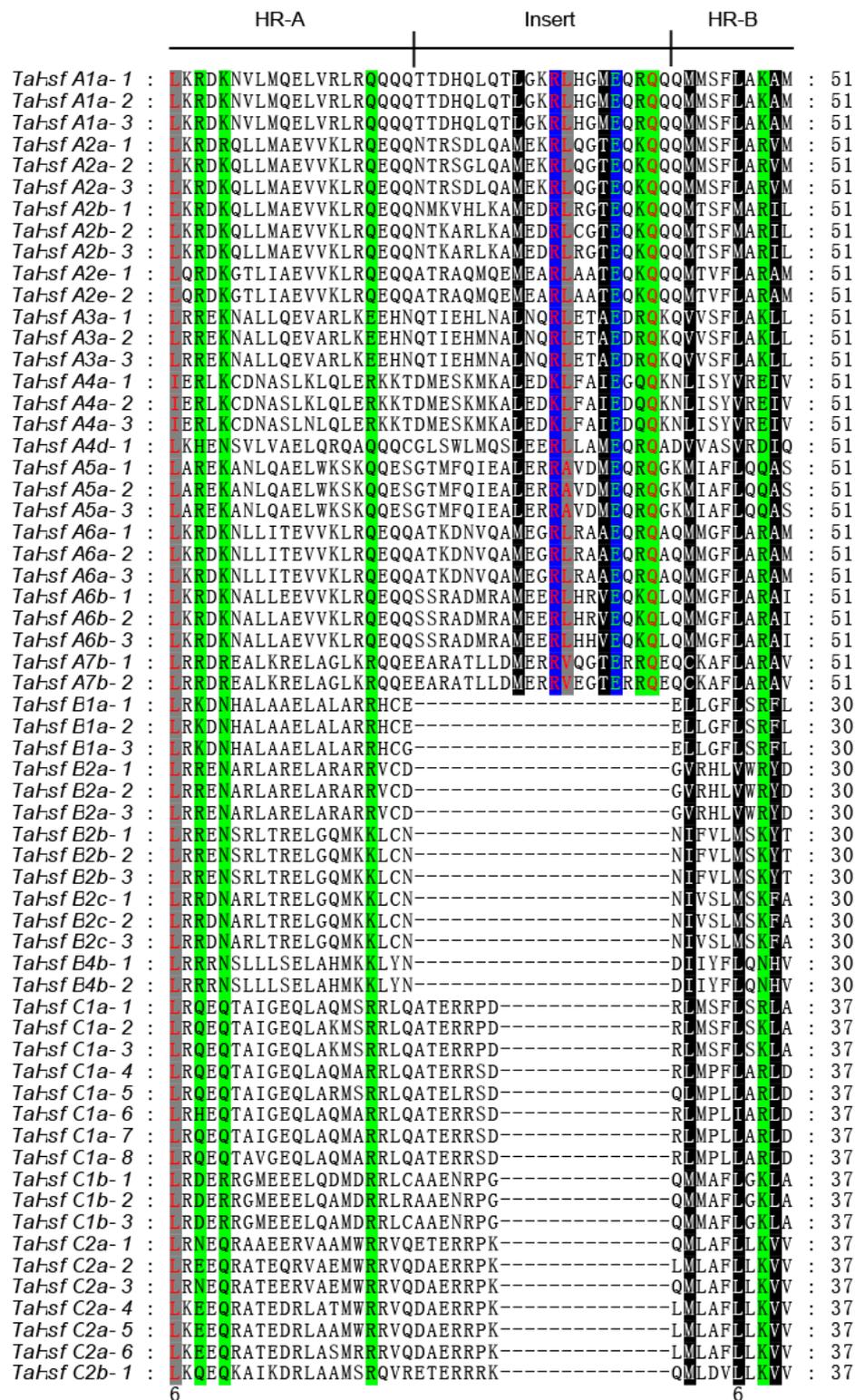
Interaction network analysis can help to understand the biological functions and molecular mechanisms of proteins. Therefore, we predicted the proteins that interact with wheat HSFs using STRING ([https://version-10-5.string-db.org/cgi/input.pl?sessionId=bHwqHDUnAsoF&input\\_page\\_show\\_search=on](https://version-10-5.string-db.org/cgi/input.pl?sessionId=bHwqHDUnAsoF&input_page_show_search=on)). The results showed that there were many interactions between *TaHsfs*, and proteins encoded by many other genes not belonging to this family also interacted with *TaHsfs*. HSP70s, HSP90s, and another class of chaperone protein called high-temperature protein G (htpG) were identified as the three major types of proteins that interact with wheat HSFs (Figure 6). In particular, HSP70 is an interesting and highly conserved protein, and the insertion of an HSP70 antisense gene fragment in sorghum and rice has been confirmed to lead to pollen abortion and male sterility [45,46].

### 2.6. Expression Profiles of TaHsfs

In order to fully understand the expression patterns of the *TaHsfs* in different tissues and during different growth periods, we analyzed RNA-seq data from the roots, stems, and leaves in the seedling, vegetative growth, and reproductive growth stages, as well as spikes from the reproductive growth stages, and the mature grain in Chinese Spring wheat (Wheat Expression Browser, <http://www.wheat-expression.com/>, developmental time-course of Chinese Spring). Except for the similar expression patterns of *TaHsfs* in the roots during reproductive growth and vegetative growth, the expression levels of *TaHsfs* differed in the other periods and tissues (Figure 7A, Table S8). Two specific *TaHsfs* (*C1a-4*, *C1a-8*) were expressed in the stems and leaves in the reproductive stage, and they were not expressed in the other tissues and periods. In particular, seven *TaHsFs* (*A2a-2*, *C1b-1*, *A2a-3*, *C1a-6*, *C2b-1*, *C1b-3*, *C1b-2*) were up-regulated in the spikes.

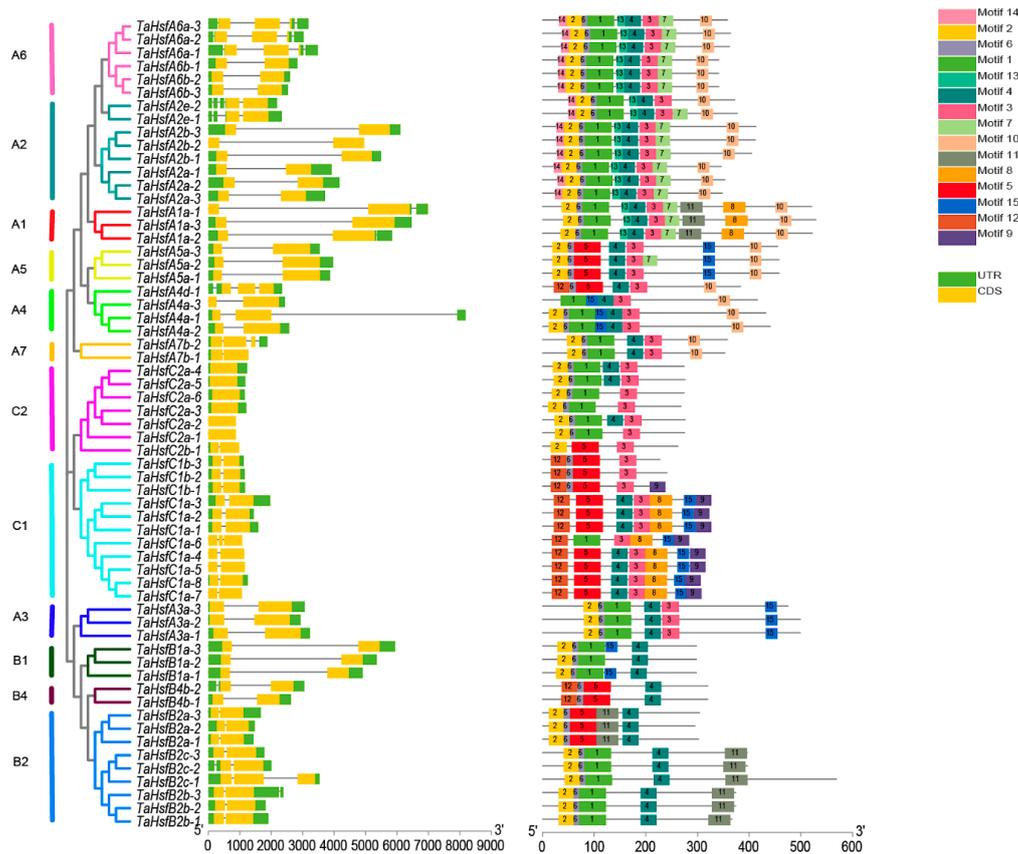


(a)  
Figure 4. Cont.



(b)

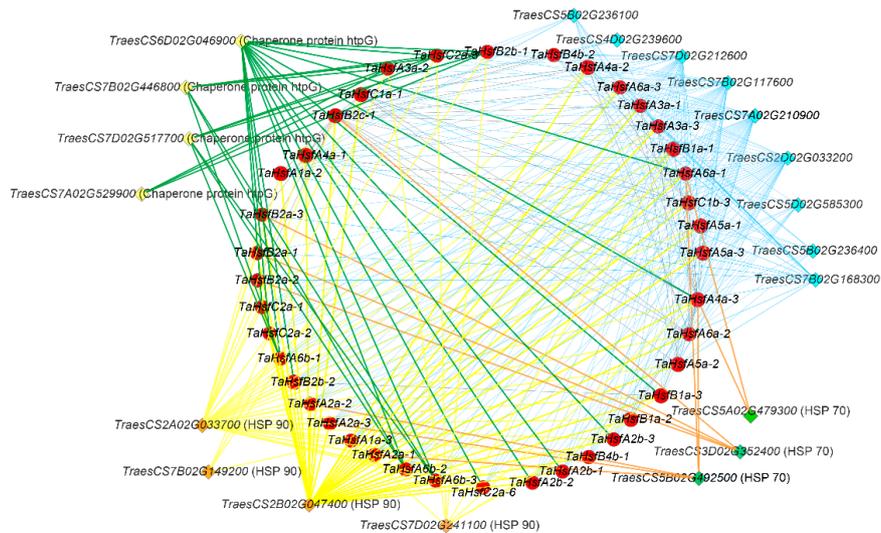
**Figure 4.** Analysis of the conserved DNA-binding domain (DBD) domains and oligomeric domain (OD) regions in the 61 *TaHsfs*. (a) Multiple sequence alignment of the DBD domains in 61 *TaHsfs*. (b) Multiple sequence alignment of the heptad repeat A/B (HR-A/B) regions in 61 *TaHsfs*. The scheme at the top shows the locations and boundaries of the HR-A core, insert, and HR-B regions within the HR-A/B regions.



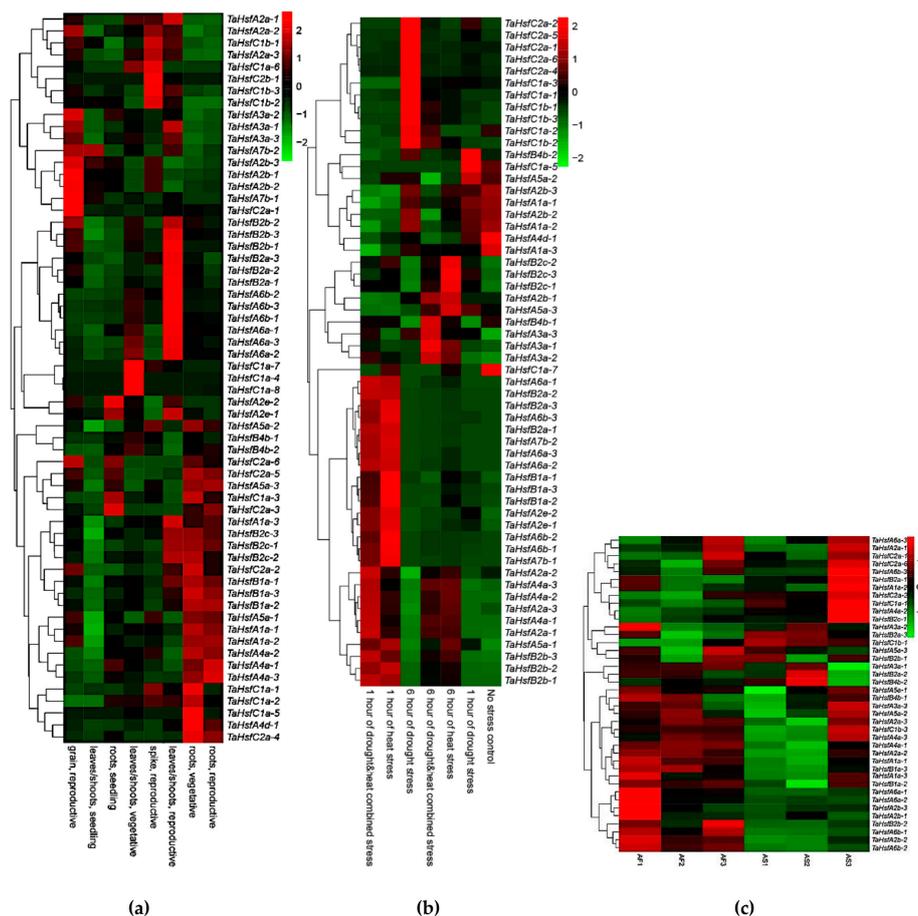
**Figure 5.** Phylogenetic tree, gene structure, and architecture of conserved protein motifs in *TaHsfs*. Details of the clusters in the phylogenetic tree are shown in different colors. The left panel represents the intron–exon structure of the *TaHsfs*, and the right panel shows the motif compositions of the *TaHsfs*. Different motifs are indicated by different colors and they are numbered from 1–15.

In addition, we analyzed the expression levels of *TaHsfs* under different physiological conditions using RNA-seq data from the expVIP database (Wheat Expression Browser, <http://www.wheat-expression.com/genes/heatmap>, drought and heat stress time-course in seedlings). The heatmap showed that compared with almost all *TaHsfs* of the no stress control, almost all *TaHsfs* did not exhibit differences in their expression levels in response to drought stress for 1 h, but 10 members of the C1a, C2a, and C1b sub-classes of *TaHsfs* were significantly up-regulated under drought stress for 6 h. Under heat stress for 1 h and the combination of heat stress and drought stress for 1 h, the expression levels of the *TaHsfs* were similar, where the main up-regulated *TaHsfs* belonged to classes A and B. *TaHsfs* from B2c and A3a were up-regulated under heat stress for 6 h and the combination of heat stress and drought stress for 6 h, respectively (Figure 7B, Table S9). Our analysis showed that class A and B *TaHsfs* were significantly up-regulated under heat stress, whereas class C *TaHsfs* were significantly up-regulated under drought stress. Subclass B1, B2, and B4 *Hsfs* were induced by the heat shock reaction, and they may act as synergistic factors that allow A class *Hsfs* to resist heat more effectively [30,47,48].

To determine the responses of *TaHsfs* in thermo-sensitive male sterile wheat to different fertility conditions, we analyzed the RNA-seq data for 40 *TaHsfs* under two fertility conditions in three different periods (uninucleate, binucleate, and trinucleate stages). We designated the sterile and fertile anthers in these three periods as AS1–3 and AF1–3, respectively. The results showed that compared with the sterile conditions, eight *TaHsfs* from A6a (-1, -2), A2b (-1, -2, -3), A6b (-1, -2), and B2b (-2) were always up-regulated in AF. However, no *TaHsfs* in subclass C were up-regulated under fertile conditions during these three stages (Figure 7C, Table S10).



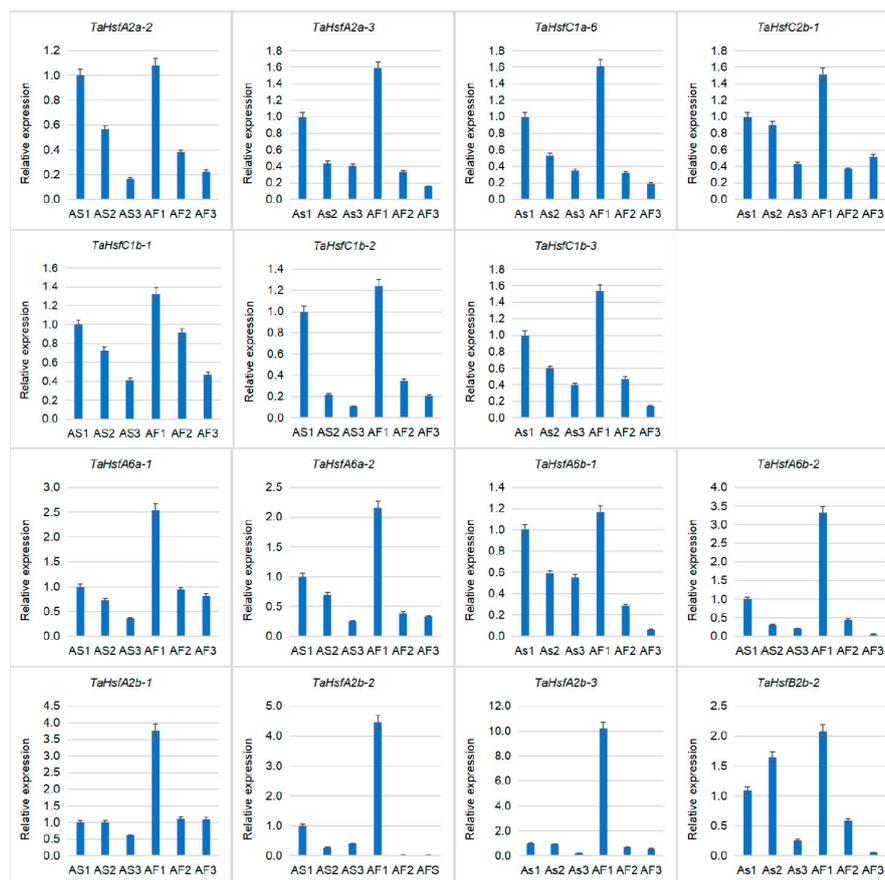
**Figure 6.** Interaction network between *TaHsfs* and related genes. The red circles represent *TaHsfs* and the diamonds in other colors represent the genes that interact with them. Lines connecting two genes show that an interaction exists between them.



**Figure 7.** Expression profiles for *TaHsfs* in different conditions. (a) Heat map of *TaHsfs* in nine different tissues and growth stages under normal conditions. (b) Heat map of *TaHsfs* under five different abiotic stress conditions and in a normal control. (c) Heat map of *TaHsfs* in the thermosensitive male sterile wheat KTM3315A at two different temperatures. Heatmap obtained using the count values from RNA-seq analysis.

### 2.7. Validation of *TaHsfs* Involved in Anther Protection Mechanisms

Based on the RNA-seq data, we identified seven up-regulated genes in the spikes of Chinese Spring and eight up-regulated genes in the fertile anthers of KTM3315A. In order to identify differentially expressed *TaHsfs* in fertile and sterile anthers, we performed quantitative real-time PCR (qRT-PCR) for the 15 candidate *TaHsfs* in three stages of anther development. The results showed that the expression levels of all 15 *TaHsfs* were highest in the uninucleate stage and they gradually decreased in the following two stages (binucleate stage and trinucleate stage) (Figure 8). There was no quantitative difference in the expression levels of the seven up-regulated *TaHsfs* in the Chinese Spring spikes during the three stages for AS and AF. However, the qRT-PCR results for the other eight *TaHsfs* were consistent with the RNA-seq results, where they were up-regulated in the uninucleate stage for AF compared with AS, and three *TaHsfA2bs* (-1, -2, -3) were significantly up-regulated. The regulation of *HsfA2* genes is known to be involved with redox homeostasis and metabolic, which are critical pathways for normal anther development [49]. In addition, the expression levels of *HsfA2s* are high in the early stages of anther and pollen development, and their important roles in the regulation of pollen protection genes have been confirmed in tomatoes [27,40,50]. Therefore, the three *TaHsfsA2bs* with high level expression levels in the fertile anthers may be candidate genes involved in anther protection during fertility conversion in the thermo-sensitive male sterile wheat line KTM3315A.



**Figure 8.** qRT-PCR results for *TaHsfs*. up-regulated *TaHsfs* in the spikes of Chinese spring wheat according to RNA-seq: *TaHsfA2a-2*, *TaHsfA2a-3*, *TaHsfC1a-6*, *TaHsfC1b-1*, *TaHsfC1b-2*, *TaHsfC1b-3*, and *TaHsfC2b-1*. Up-regulated *TaHsfs* in the fertile anthers of KTM3315A according to RNA-seq: *TaHsfA6a-1*, *TaHsfA6a-2*, *TaHsfA2b-1*, *TaHsfA2b-2*, *TaHsfA2b-3*, *TaHsfA6b-1*, *TaHsfA6b-2*, and *TaHsfB2b-2*. The x-axis represents six materials comprising sterile anthers and fertile anthers from the uninucleate, binucleate, and trinucleate stages. The y-axis represents the relative expression levels of *TaHsfs*. All data are the means  $\pm$  SE of three independent experiments.

### 3. Discussion

Wheat is an important food crop throughout the world, and thus increasing its stress resistance, yield, and quality are of great economic importance. The release of the wheat genome sequence has provided important information to facilitate the identification of excellent agronomic traits and stress resistance genes in wheat at the genome level, thereby contributing to develop high quality and stress-resistant wheat varieties.

In plants, HSFs are involved in various molecular processes and they have essential roles in the responses to abiotic stresses. Previous studies identified 21, 25, 25, 25, 25, and 22 *Hsf* genes in *Arabidopsis*, rice, maize, sorghum, tomato, and chickpea, respectively [13–16,51]. The number of HSF family members appears to be unrelated to the genome size, e.g., 35 *Hsfs* were found in Chinese cabbage (485 Mb) and 137 *Hsfs* in pear (527 Mb) [21,52]. Previous studies identified 56 and 82 members of the wheat HSF family [32,53], but we identified 61 *TaHsfs* in this study. To understand these differences in the numbers of wheat HSF members identified, we blasted the protein sequences identified by Xue et al. [32] and Duan et al. [53] against the latest wheat reference genome (IWGSC V1.1) [43]. Using the latest version gene IDs as standards, 77 genes were identified in all three studies, where four unique genes were identified in the present study, whereas six unique genes were identified by Duan et al. [53] and one by Xue et al. [32] (Figure S3, Table S11). The 16 protein sequences not identified as HSFs in the present study were analyzed. Although the DBD domains were detected, no high threshold coiled-coil structure was found by using MARCOIL software (<https://bcf.isb-sib.ch/webmarcoil/webmarcoilC1.html>) [54]. These findings matched with the identification criteria employed in the present study, so they were excluded. In total, 41 common HSFs were identified in all three studies, but the classifications and names of some of these proteins differed in the studies by Xue et al. [32] and Duan et al. [53]. For example, *TraesCS1A02G375600*, *TraesCS1D02G382900*, and *TraesCS1B02G396000* were renamed as *TaHsfA6a-1*, *TaHsfA6a-2*, and *TaHsfA6a-3* in the present study because they belong to a homeologous group and are orthologous to rice *Os10g28340* (A6a). However, Duan et al. [53] classified these genes in classes A4 and C1, and *TraesCS1B02G396000* was not even identified in the study of Xue et al. [32]. The main reasons for the differences in the three studies are as follows. First, the wheat genome sequence was not completed when these previous studies were conducted, and thus the assembly of incomplete and incorrect genes resulted in the failure to find all of the HSF gene family members. Second, the different threshold settings and strategies employed as identification criteria resulted in differences in the number of wheat HSF family members identified.

In the present study, in addition to employing two approaches described in the method section for identifying *TaHsfs*, we used the wheat HSF sequence in the plantTFDB database (<http://planttfdb.cbi.pku.edu.cn/family.php?sp=Tae&fam=HSF>) and NCBI (<https://www.ncbi.nlm.nih.gov/>) as query sequences to search for all of the wheat protein sequences. However, after verification based on Pfam (<http://pfam.xfam.org/>) [55], SMART (<http://smart.embl.de/>), Conserved Domain Database (CDD) (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) [56], MARCOIL, and HEASTER (<https://applbio.biologie.uni-frankfurt.de/hsf/heatster/analyse.php#menu>) [9], the final results were exactly the same as those obtained using the two approaches employed in the present study. Thus, the method employed for constructing wheat-specific HSF HMM files with HSF HMM files (PF00447) in the present study was highly effective and reasonable, and the combination of multiple approaches avoided the failure to identify *TaHsfs*. The gene family member identification strategy presented in this study will greatly facilitate the identification of HSF family in other species.

A gene family is a group of genes derived from the same ancestor, where two or more copies of the gene may be present. Whole genome duplication is a major driver of gene family expansion to increase the number of gene family members in a single species [57]. Genes with products that have close interactions are usually retained after genome-wide duplication, such as genes involved in signaling pathways, regulatory networks, and the formation of macromolecular complexes. These genes are called dose-sensitive genes, and they include transcription factors and ribonucleoprotein family members [57,58]. If these genes are lost after genome-wide duplication, the metabolic stability

of the organism can be disrupted with fatal consequences [59,60]. Segmental and tandem duplications are the other two main causes of the expansion of gene families in plants [61]. Tandem duplication is closely related to the amplification of biotic and abiotic stress-related genes [62,63]. Tandem duplication events only amplify a class of genes, so any redundant dose-sensitive genes produced by these duplication events will disrupt the balance of the related biological pathways. Therefore, these types of duplication events tend to amplify genes at the apical or terminal ends of metabolic pathways as well as dose-insensitive genes [61]. *TaHsfs* are typical dose-sensitive genes because they regulate the expression of HSPs by encoding transcription factors. Thus, during the long process of evolution, wheat has retained 14 groups of complete homeologous triplet HSF gene family members in the ABD sub-genomes, and only a very small number of genes were derived by tandem duplication. This balanced state maintains the metabolic stability of the organism and it is also required for the fine regulation of mechanisms in plants.

The evolution of gene families is often accompanied by increases or losses of exons, which have played key roles in the evolution of gene families [64]. Thus, we analyzed the number and distribution of exons and introns in each member of the *TaHsf* family and found that the 61 *TaHsfs* contained 1–3 exons and 0–4 introns. The lengths and positions of the *TaHsf* gene introns in the same subclass were relatively well conserved, but the introns varied greatly among the different *TaHsf* subclasses. The presence of introns allows a variety of splicing modes and the same DNA sequence can produce different protein products via alternative splicing of the gene after transcription [65]. This is one explanation for the evolution of the functional diversity of the wheat HSF family and the different responses of *TaHsfs* to temperature [66].

In plants, transcription factor HSFs are expressed in male reproductive cells during heat stress. In particular, *AtHsfA2* (AT2G26150) and *AtHsfA5* (AT4G13980) were identified as critical during pollen reproduction in *Arabidopsis* [67,68]. Five class A *Hsfs* and three class B *Hsfs* are up-regulated in rice spikes [40]. In addition, increased levels of *HsfA2* and *HsfA3* were detected during tomato pollen development, and other *Hsfs* were also studied in the anthers [27,41,50]. The molecular roles of *Hsfs* in anther development are not yet known but their downstream regulated HSPs are important molecular chaperones involved in the development of male gametophytes in plants [69–71]. Chen et al. showed that cytoplasmic male sterility in sorghum was caused by insufficient mitochondria. When the sterile sorghum line 3A was heated for 4 h, the nuclear-encoded protein HSP70 was expressed in the mitochondria, which promoted an increase in the mitochondria content from 7 mg/g (mitochondria weight/fresh ear weight) to 19 mg/g, and the sterile plants were transformed into fertile plants. However, the mitochondria content was more than 19 mg/g in the maintainer line 3B and it was fertile regardless of whether it was subjected to heat treatment. Thus, HSP70 could regulate fertility by increasing the content of mitochondria to meet the demand for large amounts of energy during pollen development. According to this hypothesis, a rice *Hsp70* antisense expression vector driven by the anther-specific promoter *Osg6B* was constructed to inhibit *Hsp70* expression during pollen development and block male gamete development, thereby resulting in male sterility [45]. Frank et al. [41] and Giorno et al. [27] demonstrated the high expression levels of *HsfA2* at 42 °C and 36 °C in tomato microspores, respectively, and we demonstrated the high expression level of *TaHsfA2* in fertile anthers cultured at 22 °C. These studies were based on heat treatment, but we hypothesize that increased temperature can induce the upregulation of HSFs to regulate the expression of downstream HSPs in the cytoplasm and increase the number of mitochondria to ensure the energy supply is adequate for pollen development. Thus, the findings obtained in the present study indicate likely roles of wheat HSF family members in anther development, and provide a new theoretical basis for exploring the mechanism responsible for male sterility in wheat.

## 4. Materials and Methods

### 4.1. Plant Materials

KTM3315A is a type of thermo-sensitive male sterile wheat with various advantages and it was selected by our research group in 2001. It is completely male sterile during the normal wheat-growing season, whereas its fertility can be restored in a high-temperature environment [37]. Ten pots of KTM3315A were placed in two artificial climate incubators until the flowering stage and they were treated at these conditions: 14 h light (day) and 10 h dark (night) and two different temperatures (day/night temperatures of 17 °C/15 °C for the sterile conditions and 22 °C/20 °C for the fertile conditions). RNA was extracted from the anthers of sterile (AS1, AS2, and AS3) and fertile (AF1, AF2, and AF3) plants during the uninucleate, binucleate, and trinucleate stages, and six samples were used for RNA-seq and qRT-PCR.

### 4.2. Genome-Wide Identification of HSF Family Members in Wheat

The latest wheat reference genome and the HMM profile (PF00447) of the HSF family were downloaded from the wheat genome database IWGSC (IWGSC RefSeq annotation v1.1, <https://wheat-urgi.versailles.inra.fr/Seq-Repository/Annotations>) [44] and Pfam database (<http://pfam.xfam.org/family/PF00447>) [72]. The HSF HMM profile was searched in the wheat genome with the hmmsearch program in HMMER software [73] and reliable results were screened based on an E-value less than or equal to  $1 \times 10^{-20}$ . The results obtained were used to construct a wheat-specific HSF HMM profile by hmmbuild program, before searching the wheat reference genome again with an E-value threshold of  $1 \times 10^{-3}$ . In order to avoid missing other HSF members, the HSF protein sequences for rice and *Arabidopsis* were downloaded from phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>) and used as queries to BLAST the wheat reference genome, where the E-value threshold was  $1 \times 10^{-10}$ . The results obtained were identified as wheat candidate genes and verified by online software CD-search, SMART, and pfam for presence of protein domain. Coiled-coil structures were detected using the MARCOIL program. Protein sequences that lacked the DBD domain or coiled-coil structure were removed. All high confidence level HSF protein sequences were submitted to Expasy ([https://www.expasy.org/structural\\_bioinformatics](https://www.expasy.org/structural_bioinformatics)) to calculate the number of amino acids, molecular weight, and theoretical isoelectric point. NetNES (<http://www.cbs.dtu.dk/services/NetNES/>), cNLS Mapper ([http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS\\_Mapper\\_form.cgi#opennewwindow](http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi#opennewwindow)), and HEATSTER were used to search for the NESs and NLSs in the TaHsfs [9,74,75]. The gene annotation information for the TaHsfs was submitted to the online site MapGene2 Chrom web V2 ([http://mg2c.iask.in/mg2c\\_v2.0/](http://mg2c.iask.in/mg2c_v2.0/)) to map their chromosomal locations.

### 4.3. Multiple Sequence Alignment, Phylogenetic Analysis, and Classification of TaHsfs

The full-length amino acid sequences of the HSFs in rice and *Arabidopsis* were used for evolutionary analysis with the full-length amino acid sequences of the TaHsfs [13]. All of the sequences were imported into MEGA7 for sequence comparison using the default parameters, and then a phylogenetic tree was constructed using the conserved DBD and OD domains and the connection between them. The phylogenetic tree was constructed using the neighbor-joining method and verified based on the maximum likelihood method, where the parameters were set to the Poisson distribution mode and the bootstrap tests were conducted with 1000 replicates. The classifications of HSFs in wheat were based on the topology of the phylogenetic tree and the classification of HSFs in the other two plants.

### 4.4. Analysis of Gene Duplication and Synteny Relationships of TaHsfs

Segmental and tandem duplications were determined for the TaHsfs using McscanX software [76]. Sequences on the same chromosome with a shared similarity greater than 75% and fragment lengths in the alignment that exceeded 75% of the length of the longer sequence were confirmed as tandem duplications. The KA/KS values were calculated for the tandem duplications using

KAKS\_calculator 2.0 [77]. A segmental duplication map of the genome was prepared using Circos software [78]. In addition to rice and *Arabidopsis*, the genome-wide CDS sequences and genome annotation information for Chinese cabbage ([http://brassicadb.org/brad/datasets/pub/Genomes/Brassica\\_rapa/V1.0/Scaffold1.0/](http://brassicadb.org/brad/datasets/pub/Genomes/Brassica_rapa/V1.0/Scaffold1.0/)), sorghum (<http://www.plantgdb.org/BdGDB/>), maize ([http://interface.maizegdb.org/gbrowse/maize\\_v4](http://interface.maizegdb.org/gbrowse/maize_v4)), and *Brachypodium* ([https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\\_BdistachyonBd21\\_3\\_er](https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_BdistachyonBd21_3_er)) were obtained, and the synteny relationships between the HSF family member in wheat and these species were determined using Mcscanx.

#### 4.5. Structural Characterization, Conserved Motifs, and Construction of the Interaction Network for TaHsfs

ClustalW was used to align the conserved DBD and HR-A/B domains in the TaHsfs. The gene structures of the TaHsfs were visualized using the online drawing tool Gene Structure Display Server 2.0 (GSDS, <http://gsds.cbi.pku.edu.cn/>) [79]. The MEME software [80] obtained 15 conserved motifs in the TaHsfs with lengths of 6–50 aa. The gene structure and motifs in the TaHsfs were visualized using TB tools [81]. STRING was employed to obtain the regulatory networks between TaHsfs and their interacting genes, which were visualized with Cytoscape (<https://cytoscape.org/>) [82–84].

#### 4.6. RNA-seq Data Analysis

Part of the wheat transcriptome data was obtained from the Wheat Expression Browser powered by expVIP (<http://www.wheat-expression.com/>) to study the expression patterns of the TaHsfs. The RNA-seq data were derived from nine different growth periods and tissues in Chinese Spring under normal conditions (leaves and shoots in seedling stage, roots in seedling stage, leaves and shoots in vegetative stage, roots in vegetative stage, leaves and shoots in reproductive stage, roots in reproductive stage, spike in reproductive stage, and grain in reproductive stage) [85], as well as under six types of stress and a normal control in the heat and drought tolerant wheat variety TAM107 (heat stress for 1 h, heat stress for 6 h, drought stress for 1 h, drought stress for 6 h, combined drought and heat stress for 1 h, combined drought and heat stress for 6 h, and no stress control) [86]. In addition, in order to study the mechanism responsible for regulating TaHsfs in thermo-sensitive male sterile wheat, transcriptome data were analyzed from temperature-sensitive materials under different fertility conditions (AF1, AF2, AF3, AS1, AS2, and AS3) [37]. The heat maps were drawn using the heatmap application in omicshare (<http://www.omicshare.com/tools/index.php/Home/help/index/html/combinemore>).

#### 4.7. qRT-PCR Verification

Total RNA extraction and reverse transcription of cDNA were performed using TRIGene and StarScript II First-strand cDNA Synthesis Mix (Genestar), respectively, according to the manufacturers' instructions. qRT-PCR was conducted with an Applied Biosystems 7500 Real-Time PCR System and 2×RealStar Green Fast Mixture with ROX II according to the operating instructions and program default settings with three replicates per group. ROX II was added to correct fluorescence signal errors between wells during qRT-PCR reactions. The wheat Actin gene was used as an internal reference, the final relative quantitative results were calculated using the  $2^{-\Delta\Delta C_t}$  method [87]. At least three replicates of each cDNA sample were used to perform the qRT-PCR of each gene. Primer 5.0 was used to design the specific primers according to the TaHsfs sequences (Table S12).

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/21/2/608/s1>.

**Author Contributions:** Conceptualization, J.Y.; data curation, J.Y. and Q.L.; formal analysis, J.Y. and X.Y.; funding acquisition, X.S.; investigation, J.Y.; methodology, W.L.; project administration, X.S.; resources, J.Y.; software, J.Y.; supervision, L.Z.; validation, J.Y. and G.H.; visualization, X.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (31771874) and the Program in Science and Technology of Yangling State Demonstration Zone of Agricultural High-tech Industries (2018NY-19).

**Acknowledgments:** Special thanks to Ingo Ebersberger (Senckenberg Biodiversität und Klima Forschungszentrum) for providing professional guidance and advice, and English-speaking scientific editor Duncan E. Jackson for professional language editing services.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

HSF	Heat shock transcription factor
<i>Hsf</i>	HSF gene
<i>TaHsf</i>	Wheat HSF gene
DBD	DNA binding domain
OD	Oligomeric domain

## References

- Kumar, S.V.; Wigge, P.A. H2A.Z-Containing nucleosomes mediate the thermosensory response in Arabidopsis. *Cell* **2010**, *140*, 136–147. [[CrossRef](#)]
- Sung, D.; Kaplan, F.; Lee, K.; Guy, C.L. Acquired tolerance to temperature extremes. *Trends Plant Sci.* **2003**, *8*, 179–187. [[CrossRef](#)]
- Parsell, D.A.; Lindquist, S. The function of heat-shock proteins in stress tolerance: Degradation and reactivation of damaged proteins. *Annu. Rev. Genet.* **1993**, *27*, 437–496. [[CrossRef](#)] [[PubMed](#)]
- Hartl, F.U.; Hayerhartl, M. Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science* **2002**, *295*, 1852–1858. [[CrossRef](#)] [[PubMed](#)]
- Nover, L.; Bharti, K.; Döring, P.; Mishra, S.K.; Ganguli, A.; Scharf, K. Arabidopsis and the heat stress transcription factor world: How many heat stress transcription factors do we need? *Cell Stress Chaperones* **2001**, *6*, 177–189. [[CrossRef](#)]
- Damberger, F.F.; Pelton, J.G.; Harrison, C.J.; Nelson, H.C.M.; Wemmer, D.E. Solution structure of the DNA-binding domain of the heat shock transcription factor determined by multidimensional heteronuclear magnetic resonance spectroscopy. *Protein Sci.* **1994**, *3*, 1806–1821. [[CrossRef](#)]
- Cicero, M.P.; Hubl, S.T.; Harrison, C.J.; Littlefield, O.; Hardy, J.A.; Nelson, H.C. The wing in yeast heat shock transcription factor (HSF) DNA-binding domain is required for full activity. *Nucleic Acids Res.* **2001**, *29*, 1715–1723. [[CrossRef](#)]
- Yura, T.; Nakahigashi, K. Regulation of the heat-shock response. *Curr. Opin. Microbiol.* **1999**, *2*, 153–158. [[CrossRef](#)]
- Scharf, K.; Berberich, T.; Ebersberger, I.; Nover, L. The plant heat stress transcription factor (Hsf) family: Structure, function and evolution. *Biochim. Biophys. Acta* **2012**, *1819*, 104–119. [[CrossRef](#)]
- Heerklotz, D.; Döring, P.; Bonzelius, F.; Winkelhaus, S.; Nover, L. The balance of nuclear import and export determines the intracellular distribution and function of tomato heat stress transcription factor HsfA2. *Mol. Cell. Biol.* **2001**, *21*, 1759–1768. [[CrossRef](#)]
- Kotak, S.; Port, M.; Ganguli, A.; Bicker, F.; Von Koskull, P. 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.* **2004**, *39*, 98–112. [[CrossRef](#)] [[PubMed](#)]
- Mishra, S.K.; Tripp, J.; Winkelhaus, S.; Tschiersch, B.; Theres, K.; Nover, L.; Scharf, K. In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato. *Genes Dev.* **2002**, *16*, 1555–1567. [[CrossRef](#)] [[PubMed](#)]
- Guo, J.; Wu, J.; Ji, Q.; Wang, C.; Luo, L.; Yuan, Y.; Wang, Y.; Wang, J. Genome-wide analysis of heat shock transcription factor families in rice and Arabidopsis. *J. Genet. Genom.* **2008**, *35*, 105–118. [[CrossRef](#)]
- Lin, Y.X.; Jiang, H.Y.; Chu, Z.X.; Tang, X.L.; Zhu, S.W.; Cheng, B.J. Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. *BMC Genom.* **2011**, *12*, 76. [[CrossRef](#)]
- Li, P.S.; Yu, T.F.; He, G.H.; Chen, M.; Zhou, Y.B.; Chai, S.C.; Xu, Z.S.; Ma, Y.Z. Genome-wide analysis of the Hsf family in soybean and functional identification of GmHsf-34 involvement in drought and heat stresses. *BMC Genom.* **2014**, *15*, 1009. [[CrossRef](#)]

16. Chidambaranathan, P.; Jagannadham, P.T.K.; Satheesh, V.; Kohli, D.; Basavarajappa, S.H.; Chellapilla, B.; Kumar, J.; Jain, P.K.; Srinivasan, R. Genome-wide analysis identifies chickpea (*Cicer arietinum*) heat stress transcription factors (Hsfs) responsive to heat stress at the pod development stage. *J. Plant Res.* **2018**, *131*, 525–542. [[CrossRef](#)]
17. Nagaraju, M.; Reddy, P.S.; Kumar, S.A.; Srivastava, R.K.; Kishor, P.B.K.; Rao, D.M. Genome-wide Scanning and Characterization of Sorghum bicolor L. Heat shock transcription factors. *Curr. Genom.* **2015**, *16*, 279–291. [[CrossRef](#)]
18. Ma, J.; Xu, Z.S.; Wang, F.; Tan, G.F.; Li, M.Y.; Xiong, A.S. Genome-wide analysis of HSF family transcription factors and their responses to abiotic stresses in two Chinese cabbage varieties. *Acta Physiol. Plant.* **2014**, *36*, 513–523. [[CrossRef](#)]
19. Guo, M.; Lu, J.P.; Zhai, Y.F.; Chai, W.G.; Gong, Z.H.; Lu, M.H. Genome-wide analysis, expression profile of heat shock factor gene family (CaHsfs) and characterisation of CaHsfA2 in pepper (*Capsicum annuum* L.). *BMC Plant Biol.* **2015**, *15*, 151. [[CrossRef](#)]
20. Giorno, F.; Guerriero, G.; Baric, S.; Mariani, C. Heat shock transcriptional factors in *Malus domestica*: Identification, classification and expression analysis. *BMC Genom.* **2012**, *13*, 639. [[CrossRef](#)]
21. Qiao, X.; Li, M.; Li, L.; Yin, H.; Wu, J.; Zhang, S. Genome-wide identification and comparative analysis of the heat shock transcription factor family in Chinese white pear (*Pyrus bretschneideri*) and five other Rosaceae species. *BMC Plant Biol.* **2015**, *15*, 12. [[CrossRef](#)] [[PubMed](#)]
22. Lohmann, C.; Eggersschumacher, G.; Wunderlich, M.; Schöffl, F. Two different heat shock transcription factors regulate immediate early expression of stress genes in *Arabidopsis*. *Mol. Genet. Genom.* **2004**, *271*, 11–21. [[CrossRef](#)] [[PubMed](#)]
23. Busch, W.; Wunderlich, M.; Schöffl, F. Identification of novel heat shock factor-dependent genes and biochemical pathways in *Arabidopsis thaliana*. *Plant J.* **2004**, *41*, 1–14. [[CrossRef](#)] [[PubMed](#)]
24. Ogawa, D.; Yamaguchi, K.; Nishiuchi, T. High-level overexpression of the *Arabidopsis* HsfA2 gene confers not only increased thermotolerance but also salt/osmotic stress tolerance and enhanced callus growth. *J. Exp. Bot.* **2007**, *58*, 3373–3383. [[CrossRef](#)]
25. Banti, V.; Mafessoni, F.; Loreti, E.; Alpi, A.; Perata, P. The heat-inducible transcription factor HsfA2 enhances anoxia tolerance in *Arabidopsis*. *Plant Physiol.* **2010**, *152*, 1471–1483. [[CrossRef](#)]
26. Zhang, L.; Li, Y.; Xing, D.; Gao, C. Characterization of mitochondrial dynamics and subcellular localization of ROS reveal that HsfA2 alleviates oxidative damage caused by heat stress in *Arabidopsis*. *J. Exp. Bot.* **2009**, *60*, 2073–2091. [[CrossRef](#)]
27. Giorno, F.; Woltersarts, M.; Grillo, S.; Scharf, K.; Vriezen, W.H.; Mariani, C. Developmental and heat stress-regulated expression of HsfA2 and small heat shock proteins in tomato anthers. *J. Exp. Bot.* **2010**, *61*, 453–462. [[CrossRef](#)]
28. Shim, D.; Hwang, J.; Lee, J.; Lee, S.; Choi, Y.; An, G.; Martinoia, E.; Lee, Y. Orthologs of the class A4 heat shock transcription factor HsfA4a confer cadmium tolerance in wheat and rice. *Plant Cell* **2009**, *21*, 4031–4043. [[CrossRef](#)]
29. Almoguera, C.; Prietodapena, P.; Diazmartin, J.; Espinosa, J.M.R.; Carranco, R.; Jordano, J. The HaDREB2 transcription factor enhances basal thermotolerance and longevity of seeds through functional interaction with HaHSFA9. *BMC Plant Biol.* **2009**, *9*, 75. [[CrossRef](#)]
30. Bharti, K.; Von Koskuldoring, P.; Bharti, S.; Kumar, P.; Tintschlkorbiter, A.; Treuter, E.; Nover, L. Tomato heat stress transcription factor HsfB1 represents a novel type of general transcription coactivator with a histone-like motif interacting with the plant CREB binding protein ortholog HAC1. *Plant Cell* **2004**, *16*, 1521–1535. [[CrossRef](#)]
31. Prasad, P.V.V.; Djanaguiraman, M. Response of floret fertility and individual grain weight of wheat to high temperature stress: Sensitive stages and thresholds for temperature and duration. *Funct. Plant Biol.* **2014**, *41*, 1261–1269. [[CrossRef](#)]
32. Xue, G.; Sadat, S.; Drenth, J.; McIntyre, C.L. The heat shock factor family from *Triticum aestivum* in response to heat and other major abiotic stresses and their role in regulation of heat shock protein genes. *J. Exp. Bot.* **2014**, *65*, 539–557. [[CrossRef](#)] [[PubMed](#)]
33. Zhang, S.; Xu, Z.; Li, P.; Yang, L.; Wei, Y.; Chen, M.; Li, L.; Zhang, G.; Ma, Y. Overexpression of TaHSF3 in transgenic *Arabidopsis* enhances tolerance to extreme temperatures. *Plant Mol. Biol. Rep.* **2013**, *31*, 688–697. [[CrossRef](#)]

34. Kumar, R.R.; Goswami, S.; Singh, K.; Dubey, K.; Rai, G.K.; Singh, B.; Singh, S.; Grover, M.; Mishra, D.; Kumar, S.; et al. Characterization of novel heat-responsive transcription factor (TaHSFA6e) gene involved in regulation of heat shock proteins (HSPs)—A key member of heat stress-tolerance network of wheat. *J. Biotechnol.* **2018**, *279*, 1–12. [[CrossRef](#)]
35. Hu, X.J.; Chen, D.; Lynne McIntyre, C.; Fernanda Dreccer, M.; Zhang, Z.B.; Drenth, J.; Kalaipandian, S.; Chang, H.; Xue, G.P. Heat shock factor C2a serves as a proactive mechanism for heat protection in developing grains in wheat via an ABA-mediated regulatory pathway. *Plant Cell Environ.* **2018**, *41*, 79–98. [[CrossRef](#)]
36. Meng, L.Y.; Liu, Z.H.; Zhang, L.L.; Hu, G.; Song, X.Y. Cytological characterization of a thermo-sensitive cytoplasmic male-sterile wheat line having K-type cytoplasm of *aegilops kotschyi*. *Breed Sci.* **2016**, *66*, 752–761. [[CrossRef](#)]
37. Ye, J.; Duan, Y.; Hu, G.; Geng, X.; Zhang, G.; Yan, P.; Liu, Z.; Zhang, L.; Song, X. Identification of candidate genes and biosynthesis pathways related to fertility conversion by wheat KTM3315A transcriptome profiling. *Front. Plant Sci.* **2017**, *8*, 449. [[CrossRef](#)]
38. Widlak, W.; Vydra, N. The role of heat shock factors in mammalian spermatogenesis. *Adv. Anat. Embryol. Cell Biol.* **2017**, *222*, 45–65. [[CrossRef](#)]
39. Shinka, T.; Sato, Y.; Chen, G.; Naroda, T.; Kinoshita, K.; Unemi, Y.; Tsuji, K.; Toida, K.; Iwamoto, T.; Nakahori, Y. Molecular characterization of heat shock-like factor encoded on the human Y chromosome, and implications for male infertility. *Biol. Reprod.* **2004**, *71*, 297–306. [[CrossRef](#)]
40. Zhang, X.; Li, J.; Liu, A.; Zou, J.; Zhou, X.; Xiang, J.; Rerksiri, W.; Peng, Y.; Xiong, X.; Chen, X. Expression profile in rice panicle: Insights into heat response mechanism at reproductive stage. *PLoS ONE* **2012**, *7*, e49652. [[CrossRef](#)]
41. Frank, G.; Pressman, E.; Ophir, R.; Althan, L.; Shaked, R.; Freedman, M.; Shen, S.; Firon, N. Transcriptional profiling of maturing tomato (*Solanum lycopersicum* L.) microspores reveals the involvement of heat shock proteins, ROS scavengers, hormones, and sugars in the heat stress response. *J. Exp. Bot.* **2009**, *60*, 3891–3908. [[CrossRef](#)] [[PubMed](#)]
42. Bennetzen, J.L.; Devos, K. Plant synteny, colinearity and genome evolution. *eLS* **2001**. [[CrossRef](#)]
43. Eckardt, N.A. Everything in its place: Conservation of gene order among distantly related plant species. *Plant Cell* **2001**, *13*, 723–725. [[CrossRef](#)] [[PubMed](#)]
44. Appels, R.; Eversole, K.; Stein, N.; Feuillet, C.; Keller, B.; Rogers, J.; Pozniak, C.J.; Choulet, F.; Distelfeld, A.; Poland, J.; et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, *361*, eaar7191. [[CrossRef](#)]
45. Liu, L.K.; Liu, G.Q.; Hou, N.; Liu, C.G.; Chen, J.N.; Zhang, W.H.; Fan, Y.L.; Wu, B.J. Construction and identification of HSP70 antisense RNA expression vector for genetic engineering male sterility in plant. *J. Anhui Agric. Sci.* **2008**, *36*, 6698–6700.
46. Yi, L.X.; Li, X.; Chen, J.N.; Xia, S.T. Creation of male sterile line in tobacco with HSP70 anti-sense fragment. *Agric. Sci. Technol.* **2016**, *17*, 2262–2266. [[CrossRef](#)]
47. Czarneckaverner, E.; Yuan, C.; Scharf, K.; Englich, G.; Gurley, W.B. Plants contain a novel multi-member class of heat shock factors without transcriptional activator potential. *Plant Mol. Biol.* **2000**, *43*, 459–471. [[CrossRef](#)]
48. Ikeda, M.; Mitsuda, N.; Ohmetakagi, M. Arabidopsis HsfB1 and HsfB2b act as repressors of the expression of heat-inducible Hsfs but positively regulate the acquired thermotolerance. *Plant Physiol.* **2011**, *157*, 1243–1254. [[CrossRef](#)]
49. Liu, H.C.; Charng, Y.Y. Common and distinct functions of Arabidopsis class A1 and A2 heat shock factors in diverse abiotic stress responses and development. *Plant Physiol.* **2013**, *163*, 276–290. [[CrossRef](#)]
50. Fragkostefanakis, S.; Mesihovic, A.; Simm, S.; Paupiere, M.J.; Hu, Y.; Paul, P.; Mishra, S.K.; Tschiersch, B.; Theres, K.; Bovy, A.; et al. HsfA2 controls the activity of developmentally and stress-regulated heat stress protection mechanisms in tomato male reproductive tissues. *Plant Physiol.* **2016**, *170*, 2461–2477. [[CrossRef](#)]
51. Chen, Y.B.; Wang, Y.; Zhang, H.L. Identification and expression analysis of heat shock factor (Hsf) gene family in tomato (*Solanum lycopersicum*). *J. Agric. Biotechnol.* **2015**, *23*, 492–501. [[CrossRef](#)]
52. Song, X.; Liu, G.; Duan, W.; Liu, T.; Huang, Z.; Ren, J.; Li, Y.; Hou, X. Genome-wide identification, classification and expression analysis of the heat shock transcription factor family in Chinese cabbage. *Mol. Genet. Genom.* **2014**, *289*, 541–551. [[CrossRef](#)] [[PubMed](#)]

53. Duan, S.; Liu, B.; Zhang, Y.; Li, G.; Guo, X. Genome-wide identification and abiotic stress-responsive pattern of heat shock transcription factor family in *Triticum aestivum* L. *BMC Genom.* **2019**, *20*, 257. [[CrossRef](#)] [[PubMed](#)]
54. Mauro, D.; Terry, S. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **2002**, *18*, 617–625. [[CrossRef](#)]
55. Letunic, I.; Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **2018**, *46*. [[CrossRef](#)]
56. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **2017**, *45*. [[CrossRef](#)]
57. Freeling, M. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **2009**, *60*, 433–453. [[CrossRef](#)]
58. Edger, P.P.; Pires, J.C. Gene and genome duplications: The impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res.* **2009**, *17*, 699–717. [[CrossRef](#)]
59. Birchler, J.A.; Bhadra, U.; Bhadra, M.P.; Auger, D.L. Dosage-dependent gene regulation in multicellular eukaryotes: Implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Dev. Biol.* **2001**, *234*, 275–288. [[CrossRef](#)]
60. Papp, B.; Pal, C.; Hurst, L.D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **2003**, *424*, 194–197. [[CrossRef](#)]
61. Cannon, S.B.; Mitra, A.; Baumgarten, A.; Young, N.D.; May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* **2004**, *4*, 10. [[CrossRef](#)] [[PubMed](#)]
62. Michelmore, R.W.; Meyers, B.C. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* **1998**, *8*, 1113–1130. [[CrossRef](#)] [[PubMed](#)]
63. Maere, S.; De Bodt, S.; Raes, J.; Casneuf, T.; Van Montagu, M.; Kuiper, M.; Van de Peer, Y. Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5454–5459. [[CrossRef](#)] [[PubMed](#)]
64. Rose, A.B. Intron-mediated regulation of gene expression. *Curr. Top. Microbiol. Immunol.* **2008**, *326*, 277–290. [[CrossRef](#)]
65. Barrett, L.W.; Fletcher, S.; Wilton, S.D. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. Life Sci.* **2012**, *69*, 3613–3634. [[CrossRef](#)]
66. Vu, L.D.; Gevaert, K.; De Smet, I. Feeling the heat: Searching for plant thermosensors. *Trends Plant Sci.* **2019**, *24*, 210–219. [[CrossRef](#)]
67. Reňák, D.; Gibalová, A.; Solcová, K.; Honys, D. A new link between stress response and nucleolar function during pollen development in *Arabidopsis* mediated by AtREN1 protein. *Plant Cell Environ.* **2014**, *37*, 670–683. [[CrossRef](#)]
68. Tuncozdemir, M.; Tang, C.; Ishka, M.R.; Brown, E.; Groves, N.R.; Myers, C.T.; Rato, C.; Poulsen, L.R.; McDowell, S.C.; Miller, G.; et al. A cyclic nucleotide-gated channel (CNGC16) in pollen is critical for stress tolerance in pollen reproductive development. *Plant Physiol.* **2013**, *161*, 1010–1020. [[CrossRef](#)]
69. Bukau, B.; Weissman, J.S.; Horwich, A.L. Molecular chaperones and protein quality control. *Cell* **2006**, *125*, 443–451. [[CrossRef](#)]
70. Neumann, D.; Nover, L.; Parthier, B.; Rieger, R.; Scharf, K.D.; Wollgiehn, R.; zur Nieden, U. Heat shock and other stress response systems of plants. *Results Probl Cell Differ.* **1989**, *16*, 1–155. [[PubMed](#)]
71. Baniwal, S.K.; Bharti, K.; Chan, K.Y.; Fauth, M.; Ganguli, A.; Kotak, S.; Mishra, S.K.; Nover, L.; Port, M.; Scharf, K.D.; et al. Heat stress response in plants: A complex game with chaperones and more than twenty heat stress transcription factors. *J. Biosci.* **2004**, *29*, 471–487. [[CrossRef](#)] [[PubMed](#)]
72. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2018**, *47*, D427–D432. [[CrossRef](#)] [[PubMed](#)]
73. Jaina, M.; Robert, D.F.; Sean, R.E.; Alex, B.; Marco, P. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **2013**, *41*, e121. [[CrossRef](#)]
74. La Cour, T.; Kiemer, L.; Molgaard, A.; Gupta, R.; Skriver, K.; Brunak, S. Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng. Des. Sel.* **2004**, *17*, 527–536. [[CrossRef](#)] [[PubMed](#)]

75. Kosugi, S.; Hasebe, M.; Tomita, M.; Yanagawa, H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10171–10176. [[CrossRef](#)]
76. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [[CrossRef](#)]
77. Guo, J.; Jain, R.; Yang, P.; Fan, R.; Kwok, C.K.; Zheng, J. Reliable and fast estimation of recombination rates by convergence diagnosis and parallel markov chain monte carlo. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2014**, *11*, 63–72. [[CrossRef](#)]
78. Krzywinski, M.; Schein, J.E.; Birol, I.; Connors, J.M.; Gascoyne, R.D.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [[CrossRef](#)]
79. Hu, B.; Jin, J.; Guo, A.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. [[CrossRef](#)]
80. Timothy, L.B.; Mikael, B.; Fabian, A.B.; Martin, F.; Charles, E.G.; Luca, C.; Ren, J.Y.; Wilfred, W.L.; William, S.N. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [[CrossRef](#)]
81. Chen, C.; Xia, R.; Chen, H.; He, Y. TBtools, a toolkit for biologists integrating various HTS-data handling tools with a user-friendly interface. *bioRxiv* **2018**. [[CrossRef](#)]
82. Szklarczyk, D.; Morris, J.H.; Cook, H.; Kuhn, M.; Wyder, S.; Simonovic, M.; Santos, A.; Doncheva, N.T.; Roth, A.; Bork, P.; et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **2017**, *45*, D362–D368. [[CrossRef](#)] [[PubMed](#)]
83. Jin, J.; Tian, F.; Yang, D.C.; Meng, Y.Q.; Kong, L.; Luo, J.; Gao, G. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **2017**, *45*, D1040–D1045. [[CrossRef](#)] [[PubMed](#)]
84. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
85. Clavijo, B.J.; Venturini, L.; Schudoma, C.; Accinelli, G.G.; Kaithakottil, G.; Wright, J.; Borrill, P.; Kettleborough, G.; Heavens, D.; Chapman, H.D.; et al. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res.* **2017**, *27*, 885–896. [[CrossRef](#)]
86. Liu, Z.; Xin, M.; Qin, J.; Peng, H.; Ni, Z.; Yao, Y.; Sun, Q. Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). *BMC Plant Biol.* **2015**, *15*, 152. [[CrossRef](#)]
87. Schmittgen, T.; Livak, K. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc.* **2008**, *3*, 1101–1108. [[CrossRef](#)]

