

RESEARCH

Open Access



Genome-wide identification, phylogenetic and expression pattern analysis of HSF family genes in the *Rye* (*Secale cereale* L.)

Yanyan Ren^{1†}, Rui Ma^{1†}, Muhua Xie¹, Yue Fan², Liang Feng³, Long Chen⁴, Hao Yang⁵, Xiaobao Wei⁶, Xintong Wang¹, Kouhan Liu¹, Peng Cheng^{1*} and Baotong Wang^{1*}

Abstract

Background Heat shock factor (HSF), a typical class of transcription factors in plants, has played an essential role in plant growth and developmental stages, signal transduction, and response to biotic and abiotic stresses. The *HSF* genes families has been identified and characterized in many species through leveraging whole genome sequencing (WGS). However, the identification and systematic analysis of *HSF* family genes in *Rye* is limited.

Results In this study, 31 *HSF* genes were identified in *Rye*, which were unevenly distributed on seven chromosomes. Based on the homology of *A. thaliana*, we analyzed the number of conserved domains and gene structures of *ScHSF* genes that were classified into seven subfamilies. To better understand the developmental mechanisms of *ScHSF* family during evolution, we selected one monocotyledon (*Arabidopsis thaliana*) and five (*Triticum aestivum* L., *Hordeum vulgare* L., *Oryza sativa* L., *Zea mays* L., and *Aegilops tauschii* Coss.) specific representative dicotyledons associated with *Rye* for comparative homology mapping. The results showed that fragment replication events modulated the expansion of the *ScHSF* genes family. In addition, interactions between *ScHSF* proteins and promoters containing hormone- and stress-responsive cis-acting elements suggest that the regulation of *ScHSF* expression was complex. A total of 15 representative genes were targeted from seven subfamilies to characterize their gene expression responses in different tissues, fruit developmental stages, three hormones, and six different abiotic stresses.

Conclusions This study demonstrated that *ScHSF* genes, especially *ScHSF1* and *ScHSF3*, played a key role in *Rye* development and its response to various hormones and abiotic stresses. These results provided new insights into the evolution of *HSF* genes in *Rye*, which could help the success of molecular breeding in *Rye*.

Keywords *Rye*, *HSF* gene family, Hormone, Abiotic stress

[†]Yanyan Ren and Rui Ma are co-first authors.

*Correspondence:

Peng Cheng
pengcheng@nwsuaf.edu.cn
Baotong Wang
wangbt@nwsuaf.edu.cn

¹ State Key Laboratory of Crop Stress Biology for Arid Areas, College of Plant Protection, Northwest A&F University, Yangling, Shaanxi 712100, People's Republic of China

² College of Food Science and Engineering, Xinjiang Institute of Technology, Aksu 843100, People's Republic of China

³ Chengdu Institute of Food Inspection, Chengdu 610000, People's Republic of China

⁴ Tianfu New Area General Aviation Profession Academy, Meishan 620564, China

⁵ Agricultural Service Center of Langde Town of Leishan County, Qiandongnan Miao and Dong Autonomous Prefecture 556019, China

⁶ Guizhou Provincial Center For Disease Control And Prevention, Guiyang 550025, People's Republic of China



Background

Rye (*Secale cereale*, $2n=2x=14$, RR) belongs to the genus *Secale* in the Triticeae tribe of Poaceae, the grass family [1]. It is a good source of carbohydrates and a small amount of protein, potassium, B vitamins, lignans, ferulic acid, alkylpolysechenol, and prebiotics. *Rye* is mainly used for bread production and high-quality animal feed and pasture [2, 3]. In addition, *Rye* exhibits strong tolerance to abiotic stress, such as cold tolerance, drought resistance and soil impoverishment resistance, as well as biotic stress, including resistance to fungi and other pathogens [4, 5]. Recently, numerous studies have been devoted to studying *Rye* genome, which has laid the foundation for the promotion of breeding and targeted gene editing in *Rye* [6, 7].

Transcription factors (TFs), are DNA-binding proteins that specifically interact with cis-acting elements of the eukaryotic proteins [8]. As an essential class of regulators, transcription factors are involved in almost all biological processes in plants. When plants are subjected to biotic and abiotic stresses, transcription factors bind to the specific promoter regions of genes to activate or inhibit transcription of downstream target genes for defensive responses [9, 10]. Previous studies have demonstrated that heat shock transcription factors (HSFs) can regulate plant heat stress responses [11]. *HSFs* contain five basic functional and structural domains: N-terminal DNA binding domain (DBD), oligomerization structural domain (OD), nuclear localization signaling domain (NLS), nuclear export signaling domain (NES), and C-terminal transcriptional activation domain (CAD) [12, 13]. The oligomeric domain consists of two hydrophobic heptapeptide repeats, abbreviated as HR-A and HR-B. Heat shock transcription factors are divided into three subfamilies, A, B, and C, depending on the different number of amino acids inserted in HR-A and HR-B [13–15]. HSEA subfamily inserts a long amino acid chain between HR-A and HR-B, with an aromatic-rich hydrophobic acidic amino acid tail (AHA) at the C-terminus. AHA is essential for transcriptional activation function. On the other hand, the HSEB subfamily has no amino acid sequence between HR-A and HR-B, while members of HSEC subfamily have a short amino acid sequence inserted between HR-A and HR-B. Therefore, HSEA has a transcriptional activation function, while the opposite holds true for HSEB and HSEC. Under normal conditions, HSFs transcription factors are present as monomers in cytoplasm of plants and are inactive. When plants are exposed to abiotic stresses (e.g., heat stress), HSFs are active through OD binding to DNA-binding structural domains and AHA structural domains to form heat-kill transcription factor trimers. DBD structural domain will be recognized and bind to heat stress

elements (HSEs) of the heat-kill protein promoter, which activates transcription of the heat-killed protein *Hsp* genes and enable plants to exhibit resistance to abiotic stresses [16–19].

Following the mining of genomic data, an increasing number of *HSF* gene families of plants such as *A. thaliana* [18], *Populus* L. [20], *T. aestivum* [21], *Beta. vulgaris* L. [22], and *Gossypium* spp. [23], *O. sativa* [24], *H. vulgare* [25], and *Z. mays* [26] have been identified. Scharf first identified transcription factors *HSFs* associated with heat stress response elements by exposing cell cultures of tomatoes to heat stress. Furthermore, studies on *HSFs* have been carried out in a variety of plants [27]. In *A. thaliana*, overexpression of *AtHSFB4* (*AT1G46264.1*) gene resulted in a shortened root length. Overexpression of *AtHSEA2* improves plant salt tolerance, and promotes callus growth [28]. Overexpression of *GmHSEA1* in soybean, and overexpression of *SlHSEA1* in tomato enhances heat tolerance of transgenic plants [29, 30]. In addition, numerous studies have revealed that *HSF* family has a crucial regulatory function in various physiological aspects of the growth and development in other plants, such as stress response and plant phase transition. Currently, a number of *HSF* genes have been isolated and identified in different plants, which are still poorly understood in *Rye* [11, 31, 32].

In this study, based on the newly published *Rye* genome [7], we have identified 31 *HSF* genes and compared their gene structure, motif composition, chromosomal location, and gene duplication. To further study the developmental mechanisms among species, the *HSF* genes in *Rye* were compared with closely related genera to analyze the evolutionary distance and relationships. Finally, the expression of *HSF* genes (under hormones and stress treatment) was analyzed by qRT-PCR, it was found that there was a differential expression pattern of *HSF* genes in different tissues, which initially confirmed the biological function of *HSF* genes in *Rye*. This study provided a comprehensive analysis of the *HSF* gene family in *Rye*, which provided valuable information for screening important *HSF* genes in *Rye* under different development stages and stress treatment, and provided a theoretical basis for functional analysis of *HSF* gene family in other species.

Result

Identification of the *HSF* gene in *Rye*

A total of 31 *SchHSF* genes were identified according to different chromosomes and named between *SchHSF1* and *SchHSF31*. The basic characteristics of *SchHSF*, such as molecular weight (MW), isoelectric point (pI), coding sequence length (CDS), and subcellular localization (<http://cello.life.nctu.edu.tw/>) were analyzed.

Among the 31 ScHSF proteins, ScHSF20 protein had the least amino acids (226), while ScHSF23 protein had the most amino acids (520). The molecular weight of the protein ranged from 24.64 kDa (ScHSF20) to 57.32 kDa (ScHSF23), and the pI ranged from 4.80 (ScHSF29) to 10.24 (ScHSF20), with an average of 6.14. All ScHSF proteins contained HSF_DNA-bind domains. Subcellular localization results showed that all ScHSF proteins were located in the nucleus, while four were located in the peroxisome, three were located in the cytoplasm, and one (ScHSF13) was located in the extracellular. The number of HSF genes in *Rye* (31) was higher than that in *A. tauschii* Coss (19), *A. thaliana* (21), *Z. mays* (25), and *O. sativa* (26) but lower than that in *H. vulgare* (32) and *T. aestivum* (82) (Additional file 1: Table S1).

Multiple sequence alignment, phylogenetic analysis, and classification of ScHSF proteins

To determine the phylogenetic relationships among *Rye* HSF proteins, a phylogenetic tree of *Rye* (31 ScHSFs) and *Arabidopsis* (22 AtHSFs) was constructed using MEGA 7.0 software. The 31 ScHSF proteins were divided into seven branches (groups 1–7) in the phylogenetic tree according to the previously proposed Cenci and Rouard classification method and topology [33]. Consensus exists with the taxa of HSF proteins in

Arabidopsis, indicating that these HSF genes remained stable during the evolutionary process.

Among seven subfamilies, subfamily C had the largest number of members (9 ScHSFs), and subfamilies A and A1 had the fewest members (only 1 ScHSF). All members were usually concentrated in the three subfamilies A2, B and C. Comparison with the phylogenetic tree of *Rye* revealed that some of the ScHSFs clustered closely with AtHSFs (bootstrap support ≥ 70), suggesting that these proteins in *Rye* and *A.thaliana* might be homologous and have similar biological functions (Fig. 1a; Additional files 1 and 2: Table S1 and Fig. S1).

Previous studies have shown that all HSF proteins contain DBD-conserved domain and are highly conserved. The conserved domain contains three α -helical bundles ($\alpha 1$ - $\alpha 3$) and four reverse parallel β -folds ($\beta 1$ - $\beta 4$), consisting of about 100 amino acid residues (Fig. 1b). However, there are varying degrees of insertions or deletions in these proteins. For example, in subfamily C, *ScHSF21* had an 8 amino acid sequence inserted between $\alpha 3$ and $\beta 3$, and *ScHSF18* had an 11 amino acid sequence missing between $\alpha 3$ and $\beta 3$. Overall, DBD domains in *Arabidopsis* and *Rye* were highly conserved, indicating that DBD structural domains were established early in plants.

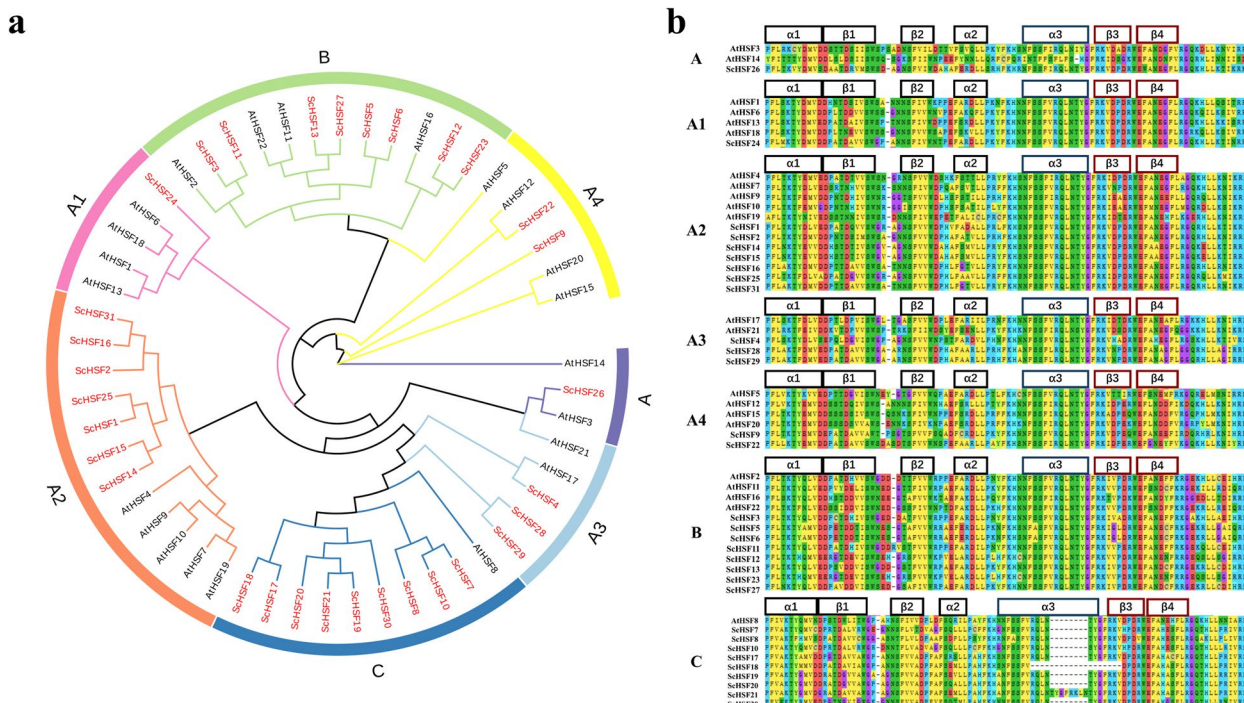


Fig. 1 The evolutionary relationship and sequence alignment of the rye HSF proteins. **a** A phylogenetic tree of *Rye* and *Arabidopsis thaliana* HSF proteins showing that HSF proteins were divided into seven subfamilies. *Rye* proteins were denoted in red, whereas *A. thaliana* proteins were indicated in black. **b** Multiple sequence alignment of the DBD domains among the seven subfamilies of the ScHSF protein family

Conserved motifs and gene structure analysis of *ScHSF* genes

The structural and taxonomic diversity of HSF genes in *Rye* were explored by comparing genomic DNA sequences. By comparing the localization and number of exon–intron structures, 31 *ScHSF* genes were found to have different numbers of exons, ranging from 1 to 4. Moreover, five *ScHSF* genes had only one exon, while most *ScHSF* genes (20, ~64.5%) contained 2 exons, and all genes contained HSF-DNA binding sites (Additional file 3: Figure S2). In addition to this, *ScHSF1* and *ScHSF14* genes belonged to A2 subfamily, which have the same intron and exon structure with 4 exons and 3 introns, the highest number of *HSF* genes (Additional file 3: Figure S2a, b). It was worth mentioning that *ScHSF21* had a very large intron structure. In general, *ScHSF* genes of the same subfamily had similar gene structures. Subfamily A2 showed greater structural differences in the number of introns. Therefore, it could be speculated that they might have more biological functions.

To further evaluate the structural diversity of *ScHSF* genes, the motifs of *ScHSF* genes were analyzed using online motif software. Ten different motifs were identified in *ScHSF* proteins (motif 1 to motif 10). Motif 1, 2, 4, and 5 were usually located together, indicating that these four motifs were closely associated with SPL proteins. Notably, *ScHSF* genes of the same subfamily usually had similar motif compositions. For example, subfamily C contained motifs 1, 2, 3, 4, and 5, while subfamily A2 contained all the same motifs (motifs 1–9). Furthermore, it was found that some motifs were located at specific positions. For instance, motif 2 was always located at the beginning of the motif, while the motifs located at the end were different. Motif 5 was always between motif 1 and motif 2 (Additional file 3: Figure S2c). Overall, these results indicated that genes from the same subfamily had similar genetic composition and structure tended to cluster together, which was consistent with the population classification of the phylogenetic tree.

Chromosomal spread and gene duplication in *ScHSF* genes

According to the newly published *Rye* genome database, *HSF* genes (31) were distributed on seven chromosomes (Chr), and each *HSF* gene was named based on its physical position on chromosomes. Chr5 contained the most *ScHSF* genes (11, ~35.5%), followed by Chr7 (8, ~25.8%). Chr1 had only 1 gene (~4.76%), while Chr4 had no distribution of *ScHSF* genes. Notably, nearly half of *HSF* genes were located at the bottom of chromosomes (Fig. 2a).

Gene duplication events, including tandem repeat events and segmental duplications, play an essential

role in gene amplification and the generation of new functions, [34]. Chromosomal regions with a range of 200 kb containing two or more genes are defined as tandem repeat events [35]. Accordingly, a duplication event analysis of *HSF* genes in *Rye* was performed to explore the evolutionary conservation of gene family. The results showed that there were no tandem duplication events in *Rye* genome, but two pairs of duplicated fragments were present (Fig. 2b, Additional file 4: Table S2). Four homologs in *HSF* genes indicated an evolutionary relationship between these genes. The highest number of *ScHSF* was found in LG5 ($n=2$), followed by LG2 and LG7 ($n=1$), and all genes were linked within subfamily B, suggesting that some *ScHSF* genes might be accompanied by fragment replications. These replication events were the main drivers of new functions of *ScHSF* genes during evolution.

Evolutionary analysis of the *ScHSF* genes and *HSF* genes of several different species

A dicotyledon (*A. thaliana*) and five monocotyledons (*H. vulgare*, *O. sativa*, *Z. mays*, *T. aestivum* and *A. tauschii* Coss) were selected to analyze the evolution of *HSF* in *Rye*. 31 of *ScHSF* genes identified were compared with 10 conserved motifs from 6 other plants. *ScHSF* genes were not evenly distributed in the phylogenetic tree. As shown in Fig. 3a, the *ScHSF* proteins tended to gather with the *HSF* proteins of *T. aestivum* and *H. vulgare*, suggesting that they are more closely related (Fig. 3a). These genes, from the same subfamily, tended to have the same themes and tended to cluster together. Remarkably, almost all *HSF* genes from these seven plants contained motifs 1, 2, 3, 6, and 7 (Fig. 3, Additional file 5: Table S3). Subfamily A2 contained the most motifs and showed a diversity of expression. In addition, all genes except *HSF* in wheat began with motif 3. And in subfamily A1, A2, A3, and C, motif 8 was always distributed at the end of the pattern. In summary, *ScHSF* genes in subfamily A3 had high homology with barley *HSF* gene clusters, while most of the *HSF* genes in other groups had homology with wheat *HSF* gene clusters, indicating a closer distance in evolution and with similar potential functions.

A homology map between *Rye* and six representative species was constructed to explore the phylogenetic mechanisms of *HSF* genes in *Rye*. These species included one dicot (*A. thaliana*) and five monocots (*H. vulgare*, *O. sativa*, *Z. mays*, *T. aestivum*, and *A. tauschii* Coss). A total of 31 *ScHSF* genes were colinear with those of *A. tauschii* Coss (19), *A. thaliana* (21), *Z. mays* (25) and *O. sativa* (26), *H. vulgare* (32), and *T. aestivum* (82). The number of homologous pairs among the other six species (*A. thaliana*, *H. vulgare*, *O. sativa*, *Z. mays*, *A. tauschii* Coss,

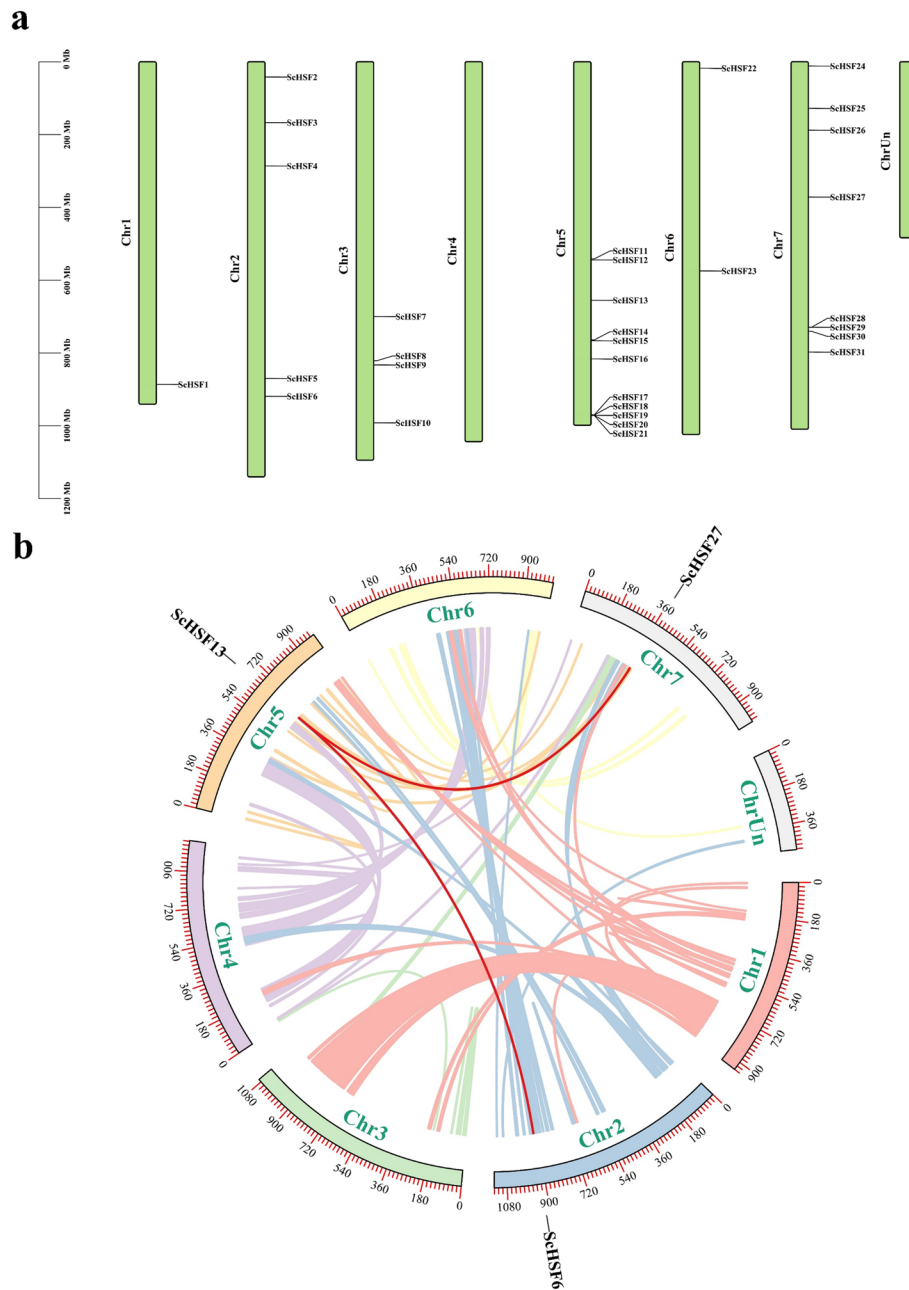


Fig. 2 The chromosomal distribution and synteny blocks of HSF genes in *Rye*. **a** Distribution of the 31 *ScHSF* genes on different chromosomes. The scale represented the length of chromosomes, whereas the green bars indicated chromosomes. The chromosome number is displayed on the left side of each green bar. **b** Analysis of interchromosomal fragment duplication of *HSF* genes in the *Rye* genome. The colored lines represented all synthetic blocks and the red lines specifically indicated the duplicated pairs among the 31 *ScHSF* genes

and *T. aestivum*) was 1, 19, 19, 21, 26, and 69, respectively (Fig. 3, Table S3).

Homology analysis of these six plants revealed at least one pair of genes homologous to *ScHSF*, such as *ScHSF25* with *AET4Gv20678400.5/ Zm00001d032923_T002/ AT3G22830.1/ HORVU4Hr1G073650.2/ Os10t0419300-01/ TraesARI4B01G298300.1*, indicating that these

homologous genes were highly conserved and might have existed prior to the ancestral divergence. Accordingly, it was speculated that they might have played a crucial role in the evolution of HSF gene family in *Rye*. Interestingly, In *H. vulgare*, *O. sativa*, *A. tauschii* Coss, *Z. mays*, and *T. aestivum*, gene pairs were found to be collinear with eight *ScHSF* genes (*ScHSF6*, *ScHSF7*, *ScHSF8*, *ScHSF12*,

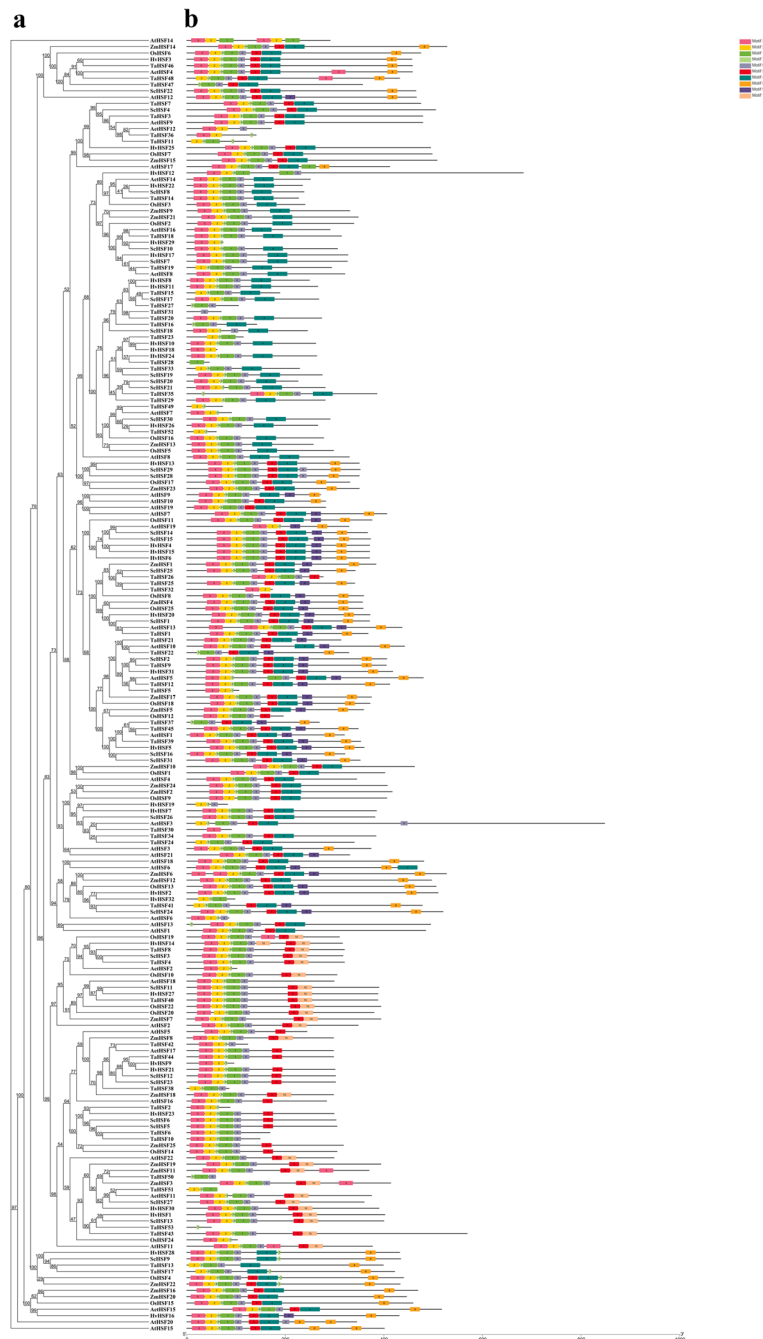


Fig. 3 Phylogenetic relationships and motif compositions of HSF proteins of seven different plant species (*Rye, A. thaliana, H. vulgare, O. sativa, Z. mays, T. aestivum, and A. tauschii* Coss). **a** An unrooted phylogenetic tree was constructed using the neighbor-joining method as implemented by Geneious R11. **b** Distribution of the conserved motifs in HSF proteins. Ten differently colored boxes represent different motifs and their position in each HSF protein (Table S3)

ScHSF13, *ScHSF24*, *ScHSF25*, and *ScHSF27*) (Fig. 4, Additional file 6: Table S4). These homologous gene pairs could have been formed by gene replication during the differentiation of dicotyledons and monocotyledons.

Analysis of cis-acting elements in *ScHSF* promoters

The promoter regions of *ScHSFs* were analyzed to provide ideas for tissue-specific expression of genes and stress response patterns. The cis-acting elements in promoter could be divided into four categories:

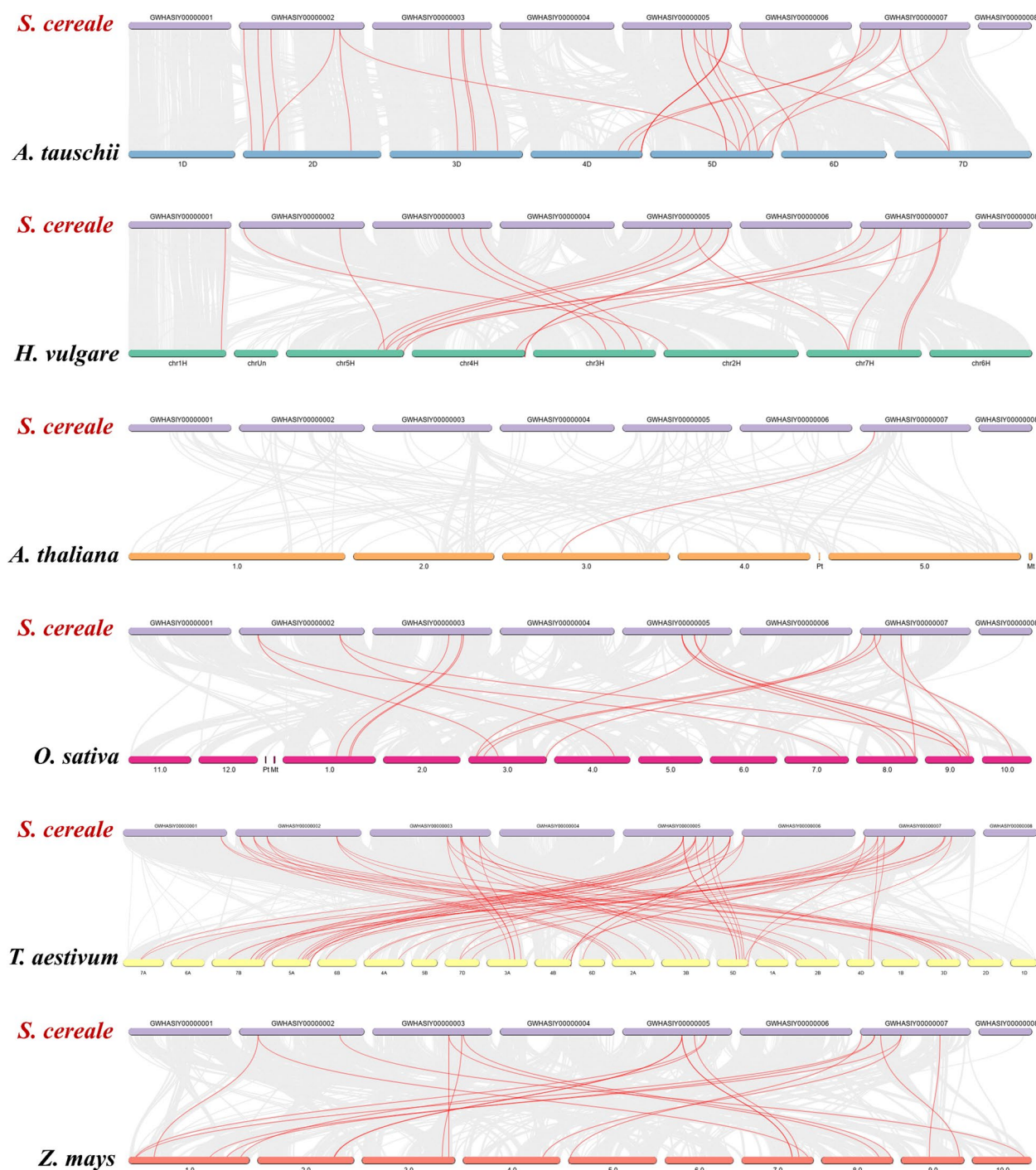


Fig. 4 Analysis of the *HSF* genes between *rye* and six representative plant species (*A. thaliana*, *H. vulgare*, *O. sativa*, *Z. mays*, *T. aestivum*, and *A. tauschii* Coss). Gray lines in the background indicate the neighboring blocks in genomes of *rye* and other plants, whereas red lines highlight the syntenic *rye* *HSF* gene pairs

light-responsive, hormone- and stress-responsive, plant growth- and development-related elements. Individual *ScHSF* gene in *Rye* covered most of phytohormone response elements, including abscisic acid

response elements (ABRE). MeJA hormone response elements (CGTCA-motif and TGACG-motif). In addition, cis-regulatory elements were associated with low temperature, drought, anaerobic conditions, and other

defenses found in all *ScHSF* genes (Fig. 5, Additional file 7: Table S5).

All genes contained light (G-BOX), drought (MYC), and stress (STRE) elements, while 93.3% of *ScHSF* genes

contained MeJA and ABA response elements. The promoters of *ScHSF1*, *ScHSF8*, *ScHSF22*, and *ScHSF26* contained growth hormone γ , ethylene γ , SA-, and gibberellin- reaction elements. All *ScHSF* promoters except

Function	Element	<i>ScHSF1</i>	<i>ScHSF3</i>	<i>ScHSF4</i>	<i>ScHSF8</i>	<i>ScHSF9</i>	<i>ScHSF13</i>	<i>ScHSF15</i>	<i>ScHSF18</i>	<i>ScHSF21</i>	<i>ScHSF22</i>	<i>ScHSF23</i>	<i>ScHSF24</i>	<i>ScHSF26</i>	<i>ScHSF28</i>		
Light responsive	Box 4	2		1								1		1	2		
	GATA-motif			3	1					1		2	2				
	G-Box	4	3	5	12	2	4	6	4	5	3	7	1	2	7	4	
	GT1-motif	1		1		1		1	1	3	2		1	1			
	I-box			1	1	2			1			1			1		
	TCCC-motif		2	1						1		2	2	1	1	1	
	TCT-motif	1	1	1					1	1		1	1				
MeJA	CGTCA-motif	3		8	4	3	2	3	5	4	1	6	3	6	2	5	
	TGACG-motif	3		8	4	3	2	3	5	4	1	6	3	6	2	5	
ABA	ABRE	2	3	6	1		5	6	6	4	1	7	2	1	3	3	
	DRE1			1			1			1				2			
Auxin	AuxRR-core		1	1											1		
	TGA-element	1		1	1	2		2			2		1	1			
Ethylene	ERE	2			1						1				1		
SA	TCA-element		1	1	1	1			1		1				1		
Gibberellin	GARE-motif			1		1				1							
	P-box	2		1	2						1						
	TATC-box								1		1					1	
Drought	as-1	3		8	4	3	2	3	5	4	1	6	3	6	2	5	
	MBS	2	1	2	2				1	1	1	1	2				
	MYC	3	3	3	2	2	3	4	4	6	5	2	2	5	3	3	
Wound/Pathogen	box S	1							1	1		1			1	1	
	WRE3	4	3			4	1	2	5	2	1	4	2	1	1	5	
Anaerobic induction	ARE		2	1	2		1	1			1	1	2		1		
Cold/Dehydration	DRE core		1	1	1	3	1		1	2		1		3			
Anoxic specific inducibility	GC-motif	3		1	1	4	2		1	2	1	4	3	3		4	
Low-temperatures	LTR	1		3			1	1	1	2	2	1	3	1			
Stress	STRE	2	5	9	2	9	4	6	6	4	6	4	8	3	3	5	
Defense/Stress	TC-rich repeats		1									1	1		1		
Wound	WUN-motif									1							
Plant growth and development	Specific day-length	A-box	6	1	1	2	1	3	1		2		6	2		1	2
	Zein metabolism	O2-site			2	4			1		1		1				1
	Meristem specific gene activation	CCGTCC-box	6	1	1	2	1	3	1		2		6	2		1	2
	MYBHv1 binding site	CCAAT-box	2	1		1		2	1		2				1		1
	Seed specific expression	AAGAA-motif	5				1		1			1		1	1	1	

Fig. 5 The distribution of cis-acting elements in promoters of *ScHSF* gene family members

the *ScHSF3* promoter contained drought-related element as-1 (Fig. 5, Additional file 7: Table S5). These results suggested that certain cis-acting elements might be involved in regulating the expression of different tissues, such as seeds and meristematic tissues. Furthermore, we speculated that *ScHSF* genes might be involved in tissue developmental processes and in response to various abiotic stresses.

Protein–protein interaction network analysis of *ScHSF* family members

The combination of promoter cis-elements and transcription factors can regulate the precise initiation and efficiency of transcription. The PlantTFDB was used to explore the potential TFs that binding to *ScHSF* promoter. The results showed that *ScHSF18* and *ScHSF15* had the most and least transcription factors, respectively (Fig. 6a). Meanwhile, All *ScHSF* genes were regulated by a large number of *ERF* transcription factors. Studies have shown that ERFs can regulate the expression of target genes JA-based target genes and defense against *Boea chinensis* in *Arabidopsis thaliana*, suggesting that *ScHSF* might indirectly participate in regulation of JA synthesis against pathogens. Meanwhile, it has been reported that *OsERF3* acts as a central switch that enables plant metabolism to respond appropriately to insects [36]. Therefore, it was speculated that *ScHSF* and participating in JA response played a potential role in insecticide traits through *ERF* regulation [37, 38].

To better understand regulatory relationships of *HSF* genes in *Rye*. We performed mutual protein prediction of *HSF* genes in *Rye* based on the most homologous wheat species. As shown in Fig. 6b, interactions existed among the nine *ScHSF* members. Moreover, *ScHSF5* and *ScHSF13* could interact with each other. Interestingly, *ScHSF3* and *ScHSF11* could interact with each other, along with *ScHSF1*, *ScHSF2*, *ScHSF4*, *ScHSF16*, and *ScHSF25*. Overexpression of *AtHSA2* (homologous *ScHSF1*, *ScHSF2*, *ScHSF16*, and *ScHSF25*) significantly increased basal and acquired heat tolerance in *Arabidopsis* plants [39]. *ScHSF3*, *ScHSF5*, *ScHSF11*, and *ScHSF13*, were identified as homologs proteins of *Arabidopsis* HSF2b, a protein involved in plant resistance to pathogens [40].

Expression patterns of *ScHSF* genes in different plant organs

To further evaluate the potential function of *ScHSF* genes, a total of 15 genes in seven subfamilies were selected and the expression of these representative genes in four plant organs (root, stem, leaf, and flower) was analyzed by qRT-PCR. *ScHSF* genes showed different expression patterns in roots, stems, leaves, and flowers, suggesting that these genes contribute to diverse regulatory roles ($P < 0.05$). All genes were expressed in different tissues; two genes (*ScHSF3*, and *ScHSF9*) had the highest expression level in fruits; seven genes (*ScHSF1*, *ScHSF4*, *ScHSF13*, *ScHSF15*, *ScHSF22*, *ScHSF26*, and *ScHSF28*) had the highest expression level in roots, while six genes

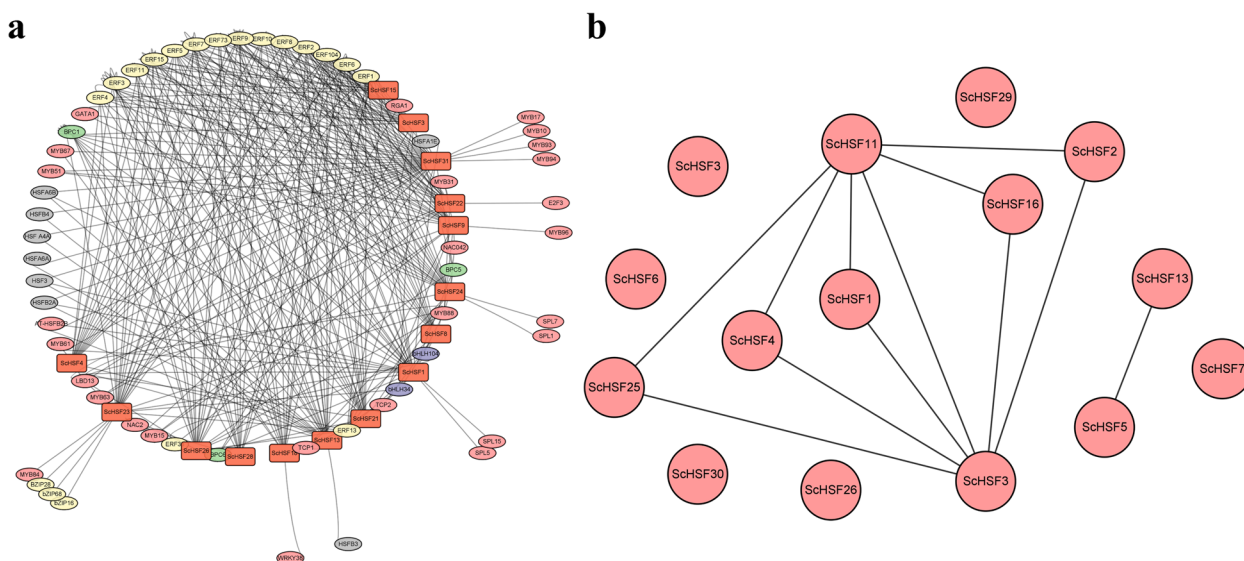


Fig. 6 Predicted interactions between rye HSF proteins. **a** Regulation network between *ScHSFs* and potential TFs. Red boxes represent *ScHSFs* genes and different colored ovals represent different TFs. Yellow shows ERF, gray shows HSF, blue shows bZIP, green shows SPL, pink shows MYB. **b** Prediction of the protein–protein interaction network among 31 *ScHSFs*

(*ScHSF8*, *ScHSF18*, *ScHSF21*, *ScHSF23*, *ScHSF24*, and *ScHSF31*) had the highest expression level in flowers (Fig. 7a). Most genes from the same subfamily had similar expression patterns, suggesting that these genes might have similar functions. By analyzing the expression of *ScHSF* genes in different tissues, it was obvious that all *HSF* genes were least expressed in leaves. We could speculate that *HSF* genes might be relevant to the development of stems, roots, and flowers in plants (Fig. 7b).

The *ScHSFs* may also regulate the fruit development of *Rye*, thus affecting its nutritional composition and development rate [41, 42]. Therefore, 15 *HSF* genes were analyzed for expression level at five different post-anthesis stages (7d, 14d, 21d, 28d, 35d) after anthesis to identify genes that could potentially regulate *Rye*

fruiting-related genes ($P < 0.05$). Most *ScHSF* genes displayed different expression patterns at five stages of fruit development. In the fruit of *Rye*, the expression of five genes (*ScHSF4*, *ScHSF15*, *ScHSF18*, *ScHSF22*, and *ScHSF31*) were significantly increased with fruit development, while the expression of most genes (*ScHSF3*, *ScHSF13*, *ScHSF23*, *ScHSF26*, and *ScHSF28*) decreased with fruit development (Fig. 7c). The expression of most genes showed a down-regulation trend in fruit expression with increasing time, and it could be speculated that *HSF* genes showed negative regulation in fruits ($P < 0.05$) (Fig. 7d). This also demonstrated that *HSF* genes played an essential role in fruit development and provided a theoretical basis for the studying of the nutritional value of *Rye*.

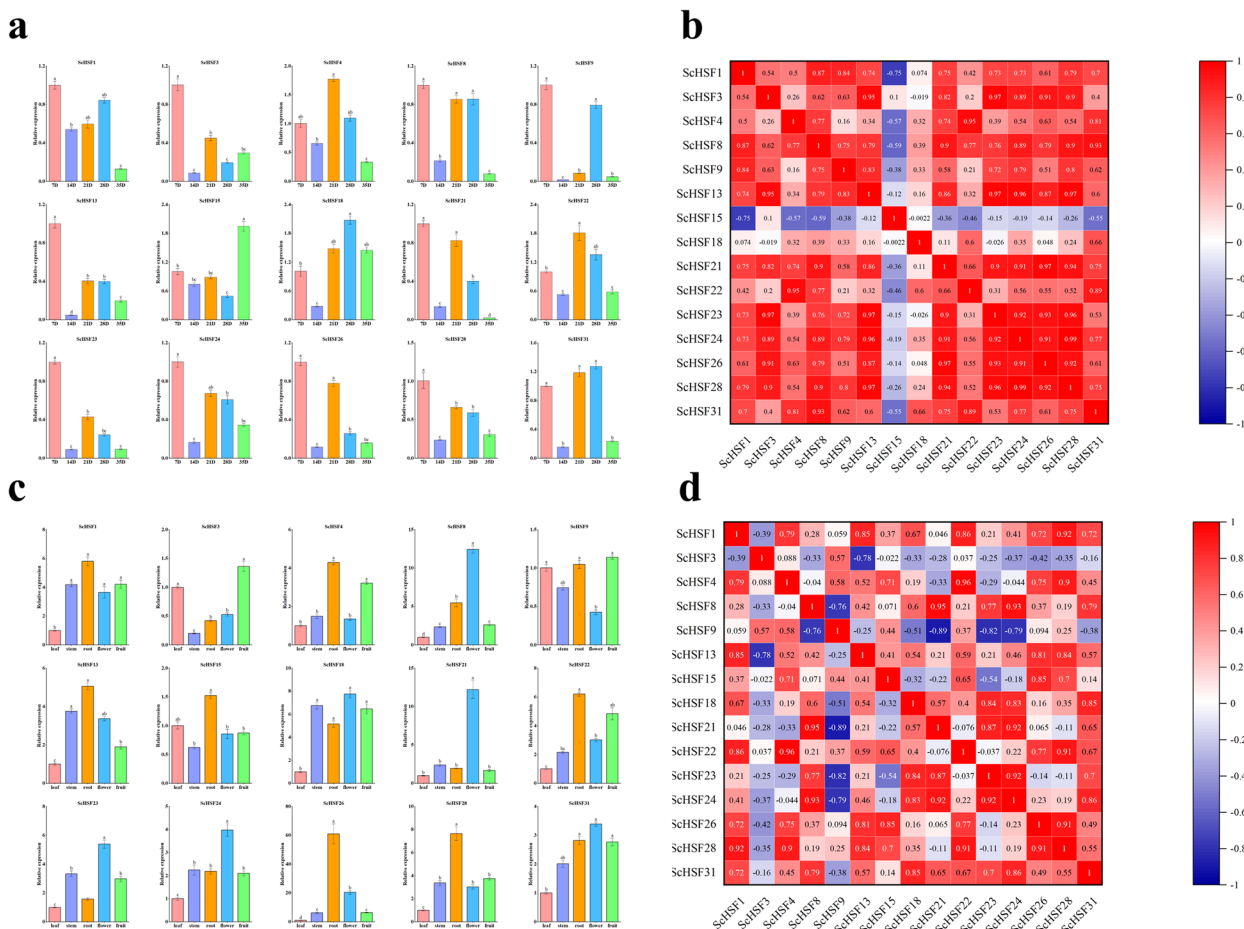


Fig. 7 Tissue-specific gene expression of the 15 *ScHSF* genes and gene expression levels during various fruit development stages. **a** Expression patterns of the 15 *ScHSF* genes in flower, leaf, root, stem, and fruit tissues were analyzed using qRT-PCR. Error bars represent the stand errors with three replications. Lowercase letters indicate significant differences among treatments ($\alpha = 0.05$, LSD). **b** Positive number = positive correlation; negative number = negative correlation. Red numbers indicate a significant correlation at the 0.05 level. **c** Expression patterns of the 15 *ScHSF* genes at different fruit developmental stages were analyzed using qRT-PCR (7 DPA, 14 DPA, 21 DPA, 28 DPA, and 35 DPA). Error bars represent the stand errors with three replications. Lowercase letters indicate significant differences among treatments ($\alpha = 0.05$, LSD). **d** Positive number = positive correlation; negative number = negative correlation. Red numbers indicate a significant correlation at the 0.05 level

Expression patterns of *ScHSF* genes under various treatments

To determine whether the expression of *ScHSF* genes was restricted by different abiotic stresses, a representative expression of 15 *ScHSF* genes were expressed under six abiotic stresses. The results showed that some *ScHSF* genes exhibited significant up-regulated and down-regulated expression patterns under different stress treatments. Most of *ScHSF* genes also displayed significant differences in diverse tissues with the treatment period ($P < 0.05$). For example, most *HSF* genes were induced by cold stress in stems and leaves, whereas most genes were induced to express in roots under heat stress. Notably, *ScHSF3*, *ScHSF9*, *ScHSF13*, *ScHSF15*, *ScHSF21*, *ScHSF23*, *ScHSF28*, and *ScHSF31* showed opposite patterns of highest expression levels in leaves and stems, compared to roots under cold and heat stress. Under flooding stress, expression level of *ScHSF1*, *ScHSF3*, and *ScHSF31* were most significantly up-regulated and mostly concentrated at 4 h treatment time. Under drought stress, *ScHSF1*, *ScHSF8*, *ScHSF18*, *ScHSF22*, *ScHSF24*, *ScHSF26*, and *ScHSF28* were significantly up-regulated in leaves. Meanwhile, most of the genes also showed a significant up-regulation under UV and NaCl stresses (Fig. 8). It should be emphasized that *ScHSF1* and *ScHSF3* were significantly highly expressed under all six different stress treatments and it could be further investigated as a potential candidate gene for stress management ($P < 0.05$).

In addition, the expression patterns of *ScHSF* genes under ABA, IAA, and GA3 were used to further explore the functions of genes. The genes exhibited different expression patterns under different hormone treatments ($P < 0.05$) (Fig. 9). Under ABA treatment, most genes showed an up-regulation trend, while *ScHSF15*, *ScHSF24*, *ScHSF26*, and *ScHSF31* showed a down-regulation trend. Moreover, *ScHSF9* showed the highest expression level under IAA treatment, while only *ScHSF5* expression level was down-regulated. Expression level of *ScHSF22* and *ScHSF28* were the highest under IAA treatment.

Discussion

ScHSF gene structure and evolutionary analyses

During the growth and development of *Rye*, *HSF*, a transcription factor involved in various stress responses, including high-temperature stress, salt stress, drought stress, and oxidation stress, played a crucial role [31, 32]. In recent years, the rapid development of metagenomics has resulted in the identification and characterization of *HSF* genes in many plants, including *O. sativa*, *A. thaliana*, *Z. mays*, *Poplar*, *tomatoes*, *T. aestivum*. However, the study of *ScHSF* family was still poorly understood to date.

In this study, 31 *HSF* genes were identified in *Rye* with *HSF* proteins ranging from 260 to 520 amino acids in length (Additional file 1: Table S1). A comparative genomic analysis of gene structure revealed that all *HSF* genes contained different numbers of introns, ranging from 0 to 4. All of the encoded proteins showed complex and variable structures, while variability might be attributable to gene duplication during evolution. The introns, as a part of plant evolution, might not only increase the length of genes as well as the frequency of recombination between genes and played a major role in the regulatory roles [43]. In contrast, genes without introns have no advantages during the evolution of the species and delay regulatory responses [44–47]. Therefore, many *ScHSF* members respond rapidly when subjected to stress treatments. The same subfamily has a similar number of motif compositions and introns, which allows us to speculate that they might share a common evolutionary origin and molecular functions. This approach can also be used to predict the function of unknown proteins.

Based on the conserved structural domains of *Arabidopsis*, they were divided into seven subfamilies. Each group contains at least one *HSF* gene from *Arabidopsis* and *Rye*, suggesting that these genes have not been missing during evolution and might have some biologically important functions (Fig. 1a). Gene amplification is the main driver for the generation of new functional genes during evolution, which is divided into two types: segmental duplications and tandem replication. Compared to segmental replication [48], tandem duplication events represent a larger proportion of plant genomes, with an approximately 10% incidence in *Arabidopsis* and *rice* [49, 50]. We found more *HSF* proteins in *Rye* compared with *A. tauschii* Coss (19), *A. thaliana* (21), *Z. mays* (25), and *O. sativa* (26). This suggested a possibility of more gene duplication events happened in *Rye*, which could also lead to the production of new functional genes in plants to adapt to harsh environment [51]. Based on physical location, 31 *ScHSF* genes were unevenly distributed on 7 chromosomes of *Rye* (Fig. 2a). Homology analysis of the *HSF* gene in *Rye* showed that no tandem duplicate gene pairs were discovered. Nevertheless, two pairs of fragment duplicates were identified (Fig. 2b). The homologous genes on different chromosomes of *Rye* might have promoted the evolution of *ScHSF* genes, resulting in a higher number of *HSF* genes in *Rye* than in other monocotyledons (*A. tauschii* Coss, *Z. mays*, and *O. sativa*).

To further speculate on phylogenetic developmental mechanisms of *HSF* genes, six comparative syntenic maps of *Rye* connections with one dicotyledon and five monocotyledons were constructed. As we can see from Fig. 3, *HSF* genes from both *Rye* and different plants were classified into seven taxa. *ScHSF* genes from

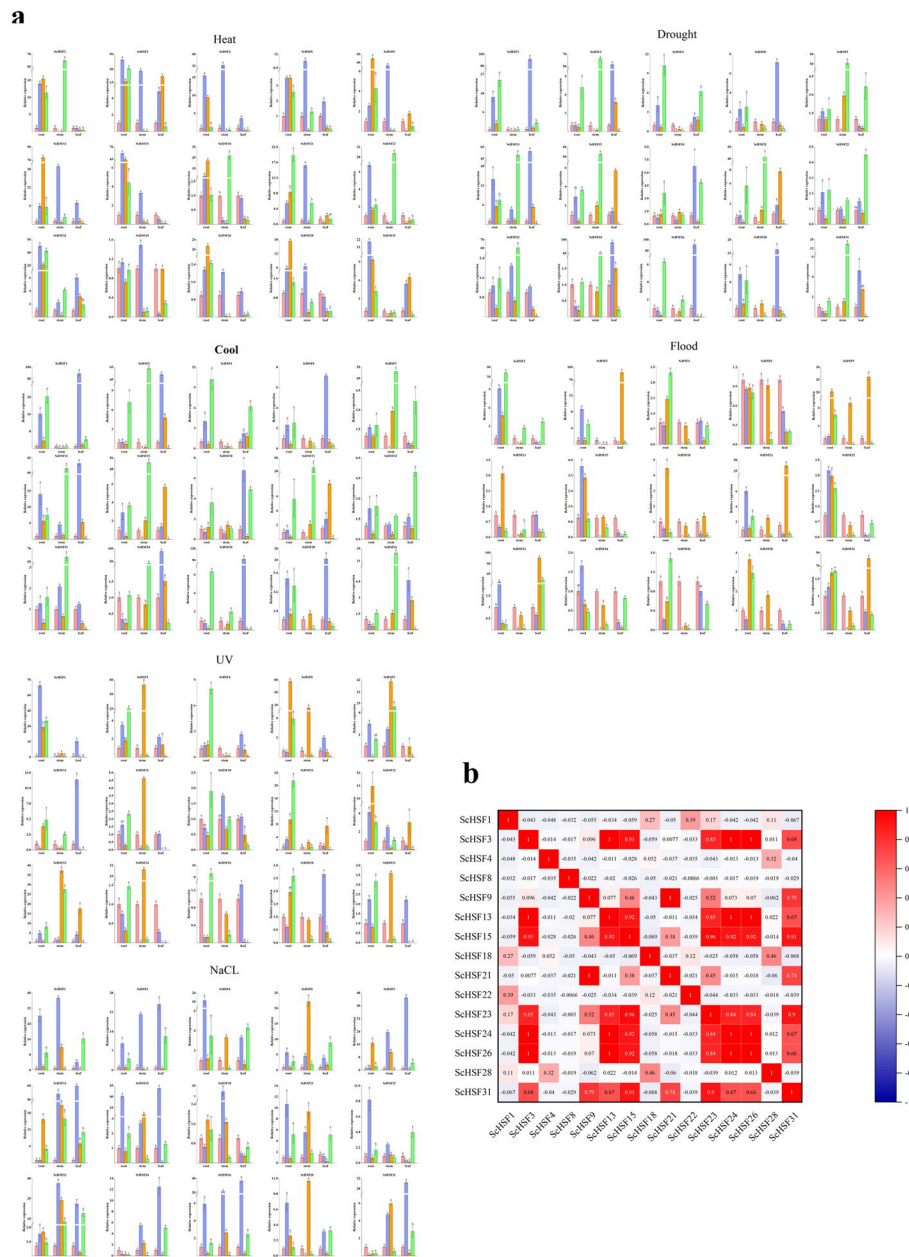
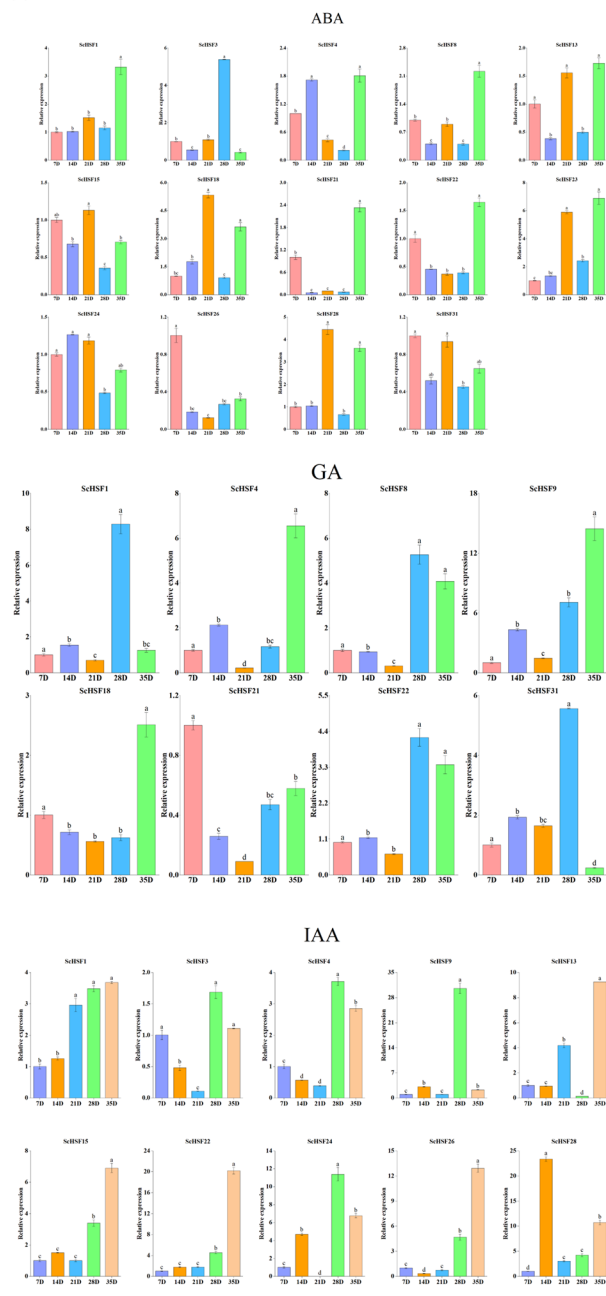


Fig. 8 Expression analysis of the 15 *ScHSF* genes in three tissues (roots, stems, and leaves) at the seedling stage under different abiotic stresses (UV radiation, flooding, PEG, NaCl, heat, and cold treatments). **a** Expression analysis of the 15 *ScHSF* genes was performed using qRT-PCR. Error bars represent the stand errors with three replications and the lowercase letter above the bar indicates a significant difference ($\alpha=0.05$, LSD) among the treatments. **b** Positive numbers = positive correlations; negative numbers = negative correlations. Red numbers indicate a significant correlation at the 0.05 level

subfamily A3 showed higher homology with barely *HSF* genes clusters, whereas most of *HSF* genes from other groups were clustered with wheat. Interestingly, there was at least one pair of co-linear genes between *ScHSF25* and *AET4Gv20678400.5/ Zm00001d032923_T002/ AT3G22830.1/ HORVU4Hr1G073650.2/ Os10t0419300-01/ TraesARI4B01G298300.1*, which might provide a

theoretical basis for understanding whether they shared a common ancestor. Analysis of orthologous genes also illustrated that *ScHSFs* had the highest number of homologous gene pairs with wheat, indicating a higher level of homology among them. In addition, by analyzing the motif composition of *HSF* genes in plants, we found that *HSF* genes contained 10 motifs, with different subfamilies

a



b

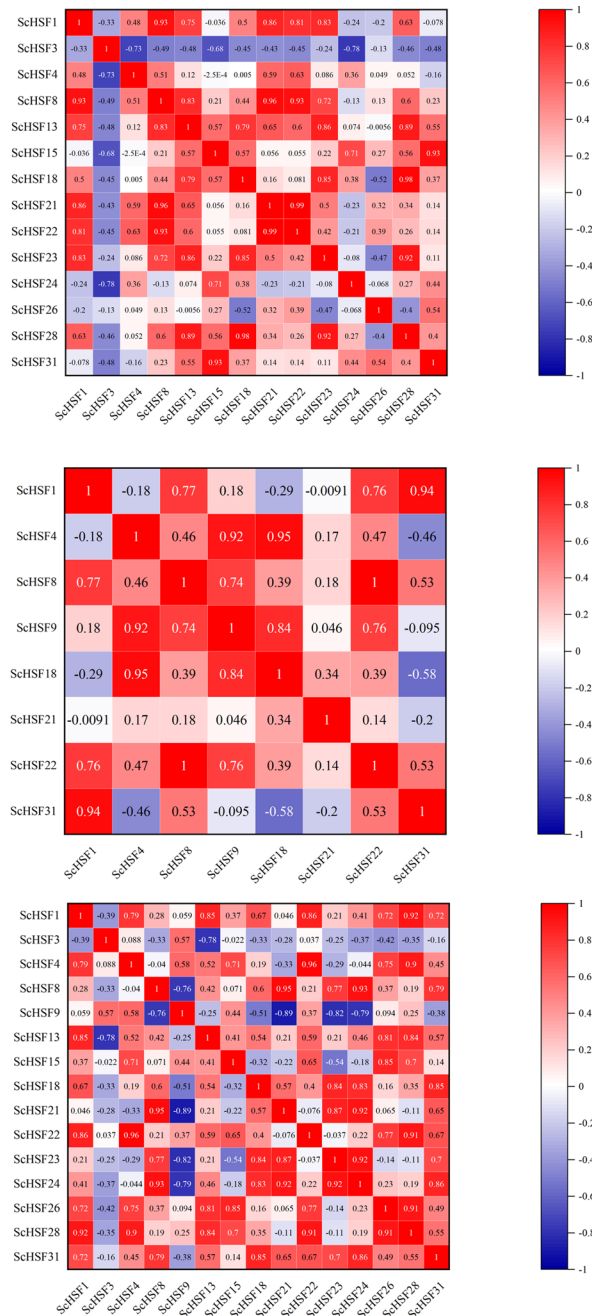


Fig. 9 Expression analysis of the 15 *ScHSF* genes in fruits under different hormones (ABA, IAA, and GA). **a** Expression analysis of the 15 *ScHSF* genes was performed using qRT-PCR. Error bars represent the standard error of three replicates, with lowercase letters above the error bars indicating significant differences among the treatments ($\alpha=0.05$, LSD). **b** Positive numbers = positive correlation; negative numbers = negative correlation. Red numbers indicate a significant correlation at the 0.05 level

containing similar motifs. Moreover, *HSF* genes contained almost all motifs of the A2 subfamily. These results also reaffirmed that *HSF* genes in *Rye* were more closely related to wheat and might have a common ancestor.

Expression patterns and function prediction of *ScHSFs*

The analysis of gene expression is often used as an essential step in providing useful clues for functional prediction [52]. In this study, the expression patterns of 15

genes, which were represented in seven subfamilies, were selected and explored in different tissues and at different developmental stages. The results showed that most of *HSF* genes were significantly expressed (more than a twofold difference). For instance, most genes were significantly up-regulated in stems and leaves under cold treatment, while all genes were significantly up-regulated under UV and drought treatment (Fig. 8). This explained the high adaptability of *Rye* crop in alpine or arid areas. Most of *HSF* genes were significantly up-regulated in response to stress in these six treatments, and the genes expression was mainly in leaves and stems. However, the expression of most of *HSF* were highest in roots, suggesting that roots played a key function under drought conditions. Notably, both *ScHSF1* and *ScHSF3* were expressed in response to all six stresses. It could be further validated as a potential candidate gene for improving crop breeding.

Previous studies have shown that *HSF* genes were mainly involved in several environmental stress responses, such as high-temperature, salt, drought, and oxidation stress [11, 31, 32]. ClassA *HSFs* are major regulators of heat stress and could induce the expression of resistance genes [53]. The high expression levels of most *HSF* genes under different stress treatments also suggested that *HSF* genes played a significant role in roots and leaves. For instance, Tahmina reports that overexpression of the *AtHSFB4* (*AT1G46264.1*) gene in *Arabidopsis* resulted in a shortened root length [28]. *ScHSF23*, the corresponding homologous gene, was up-regulated in roots for stress-responsive expressions under heat treatment and drought conditions. Based on the fact that homologous genes with similar structures may have similar functions, it is speculated that *ScHSF23* may be related to root growth and development. Overexpression of *AtHSFA2* in *A. thaliana* increases stress tolerance, and enhanced callus growth [49, 54]. The *ScHSF1*, *ScHSF 2*, *ScHSF 14*, *ScHSF 15*, *ScHSF 16*, *ScHSF 25*, and *ScHSF 31* belong to A2 subfamily, which had high similarity to *AtHSFA2*. Meanwhile, we found that A2 subfamily had the most abundant motifs, *ScHSF1* and *ScHSF15* also showed significant up-regulation in stress treatment (Fig. 8). Finally, two genes from each subfamily were screened for qRT-PCR analysis and verification of functional traits. The results showed that these genes were significantly up-regulated in different tissues during stress treatment, suggesting that these genes might respond to stress through different tissues. Interestingly, these up-regulated genes were not only expressed in roots but also dominantly in leaves and stems (Fig. 8). Thus, we speculated that this was likely due to complex protein interactions that coordinated the expression of multiple genes through a network of feedback mechanisms [55].

Conclusion

Altogether, identification and systematic analysis of *HSF* genes in *Rye* showed that the 31 *ScHSF* genes were unevenly distributed on 7 chromosomes that were classified into seven subfamilies. By comprehensively analyzing gene structures and conserved motifs of 31 putative *ScHSF* genes, we found that motifs and gene structures of the same family were similar and might have the same biological functions. Furthermore, fragments and tandem repeats were the main drivers of novel functions in *ScHSF* family. Fragment repeats might have more substantial contributions to the evolution of *Rye HSF* genes. Overall, we performed a preliminary analysis of the structure of *HSF* gene family in *Rye* and further detailed its expression pattern. The results indicated that *ScHSF* gene family played a critical role not only in stem and flower development but also in hormonal and abiotic stress response during *Rye* development.

Materials and methods

Gene identification

The whole *Rye* genome was downloaded from the Ensembl website (<http://ensemblgenomes.org>). *HSF* gene family members were obtained based on two BLASTp approaches (PFAM and SMART) [56–58]. Firstly, all possible HSF proteins were identified using BLASTp (score value ≥ 100 , e value $\leq 1e-10$) with reference to the trihelix protein sequence of *Arabidopsis*. Secondly, the PFAM protein family database (<http://pfam.sanger.ac.uk>) was used to produce a Hidden Markov Model (HMM) with HSF domains, and then an HMM model cutoff value of 0.01 in HMMER 3.0 was applied to compare HSF protein sequences of *Rye* (<http://plants.ensembl.org/hmmer/index.html>). The availability of the HSF core sequence was confirmed using PFAM and the SMART program (<http://smart.emblheidelberg.de>). A total of thirty-one *HSF* genes were identified and then used as initial sequences to confirm HSF proteins (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) with BLASTp. Finally, several characteristics of the *HSF* genes, such as the sequence length, isoelectric point (pI), molecular weight (MW), and subcellular localization, were identified using ExPasy. A 2000 bp sequence upstream of the start codon (ATG) of the *ScHSF* gene was extracted from *Rye* genome using TBtools, followed by an analysis of the cis-acting elements using PlantCare (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html>). Finally, TFs were predicted through PlantTFDB and shown by using Cytoscape [59, 60].

HSF gene structure

Multiple protein sequence alignments based on the domain sequences in characterized HSF proteins of *A. thaliana* were created by ClustalW with default settings. The deduced amino acid sequences of HSF domains of different subfamilies were manually regulated using GeneDoc software and Mega 7.0. Furthermore, Gene Structure DiHSFay Server (<http://gsds.cbi.pku.edu.cn>) online program was applied to analyze the exon–intron structure of *HSF* genes. MEME Online Applications (<http://meme.nbcr.net/meme/intro.html>) were then employed to identify the protein sequences by adjusting the optimum motif width to 6~200 and the maximum number of motifs to 10.

Chromosomal distribution and gene duplication events

All *SchSF* genes were mapped to locations on different *Rye* chromosomes by using physical location information and handled using the Circos program [61]. The multiple collinear scanning toolkits (MCScanX) were then used, with default parameters, to analyze replication events of *SchSF* genes [62]. Finally, the *HSF* genes homology between *Rye* and six other plants (*O. sativa*, *Z. mays*, *H. vulgare*, *A. thaliana*, *T. aestivum*, and *A. tauschii* Coss) was measured using Dual Synteny Plotter (<https://github.com/CJ-Chen/TBtools>).

Phylogenetic analysis and classification of the SchSF family

With regard to the classification of AtHSFs, all identified SchSF proteins were first clustered into diverse groups. Next, a neighbor-joining (NJ) tree was built using Jukes-Cantor model in MEGA 7.0 [63]. The phylogenetic tree was then generated, with a bootstrap value of 1000, and was assigned by gene R11 and BLOSUM62 cost matrix. Moreover, we generated a multi-species phylogenetic evolutionary tree that included all HSF protein sequences from *Rye* as well as six others plants species (*O. sativa*, *Z. mays*, *H. vulgare*, *A. thaliana*, *T. aestivum*, and *A. tauschii* Coss). Notably, all protein sequences were downloaded from the UniProt database (<https://www.uniprot.org>). A protein–protein interaction analysis was performed on the STRING database (<http://string-db.org>) using SchSFs as the queries and *T. aestivum* proteins as references. Promoter cis-acting elements were predicted by both PlantCare and PlantTFDB [60, 64].

Plant materials, growth conditions, and different abiotic stress in *Rye*

The rye seed used in the experiment was provided by Fan Yu from Guizhou University. Wei ning Rye is the variety we used. Rye plants were cultivated in pots containing a mixture of soil and vermiculite (1:1) in a growth room.

The growth room was maintained at a temperature regime of 25 °C during the 16-h daytime period and 20 °C during the 8-h nighttime period. The 3relative humidity in the growth room was set at 75%. After 21 days of growth, stress treatment was initiated. Fruit sampling was conducted when the first seed setting occurred, and subsequent samples were collected every other week for five consecutive harvests. After planting, leaves, roots, stems, grains, anthers, and styles were collected from five individual plants under the same growth environment. The samples were immediately stored in liquid nitrogen at -80 °C until further analysis. The expression pattern of 31 *HSF* genes under different stresses was explored. Specifically, the abiotic stress treatments, including salt treatment (5% sodium chloride), water immersion (full plant), drought treatment (30% PEG 6000), UV radiation (70W/cm², 220 V, 30W), high temperature treatment (40 °C), and low temperature treatment (4 °C), were applied at the seedling stage (after 21 days). Each stress treatment was replicated five times, and qRT-PCR analysis was performed after sampling at 1 h, 4 h, and 12 h, respectively. Hormone stress treatments were applied at seedling stage (after 21 days) with package expansions of ABA (100 μmol/L), IAA (100 μmol/L), and GA3 (100 μmol/L). Grain samples were collected at 7D, 14D, 21D, 28D, and 35D, and each treatment was replicated five times, respectively.

Total RNA extraction, cDNA reverse transcription, and qRT-PCR analysis

Total RNA was extracted using a plant RNA extraction kit (Vazyme Biotech) following the manufacturer's instructions. A cDNA library was constructed through reverse transcription of 1 mg RNA samples using 5 × HiScript[®] Reverse Transcriptase (vazymes) and 4 × gDNA (vazymes) kits in accordance with the manufacturer's protocol. The expression of some representative genes was then analyzed by qRT-PCR, with at least three biological replicates. The primers used were designed by Beacon Designer 7 (Additional file 8: Table S6). Relative mRNA expression was normalized to the actin gene (GADPH) mRNA expression as internal control and was calculated using the delta-delta Ct ($2^{-\Delta\Delta C_t}$) method [65].

Statistical analyses

JMP6.0 (SAS Institute) was used to perform analysis of variance (ANOVA) tests; multiple comparison tests of ANOVA results were performed using the least significant difference (LSD) method at two difference significance levels $p < 0.05^*$ and $p < 0.01^{**}$. Finally, histograms were generated using Origin version 8.0.

Abbreviations

<i>HSF</i>	Heat stress transcription factors
<i>SchSF</i>	Secale cereale L. HSF
qRT-PCR	Quantitative real-time polymerase chain reaction
<i>AtHSF</i>	<i>Arabidopsis thaliana</i> HSF
HMM	Hidden Markov Model
pI	Isoelectric point
LG	Linkage group
DPA	Days post anthesis

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04418-1>.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Additional file 5.

Additional file 6.

Additional file 7.

Additional file 8.

Additional file 9.

Acknowledgements

We thank all our colleagues for providing useful discussions and technical assistance. We are very grateful to the editor and reviewers for critically evaluating the manuscript and providing constructive comments for its improvement.

Authors' contributions

Yanyan Ren: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing-original draft. Rui Ma: Methodology, Software, Investigation, Writing-review & editing. Yue Fan: Methodology, Validation. Liang Feng, Muhua Xie: Validation. Long Chen, Hao Yang, Xiaobao Wei, Xintong Wang, Kouhan Liu: Investigation. Peng Cheng: Funding acquisition, Methodology, Supervision. Baotong Wang: Funding acquisition, Conceptualization, Supervision.

Funding

This research was supported by the National Key R&D Program of China (Grant No: 2021YFD1401000), National Key R&D Program of China (Grant No: 2021YFD1401005), Key Research and Development Projects of Shaanxi Province (Grant No: 2022NY-125), Key Research and Development Project of Shaanxi Province (2021ZDLNY01-01), National Natural Science Foundation of China (32072410).

Availability of data and materials

The whole genome sequence information of *rye* was obtained from the Ensembl genome website (<http://ensemblgenomes.org/>). In the experiment, the *rye* material used was provided by Yu Fan from Guizhou University. The datasets supporting the conclusions of this study are included in the article and its additional files.

Declarations**Ethics approval and consent to participate**

This article does not contain any studies involving human participants or animals performed by the authors. These methods were carried out in accordance with relevant guidelines and regulations. All experimental protocols were approved by the Northwest A&F University.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 23 February 2023 Accepted: 24 August 2023

Published online: 20 September 2023

References

- Martis MM, Zhou R, Haseneyer G, et al. Reticulate evolution of the *Rye* genome. *Plant Cell*. 2013;25(10):3685–98. <https://doi.org/10.1105/tpc.113.114553>.
- Bartłomiej S, Justyna RK, Ewa N. Bioactive compounds in cereal grains - occurrence, structure, technological significance and nutritional benefits - a review. *Food Sci Technol Int*. 2012;18(6):559–68. <https://doi.org/10.1177/1082013211433079>.
- Zhu F. Triticale: Nutritional composition and food uses. *Food Chem*. 2018;241:468–79. <https://doi.org/10.1016/j.foodchem.2017.09.009>.
- Crespo-Herrera LA, Garkava-Gustavsson L, Åhman I. A systematic review of *Rye* (*Secale cereale* L.) as a source of resistance to pathogens and pests in wheat (*Triticum aestivum* L.). *Hereditas*. 2017;154:14. Published 2017 May 25. <https://doi.org/10.1186/s41065-017-0033-5>.
- Schreiber M, Himmelbach A, Börner A, Mascher M. Genetic diversity and relationship between domesticated *Rye* and its wild relatives as revealed through genotyping-by-sequencing. *Evol Appl*. 2018;12(1):66–77. Published 2018 Mar 26. <https://doi.org/10.1111/eva.12624>.
- Bauer E, Schmutzer T, Barilar I, et al. Towards a whole-genome sequence for *rye* (*Secale cereale* L.). *Plant J*. 2017;89(5):853–69. <https://doi.org/10.1111/tpj.13436>.
- Li G, Wang L, Yang J, et al. A high-quality genome assembly highlights *Rye* genomic characteristics and agronomically important genes. *Nat Genet*. 2021;53(4):574–84. <https://doi.org/10.1038/s41588-021-00808-z>.
- Baillo EH, Kimotho RN, Zhang Z, Xu P. Transcription Factors Associated with Abiotic and Biotic Stress Tolerance and Their Potential for Crops Improvement. *Genes* (Basel). 2019;10(10):771. Published 2019 Sept 30. <https://doi.org/10.3390/genes10100771>.
- Murre C, McCaw PS, Baltimore D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell*. 1989;56(5):777–83. [https://doi.org/10.1016/0092-8674\(89\)90682-x](https://doi.org/10.1016/0092-8674(89)90682-x).
- Riechmann JL, Heard J, Martin G, et al. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science*. 2000;290(5499):2105–10. <https://doi.org/10.1126/science.290.5499.2105>.
- Guo M, Liu JH, Ma X, Luo DX, Gong ZH, Lu MH. The Plant Heat Stress Transcription Factors (HSFs): Structure, Regulation, and Function in Response to Abiotic Stresses. *Front Plant Sci*. 2016;7:114. Published 2016 Feb 9. <https://doi.org/10.3389/fpls.2016.00114>.
- Guo M, Lu JP, Zhai YF, Chai WG, Gong ZH, Lu MH. 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. Published 2015 June 19. <https://doi.org/10.1186/s12870-015-0512-7>.
- Harrison CJ, Bohm AA, Nelson HC. Crystal structure of the DNA binding domain of the heat shock transcription factor. *Science*. 1994;263(5144):224–7. <https://doi.org/10.1126/science.8284672>.
- Nover L, Scharf KD, Gagliardi D, Vergne P, Czarnecka-Verner E, Gurley WB. The Hsf world: classification and properties of plant heat stress transcription factors. *Cell Stress Chaperones*. 1996;1(4):215–23. [https://doi.org/10.1379/1466-1268\(1996\)001%3c0215:thwcap%3e2.3.co;2](https://doi.org/10.1379/1466-1268(1996)001%3c0215:thwcap%3e2.3.co;2).
- Schultheiss J, Kunert O, Gase U, Scharf KD, Nover L, Rüterjans H. Solution structure of the DNA-binding domain of the tomato heat-stress transcription factor HSF24. *Eur J Biochem*. 1996;236(3):911–21. <https://doi.org/10.1111/j.1432-1033.1996.00911.x>.
- Giorno F, Guerriero G, Baric S, Mariani C. Heat shock transcriptional factors in *Malus domestica*: identification, classification and expression analysis. *BMC Genomics*. 2012;13:639. Published 2012 Nov 20. <https://doi.org/10.1186/1471-2164-13-639>.
- Kotak S, Port M, Ganguli A, Bicker F, von Koskull-Döring 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(1):98–112. <https://doi.org/10.1111/j.1365-313X.2004.02111.x>.
18. Nover L, Bharti K, Döring P, Mishra SK, Ganguli A, Scharf KD. Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need? *Cell Stress Chaperones.* 2001;6(3):177–89. [https://doi.org/10.1379/1466-1268\(2001\)006%3c0177:aathst%3e2.0.co;2](https://doi.org/10.1379/1466-1268(2001)006%3c0177:aathst%3e2.0.co;2).
 19. Scharf KD, Berberich T, Ebersberger I, Nover L. The plant heat stress transcription factor (Hsf) family: structure, function and evolution [published correction appears in *Biochim Biophys Acta.* 2017 Dec 13;]. *Biochim Biophys Acta.* 2012;1819(2):104–19. <https://doi.org/10.1016/j.bbaggm.2011.10.002>.
 20. Wang F, Dong Q, Jiang H, Zhu S, Chen B, Xiang Y. Genome-wide analysis of the heat shock transcription factors in *Populus trichocarpa* and *Medicago truncatula*. *Mol Biol Rep.* 2012;39(2):1877–86. <https://doi.org/10.1007/s11033-011-0933-9>.
 21. 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 Genomics.* 2019;20(1):257. Published 2019 Apr 1. <https://doi.org/10.1186/s12864-019-5617-1>.
 22. Talaei L, Fathipour Y, Talebi AA, Khajehali J. Screening of Potential Sources of Resistance to *Spodoptera exigua* (Lepidoptera: Noctuidae) in 24 Sugar Beet Genotypes. *J Econ Entomol.* 2017;110(1):250–8. <https://doi.org/10.1093/jee/tow257>.
 23. Liu GT, Chai FM, Wang Y, et al. Genome-wide Identification and Classification of HSF Family in Grape, and Their Transcriptional Analysis under Heat Acclimation and Heat Stress. *Horticultural Plant J.* 2018;4(4):133–43.
 24. Mittal D, Chakrabarti S, Sarkar A, Singh A, Grover A. Heat shock factor gene family in rice: genomic organization and transcript expression profiling in response to high temperature, low temperature and oxidative stresses. *Plant Physiol Biochem.* 2009;47(9):785–95. <https://doi.org/10.1016/j.plaphy.2009.05.003>.
 25. Reddy PS, Kavi Kishor PB, Seiler C, et al. Unraveling regulation of the small heat shock proteins by the heat shock factor HvHsfB2c in barley: its implications in drought stress response and seed development. *PLoS One.* 2014;9(3):e89125. Published 2014 Mar 4. <https://doi.org/10.1371/journal.pone.0089125>.
 26. Jiang L, Hu W, Qian Y, Ren Q, Zhang J. Genome-wide identification, classification and expression analysis of the Hsf and Hsp70 gene families in maize. *Gene.* 2021;770:145348. <https://doi.org/10.1016/j.gene.2020.145348>.
 27. Scharf KD, Rose S, Zott W, Schöffl F, Nover L. Three tomato genes code for heat stress transcription factors with a region of remarkable homology to the DNA-binding domain of the yeast HSF [published correction appears in *EMBO J* 1991 Apr;10(4):1026. Schöffl F [corrected to Schöffl F]]. *EMBO J.* 1990;9(13):4495–501. <https://doi.org/10.1002/j.1460-2075.1990.tb07900.x>.
 28. Begum T, Reuter R, Schöffl F. Overexpression of AtHsfB4 induces specific effects on root development of Arabidopsis. *Mech Dev.* 2013;130(1):54–60. <https://doi.org/10.1016/j.mod.2012.05.008>.
 29. Mishra SK, Tripp J, Winkelhaus S, et al. 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(12):1555–67. <https://doi.org/10.1101/gad.228802>.
 30. Zhu B, Ye C, Lü H, et al. Identification and characterization of a novel heat shock transcription factor gene, GmHsfA1, in soybeans (*Glycine max*). *J Plant Res.* 2006;119(3):247–56. <https://doi.org/10.1007/s10265-006-0267-1>.
 31. Fragkostefanakis S, Mesihovic A, Simm S, et al. HsfA2 Controls the Activity of Developmentally and Stress-Regulated Heat Stress Protection Mechanisms in Tomato Male Reproductive Tissues. *Plant Physiol.* 2016;170(4):2461–77. <https://doi.org/10.1104/pp.15.01913>.
 32. Zhuang L, Cao W, Wang J, Yu J, Yang Z, Huang B. Characterization and Functional Analysis of FaHsfC1b from *Festuca arundinacea* Conferring Heat Tolerance in Arabidopsis. *Int J Mol Sci.* 2018;19(9):2702. Published 2018 Sept 11. <https://doi.org/10.3390/ijms19092702>.
 33. Cenci A, Rouard M. Evolutionary Analyses of GRAS Transcription Factors in Angiosperms. *Front Plant Sci.* 2017;8:273. Published 2017 Mar 2. <https://doi.org/10.3389/fpls.2017.00273>.
 34. Fan Y, Yan J, Lai D, et al. Genome-wide identification, expression analysis, and functional study of the GRAS transcription factor family and its response to abiotic stress in sorghum [*Sorghum bicolor* (L.) Moench]. *BMC Genomics.* 2021;22(1):509. Published 2021 July 6. <https://doi.org/10.1186/s12864-021-07848-z>.
 35. Chen F, Hu Y, Vannozzi A, et al. The WRKY Transcription Factor Family in Model Plants and Crops. *Crit Rev Plant Sci.* 2018;36(5):1–25.
 36. Lu J, Ju H, Zhou G, et al. An EAR-motif-containing ERF transcription factor affects herbivore-induced signaling, defense and resistance in rice. *Plant J.* 2011;68(4):583–96. <https://doi.org/10.1111/j.1365-313X.2011.04709.x>.
 37. Lorenzo O, Piqueras R, Sánchez-Serrano JJ, Solano R. ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. *Plant Cell.* 2003;15(1):165–78. <https://doi.org/10.1105/tpc.007468>.
 38. Pré M, Atallah M, Champion A, De Vos M, Pieterse CM, Memelink J. The AP2/ERF domain transcription factor ORA59 integrates jasmonic acid and ethylene signals in plant defense. *Plant Physiol.* 2008;147(3):1347–57. <https://doi.org/10.1104/pp.108.117523>.
 39. 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(12):3373–83. <https://doi.org/10.1093/jxb/erm184>.
 40. Kumar M, Busch W, Birke H, Kemmerling B, Nürnberger T, Schöffl F. Heat shock factors HsfB1 and HsfB2b are involved in the regulation of Pdf1.2 expression and pathogen resistance in Arabidopsis. *Mol Plant.* 2009;2(1):152–65. <https://doi.org/10.1093/mp/ssn095>.
 41. Liao Y, Liu Z, Gichira AW, et al. Deep evaluation of the evolutionary history of the Heat Shock Factor (HSF) gene family and its expansion pattern in seed plants. *PeerJ.* 2022;10:e13603. Published 2022 Aug 9. <https://doi.org/10.7717/peerj.13603>.
 42. Chauhan H, Khurana N, Agarwal P, Khurana JP, Khurana P. A seed preferential heat shock transcription factor from wheat provides abiotic stress tolerance and yield enhancement in transgenic Arabidopsis under heat stress environment. *PLoS One.* 2013;8(11):e79577. Published 2013 Nov 12. <https://doi.org/10.1371/journal.pone.0079577>.
 43. Shabalina SA, Ogurtsov AY, Spiridonov AN, Novichkov PS, Spiridonov NA, Koonin EV. Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes. *Mol Biol Evol.* 2010;27(8):1745–9. <https://doi.org/10.1093/molbev/msq086>.
 44. Sang Y, Liu Q, Lee J, Ma W, McVey DS, Blecha F. Expansion of amphibian intronless interferons revises the paradigm for interferon evolution and functional diversity. *Sci Rep.* 2016;6:29072. Published 2016 June 30. <https://doi.org/10.1038/srep29072>.
 45. Jain M, Khurana P, Tyagi AK, Khurana JP. Genome-wide analysis of intronless genes in rice and Arabidopsis. *Funct Integr Genomics.* 2008;8(1):69–78. <https://doi.org/10.1007/s10142-007-0052-9>.
 46. Rogozin IB, Sverdlov AV, Babenko VN, Koonin EV. Analysis of evolution of exon-intron structure of eukaryotic genes. *Brief Bioinform.* 2005;6(2):118–34. <https://doi.org/10.1093/bib/6.2.118>.
 47. Del Campo EM, Casano LM, Barreno E. Evolutionary implications of intron-exon distribution and the properties and sequences of the RPL10A gene in eukaryotes. *Mol Phylogenet Evol.* 2013;66(3):857–67. <https://doi.org/10.1016/j.ympev.2012.11.013>.
 48. Mascagni F, Usai G, Cavallini A, Porceddu A. Structural characterization and duplication modes of pseudogenes in plants. *Sci Rep.* 2021;11(1):5292. Published 2021 Mar 5. <https://doi.org/10.1038/s41598-021-84778-6>.
 49. Zhou Q, Zhang S, Chen F, et al. Genome-wide identification and characterization of the SBP-box gene family in *Petunia*. *BMC Genomics.* 2018;19(1):193. Published 2018 Mar 12. <https://doi.org/10.1186/s12864-018-4537-9>.
 50. Xie K, Wu C, Xiong L. Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiol.* 2006;142(1):280–93. <https://doi.org/10.1104/pp.106.084475>.
 51. Zhang SD, Ling LZ, Yi TS. Evolution and divergence of SBP-box genes in land plants. *BMC Genomics.* 2015;16:787. Published 2015 Oct 14. <https://doi.org/10.1186/s12864-015-1998-y>.
 52. Zhang Q, Geng J, Du Y, et al. Heat shock transcription factor (Hsf) gene family in common bean (*Phaseolus vulgaris*): genome-wide identification, phylogeny, evolutionary expansion and expression analyses at the sprout stage under abiotic stress. *BMC Plant Biol.* 2022;22(1):33. Published 2022 Jan 14. <https://doi.org/10.1186/s12870-021-03417-4>.

53. Jiao SZ, Yao WK, Zhang NB, et al. Research progress of heat stress transcription factors(Hsfs) in horticultural plants. *J Fruit Sci.* 2020;37(3):419–30.
54. Chang YY, Liu HC, Liu NY, et al. A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermotolerance in Arabidopsis. *Plant Physiol.* 2007;143(1):251–62. <https://doi.org/10.1104/pp.106.091322>.
55. Li J, Mahajan A, Tsai MD. Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry.* 2006;45(51):15168–78. <https://doi.org/10.1021/bi062188q>.
56. Liu M, Ma Z, Wang A, et al. Genome-Wide Investigation of the Auxin Response Factor Gene Family in Tartary Buckwheat (*Fagopyrum tataricum*). *Int J Mol Sci.* 2018;19(11):3526. Published 2018 Nov 9. <https://doi.org/10.3390/ijms19113526>.
57. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL. The Pfam protein families database. *Nucleic Acids Res.* 2000;28(1):263–6. <https://doi.org/10.1093/nar/28.1.263>.
58. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 2018;46(D1):D493–6. <https://doi.org/10.1093/nar/gkx922>.
59. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504. <https://doi.org/10.1101/gr.1239303>.
60. Jin J, Tian F, Yang DC, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 2017;45(D1):D1040–D1045. <https://doi.org/10.1093/nar/gkw982>.
61. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos an information aesthetic for comparative genomics. *Genome Res.* 2009;19(9):1639–45.
62. Wang Y, Tang H, Debarry JD, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40(7):e49. <https://doi.org/10.1093/nar/gkr1293>.
63. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 2013;30(12):2725–9. <https://doi.org/10.1093/molbev/mst197>.
64. Lescot M, Déhais P, Thijs G, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 2002;30(1):325–7. <https://doi.org/10.1093/nar/30.1.325>.
65. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔC_T} method. *Meth-ods.* 2001;25(4):402–8. <https://doi.org/10.1006/meth.2001.1262>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

