

# Genome-wide characterization and expression analysis of the heat shock transcription factor family in pumpkin (*Cucurbita moschata*)

Changwei Shen

Henan Institute of Science and Technology

Jingping Yuan (✉ [jpyuan666@163.com](mailto:jpyuan666@163.com))

Henan Institute of Science and Technology

---

## Research article

**Keywords:** Cucurbita moschata, Heat shock transcription factor, Gene duplication, Conserved domain, Cis-acting elements, Expression pattern

**Posted Date:** September 28th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-38574/v4>

**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 at BMC Plant Biology on October 14th, 2020. See the published version at <https://doi.org/10.1186/s12870-020-02683-y>.

## Abstract

**Background:** Crop quality and yield are affected by abiotic and biotic stresses, and heat shock transcription factors (*Hsfs*) are considered to play important roles in regulating plant tolerance under various stresses. To investigate the response of *Cucurbita moschata* to abiotic stress, we analyzed the genome of *C. moschata*.

**Results:** In this research, a total of 36 *C. moschata Hsf* (*CmHsf*) members were identified and classified into three subfamilies (I, II, and III) according to their amino acid sequence identity. The *Hsfs* of the same subfamily usually exhibit a similar gene structure (intron-exon distribution) and conserved domains (DNA-binding and other functional domains). Chromosome localization analysis showed that the 36 *CmHsfs* were unevenly distributed on 18 of the 21 chromosomes (except for Cm\_Chr00, Cm\_Chr08 and Cm\_Chr20), among which 18 genes formed 9 duplicated gene pairs that have undergone segmental duplication events. The Ka/Ks ratio showed that the duplicated *CmHsfs* have mainly experienced strong purifying selection. High-level synteny was observed between *C. moschata* and other *Cucurbitaceae* species.

**Conclusions:** The expression profile of *CmHsfs* in the roots, stems, cotyledons and true leaves revealed that the *CmHsfs* exhibit tissue specificity. The analysis of *cis*-acting elements and quantitative real-time polymerase chain reaction (qRT-PCR) revealed that some key *CmHsfs* were activated by cold stress, heat stress, hormones and salicylic acid. This study lays the foundation for revealing the role of *CmHsfs* in resistance to various stresses, which is of great significance for the selection of stress-tolerant *C. moschata*.

## Background

Plants are constantly subjected to all kinds of adverse environmental pressures during growth and development stages, thus, they have developed special mechanisms to cope with adverse conditions [1, 2]. Transcription factors usually play an important role in the regulation of stress responses [3]. Heat shock transcription factors (*Hsfs*) are the most important transcription regulators [4]. They are the terminal components of signal transduction chains and can mediate the activation of genes that respond to various abiotic pressures (drought stress, heat stress and a large number of chemical stress factors) [4].

The first *Hsf* gene was cloned from yeast [5, 6], followed by some mammals [7-10]. The first plant *Hsf* gene was cloned from tomato [11]. With the sequencing of the *Oryza sativa* and *Arabidopsis thaliana* genomes, *Hsf* genes have also been identified in *O. sativa* and *A. thaliana* [12, 13]. Subsequently, researchers identified 31, 25, 21, 26, 35, 29, 27, 19 and 35 *Hsf* genes in the *Populus trichocarpa* [14], *Zea mays* [15], *Cucumis sativa* [16], *Glycine max* [17], *Brassica rapa* ssp. *pekinensis* [18], *Pyrus bretschneideri* [19], *Solanum tuberosum* [20], *Vitis vinifera* [21] and *Brassica oleracea* [22] genomes, respectively.

A typical *Hsf* usually contains four conserved domains: a DNA-binding domain (DBD) at the N-terminus, a hydrophobic oligomerization domain (HR-A/B or OD), a nuclear localization signal (NLS), and a nuclear export signal (NES) [23]. The DBD is the most conserved domain structure in *Hsfs* and is mainly responsible for binding to the heat shock elements (HSEs) of the target gene promoter, while the HR-A/B domain is a hydrophobic heptad repeat forming a spiral coil structure, which is a prerequisite for transcription [23]. The NLS is rich in Arg (R) and Lys (K) residues, while the NES is rich in Leu (L). NLS is recognized by the corresponding NES, which interacts with nucleoporins to help protein containing nuclear localization signal reach the nucleus through the nuclear pore [24-26]. There is a flexible link between the DBD and the HR-A/B domain. Based on the structural characteristics of the conserved DBD and HR-A/B domain, the *Hsfs* have been divided into three groups (A, B and C). The main differences between the three groups are as follows: group B proteins exhibit 7 amino acid residues in their HR-A/B domain, while group A has 28 amino acid residues in the relevant domain and group C had 14 amino acid residues in the same domain. In addition, the transcription activation domain (AHA) at the C-terminus is characteristic of group A, which guarantees the normal transcription of the *Hsfs* by binding to some basic transcription protein complexes. However, the *Hsfs* of group B and group C cannot maintain their activation activity due to the lack of an AHA motif [26, 27]. The repression domain (RD) is a peptide containing conserved amino acids (LFGV) at the C-terminus and mainly exists in group B [28].

*Hsfs* can specifically regulate the transcription of heat shock protein (*Hsp*) genes by specifically binding to the HSE in the promoter of an *Hsp* gene, and the *Hsp*, in turn, protect cells from stress and participate in protein folding [29, 30]. Some studies have confirmed that *Hsfs* are involved in the heat stress response. For example, the silencing of *HsfA1a* in tomato reduces the synthesis of heat stress-induced chaperone and *HsfA1a* proteins, thereby increasing the sensitivity of *HsfA1a*-silenced tomato plants to heat stress [31]. At 37 °C, *A. thaliana HsfA2*-mutant plants are more sensitive to heat stress than wild-type plants, which can be reversed by introducing the *HsfA2* gene [32]. The *OsHsfA4d*-mutant shows a phenotype of necrotic damage under high-temperature stress [13]. The expression of *OsHsfA2e* enhances high temperature and salt tolerance in *A. thaliana* [33]. In addition to heat stress, *Hsfs* are involved in plant growth and other biotic and abiotic stress responses. It is found that *HsfA9* is involved in embryo development and seed maturation in *A. thaliana* and *Helianthus annuus* [34]. Four *Hsf* genes (*HsfA1e*, *HsfA3*, *HsfA4a*, *HsfB2a* and *HsfC1*) in *A. thaliana* are strongly induced by salt, cold and osmotic stress [35-37]. The *HsfA2* in *A. thaliana* is involved in the response to oxidative stress [38]. The *HsfA4a* in *A. thaliana* can be used as an H<sub>2</sub>O<sub>2</sub> sensor [39,35]. The *OsHsfA4a* in *O. sativa* is associated with cadmium tolerance [40]. To date, there have been no reports of the cloning and functional analysis of *Cucurbita moschata Hsfs*.

*C. moschata* is rich in a variety of amino acids, vitamins, polysaccharides, pectin, and minerals and contains trigonelline, carotenoids and other biologically active substances and nutrients [41]. According to the Food and Agriculture Organization of the United Nations (<http://www.fao.org/home/en/>), pumpkin ranks the ninth in the output value of different vegetable crops in the world, with an annual sales value of 4 billion US dollars. China and India are the two main pumpkin producing countries in the world. China's cultivation area ranks second in the world, and its total output ranks first in the world [42]. During growth and development, unfavorable stress often causes great harm to the growth of pumpkin, resulting in a decline in pumpkin yield and quality [41]. Therefore, research on pumpkin resistance-related genes is increasingly important for pumpkin breeding and production. Because the *C. moschata* (*Rifu*) genome has been published [43], the *Hsf* family in *C. moschata* can now be subjected to systematic and comprehensive analysis. In this study, we provide information about the gene structural characteristics, gene duplications, chromosomal locations, evolutionary divergence and phylogenetic relationships of 36

*C. moschata* Hsf Furthermore, we analyze the digital expression profiles of 36 *CmHsfs* in response to numerous stresses. This study emphasizes the function of the *Hsfs* in various stress conditions and improves our understanding of the effects of polyploidization events on the evolution of the Hsf family.

## Results

### Identification of *Hsf* genes in *C. moschata* and their physical and chemical characteristics

A total of 36 *CmHsf* genes were identified after the removal of false positives and the same genes (Table 1), and they were designated *CmHsf1* to *CmHsf36* according to the starting positions of these genes on the chromosomes (from *Cmo\_Chr00* to *Cmo\_Chr20*, from top to bottom). The physicochemical parameters of each *CmHsf* were generated, and the predicted open reading frames (ORFs) ranged from 543 bp (*CmHsf32*) to 4380 bp (*CmHsf13*), with predicted proteins of 179-1458 amino acids. The physical and chemical parameters of these genes are similar to those seen in *A. thaliana* and *O. sativa* [44]. Furthermore, the molecular weights (MW) of these *CmHsfs* ranged from 20.5642 to 161.5554 kiloDaltons (*kDa*) (Table 1). Although the deduced heat shock transcription factors presented diversity in terms of the parameters mentioned above, most of the *CmHsfs* exhibited low isoelectric points (*pI*) (average 6.3) (Table 1). Subcellular localization prediction indicated that only 2 heat shock transcription factors (*CmHsf12* and *CmHsf17*) were predicted to be localized to the cell membrane, cytoplasm and nucleus, while the remaining *CmHsfs* were predicted to be localized to the nucleus.

### Classification and conserved domain analysis of 36 *CmHsfs*

To identify the phylogenetic relationships of the 36 *CmHsfs*, an unrooted phylogenetic tree was produced. These *CmHsfs* can be divided into three subfamilies (subfamily I, subfamily II and subfamily III; Fig. 1A) according to the amino acid sequence identity. Subfamily I (containing 21 members) was the largest group, and subfamily III included 13 members, while subfamily II presented the fewest members (2 members) (Fig. 1A). Furthermore, based on the structural characteristics of the conserved DBDs and HR-A/B domains, we can divide the 36 *CmHsfs* into three groups (A, B, and C) (Table 2). All *CmHsfs* contained a DBD and an HR-A/B domain (Table 2), and the DBD was composed of approximately 100 conserved amino acids (Additional file 2: Fig. S1). In addition, except for *CmHsf27* and *CmHsf32*, all of the *CmHsfs* contained an NLS. The *CmHsfs* in group A contained an AHA domain, while the *CmHsfs* in groups B and C did not contain an AHA domain, and only the proteins in Group B contained an RD (Table 2). To further reveal conserved domains, all *CmHsfs* were submitted to MEME, and 10 different motifs were identified (Fig. 1B; Additional file 2: Fig. S2). Overall, the *CmHsfs* exhibited 4-9 motifs, and motifs 1, 2 and 4 were present in all *CmHsf* proteins. Motif 3 was present in all proteins except for *CmHsf20* and *CmHsf5*. In addition, we found that motif 5 existed only in subfamily I, while motif 9 appeared only in subfamily III (Fig. 1B). The *CmHsfs* from the same clade usually present conserved domains or similar motif compositions, suggesting functional similarities among these proteins.

### Exon-intron analysis of 36 *Hsfs* in *C. moschata*

An exon-intron organization map of the 36 *CmHsf* genes was also produced (Fig. 2). Different numbers of exons (from 2 to 26) were found in the 36 *CmHsf* genes, suggesting that *CmHsfs* are quite diverse. In subfamily III, except for *CmHsf1*, *CmHsf10* and *CmHsf35*, which contained 9, 8 and 3 exons, respectively, the other *CmHsf* genes all contained 2 exons. *CmHsf* genes on the same branch usually presented similar intron-exon distributions, such as *CmHsf26\_CmHsf9*. Some genes in the same family exhibited significantly different intron-exon distributions. For example, *CmHsf12* contained 26 exons, which was different from the other *CmHsfs*, indicating that *CmHsf12* may have a special function.

### Chromosomal distribution and gene duplication of *Hsf* genes in *C. moschata*

Chromosomal distribution analysis in the genome revealed that the 36 *CmHsf* genes were unevenly distributed on 19 of the 21 chromosomes (Fig. 3). The chromosome *Cm\_Chr06* exhibited the most *CmHsf* genes, with 5 genes, followed by chromosome *Cm\_Chr05*, with 4 genes. A total of 3 genes were present on each of chromosomes *Cm\_Chr03*, *Cm\_Chr07* and *Cm\_Chr14*, and 2 genes were present on each of chromosomes *Cm\_Chr02*, *Cm\_Chr04*, *Cm\_Chr10*, *Cm\_Chr11* and *Cm\_Chr16*, while no genes were distributed on chromosomes *Cm\_Chr00*, *Cm\_Chr08* and *Cm\_Chr20*.

Two genes, whose putative amino acid identity is >85% and gene alignment coverage is >0.75, were defined here as a recently duplicated gene pair [45-46]. A total of 18 duplicated genes were identified and divided into nine groups, each of which contained two duplicated genes. Eight duplicated gene pairs were distributed on different chromosomes (Fig. 3), which demonstrated that segmental duplication events were involved in the expansion of the *CmHsf* genes. *CmHsf10* and *CmHsf12* were separated by a region of more than 100 kb, indicating that all duplicated gene pairs had undergone segmental duplication events. The *Ka* and *Ks* ratios were less than 1.0, which suggested that the pairs had evolved mainly under functional constraints with negative or purifying selection (Table 3). We also calculated evolutionary times and divergence times of the duplicated *C. moschata* *Hsf* gene pairs ranging from 10.17 to 65.74 million years ago (Mya), averaging 21.11 Mya (Table 3).

### Phylogenetic relationship of *Hsfs* in *C. moschata*, *C. sativa* and *A. thaliana*

To better evaluate the molecular evolution and phylogenetic relationship of plant Hsf, a phylogenetic tree of 79 Hsf proteins in *C. moschata*, *C. sativa* and *A. thaliana* was established. Based on the previous classification of *C. moschata* Hsf proteins (Fig. 1A), they were divided into 9 clades (Clade Ia-b, Clade II and Clade IIIa-e) (Fig. 4). Subfamily I was divided into Clade Ia and Clade Ib, and subfamily III was divided into Clade IIIa-e. This classification was consistent with the phylogenetic classification of *AtHsf* proteins [44]. In general, genes from subfamily I (Clade Ia and Clade Ib) (including 51 *Hsfs*) constituted the largest branch and accounted for 65% of the total *Hsfs*. Subfamily II contained 2 proteins. The remaining *Hsfs* belong to subfamily III and contain a total of 26 Hsf proteins. From the perspective of phylogenetic branch, the homology of *Hsfs* between *C. moschata* and *C. sativa* was higher than that between *C. moschata* and *A. thaliana*, which was consistent with the evolutionary rules of the three species.

### Synteny analysis of *Hsf* genes in *C. moschata*

According to the synteny analysis of *Hsfs* in *C. moschata* and 5 other species (*A. thaliana*; *Lagenaria siceraria*; *Cucumis sativus*; *Cucurbita maxima*; *Citrullus lanatus*), we found that *C. lanatus* exhibited the most *Hsf* homologous genes (56), followed by *L. siceraria* (52), *C. maxima* (51) and *C. sativus* (51). *A. thaliana* presented the fewest (18) homologous genes (Fig. 5). Furthermore, the syntenic genes of the *CmHsfs* could be found on all chromosomes of *A. thaliana*, *L. siceraria*, *C. sativus*, *C. maxima*, and *C. lanatus*, indicating that the *CmHsfs* have remained closely related to those of these five species during the process of evolution. In addition, we found that certain *CmHsf* genes on chromosomes Cm\_Chr02, Cm\_Chr06, Cm\_Chr08, and Cm\_Chr016 corresponded to two or more *Hsf* genes in *A. thaliana*. This phenomenon was more fully reflected in the collinear diagram of *C. moschata* with *L. siceraria*, *C. sativus*, *C. maxima* and *C. lanatus*. In general, the collinear relationship between *C. moschata* and (*L. siceraria*, *C. sativus*, *C. maxima* or *C. lanatus*) was closer than that for *A. thaliana*, suggesting that these species may have originated from the same ancestor. The collinear analysis showed that *C. moschata* and (*L. siceraria*, *C. sativus*, *C. maxima*, and *C. lanatus*) had frequent collinearity (Fig. 5), indicating that genes with collinear relationship may have similar functions.

### Expression pattern of *Hsf* genes in *C. moschata*

To understand the physiological role of *CmHsfs*, we analysed the expression patterns of 36 heat shock transcription factors in the roots, stems, cotyledons and true leaves of *C. moschata* via quantitative *real-time PCR*. The transcriptional abundance of 36 *C. moschata* heat shock transcription factors can be obtained from at least one of the four tissues (Fig. 6; Additional file 1: Table S1). Heat map and cluster analyses showed that 21 *CmHsfs* were highly expressed in cotyledons and true leaves, such as *CmHsf4*, *CmHsf32*, *CmHsf35*, *CmHsf19* and *CmHsf15*. Two genes (*CmHsf9* and *CmHsf10*) were expressed more highly in the roots and stem than in the cotyledons and true leaves. Some genes were highly expressed only in one tissue. For example, *CmHsf23* was mainly expressed in the roots, and its relative expression level was 100-258 times that in other tissues. Based on the above analysis, 36 heat shock transcription factors showed tissue specificity.

### *Cis*-acting element analysis of *Hsf* genes in *C. moschata*

To explore the potential function of *Hsfs*, the *cis*-elements in the promoters (2 kb before the start codon) of the 36 *Hsf* genes in *C. moschata* were predicted. A total of 429 *cis*-elements were found among all *CmHsfs*. They were involved in 9 abiotic stresses, including showing salicylic acid responsiveness, defence and stress responsiveness, low-temperature responsiveness, abscisic acid responsiveness, gibberellin responsiveness, MeJA responsiveness, auxin responsiveness, drought inducibility and wound responsiveness (Fig. 7A; Additional file 1: Table S2). A total of 31% of the 429 *cis*-acting elements were involved in abscisic acid responsiveness, which existed in 32 of the 36 *CmHsfs* (Fig. 7B, 7C). In addition, 27% and 45% of the *cis*-acting elements were MeJA response elements (harboring CGTCA and TGACG motifs) and auxin response elements, respectively (Fig. 7B). Among the 36 heat shock transcription factors, 28 genes were involved in the MeJA response, and 22 genes were involved in the auxin response. A total of 14 heat shock transcription factors exhibited low-temperature response elements. Since the *Hsf* genes involved in abscisic acid responsiveness, low-temperature responsiveness, MeJA responsiveness and auxin responsiveness account for a high proportion of these genes, we speculated that these genes might play important roles in these stresses.

By analyzing the *cis*-acting elements of individual genes, we found that both *CmHsf34* and *CmHsf27* contained 12 abscisic acid response elements (Additional file 1: Table S2). In addition, *CmHsf17*, *CmHsf26*, *CmHsf9* and *CmHsf35* contained 8 MeJA response elements, and *CmHsf23* and *CmHsf35* contained the greatest number (3) of low-temperature response elements, which indicates that these key *CmHsfs* may play an important role in the corresponding stress response.

### The response of *CmHsf* genes to temperature stress

To explore the response of *CmHsfs* to temperature stress, we cultured *C. moschata* seedlings at 4 °C and 38 °C. Under cold treatment, 44% of the *CmHsfs* (16 genes) were significantly upregulated, and 27% of the *CmHsfs* (10 genes) were significantly downregulated (Fig. 8; Additional file 1: Table S3). For instance, *CmHsf3*, *CmHsf5*, *CmHsf23*, *CmHsf24*, *CmHsf27*, *CmHsf35* and *CmHsf36* were highly expressed under cold stress. In addition, the *CmHsf4*, *CmHsf15*, *CmHsf31* and *CmHsf32* genes exhibited low expression levels under cold stress. At the same time, two genes (*CmHsf28* and *CmHsf30*) were not expressed under cold stress, indicating that the expression of these genes may be limited under cold stress. Under heat treatment, 24 genes were significantly upregulated, and 12 genes were significantly downregulated (Fig. 8; Additional file 1: Table S3). The expression levels of *CmHsf9* and *CmHsf31* under heat stress were 128.38 and 66.39 times those in the control plants, respectively, suggesting that these two genes may play important roles under heat stress. Some genes presented low expression levels under heat treatment, such as *CmHsf17*, *CmHsf11*, *CmHsf21*, *CmHsf22*, *CmHsf23* and *CmHsf35*. Considering the expression levels of the *CmHsf* genes under cold and heat stress together, we found that *CmHsf9*, *CmHsf11*, *CmHsf21*, *CmHsf23*, *CmHsf31*, *CmHsf34* and *CmHsf35* showed opposite trends under the two stresses, so we speculate that these genes may play important roles in temperature stress.

### The response of *CmHsf* genes to hormones and salicylic acid

According to the prediction of *cis*-acting elements in the *CmHsfs* promoter, a total of 28, 32, and 19 *CmHsf* genes were found to be involved in the MeJA response, abscisic acid responsiveness and salicylic acid responsiveness, respectively (Fig. 7; Additional file 1: Table S2). Therefore, we analysed the responses of these genes to MeJA, ABA, and SA. The results of qRT-PCR analysis showed that 31 *CmHsfs* responded to MeJA to varying degrees, and the expression of *CmHsf20* was 5.1 times that in the control (Fig. 9; Additional file 1: Table S4). Under ABA treatment, 21 *CmHsfs* were significantly upregulated, and 15 genes were significantly downregulated. The expression levels of *CmHsf3*, *CmHsf4*, *CmHsf5*, *CmHsf6*, *CmHsf7*, *CmHsf8*, *CmHsf12*, *CmHsf25*, *CmHsf29* and *CmHsf31* under ABA stress were 20~86 times those of the control plants, indicating that these genes play important roles under ABA stress. All *CmHsfs* responded to SA, among which *CmHsf25*, *CmHsf27*, *CmHsf29* and *CmHsf32* were significantly increased under SA treatment, while *CmHsf1*, *CmHsf2*, *CmHsf23* and *CmHsf28* were significantly decreased under SA treatment. Based on the above analysis, we conclude that *CmHsf* family genes are involved in multiple stresses and may play different roles in these stresses.

## Discussion

Heat shock transcription factors are broadly present in all plants and are considered to be important regulators of abiotic stress. The Hsf family has been comprehensively and systematically analyzed in *G. max* [17], *B. rapa* [18], *P. breitschneideri* [19], *S. tuberosum* [20], *V. vinifera* [21] and *B. oleracea* [22]. However, the Hsf family has not been extensively studied in *C. moschata*.

In our study, we identified 36 *Hsf* genes in *C. moschata* via genome-wide analysis (Table 1). The analysis of the physical and chemical properties of the gene family can show the diversity of each member in the process of evolution [45]. Our results showed that the MW and the number of amino acids of 36 *CmHsfs* vary widely (Table 1), which indicates that *C. moschata* changes in the process of evolution. Most of the *CmHsfs* exhibited low isoelectric points (*pI*) (average 6.3), this result is similar to the report of Hsf in *C. sativa* [16]. Regardless of the size and domain composition of proteins, the characteristics of low *pI* are preserved, indicating that *CmHsf* proteins should be negatively charged at physiological pH. Through predictive analysis of subcellular location, it was found that most of the *CmHsfs* were predicted to be localized to the nucleus (Table 1), which indicated that their functions were indeed as transcription factors. But *CmHsf12* and *CmHsf17*, from the same subfamily, were predicted to be localized to cell membrane, cytoplasm and nucleus, indicate that *CmHsf* members in the same subfamily do not necessarily correspond to the same subcellular location, and they might have other special function.

The phylogenetic tree divided 36 *CmHsfs* into 3 subfamilies (Subfamily I, Subfamily II and Subfamily III), most of the genes within the same subfamily shared similar gene structures in terms of either exon length or intron number (Fig. 2). Therefore, we speculated that the *CmHsfs* in one branch may have similar functions, and this feature was similar to that previously reported in other species [16, 18]. The structural characteristics of some *CmHsfs* in the same branch are different from those of other *CmHsf* genes, indicating that these genes may have functional diversity. In addition, The *CmHsfs* were also divided into three groups (groups A, B and C) based on the conserved structural characteristics of the DBD and the HR-A/B domain (Table 2). Subfamily II corresponded to group C and subfamily III corresponds to group B, subgroup I contained not only group A genes but also group C genes. Due to the close homology of the genes on the same branch, we speculate that the evolutionary path of the *CmHsfs* has been changing.

The conserved motifs of *CmHsfs* protein were also predicted and analyzed (Fig. 1). It was found that motifs 1, 2 and 4 exist in all *CmHsf* proteins (Fig. 1). According to the comprehensive analysis of the motif position and conserved domain position of 36 *CmHsf* protein, we found that motif 1, motif 2 and motif 3 together formed the DBD, and motif 4 formed the HR-A/B domain (Fig. 1; Table 2). The result is consistent with the previous reports in *Z. mays* [15], *C. sativa* [16], *B. rapa* ssp. *pekinensis* [18], *S. tuberosum* [20], which indicates that *CmHsfs* may have similar functional characteristics.

In some species, the number of members of a specific gene family is considered to be the result of natural evolution. At the same time, the diversity of gene family members is generally due to genome recombination and amplification [46]. Chromosomal segmental duplications and individual gene duplications are a major driving force in the genome evolution process [47]. Compared with the 25 reported *ZmHsfs* [15], 21 *CsHsfs* [16] and 31 *PtHsfs* [14], we found that the number of *Hsf* genes in *C. moschata* is greater than those in *Z. mays*, *C. sativa* and *P. trichocarpa*. Genome sizes vary significantly in these species; for instance, the genome size of *C. moschata* is 197.83 Mb, and that of *Z. mays* is 2300 Mb. The maize genome size is 11 times that of *C. moschata*. However, the number of maize *Hsf* genes is much lower than the number of *Hsf* genes in *C. moschata*. The reason for this difference might be that although two rounds of gene duplication occurred in the *Z. mays* genome during its evolution [48, 49], the *Hsf* genes of *Z. mays* underwent large gene losses. In addition, the genome of *C. moschata* also underwent a whole-genome duplication (WGD) event during the phylogeny of the species [43]. For *C. sativa*, the genome size is 350 Mb, but 21 *CsHsfs* was less than the number of *CmHsfs*. We speculated that gene duplication promotes the amplification of *CmHsf* genes [43] or gene degeneration and mutation promotes the reduction of the number of *CsHsf* genes [16], ultimately resulting in the number of *CmHsf* genes more than that of other plants.

In this study, all *CmHsf* gene pairs were found to have experienced segmental duplication events, with no tandem duplication events, indicating that segmental duplication has played an important role in the evolution of the *C. moschata* *Hsf* gene family (Fig. 3). The *Ka* and *Ks* ratios of all duplicated pairs indicated that these gene pairs were under purifying selection. Additionally, the relatively high *Ka/Ks* ratios for *CmHsf12-CmHsf10* suggested that they have experienced rapid evolution (Table 3).

A study proposes three hypotheses to explain the fate of duplicated genes: (1) In the process of plant evolution, sometimes gene degeneration and mutation occur, which often leads to the loss of copy function of some duplicated genes. (2) Due to the diversity and directionality of mutations, one copy of the duplicated gene may mutate and retain its new function during evolution, while the other copy retains its original function. This process is called new functionalization. (3) Two copies of the duplicated gene may mutate to obtain different functions, which is called subfunctionalization [50]. According to the different expression patterns of *CmHsf26* and *CmHsf9* genes, it can be inferred that there are differences between the duplicated genes. *CmHsf26* is highly expressed in cotyledon and true leaf, while *CmHsf9* gene is highly expressed in root and stem (Fig. 6). Their gene structure and motif composition are similar, which indicates that the subfunctionalization of duplicated genes in *CmHsf* gene family may change the gene expression pattern (Fig. 1; Fig. 2). In addition, the duplicated genes *CmHsf30* and *CmHsf16* have similar intron-exon structure, the same motif component, and the similar tissue expression pattern, but there are obvious differences in temperature stress and hormone treatment (Fig. 2; Fig. 6; Fig. 8; Fig. 9), which indicates that the new functionalization of the duplicated genes in the *CmHsf* gene family may play a key role. The collinear analysis showed that *C. moschata* had frequent collinearity with *L. siceraria*, *C. sativus*, *C. maxima*, and *C. lanatus* (Fig. 5), indicating that genes with collinear relationship may have similar functions.

*Cis*-acting elements are essential for gene expression, and their numbers are correlated with gene expression intensity [51, 52]. *CmHsf23* and *CmHsf35* contain three low-temperature response elements (Fig. 7), which mean that *CmHsf23* and *CmHsf35* may play key roles under low-temperature stress. The qRT-PCR results showed that *CmHsf23*, *CmHsf21*, *CmHsf11*, and *CmHsf35* were significantly upregulated under low temperature, and the expression profiles of these genes showed opposite trends under high-temperature stress, which further verified the response of these genes to temperature stress (Fig. 8). However, *CmHsf13*, *CmHsf36*, *CmHsf3* and *CmHsf5* were significantly induced under cold stress and heat stress (Fig. 8), and their responses were more prominent under cold stress, which indicated that these genes were highly sensitive to temperature and might play a key role under temperature stress. The prediction of *cis*-acting elements showed that the promoters of 28 *CmHsf* genes contained MeJA response elements (Fig. 7), and qRT-PCR analysis showed that the

expression levels of 31 genes changed to varying degrees under MeJA treatment (Fig. 9). However, from the relative expression values, we found that the *CmHsfs* responded less to MeJA than to ABA and SA (Fig. 9). Therefore, we concluded that *C. moschata* Hsf family genes were mainly involved in the response to ABA and SA.

## Conclusions

In summary, we identified 36 *Hsfs* in the *Cucurbita moschata* genome based on a thorough analysis and provided genetic information such as chromosome locations and exon-intron structures, conserved domains, and duplicated genes. We specifically examined the expression profiles of these *CmHsfs* in different tissues. At the same time, we examined the responses of *CmHsfs* to multiple stresses, and several key genes were found to respond to adverse environments.

## Methods

### Sequence retrieval from the *Cucurbit* genomics database and physicochemical characterization

To identify the heat shock transcription factor family in *C. moschata*, the genome was downloaded from the *Cucurbit* genomics database (CuGenDB, <http://cucurbitgenomics.org/>) [43]. A total of 25 *A. thaliana* Hsf genes were obtained from the NCBI database by using their gene IDs from *A. thaliana* references [26]. We used 25 AtHsf proteins as queries to search against the *Cucurbit* genomics database using BLASTP with an *e*-value cut-off of  $1 \times 10^{-10}$ . To eliminate false positives, sequences were discarded if they constituted <70% of the corresponding *A. thaliana* Hsf protein. SMART (<http://smart.embl-heidelberg.de/>) [53] and MARCOIL (<http://toolkit.tuebingen.mpg.de/marcoil>) [19] were used to predict the DBDs and HR-A/B domains. After the removal of the same genes, the remaining genes were identified as *CmHsf* genes. The coding sequence and protein sequence information for each of the *CmHsfs* were shown in Additional file 1: Table S6.

The physical and chemical characteristics of the heat shock transcription factors, including their theoretical molecular weight (MW), theoretical isoelectric point (*pI*) and the number of amino acids, were analyzed with ExPASy (<http://web.expasy.org/tools/>) [54]. Information on *CmHsf* genes including their chromosomal distribution, their start and the end positions on the chromosomes were extracted from the *Cucurbit* genomics database, and their subcellular locations were predicted with Plant-mPLoc [55].

### Phylogenetic tree construction

To reveal the phylogenetic relationships of *Hsf* genes in *C. moschata*, an unrooted phylogenetic tree was constructed with MEGA 5.0 [56] according to the similarity of full-length amino acid sequence of 36 *CmHsfs*. In addition, the phylogenetic relationship of Hsf protein from *C. moschata*, *C. sativa* and *A. thaliana* was also constructed by MEGA 5.0. The protein sequences of 21 CsHsfs and 22 AtHsfs were obtained based on previous literature [44, 57]. The unrooted Neighbor-Joining (NJ) method was used to construct the phylogenetic tree, and the bootstrap values were obtained using 1000 replicates with the pairwise deletion option.

### Analysis of conserved domains and gene structure

The conserved motifs of Hsf in *C. moschata* were obtained on the Multiple Expectation Maximization or Motif Elicitation (MEME, <http://meme-suite.org/>) [58] using the protein sequences, and the LOGOs (Additional file 2: Fig. S2) of the protein motifs were also obtained with MEME. The NLSs and NESs of the heat shock transcription factors were predicted by using cNLS Mapper [59] and the NetNES 1.1 Server [60], respectively. The exon-intron structures were obtained from GSDS (Gene Structure Display Server, <http://gsds.cbi.pku.edu.cn/>) [61] by comparing the cDNA sequences and its corresponding genomic DNA sequences of *CmHsfs* members.

### Gene duplication and gene collinearity analysis

The chromosomal locations of the *CmHsf* genes were mapped and imaged with visualization tools (<http://visualization.ritchielab.psu.edu/home/index>) based on their initial positional information obtained from *C. moschata* (CuGenDB, <http://cucurbitgenomics.org/>). To identify gene duplications, all CDS sequences of *C. moschata* Hsf genes were subjected to BLAST searches against each other (Identity >85%, E-value < $1 \times 10^{-10}$ ) by using the Local Blast program. Gene alignment coverage was then acquired by pair-wise alignment using the previously calculated method: Gene alignment coverage = (alignment length - mismatches)/length of the longer gene. Pairs were considered duplications when the gene alignment coverage was greater than 0.75. Moreover, two genes that were separated by several genes in a 100-kb were named as tandemly duplicated genes [62]. To estimate the divergence of these duplicated *CmHsf* genes, we used the KaKs calculator to calculate the synonymous substitution ratio (*Ks*) according to the method of Gojo-bori and Nei [63]. To avoid the saturation of substitutions, we required that *Ks* values >2.0 must be discarded [64, 65]. The divergence time (*T*) was computed according to the formula ( $T = Ks/2\lambda \times 10^{-6}$  million years ago (Mya),  $\lambda = 1.5 \times 10^{-8}$ ) in the previous literature [66]. The criteria for identifying gene collinearity were based on previous reports [67], and the synteny relationships between the heat shock transcription factors of *C. moschata* and those of other species (*A. thaliana*, *C. sativus*, *C. maxima*, *C. lanatus*, *L. siceraria*) were constructed using Advanced Circos program in TBtools [68].

### Analysis of *cis*-acting elements of *CmHsf* gene promoters

The promoter sequences (2 kb before the start codon) of all *CmHsf* genes were extracted from the *Cucurbit* genome database (<http://cucurbitgenomics.org/>), and we predicted the promoter *cis*-acting elements of *CmHsfs* by using PlantCare program (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) [69] and visualized by Simple BioSequence Viewer in TBtools [68].

### Plant material, growth conditions and stress treatment

The *C. moschata* variety “Tianmi 1” was used as the study material. The seeds were provided by the pumpkin team of School of Horticulture and Landscape Architecture, Henan Institute of Science and Technology. The seeds were sown in a tray containing a vermiculite-matrix (2:1) mixture and grown in a plant growth chamber. The artificial growth conditions were set as light intensity of 350  $\mu\text{mol}/\text{m}^2/\text{sec}$ , 25 °C 16 h light / 16 °C 8 h dark and 65% relative humidity. We sampled and analyzed different tissues (roots, stems, cotyledons and true leaves) of two-month-old seedlings. In addition, some of the seedlings were transferred to 38 °C for 6 h heat treatment, or transferred to 4 °C for 6 h cold treatment. Another portion of the seedlings was cultured in 1/2 Hoagland solution, pH 6.5. After 5 days of adaptation, the plants were cultured with the following treatments: (1) control (untreated plants); (2) 1 mM MeJA; (3) 5 mM salicylic acid (SA); (4) 100  $\mu\text{M}$  abscisic acid (ABA) [70]. Leaf samples were collected at 10 h after the above treatments. Control and stress-treated samples were frozen in liquid nitrogen and stored at -70 °C for further analysis.

### RNA extraction, reverse transcription and qRT-PCR analysis

Total RNA was extracted from the frozen samples according to the instructions of the RNA kit (Tiangen, Beijing). Moreover, the RNA was isolated and then reverse transcribed into cDNA using a Prime Script RT reagent kit (TaKaRa, Dalian, China). Finally, quantitative real-time PCR was performed using the SYBR Premix ExTaq kit (TaKaRa, Dalian). To verify the specificity of gene primers, the target genes and the reference gene ( $\beta$ -Actin) primers (Additional file 1: Table S5) were aligned at the *Cucurbit* genome database. The qRT-PCR analysis was performed on an ABI7500 Real-Time PCR System (Applied Biosystems) with the following cycling profile: stage 1, 95 °C 20 s; stage 2, 95 °C 3 s, 60 °C 30 s (40 cycles); stage 3, 95 °C 15 s, 60 °C 1 min, 95 °C 15 s. Stage 3 was used to perform a melting curve. Experimental repeats were run for three technical and three biological replicates. The relative gene expression was calculated according to the  $2^{-\Delta\Delta\text{Ct}}$  method.

## Abbreviations

*Hsfs*: Heat shock transcription factors; DBD: DNA-binding domain; OD: Oligomerization domain; NLS: Nuclear localization signal; NES: Nuclear export signal; HSE: Heat shock element; RD: Repressor domain; Hsp: Heat shock protein; *Cm*: *C. moschata*; kDa: KiloDaltons; *pI*: Isoelectric points; WGD: Whole-genome duplication; Mw: Molecular weight; MEME: Multiple expectation maximization or motif elicitation; CDS: Coding domain sequence; *Ks*: Synonymous substitution ratio; Mya: Million years ago; SA: Salicylic acid; ABA: Abscisic acid

## Declarations

### Acknowledgement

Not applicable.

### Authors' contributions

CS and JY conceived, designed and supervised the experiment; CS and JP wrote the manuscript; CS and JP performed the experiment; CS and JY provided support in lab experiment and data analysis. CS and JY analyzed the data. All authors read and approved the manuscript.

### Funding

This work was funded by the Scientific Research Foundation for High - level Talent (103010620001/015 and 2017034).

### Availability of data and materials

All data generated or analysed during this study are included in this published article and its Additional files.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>School of Resources and Environmental Sciences, Henan Institute of Science and Technology, Xinxiang 453003, China; <sup>2</sup>School of Horticulture and Landscape Architecture, Henan Institute of Science and Technology, Xinxiang, China. <sup>3</sup>Henan Province Engineering Research Center of Horticultural Plant Resource Utilization and Germplasm Enhancement, Xinxiang, 453003, China

## References

1. Wang M, Vannozzi A, Wang G, Liang YH, Tornielli GB, Zenoni S, Cavallini E, Pezzotti M, Cheng ZM. Genome and transcriptome analysis of the grapevine (*Vitis vinifera* L.) WRKY gene family. *Hortic Res.* 2014;1(1):1-16.
2. Gomez-Pastor R, Burchfiel ET, Thiele DJ. Regulation of heat shock transcription factors and their roles in physiology and disease. *Nat Rev Mol Cell B* 2018;19(1):4-19.
3. Kotak S, Larkindale J, Lee U, Koskull-Döring PV, Vierling E, Scharf KD. Complexity of the heat stress response in plants. *Curr Opin Plant Biol.* 2007;10(3):310-6.
4. Baniwal S K, Bharti K, Chan KY, Fauth M, Ganguli A, Kotak S, Mishra SK, Nover L, Port M, Scharf KD, Tripp J, Weber C, Zielinski D, Koskull Doring P. von. Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. *J Biosciences.* 2004;29(4):471-87.
5. Wiederrecht G, Seto D and Parker CS. Isolation of the gene encoding the *cerevisiae* heat shock transcription factor. *Cell.* 1988;54(6):841-53.
6. Sorger PK and Pelham HRB. Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. *Cell.* 1988;54:855-64.
7. Clos J, Westwood JT, Becker PB, Wilson S, Lambert U, Wu C. Molecular cloning and expression of a hexameric drosophila heat stress factor subject to negative regulation. *Cell.* 1990;63(5):1085-97.
8. Rabindran SK, Giorgi G, Clos J, Wu C. Molecular cloning and expression of a human heat stress factor. *P Natl Acad Sci USA.* 1991;88(16):6906-10.
9. Sarge KD, Zimarino V, Holm K, Wu C, Morimoto RI. Cloning and characterization of two mouse heat stress factors with distinct inducible and constitutive DNA binding ability. *Gene Dev.* 1991;5(10):1902-11.
10. Schuetz TJ, Gallo GJ, Sheldon L, Tempst P, Kingston RE. Isolation of a cDNA for *HSF2*: evidence for two heat stress factor genes in humans. *P Natl Acad Sci USA.* 1991;88(16):6911-5.
11. Scharf K D, 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. *EMBO J.* 1990;9(13):4495-501.
12. Huhel A, Schoffl F. *Arabidopsis* heat shock factor: isolation and characterization of the gene and the recombinant protein. *Plant Mol Biol.* 1994;26(1):353-62.
13. Yamanouchi U, Yano M, Lin H, Ashikari M, Yamada K. A rice spotted leaf gene, *Sp17*, encodes a heat stress transcription factor protein. *P Natl Acad Sci USA.* 2002;99(11):7530-5.
14. Zhang HZ, Yang JL, Chen YL, Mao XL, Wang ZC, Li CH. Identification and expression analysis of the heat shock transcription factor (*HSF*) gene family in *P. trichocarpa*. *Plant Omics.* 2013;6(6):415-24.
15. 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;12:1-14.
16. Chen XZ, Wang Y, Shi JL, Zhu LJ, Wang KL, Xu J. Genome-wide identification, sequence characteristic and expression analysis of heat shock factors (HSFs) in cucumber. *Hereditas.* 2014;36(4):376-86. (In Chinese).
17. Chung E, Kim KM, Lee JH. Genome-wide analysis and molecular characterization of heat shock transcription factor family in *Glycine max*. *J Genet Genomics.* 2013;40(3):127-35.
18. 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):541-51.
19. Qiao X, Li M, Li L, Yin H, Wu J, Zhang S. Genome-wide identification and comparative analysis of the heat shock transcription factor family in Chinese white pear (*Pyrus bretschneideri*) and five other *Rosaceae* BMC Plant Biol. 2015;15(1):1-16.
20. Tang RM, Zhu WJ, Song XY, Lin XZ, Cai JH, Wang M, Yang Q. Genome-wide identification and function analyses of heat shock transcription factors in potato. *Front Plant Sci.* 2016;7:1-18.
21. Liu GT, Chai FM, Wang Y, Jiang J Z, Duan W, Wang YT, Wang FF, Li SH, Wang LJ. Genome-wide identification and classification of HSF family in grape and their transcriptional analysis under heat acclimation and heat stress. *Hortic Plant J.* 2018;4:7-17.
22. Lohani N, Golicz AA, Singh MB, Bhalla PL. Genome-wide analysis of the *Hsf* gene family in *B. oleracea* and a comparative analysis of the *Hsf* gene family in *oleracea*, *B. rapa* and *B. napus*. *Funct Integr Genomic.* 2019;19:515-31.
23. Scharf KD, Berberich T, Ebersberger I, Nover L. The plant heat stress transcription factor (*Hsf*) family: structure, function and evolution. *BBA-Gene Regul Mech.* 2012;1819(20):104-19.
24. Gorlich D, Kutay U. Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Bi.* 1999;15(1):607-60.
25. Heerklotz D, Doring 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 Cel Biol.* 2001;21(5):1759-68.
26. Kotak S, Port M, Ganguli A, Bicker F, Koskull-Doring 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.
27. Doring P, Treuter E, Kistner C, Lyck R, Chen A, Nover L. The role of AHA motifs in the activator function of tomato heat stress transcription factors *HsfA1* and *HsfA2*. *Plant Cell.* 2000;12(2):265-78.
28. Ikeda M, Ohme-Takagi M. A novel group of transcriptional repressors in *Arabidopsis*. *Plant Cell Physiol.* 2009;50(5):970-5.

29. Sakuma Y, Maruyama K, Qin F, Osakabe Y, Shinozaki K, Yamaguchi Shinozaki K. Dual function of an *Arabidopsis* transcription factor *DREB2A* in water-stress-responsive and heat-stress-responsive gene expression. *Proc Natl Acad Sci USA*. 2006;103(49):18822-7.
30. Qin F, Kakimoto M, Sakuma Y, Maruyama K, Osakabe Y, Tran LS, Shinozaki K, Yamaguchi-Shinozaki K. Regulation and functional analysis of *ZmDREB2A* in response to drought and heat stress in *Zea mays* *Plant J*. 2007;50(1):54-9.
31. Mishra SK, Tripp J, Winkelhaus S, Tschiersch B, Theres K, Nover L, Scharf KD. 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.
32. Chang YY, Liu HC, Liu NY, Chi WT, Wang CN, Chang SH, Wang TT. A heat-inducible transcription factor, *HsfA2*, is required for extension of acquired thermotolerance in *Arabidopsis*. *Plant Physiol*. 2007;143(1):251-62.
33. Yokotani N, Ichikawa T, Kondou Y, Matsui M, Hirochika H, Iwabuchi M, Oda K. Expression of rice heat stress transcription factor *OsHsfA2e* enhances tolerance to environmental stresses in transgenic *Arabidopsis*. *Planta*. 2008;227(5):957-67.
34. Almoquera C, Rojas A, Diaz-Martin J, Prieto-Dapena P, Carranco R, Jordano J. A seed-specific heat-shock transcription factor involved in developmental regulation during embryogenesis in sunflower. *J Biol Chem*. 2002;277(46):43866-72.
35. Miller G, Mittler R. Could heat shock transcription factors function as hydrogen peroxide sensors in plants? *Ann Bot*. 2006;98(2):279-88.
36. Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J*. 2007;50(2):347-63.
37. Zhang J, Liu B, Li JB, Zhang L, Wang Y, Zheng HQ, Lu MZ, Chen J. *Hsf* and *Hsp* gene families in *Populus*: genomewide identification, organization and correlated expression during development and in stress responses. *BMC Genomics*. 2015;16(1):1-16.
38. Nishizawa A, Ybata Y, Yoshida E, Maruta T, Yoshimura K, Shigeoka S. *Arabidopsis* heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. *Plant J*. 2006;48(4):535-47.
39. Davletova S, Rizhsky L, Liang H, Shengqiang, Z, Oliver DJ, Couto J, Shulaev V, Schlauch K, Mittler R. Cytosolic ascorbate peroxidase 1 is a central component of the reactive oxygen gene network of *Arabidopsis*. *Plant Cell*. 2005;17(1):268-81.
40. Shim D, Hwang JU, Lee J, Lee S, Shoi Y, An G, Martinoia E, Lee Y. Orthologs of class A4 heat shock transcription factor *HsfA4a* confer cadmium tolerance in wheat and rice. *Plant Cell*. 2009;21(12):4031-4034.
41. Zhou J Studies on screening and physiological characteristics of salt-tolerant root stocks from *Cucurbita moschata* Duch. Germplasm. Nanjing Agricultural University. 2008 (In Chinese).
42. Wang JP, Sun PC, Li YX, Liu YZ, Yang NS, Yu JG, Ma XL, Sun SR, Xia RY, Liu XJ, Ge DC, Luo SN, Liu YM, Kong YT, Cui XB, Lei TY, Wang L, Wang ZY, Ge WN, Zhang L, Song XM, Yuan M, Guo D, Jin DC, Chen W, Pan YX, Liu T, Yang GX, Xiao Y, Sun JS, Zhang C, Li ZB, Xu HQ, Duan XQ, Shen SQ, Zhang ZH, Huang SW, Wang XY. An overlooked paleotetraploidization in *Cucurbitaceae*. *Mol Biol Evol*. 2018;35:16-26.
43. Sun HH, Wu S, Zhang GY, Jiao C, Guo SG, Ren Y, Zhang J, Zhang HY, Gong GY, Jia ZC, Zhang F, Tian JX, Lucas WJ, Doyle JJ, Li HZ, Fei ZJ, Xu Y. Karyotype stability and unbiased fractionation in the paleo-allotetraploid *cucurbita* *Mol Plant*. 2017;10(10):1293-306.
44. 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 J Genet Genomics](#). 2008;35:105-18.
45. Song XM, Liu TK, Duan WK, Ma QH, Ren J, Wang Z, Li Y, Hou XL. Genome-wide analysis of the gras gene family in Chinese cabbage (*Brassica rapa pekinensis*). *Genomics*. 2014;103(1):135-46.
46. Wang C, Duan WK, Riquicho AR, Jing ZG, Liu TK, Hou XL, Li Y. Genome-wide survey and expression analysis of the PUB family in Chinese cabbage (*Brassica rapa pekinensis*). *Mol Gen Genet*. 2015;290:2241-60.
47. Du JC, Tian ZX, Sui Y, Zhao MX, Song QJ, Cannon SB, Cregan P, Ma JX. Pericentromeric effects shape the patterns of divergence, retention, and expression of duplicated genes in the paleopolyploid soybean. *Plant Cell*. 2012;24(1):21-32.
48. Swigoňová Z, Lai J, Ma J. Close split of sorghum and maize genome progenitors. *Genome Res*. 2004;14(10):1916-23.
49. Wei FH, Coe EH, Nelson W, Bharti AK, Engler F, Butler E, Kim H, Goicoechea JL, Chen MS, Lee S, Fuks G, Sanchezvilleda H, Schroeder SA, Fang ZW, McMullen MS, Davis GL, Bowers JE, Paterson AH, Schaeffer ML, Gardiner JM, Cone KC, Messing J, Soderlund C, Wing Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLoS Genet*. 2007;3(7):1254-63.
50. Makkena S, Lamb RS. The bHLH transcription factor SPATULA regulates root growth by controlling the size of the root meristem. *BMC Plant Biol*. 2013;13:11.
51. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Quackenbush. TM4 microarray software suite. *Method Enzymol*. 2006;411(2):134-93.
52. Peng S, Huang ZC, Ou YLJ, Cheng J, Zeng FH. Research progress of artificial promoter in plant genetic engineering. *J Plant Physiol*. 2011;47:141-6.
53. Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res*. 2012;40:302-5.
54. Elisabeth G, Alexandre G, Christine H, Ivan I, Appel RD, Amos B. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 2003;31(13):3784-8.
55. Chou KC, Shen HB. Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS ONE*. 2010;5(6):e11335.
56. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33(7):1870-4.
57. Zhou S, Zhang P, Jing Z, Shi J. Genome-wide identification and analysis of heat shock transcription factor family in cucumber (*Cucumis sativus*). *Plant Omics J*. 2013;6(6):449-55.
58. Bailey TL, Elkan C. The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol*. 1995;3:21-9.

59. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Systematic identification of yeast cell cycle-dependent nucleocytoplasmic shuttling proteins by prediction of composite motifs. *P Natl Acad Sci USA*. 2009;106(25):10171-6.
60. La Cour T, Kierner L, Molgaard A, Gupta R, Skriver K, Brunak S. Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng Des Sel*. 2004;17(6):527-36.
61. Hu B, Jin J, Guo AY, Zhang H, Luo JC, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 2015;31(8):1296-7.
62. Wang LQ, Guo K, Li Y, Tu YY, Hu HZ, Wang BR, Cui XC, Peng LC. Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. *BMC Plant Biol*. 2010;10(1):1-16.
63. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. KaKs calculating *Ka* and *Ks* through model selection and model averaging. *Genom Proteom Bioinf*. 2006;4(4):259-63.
64. Blanc G, Wolfe K Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*. 2004;16(7):1667-78 .
65. Li Z, Jiang HY, Zhou LY, Deng L, Lin YX, Peng XJ, Yan HW, Cheng BJ. Molecular evolution of the HD-ZIP I gene family in legume genomes. *Gene*. 2014;533(1):218-28.
66. Emanuelsson O, Nielsen H, Brunak S, Heijne GV. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*. 2000;300(4):1005-16.
67. Wu M, Li Y, Chen D, Liu H, Zhu D, Xiang Y. Genome-wide identification and expression analysis of the *IQD* gene family in moso bamboo (*Phyllostachys edulis*). *Sci Rep*. 2016;6(1):1-14.
68. Chen CJ, Xia R, Chen H, He Y. TBtools, a Toolkit for biologists integrating various HTS-data handling tools with a user-friendly interface. *BioRxiv*. 2018;289660.
69. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S. 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.
70. Bari R, Jones JDG. Role of plant hormones in plant defence responses. *Plant Mol Biol*. 2009;69(4):473-88

## Tables

Table 1 Physical and chemical characteristics of the 36 *Hsf* genes identified in *Cucurbita moschata*.

Gene ID	Gene name	Cmo_Chr <sup>*1</sup>	Start <sup>*2</sup>	End <sup>*3</sup>	ORF length (bp)	AA <sup>*4</sup>	pI <sup>*5</sup>	Mw <sup>*6</sup> (Da)	Loc <sup>*7</sup>
CmoCh01G018910.1	<i>CmHsf01</i>	01	13630401	13636203	1701	565	7.32	63908.05	Nucleus.
CmoCh02G000520.1	<i>CmHsf02</i>	02	279098	280430	945	313	6.23	35866.58	Nucleus.
CmoCh02G015130.1	<i>CmHsf03</i>	02	8829467	8831346	1017	337	4.79	37042.37	Nucleus.
CmoCh03G000560.1	<i>CmHsf04</i>	03	917233	919195	723	239	9.35	27525.06	Nucleus.
CmoCh03G009950.1	<i>CmHsf05</i>	03	7477236	7479691	900	298	5.6	33388.41	Nucleus.
CmoCh03G012560.1	<i>CmHsf06</i>	03	9632303	9635635	1392	462	7.55	52807.71	Nucleus.
CmoCh04G000850.1	<i>CmHsf07</i>	04	461682	465859	1218	404	4.88	46844.9	Nucleus.
CmoCh04G011130.1	<i>CmHsf08</i>	04	5675420	5678524	1134	376	4.95	43681.38	Nucleus.
CmoCh05G000960.1	<i>CmHsf09</i>	05	393383	395093	1110	368	4.93	41839.89	Nucleus.
CmoCh05G001750.1	<i>CmHsf10</i>	05	759147	761562	1362	452	7.64	50399.85	Nucleus.
CmoCh05G013450.1	<i>CmHsf11</i>	05	10456658	10458207	993	329	6.12	37483.5	Nucleus.
CmoCh05G014000.1	<i>CmHsf12</i>	05	10787694	10799787	3714	1236	6.8	139325.5	Cell membrane. Cytoplasm. Nucleus.
CmoCh06G004420.1	<i>CmHsf13</i>	06	2118798	2130108	4380	1458	5.55	161555.42	Nucleus.
CmoCh06G006450.1	<i>CmHsf14</i>	06	3242367	3246508	1566	520	5.12	57039.44	Nucleus.
CmoCh06G009230.1	<i>CmHsf15</i>	06	6678383	6679150	687	227	8.85	26518.21	Nucleus.
CmoCh06G012330.1	<i>CmHsf16</i>	06	9329887	9333367	1416	470	6.48	52376.05	Nucleus.
CmoCh06G013840.1	<i>CmHsf17</i>	06	10166157	10173534	1650	548	5.35	64159.97	Cell membrane. Cytoplasm. Nucleus.
CmoCh07G001570.1	<i>CmHsf18</i>	07	853089	854975	1227	407	5.45	46937.23	Nucleus.
CmoCh07G002420.1	<i>CmHsf19</i>	07	1191784	1192862	579	191	8.38	22514.5	Nucleus.
CmoCh07G007220.1	<i>CmHsf20</i>	07	3258238	3259253	873	289	6.36	32815.78	Nucleus.
CmoCh09G002330.1	<i>CmHsf21</i>	09	1070417	1071523	993	329	8.56	37629.07	Nucleus.
CmoCh10G006520.1	<i>CmHsf22</i>	10	2987379	2988593	855	283	6.07	32230.65	Nucleus.
CmoCh10G009220.1	<i>CmHsf23</i>	10	4574443	4576160	750	248	8.75	28527.16	Nucleus.
CmoCh13G006110.1	<i>CmHsf24</i>	11	6682088	6683686	1239	411	5.21	46658.15	Nucleus.
CmoCh11G009050.1	<i>CmHsf25</i>	11	4658284	4659725	708	234	7.96	27047.37	Nucleus.
CmoCh12G005810.1	<i>CmHsf26</i>	12	3595429	3596964	1074	356	4.88	40558.39	Nucleus.
CmoCh11G006110.1	<i>CmHsf27</i>	13	2932186	2933416	879	291	5.61	33180.33	Nucleus.
CmoCh14G002670.1	<i>CmHsf28</i>	14	1203588	1210628	2073	689	5.78	76886.63	Nucleus.
CmoCh14G017830.1	<i>CmHsf29</i>	14	13739183	13747848	2604	866	5.34	97900.69	Nucleus.
CmoCh14G019680.1	<i>CmHsf30</i>	14	14515610	14518497	1350	448	6.53	50396.1	Nucleus.
CmoCh15G012680.1	<i>CmHsf31</i>	15	8690633	8692333	1059	351	4.64	39145.25	Nucleus.
CmoCh16G001410.1	<i>CmHsf32</i>	16	644769	646828	543	179	8.42	20564.2	Nucleus.
CmoCh16G012250.1	<i>CmHsf33</i>	16	8775979	8782264	1572	522	4.9	57215.79	Nucleus.
CmoCh17G011810.1	<i>CmHsf34</i>	17	9496232	9498290	1140	378	4.9	43615.67	Nucleus.
CmoCh18G012590.1	<i>CmHsf35</i>	18	12324683	12328013	1059	351	5.77	39472.8	Nucleus.
CmoCh19G000190.1	<i>CmHsf36</i>	19	124488	127854	1164	386	5.78	44382.64	Nucleus.

Note: Information on including their chromosomal distribution, their start and the end positions on the chromosomes, nucleic acid sequence and amino acid sequence were extracted from *Cucurbit* genomics database, and all the data in the table is predicted or theoretical.

\*1 Cmo\_Chr, The name of the CmHsf chromosome corresponding to the gene;

\*2 Start, Predicted starting position of mRNA;

\*3 End, Predicted termination position of mRNA;

\*4 AA, Amino acid number in CmHsf protein sequences;

\*5 *pI*, Theoretical Isoelectric point;

\*6 MW, Molecular weight (Mw) predicted by ExpASY (<http://web.expasy.org/tools/>);

\*7 Loc, Subcellular location of the CmHsf proteins predicted by Plant-mPLoc

Table 2 Functional domain analysis of the 36 Hsf proteins identified in *Cucurbita moschata*.

Subfamily Name	Gene ID	Gene Name	Group	DBD	HR- A/B	NLS	NES	AHA	RD
Subfamily I	CmoCh06G013840.1	CmHsf17	A	243-336	359-407	(428)QKDKHKELEEAINRKRRII	nd	DDGFWENLL	nd
	CmoCh14G017830.1	CmHsf29	A	42-135	162-205	(225)GFRKVDPKWEFAHESFLRGQRHLLKLIRR	IEELCLSL	SDFWQKLIEL	nd
	CmoCh17G011810.1	CmHsf34	A	43-136	163-206	(243)ITRKRRIPIQ	TELEALALEMQGL	EGFWEEFSE	nd
	CmoCh19G000190.1	CmHsf36	A	79-172	198-242	(278)ATKRRWPID	LEALAMEM	EGFWEEFFSE	nd
	CmoCh05G014000.1	CmHsf12	C	458-551	573-599	(1191)QRRPPVGPEDPKRSASGRHTGYVKNYD	nd	nd	nd
	CmoCh12G005810.1	CmHsf26	C	39-132	153-205	(236)RKRRLTASPSLENLQDETILAAVKQEQL	nd	nd	nd
	CmoCh05G000960.1	CmHsf9	C	41-134	155-208	(236)EIGKRRLTSS	nd	nd	nd
	CmoCh04G011130.1	CmHsf8	C	12-105	137-184	(323)IDHEKRSVDNEDDELDMETIDTRTHEEKSQD	nd	nd	nd
	CmoCh04G000850.1	CmHsf7	C	12-105	137-184	(369)RLDESYIEKSNTVNLMEMLMASDQEILYETPAKMQG	nd	nd	nd
	CmoCh05G013450.1	CmHsf11	A	53-146	162-195	(221)RRVRRRVTRMRPPSPVKFKA	VKREDDGELALEISKLKQEQI	SNFWDDLLVQ	nd
	CmoCh13G006110.1	CmHsf24	A	11-104	124-172	(187)RFLHKPGLRLDLLPQLETSRKRRLP	LKRDKQEQLLELRKHEQ	DVFWQQLFTE	nd
	CmoCh10G006520.1	CmHsf22	A	9-102	122-155	(189)PDKKRRFMTS	nd	EGFWEEFSE	nd
	CmoCh16G012250.1	CmHsf33	A	32-125	154-205	(236)EANKKRRLKQD	MKVLLDEKLCCLDNH	SNFWDDLLVQ	nd
	CmoCh06G006450.1	CmHsf14	A	32-125	154-205	(236)EANKKRRLKQD	LQDFELLIKQM	SNFWNDLLVH	nd
	CmoCh06G004420.1	CmHsf13	A	968-1124	1144-1195	(1225)PRMKRKFVKQ	LQLALALRL	LSPFWDLGSL	nd
	CmoCh14G002670.1	CmHsf28	A	239-387	407-438	(474)FLLKRKKEPKDIDSERIKRKFVK	nd	DVFWEQFLTE	nd
	CmoCh02G000520.1	CmHsf2	A	11-104	117-155	(173)RMGNQKQLIAIMAAELQKDQSRKRRK	LSELERQELELKI	DVFWEQFLTE	nd
	CmoCh11G006110.1	CmHsf27	A	11-104	117-155	nd	LEEELEGM	DVFWEQFLTE	nd
	CmoCh07G001570.1	CmHsf18	A	11-104	123-179	(205)HERKRRLATV	LQLQMQL	DVFWEQFLTE	nd
	CmoCh14G019680.1	CmHsf30	A	9-102	119-174	(201)FNKKRRLPS	LQLQELTM	DVFWEQFLTE	nd
CmoCh06G012330.1	CmHsf16	A	14-107	124-179	(206)FNKKRRLPS	IQLQDLTV	DVFWEQFLTE	nd	
Subfamily II	CmoCh03G012560.1	CmHsf6	C	276-391	320-352	(432)RRQKLELQAQIAQFKALHIRLLDCVGRRIEK	nd	nd	nd
	CmoCh07G002420.1	CmHsf19	C	8-120	147-178	(182)KTRNPAPFLSKTY	nd	nd	nd
Subfamily III	CmoCh09G002330.1	CmHsf21	B	21-114	117-155	(287)IHSKKRLHPEYASNNGKENNNKARFV	LEKDDLGLNL	nd	KLFG <sup>3</sup>
	CmoCh01G018910.1	CmHsf1	B	120-213/21-114	117-156	(374)GSSKSFVTIVEEPKTKLFGVSLQSKRVHPE	VLEKDDLGLNL	nd	KLFG <sup>3</sup>
	CmoCh05G001750.1	CmHsf10	B	21-114	117-157	(337)KKRQHPDITNYVSTSSNVSDTNKNSRGS	LLLLFKPRL	nd	KLFG <sup>3</sup>
	CmoCh10G009220.1	CmHsf23	B	21-114	117-158	(273)RGKKRMHHE	KQLLLAI	nd	KLFG <sup>3</sup>
	CmoCh11G009050.1	CmHsf25	B	11-104	124-172	(221)RGKKRGASDEE	nd	nd	KLFG <sup>3</sup>
	CmoCh03G000560.1	CmHsf4	B	32-125	150-187	(197)GSRKEDEDERPKLFGVRLEVEGERRRKTQR	nd	nd	KLFG <sup>3</sup>
	CmoCh06G009230.1	CmHsf15	B	19-112	144-180	(196)EMMVMKPNLKLFGVKLEVEGEDEMVRQSKR	LKLFGVKLEV	nd	KLFG <sup>3</sup>
	CmoCh07G007220.1	CmHsf20	B	6-99	144-183	(244)EKNNDNKNTKREEEKVEVCNEPEAKVMKT	nd	nd	KLFG <sup>3</sup>
	CmoCh03G009950.1	CmHsf5	B	6-99	150-188	(258)EKKKMKRVREEKIGCSNAPHAKAMK	nd	nd	KLFG <sup>3</sup>
	CmoCh02G015130.1	CmHsf3	B	21-114	177-207	(254)FLTQTYQLVDDPDVDDLISWNEDGSTFIVW	nd	nd	KLFG <sup>3</sup>
	CmoCh15G012680.1	CmHsf31	B	21-114	176-206	(279)IGVKRRREEE	nd	nd	KLFG <sup>3</sup>
	CmoCh16G001410.1	CmHsf32	B	19-112	134-173	nd	LASAKSLDL	nd	KLFG <sup>3</sup>
	CmoCh18G012590.1	CmHsf35	B	85-178	226-260	(269)ENQLKSSCKVRESVLASAKSLDLFPLKRRSEE	LASAKSLDL	nd	KLFG <sup>3</sup>

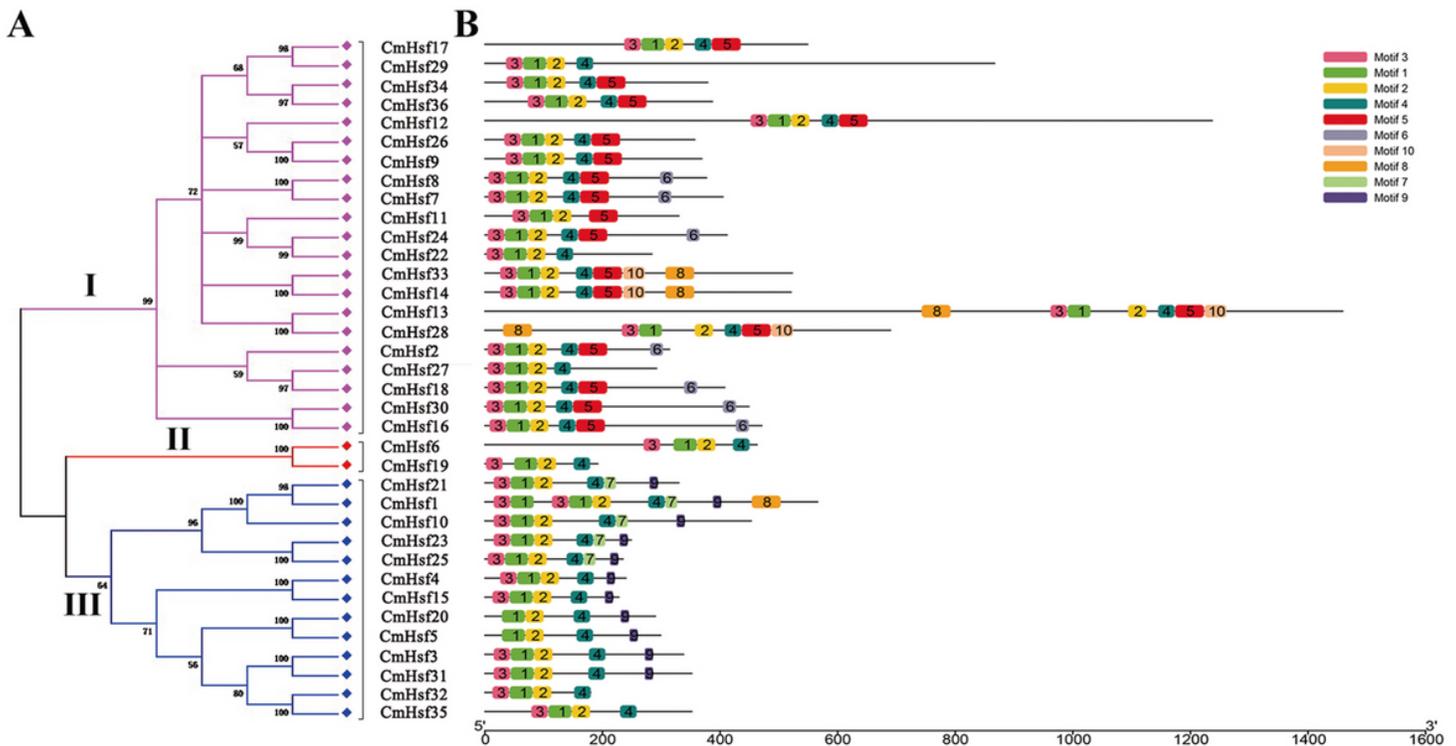
Note: The amino acid sequences of Hsf in the table came from *Cucurbit* genomics database. DBD, DNA-binding domain; HR-A/B, oligomerization domain; NLS, nuclear localization signal; NES, nuclear export signal; AHA, transcriptional activation domain; RD, repressor domain; nd, no motifs detectable by sequence similarity search. For the NLS column, the numbers in parenthesis are the start site of the functional domain.

Table 3 KaKs calculation and estimated divergence time for the duplicated *CmHsf* gene pairs.

Duplicated <i>CmHsf</i> gene pairs	Identity (%)	E-value	Gene alignment coverage	<i>Ka</i>	<i>Ks</i>	<i>Ka/Ks</i>	Divergence time (MYA)
CmHsf12-CmHsf10	95.12	1.00E-12	0.975	0.832	1.972	0.422	65.742
CmHsf26-CmHsf9	95.12	1.00E-12	0.910	0.126	0.462	0.273	15.416
CmHsf22-CmHsf27	85.09	0	0.893	0.145	0.673	0.215	22.432
CmHsf13-CmHsf28	85.09	0	0.946	0.232	0.436	0.531	14.535
CmHsf30-CmHsf16	86.13	0	0.931	0.074	0.305	0.242	10.168
CmHsf6-CmHsf19	86.13	0	0.863	0.080	0.336	0.238	11.204
CmHsf21-CmHsf1	87.83	0	0.812	0.066	0.542	0.121	18.083
CmHsf3-CmHsf31	87.83	0	0.814	0.106	0.431	0.246	14.354
CmHsf32-CmHsf35	87.97	0	0.981	0.181	0.542	0.335	18.053

Note: We used the KaKs calculator to calculate the *Ka/Ks*. *Ks*, synonymous substitutions; *Ka*, nonsynonymous substitutions.

## Figures



**Figure 1**  
 Classification and conserved motifs of 36 *CmHsfs*. **A**. The unrooted phylogenetic tree of 36 *CmHsfs* was constructed using the Neighbor-joining (NJ) method with 1,000 bootstrap replicates, and a 60% cut-off value was used for the condensed tree. Three different subfamilies (I-III) were highlighted with different colored branch lines. **B**. Schematic representation of conserved motifs in 36 *CmHsfs*. Each motif was represented by a numbered colored box on the right. The same number in different proteins referred to the same motif. Motif 1, motif 2 and motif 3 together formed the DBD, and motif 4 formed the HR-A/B domain. The function of other motifs was unknown.

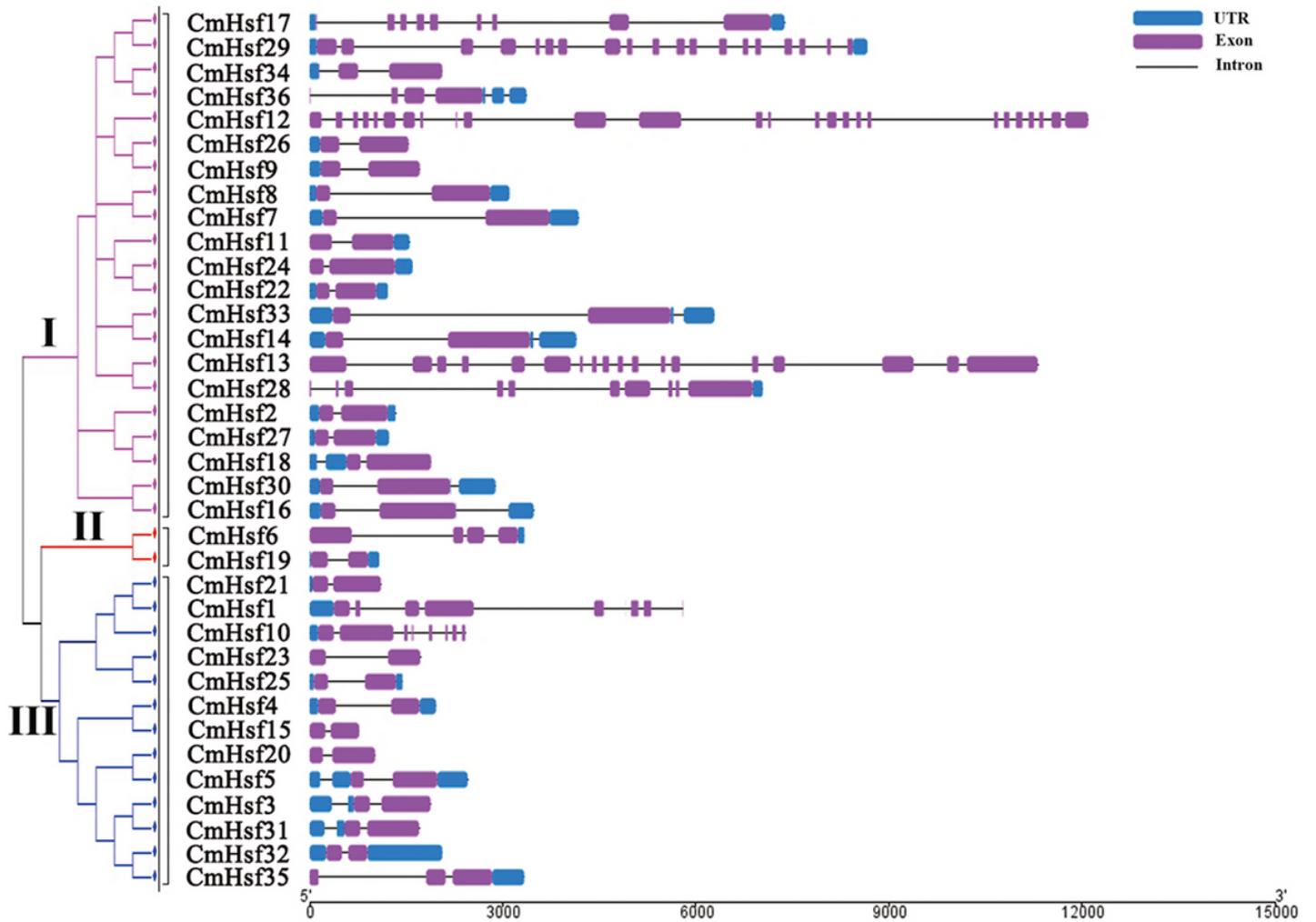


Figure 2 Exon-intron organization of 36 CmHsfs constructed by GSDS (Gene structure display server). The exons and introns were represented by pink boxes and grey lines, respectively. Untranslated regions (UTRs) were indicated by blue boxes. The sizes of the exons and introns can be estimated using the scale at the bottom.

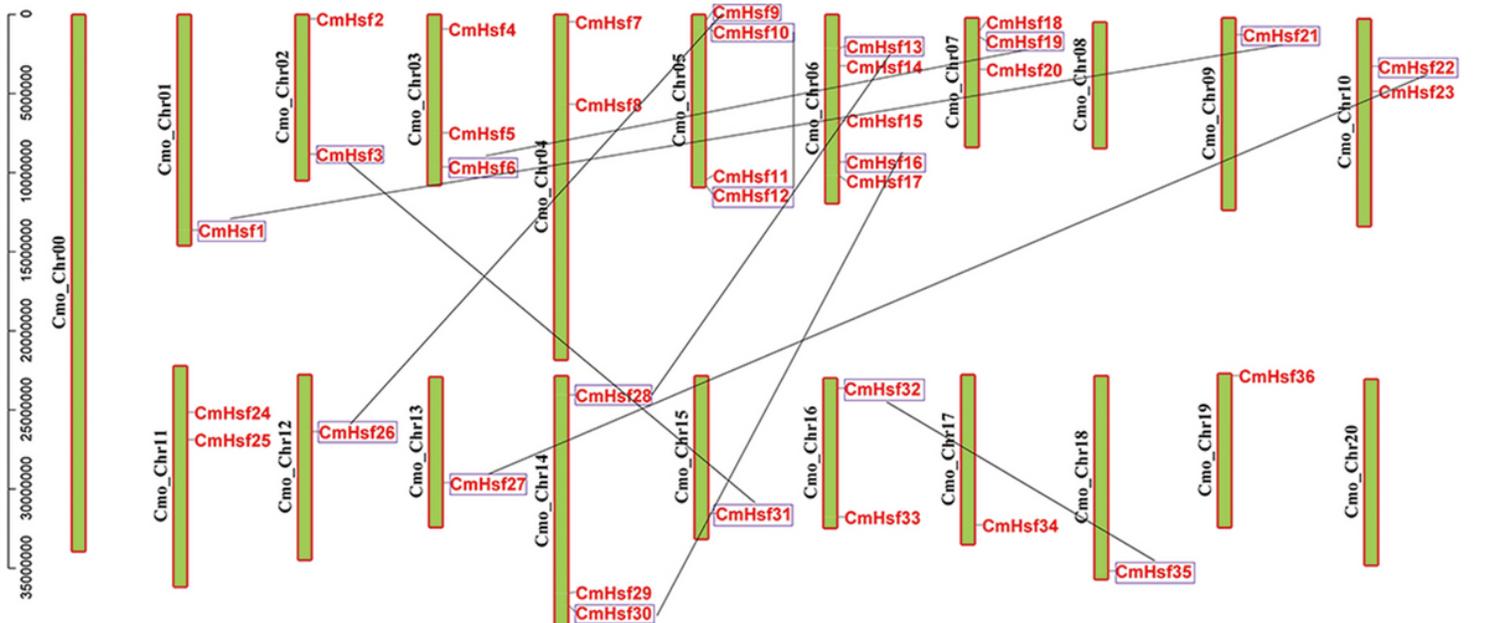
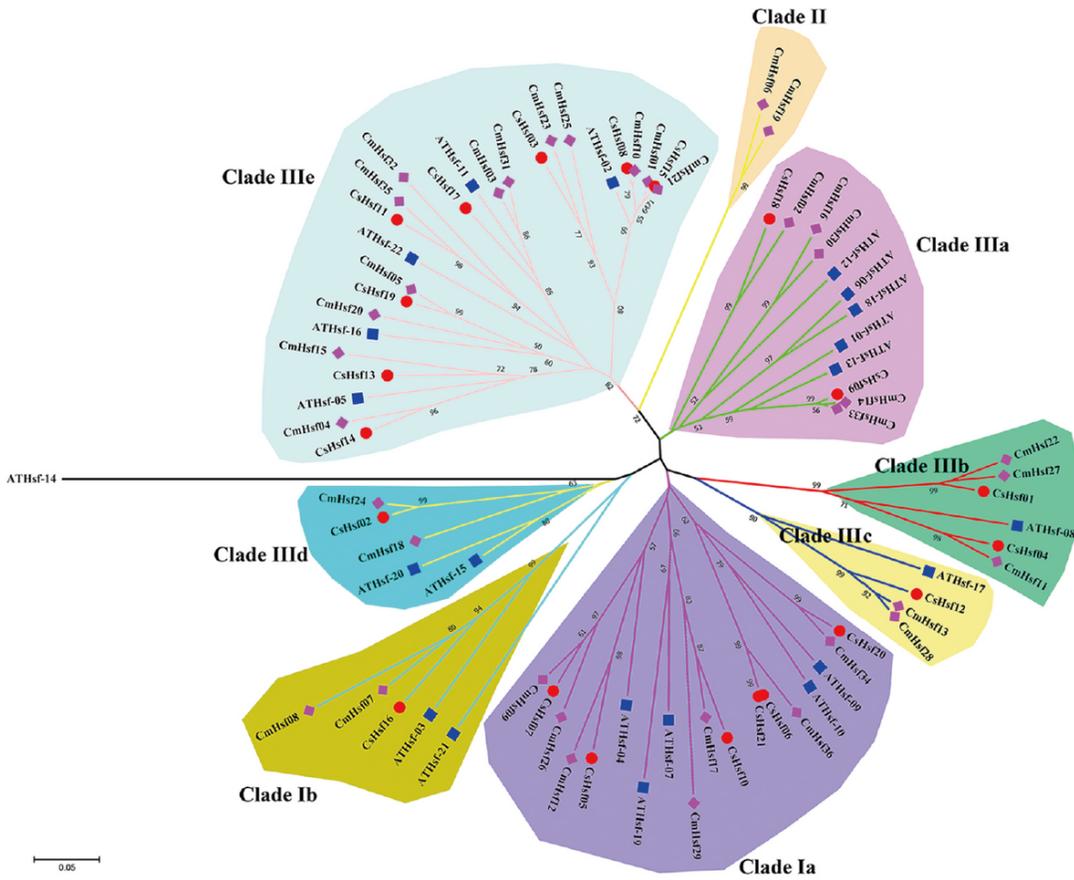
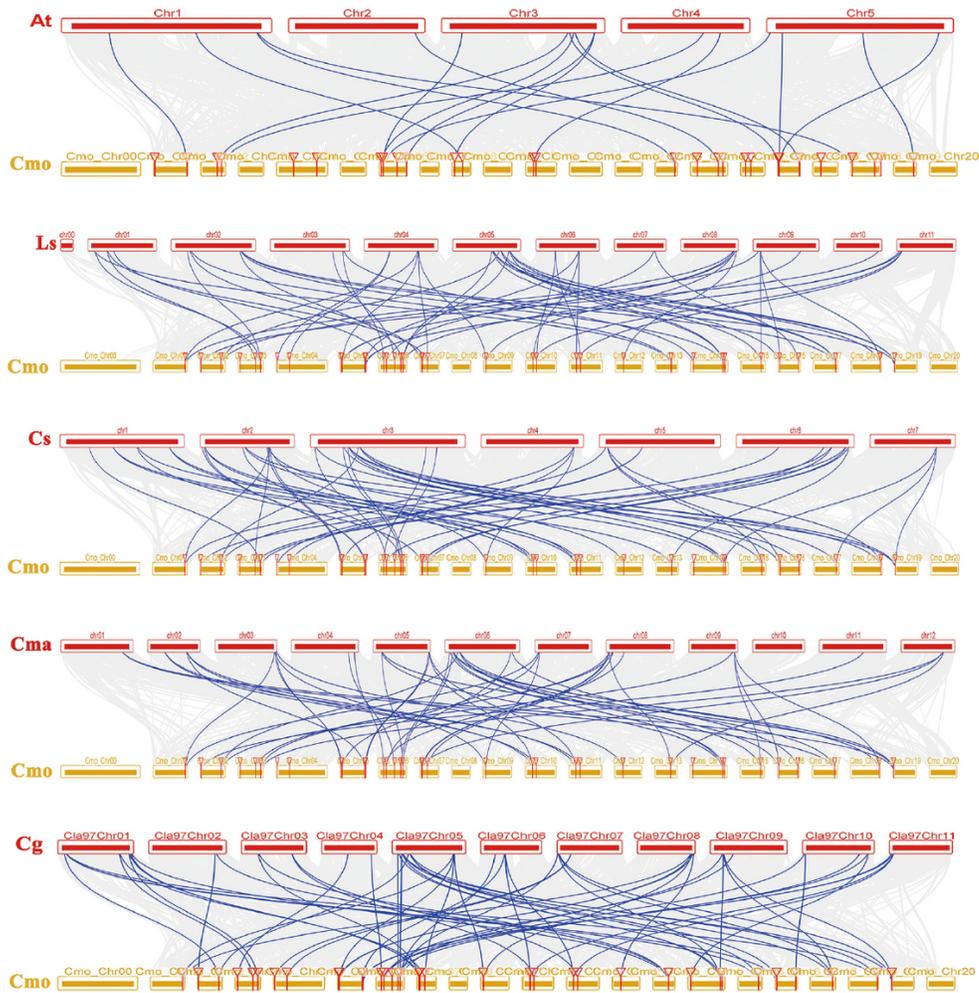


Figure 3

Chromosomal distribution and duplication events of Hsf genes in *C. moschata*. The chromosomal locations of the CmHsf genes were mapped with visualization tools. The duplicated CmHsf genes were shown in blue boxes and black lines.



**Figure 4**  
 Phylogenetic trees of the Hsf gene family in *C. moschata*, *C. sativa* and *A. thaliana*. The 9 clades (Clade Ia-b, Clade II and Clade IIIa-e) were displayed with different background colors. The phylogenetic tree was constructed with MEGA 5.0 software using the Neighbor-joining (NJ) method with 1000 bootstrap replicates. Cm, *C. moschata*; Cs, *C. sativa*; At, *A. thaliana*.



**Figure 5**  
 Synteny analysis of the Hsf genes between *C. moschata* and five other species. The synteny relationship maps were constructed using the Advanced Circos program in TBtools. At, *A. thaliana*; Ls, *L. siceraria*; Cs, *C. sativus*; Cma, *C. maxima*; Cg, *C. lanatus*; Cmo, *C. moschata*. The gray lines in the background indicated the collinear blocks in the genome of *C. moschata* and other plants, while blue lines in the background highlighted syntenic Hsf gene pairs. All the data for the various species was extracted from Cucurbit genomics database.

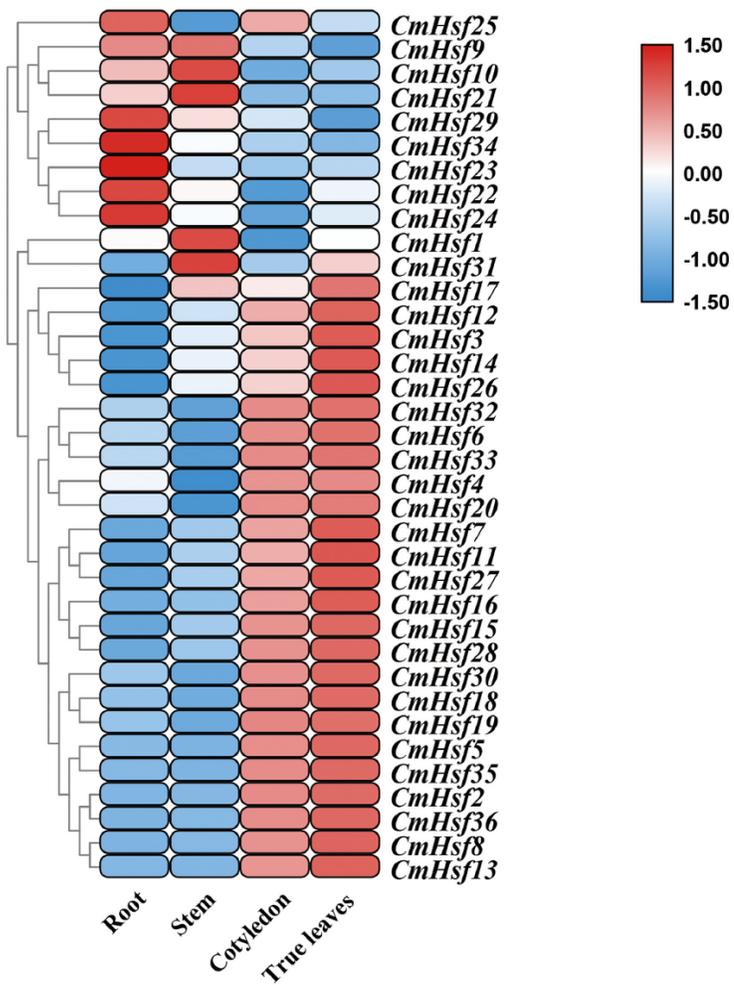
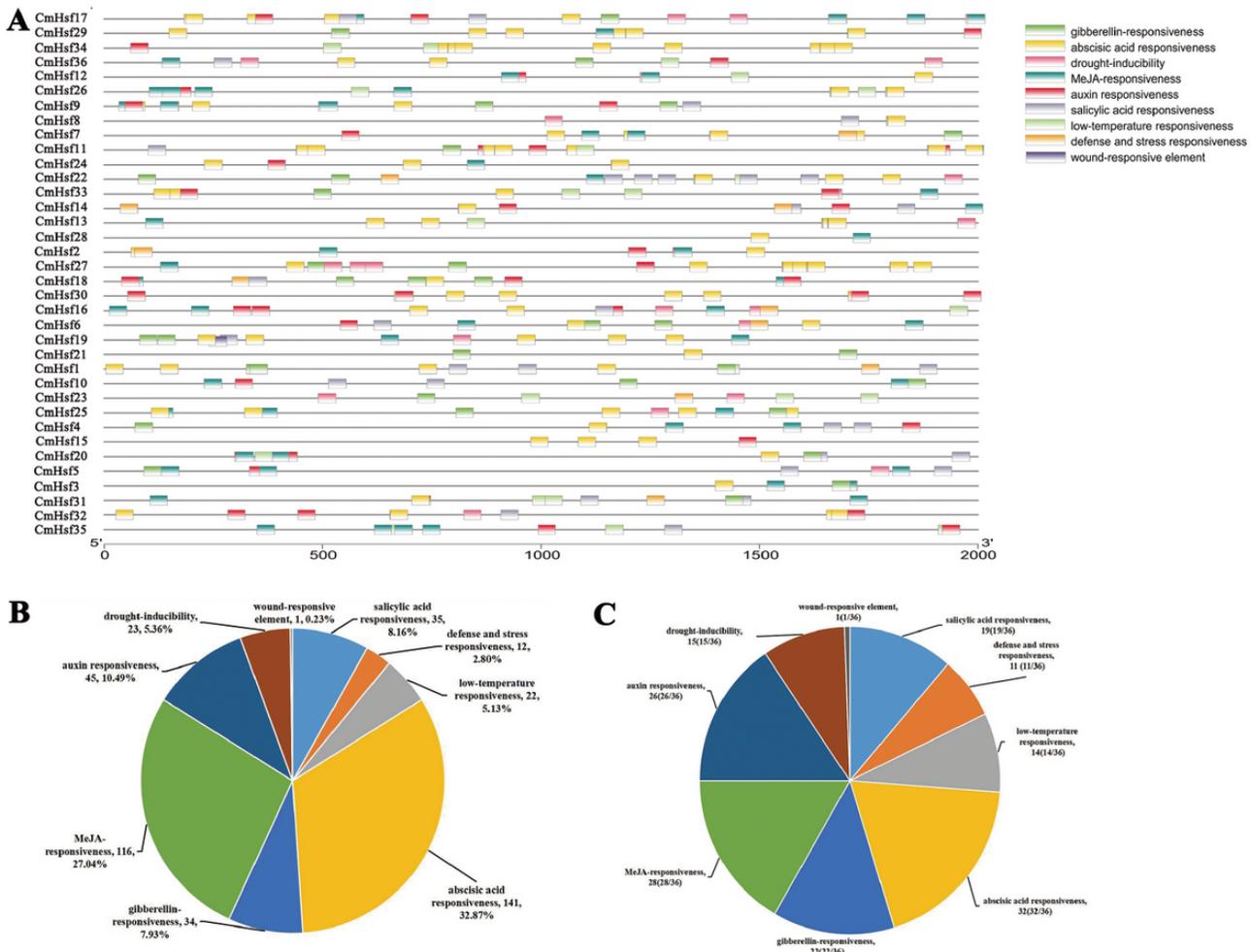


Figure 6

Heat map and hierarchical clustering of 36 CmHsf genes in the roots, stems, cotyledons and true leaves. Quantitative real-time PCR was performed in three biological replicates and three technical replicates, and the heat map and hierarchical clustering were constructed by TBtools. The results were calculated via the  $2^{-\Delta\Delta Ct}$  method, and the reference gene ( $\beta$ -Actin) was used to correct the expression level of target genes. All data were standardized by  $\text{Log}_{10}(2^{-\Delta\Delta Ct})$ . The bar on the right of the heat map represented the data that has been converted to  $\text{Log}_{10}(2^{-\Delta\Delta Ct})$ .



**Figure 7**

Distribution of cis-acting elements in 36 CmHsfs and the proportions of corresponding genes in 9 stress response elements. A. The cis-acting elements of 36 heat shock transcription factors in *C. moschata*. They were predicted by PlantCare program (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) and visualized by Simple BioSequence Viewer in TBtools. The squares on the right represented cis-acting elements that respond to a total of 9 stresses. Different colors indicated cis-acting elements that participate in different stresses. The coordinates at the bottom of the figure indicated the length of the gene promoter. The promoter sequence was defined as 2 kb before the start codon. B. The distribution of 429 cis-acting elements related to 9 abiotic stresses. C. The proportion of 36 CmHsfs related to 9 abiotic stresses.

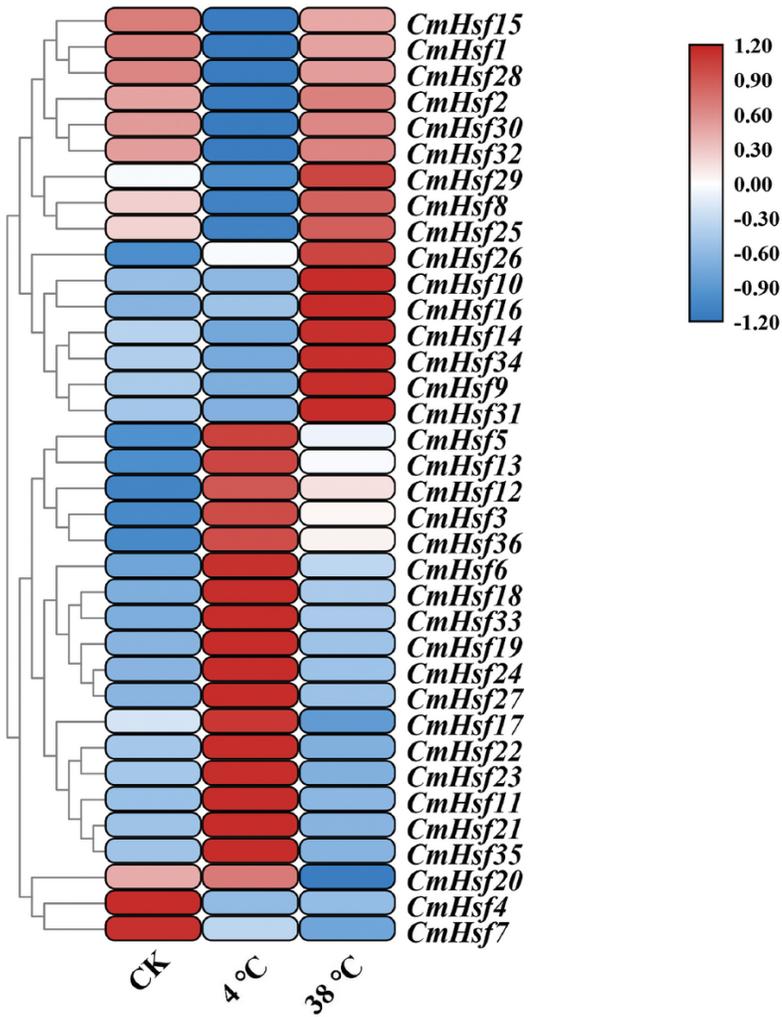
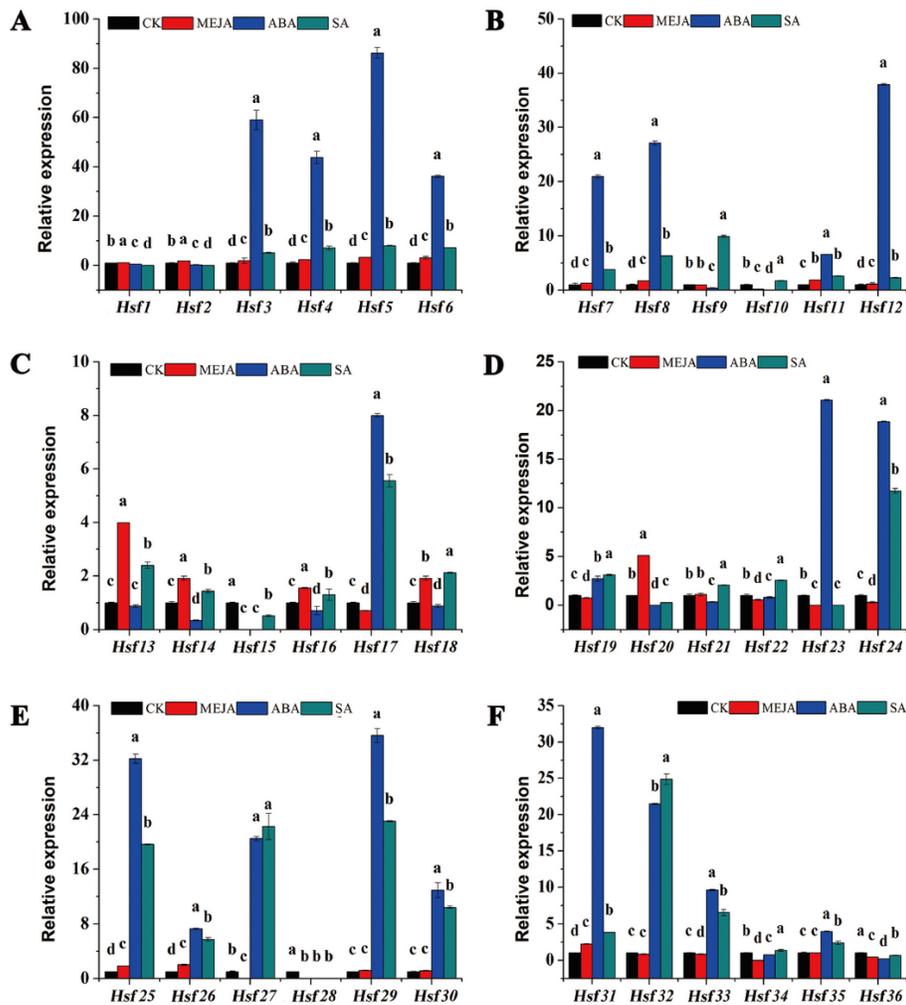


Figure 8

Heat map and hierarchical clustering of 36 CmHsf genes in true leaves under cold stress and heat stress. Quantitative real-time PCR and hierarchical clustering were performed in three biological replicates and three technical replicates, and the heat map was constructed by TBtools. The results were calculated via the  $2^{-\Delta\Delta Ct}$  method, and the reference gene ( $\beta$ -Actin) was used to correct the expression level of target genes. All data were standardized by  $\text{Log}_{10}(2^{-\Delta\Delta Ct})$ . The bar on the right of the heat map represented the data that has been converted to  $\text{Log}_{10}(2^{-\Delta\Delta Ct})$ .



**Figure 9**

Expression profiles of 36 CmHsf genes in true leaves under MeJA, ABA and SA treatments. The data represented the expression levels of CmHsf genes at 10 h after the MeJA, ABA and SA treatments. CK referred to untreated plants (control plants) under normal conditions. The results were calculated via the  $2^{-\Delta\Delta Ct}$  method, and the reference gene ( $\beta$ -Actin) was used to correct the expression level of target genes. The expression level of CK was set as 1. The data were presented as the means of three biological replicates and three technical replicates, and the error bars represented the standard deviations of the means. According to Welch's t-test, different letters above the bars indicated significant differences ( $p < 0.05$ ) between different treatments.

## Supplementary Files

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

- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [Fig.S1.pdf](#)
- [Fig.S2.pdf](#)