

# Genome-wide characterization and evolutionary analysis of heat stress transcription factors (HSFs) to reveal their potential role under abiotic stresses in radish (*Raphanus sativus* L.)

**Mingjia Tang**

Nanjing Agricultural University

**Liang Xu**

Nanjing Agricultural University

**Yan Wang**

Nanjing Agricultural University

**Wanwan Cheng**

Nanjing Agricultural University

**Xiaobo Luo**

Nanjing Agricultural University

**Yang Xie**

Nanjing Agricultural University

**Lianxue Fan**

Nanjing Agricultural university

**Liwang Liu** (✉ [nauliulw@njau.edu.cn](mailto:nauliulw@njau.edu.cn))

College of Horticulture, Nanjing Agricultural University <https://orcid.org/0000-0001-7355-4408>

---

## Research article

**Keywords:** Abiotic stress, heat shock transcription factors (HSFs), radish, Reverse-transcription quantitative PCR (RT-qPCR)

**Posted Date:** September 19th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.10724/v2>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on October 24th, 2019. See the published version at <https://doi.org/10.1186/s12864-019-6121-3>.

# Abstract

Background Abiotic stresses due to climate change pose a great threat to crop production. Heat shock transcription factors (HSFs) are vital regulators that play key roles in protecting plants against various abiotic stresses. Therefore, the identification and characterization of HSFs is imperative to dissect the mechanism responsible for plant stress responses. Although the HSF gene family has been extensively studied in several plant species, its characterization, evolutionary history and expression patterns in the radish (*Raphanus sativus* L.) remain limited. Results In this study, 33 RsHSF genes were obtained from the radish genome, which were classified into three main groups and 12 subgroups based on HSF protein domain structure. Chromosomal localization analysis revealed that 28 of 33 RsHSF genes were located on nine chromosomes, and 10 duplicated RsHSF genes were grouped into eight gene pairs by whole genome duplication (WGD). Moreover, there were 23 or 9 pairs of orthologous HSFs were identified between radish and *Arabidopsis* or rice, respectively. Comparative analysis revealed a close relationship among radish, Chinese cabbage and *Arabidopsis*. RNA-seq data showed that eight RsHSF genes, including RsHSF-03, were highly expressed in the leaf, root, cortex, cambium and xylem, results that these genes might be involved in plant growth and development. Further, quantitative real-time polymerase chain reaction (RT-qPCR) indicated that the expression patterns of 12 RsHSF genes varied upon exposure to different abiotic stresses, including heat, salt, and heavy metals. This data indicated that the RsHSFs may be involved in abiotic stress response. Conclusions These results could provide fundamental insights into the characteristics and evolution of the HSF family and facilitate further dissection of the molecular mechanism responsible for radish abiotic stress responses.

## Background

Many inevitable environmental factors (e.g., heat, drought, flooding, salinity and heavy metals) trigger abiotic stress, affect plant growth and consequently increase crop losses [1-3]. Heat stress (HS), one of the major abiotic stresses, severely inhibits crop growth and development. These effects have had devastating economic impacts on the yield and quality of rice, wheat, maize and vegetable crops [4-6]. Heat shock transcription factors (HSFs) are important regulators that could contribute to controlling the differential expression of heat shock proteins (HSPs) and other functional genes in the process of protecting plants from heat and other stresses including chilling, salinity, drought and heavy metal [7,8]. It is imperative to clarify the molecular mechanisms that govern how plants respond and adapt to HS.

Plants have developed multiple defense mechanisms and strategies to cope with adverse conditions and respond accordingly [9-11]. Under abiotic stresses, the induction of myriad proteins, including transcription factors (TFs), could regulate the expression of specific functional genes to enhance plant resistance through signal transduction pathways. Reactive oxygen species (ROS)-scavenging enzymes and HSPs are important functional proteins induced by HS, and their corresponding genes are targets of several HS-responsive TFs [12-14]. Previous studies indicate that *HSFA6b* vitally acts as a downstream regulator of the ABA-mediated heat stress response (HSR) in *Arabidopsis thaliana* [15]. Additionally, in the absence of HS, HSP genes expression is induced by the overexpression of constitutively active HsfA1d

without temperature-dependent repression (TDR) domain, thus conferring effective thermotolerance in *Arabidopsis* [16]. In the tomato, cadmium (Cd) tolerance is conferred by HsfA1a through activating *COMT1* gene transcription, which partially upregulates HSP expression by inducing melatonin in accumulation [8].

In plants, the HSFs have a conserved modular structure with considerable size and sequence variability [9, 17-19]. For example, all HSFs have an N-terminal DNA-binding domain (DBD) and an adjacent bipartite oligomerization domain (HR-A/B), while only a proportion contain a nuclear localization signal (NLS) domain, nuclear export signal (NES) domain and C-terminal activator domain (CTAD) [9]. According to the peculiarities of the adjacent hydrophobic amino acid residues (HR-A/B), plant HSFs can be divided into three types, class A, B and C [21,22]. Numerous HSF families have been isolated from a range of plant species, including Chinese cabbage (*Brassica.rapa ssp. pekinensis*) [23], *Arabidopsis*[24], carrot (*Daucus carota*) [13], peanut (*Arachis hypogaea*) [7], rice (*Oryza sativa*) [25], maize (*Zea mays*) [26], Chinese white pear (*Pyrus bretschneideri*) [27], poplar (*Populus trichocarpa*) and barrelclover (*Medicago truncatula*) [28]. However, the comprehensive identification of HSF genes remains unknown in root vegetable crops, including radish.

Radish (*Raphanus sativus* L.,  $2n = 18$ ), belonging to the family Brassicaceae, is one of the most economically important annual or biennial root vegetable crops [29]. Previous studies showed that 26 known and 19 novel microRNAs (miRNAs) are differentially expressed under HS in radish roots [4]. Moreover, there are several differentially expressed genes (DEGs) involved in heat stress response process in radish [30]. Recently, the available radish genome sequence provided a useful resource for whole-genome identification of TF families [31]. Nevertheless, little information on systematic characterization of HSF genes and their families in response to abiotic stresses is available in the radish. This study aimed to to isolate the full-length RsHSF sequences from the radish genome, map the RsHSFs onto chromosomes and explore their expression in response to abiotic stresses. Moreover, RsHSF orthologs and paralogs were obtained and the expression pattern of RsHSFs in different radish tissues was investigated. The outcomes of this study provide an opportunity to further explore the roles of HSF genes involved in abiotic stresses and present the expansion and evolutionary history of the HSF gene family in the radish.

## Results

## Results

# Identification and characterization of radish HSF proteins

To comprehensively identify the candidate HSF proteins in radish, a profile Hidden Markov Model (HMM) [32] search against NODAI radish genome protein sequences with the HSF domain (Pfam: PF00447) was performed. A total of 55 putative RsHSFgene sequences were obtained from the whole genome. SMART

(<http://smart.embl-heidelberg.de/>) and Pfam (<http://pfam.xfam.org/>) were used to remove proteins with incomplete hsf-type DBD domain, while MARCOIL (<http://toolkit.tuebingen.mpg.de/marcoil>) was used to confirm OD (HR-A/B) domain presence. After removing sequences that encoded proteins without the complete DBD, OD (HR-A/B) and/or start/stop codons, 33 RsHSF genes remained for further analysis (Additional file 1: Table S1-S3). Subsequently, the conserved domains of all retained genes were identified by Heatster. The DBD, which consists of approximately 100 amino acids (AAs) at the N-terminus, was the most conserved domain. Overall, 31 AAs of the DBD domain were highly conserved among all RsHSFs (Additional file 2: Figure S1). NLS and NES are crucial for HSF intracellular distribution between the nucleus and cytoplasm. All RsHSFs had an NES, whereas only 13 contained the NLS (Additional file 3: Table S4).

Furthermore, the physical and chemical properties of 33 RsHSF proteins were analyzed with ExPASy (Additional file 3: Table S4). Among these, the RsHSF protein sizes varied from 238 to 2,427 AAs with molecular weights (MWs) from 27.80 to 274.05 KDa, respectively. Additionally, the theoretical pI of most RsHSF proteins was < 7, with the exception *RsHSF-04*, *RsHSF-05*, *RsHSF-06* and *RsHSF-25*. All RsHSFs were classified as unstable proteins according to the instability index. The aliphatic index may be a positive factor that increased globular protein thermostability. *RsHSF-25* had the largest aliphatic index (77.14). Most RsHSFs showed similar physical and chemical properties, while different AA sequences in non-conserved regions may alter some of the molecular characteristics.

## Intron–exon structure and conserved motif distribution

To obtain information about the diversity of RsHSF gene structure, full-length complementary DNA (cDNA) sequences were compared with the corresponding genomic DNA sequences through GSDS. To further analyze RsHSF protein motif distribution, whole sequences were subjected to the MEME web server and a total of 25 AA motifs were generated (Additional file 2: Figure.S2).

Overall, 25 of 33 RsHSF genes had one intron, while the remaining members have two or more introns. Additionally, proteins within the same subgroups had similar motif structures. All RsHSF proteins harbored motifs 1, 2 and 4, all of which highly correspond to the conserved DBD. Moreover, *RsHSF-07*, *RsHSF-14* and *RsHSF-26* in subgroup A6 presented a similar gene and motif structure. However, some motifs were only detected in specific RsHSF protein classes. For example, motif 3 was identified in classes A and C, while motifs 9 and 17 only existed in class A1. A proportion of RsHSF proteins in the same class exhibited similar motif distribution, indicating that these proteins might have conserved functions (Fig.1).

## Comparative and phylogenetic analysis of RsHSF genes

To investigate the distribution of HSF subfamilies and the evolutionary relationship among different species, 33 RsHSFs combined with 35 BrHSFs, 20 AcHSFs, 25 OsHSFs and 21 AtHSFs identified from

Chinese cabbage (*B. rapa* ssp. *pekinensis*) [33], pineapple (*A.comosus*) [34,35], rice (*O. sativa*) [36] and *Arabidopsis* [37] were employed for comparative analysis (Additional file 4: Table S5). Based on the number of AA residues between the A and B portions of the HR-A/B region, RsHSFs were classified into three classes, namely A, B and C. Intriguingly, the motifs 5 and 19 were only identified in class B (Fig.1).

A phylogenetic tree was constructed using neighbor-joining and maximum likelihood methods by MEGA6 [38] with the full sequences of the 134 HSF proteins (Fig.2). All proteins were categorised into three groups corresponding to the HSF classes A, B and C. Additionally, there were 9 RsHSF class A subgroups clearly obtained on the basis of the bootstrap values and phylogenetic analysis of *Arabidopsis* and rice. According to the phylogenetic tree, class A consisted of 8 subclasses (A1, A3, A4, A5, A6, A7, A8, and A9) and contained the most RsHSF numbers, while class C accounted for the least proportion of HSFs among these five species. Interestingly, no RsHSFs, AtHSFs or BrHSFs clustered into classes A2, whereas AcHSFs did not appear in classes A6 and A7. This finding suggests that most RsHSF genes are more closely related to their corresponding homologous genes in Chinese cabbage and *Arabidopsis*.

## Chromosomal location and duplication of RsHSF genes

To obtain the chromosomes distributions of RsHSF genes, their DNA sequences were physically plotted on the chromosomes through blast searches against the genomic sequences (Additional file 5: Table S6). In total, 28RsHSF genes were mapped on the chromosomes with uneven distribution, with exception of five genes (*RsHSF-22*, *23*, *25*, *29* and *33*) (Fig.3). Chromosome (Chr.) 5 had the largest number of RsHSF genes among three classes, followed by Chr. 4 (5 RsHSF genes from class A and B). Only one RsHSF gene was present in Chr. 8, while Chr. 1, 6 and 9 harbored two genes. Moreover, all the class B RsHSF genes were mapped on the chromosomes.

Genome duplications have been recognized and contributed to the expansion of gene family in plants [39, 40]. MCScanX was used to obtain information about origins of duplicate RsHSF genes in radish genome [41]. 13 (46.4%) RsHSF genes were duplicated and retained from a whole genome duplication (WGD) event and 11 (39.3%) were duplicated and retained from a singleton event (Additional file 6: Table S7). Eight WGD events of 10 duplicated RsHSF genes were identified and classified into four groups. Among these four groups, two harbored three genes (*RsHSF-07*, *RsHSF-14* and *RsHSF-26*; *RsHSF-08*, *RsHSF-16* and *RsHSF-17*), and these six genes were located on five chromosomes (Chr.2, 4, 5, 6 and 9). The other two groups contained two WGD events of four genes, and one duplication event harbored *RsHSF-01* and *RsHSF-03* which were located on Chr. 2 and 4, respectively (Fig.3). Only one WGD event (*RsHSF-07*and *RsHSF-14*) took place within the same chromosome (Fig.3, Additional file 7: Table S8). These results revealed that WGD or segmental duplication played an important role in the expansion of the radish HSF gene family.

## Identification of orthologous and paralogous HSF genes

Polyploidization events contributed significantly to the evolution of flowering plants [42]. Previous studies showed that the common ancestor of *Brassica* and *Raphanus* have experienced ' whole-genome triplication (WGT) event since its divergence from *Arabidopsis*. However, the gene losses of orthologous groups between *Arabidopsis*, *Brassica* and *Raphanus* for protein-coding genes occurred in both the *Brassica* and *Raphanus* lineages [39, 42]. Orthologous HSF genes among radish, *Arabidopsis* and rice were identified for triplication assessment using the Orthomcl-pipeline [43]. In total, there were 23 pairs of orthologous HSFs between radish and *Arabidopsis*, and eight paralogous gene pairs were identified in radish (Additional file 8: Table S9). It was reported that there are only six orthologous gene pairs between *Arabidopsis* and rice [23]. In addition, seven orthologous HSF gene pairs were identified between radish and rice (Additional file 8: Table S9). Among the orthologous HSF gene pairs between radish and *Arabidopsis*, *AtHSF-11* and *AtHSF-20* had three orthologous genes, three *AtHSFs* have two orthologous genes and 11 *AtHSFs* only had one orthologous gene (Additional file 2: Fig S3). These results indicate that many orthologous groups experienced gene losses. Moreover, five genes (*RsHSF-01*, *RsHSF-12*, *RsHSF-13*, *RsHSF-20* and *RsHSF-32*) had no orthologous gene in either *Arabidopsis* or rice. These results provide useful resource and reference for further exploring the functions of RsHSF genes in radish (Fig.4).

## Expression pattern of RsHSF genes in different tissues

To determine spatiotemporal expression patterns of RsHSF genes, the Reads Per Kilobase Per Million (RPKM) values of the 33 RsHSF genes in different tissues and developmental stages were collected from RNA-Seq data and presented in a heatmap (Fig.5, Additional file 9: Table S10). In total, the expression profiling showed that not all RsHSF genes were expressed. The number of RsHSF genes in different tissues and developmental stages with RPKM >1 was 14 (42.4%). Almost all RsHSFA1s RPKM values were < 10 among different tissues and stages, except for *RsHSF-18* and *RsHSF-24* that were highly expressed in the cortex, cambium and xylem. The majority of RsHSF genes exhibited differential expression patterns in the various tissues and developmental stages. *RsHSF-21* and *RsHSF-29* (class C) were relatively lowly expressed during all developmental stages. *RsHSF-21* and *RsHSF-29* (class C) were relatively lowly expressed during all developmental stages. In addition, *RsHSF-13* expression increased significantly in both root and leaf after 7 days after sowing (DAS), and it was in leaves after 40 DAS. *RsHSF-30* (class A8) expression was higher compared to *RsHSF-28* (class A3) during all developmental stages. Besides, RsHSFB genes exhibited stage-specific expression patterns. Among class B, both *RsHSF-04* and *RsHSF-10* were lowly expressed during all stages, while *RsHSF-03* was highly expressed.

## RsHSF gene expression profile under abiotic stresses

RNA-Seq data showed that some HSF genes were differentially expressed under disparate stresses. The transcription levels ( $\log_2$  (Fold Change) and  $\log_2$ TPM value) of certain RsHSF genes from the NCBI SRA were further analyzed (Additional file11: Table S12; Additional file 2: Fig S5). Several HSF genes, including HSFA1s were up-regulated in response to heat and salt stresses. RsHSFA4s (*RsHSF-08*, *RsHSF-*

*16* and *RsHSF-22*) were up-regulated under salt and lead (Pb) stresses, while they were down-regulated under HS. Additionally, HSFC1 (*RsHSF-21*) expression increased after Cd and salt treatments. HSFA1s as master regulators play vital roles in the HSR and activation of transcriptional networks. Expression levels of genes that encode HS responsive TFs including *DREB2A*, HSFA7s and HSFs, are directly regulated by HSFA1s [44]. *HsfA4a* is induced by oxidative stress, including HS, and regulates *APX1* expression [45]. To gain insight into the expression patterns of HSF genes in radish, 12 RsHSF genes that belonged to different groups were verified by RT-qPCR analysis under heat, salt, Pb, and Cd stress (Additional file 10: Table S11). Overall, the gene expression patterns varied significantly under the various treatments (Fig.6). Most genes were up-regulated after 24h HS exposure, while some genes were down-regulated after HS for 6h (Fig.6), such as *RsHSF-16*, *RsHSF-22* and *RsHSF-29*. Moreover, *RsHSF-18* expression was highly increased after 6h HS, implying that it can quickly respond to this stress. Further, *RsHSF-18* was rapidly up-regulated under Cd treatment and during all heat and salt stress time points, while *RsHSF-33* was only up-regulated in response to Pb stress. *RsHSF-05*, *RsHSF-06*, and *RsHSF-33* were up-regulated under 200mg/L Pb treatment and are likely involved in the Pb stress response. *RsHSF-11* and *RsHSF-22*, which were down-regulated under Cd stress, exhibited relatively high expression level under salt stress and thus they may contribute to salt tolerance.

## Discussion

Radish (*Raphanus sativus* L.), an ancient cultivated crop worldwide, is an important human dietary component. Industrial development, climate change and increased areas of contaminated soil considerably affect the expansion and healthy cultivation of the radish. The HSFs family allows plants to cope with abiotic stresses (e.g. heat, Cd and high light) by regulating gene expression to prevent damage [15, 46]. Recently, the availability of increasingly complete genome sequencing has enabled the application of a bioinformatics approach to identify various gene families, including HSF in different species [27, 28]. However, the available information about the RsHSFs is limited. This study is the first comprehensive overview of the HSF gene family within the radish genome.

## Genome-wide identification and characterization of RsHSF genes

It is widely accepted that the HSF-type DBD and OD domains are necessary HSF structural components. The highly conserved DBD is composed of a three-helical bundle (H1, H2 and H3) and a four-stranded antiparallel  $\beta$ -sheet [9]. The HR-A/B, connected to the DBD by a flexible linker of variable length (15–80 AAs), is comprised of hydrophobic heptad repeats [47, 48]. In brief, the DBD ensures HSFs combine with HSE, and the OD is the basis for differentiating the three HSF classes. It is reported that the DNA-binding domain of plant HSF is separated by a single intron. Although the position of this intron is unanimous in all HSF DBDs, the intron size is variable [21]. Therefore, we predicted the conserved domain using Heatster online tools [49] after performing Pfam and MARCOIL analysis. Consequently, 33 RsHSF genes were identified through genome-wide analysis (Additional file 1: Table S1-S3).

Among five plant species, 134 HSFs divided into three major classes were used to analyze the relationships. The results suggested that HSFs originated prior to the divergence of these species. Phylogenetic analysis evidenced that RsHSFs were more similar to BrHSFs and AtHSFs than OsHSFs, which corresponds with the fact that the radish, Chinese cabbage and *Arabidopsis* belong to Brassicaceae family (Fig.2). The radish HSF number (33) was larger than that in *Arabidopsis* (21) and pineapple (20). Previous studies showed that distinct plants species harbor different numbers of HSFs: land plants have more TFs than algae. These results indicate that the number of HSFs is probably related to the evolution and the growth environment [13].

## The evolutionary characterization of the RsHSF family

Gene duplication plays a major role in the evolution of novel gene functions and the expansion of gene families [36]. Among the evolutionary process in plants, gene duplication that increases the genome size and relaxes selection on one gene copy, is regarded as the primary driving force and allows the acquisition of new function, that enhance environmental adaptability [13, 50]. In Brassicaceae, there were three WGDs after an *Arabidopsis* lineage diverged from the monocot lineage. The most recent WGD event occurred 50-65 million years ago, which was earlier than the divergence of plants in the Brassicaceae family [40, 51, 52]. Many gene families are expanded in higher plants, including *CML* [53], *ALDHs* [54], *MADS-Box* [55], *NAC* [56] and *bHLH* [57]. Current evidence shows that the evolution patterns of ' duplicates are similar in *Raphanus* and *Brassica* [40]. For RsHSFs evolution, the expansion was due to eight duplicated pairs, which are consistent with those in Chinese cabbage. Previous data showed that HSF gene family evolution is related to the polyploid copy [23].

Polyploidization events apparently occurred 3-12 million years after independence of the *Arabidopsis*–*Brassica* lineages, and the time of the *Raphanus* genus diverging from *Brassica* was longer than previously predicted [58, 59]. Furthermore, recent reports revealed that ~60% of genes that belong to the neopolyploid ancestor of *Raphanus* and *Brassica* disappeared due to a WGT event. However, a considerable number of genes still exist within the *Raphanus* and *Brassica* genomes, a phenomenon that represents the benefit of this evolutionary novelty [40]. Phylogenetic analysis of HSF genes among radish, *Arabidopsis* and Chinese cabbage indicated that many RsHSF genes are highly similar to their corresponding *Arabidopsis* and Chinese cabbage homologs. Moreover, a large number of the identified RsHSFs were detected as orthologous genes in *Arabidopsis*, including *RsHSF-04*, *RsHSF-05* and *RsHSF-06*, which were the orthologous genes of *AtHSF-20* (Fig.4). These results provide valuable cues on further understanding the functions of these highly homologous genes in radish. For instance, *RsHSF-07*, *RsHSF-14* and *RsHSF-26* were highly similar to *AtHSF-11*, which is vital for HS tolerance and is a downstream regulator during the ABA-mediated stress response [15]. In this study, we identified eight pairs of RsHSF paralogs that might be related to the extended genome triplication of the radish. Collectively, the RsHSF gene family expansion may be largely related with gene duplication, and the identification of the orthologous gene pairs between radish and *Arabidopsis* provide a reference for exploring the roles of HSFs under different stresses in radish.

# The potential roles of differentially expressed RsHSF genes

HSFs play important roles in plants in response to abiotic stresses by regulating the expression of different genes [13, 23, 60]. Among HSF genes in plants, *HsfA1* (a major transcriptional activator) is necessary to evoke the HSR. Indeed, the *HsfA1*-knockout mutant shows reduced expression of many HS-induced genes [45, 61]. Moreover, *HsfA1a* is associated with Cd tolerance by improving melatonin biosynthesis [8]. In this study, RsHSF genes were significantly differentially expressed in response to various stresses. RsHSFA1s (*RsHSF-18* and *RsHSF-25*) showed converse expression levels upon 24 h HS or 200mg/L Pb, which was significantly down-regulated under Pb stress, indicating that the expression of *HSFA1* may be repressed under 200mg/L Pb stress. In addition, *RsHSF-18* and *RsHSF-25* were differentially expressed under 50mg/L Cd stress, which may be related to the specific function in the Cd stress response. *HsfA4A* enhances Cd tolerance in wheat and rice and salt protection in *Arabidopsis* [62]. Our results showed that RsHSFA4s (*RsHSF-08*, *RsHSF-16*, *RsHSF-17* and *RsHSF-22*) were up-regulated: only *RsHSF-16* was highly expressed upon salt and Cd treatment. These up-regulated genes might play specific roles in coping with stresses in radish plants. Unlike HSFAs, a considerable number of HSFBs and HSFCs are not reported to act as transcription activators, although tomato HSF1 apparently plays a role as a transcriptional co-activator of HsfA1 under HS [17, 44, 63]. In this study, the expression of RsHSFBs was distinct. For example, *RsHSF-08* and *RsHSF-18* had high expression level while *RsHSF-11* and *RsHSF-33* were down-regulated upon 24h HS. Intriguingly, the expression level of *RsHSF-29* (an HSFC), was down-regulated under all abiotic stresses except 200mg/L Pb treatment. *HSFA1d* and *HSFA3* were significantly up-regulated in radish under HS [30], while several HSFs, including *HSFC1* and *HSFB2a*, were notably up-regulated under salt stress [64]. Some heavy metal stress responsive signaling molecules are activated on detecting Cd<sup>2+</sup>, such as calcium-dependent protein kinases (CDPKs) and mitogen-activated protein kinase (MAPKs), both of which consequently regulate various metal-responsive TF families (eg., NACs and HSFs) in radish nucleus [65]. Moreover, four radish HSF genes were up-regulated under Cr stress in radish [66]. Taken together, several RsHSF genes were differentially expressed upon abiotic stresses including heat, salt or heavy metal stress, and these results indicate that they might be involved in the plant response to abiotic stresses.

RsHSF gene expression during different stages and in tissues is associated with abiotic stresses tolerance. In this study, RsHSF genes exhibited tissue-specific expression patterns. At the seedling stage, *RsHSF-04* and *RsHSF-33* were lowly expressed in 20 and 14 DAS root, respectively. Throughout development, *RsHSF-10* was only detected in 7 and 14 DAS root. *AtHSF-16* (At4g36990) is a direct target gene of AtHsfA1s, which are critical for plants to initiate positive activities when coping with HS [16, 67]. *RsHSF-03* is highly similar to *AtHSF-16* (At4g36990) and exhibited high expression level during seedling and taproot-thickening stages. Notably, *RsHSF-19*, the orthologous gene of *AtHSF-02*(At5g1682), was not expressed during all stages. These findings imply that the highly expressed genes among different stages and tissues may contribute to enhance stress tolerance in radish. Further characterization of these differentially expressed RsHSF genes would facilitate investigating the abiotic stresses response regulatory networks in radish.

## Conclusion

In summary, a total of 33 RsHSF genes were identified at the genome-wide level in radish. The protein motifs in most HSF members within one class were similar, and most HSF genes had only one intron. The expression patterns of several RsHSF genes were characterized under different stresses: the patterns were more similar within the same class than among different classes. Moreover, comparative analysis revealed a series of paralogous RsHSF genes in radish and orthologous RsHSF genes in *Arabidopsis* and rice. These results may enhance our understanding of RsHSF functions and their involvement in the radish stress responses. Identification of RsHSF genes provides a rich resource for comprehensive understanding of the roles of HSF genes in abiotic stresses response regulatory networks. These findings will facilitate further functional characterization of RsHSF genes, and provide valuable information to clarify the molecular mechanisms underlying abiotic stresses response in radish.

## Material And Methods

### Sequence collection and HSF identification

Whole genome sequences of the radish HSF family were identified from the NODAI Radish genome database [31]. The HSF family sequences identified from *Arabidopsis* and Chinese cabbage were downloaded from the *Arabidopsis* Information Resource [37] and *Brassica* database (BRAD) [33], respectively. The HSF family members for rice (*O. sativa*) and pineapple (*A. comosus* L.) were obtained from the Rice Genome Annotation Project [36] and Phytozome [34], respectively.

To confirm the radish HSF family candidates, proteins with the HSF-type DBD domain (Pfam accession number: PF00447) were searched against the genome protein sequences using the HMM search tool with an E-value cut-off of 0.01 [32].

### Phylogenetic analysis and characterization of RsHSFs

RsHSF DBD and oligomerization (HR-A/B or OD) domains were detected using SMART (<http://smart.embl-heidelberg.de/>) [68], Pfam [69], MARCOIL (<http://toolkit.tuebingen.mpg.de/marcoil>) and Heatster online tools (<http://www.cibiv.at/services/hsf/info>) [49]. NLS and NES domains were predicted using NLStradamus (<http://www.moseslab.csb.utoronto.ca/NLStradamus/>) and NESs (<http://www.cbs.dtu.dk/services/NetNES/>), respectively. RsHSF protein properties including molecular weight, theoretical pI and instability index were analyzed with the ExPASy ProtParam tool (<https://www.expasy.org/>). The neighbor-joining (NJ) method with parameters (substitutions type, amino acid; model/method, Poisson model; rates among sites, uniform rates; gaps/missing data treatment, complete deletion) was performed to generate phylogenetic trees with MEGA6.0 [38]. In addition, with specific parameters (substitutions type, amino acid; model/method, Jones – Taylor-Thornton model; rates among sites, gamma distributed (G); gamma parameters, 5; gaps/missing data treatment, complete deletion), phylogenetic trees were reconstructed using the maximum likelihood method. Bootstrap

analysis was conducted with 1000 replicates to assess the statistical support for each node. The phylogenetic trees were displayed using EVOLVIEW (<http://www.evolgenius.info/evolview/>).

## Gene structure and conserved motif analysis

RsHSFgene exon/intron organisation was obtained from GSDS (<http://gsds.cbi.pku.edu.cn/>) [70] by aligning coding sequences with their corresponding genomic DNA sequences. The conserved motifs within the determined HSF groups were identified by MEME (version4.11.1) (<http://meme-suite.org/>) based on the following parameters: number of repetitions, any; maximum number of motifs, 25; and the optimum motif widths, 6–50 AAs [71]. The conserved motifs were visualized by Tertools [72].

## Chromosomal localization and identification of orthologous and paralogous *HSFs*

RsHSFgene sequences in scaffolds were obtained from the NODAI Radish Genome Database [31], and local Blasts were performed against Radish Genome Database [73]. Gene sequences with  $\geq 98\%$  identity and length difference  $\leq 5$  base pairs were considered to be the same genes between two genomes, and the corresponding location of RsHSF genes in chromosomes was localized using MapChart Software [74]. The Multiple Collinearity Scan toolkit (MCScanX) was used to identify the RsHSF duplication events [41, 73]. BLASTP was performed to identify the intra-species paralogous pairs using protein sequences with the following parameters settings: 1) alignment significance: E\_VALUE (default: 1e-05); 2) MATCH\_SCORE: final score (default: 50); 3) MATCH\_SIZE: number of genes required to call a collinear block (default: 5) and the maximum gaps (default: 25).

The OrthoMCL pipeline [43] was used with standard settings to identify potential orthologous and paralogousHSFgenes in radish, *Arabidopsis*, and rice. The relationships of orthologous and paralogous genes among the three species were plotted using Circos software [75].

## Expression analyses of RsHSF genes by RNA-Seq

To analyze the expression patterns of the RsHSF genes, the Illumina RNA-Seq data from five tissues (cortex, cambium, xylem, root tip and leaf) and six stages [7, 14, 20, 40, 60 and 90 days after sowing (DAS)] were collected from radish reference genome [31] (Additional file 10: Table S12). The expression level for each RsHSF gene was presented with the RPKM method. Finally, heat maps were generated using TBtools [72].

## Abiotic stress treatments

'NAU-YH' seeds, a radish advanced in a breeding line with small red global roots, were sterilised, rinsed and incubated for three days. Germinated seeds were transferred into a plug tray and cultured in a growth chamber at 25°C day/18°C night with 14 h light/10 h dark, 60% humidity and 12,000 lx light. Three weeks later, seedlings with 6–7 true leaves at the cortex split stage were exposed to 38 °C for 0, 6 or 24 h (termed control, Heat6 or Heat24, respectively). For salt and heavy metal (Cd and Pb) treatment, seedlings of identical sizes were grown in a plastic container with modified half-strength Hoagland nutrient solution as described previously [76]. The nutrient solution was changed every three days. One week later, the seedlings were treated with CdCl<sub>2</sub>·2.5H<sub>2</sub>O (50mg/L), Pb(NO<sub>3</sub>)<sub>2</sub> (100mg/L or 200mg/L) and NaCl (150mM or 250mM), respectively [29]. Seedlings grown in Cd/Pb-free nutrient solution were used as the control. In this study, three similar seedlings with 6 true leaves at the cortex split stage were used for each treatment, and each treatment was performed in triplicate. Additionally, 0.1g root tissue was used for RNA isolation from each sample. These samples were immediately frozen in liquid nitrogen and stored at -80°C for further analysis.

## RNA-Seq data and RT-qPCR analysis

In this study, the raw RNA-Seq data was collected from NCBI Sequence Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/Traces/sra>) using the SRA Toolkit (ver.2.8.2). After removing the adaptor sequences, contaminants and low-quality reads, the clean reads were mapped to the radish reference genome using TopHat2 [77]. Transcripts per Kilobase Million (TPM) values were estimated using Salmon [78] for different sample comparisons. Differential gene expression and the adjusted *p* values were calculated using the edgeR package [79]

Total RNA was isolated from the control and treated radish roots using RNAsimple total RNA kit (Tiangen, Beijing, China). Then, the RNA was reverse transcribed into cDNA using PrimeScript™ II 1st Strand cDNA Synthesis Kit (Takara, Dalian, China) according to the manufacturer's instructions. RT-qPCR analysis was conducted using SYBR Green PCR Master Mix with ROX (Takara, Dalian, China) [29]. *RsActin* was employed as the internal standard to normalise expression [80]. Each 20 µl reaction contained 10 µl of 2×SYBR Green PCR Master Mix (Takara, Dalian, China), 0.2 µM of each primer and 2 µl diluted cDNA. PCR was performed out on an iCycler iQ real-time PCR detection system (BIO-RAD, USA) with the following thermal cycling conditions: 95°C for 3 min, and 40 cycles of 95°C for 5 s, 58°C for 30 s, and 72°C for 10 s. The 2<sup>-CT</sup> method was used to calculate relative expression levels [81]. Three biological and technical replicates were performed. The significance of differences between groups was evaluated using a Student's t test. Analyses were performed with GraphPad Prism software (GraphPad Software, San Diego, California).

## Abbreviations

HSF: heat stress transcription factor; RT-qPCR: Reverse-transcription quantitative PCR; HSR: heat stress response; TDR: temperature-dependent repression; HSPs: heat shock proteins; DEGs: differentially

expressed genes; HMM: Hidden Markov Model; HR-A/B: hydrophobic amino acid residues; OD: oligomerization domain; DBD: DNA-binding domain; NLS: nuclear localization Signal; NES: nuclear export signal; MWs: molecular weights; pI: theoretical isoelectric points; DAS: days after sowing; CDS: coding sequence; TF: transcription factor; WGD: whole-genome duplication; WGT: whole genome triplication.

## **Declarations**

## **Ethics approval and consent to participate**

In this study, all the radish lines were collected from germplasm bank of National Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University. Collection of plant materials complied with the institutional, national and international provisions.

## **Consent for publication**

**Not applicable.**

## **Competing interests**

**The authors declare that they have no competing interests for this research.**

## **Funding**

This work was in part supported by grants from the Natural Science Foundation of China (31601766,31501759), National Key Technology R&D Program of China (2017YFD0101803; 2017YFD0101806), Key Technology R&D Program of Jiangsu Province (BE2016379), the Jiangsu Agricultural S&T Innovation Fund [CX (16)1012] and the Jiangsu Agricultural Key Project [JATS[2018]285].

## **Author' contributions**

MT performed the data analysis and wrote the manuscript. LX and YW contributed powerful analytical tools. WC and XL performed the RNA extractions. YX and LF helped with the revise of manuscript. LL conceived and designed the research. All authors read and approved the final manuscript.

## **Availability of data and materials**

**The datasets supporting the conclusions of this article are included within the article and its additional files.**

## Acknowledgments

**Not applicable.**

## References

- Tester M, Bacic A. Abiotic stress tolerance in Grasses. From model plants to crop plants. *Plant Physiol.* 2005; 137, 791–793.
- Yokotani N, Ichikawa T, Kondou Y, Iwabuchi M, Matsui M, Hirochika H, Oda K. Role of the rice transcription factor JAmyb in abiotic stress response. *J Plant Res.* 2013; 126:131–139.
- Mickelbart MV, Hasegawa PM, Bailey-Serres J. Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nat Rev Genet.* 2015; 16(4): 237-251.
- Wang RH, Xu L, Zhu XW, Zhai LL, Wang Y, Yu RG, Gong YQ, Limera C, Liu LW. Transcriptome-wide characterization of novel and heatstress-responsive microRNAs in radish (*Raphanus sativus* L.) using next-generation sequencing. *Plant Mol Biol Rep.* 2015; 33(4):867–880.
- Lobell DB, Schlenker W, Costa-Roberts J. Climate trends and global crop production since 1980. *Science.* 2011; 333(6042):616–620.
- Jung KH, Ko HJ, Nguyen MX, Kim SR, Ronald P, An G. Genome-wide identification and analysis of early heat stress responsive genes in rice. *J Plant Biol.* 2012; 55(6):458–468.
- Wang PF, Song H, Li CS, Li AQ, Guan HS, Hou L, Wang XJ. Genome-wide dissection of the heat shock transcription factor family genes in *Arachis*. *Front Plant Sci.* 2017; 8:106.
- Cai SY, Zhang Y, Xu YP, Qi ZY, Li MQ, Ahammed GJ, Xia XJ, Shi K, Zhou HY, Reiter RJ, Yu JQ, Zhou J. HsfA1a upregulates melatonin biosynthesis to confer cadmium tolerance in tomato plants. *J Pineal Res.* 2017; 62(2): e12387.
- Scharf KD, Berberich T, Ebersberger I, Nover L. The plant heat stress transcription factor (Hsf) family: structure, function and evolution. *BBA-Gene Regul Mec.* 2012; 1819(2): 104-119.
- Hossain Z, Nouri MZ, Komatsu S. Plant cell organelle proteomics in response to abiotic stress. *J Proteome Res.* 2012; 11(1), 37–48.

- Singh R, Jwa NS. Understanding the responses of rice to environmental stress using proteomics. *J Proteome Res.* 2013; 12(11): 4652-4669.
- Nakashima K, Ito Y, Yamaguchi-Shinozaki K. Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. *Plant physiology*, 2009, 149(1): 88-95
- Huang Y, Li MY, Wang F, Xu ZS, Huang W, Wang GL, Ma J, Xiong AS. Heat shock factors in carrot: genome-wide identification, classification, and expression profiles response to abiotic stress. *Mol Biol Rep.* 2015; 42(5): 893.
- Todaka D, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K. Toward understanding transcriptional regulatory networks in abiotic stress responses and tolerance in rice. *Rice*, 2012, 5(1): 6.
- Huang YC, Niu CY, Yang CR, Jinn TL. The heat stress factor HSFA6b connects ABA signaling and ABA-mediated heat responses. *Plant Physiol.* 2016; 172(2): 1182-1199.
- Ohama N, Kusakabe K, Mizoi J, et al. The transcriptional cascade in the heat stress response of *Arabidopsis* is strictly regulated at the level of transcription factor expression. *Plant Cell.* 2016; 28(1): 181-201.
- Koskull-Döring PV, Scharf KD, Nover L. The diversity of plant heat stress transcription factors. *Trends Plant Sci.* 2007; 12(10): 452-457.
- Björk JK, Sistonen L. Regulation of the members of the mammalian heat shock factor family. *FEBS J.* 2010; 277(20): 4126-4139.
- Cicero MP, Hubl ST, Harrison CJ, Littlefield O, Hardy JA, Nelson HCM. The wing in yeast heat shock transcription factor (HSF) DNA-binding domain is required for full activity. *Nucleic Acids Res.* 2001; 29(8): 1715-1723.
- Baniwal SK, Bharti K, Chan KY, Fauth M, Ganguli A, Kotak S, Mishra SK, et al. Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. *J Biosci.* 2004; 29(4):471–87.
- Nover L, Bharti K, Doring 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 Chaperons.* 2001; 6:177–189.
- Yang Z, Wang Y, Gao Y, Zhou Y, Zhang E, Hu Y et al. Adaptive evolution and divergent expression of heat stress transcription factors in grasses. *BMC Evol Biol.* 2014; 14:147.
- Song XM, Liu GF, Duan WK, Liu TK, Huang ZN, Ren J, Li Y, Hou XL. Genome-wide identification, classification and expression analysis of the heat shock transcription factor family in Chinese cabbage. *Mol Genet Genomics.* 2014; 289(4): 541-551.

- Guo JK, Wu J, Ji Q, Wang C, Luo L, Yuan Y, Wang YH, Wang J. Genome-wide analysis of heat shock transcription factor families in rice and *Arabidopsis*. *J Genet Genomics*.2008; 35(2), 105-118.
- Chauhan H, Khurana N, Agarwal P, Khurana P. Heat shock factors in rice (*Oryza sativa* L.): genome-wide expression analysis during reproductive development and abiotic stress. *Mol Genet Genomics*. 2011; 286(2):171–187.
- Lin YX, Jiang HY, Chu ZX, Tang XL, Zhu SW, Cheng BJ. Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. *BMC Genomics*. 2011; 12(1): 76.
- Qiao X, Li M, Li LT, Yin H, Wu JY, Zhang SL. 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(1): 12.
- Wang FM, Dong Q, Jiang HY, Zhu SW, Chen BJ, Xiang Y. Genome-wide analysis of the heat shock transcription factors in *Populus trichocarpa* and *Medicago truncatula*. *Mol Biol Rep*. 2012; 39(2):1877–1886.
- Xu L, Wang Y, Liu W, Wang J, Zhu XW, Zhang KY, Yu RG, Wang RH, Xie Y, Zhang W, Gong YQ, Liu LW. De novo sequencing of root transcriptome reveals complex cadmium-responsive regulatory networks in radish (*Raphanus sativus* L.). *Plant Sci*. 2015; 236: 313-323.
- Wang RH, Mei Y, Xu L, Zhu XW, Wang Y, Guo J, Liu LW. Genome-wide characterization of differentially expressed genes provides insights into regulatory network of heat stress response in radish (*Raphanus sativus* L.). *Functional & integrative genomics*. 2018; 18(2): 225-239.
- Mitsui Y, Shimomura M, Komatsu K, et al. The radish genome and comprehensive gene expression profile of tuberous root formation and development. *Sci Rep*. 2015; 5: 10835.
- Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011; 39(suppl\_2): W29-W37.
- Cheng F, liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X. BRAD, the genetics and genomics database for *Brassica* plants. *BMC Plant Biol*. 2011; 11:136.
- Ming R, VanBuren R, Wai CM, et al. The pineapple genome and the evolution of CAM photosynthesis. *Nat Genet*. 2015; 47(12): 1435.
- Phytozome: [<https://phytozome.jgi.doe.gov/pz/portal.html>].
- Rice Genome Annotation Project: [<http://rice.plantbiology.msu.edu/>].
- TAIR: [<http://www.arabidopsis.org/>].

- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013; 30:2725–2729.
- Wang X, Wang H, Wang J, Sun r, Wu J, liu S, Bai Y, Mun JH, Bancroft I and Cheng F. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet.* 2011, 43:1035–1039.
- Cannon SB, Mitra A, Baumgarten A, Young ND, 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.
- Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC and Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012; 40(7):e49.
- Moghe GD, Hufnagel DE, Tang HB, Xiao YL, Dworkin I, Town CD, Conner JK and Shiu SH. Consequences of whole-genome triplication as revealed by comparative genomic analyses of the wild radish *Raphanus raphanistrum* and three other Brassicaceae species. *Plant Cell.* 2014; 26(5): 1925-1937.
- Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003; 13 2178–2189.
- Ohama N, Sato H, Shinozaki K, Yamaguchi-Shinozaki K. Transcriptional regulatory network of plant heat stress response. *Trends Plant Sci.* 2017; 22(1), 53-65.
- Yoshida T, Ohama N, Nakajima J, Kidokoro S, et al. *Arabidopsis* HsfA1 transcription factors function as the main positive regulators in heat shock-responsive gene expression. *Mol Genet Genomics.* 2011; 286(5–6):321–332. Nishizawa-Yokoi A, Nosaka R, Hayashi H, et al. HsfA1d and HsfA1e involved in the transcriptional regulation of HsfA2 function as key regulators for the Hsf signaling network in response to environmental stress. *Plant Cell Physiol.* 2011; 52(5): 933-945. 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(5): 1759-1768.
- Kotak S, Port M, Ganguli A, Bicker F, Koskull-Döring PV. 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.
- Berz J, Simm S, Schuster S, et al. HEATSTER: A Database and Web Server for Identification and Classification of Heat Stress Transcription Factors in Plants. *Bioinformatics and biology insights*, 2019, 13: 1177932218821365. Wendel JF. Genome evolution in polyploids. In: Doyle J, Gaut B (editors). *Plant molecular evolution*. Dordrecht: Springer; 2000. p. 225–249.
- Bowers JE, Chapman BA, Rong J, Paterson AH. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature.* 2003; 422: 433–438.

- Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. Proc Natl Acad Sci USA. 2010; 107: 18724–18728.
- Nie SS, Zhang MJ, Zhang LG. Genome-wide identification and expression analysis of *calmodulin-like* (CML) genes in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). BMC Genomics. 2017; 18(1): 842.
- Wang W, Jiang W, Liu J, Li Y, Gai JY, Li Y. Genome-wide characterization of the aldehyde dehydrogenase gene superfamily in soybean and its potential role in drought stress response. BMC Genomics. 2017; 18(1): 518.
- Li C, Wang Y, Xu L, Nie SS, Chen YL, Liang DY, Sun XC, Karanja BK, Luo XB, Liu LW. Genome-wide characterization of the MADS-Box gene family in Radish (*Raphanus sativus* L.) and assessment of its roles in flowering and floral organogenesis. Front Plant Sci. 2016; 7:1390.
- Ma J, Wang F, Li MY, Jiang Q, Tan GF, Xiong AS. Genome wide analysis of the NAC transcription factor family in Chinese cabbage to elucidate responses to temperature stress. Sci Hortic. 2014; 165:82–90
- Song XM, Huang ZN, Duan WK, Ren J, Liu TK, Li Y, Hou XL. Genome-wide analysis of the bHLH transcription factor family in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). Mol Genet Genomics. 2014; 289(1):77–91.
- Yang YW, Tai PY, Chen Y, Li WH. A study of the phylogeny of *Brassica rapa*, *B. nigra*, *Raphanus sativus*, and their related genera using noncoding regions of chloroplast DNA. Mol Phylogenet Evol. 2002; 23(2): 268–275.
- Lysak MA, Koch MA, Pecinka A, Schubert I. Chromosome triplication found across the tribe Brassiceae. Genome Res. 2005; 15: 516–525.
- Hu WH, Hu GC, Han B. Genome-wide survey and expression profiling of heat shock proteins and heat shock factors revealed overlapped and stress specific response under abiotic stresses in rice. Plant Sci. 2009; 176(4): 583-590.
- Liu HC, Liao HT, Charng YY. The role of class A1 heat shock factors (HSFA1s) in response to heat and other stresses in *Arabidopsis*. Plant Cell Environ, 2011. 34(5):738.
- Pérez-Salamó I, Papdi C, Rigó G, Zsigmond L, Vilela B, Lumberras V, et al. The heat shock factor A4A confers salt tolerance and is regulated by oxidative stress and the mitogen-activated protein kinases MPK3 and MPK6. Plant Physiol. 2014; 165(1), 319-334.
- Czarnecka-Verner E, Pan S, Salem T, Gurley WB. Plant class B HSFs inhibit transcription and exhibit affinity for TFIIB and TBP. Plant Mol Biol. 2004; 56(1), 57-75.
- Sun XC, Xu L, Wang Y, Luo XB, Zhu XW, Kinuthia KB, Nie SS, Feng HY, Li C and Liu LW. Transcriptome-based gene expression profiling identifies differentially expressed genes critical for salt stress response in

radish (*Raphanus sativus* L.). Plant Cell Rep, 2016, 35(2): 329-346.

Xu L, Wang Y, Zhang F, et al. Dissecting root proteome changes reveals new insight into cadmium stress response in radish (*Raphanus sativus* L.). Plant Cell Physiol, 2017, 58(11): 1901-1913.

Xie Y, Ye S, Wang Y, Xu L, Zhu XW, Yang JL, Feng HY, Yu RG, Karanja B, Gong YQ and Liu LW. Transcriptome-based gene profiling provides novel insights into the characteristics of radish root response to Cr stress with next-generation sequencing. Front Plant Sci, 2015, 6: 202. Chang YY, Liu HC, Liu NY, Chi WT, Wang CN, Chang SH and Wang TT . A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermotolerance in *Arabidopsis*. Plant Physiol. 2007.143(1):251-62.

Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 2012; 40:D302–D305.

Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, Bateman A. The Pfam protein families database. Nucleic Acids Res. 2010; 38 Database: 211-222.

Hu B, Jin JP, Guo AY, Zhang H, Luo JC and Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics, 2015, 31(8):1296-1297.

Bailey TL, Elkan C. The value of prior knowledge in discovering motifs with MEME. Proc Int Conf Intell Syst Mol Biol. 2005; 3:21-29.

Chen CJ, Xia R, Chen H and He YH. TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. BioRxiv, 2018: 289660.

Jeong YM, Kim N, Ahn BO, Oh M, Chung WH, Chung H, et al. Elucidating the triplicated ancestral genome structure of radish based on chromosome-level comparison with the *Brassica* genomes. Theor Appl Genet. 2016; 129(7): 1357-1372.

Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered. 2002; 93(1):77–8.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ and Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; 19:1639–45.

Xu L, Wang Y, Zhai LL, Xu Y, Wang LJ, Zhu XW, Gong YQ, Yu RG, Limera C, Liu LW. Genome-wide identification and characterization of cadmium-responsive microRNAs and their target genes in radish (*Raphanus sativus* L.) roots. J Exp Bot. 2013; 64:4271–4287.

Kim D, Pertea G, Trapnell C, Pimentel H, Ryan Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36

Patro R., Duggal G., Love M.I., Irizarry R.A., Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Meth.* 2017; 14(4), 417–419 (2017).

Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26(1):139–40.

Xu YY, Zhu XW, Gong YQ, Xu L, Wang Y and Liu LW. Evaluation of reference genes for gene expression studies in radish (*Raphanus sativus* L.) using quantitative real-time PCR. *Biochem Bioph Res Co.* 2012; 424(3): 398-403.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-CT}$  method. *Methods.* 2001; 25:402–408.

## Supplementary Legends

### Additional files

Additional file 1: Table S1. Amino acid sequences of the HSF proteins from NODAI Radish genome database. Table S2. CDS sequences of the *HSF* genes from NODAI Radish genome database. Table S3. DNA sequences of the *HSF* genes from NODAI Radish genome database.

Additional file 2: Figure.S1 Protein sequence alignment of the DBD domain identified in for all BrHsf genes. Figure.S2 The LOGO of 25 amino acid motifs in HSF proteins. Figure.S3 Phylogenetic tree constructed using the maximum likelihood method by MEGA6, with HSF proteins in radish, Chinese cabbage, pineapple, *Arabidopsis*, and rice. The numbers are bootstrap values based on 1,000 iterations. Figure.S4 The networks of HSF genes in radish and *Arabidopsis*. This interrelation network has been constructed using radish and *Arabidopsis* orthologous gene pairs. Figure.S5 The differential expression profiles of some HSF genes in radish under the Heat, salt, Pb and Cd treatments in radish. The color scale of heatmap is based on the  $\log_2$ Foldchange (a) and  $\log_2$ TPM value (b), respectively.

Additional file 3: Table S4. Summary information of the RsHSF proteins.

Additional file 4: Table S5. HSFs in *Arabidopsis*, rice, and pineapple.

Additional file 5: Table S6. Summary information of chromosomal location of RsHSFs in radish.

Additional file 6: Table S7. The different origins of HSF genes in radish.

Additional file 7: Table S8. Duplicated genes of *HSF* in radish.

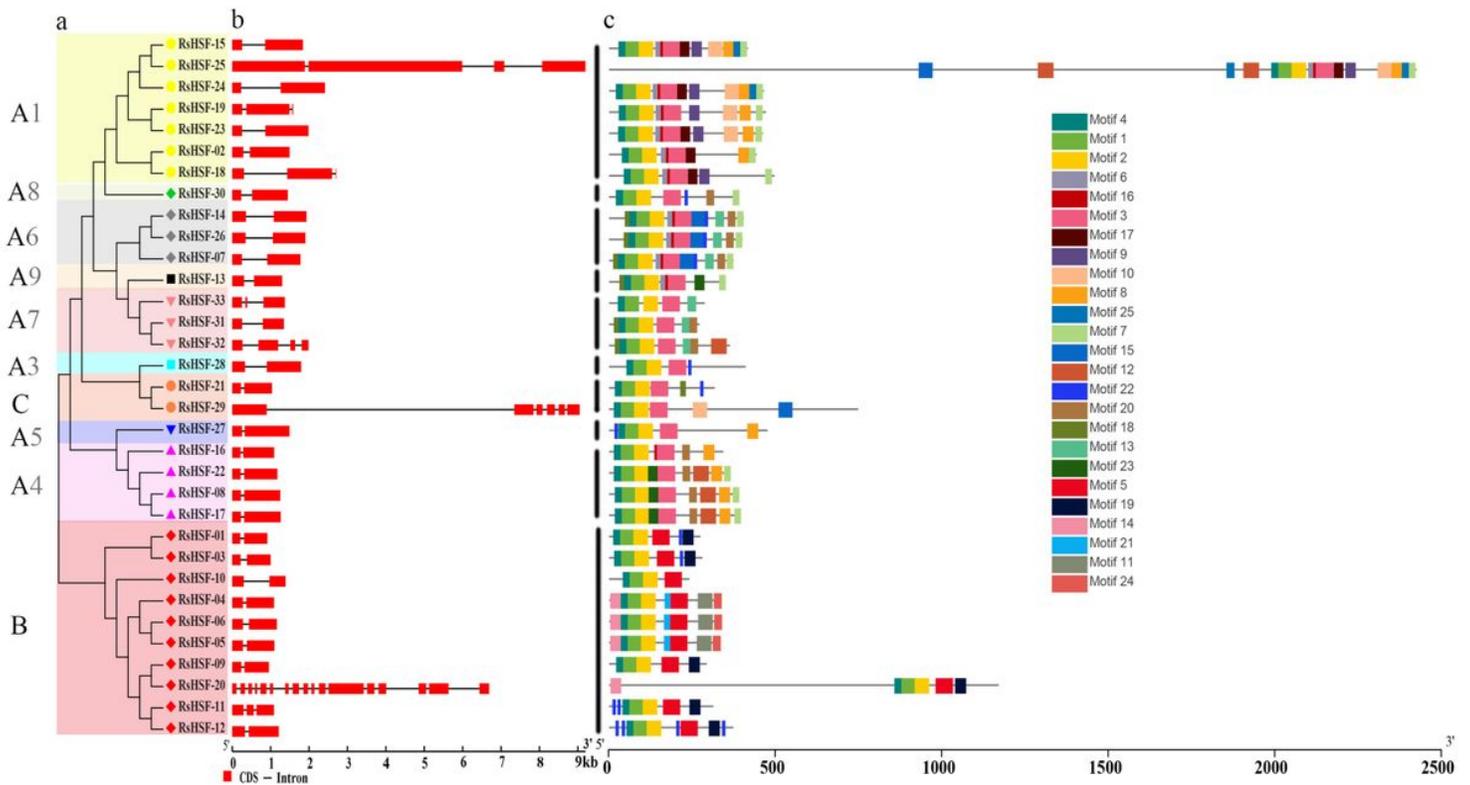
Additional file 8: Table S9. The orthologous and paralogous gene pairs of HSF proteins among the radish, *Arabidopsis* and rice.

Additional file 9: Table S10. Tissue-specific expression of radish *HSF* genes.

Additional file 10: Table S11. Relative expression of RsHSF genes under different treatments determined by RT-qPCR.

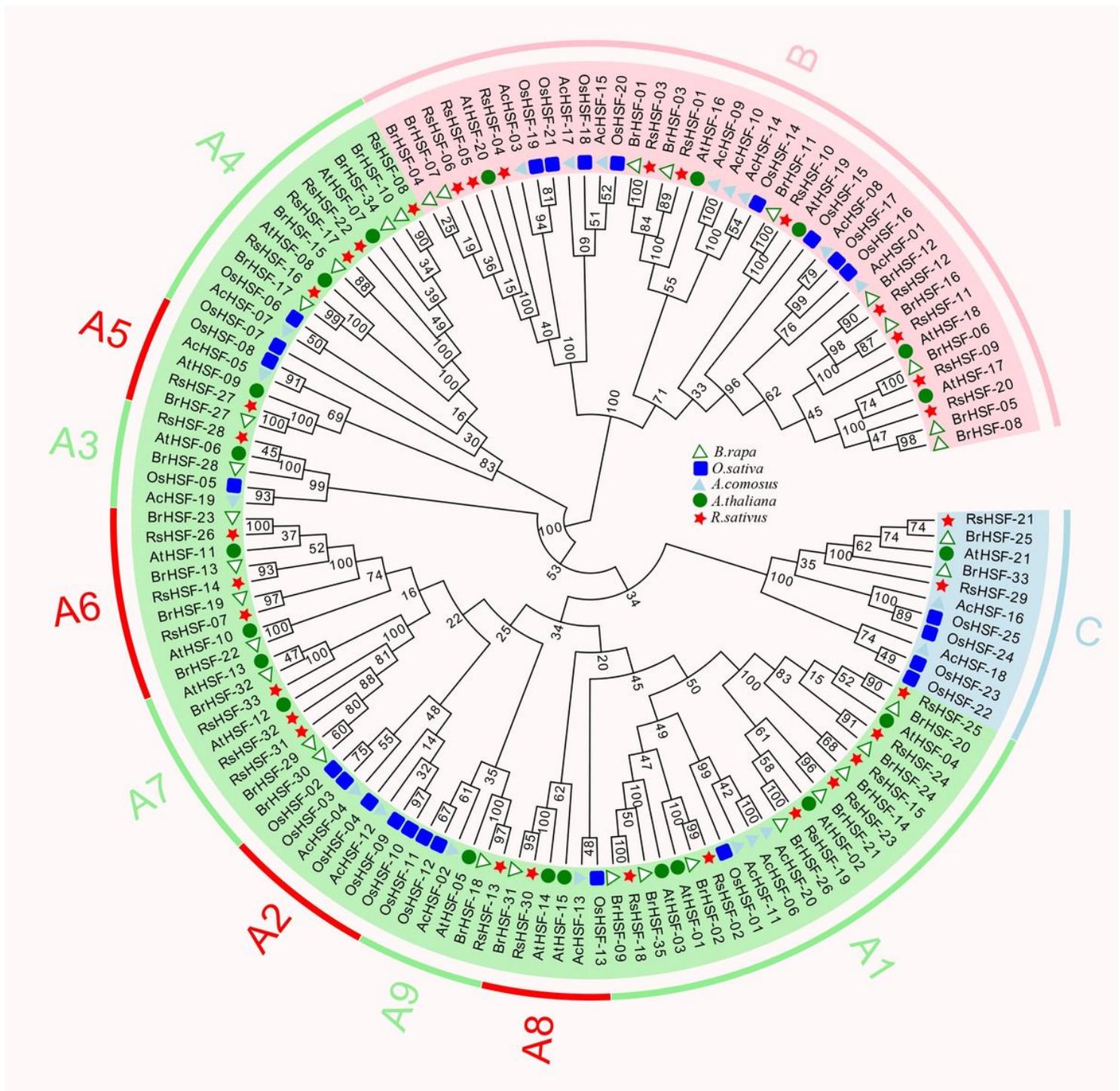
Additional file 11: Table S12. The accession number of RNAseq data

## Figures



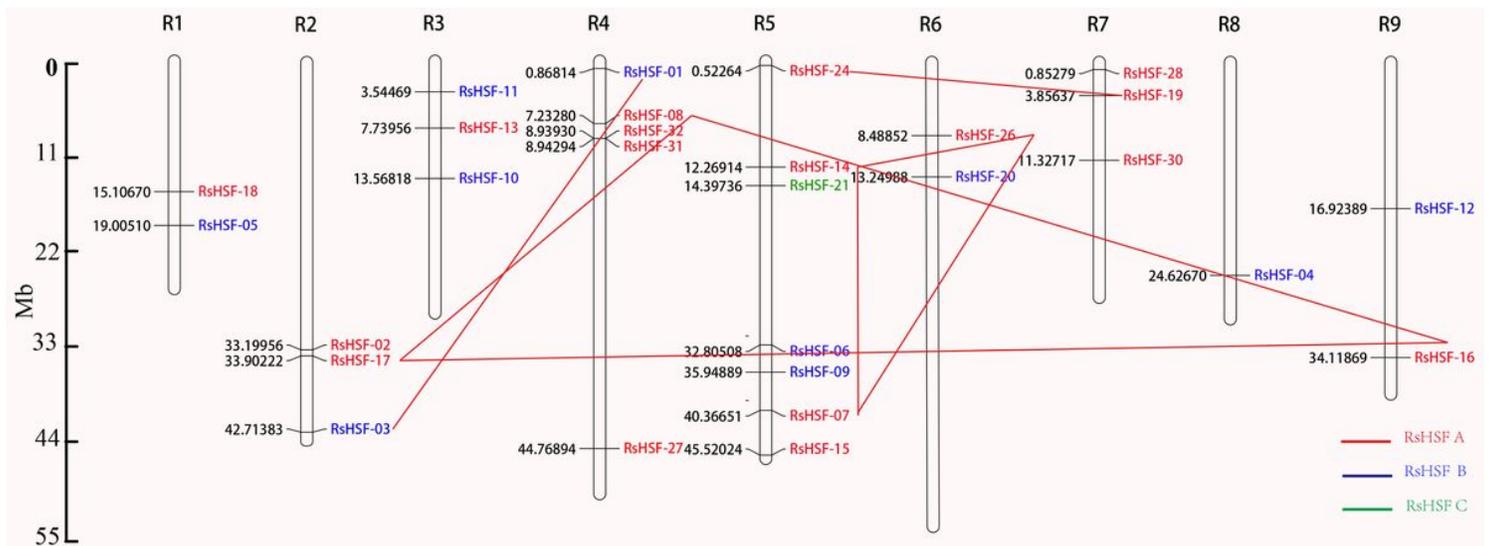
**Figure 1**

Gene structures and protein motifs of HSF family. (a) The neighborjoining phylogenetic tree of RsHSFs. (b) The exon/intron structures of RsHSF genes. The relative position is proportionally displayed based on the kilobase scale at the bottom of the figure. Red boxes and gray lines represent exons and introns, respectively. (c) The conserved motifs in radish HSFs. Different motifs and their relative positions are represented by the colored boxes



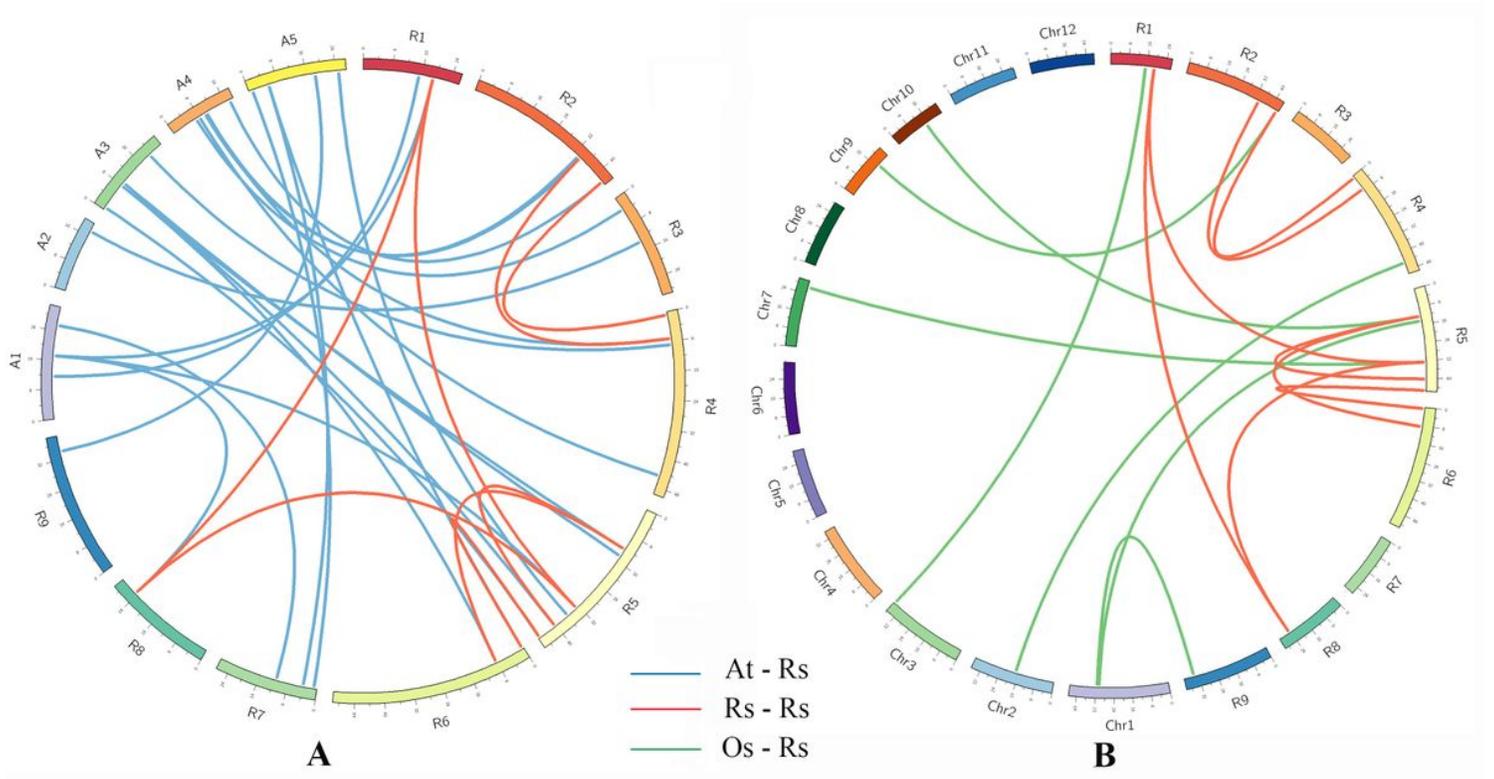
**Figure 2**

Phylogenetic tree constructed using the neighbor-joining method by MEGA6, with HSF proteins in radish, Chinese cabbage, pineapple, Arabidopsis, and rice. The numbers are bootstrap values based on 1,000 iterations.



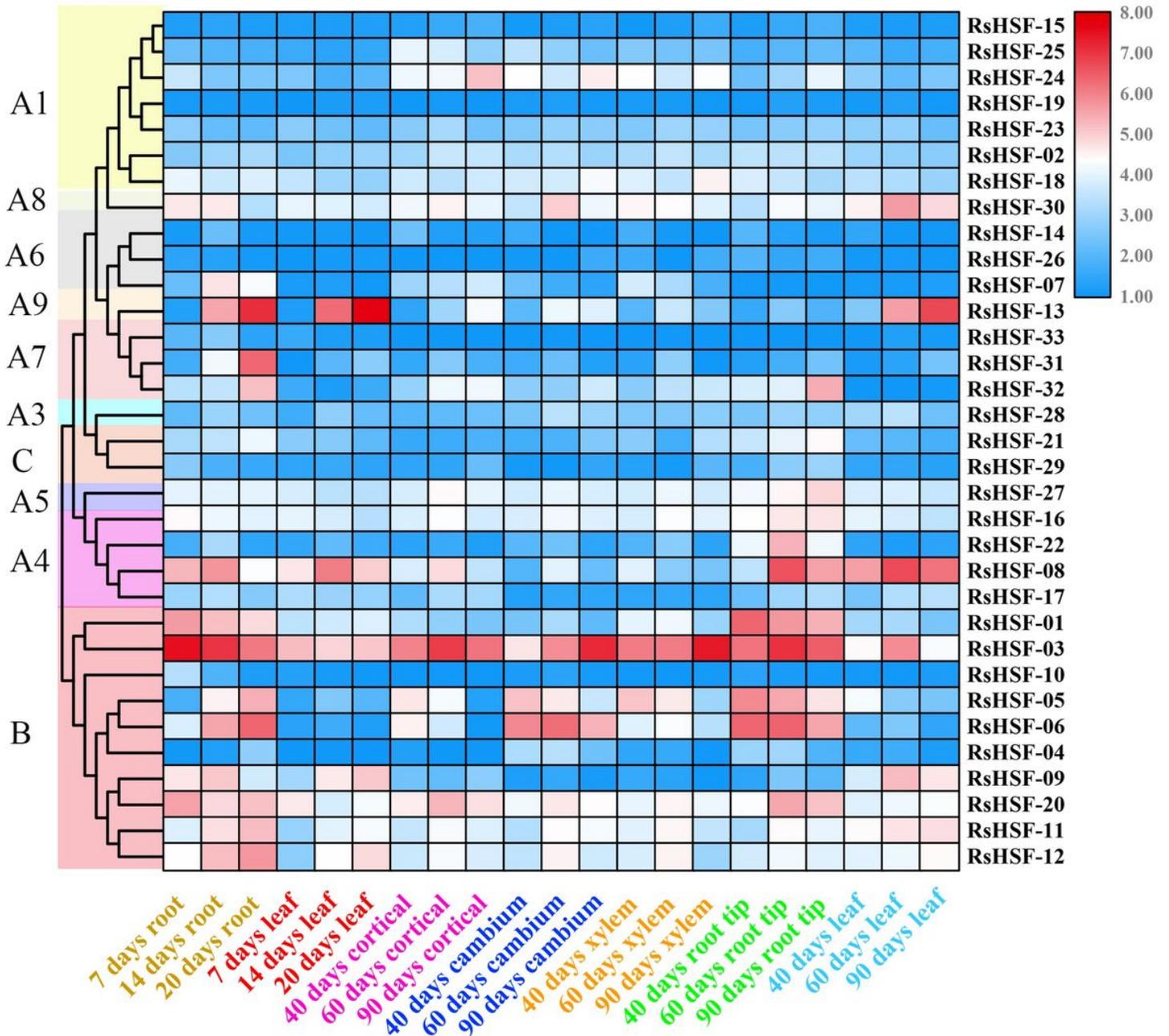
**Figure 3**

Distribution of RsHSF genes on nine chromosomes. Red, blue and green word represent class A, B and C, respectively. Red lines connect the RsHSF duplicated genes.



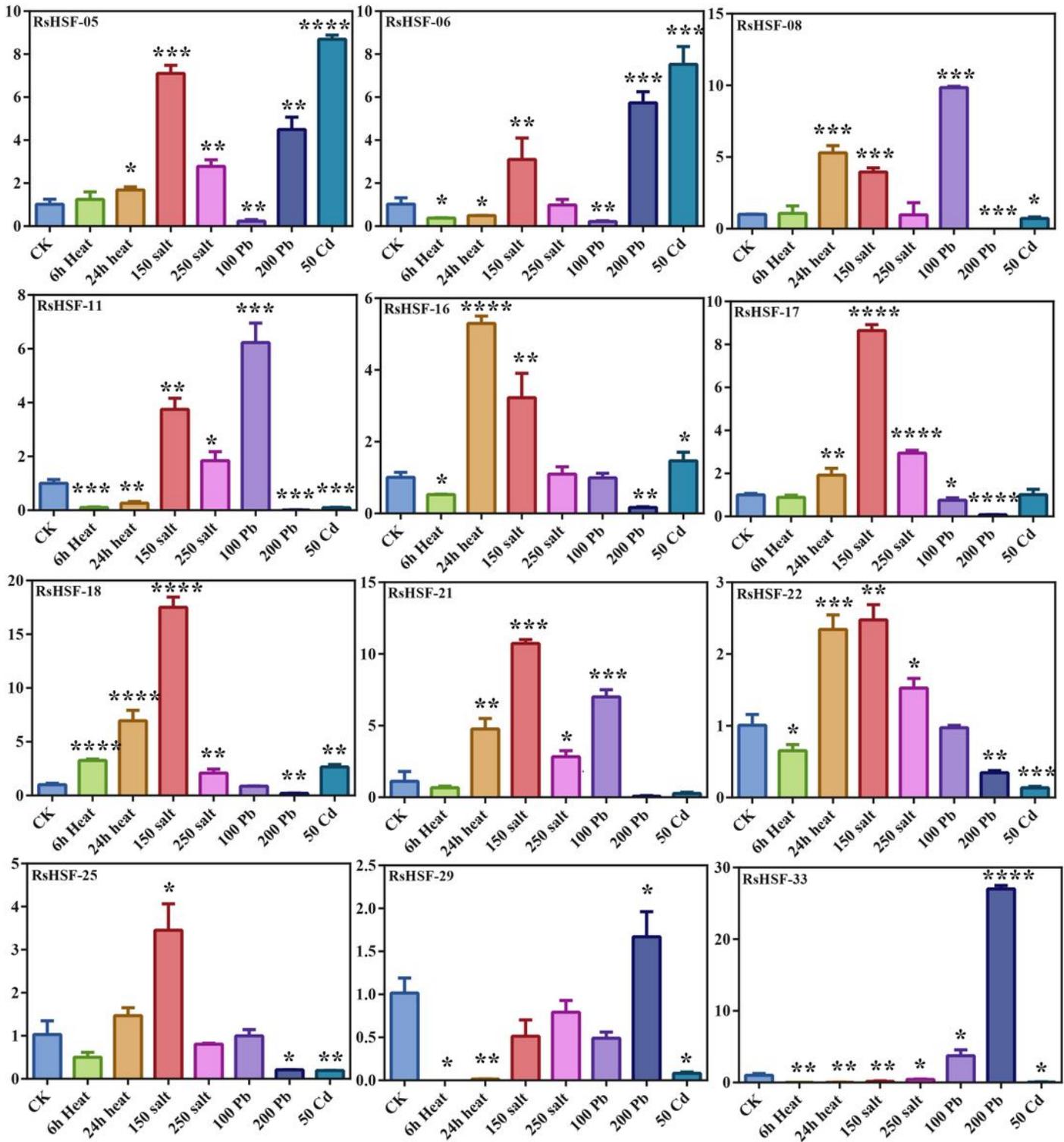
**Figure 4**

Orthologous and paralogous genes of HSF. (A) Radish (R01–R09) and five Arabidopsis chromosome (A1–A5) maps were based on orthologous and paralogous pair positions; (B) Radish and rice chromosome (Chr1–Chr12) maps were based on orthologous and paralogous pair positions with Circos.



**Figure 5**

Heatmap of RsHSF genes in different stages and tissues. The color scale is based on the log<sub>2</sub> (RPKM + 1) values.



**Figure 6**

The expression levels of representative RsHSF genes under different stresses (cadmium, lead, heat, and salt stress). The subgroup is marked in different colors under the gene name. Each bar shows the mean  $\pm$  SE of the triplicate assay. GraphPad Prism software was used to analysis and asterisks reveal the gene significantly upregulated or downregulated under abiotic stresses by t test (\*P<0.05, \*\*P<0.01, \*\*\*<0.001).

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS5.jpg](#)
- [TableS4.xlsx](#)
- [TableS1.xlsx](#)
- [TableS12.xlsx](#)
- [TableS11.xlsx](#)
- [FigureS1.jpg](#)
- [FigureS3.jpg](#)
- [FigureS4.jpg](#)
- [FigureS2.jpg](#)
- [TableS10.xlsx](#)
- [TableS9.xlsx](#)
- [TableS8.xlsx](#)
- [TableS6.xlsx](#)
- [TableS7.xlsx](#)
- [TableS5.xlsx](#)