Genome-wide Characterization and Functional Analysis of the Heat Shock Transcription Factor Family in Pumpkin

Changwei Shen
Henan Institute of Science and Technology
Jingping Yuan (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: July 14th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-38574/v1

Version of Record: A version of this preprint was published 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 pumpkin to abiotic stress, we analysed the genome of pumpkin.

Results: In this research, a total of 36 *Cucurbita 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 *Cucurbita 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 *CmHsf* genes 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 pumpkin varieties.

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 among of the most important transcription regulators. 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 [15]. With the sequencing of the rice and *Arabidopsis* genomes, *Hsf* genes have also been identified in rice and *Arabidopsis* [11, 12]. subsequently, researchers identified 25, 31, 21, 26, 35, 29, 30, 27, 19 and 35 *Hsf* genes in the *Populus trichocarpa* [13], maize [14], cucumber [13], *Glycine max* [16], Chinese cabbage [17], *Pyrus bretschneideri* [18], potato[19], grape [20] and *Brassica oleracea* [21] 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). 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 [22]. The NLS is rich in Arg (R) and Lys (K) residues, while the NES is rich in Leu (L). These residues together determine the dynamic distribution of *Hsfs* between the cytoplasm and the nucleus [23–25]. 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 and C proteins exhibit an additional 21 or 7 residues, respectively. In addition, the transcriptional 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, 25]. The repressor domain (RD) is a peptide containing conserved amino acids (LFGV) at the C-terminus and mainly exists in group B [27].

Hsfs can specifically regulate the transcription of heat shock protein (Hsp) genes by specifically binding to the heat shock element (HSE) in the promoter of an Hsp gene, and the Hsps in turn protect cells from stress and participate in protein folding [28, 29]. 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 [30]. At 37 ℃, *Arabidopsis HsfA2*-mutant plants are more sensitive to heat stress than wild-type plants, which can be reversed by introducing the *HsfA2* gene [31]. The *OsHsfA4d* mutant shows a phenotype of spontaneous necrotic damage under high-temperature stress [12]. The expression of *OsHsfA2e* enhances high temperature and salt tolerance in *Arabidopsis* plants [32]. In addition to heat stress, *Hsfs* are involved in plant growth and other biotic and abiotic stress responses. It is found the *HsfA9* is involved in embryo development and seed maturation in *Arabidopsis* and sunflower [33]. Four *Hsf* genes (*HsfA1e*, *HsfA3*, *HsfA4a*, *HsfB2a* and *HsfC1*) in *Arabidopsis* are strongly induced by salt, cold and osmotic stress [34–36]. The *HsfA2* in *Arabidopsis* is involved in the response to oxidative stress [37]. The *HsfA4a* in *Arabidopsis* can be used as an H_2O_2 sensor [38, 34]. The *OsHsfA4a in* rice is associated with cadmium tolerance [39]. To date, there have been no reports of the cloning and functional analysis of pumpkin *Hsfs*.

Pumpkin is rich in a variety of amino acids, vitamins, polysaccharides, pectin, and minerals and contains trigonelline, carotenoids and other biologically active substances and nutrients. Chinese pumpkin (*Cucurbita moschata*) belongs to the *Cucurbitaceae* and is an important vegetable species that is widely grown around the world. During growth and development, unfavourable stress often causes great harm to the growth of pumpkin, resulting in a decline in pumpkin yield and quality. Therefore, research on pumpkin resistance-related genes is increasingly important for pumpkin breeding and production. Because the *Cucurbita moschata* (*Rifu*) genome has been published [40], the Hsf family in pumpkin 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 pumpkin *Hsf* genes. Furthermore, we analyse the digital expression profiles of these 36 genes in response to numerous stresses. This study emphasizes the function of the *Hsf* genes in various stress conditions and improves our understanding of the effects of polyploidization events on the evolution of the Hsf family.

Results

Identification of *CmHsf* genes in pumpkin and physical and chemical characteristics of *CmHsfs*

A total of 36 *CmHsf* genes were identified after the removal of redundant sequences (Table 1), and they were designated *CmHsf1* to *CmHsf36* according to the starting positions of the 36 genes on the chromosomes (from Cmo_Chr00 to Cmo_Chr20, from top to bottom). The physicochemical parameters of each *CmHsf* were generated, and the open reading frames (ORFs) ranged from 543 bp (*CmHsf32*) to 4380 bp (*CmHsf13*), with predicted proteins of 179–1458 amino acids. Furthermore, the molecular weights 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 factor genes (*CmHsf12* and *CmHsf17*) were expressed in the cell membrane, cytoplasm and nucleus, while the remaining *CmHsf* genes were expressed in the nucleus.

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

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

*1 Cmo_Chr,The name of the chromosome corresponding to the gene;

*2 Start, Starting position of mRNA;

*3 End Termination position of mRNA;

*4 Isoelectric points.

Gene ID	Gene name	Cmo_Chr *1	Start*2	End*3	ORF length (bp)	Number of amino acid	*pI*4	Molecular weight Mw/Da	Subcellular localization
CmoCh18G012590.1	*CmHsf35*	18	12324683	12328013	1059	351	5.77	39472.8	Nucleus.
CmoCh19G000190.1	*CmHsf36*	19	124488	127854	1164	386	5.78	44382.64	Nucleus.

*1 Cmo_Chr,The name of the chromosome corresponding to the gene;

*2 Start, Starting position of mRNA;

*3 End Termination position of mRNA;

*4 Isoelectric points.

Classification and conserved domain analysis of 36 *CmHsfs* in pumpkin

To identify the phylogenetic relationships of the 36 *CmHsfs*, an unrooted phylogenetic tree was produced. These pumpkin *CmHsfs* can be divided into three subfamilies (subfamily I, subfamily II and subfamily III; Fig. 1A) according to sequence similarity. 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 (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 genes 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; 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. Motif 3, motif 1 and motif 2 together formed the DBD, and motif 4 formed the HR-A/B domain. In addition, we found that motif 5 existed only in subfamily I, while motif 9 appeared only in subfamily III (Fig. 1B). It is clear that the CmHsfs from the same clade usually present conserved domains or similar motif compositions, suggesting functional similarities among these proteins.

Table 2
Functional domain analysis of the 36 *Hsf* genes identified in pumpkin.

Gene ID	Gene Name	Group	DBD	HR-A/B	NLS	NES	AH
CmoCh06G013840.1	CmHsf17	A	243–336	359–407	(428) QKDKHKELEEAINRKRRRHI	nd	DD
CmoCh14G017830.1	CmHsf29	A	42–135	162–205	(225) GFRKVDPDKWEFAHESFLRGQRHLLKLIRR	IEELCLSL	SD
CmoCh17G011810.1	CmHsf34	A	43–136	163–206	(243) ITRKRRRPIQ	TELEALALEMQGL	EG
CmoCh19G000190.1	CmHsf36	A	79–172	198–242	(278) ATKKRRWPID	LEALAMEM	EG
CmoCh05G014000.1	CmHsf12	C	458–551	573–599	(1191) QRRPPVGPEDPKRSASGRHTGYVKNYD	nd	nd
CmoCh12G005810.1	CmHsf26	C	39–132	153–205	(236) RKRRLTASPSLENLQDETILAAVKQEQLE	nd	nd
CmoCh05G000960.1	CmHsf9	C	41–134	155–208	(236) EIGRKRRLTSS	nd	nd
CmoCh04G011130.1	CmHsf8	C	12–105	137–184	(323)IDHEKRSVDNEDDELDMETIDTRTHEEKSQD	nd	nd
CmoCh04G000850.1	CmHsf7	C	12–105	137–184	(369) RLDESYIEKSNTVNLMELMASDQEILYETPAKMQG	nd	nd
CmoCh05G013450.1	CmHsf11	A	53–146	162–195	(221) RRVRRRVTMRPPPSPVKFVKA	VKREDDGELALEISKLKQEQI	SN
CmoCh13G006110.1	CmHsf24	A	11–104	124–172	(187) RFLHKPGLRLDLLPQLETSDRKRRLP	LKRDKEQLLLELRKHEQ	DV
CmoCh10G006520.1	CmHsf22	A	9-102	122–155	(189) PDKKRRFMTS	nd	EG
CmoCh16G012250.1	CmHsf33	A	32–125	154–205	(236) EANKKRRLKQD	MKVLLDEKLCLDNH	SN
CmoCh06G006450.1	CmHsf14	A	32–125	154–205	(236) EANKKRRLKQD	LQDFELLIKQM	SN
CmoCh06G004420.1	CmHsf13	A	968–1124	1144–1195	(1225) PRMKRKFVKQ	LQLALALRL	LS
CmoCh14G002670.1	CmHsf28	A	239–387	407–438	(474) FLLKRKKEPKDIDSERIKRKFVK	nd	DV
CmoCh02G000520.1	CmHsf2	A	11–104	117–155	(173) RMGNQQKQLIAIMAAELQKDQSRKRRK	LSELERQELELKI	DV
CmoCh11G006110.1	CmHsf27	A	11–104	117–155	nd	LEEELEGM	DV
CmoCh07G001570.1	CmHsf18	A	11–104	123–179	(205) HERKRRLATV	LQLQMQL	DV
CmoCh14G019680.1	CmHsf30	A	9-102	119–174	(201) FNKKRRLPS	LQLQELTM	DV
CmoCh06G012330.1	CmHsf16	A	14–107	124–179	(206) FNKKRRLPS	IQLQDLTV	DV
CmoCh03G012560.1	CmHsf6	C	276–391	320–352	(432) RRQKLELQAQIAQFKALHIRLLDCVGRRIEK	nd	nd
CmoCh07G002420.1	CmHsf19	C	8-120	147–178	(182) KTRNPAPFLSKTY	nd	nd
CmoCh09G002330.1	CmHsf21	B	21–114	117–155	(287) IHSKKRLHPEYASNNIGKENNNKARFV	LEKDDLGLNL	nd
CmoCh01G018910.1	CmHsf1	B	120–213/21–114	117–156	(374) GSSKSFVTIVEEPKTKLFGVSLQSKKRVHPE	VLEKDDLGLNL	nd
CmoCh05G001750.1	CmHsf10	B	21–114	117–157	(337) KKRQHPDTTNYVSTSSNVSDTNKNSRGS	LLLLFKPRL	nd

DBD, DNA-binding domain; HR-A/B, oligomerization domain; NLS, nuclear localization signal; NES, nuclear export signal; AHA, transcriptional activation doma domain; nd, not found.

Gene ID	Gene Name	Group	DBD	HR-A/B	NLS	NES	AH
CmoCh10G009220.1	CmHsf23	B	21–114	117–158	(273) RGKKRMHHE	KQLLLAI	nd
CmoCh11G009050.1	CmHsf25	B	11–104	124–172	(221) RGKKRGASDEE	nd	nd
CmoCh03G000560.1	CmHsf4	B	32–125	150–187	(197) GSRKEDEDERPKLFGVRLEVEGERRRKTKR	nd	nd
CmoCh06G009230.1	CmHsf15	B	19–112	144–180	(196) EMMVMKPNLKLFGVKLEVGEEDEMVRQSKR	LKLFGVKLEV	nd
CmCh07G007220.1	CmHsf20	B	6–99	144–183	(244) EKNNDKNKTKREEEEKVEVCGNEPEAKVMKT	nd	nd
CmoCh03G009950.1	CmHsf5	B	6–99	150–188	(258) EKKKMKRVREEKIGCSNAPHAKAMK	nd	nd
CmoCh02G015130.1	CmHsf3	B	21–114	177–207	(254) FLTKTYQLVDDPDVDDLISWNEDGSTFIVW	nd	nd
CmoCh15G012680.1	CmHsf31	B	21–114	176–206	(279) IGVKRRREEE	nd	nd
CmoCh16G001410.1	CmHsf32	B	19–112	134–173	nd	LASAKSLDL	nd
CmoCh18G012590.1	CmHsf35	B	85–178	226–260	(269) ENQLKSSCKVRESVLASAKSLDLFPLKRRSEE	LASAKSLDL	nd

DBD, DNA-binding domain; HR-A/B, oligomerization domain; NLS, nuclear localization signal; NES, nuclear export signal; AHA, transcriptional activation doma domain; nd, not found.

Exon-intron analysis of 36 *Hsfs* in pumpkin

An exon-intron organization map of the 36 *CmHsf g*enes 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 pumpkin

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 > 80% and gene alignment coverage is > 0.75, were defined here as a recently duplicated gene pair. 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 pumpkin *Hsf* gene pairs ranging from 10.17 to 65.74 million years ago (Mya), averaging 21.11 Mya (Table 3).

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

Sequence	Gene alignment coverage	*Ka*	*Ks*	*Ka*/*Ks*	Divergence time (MYA)
CmHsf12-CmHsf10	0.975	0.832	1.972	0.422	65.742
CmHsf26-CmHsf9	0.910	0.126	0.462	0.273	15.416
CmHsf22-CmHsf27	0.893	0.145	0.673	0.215	22.432
CmHsf13-CmHsf28	0.946	0.232	0.436	0.531	14.535
CmHsf30-CmHsf16	0.931	0.074	0.305	0.242	10.168
CmHsf6-CmHsf19	0.863	0.080	0.336	0.238	11.204
CmHsf21-CmHsf1	0.812	0.066	0.542	0.121	18.083
CmHsf3-CmHsf31	0.814	0.106	0.431	0.246	14.354
CmHsf32-CmHsf35	0.981	0.181	0.542	0.335	18.053
Note: *Ks*, synonymous substitutions; *Ka*, nonsynonymous substitutions.					

Synteny analysis of *Hsf* genes in pumpkin

According to the synteny analysis of *Hsfs* in pumpkin and 5 other species (*Arabidopsis thaliana*; *Lagenaria siceraria*; *Cucumis sativus*; *Cucurbita maxima*; *Charleston gray*), we found that *Charleston gray* exhibited the most *Hsf* homologous genes (56), followed by *Lagenaria siceraria* (52), *Cucurbita maxima* (51) and *Cucumis sativus* (51). *Arabidopsis thaliana* presented the fewest (18) homologous genes (Fig. 4). Furthermore, the syntenic genes of the pumpkin *Hsfs* could be found on all chromosomes of *Arabidopsis thaliana*, *Lagenaria siceraria*, *Cucumis sativus*, *Cucurbita maxima*, and *Charleston gray*, indicating that the pumpkin *Hsfs* have remained closely related to those of these five species during the process of evolution. In addition, we found that certain *Hsf* genes on chromosomes Cm_Chr02, Cm_Chr06, Cm_Chr08, and Cm_Chr016 of pumpkin corresponded to two or more *Hsf* genes in *Arabidopsis*. This phenomenon was more fully reflected in the collinear diagram of pumpkin with *Lagenaria siceraria*, *Cucumis sativus*, *Cucurbita maxim*a and *Charleston gray*. In general, the collinear relationship between pumpkin and *Lagenaria siceraria*, *Cucumis sativus*, *Cucurbita maxima* or *Charleston gray*) was closer than that for *Arabidopsis*, suggesting that these species may have originated from the same ancestor.

Expression pattern of *Hsf* genes in pumpkin

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 pumpkin via quantitative real-time PCR. The transcriptional abundance of 36 pumpkin heat shock transcription factors was obtained from at least one tissue (Fig. 5; Table S1). Heat map and cluster analyses showed that *21 CmHsf* genes 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 pumpkin

To explore the potential function of *Hsfs*, the *cis*-elements in the promoters (2 kb upstream) of the 36 *Hsf* genes in pumpkin were predicted. A total of 429 *cis*-elements were found among all *CmHsfs*, which responsed to 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. 6A; 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. 6B, 6C). In addition, 27% and 45% of the *cis*-acting elements were MeJA response elements (harbouring CGTCA and TGACG motifs) and auxin response elements, respectively (Fig. 6B). 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 analysing the *cis*-acting elements of individual genes, we found that both *CmHsf34* and *CmHsf27* contained 12 abscisic acid response elements (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.

Response of *CmHsf* genes to temperature stress

To explore the response of *CmHsfs* to temperature stress, we cultured pumpkin seedlings at 4 °C and 38 °C. Under cold treatment, 44% of the *CmHsf* genes (16 genes) were significantly upregulated, and 27% of the *CmHsf* genes (10 genes) were significantly downregulated (Fig. 7; Table S3). For instance, *CmHsf3*, *CmHsf5*, *CmHsf23*, *CmHsf24*, *CmHsf27*, *CmHsf35* and *CmHsf36* were highly expressed under cold stress, at 23.06–99.04 times the control level. 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 in time or space. Under heat treatment, 24 genes were significantly upregulated, and 12 genes were significantly downregulated (Fig. 7; 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.

Response of *CmHsf* genes to hormones and salicylic acid

According to the prediction of *cis*-acting elements in the *CmHsf*s 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. 6; Table S2). Therefore, we analysed the responses of these genes to MeJA, ABA, and SA. The results of qRT-PCR analysis showed that 31 *CmHsf*s responded to MeJA to varying degrees, and the expression of *CmHsf20* was 5.1 times that in the control (Fig. 8; ; Table S4). Under ABA treatment, 21 *CmHsf*s 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 *CmHsf*s 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 analysed *Glycine max* [16], Chinese cabbage [17], *Pyrus bretschneideri* [18], potato [19], grape [20] and *Brassica oleracea* [21]. However, the Hsf family has not been extensively studied in *Cucurbita moschata*.

In our study, we identified 36 *Hsf* genes in *Cucurbita moschata* via genome-wide analysis. Compared with the 25 reported *ZmHsf*s [14] and 31 *PtHsf*s [13], we found that the number of *Hsf* genes in *Cucurbita moschata* is greater than those in *Zea mays* and *Populus trichocarpa*. Genome sizes vary significantly in these species; for instance, the genome size of *Cucurbita moschata* is 197.83 Mb, and that of *Zea mays* is 2300 Mb. The maize genome size is 11 times that of *Cucurbita moschata*. However, the number of maize *Hsf* genes is much lower than the number of *Hsf* genes in *Cucurbita moschata*. The reason for this difference might be that although two rounds of gene duplication occurred in the *Zea mays* genome during its evolution [55, 56], the *Hsf* genes of *Zea mays* underwent large gene losses. In addition, the genome of *Cucurbita moschata* also underwent a whole-genome duplication (WGD) event during the phylogeny of the species [40].

Chromosomal segmental duplications and individual gene duplications are a major driving force in the genome evolution process [57]. In this study, all *Hsf* gene pairs were found to have experienced segmental duplication events, with no tandem duplication events, indicating that segmental duplication has played a important role in the evolution of the *Cucurbita moschata Hsf* gene family. 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.

The *CmHsfs* were divided into three groups (Groups A, B and C) based on the conserved structural characteristics of the DBD and the HR-A/B domain. In addition, phylogenetic tree analysis indicated that the *CmHsfs* could be divided into three subfamilies (subfamily I, subfamily II and subfamily III). While subfamily II corresponded to group C and subgroup 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.

Cis-acting elements are essential for gene expression, and their numbers are correlated with gene expression intensity [58, 59]. *CmHsf23* and *CmHsf35* contain three low-temperature response elements, which mean that *CmHsf23* and *CmHsf35* may play key roles under low-temperature stress. The qRT-PCR results showed that *CmHsf23* and *CmHsf35* were significantly upregulated under low temperature, and the expression profiles of these two genes showed opposite trends under high-temperature stress, which further verified the response of the two genes to temperature stress. The prediction of *cis*-acting elements showed that the promoters of 28 *CmHsf* genes contained MeJA response elements, and qRT-PCR analysis showed that the expression levels of 31 genes changed to varying degrees under MeJA treatment. However, from the relative expression values, we found that the *CmHsfs* responded less to MeJA than to ABA and SA. Therefore, we concluded that pumpkin Hsf family genes were mainly involved in the responses to ABA and SA.

Conclusions

In summary, we identified 36 *CmHsfs* in the pumpkin 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 regulate pumpkin resistance.

Methods

Sequence retrieval from the *Cucurbita moschata* database and physicochemical characterization

To identify the heat shock transcription factor family in pumpkin, the genome of pumpkin was downloaded from the *Cucurbita moschata* database (CuGenDB, http://cucurbitgenomics.org/) [40]. A total of 25 *Arabidopsis thaliana Hsf* genes were obtained from the NCBI database by using their gene IDs from

*Arabidop*sis references [25]. We used 25 AtHsf proteins as queries to search against the *Cucurbita moschata* database using BLASTP with an *e*-value cut-off of $1 \times e^{-10}$. To eliminate false positives, sequences were discarded if they constituted < 70% of the corresponding *Arabidopsis* Hsf protein. SMART (http://smart.embl-heidelberg.de/) [41] and MARCOIL (http://toolkit.tuebingen.mpg.de/marcoil) were used to predict the DBDs and HR-A/B domains. After the removal of redundant genes, the remaining genes were identified as pumpkin Hsf (CmHsf) family genes.

The physical and chemical characteristics of the heat shock transcription factors, including their molecular weight (Mw), isoelectric point (*pI*) and the number of amino acids, were analysed with ExPASy (http://web.expasy.org/tools/). Information on *CmHsf* genes including their chromosomal distribution, their start and the end positions on the chromosomes were extracted from *Cucurbita moschata* database, and their subcellular locations was predicted with Plant-mPLoc [42].

Phylogenetic tree construction

To reveal the phylogenetic relationships of *Hsf* genes in pumpkin, an unrooted phylogenetic tree was constructed with MEGA 5.0 [43] according to the similarity of full-length amino acid sequence of 36 *CmHsf* genes. 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* genes in pumpkin were analysed by using Multiple Expectation Maximization or Motif Elicitation (MEME, http://meme-suite.org/) [44], and the LOGOs 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 [45] and the NetNES 1.1 Server [46], respectively. The exon-intron structures of individual *CmHsf* family genes were obtained from GSDS (Gene Structure Display Server, http: //gsds.cbi. pku.edu.cn/) [47] by comparing the coding domain sequences (CDSs) and the corresponding genomic sequences of pumpkin.

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 *Cucurbita moschata* (CuGenDB, http://cucurbitgenomics.org/). To identify gene duplications, all CDS sequences of pumpkin *Hsf* genes were subjected to BLAST searches against each other (identity > 80%, e value < $1e^{-10}$) 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 which were separated by several genes in a 100-kb were named as tandemly duplicated genes [48]. 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 [49]. To avoid the saturation of substitutions, we required that Ks values > 2.0 must be discarded [50, 51]. The divergence time (T) was computed according to the formulathe ($T = Ks/2\lambda \times 10^{-6}$ million years ago (Mya), $\lambda = 1.5 \times 10^{-8}$) in the previous literature [52]. The criteria for identifying gene collinearity were based on previous reports [53], and the synteny relationships between the heat shock transcription factors of pumpkin and those of other species (*Arabidopsis thaliana*, *Cucumis sativus*, *Cucurbita maxima*, *Charleston gray*, *Lagenaria siceraria*) were constructed using Dual Synteny Plotter software (https://github.com/CJ-Chen/TBtools) [54].

Analysis of *cis*-acting elements of CmHsf gene promoters in pumpkin

The promoter sequences (2 kb before the start codon) of all *CmHsf* genes were extracted from the *Cucurbita moschata* database, and we predicted the promoter *cis*-acting elements of *CmHsfs* by using Dual Synteny Plotter software [54].

Plant material, growth conditions and stress treatment

The pumpkin 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 matrix-meteorite (3:1) mixture and grown in a plant growth chamber. The artificial growth conditions were set to 25 ℃ / 16 ℃, 16 h light / 8 h dark and 65% relative humidity. We sampled and analysed different tissues (roots, stems, cotyledons and true leaves) of two-month-old seedlings. In addition, some of the seedlings were transferred to 38 ℃ for 6 h heat treatment, or transferred to 4 ℃ for 6 h cold treatment. Another portion of the seedlings were cultured in 1/2 Hoagland solution, pH 6.5. After 5 days of adaptation, the plants were cultured with the following treatments: (1) control; (2) 1 mM MeJA; (3) 5 mM salicylic acid (SA); (4) 100 μM abscisic acid (ABA). Leaf samples were collected at 10 h after the above treatments. Three independent biological replications formed one sample. Control and stress-treated samples were frozen in liquid nitrogen and stored at -70 ℃ 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 in three biological replicates using the SYBR Premix ExTaq kit (TaKaRa, Dalian). In order to verify the specificity of gene primers, the target genes and the reference gene (*β*-Actin) primers were aligned at the pumpkin genome database. qRT-PCR analysis was performed on an ABI7500 Real-Time PCR System (Applied Biosystems) with the following cycling profile: stage 1, 95 ℃ 20 s; stage 2, 95 ℃ 3 s, 60 ℃ 30 s (40 cycles); stage 3, 95 ℃ 15 s, 60 ℃ 1 min, 95 ℃ 15 s. Stage 3 was used to perform a melting curve. Each reaction was performed three times, and the relative gene expression was calculated according to the $2^{-\Delta\Delta Ct}$ 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*:*Cucurbita 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 JYconceived, designed and supervised the experiment; CS and JP wrote the manuscript; CS and JP performed theexperiment; 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

Not applicable.

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

[1]School of Resources and Environmental Sciences, Henan Institute of Science and Technology, Xinxiang 453003, China; [2]School of Horticulture and Landscape Architecture, Henan Institute of Science and Technology, Xinxiang, 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 Bio. 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):310-316.
4. Baniwal SK, Bharti K, Chan KY, Fauth M, Ganguli A, Kotak S, Mishra SK, Nover L, Port M, Scharf KD, Tripp J, Weber C, Zielinski D, 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):471-487.
5. Wiederrecht G, Seto D and Parker CS. Isolation of the gene encoding the S. cerevisiae heat shock transcription factor. Cell. 1988; 54(6):841-853.
6. SorgerPK and PelhamHRB. Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. Cell. 1988; 54:855-864.
7. Clos J, Westwood JT, Becker PB, Wilson S, Lambert U, Wu C. Molecular cloning and expression of ahexameric drosophila heat stress factor subject to negative regulation. Cell. 1990; 63(5):1085-1097.
8. Rabindran SK, Giorgi G, Clos J, Wu C. Molecular cloning and expression of a human heat stress factor. P Natl Acad SciUSA. 1991; 88(16):6906-6910.
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):1902-1911.
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 SciUSA. 1991; 88(16):6911-6915.
11. Huhel A, Schoffl F. *Arabidopsis* heat shock factor: isolation and characterization of the gene and the recombinant protein. Plant Mol Biol. 1994; 26(1):353-362.
12. Yamanouchi U, Yano M, Lin H, Ashikari M, Yamada K. A rice spotted leaf gene, *Spl7*, encodes a heat stress transcription factor protein. P Natl Acad SciUSA. 2002; 99(11):7530-7535.

13. 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 *Populus trichocarpa*. Plant Omics. 2013; 6(6):415-424.
14. 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.
15. Scharf KD, Rose S, Zott W, Schŏffl 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):4495-4501.
16. 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):127-135.
17. Song XM, Liu GF, Duan WK, Liu TK, Huang ZN, Ren J, Li Y, Hou XL. Genome-wide identifcation, classifcation and expression analysis of the heat shock transcription factor family in Chinese cabbage. Mol Genet Genomics. 2014; 289(4):541-551.
18. Qiao X, Li M, Li L, Yin H, Wu J, Zhang S. Genome-wide identification and comparative analysis of the heat shock transcription factor family in Chinese white pear (*Pyrus bretschneideri*) and five other *Rosaceae* species. BMC Plant Biol.2015; 15(1):1-16.
19. 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.
20. Liu GT, Chai FM, Wang Y, Jiang JZ, 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.
21. Lohani N , Golicz AA , Singh MB , Bhalla PL. Genome-wide analysis of the *Hsf* gene family in *Brassica oleracea* and a comparative analysis of the *Hsf* gene family in *B. oleracea*, *B. rapa* and *B. napus*. Funct Integr Genomic. 2019; 19:515-531.
22. Scharf KD, Berberich T, Ebersberger I, Nover L. The plant heat stress transcription factor (*Hsf*) family: structure, function and evolution.BBA-Gene RegulMech. 2012; 1819(20):104-119.
23. Gorlich D, Kutay U. Transport between the cell nucleus and the cytoplasm. AnnuRevCellDevBi.1999; 15(1):607-660.
24. 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*. MolCelBiol. 2001; 21(5):1759-1768.
25. 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.
26. 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):265-278.
27. Ikeda M, Ohme-Takagi M. A novel group of transcriptional repressors in *Arabidopsis*.Plant Cell Physiol.2009; 50(5):970-975.
28. 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.PNatl AcadSciUSA. 2006; 103(49):18822-18827.
29. 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 maysL*.Plant J.2007; 50(1):54-59.
30. 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):1555-1567.
31. Charng 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):251-262.
32. 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):957-967.
33. Almoguera 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.JBiolChem. 2002; 277(46):43866-43872.
34. Miller G, Mittler R. Could heat shock transcription factors function as hydrogen peroxide sensors in plants?AnnBot.2006; 98(2):279-288.
35. 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):347-363.
36. 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.
37. Nishizawa A, Ybuta 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):535-547 .
38. Davletova S, Rizhsky L, Liang H, Shengqiang, Z, Oliver DJ, Coutu 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):268-281.
39. 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.
40. 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 genomes. Mol Plant.2017; 10(10):1293-1306.
41. Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 2012; 40:302-305.

42. 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.
43. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016; 33(7):1870-1874.
44. Bailey TL, Elkan C. The value of prior knowledge in discovering motifs with MEME. Proceeding of the Third International Conference on Intelligent Systems for Molecular Biology 1995; 3: 21-29.
45. 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 SciUSA. 2009; 106(25):10171-10176.
46. La Cour T, Kiemer L, Molgaard A, Gupta R, Skriver K, Brunak S. Analysis and prediction of leucine-rich nuclear export signals. Protein EngDesSel. 2004; 17(6):527-536.
47. 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):1296-1297.
48. 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.
49. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. KaKs calculating Ka and Ks through model selection and model averaging. GenomProteomBioinf. 2006; 4(4):259-263.
50. Blanc G, Wolfe KH. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell. 2004; 16(7):1667-1678 .
51. 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):218-228.
52. Emanuelsson O, Nielsen H, Brunak S, Heijne GV. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J MolBiol. 2000; 300(4):1005-1016.
53. 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.
54. Chen C, Xia R, Chen H, He Y. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. BioRxiv. 2018;289660.
55. Swigoňová Z, Lai J, Ma J. Close split of sorghum and maize genome progenitors. Genome Res. 2004; 14(10):1916-1923.
56. WeiFH, 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, McmullenMS, Davis GL, Bowers JE, Paterson AH, Schaeffer ML, Gardiner JM, Cone KC, Messing J, Soderlund C, Wing RA. Physical and genetic structure of the maize genome reflects its complex evolutionary history. PLoS Genet. 2007; 3(7):1254-1263.
57. 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.
58. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Quackenbush. TM4 microarray software suite. MethodEnzymol.2006; 411(2):134-193.
59. Peng S, Huang ZC, Ou YLJ, Cheng J, Zeng FH. Research progress of artificial promoter in plant genetic engineering.JPlant Physiol.2011; 47:141-146.

Figures

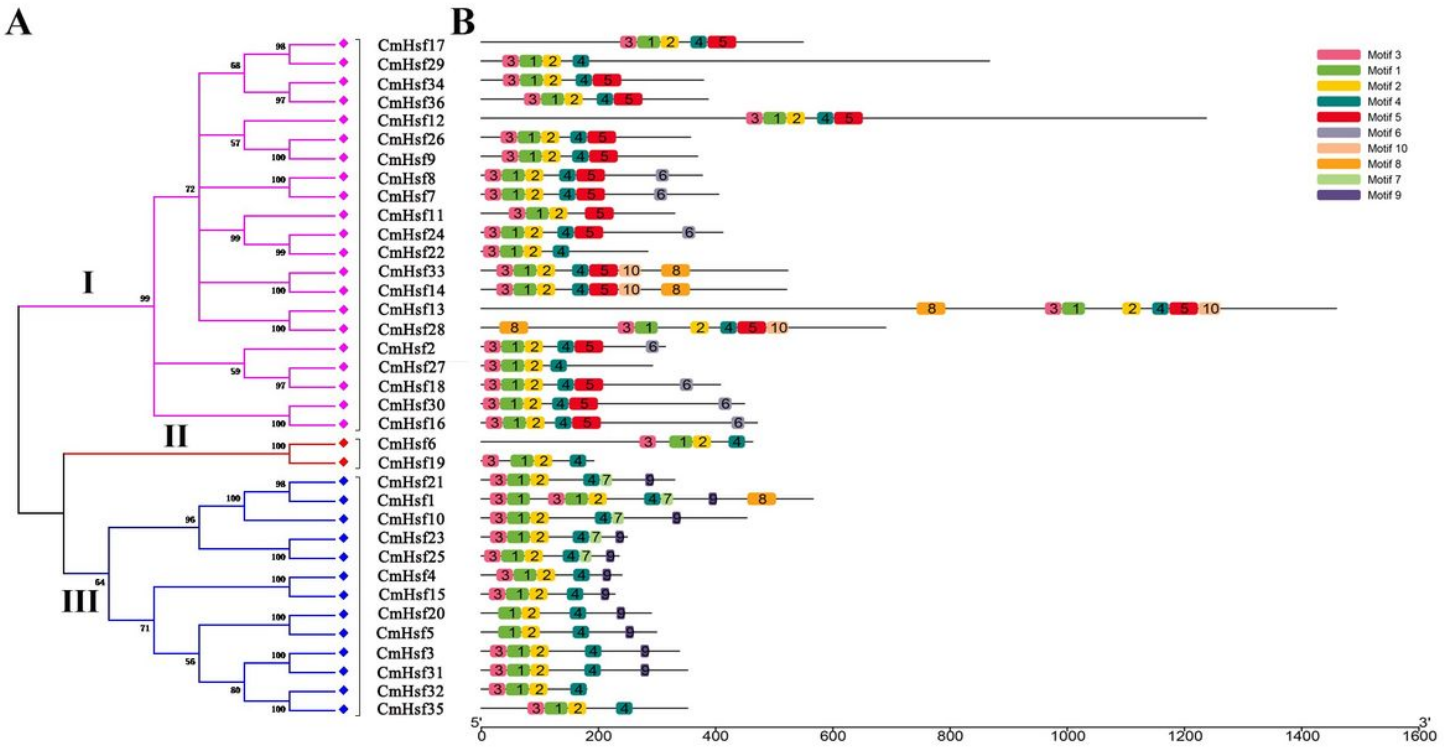

Figure 1

Classification and conserved motifs of 36 Hsfs in pumpkin. 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) are highlighted with different coloured branch lines. B. Schematic representation of conserved motifs in 36 CmHsfs.Each motif is represented by a numbered coloured box on the right. The same number in different proteins refers to the same motif.

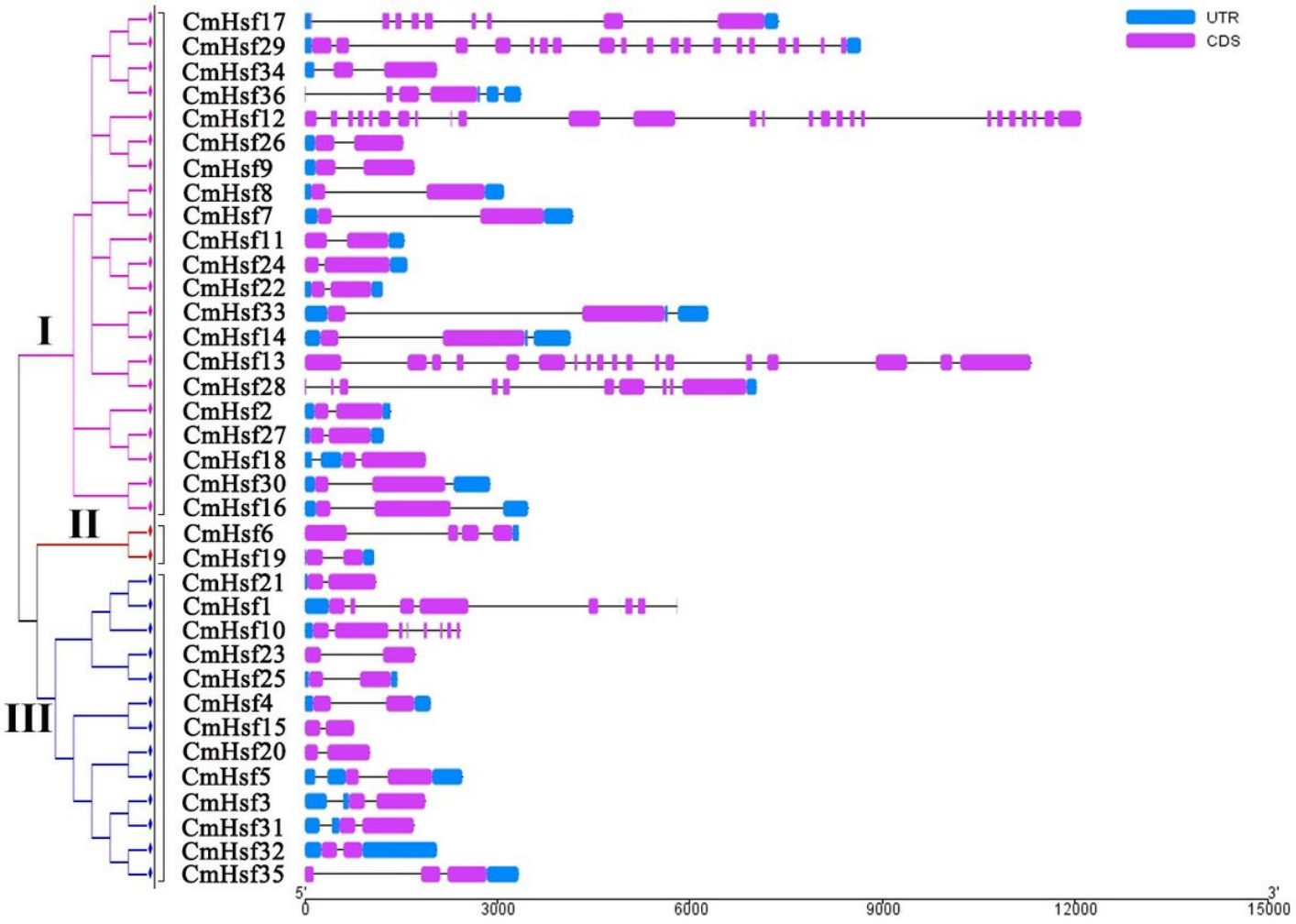

Figure 2

Exon-intron organization of 36 Hsfs in pumpkin. The exons and introns are represented by pink boxes and grey lines, respectively. Untranslated regions (UTRs) are indicated by blue boxes. The sizes of the exons and introns can be estimated using the scale at the bottom.

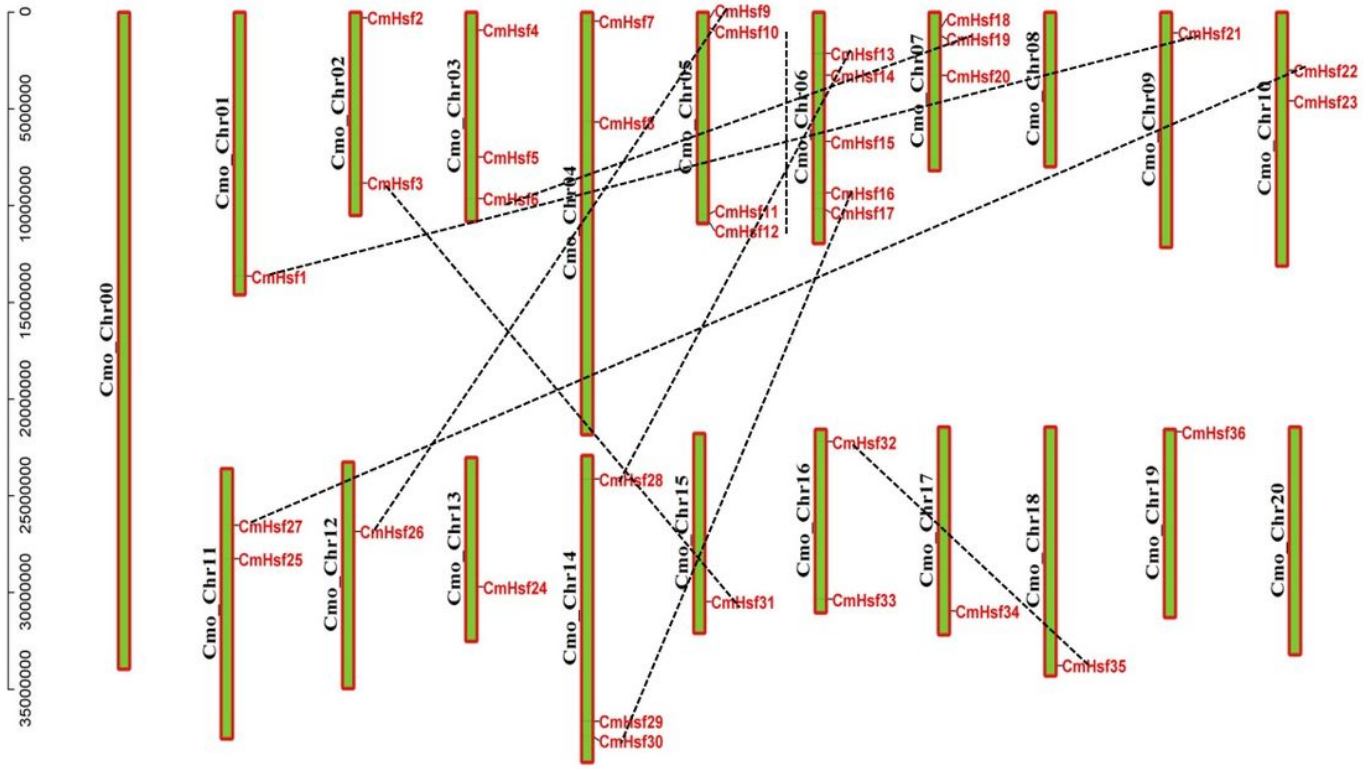

Figure 3

Chromosomal distribution and duplication events of Hsf genes in pumpkin. The duplicated pumpkin Hsf gene pairs are connected with black lines.

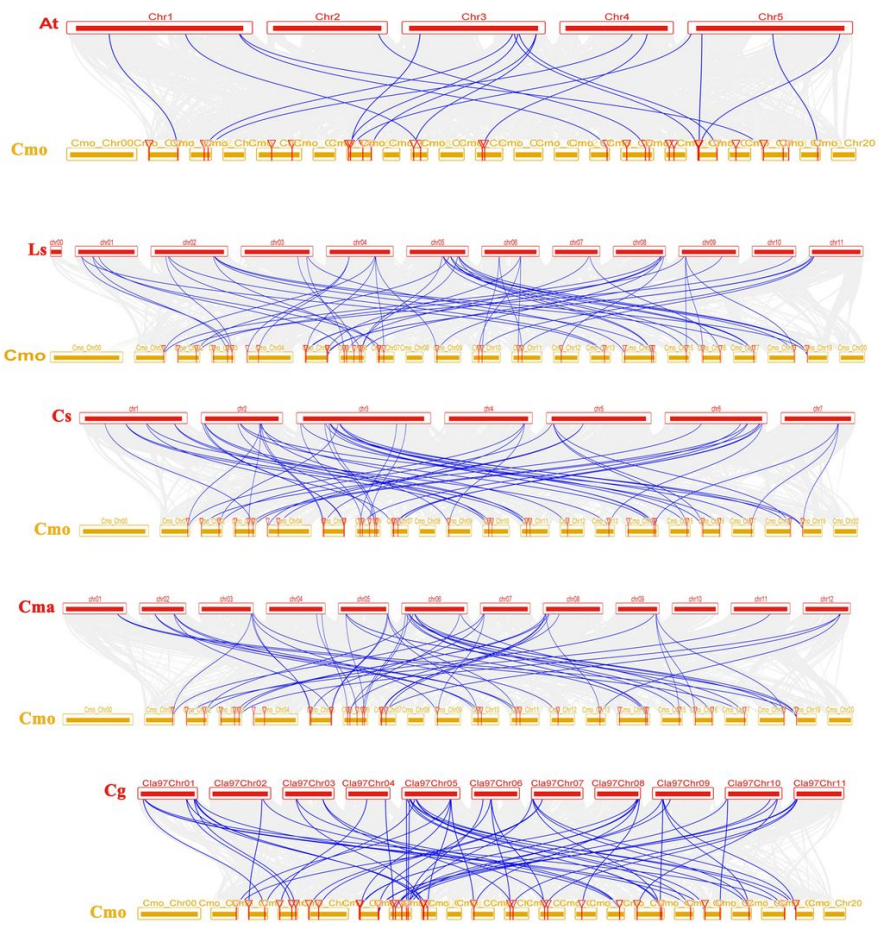

Figure 4

Synteny analysis of the Hsfgenes between pumpkin and five other species.At, Arabidopsis thaliana; Ls, Lagenaria siceraria; Cs, Cucumis sativus; Cma, Cucurbita maxima; Cg, Charleston gray grey; Cmo, Cucurbita moschata. The gray lines in the background indicate the collinear blocks in the genome of pumpkin and other plants, while blue lines in the background highlight syntenic Hsf gene pairs.

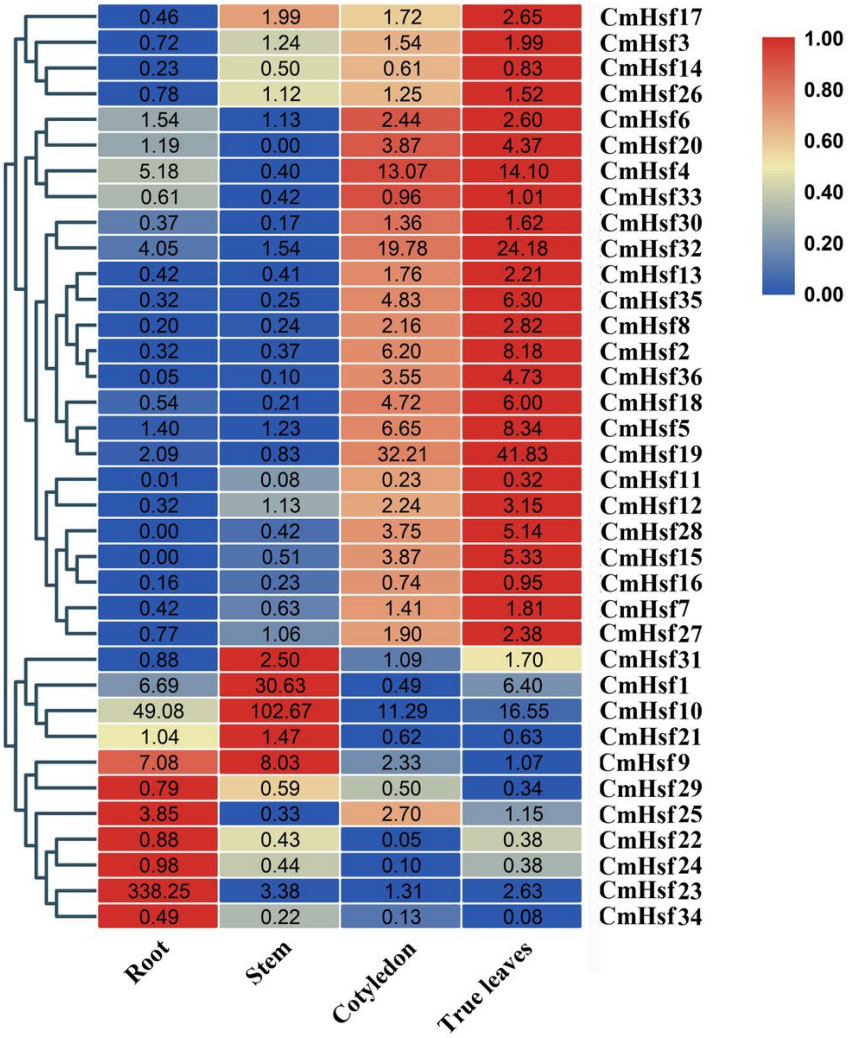

Figure 5

Heat map and hierarchical clustering of 36 pumpkinHsf genes in the roots, stems, cotyledons and true leaves. The results were calculated via the 2−ΔΔCt method, and the reference gene(β-Actin) was used to correct the expression level of target genes.

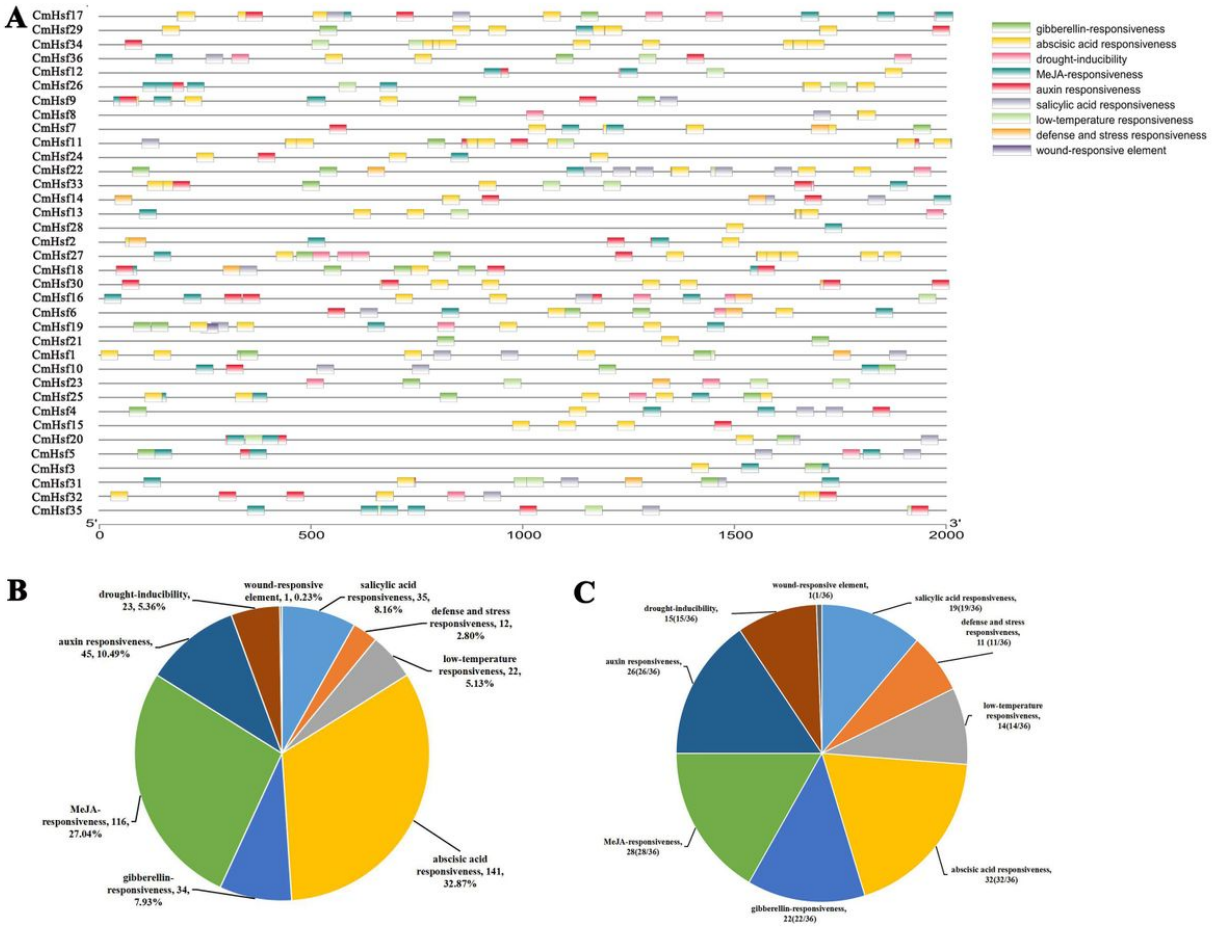

Figure 6

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 pumpkin. The squares on the right represent cis-acting elements that respond to a total of 9 stresses. Different colours indicate cis-acting elements that participate in different stresses. The coordinates at the bottom of the figure indicate the length of the gene promoter. B. The distribution of 429 cis-acting elements related to 9 abiotic stresses. C. The proportion of 36 CmHsfs related to 9 abiotic stresses.

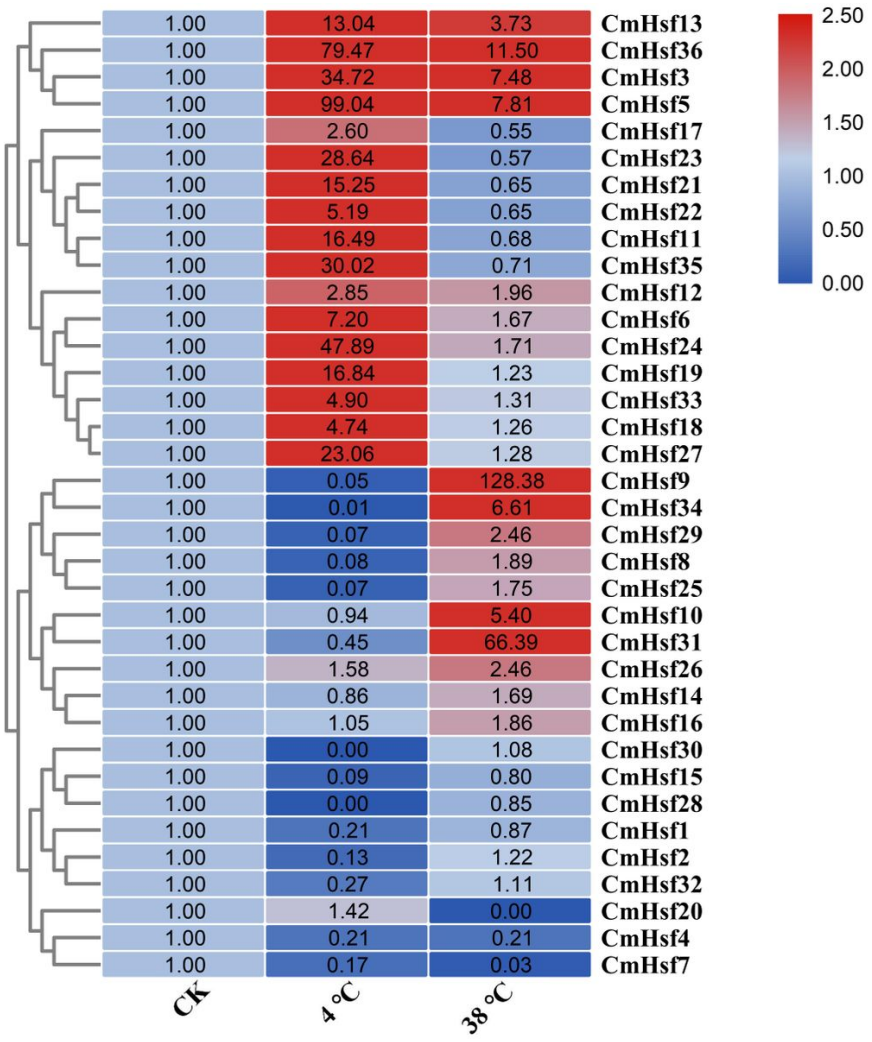

Figure 7

Heat map and hierarchical clustering of 36 CmHsf genes in true leaves under cold stress and heat stress. The relative expression levels of CmHsf genes at 10 h after cold stress (4 °C) and heat stress (38 °C)werequantifed against the seedlings without stress (CK), and the reference gene(β-Actin) was used to correct the expression level of target genes, and the expression level of the control was set as 1.The bar on the right of the heat map represents the relative expression levels.

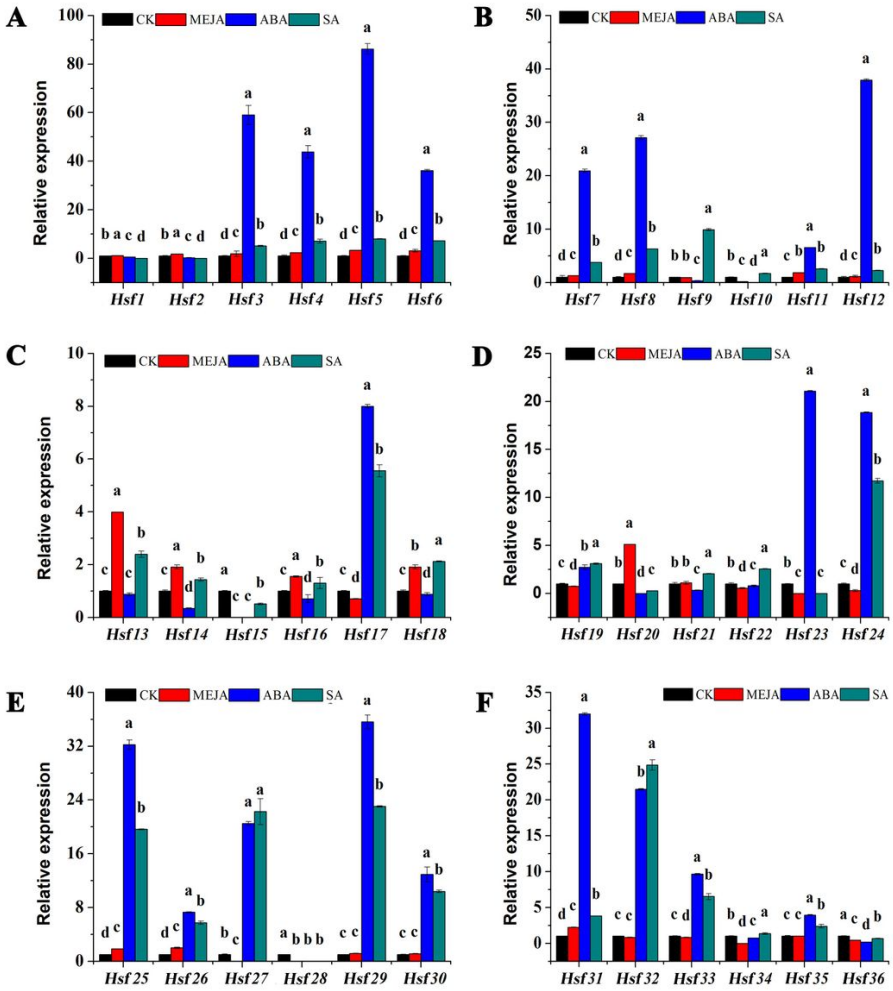

Figure 8

Expression profiles of 36 CmHsf genes in true leaves under MeJA, ABA and SA treatments. The data represents the expression levels of CmHsf genes at 10 h after the MeJA, ABA and SA treatments. CK refers to untreated plants in this figure. The results were calculated via the 2−ΔΔCt method, and the reference gene(β-Actin) was used to correct the expression level of target genes. The expression level of the control was set as 1. The data are presented as the means of three replicates, and the error bars represent the standard deviations of the means. According to Welch's t-test, different letters above the bars indicate significant differences ($p < 0.05$) between different treatments.

Supplementary Files

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

- Fig.S2.tif
- Fig.S1.tif
- TableS5.docx
- TableS4.docx
- TableS3.docx
- TableS2.doc
- TableS1.docx