Plant Omics Journal POJ

POJ 6(6):449-455 (2013) ISSN:1836-3644

Genome-wide identification and analysis of heat shock transcription factor family in cucumber (*Cucumis sativus* L.)

Shengjun Zhou[1*], Peng Zhang[1], Zange Jing[2], Jianlei Shi[2]

[1]Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, People's Republic of China
[2]Southern Zhejiang Key Laboratory of Crop Breeding, Department of Agriculture and Biotechnology, Wenzhou Vocational College of Science and Technology, Wenzhou 325006, People's Republic of China

***Correspondence author: sjzhou_zj@sina.com**

Abstract

Heat shock transcription factor (*HSF*) plays an important role in the expression regulation of thermal response genes in plants. However, little is known about this family in cucumber. Recently, the availability of cucumber genome sequences has provided an opportunity for identifying these *Hsfs*. In the present study, all members of the *HSF* gene family of cucumber (*Cucumis sativus* L.) were identified from the sequenced genome and analyzed *in silico*. We identified at least 21 *HSF* genes in the cucumber, which encode proteins with length between 184 and 560 amino acids. Multiple sequence alignments showed that cucumber *HSF* proteins possess highly conserved DNA binding domains and other extensively conserved motifs. These *HSF* genes are unevenly distributed in all seven chromosomes, with chromosomes 2 and 3 both having the greatest number of *HSF* genes. In addition, phylogenetic analysis of the *HSFs* from cucumber and *Arabidopsis thaliana* showed that these proteins could be divided into three families (A, B and C). The A family was then further divided into nine subfamilies, and there were eight pairs of orthologous genes and five pairs of paralogous genes. *HSF* members in some subfamilies were from cucumber and *Arabidopsis*, indicating that the *HSF* gene family existed before the separation of cucumber and *Arabidopsis thaliana*.

Keywords: cucumber; heat stress; heat shock transcription factor; bioinformatics; phylogenetic relationships.
Abbreviations: HSF_heat shock transcription factor; HSP_ heat shock proteins; DBD_DNA binding domain; HR-A/B_hydrophobic amino repeat; NLS_nuclear localization signal; NES_nuclear export signal.

Introduction

Plants display a series of stress responses under heat shock stress conditions; heat shock proteins (*HSP*) are one of the major accumulated components in plants under these conditions, and the regulation at the transcription levels is the main regulation method for plant heat shock protein expression (Kimpel et al., 1990; Kotak et al., 2004; Lin et al., 2011). Heat shock transcription factors (*HSF*) are major factors initiating the expression of downstream *HSP* genes under heat stress conditions (Baniwal et al., 2004; Kotak et al., 2007; Ahmad, 2012). *HSFs* activate gene expression of *HSPs* by binding to the conserved heat shock element (HSE), which further induce heat stress responses to resist damages from adverse environments. Since *HSF* genes were first cloned from yeast in the 1980s (Sorger and Pelham, 1988; Wiederrecht et al., 1988), many mammalian *HSF* genes have been isolated (Clos et al., 1990; Rabindran et al., 1991; Sarge et al., 1991; Schuetz et al., 1991). The plant *HSF* gene was first cloned from *Solanum lycopersicum* (Scharf et al., 1990). With more and more plant genomics being sequences, the corresponding *HSF* genes in *Arabidopsis* and rice were also identified (Huhel et al., 1994; Yamanouchi et al., 2002). Subsequently, at least 52 and 30 *HSF* genes were found in soybean (Chung et al., 2013) and maize (Lin et al., 2001), respectively. Therefore, plant *HSF* genes are shown to be a large gene family (Nover et al., 1996; Kotak et al., 2004). Previous studies have demonstrated that *HSFs* possess highly conserved domains. Typical *HSFs* generally include four components: a DNA binding domain (DBD) at the N-terminal, a bidirectional oligomerization domain composed of hydrophobic amino repeat (HR-A/B), a cell nuclear localization signal (NLS), and a nuclear export signal (NES). In rare cases, there was also an additional acidic C-terminal activation domain (CTAD) (Peteranderl et al., 1999). According to the structural characteristics of the conserved DBD and HR-A/B regions, *HSFs* are also divided into three subfamilies (A, B and C). The main differences between the three subfamilies were as follows: the B subfamily has 7 amino acid residues in the HR-A/B structural domain, whereas the A and C subfamilies have extra 21 and 7 residues, respectively, in addition to the 7 residues. In addition, there was no CTAD structural domain in the B and C subfamilies (Klaus-Dieter et al., 2012). Cucumber (*Cucumis sativus* L.) is one of the world's most important economic vegetable crops, and a model system for sex determination studies and plant vascular biology (Huang et al., 2009). Recently, sequencing of the cucumber genome has been completed (Huang et al., 2009), providing great advantages for studying *HSFs* in cucumbers. In the present study, identification and comparative analysis of the *HSFs* in cucumber were conducted using bioinformatics methods. Moreover, sequence characteristics, phylogenetic relationships, chromosomal localization and conserved motifs were also performed. These results will be helpful in better understanding the complexity of the *HSF* gene family to

build a basis for further deciphering the functions of *HSF*.

Results and Discussion

Identification of heat shock transcription factor gene family in cucumber

In order to comprehensively acquire cucumber *HSF* genes, the genome data of cucumber (http://cucumber.genomics.org.cn/page/cucumber/index.jsp) were searched using the BLASTP method. The acquired genes were examined in the Pfam database (http://pfam.sanger.ac.uk). In the meantime, to confirm the accuracy of the obtained *HSF* genes, the coil structures characteristic of the *HSF* genes were detected in the MARCOIL database (http://toolkit.tuebingen.mpg.de/sections/seqanal). Twenty one cucumber *HSF* genes were eventually identified and named as *CsHSF01*-*CsHSF21*. The length of *CsHSF* proteins were between 184 (*CsHSF14*) and 560 (*CsHSF12*) amino acids. Their theoretical isoelectric points spanned from 4.70 (*CsHSF17*) and 9.10 (*CsHSF13*), indicating the presence of both basic and acidic proteins. The molecular weights varied between 21.2 kDa (*CsHSF14*) and 62.3 kDa (*CsHSF12*) (Table 1).

Multiple sequence alignments of HSF genes in cucumber

CsHSF had typical structures of transcription factors, including conserved DNA binding domain (DBD) (Scharf et al., 1990). The DBD domain in the N terminal of the *HSFs* contained three α helixes (α_1~α_3) and four β folds (β_1~β_4), that had the functions of specific recognition and accurate location of heat stress elements. To analyze the DBD structure in *CsHSF*, multiple sequence alignments of the DBD domains were conducted by using Clustal X (Fig. 1), and the results demonstrated that the DBD domains are highly conserved in most *CsHSF* proteins. The DBD domain is 94 residues in length. However, the DBD domains of *CsHSF14* and *CsHSF17* were partially truncated, with 32 residues in the β_3 and β_4 of *CsHSF14* and 21 residues in the β_4 of *CsHSF17* being deleted. These deletions may have been attributed to the genetic diversity in *CsHSFs*. In order to further analyze the structural characteristics of *HSF* genes in cucumber, a *CsHSF* phylogenetic tree and intron-exon structure diagram were constructed (Fig. 2). The *CsHSF* gene family could be divided into two subfamilies (I-II), then each subfamily was further subdivided into two groups. The numbers of gene members in each group were different. The intron-exon structure diagram showed that there were few number variations of introns in all of the *CsHSF* genes. Most *CsHSF* contained one intron; only *CsHSF10* and *CsHSF03* contained two introns, and *CsHSF17* contained three introns.

Distribution of conserved motifs in cucumber HSFs

Conserved motifs in the 21 *CsHSF* proteins were identified using the MEME online tool (Fig. 3). The results demonstrated that these proteins contained 15 conserved motifs with lengths of 6-40 amino acids (Table 2). The distribution of the 15 conserved motifs in the *CsHSF* proteins was further analyzed. The number of the conserved motifs in each gene varied between 3 and 12. *CsHSF20* had the least number of motifs (Motif1, Motif2 and Motif14), whereas all *CsHSF01*, *CsHSF02* and *CsHSF03* had 12 motifs. Two motifs, namely Motif1 and Motif2, were completely conserved in all of the *CsHSF* proteins. In general, they were highly conserved in all of the *CsHSF* gene family members.

Chromosomal localizations of cucumber HSF genes

The chromosomal localizations of the 21 *CsHSF* genes were analyzed according to the data information of cucumber genome sequencing (Fig. 4). Twenty *CsHSF* genes were located in the seven chromosomes, with the exception that one *HSF* (*CsHSF21*) was located in Scaffold_repeat037858 that has not been fully assembled. The chromosomal localizations of 20 *CsHSF* genes were not distributed evenly. Chromosomes 2 and 3 each had five *CsHSFs*, collectively accounting for 50% of the gene family. Six *CsHSFs*, namely *CsHSF14* and *CsHSF15*, *CsHSF16* and *CsHSF17*, and *CsHSF18* and *CsHSF19* were located on chromosomes 4, 5 and 6, respectively. Thereamining *CsHSF* genes were located on Chromosome 1 (*CsHSF01*, *CsHSF02* and *CsHSF03*) and Chrosomose 7(*CsHSF20*).

Construction and analysis of phylogenetic tree of cucumber HSF genes

In order to analyze the evolutionary relationship of the cucumber *HSF* gene family, the *HSF* genes of the model plant *Arabidopsis thaliana* were selected to construct the phylogenetic trees (Fig. 5). According to the *HSF* classification methods of *Arabidopsis thaliana* used by previous researchers (Nover et al., 1996), the HSF gene family members of the two species were divided into three subfamilies (A, B and C). Each subfamilies contained members from the two species, in which A contained 26 HSF genes, B contained 13 genes, and C only had 3 genes, the least of the three subfamilies. A was further divided into nine subgroups (A1-A9). A phylogenetic tree is used not only to analyze the phylogenetic relationships in every gene family, but also to recognize homologous genes. As illustrated in Fig. 5, there were eight pairs of orthologous genes, namely At1g32330 and CsHSF09, At2g26150 and CsHSF07, CsHSF10 and At3g22830, CsHSF16 and At1g67970, CsHSF05 and At5g54070, CsHSF12 and At5g03720, CsHSF17 and At4g11660, CsHSF19 and At4g36990; and five pairs of paralogous genes, including three pairs from *Arabidopsis thaliana* (namely At5g16820 and At3g02990, At3g63350 and At3g51910, At3g63350 and At3g51910), and two pairs from cucumber (namely Csa009386 and Csa011478, Csa014653 and Csa026480). Although similar numbers of HSFs were found in these two species, multiple pairs of paralogous genes among them identified by the phylogenetic tree indicated that the *HSF* genes may have underwent different evolution processes after the divergence of *Arabidopsis thaliana* and cucumber. In addition, there was no orthologous gene of the *Arabidopsis thaliana* HSF gene At4g18870 and At5g43840 in cucumber, indicating that there may have been a deletion of homologous genes in cucumber genome. In the present investigation, 21 *HSF* genes were identified from the cucumber genome by using bioinformatics methods, which were divided into three subfamilies, namely groups A, B and C, which contained 11, 8 and 2 *HSF* genes, respectively. The gene number in each group was different, suggesting that cucumber *HSF* gene family members were unevenly distributed and extensively diversified. This provided gene resources for the functional studies of cucumber HSF genes. Phylogenetic relationships may provide help in searching for orthologous genes and paralogous genes within species. In the present study, there were only eight pairs of orthologous genes and five pairs of paralogous *HSF* genes, indicating that most of these gene

Table 1. Basic characteristics of the HSFs gene family members in the cucumber genome.

Gene	Cucumber Genomics ID[a]	Chromosomal location	Length of genomic sequence	ORF Length (bp)	Length (aa)[b]	Isoelectric point[c]	Molecule weight[d] (kDa)
CsHSF01	Csa021238	Chr1:11832007-11833006	1000	876	291	5.21	33.2
CsHSF02	Csa017341	Chr1:15259860-15261199	1304	1230	409	5.20	46.7
CsHSF03	Csa006414	Chr1:21622359-21625154	2796	891	296	8.94	34.3
CsHSF04	Csa006876	Chr2:214174-215205	1032	915	304	6.15	34.2
CsHSF05	Csa009195	Chr2:5793154-5794838	1685	1149	382	5.66	44.0
CsHSF06	Csa014653	Chr2:10689909-10691168	1263	1047	348	5.78	40.1
CsHSF07	Csa011545	Chr2:15110057-15111445	1389	1095	364	5.12	41.5
CsHSF08	Csa011478	Chr2:15757198-15758515	1318	1125	374	7.28	41.7
CsHSF09	Csa019047	Chr3: 290397-294690	4294	1557	518	5.04	57.0
CsHSF10	Csa000108	Chr3:9135193-9137428	2236	1092	363	5.48	42.5
CsHSF11	Csa008853	Chr3:11762367-11763348	982	882	293	5.56	33.3
CsHSF12	Csa005086	Chr3:25459662-25462183	2522	1683	560	5.09	62.3
CsHSF13	Csa002380	Chr3:28298304-28299992	1689	759	252	9.10	29.5
CsHSF14	Csa019876	Chr4:7370508-7373291	2784	555	184	8.44	21.2
CsHSF15	Csa009386	Chr4:20950404-20951491	1088	990	329	7.35	38.0
CsHSF16	Csa010629	Chr5:9885091-9888254	3164	1191	396	4.77	45.9
CsHSF17	Csa012865	Chr5:15335879-15337074	1196	765	254	4.70	28.2
CsHSF18	Csa011800	Chr6:10167494-10168488	995	903	300	7.78	34.8
CsHSF19	Csa010236	Chr6:21145571-21147058	1488	873	290	5.72	32.7
CsHSF20	Csa016743	Chr7:12554732-12556255	1524	1113	370	5.13	42.7
CsHSF21	Csa026480	Scaffold_repeat037858: 1545-2807	1263	1050	349	5.78	40.2

[a]Available at http://cucumber.genomics.org.cn/page/cucumber/index.jsp. [b]length (number of amino acids) . [c]molecular weight (kilodaltons) . [d]isoelectric point (pI) of the deduced polypeptide

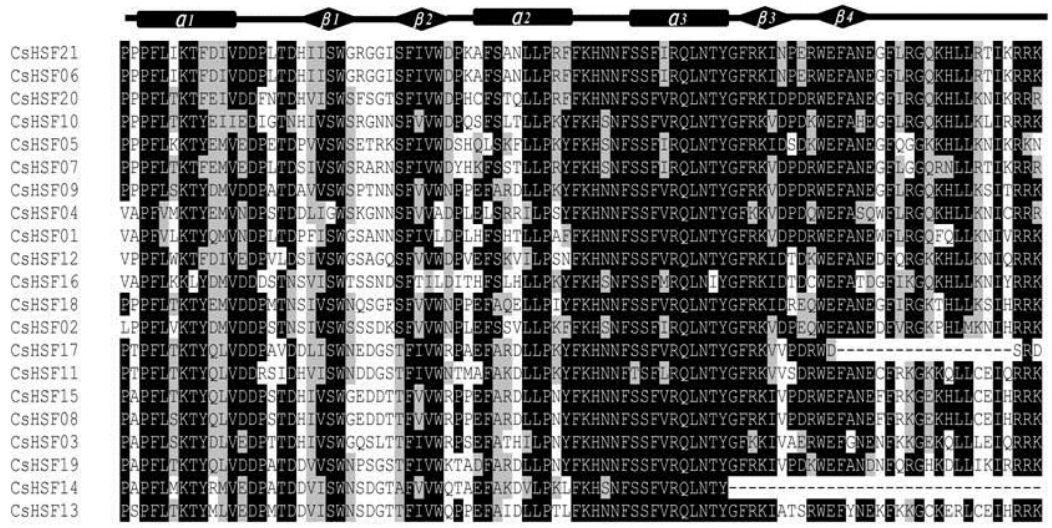

Fig 1. Multiple sequence alignment of the DBD domains of the HSF protein family in cucumber. The multiple alignment results clearly show the highly conserved DBD domains among cucumber HSF genes. The secondary structure elements of DBD (α1- β 1- β 2-α2-α3- β 3- β 4) are shown above the alignment. Cylindrical tubes represent a-helices and diamond arrows represent b-sheets.

family members were expanded in cucumber and *Arabidopsis thaliana* according to their species-specific modes. The phenomenon has been extensively confirmed in gene family analysis of other plant species (Bai et al., 2002; Zhang et al., 2005; Jain et al., 2006). Furthermore, there were no orthologous *HSF* genes for At4g18870 and At5g43840 present in cucumber, suggesting that gene loss events may occur in the evolution process of cucumber *HSF*. Previous researches showed that MdHsfs transcripts were detected in several apple organs, and expression changes were observed in developing flowers and fruits as well as in leaves in the field and exposed to the naturally increased temperatures (Giorno et al., 2012). Furthermore, the transcripts of all ZmHsf genes exhibit different expression levels in heat stress treatment (Lin et al., 2011). These results indicated that HSF genes play an important role in plant in response to heat response. In the present study, HSF was identified and analyzed at the whole genome level using bioinformatics methods with cucumber genome sequence as the materials, and 21 HSF sequences were obtained. Then sequence characteristics, gene structure and molecule evolution were analyzed, which will form the basis for further investigations of gene expression and function.

Table 2. Motif sequences identified by MEME tools.

Motif	Width	Best possible match
1	40	WDPHEFARDLLPKYFKHNNFSSFVRQLNTYGFRKIDPDRW
2	29	PFLTKTYDMVDDPATDHIISWGRDGTSFI
3	27	EFANEGFLRGQKHLLCNIHRRKPIHNH
4	40	LWEEIDRLRRDKQMLMMELVKMRQQCQNTRYYIQTMCQRL
5	29	TEMKQQQMMYFLARAMQNPDFIHQLIQQK
6	11	PMEGLHDVGPP
7	8	CVEIGRFG
8	32	CGYEVSELEALALEMQGLGRAVKKETKVKEEM
9	20	KKKDIEEAITKKRRRPIDQG
10	29	ENGDAELDEGFWEEFFSGRIEEGENDDMV
11	12	KLFGVWIGGKKR
12	21	MNPQYPVKEEDWGPSSSEFGG
13	10	GRFGYLGSI
14	6	HYHHHH
15	7	GQIVKYQ

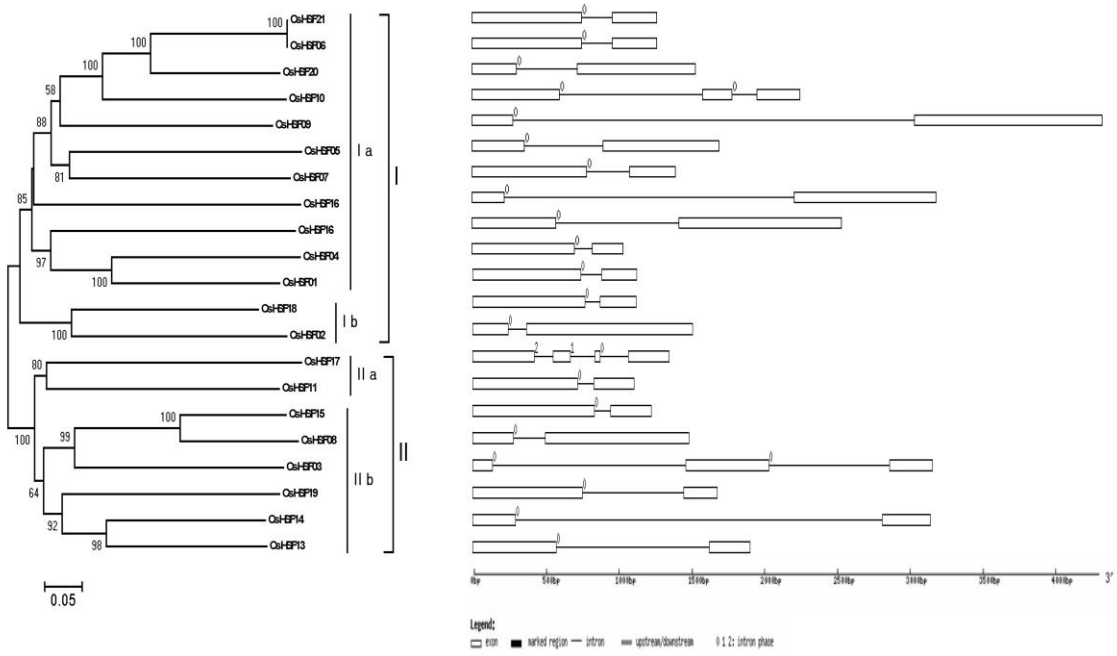

Fig 2. Phylogenetic analysis and intron/exon configurations of HSFs in cucumber. The phylogenetic tree of CsHSFs was constructed using MEGA 5.0. Introns and exons are drawn to scale with the full encoding regions of their respective genes. Boxes indicate the exon, and lines indicate the intron. 0 = intron phrase 0; 1 = intron phrase 1; 2 = intron phrase 2.

Materials and methods

Plant genomics materials

The related information of the HSF family members in cucumber was mainly obtained from the database of genomes in cucumber (http://cucumber.genomics.org.cn/page/cucumber/index.jsp), and the data of the HSF family members in *Arabidopsis thaliana* were obtained from the database at http://datf.cbi.pku.edu.cn/.

Identification of HSF transcription factor in cucumber

In order to comprehensively identify the HSF transcription factor in cucumber, the cucumber genome database was conducted by BLAST research, using the amino acid sequences of the HSF members in the model plant *Arabidopsis thaliana*. The DBD structural domain of the obtained HSF transcription factor was identified by using PFAM (http://pfam.janelia.org/). Moreover, the coil structure specific to the HSF protein was detected by the online tool MARCOIL (http://toolkit.tuebingen.mpg.de/marcoil), to obtain candidate genes for further analysis.

Sequence analysis and construction of phylogenetic tree

The basic physicochemical parameters of HSF proteins in cucumber were evaluated by the online tool ExPASy (http://web.expasy.org/compute_pi/). The nucleotide sequences of the searched HSF genes in cucumber and Arabidopsis were translated into amino acid sequences using the BioXM2.6 software program, and the file format was transformed into the FASTA format. Sequence alignment was conducted using the Clustal X software program (Ramu et al.,

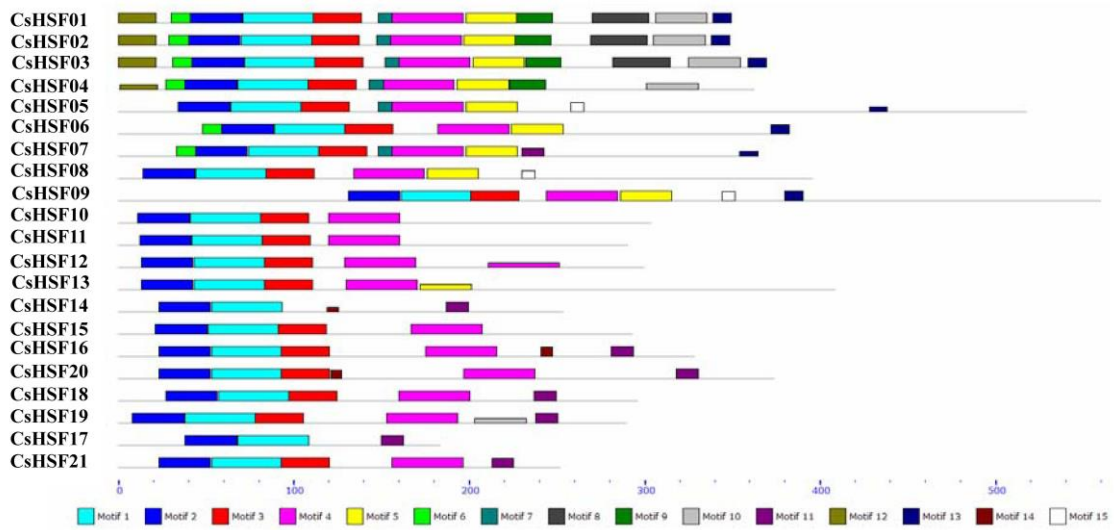

Fig 3 Distribution of conserved motifs in the CsHSF family members. All motifs were identified by MEME using the complete amino acid sequences of 21 cucumber *HSF* genes. Motif sizes are indicated at the bottom of the figure. Different motifs are indicated by different colors numbered 1-15.

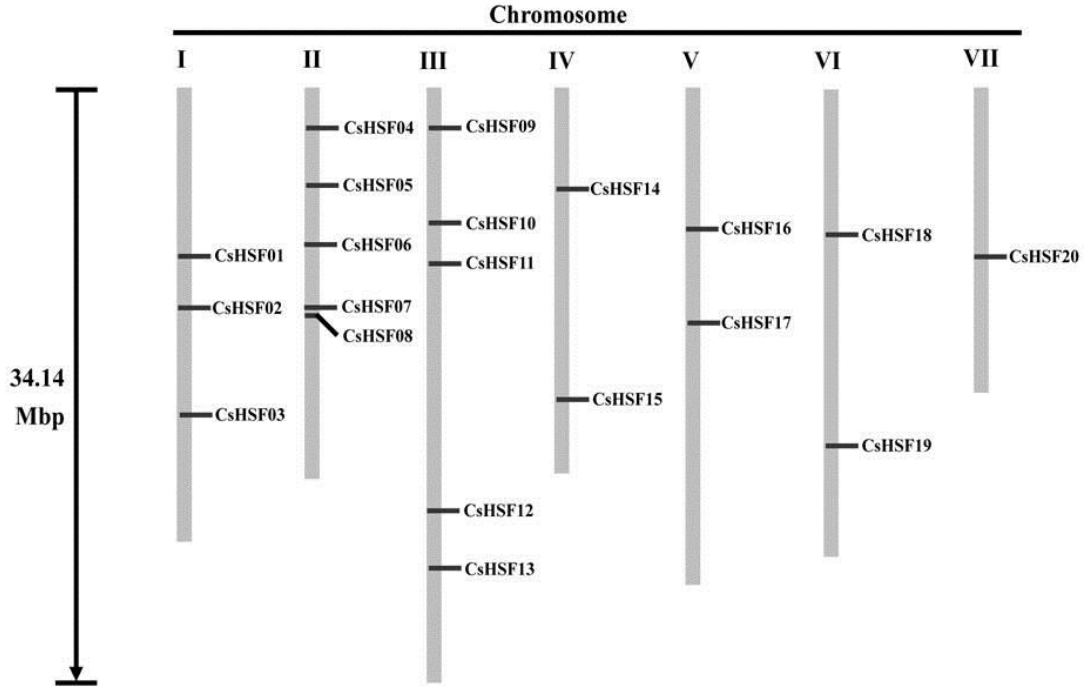

Fig 4. Chromosomal distribution of cucumber Hsf genes. Chromosomal mapping was based on the physical position (Mb) in 7 tomato chromosomes. The chromosome number is indicated at the top of each chromosome. Chromosomal positions of the cucumber Hsf genes are indicated by gene name (assigned in Table 1).

2003), and the file format was further transformed into the format of "CLUSTAL". The phylogenetic tree was constructed with the amino acid sequences of the HSFs by the Neighbor-Joining Method in the MEGA5.0 software program (Tamura et al., 2011). Bootstrapping (1000 replicates) was used to evaluate the degree of support for a particular grouping pattern in the phylogenetic tree and the joints with bootstrap supporting values less than 50% were removed, to demonstrate the lengths of various branches.

Analysis of conserved motifs of HSF in cucumber

The conserved motifs of cucumber HSF proteins were analyzed using the Multiple Em for Motif Elicitation (MEME) tool (http://meme.nbcr.net/meme/cgi-bin/meme.cgi), and the

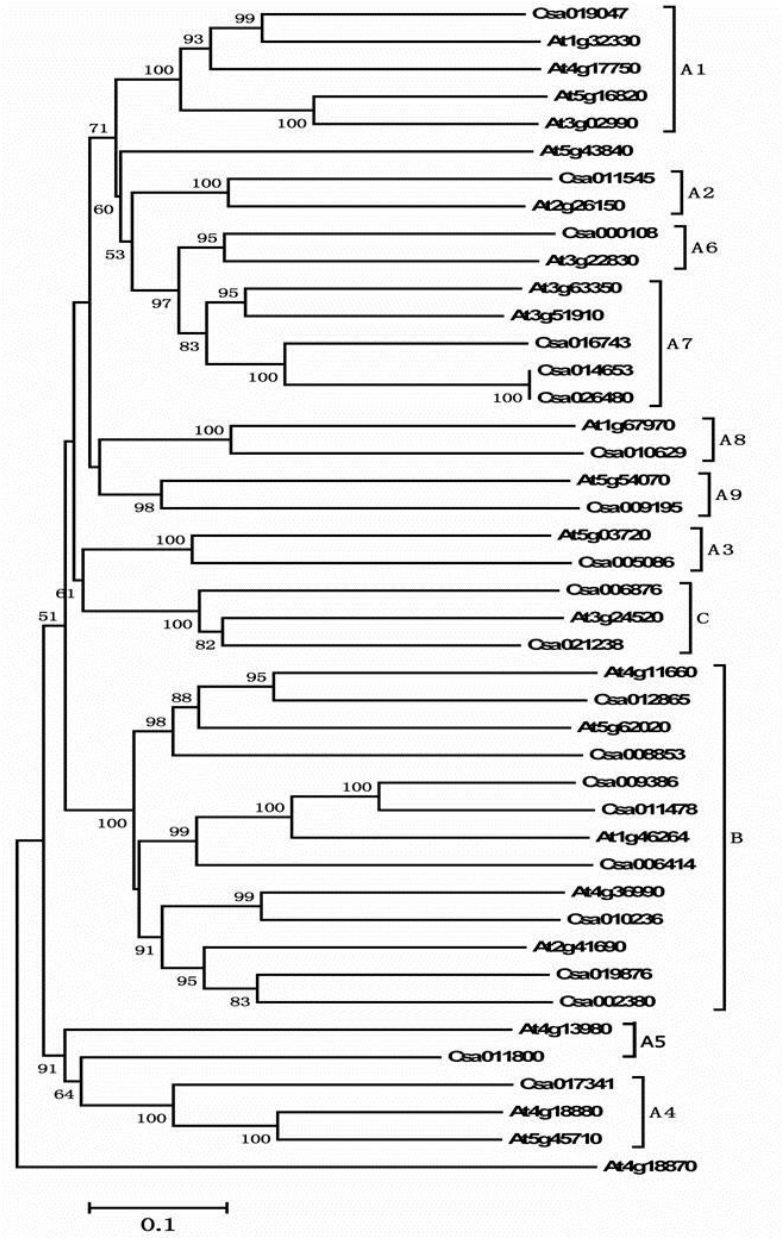

Fig 5. The phylogenetic analysis of all the Hsf genes in Arabidopsis and cucumber. The phylogenetic tree of all the Hsf genes in two species was constructed from a complete alignment of these Hsf proteins by the neighbor-joining method with bootstrap analysis (1, 000 replicates). Arabidopsis and cucumber genes are indicated at the end of branches, respectively. Bootstrapping values are indicated as percentages (when > 50%) along the branches. The resulting eleven clusters of orthologous genes according to the report by (Nover et al., 1996) are shown on the right.

parameters were set as follows: the minimum length of the conserved motif was 6, the maximum length of the conserved motif was 40, and the largest number of the discovered and conserved motifs was 15; other parameters retained their default settings.

Chromosome location of HSFs in cucumber

The cucumber chromosome map was drawn according to the chromosome data from the Cucumber Genome Database (http://cucumber.genomics.org.cn/page/cucumber/index.jsp). The chromosome location image of *HSF* genes was generated using the MapDraw V2.1 software program (Liu et al., 2003).

Conclusion

In this study, a total of 21 *HSF* genes were identified from the sequenced cucumber (*Cucumis sativus* L.) genome. These *HSF* proteins possess highly conserved DNA binding domains and other extensively conserved motifs, and are unevenly distributed in all seven chromosomes. The phylogenetic tree showed that these proteins could be divided into the A, B and C families. Evolutionary analysis indicated that the *HSF* gene family existed before the separation of cucumber and *Arabidopsis thaliana*.

Acknowledgements

The authors would like to thank the Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences for providing the real-time PCR equipment for this study.

References

Ahmad P, Prasad MNV (2012) Environmental adaptations and stress tolerance of plants in the era of climate change, Springer New York

Bai J, Pennill LA, Ning J, Lee SW, Ramalingam J, Webb CA, Zhao B, Sun Q, Nelson JC, Leach JE, Hulbert SH (2002) Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. Genome Res. 12: 1871-1884.

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, von Koskull-Döring P (2004) Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. J Biosci. 29(4): 471-487.

Chung E, Kim KM, Lee JH (2013) Genome-wide analysis and molecular characterization of heat shock transcription factor family in *Glycine max*. J Genet Genomics 3(3): 127-135.

Clos J, Westwood JT, Becker PB, Wilson S, Lambert U, Wu C (1990) Molecular cloning and expression of aheaxa meric *Drosophila* heat stress factor subject to negative regulation. Cell 63: 1085-1097.

Giorno F, Guerriero G , Baric S Mariani C (2012) Heat shock transcriptional factors in *Malus domestica*: identification, classification and expression analysis. BMC Genomics 13: 639

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EA, Wu Y, Guo J, He J, Jia Z, Ren Y, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan, Wu Z, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Liu S, Li J, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Li G, Fang L, Li Y, Liu D, Zheng H, Zhang Y, Qin N, Li Z,

Yang G, Yang S, Bolund L, Kristiansen K, Zheng H, Li S, Zhang X, Yang H, Wang J, Sun R, Zhang B, Jiang S, Wang J, Du Y, Li S (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet. 41: 1275-1281.

Huhel A, Schoffl F (1994) Arabidopsis heat shock factor: isolation and characterization of the gene and the recombinant protein. Plant Mol Biol. 26: 353-362.

Jain M, Tyagi AK, Khurana JP (2006) Genome-wide analysis, evolutionary expansion, and expression of early auxin-responsive SAUR gene family in rice (*Oryza sativa*). Genomics 88: 360-371.

Kimpel, JA, Nagao RT, Goekjian V, Key JL (1990) Regulation of the heat shock response in soybean seedlings. Plant Physiol. 94(3): 988-995.

Kotak S, Larkindale J, Lee U, von Koskull-Döring P, Vierling E, Scharf KD (2007) Complexity of the heat stress response in plants. Curr Opin Plant Biol. 10(3): 310-316.

Kotak S, Port M, Ganguli A, Bicker F, von Koskull-Dêring P (2004) Characterization of C-terminal domains of Arabidopsis heat stress transcription factors (HSF) 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. 39: 98-112.

Lin YX, Jiang HY, Chu ZX, Tang XL, Zhu SW, Cheng BJ (2011) Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. BMC genomics 12(1): 76.

Liu RH and Meng JL (2003) MapDraw: a Microsoft Excel macro for drawing genetic linkage maps based on given genetic linkage data. Heraditas (Beijing), 25 (3): 317-321.

Peteranderl R, Rabenstein M, Shin YK, Liu CW, Wemmer DE, King DS, Nelson HC (1999) Biochemical and biophysical characterization of the trimerization domain from the heat shock transcription factor. Biochemistry 38(12):3559-3569.

Rabindran S K, Giorgi G, Clos J, Wu C (1991) Molecular cloning and expression of a human heat stress factor. Proc Natl Acad Sci. USA 88: 6906-6910.

Ramu C, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thomppson JD (2003) Multiple sequences alignment with the Clustal series of programs. Nucleic Acids Res. 31(13): 3497-3500.

Sarge KD, Zimarino V, Holm K, Wu C, Morimoto RI (1991) Cloning and characterization of two mouse heat stress factors with distinct inducible and constitutive DNA binding ability. Genes Dev. 5: 1902-1911.

Schuetz TJ, Gallo GJ, Sheldon L, Tempst P, Kingston RE (1991) Isolation of a cDNA for HSF2: evidence for two heat stress factor genes in humans. Proc Natl Acad Sci. USA 88: 6911–6915.

Scharf KD, Rose S, Zott W, Schŏffl F, Nover L (1990) 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. 9(13): 4495-4501 .

Scharf KD, Berberich T , Ebersberger I, Nover L (2012) The plant heat stress transcription factor (HSF) family: structure, function and evolution. Biochim Biophys Acta. 1819(2): 104–119.

Sorger PK, Pelham HR (1988) Yeast heat stress factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. Cell 54: 855–864.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28: 2731-2739.

Wiederrecht G, Seto D, Parker CS (1988) Isolation of the gene encoding the *S. cerevisiae* heat stress transcription factor. Cell 54: 841-853.

Yamanouchi U, Yano M, Lin H, Ashikari M, Yamada K (2002) A rice spotted leaf gene, *Spl7*, encodes aheat stress transcription factor protein. Proc Natl Acad Sci USA. 99: 7530- 7535.

Zhang SB, Chen C, Li L, Meng L, Singh J, Jiang N, Deng XW, He ZH, G. Lemaux P (2005) Evolutionary expansion, gene structure, and expression of the rice wall-associated kinase gene family. Plant Physiol. 139: 1107-1124.