Genome-Wide Identification and Expression Analysis of Hsf and Hsp Gene Families in Cucumber (Cucumis Sativus L.)

Xueqian Chen
Shandong Agricultural University
Zhiyuan Wang
Shandong Agricultural University
Rui Tang
Shandong Agricultural University
Lina Wang
Shandong Agricultural University
Chunhua Chen
Shandong Agricultural University
Zhonghai Ren (zhren@sdau.edu.cn)
Shandong Agricultural University https://orcid.org/0000-0002-0759-3668

Research Article

Keywords: Cucumber, Genome-wide, Gene expression, Heat shock protein, Heat shock factor

Posted Date: February 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-189743/v1

Version of Record: A version of this preprint was published at Plant Growth Regulation on August 12th, 2021. See the published version at https://doi.org/10.1007/s10725-021-00739-z.

Abstract

Heat shock proteins (Hsps) are molecular chaperones that participate in plants' response to environmental stresses, including heat stress, and also play an essential role in plant growth and development. *Hsps* expression is monitored and regulated by specific types of transcription factors known as heat shock factors (Hsfs). Although the role of Hsfs and Hsps in stress response has been investigated in some plants, their role is still poorly understood in cucumber (*Cucumis sativus* L.). To reveal the mechanisms of cucumber Hsf and Hsp coping with various stresses, the analyses of cucumber *Hsf* and *Hsp* gene families were conducted using bioinformatics-based methods. A total of 23 Hsfs and 72 Hsps were identified in the cucumber genome (v3.0), and the gene structure and motif composition are relatively conserved in each subfamily. At least 23 pairs of heat shock genes underwent gene duplication in cucumber. A cis-element analysis is implicit that *CsHsfs* and *CsHsps* possessed at least one hormone or stress response cis-element, suggesting that *CsHsf* and *CsHsp* genes could respond to different stress conditions. Heatmaps of the *CsHsf* and *CsHsp* gene families indicated that most heat shock genes were expressed in different tissues and organs. RNA-seq showed that most of the cucumber *Hsf* and *Hsp* genes are differentially expressed upon exposure to biotic and abiotic stresses. These results provide valuable information to clarify the evolutionary relationship between the CsHsf and CsHsp family and to facilitate the functional characterization of the *CsHsf* and *CsHsp* genes in future studies.

Introduction

During growth and development, plants were exposed to a variety of strenuous stresses from the surrounding environment, including not only abiotic stresses such as high temperatures, drought, salinity, cold, and high light intensity but also biotic stresses such as animal attack and pathogen invasion (Nejat and Mantri 2017). Plants, as sessile organisms, could not avoid the damage of adverse environments by changing their position. Therefore, a complex adversity response and regulatory network has evolved at the biochemical, physiological, and molecular levels to adapt to adversity (Shao et al. 2007).

Under the trend of global warming, high temperature has become a major issue for agriculture worldwide (Driedonks et al. 2016). To prevent the negative effects of high temperature on plants, such as water loss, growth reduction, photosynthesis efficiency reduction, some genes are sensitive to heat stress and can respond quickly (Xiong and Ishitani 2006).

Heat shock factors (Hsfs), a specific type of transcription factor, normally exist as inactive proteins and can be induced rapidly to modulate growth responses under high temperature (Scharf et al. 2012). Previous studies have shown that the protein structure of Hsfs is very conserved, in which the N-terminus was mainly composed of DNA-binding domain (DBD) and adjacent bipartite oligomerization domain (HR-A/B or OD), and the C-terminal of Hsfs has nuclear localization signal (NLS), a nuclear output signal (NES) and activation peptide motif (AHA) (Guo et al. 2016; Scharf et al. 1990). Besides, according to the number of amino acid residues inserted between HR-A and HR-B, plant Hsfs can be divided into three main classes, namely: A, B, and C (Guo et al. 2016).

Regulated by Hsfs, heat shock proteins (Hsps) synthesize in large quantities, working as molecular chaperones in both plant and animal species, to prevent incorrect aggregation of proteins when plants subject to heat stress (Feder and Hofmann 1999). *Hsps* were not only transcriptionally regulated by Hsfs but also able to regulate the activity of Hsfs via feedback loop caused by their physical interaction, such as *Hsp70* and *Hsp90* inhibiting *HsfA1* activity under normal conditions to maintaining protein homeostasis (Hahn et al. 2011). Under their apparent molecular weight, plant Hsps are divided into five families: small Hsps (sHsps), Hsp60, Hsp70, Hsp90, and Hsp100 (Pratt and Toft 2003; Wang et al. 2004), which are distributed in many cell components, such as cytoplasm, nucleus, plastids, mitochondria, chloroplasts, endoplasmic reticulum, and peroxisomes, protecting different organelle structures (Jacob et al. 2017).

So far, the functions of *Hsf* and *Hsp* genes in various stress responses have been reported in many plants. In *Arabidopsis thaliana*, it has been proved that *AtHsfA1s* is the central regulator of heat stress regulation and response network, which could induce the expression of many downstream heat shock genes, for example, other *Hsf* subclasses (A2, A3, a7, B1, and B2) (Liu and Charng 2012; Yoshida et al. 2011). *AtHsfA2* not only performs an essential function in heat and osmotic stress response but also exert an important part in the growth and development of plants (Charng et al. 2007; Liu and Charng 2013; Ogawa et al. 2007). *OsMSR3*, coding a small heat shock protein, confers enhanced tolerance to cadmium and copper stresses in Arabidopsis (Cui et al. 2019). *GhHsp24.7*, a mitochondrial small heat shock protein in cotton regulates seed germination by inducing reactive oxygen species (ROS) via thermal sensing (Ma et al. 2019). *CaHsp26* and *CaHsp22.5* of sweet pepper (*Capsicum annuum* L.) play a significant role in the protection of PSII by enhancing photochemical activity and oxidation resistance during chilling stress (Li et al. 2018; Li et al. 2012). In tobacco (*Nicotiana tabacum* L.), potentiation of *Hsp* gene expression increased the resistance exposed to NaCl stress (Li et al. 2014b). Besides, *ZmHsp16.9*, a cytosolic class I small heat shock protein of maize, confers heat tolerance in transgenic tobacco (Sun et al. 2012). *Hsp100/ClpB* located in the chloroplast played an active role in regulating the heat resistance of tomato (*Solanum lycopersicum*) (Yang et al. 2006). Overexpression of *SlHSP70-1* resulted in internode elongation, which is likely to function in the shoot growth of tomato (Vu et al. 2019). *MeHSP90.9* is critical for drought-induced stress resistance in cassava by regulating hydrogen peroxide (H2O2) and abscisic acid (ABA) (Wei et al. 2020). *Hsp90* genes in barley play roles in regulating hypersensitive responses to stripe rust (Pei et al. 2015). Above all, *Hsfs* and *Hsps* are essential that confer stress tolerance and regulate the physiological balance during plant growth and development.

Cucumber (*Cucumis sativus* L.) is a thermophilic species, but is vulnerable to heat stress (Li et al. 2014a). High temperature is one of the main abiotic stress factors affecting cucumber yield and quality in summer cultivation or protected area production (Ding et al. 2016). The completion of the cucumber genome sequencing project provides important information for genetic breeding and biological function identification of related genes (Huang et al. 2009). In this study, based on the Cucumber 'Chinese Long' v3 genome database (Li et al. 2019), we identified 23 *Hsf* and 72 *Hsp* genes in the cucumber genome, including *Hsp20*, *Hsp60*, *Hsp70*, *Hsp90*, and *Hsp100*. Then we provide a comprehensive analysis of *Hsf* and *Hsp* genes in cucumber, including their sequence features, chromosomal locations, phylogenetic

relationships, and dynamic expression patterns in response to heat, NaCl, and downy mildew stresses, which not only help to identify the function of cucumber *Hsf* and *Hsp* genes, but also provide some candidate genes for breeding new stress-resistant cucumber varieties.

Materials And Methods

Identification of Hsf and Hsp genes in the cucumber genome

The cucumber genome, gene, and protein sequences were downloaded from the Cucurbit Genomics Database (http://cucurbitgenomics.org/, Cucumber 'Chinese Long' genome v3). To identify putative Hsf and Hsp members in cucumber, Protein Basic Local Alignment Search Tool (BLASTP) searches were performed with an e-value ≤1e-2 in the Cucurbit Genomics Database using the Published Hsf and Hsp protein sequences from Arabidopsis, rice, and tomato as a query sequence (Agarwal et al. 2001; Chauhan et al. 2011; Guo et al. 2008; Hu et al. 2009; Krishna and Gloor 2001; Lee et al. 2007; Lin et al. 2001; Liu et al. 2014; Sarkar et al. 2009; Sarkar et al. 2013; Scharf et al. 2001; Singh et al. 2010; Swindell et al. 2007; Wang et al. 2014; Yang et al. 2016; Yu et al. 2016). Subsequently, we downloaded the Hsf (PF00447), Hsp20 (PF00011), Hsp60 (PF00118), Hsp70 (PF00012), Hsp90(PF00183), and Hsp100(PF02861) Hidden Markov Model (HMM) configuration files from the Pfam database (http://pfam.xfam.org/). Hsfs and Hsps in the cucumber genome database were searched using HMMER 3.0. With the default parameters, the cutoff value was set to 0.01. Additionally, name search using the words "Hsf", "Hsp20", "Cpn60", "Hsp70", "Hsp90" and "ClpB" as a keyword also applied to retrieve in Cucurbit Genomics Database. Following the removal of the repetitive sequences, all sequences were reserved and submitted to CDD (https://www.ncbi.nlm.nih.gov/cdd/), Pfam, and SMART (http://smart.embl-heidelberg.de/) to confirm the Hsf/ Hsp20/ Hsp60/ Hsp70/ Hsp90/ Hsp100 domain. All of the high-confidence and non-redundant genes were assigned as cucumber *Hsfs* and *Hsps* (*CsHsps* and *CsHsfs*). These *CsHsp* and *CsHsf* genes were named based on their positions on pseudomolecules (Su et al. 2013).

Sequence analysis and structural characterization

Using cNLS-Mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) and NetNES 1.1 (http://www.cbs.dtu.dk/services/NetNES/) to predict the NLS and NES domains of cucumber Hsfs (Kotak et al. 2004; la Cour et al. 2004). Predict the domain of AHA based on the conserved AHA motif sequence (FWxxF/L, F/I/L) (Döring et al. 2000). Protein isoelectric point (pI), number of amino acids (AA), and molecular weight (MW) were calculated using ExPasy (https://web.expasy.org/compute_pi/) (Cai et al. 2016). WoLF PSORT (https://www.genscript.com/wolf-psort.html) was used to predict their protein subcellular localizations (Emanuelsson et al. 2000). The MEME program (Version 5.0.4,http://alternate.meme-suite.org/tools/meme) was used to identify the conserved motifs in the CsHsf and CsHsp protein sequences, with the following parameters: any number of repetitions, the maximum number of motifs: 10, and an optimum motif width of 6 to 200 amino acid residues (Bailey et al. 2009). The structures of all *CsHsf* and *CsHsp* genes were analyzed via the Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/) (Hu et al. 2015).

Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignments of *CsHsf* and *CsHsp* genes were performed by using the ClustalX (version 1.83) program with default parameters (Thompson 1997). The full-length amino acid sequences of Hsfs and Hsps in cucumber, Arabidopsis, rice, and tomato were uploaded to the MEGA v. 7.0 software (Kumar et al. 2016). The ClustalW alignment method was applied to the sequence alignment. Phylogenetic relationships between all *Hsf* and *Hsp* genes were examined using the Neighbour-Joining method with 1,000 bootstrap replicates. The phylogenetic tree codification is obtained from the software with. nwk extension was uploaded to the Evolview (https://evolgenius.info//evolview-v2/#login) and groups for *Hsf* and *Hsp* genes were established in a circular format with different colors.

Chromosomal Localization, Gene Duplication, and Evolutionary Analysis

The chromosomal positions of the *Hsf* and *Hsp* genes in cucumber were based on the Cucurbit Genomics Database. All the chromosome locations of *CsHsf* and *CsHsp* genes were performed with the MapChart software to visualize positions and relative distances of genes on cucumber chromosomes (Voorrips 2002). Multiple Collinearity Scan toolkit (MCScanX) was used to analyze the gene duplication events, with the default settings (Wang et al. 2012). The duplication of the *CsHsf* and *CsHsp* gene was determined according to two criteria: (a) the length of the shorter aligned sequence covered >70% of the longer sequence, and (b) the two aligned sequences shared > 70% amino acid sequence similarity (Yang et al. 2008). In the 100kb chromosomal fragment, two genes separated by less than five intermediate genes are considered to have undergone tandem duplication (Zhang et al. 2018). Codon alignment of duplicated genes was done by MEGA v.7.0 using the ClustalW codon alignment tool, and the homologous (Ks) and non-homologous (Ka) rates of the tandem and segmental duplications of cucumber *Hsf* and *Hsp* genes were calculated using KaKs_Calculator 2.0 (Wang et al. 2010). The divergence time (T) of each duplicated gene pair was calculated as T = Ks / 2r (r = 6.56 × 10 − 9) × 10−6 million years ago (Mya), with Ks being the synonymous substitutions per site and r being the rate of divergence for nuclear genes from plants (Koch et al. 2000). To reveal the synteny relationship between the orthologous *Hsf* and *Hsp* genes of cucumber, Arabidopsis, and rice, syntenic analysis maps were constructed using the Dual Systeny Plotter software (https://github.com/CJ-Chen/TBtools) (Liu et al. 2017).

Cis-Element Prediction and Protein-Protein Interaction Analysis

The upstream sequences (1.5 kb) of the *CsHsf* and *CsHsp* genes-coding sequences were retrieved from the Cucurbit Genomics Database and then submitted to PlantCARE(http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to identify hormone response elements and stress-induced components (Lescot et al. 2002). The hormone response elements include abscisic acid (ABA)-responsive element (ABRE), ethylene (ER)-responsive elements (ERE), gibberellin (GA)-responsive elements (P-box, GARE-motif, and TATC-box), salicylic acid (SA)-responsive elements (TCA-element and SARE), auxin (IAA)-responsive elements (TGA-element, AuxRR-core, and TGA-box) and jasmonate acid (MeJA)-responsive elements (CGTCA-motif and TGACG-motif). The stress-induced components include heat stress element (HSE), low-temperature responsive element (LTR), dehydration-

responsive elements (DRE), drought-responsive element (MBS), defense-responsive element (W box), defense and stress-responsive element (TC-rich repeats), wound-responsive element (WUN-motif), hypoxia-inducible response element (ARE) and anoxic specific inducible response element (GC-motif). Protein-protein interaction (PPI) among CsHsf and CsHsp was detected by STRING ((Version 11.0, https://string-db.org/cgi/input.pl) with the confidence score > 0.7 and an interaction network of CsHsf and CsHsp proteins was drawn by Cytoscape_v3.7.2.

Plant materials and application of stress

The 'Chinese long' 9930□a cucumber typical of northern China, was used for expression analyses. The selected seeds after immersion and imbibition for 12h were seeded in pots containing vermiculite and matrix and then cultivated in a growth chamber in temperature-controlled greenhouses under day/night temperatures of 25/18 ℃, with a 16h/8h photoperiod/dark period. Three-week-old seedlings were subjected to heat treatment at 42℃ and the leaves were collected at 0 (control), 3, and 6 h. For NaCl treatment, the seedlings were subjected to 75 mM NaCl stress for 24 hours. For Downy mildew treatment, the seedlings were inoculated with *Pseudoperonospora cubensis* for 6 and 24 hours. All the collected samples were quickly frozen in liquid nitrogen and stored in a −80 °C refrigerator until RNA isolation.

RNA isolation and RNA-Seq analysis

After the heat treatment, the leaves were immediately collected and frozen in liquid nitrogen, and stored at −80°C before analysis. Three independent biological samples were used in the analysis to ensure the accuracy of the analysis. Use the total RNA extraction kit (TRIzol, B511311, Sangon, China) to separate the total RNA of each sample, and use RNase-free DNase I to remove genomic DNA contamination. The RNA Nano 6000 analysis kit of the Bioanalyzer 2100 System (Agilent Technologies, California, USA) was used to evaluate RNA integrity. The total RNA amount of each sample is 1 μg, which is used as input material for RNA sample preparation. According to the manufacturer's recommendations, use the NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) to generate a sequencing library, and add the index code to the attribute sequence of each sample. The PCR products were purified on the Agilent Bioanalyzer 2100 system (AMPure XP system) and the library quality was evaluated. According to the manufacturer's instructions, useing TruSeq PE Cluster Kit v3-cBot-HS (Illumia) to cluster the index code samples on the cBot cluster generation system. After the clusters were generated, the library preparation was sequenced on the Illumina Novaseq platform, and 150 bp paired-end reads were generated. HISAT2 (version 2.0) was used to map clean reads to the cucumber reference genome (v3, http://cucurbitgenomics.org/organism/20) using default parameters. StringTie (v1.3.3b) uses a reference-based method to assemble the mapped reads for each sample. featureCounts v1.5.0-p3 was used to calculate the number of reads corresponding to each gene. Then calculate the FPKM of each gene according to the length of each gene, and count the reads of the gene. Differentially expressed genes (DEGs) were identified by the DESeq2 R package (1.16.1). The transcript abundance of cucumber Hsf and Hsp genes was calculated based on the number of fragments per kilobase (FPKM) in exon patterns. The RNA-seq data of *CsHsf* and *CsHsp* genes in different tissues and responses to NaCl and

downy mildew stresses were obtained from the Cucurbit Genomics Database (http://cucurbitgenomics.org/rnaseq/home, PRJNA80169, PRJNA437579, and PRJNA388584). Based on the converted data of log2 (FPKM+1) value, heatmaps were generated with HemI1.0.

Results

Identification and analysis of *Hsf* and *Hsp* genes in the cucumber genome

After HMM analysis, BLASTP and keyword search against the Cucumber Genomic Database (Version 3.0), a total of 23 *Hsf* and 72 *Hsp* genes in cucumber were identified, among which 33, 15, 12, 6 and 6 genes belonged to *Hsp20*, *Hsp60*, *Hsp70*, *Hsp90* and *Hsp100* families, respectively. These *Hsf* and *Hsp* genes from *Cucumis Sativus* were abbreviated as *CsHsfs* and *CsHsps*. These *CsHsp* and *CsHsf* genes were also named according to their chromosome orders. Detailed information of each *CsHsf* and *CsHsp* genes was shown in Table S1, including the gene name, gene ID, chromosome location, length of the open reading frame, number of exons, number of amino acids, molecular weight, protein isoelectric point, and subcellular localization prediction.

Structure and conserved motifs of CsHsf and CsHsp

According to the exon-intron arrangement of the coding sequence, the structural diversity in *CsHsf* and *CsHsp* genes was compared. In terms of intron number, intron, and exon length, the most closely related members of the same *Hsf* or *Hsp* subfamily share a similar gene structure (Fig. 1). Most *CsHsfs* have only one intron, and *CsHsf-11*, *CsHsf-4* and *CsHsf-18* have two introns, while *CsHsf-17* has 11 introns, which is quite different from other *CsHsfs* (Fig. 1A). In the *CsHsp20* family,13 (39%) *CsHsp20s* are intronless, 17 (52%) have only one intron, 2 (6%) have two introns, and only *CsHsp20-32* has five introns (Fig. 1B). The number of introns in different members of the *CsHsp60* family varied widely (0 to 17 introns), with *CsHsp60-8*, *CsHsp60-14*, and *CsHsp60-15* which are most likely located in the mitochondria, have the largest number of introns (16 or 17). It was worth noting that *CsHsp60-1* has no intron which is different from other *CsHsp60* members (Fig. 1C). Except for *CsHsp70-3*, all *CsHsp70s* contained introns in the gene sequences (Fig. 1D). The *CsHsp70s* (*CsHsp70-9*, *CsHsp70-6*, *CsHsp70-10*, and *CsHsp70-1*) belonging to group I are mainly located in the cytoplasm and have only one intron. The *CsHsp70-7* has a relatively complex gene structure and contained 13 introns. The *CsHsp90s* located in different subcellular positions have different intron numbers (Fig. 1E). Among them, *CsHsp90-1*, *CsHsp90-2*, and *CsHsp90-4* which are mainly located in the cytoplasm have fewer introns (2 to 3), while *CsHsp90-6* and *CsHsp90-5* which mainly located in the mitochondria or chloroplast have more introns (18 to 19). Except for *CsHsp100-3* with 16 introns, the number of introns of other *CsHsp100s* is less than 10, and only *CsHsp100-3* is mainly located in the nucleus (Fig. 1F). In conclusion, it was found that the number of introns in *CsHsps* is closely related to their subcellular localization and evolutionary relationship.

Then, we conducted a prediction of conserved motifs shared among the related proteins in each subfamily using the Multiple Expectation Maximization for Motif Elicitation (MEME) and identified 10

putative motifs in each family. In general, most closely related members of the phylogenetic trees in different subfamilies had similar motifs (Fig.1). Details of these motifs were shown in Table S2.

Like Hsfs in other plants, the protein structure of CsHsfs is also very conserved. Based on analyses of Pfam, CDD, and SMART, we found all the CsHsfs proteins contain the DBD domain composed of motif 1 and motif 2. Multiple alignments of the CsHsf protein sequences revealed a highly conserved DBD domain existed in all CsHsfs (Fig. 2A). The HR-A/B region was an essential domain in Hsfs, which was characterized by the predicted coiled-coil structure (Guo et al. 2016). The HR-A/B domain in CsHsfs was composed of two typical motifs (motif 3 and motif 4) (Fig.1). Besides, 21 amino acids were inserted between the HR-A and HR-B regions in class A CsHsf proteins and an insertion of 7 amino acids was found in class C CsHsf proteins (Fig. 2B). However, there was no insertion between the HR-A and HR-B regions in Class B CsHsf proteins. AHA motif (motif 6) was distinctive in the great majority of class A CsHsf proteins, and 8 (66.7%) CsHsfA proteins have AHA motif. NLS and NES were of vital importance for intercellular distribution and interactions of Hsf proteins in the nucleus and the cytoplasm (Heerklotz et al. 2001). NLS of CsHsfs was predicted using the cNLS Mapper software and NES was predicted by NetNES. 17 (73%) CsHsfs contained NLS domains, and 12 (52%) CsHsfs contained NLS domains (Table S1-1).

All the CsHsp20s have a highly conserved α -crystallin domain (ACD) at the C-terminus. Multiple sequence alignment analysis and sequence logo showed that the ACD domain consisted of two conserved regions, a conserved region I and conserved region II (Fig. 3). Besides, the consensus region I contained motif 3, and the consensus region II consisted of motif 1, while motif 6 and motif 4 were inserted between the consensus region I and the consensus region II of the ACD domain (Fig. 1B).

Phylogenetic analysis of CsHsf and CsHsp

We further investigated the evolutionary relationships of Hsf and Hsp in cucumber, Arabidopsis, tomato, and rice, and based on the full-length amino acid sequences, we generated an unrooted phylogenetic tree by MEGA 7.0 using the neighbor-joining method. The *CsHsf* gene family could be divided into three classes: Class A (12 genes), class B (9 genes), and class C (2 genes), and each class can be further subdivided into subclasses according to the branches (Fig. S1). Besides, the subfamily of each *CsHsp* gene family could be allocated according to the predicted subcellular location of the protein. For example, 130 *Hsp20s* from cucumber, Arabidopsis, rice, and tomato were divided into 17 different subfamilies, including CI (Cytosol I) (42 genes), CII (7 genes), CIII (3 genes), CIV (3 genes), CV (4 genes), CVI (4 genes), CVII (3 genes), CVIII (10 genes), CIX (5 genes), CX (9 genes), CXI (5 genes), MI (mitochondria I) (6 genes), MII (7 genes), MIII (4 genes), P (plastids) (7 genes), Po (peroxisome) (3 genes) and ER (endoplasmic reticulum) (8 genes) (Fig. S2). The *Hsp60* family consisted of four subfamilies. According to the prediction, 24 and 6 *Hsp60s* of the I and IV subfamilies were mainly located in the cytoplasm, 9 *Hsp60s* belonging to the II subfamily were mainly located in mitochondria, and 15 *Hsp60s* of the III subfamily were mainly located in the chloroplast (Fig. S3). Besides, 81 *Hsp70s* were divided into seven groups (I to VII), among which the I to V groups were clustered in the DnaK subfamily, and the VI and VII groups were

clustered in the HSP110 / SSE subfamily (Fig. S4). The group I contained four *CsHsp70s*, which was the largest group, and these members were mainly located in the cytoplasm. Only *CsHsp70-8* belongs to group II, which was involved in the metabolism of the endoplasmic reticulum. Members (*CsHsp70-2* and *CsHsp70-11*) of group III functioned mainly in the chloroplast and only *CsHsp70-5* belonged to group IV, which participate in metabolism occurring in the mitochondrion. *CsHsp70-3* belonged to group V and was widely distributed in various subcellular regions (cytoplasm, chloroplast, mitochondria, plastid, and unknown regions). The *Hsp90* family could be divided into five classes, among which the members of class I (*CsHsp90-1* and *CsHsp90-2*) and II (*CsHsp90-4*) were mainly effective in the cytoplasm (Fig. S5). According to subcellular localization prediction, members of class III (*CsHsp90-3*), IV (*CsHsp90-6*), and V (*CsHsp90-5*) played a role in the endoplasmic reticulum, mitochondria, and chloroplast, respectively. The *Hsp100* family consisted of four groups, in which members of groups I (*CsHsp100-5*) and IV (*CsHsp100-1* and *CsHsp100-4*) produced a marked effect in chloroplasts, and members of groups II (*CsHsp100-6*) and III (*CsHsp100-2* and *CsHsp100-3*) made a contribution to the homeostasis of mitochondria and cytoplasm, respectively (Fig. S6). All phylogenetic tree analysis showed that there was a common ancestor before monocotyledons and dicotyledons differentiated. Although the role of most heat shock genes in cucumber is yet to be elucidated, *Hsf* and *Hsp* genes with conserved functions in different plants may show a tendency to aggregate to the same subgroup and may have a recent common evolutionary origin.

Chromosomal location and synteny analysis of *CsHsf* and *CsHsp* genes

According to the cucumber genome database, 23 *CsHsf* and 72 *CsHsp* genes were located on 7 chromosomes (Chr) (Fig. 4). Although *CsHsf* and *CsHsp* genes were contained on each of the 7 chromosomes of cucumber, the distribution appeared to uneven. A relatively small density of *CsHsf* and *CsHsp* genes was found on chromosome 4 (10 genes), 6 (9 genes), and 7 (6 genes), while more *CsHsf* and *CsHsp* genes were located on chromosome 1 (22 genes), 2 (12 genes), 3 (18 genes) and 5 (18 genes), and most of *CsHsps* were distributed at both ends of chromosomes.

During the process of plant evolution, gene duplication, especially tandem and segmental duplication events, was the main mechanism for the expansion of gene families, made great contributions to the diversity of gene families (Kotak et al. 2004; Liu et al. 2012). In the analysis of duplication events of *CsHsf* and *CsHsp* genes, only *CsHsp20* genes were identified for tandem duplication. Among the 33 *CsHsp20* genes, 16 (49%) *CsHsp20* genes had a tandem duplication event, resulting in the formation of 6 tandem duplication clusters (Fig. 4). On chromosomes 1 and 3, a total of three tandem duplication clusters were composed of three different genes in pairs (*CsHsp20-7/CsHsp20-8*, *CsHsp20-9/CsHsp20-10* and *CsHsp20-14/CsHsp20-15*). In addition, two groups of tandem duplicated genes were located on chromosome 1 and 5, each of which included three genes (*CsHsp20-3/CsHsp20-4/CsHsp20-5* and *CsHsp20-21/CsHsp20-22/CsHsp20-23*). Only one tandem duplication cluster was composed of four similar genes (*CsHsp20-24/CsHsp20-25/CsHsp20-26/CsHsp20-27*), which were present on chromosome 5. The above results showed that tandem duplications greatly promoted the expansion of the *CsHsp20* gene family. In addition to tandem duplication events, using BlastP and MCScanX, we also identified 13

segmental duplication events including 10 *CsHsf* and 14 *CsHsp* genes, all of which improved the diversity of heat shock genes in cucumber (Fig. 4).

Furthermore, to analyze the selection of the above-duplicated gene pairs, the non-synonymous to synonymous substitution ratios (Ka/Ks) were calculated (Table 1). Ka and Ks values were worthy to analyze the selective pressure on a protein-encoding gene as well as to estimate the approximate date of duplication events. Ka/Ks ratio = 1 was commonly used to identify genes under the neutral mutation or no selection, and Ka/Ks >1 indicated the genes evolved under positive selection, while Ka/Ks <1 indicated the negative purifying selection. In this study, the Ka/Ks values of 22 pairs (96%) of duplicated genes were < 1, indicating that they had experienced strong purifying selection, and only one pair of tandem duplicated genes had a Ka/Ks ratio > 1, suggesting that these genes may have experienced positive selection. The Ks values of these duplicated gene pairs ranged from 0.2153 to 6.1528, corresponding to divergence times of 16.41 to 468.96 Mya (Table 1).

To further infer the phylogenetic mechanism of the *CsHsf* and *CsHsp* families, we performed a synteny analysis of the heat shock genes in cucumber, Arabidopsis, and rice (Fig. 5). A total of 52 pairs of heat shock genes showed the synonymous relationship between cucumber and Arabidopsis, followed by rice was 33. Many cucumber heat shock genes were homologous to both Arabidopsis and rice, and most of the homologous genes had Ka / Ks <1 (Table S3), suggesting that these genes were essential in plant evolution and contributed greatly to maintaining the function of heat shock genes.

Analysis of putative cis-acting elements in the promoters of *CsHsf* and *CsHsp* genes

To identify potential cis-acting elements located on the promoter regions of *CsHsf* and *CsHsp* genes, 1500 bp upstream sequences from translational start sites extracted from the cucumber genome database, were submitted to the PlantCARE database. As shown in Table 2, we analyzed 12 hormone response elements and 9 stress-induced components. Among the 95 genes, 60, 34, 55, 36, 50 and 38 (63%,36%,58%,38%,53% and 40%) genes had at least one type of abscisic acid (ABA) -responsive element, auxin (IAA) -responsive elements, ethylene (ER) -responsive elements, gibberellin (GA) -responsive elements, Jasmonic acid (MeJA) -responsive elements and salicylic acid (SA) -responsive elements, respectively (Fig. 6). For stress-induced components, one or more ARE existed in 14 (61%) *CsHsf* and 61 (87%) *CsHsp* genes, which involved in Hypoxia-inducible response. What's more, heat shock element (HSE) was found in 12 (52%) *CsHsf* and 40 (56%) *CsHsp* genes, and WUN-motif was presented in 9 (39%) *CsHsf* and 31 (43%) *CsHsp* genes, participating in wound response. Other elements were also predicted, such as MBS, LTR, TC-rich repeats, and W-box, known to function as stress-induced components in *CsHsf* and *CsHsp* genes, were effective at variable positions and can effectively respond to drought, low temperature, and biotic stress response, while GC-motif only existed in 4 *CsHsp* genes which were useful in anoxic specific inducible response (Fig. 6). The above analysis of cis-elements showed that *CsHsf* and *CsHsp* genes could respond rapidly under different stress conditions, maintaining physiological and metabolic balance, reducing the damage caused by unfavorable environments, to promote the normal growth of cucumber.

Expression patterns of *CsHsf* and *CsHsp* genes in different tissues

Cucumber Illumina RNA-seq data were obtained from the Cucurbit Genomics Database (http://cucurbitgenomics.org/rnaseq/home). Using the RNA-seq data, we analyze the expression patterns of *CsHsf* and *CsHsp* genes in different tissues including root, stem, leaf, male flower, female flower, ovary(unexpanded), expanded ovary (fertilized), expanded ovary (unfertilized), and tendril (Fig. 7). Except for that *CsHsp20-6*, *CsHsp20-24*, *CsHsp20-26,* and *CsHsp20-30* were almost not expressed in any tissue or organ, most of the *CsHsf* and *CsHsp* genes were expressed in at least one tissue. Almost all of the *CsHsp60s* had low transcript levels in male flowers. Some genes had high levels in tendril, such as *CsHsf-23*, *CsHsp20-4*, *CsHsp20-5* and *CsHsp70-10*. The wide distribution of *CsHsf* and *CsHsp* genes in various tissues ensures the normal morphology of cucumber tissues under stress.

Expression profiles of *CsHsf* and *CsHsp* genes in response to abiotic and biotic stresses

To explore the responses of *CsHsf* and *CsHsp* genes to biotic and abiotic stresses, the expression patterns of *CsHsf* and *CsHsp* genes in response to stresses were investigated using RNA-seq data. Our RNA-seq data were used to analyze the response of *CsHsf* and *CsHsp* genes to heat stress, and the RNA-seq data for *CsHsf* and *CsHsp* genes in response to NaCl and downy mildew stresses were obtained from the Cucurbit Genomics Database (http://cucurbitgenomics.org/rnaseq/home). According to the log2 (FPKM+ 1) value, the expression of many *CsHsf* and *CsHsp* genes was increased after heat treatment (Fig. 8). Interestingly, *CsHsf* and *CsHsp* genes which were up-regulated under heat stress were usually also up-regulated under NaCl treatment, indicating that cucumbers have similarities in response to heat and salt stress. Compared with the control, after 3 hours of heat treatment, the expression of *CsHsf -7* belonging to *HsfA2* increased sharply, indicating that it was a very sensitive receptor during the heat stress response in cucumber. Previous studies have shown that *HsfA1* is a master regulator that triggers the thermal response, leading to the acquired thermotolerance in tomatoes and Arabidopsis (Mishra 2002; Yoshida et al. 2011), but the regulation of *CsHsf-9* belonging to *HsfA1* was not significant under heat stress in this study (Fig. 8A). Compared with the control, the expression level of most *CsHsp20s* increased significantly in the first 3 hours after heat treatment and then decreased gradually in the next 3 hours (Fig. 8B). After heat or NaCl treatment, the expression level of many *CsHsp60s* decreased, and their expression changed were not so obvious compared with other heat shock proteins (Fig. 8C). The infection of downy mildew reduced the expression level of most heat shock genes in cucumber, which may be related to the immune response of cucumber to pathogens. In conclusion, when cucumber is subjected to biotic and abiotic stresses, *CsHsfs* and *CsHsps* genes will respond rapidly, forming a complex reaction regulation network to obtain resistance.

Potential protein-protein interaction between CsHsf and CsHsp

To further explore the possible protein-protein interaction between CsHsf and CsHsp, an interaction network was constructed using the STRING program. As one of the most important heat shock transcription factors, CsHsf7 had potential interactions with CsHsp20s, CsHsp70s, CsHsp90s, and CsHsp100s, implying that CsHsfA2 is an activator of the downstream CsHsps and acts as a key regulator

to improving adaptability to various stresses for cucumber (Fig. 9). According to predictions, there is a strong interaction between CsHsf9 and many CsHp70s (CsHsp70-6, CsHsp70-9, and CsHsp70-10), reflecting previous studies that Hsp70 and Hsp90 interact with HsfA1 under normal conditions and inhibit HsfA1 activity (Hahn et al. 2011; Ohama et al. 2017). In addition to the potential regulatory relationship between CsHsf and CsHsp, we have also detected many other interactions among CsHsps of different subfamilies, such as interactions among CsHsp60, CsHsp70, and CsHsp90, suggesting that the CsHsps of different subfamilies might also be activated or inhibited through interactions with each other when cucumber was subjected to various stresses.

Discussion

The high temperature usually causes changes in the expression of related genes in organisms (Ohama et al. 2017). Among them, Hsp is a type of conserved protein, which is synthesized rapidly by organisms under heat stress, and was first found in the salivary glands of *Drosophila melanogaster* (Tissiéres et al. 1974). Studies have shown that when plants are subjected to heat stress, plants can develop tolerance to lethal high temperatures, of which *Hsps* play an important role. In addition, low temperature, drought, heavy metals, salinization, and ABA can induce plants to produce *Hsp* (Wang et al. 2004). The expression of *Hsps* is regulated by heat shock factors (Hsfs), which are the central regulators of *Hsps* expression and participates in heat stress response (Guo et al. 2016).

However, the study on the *Hsf* and *Hsp* gene family in cucumber is still lacking. This study identified 95 heat shock genes in cucumber, including 23 *CsHsfs* and 72 *CsHsps* (33 *CsHsp20s*, 15 *CsHsp60s*, 12 *CsHsp70s*, 6 *CsHsp90s*, 6 *CsHsp100s*), and then analyzed their gene structure, conserved motifs, phylogenetic relationships, chromosome distribution, duplication events, cis-acting elements, expression patterns under stress and potential protein-protein interaction among *CsHsfs* and *CsHsps*, which will stand in comprehension the functional divergence of *Hsf* and *Hsp* genes in cucumber.

Phylogenetic analysis is often used to provide insights into the evolutionary relationships of species and to help identify orthologs between species and paralogs within species. Based on the full-length protein sequences of Arabidopsis, rice, tomato, and cucumber, we constructed unrooted phylogenetic trees of Hsp and Hsf, respectively. Though the total number of *Hsf* genes was similar to that of Arabidopsis and tomato (Guo et al. 2008), the members of some specific *Hsf* subclasses in cucumber were different from the other two species. For instance, the number of subclass *HsfA1* members in cucumber was less, only *CsHsf9* belong to the *HsfA1* subclass, but the number in subclass *HsfA4* was more than in tomato and Arabidopsis (Fig. S1). Most of the *CsHsp20* genes were classified as nucleocytoplasmic subfamilies (Fig.S2), which are similar to the distribution in Arabidopsis, rice, and tomato (Sarkar et al. 2009; Scharf et al. 2001; Yu et al. 2016). Among these subfamilies, CI was the largest subfamily, containing 14 *CsHsp20* genes. Since the protein is mainly transported to the cytoplasm (Nelson et al. 2014), we speculate that cytoplasm may be the main place for Hsp20 protein as a molecular chaperone to interact with the target proteins, regulating the folding, positioning, accumulation, and degradation of protein molecules. *Hsp70* gene family was divided into 7 groups, and each group contains *Hsp70s* from Arabidopsis, rice, and

cucumber, indicating that the *Hsp70* family genes occurred in the early stage of plant evolution, earlier than the divergence of dicots and monocots (Fig. S4).

The structure of genes determines their function. In terms of intron number, intron, and exon length, the most closely related members of the same *Hsf* or *Hsp* subfamily have similar gene structures (Fig. 1). The number of introns usually affects the activity of gene transcription regulation (Le Hir et al. 2003). Some studies suggest that under various stress conditions, the splicing mechanism of RNA is often disturbed, and transcripts without introns can be transferred from the nucleus to the cytoplasm rapidly without splicing, which improves the response efficiency of plants, thus leading to the tendency to select intronless genes during evolution (Le Hir et al. 2003). The fewer introns in stress genes, the more sensitive they are, leading to the stronger adaptability of plants to various developmental conditions and environmental stimuli (Chung et al. 2006). Compared with other *Hsp* families, the *CsHsp20* family genes were shorter in length and had fewer introns, 39% of *CsHsp20s* were intronless (Fig. 1), and most of them were highly induced under high-temperature stress (Fig. 8), which may prove the above standpoints.

The expansion of gene families in angiosperms may be the result of gene duplication and whole-genome duplication (WGD) at different time points in the evolutionary process (Liu et al. 2012). Like most gene families, *Hsf* and *Hsp* seem to have undergone a complex evolutionary process (Swindell et al. 2007). The diversity of *Hsf* and *Hsp* genes in cucumber may be the result of multiple gene duplication events, which including multiple segmental and tandem duplications. In this research, 23 *CsHsf* and 72 *CsHsp* genes were unevenly distributed on 7 cucumber chromosomes, and the analysis showed that 10 *CsHsf* and 23 *CsHsp* genes were duplicated in the Cucumber genome, forming 13 segmentally duplicated gene pairs and 5 tandem duplicated gene groups (Fig. 4), indicating that both tandem and segmental duplications are conducive to the evolution of *Hsf* and *Hsp* genes in cucumber. Although the *Hsp* gene identified in cucumber was relatively fewer than that in rice (Hu et al. 2009), all members of the *Hsp* gene family existed in the Cucumber genome, and a total of 52 pairs of heat shock genes showed a syntenic relationship between cucumber and Arabidopsis, followed by rice (33) (Fig. 5), which shows that the evolutionary process of heat stress response (HSR) system is conserved in composition but differentiated in duplicated numbers. Many cucumber heat shock genes are homologous to both Arabidopsis and rice, indicating that these genes play an important role in plant evolution. The Ka / Ks value of most duplicated gene pairs was less than one, indicating that most of them undergone purifying selection (Table 1 and Table S3), which might be the key point for maintaining the conserved structure of *Hsf* and *Hsp* genes throughout the evolution. Similar expression patterns under heat stress were found within the tandem duplicated gene groups (Fig. 8), indicating that tandem duplicated *CsHsp20* genes with similar structure have functional redundancy, reflecting shared induction mechanisms.

A large of studies have shown that *Hsf* and *Hsp* genes produced a marked effect in response to a variety of abiotic stresses (Guo et al. 2016; Scharf et al. 2012). In this study, we analyzed the dynamic expression levels of *CsHsf* and *CsHsp* genes under heat, NaCl, and downy mildew stresses by RNA-seq (Fig. 8). We have observed that most *CsHsf* and *CsHsp* genes were sensitive to heat stress, but not all. Interestingly, *CsHsf* and *CsHsp* genes which were up-regulated under heat stress were usually also up-regulated under

NaCl treatment, indicating that cucumbers have similarities in response to heat and salt stress. The results showed that the expression of most heat shock genes was up-regulated, while some *CsHsp60* genes were down-regulated after heat treatments, which may be related to mitochondrial activity and programmed cell death (PCD) (Rikhvanov et al. 2007). Different *Hsps* were induced to different levels under heat stress, indicating that *CsHsps* in different subfamilies may have distinct regulatory properties. In addition, the analysis of cis-elements (Fig. 6) showed that hat *CsHsf* and *CsHsp* genes had multiple stress response components, and potential protein-protein interaction (Fig. 9) suggested that a complex regulatory network could be formed to respond to different stress conditions through interactions, thereby improving the opportunity for cucumber to escape or better cope with the damaging effects of adverse environmental conditions.

Conclusion

Based on bioinformatics methods, we identified 95 members of cucumber *Hsf* and *Hsp* (including *Hsp20*, *Hsp60*, *Hsp70*, *Hsp90*, and *Hsp100*) gene families in this study, and then analyzed their gene structures, conserved motifs, phylogenetic relationships, chromosome distribution, duplication events, cis-acting elements, and potential protein-protein interaction. Also, the expression patterns of these heat shock genes in heat stress were also determined by RNA-seq methods. At least 23 pairs of heat shock genes had been duplicated in cucumber, and the motif composition and gene structure in each subfamily share a high similarity, showing that the functions of most *Hsfs* and *Hsps* might be conserved during evolution. Most of the *CsHsf* and *CsHsp* genes were expressed in one tissue at least, and some *CsHsf* and *CsHsp* genes responded obviously to heat, NaCl, and downy mildew stresses, and were up-regulated or down-regulated significantly. Combined with protein interaction analysis, we concluded that *CsHsf* and *CsHsp* formed a complex regulatory network through activation or inhibition mechanism to improve the heat tolerance of cucumber. This study provided comprehensive information on the *Hsf* and *Hsp* gene families in cucumber and will aid in further research on the functions of *CsHsf* and *CsHsp* genes under biotic and abiotic stresses.

Declarations

Funding

This work was supported by funding from the National Natural Science Foundation of China (31872950 and 31672170), the Natural Science Foundation of Shandong Province (JQ201309), the Shandong “Double Tops” Program (SYL2017YSTD06), and the ‘Taishan Scholar’ Foundation of the People’s Government of Shandong Province (ts20130932).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

ZR conceived the project. ZR, CC, XC, and ZW designed the experiments. XC, ZW, RT, and LW performed the experiments and analyzed the data. XC and ZW wrote the manuscript.

References

1. Agarwal M, Katiyar-Agarwal S, Sahi C, Gallie DR, Grover A (2001) &cestflwr; Arabidopsis thaliana Hsp100 proteins: kith and kin Cell Stress & Chaperones 6:219-224 doi:10.1379/1466-1268(2001)006<0219:Athpka>2.0.Co;2
2. Bailey TL et al. (2009) MEME SUITE: tools for motif discovery and searching Nucleic Acids Res 37:W202-208 doi:10.1093/nar/gkp335
3. Cai B, Li Q, Xu Y, Yang L, Bi H, Ai X (2016) Genome-wide analysis of the fructose 1,6-bisphosphate aldolase (FBA) gene family and functional characterization of FBA7 in tomato Plant Physiol Biochem 108:251-265 doi:10.1016/j.plaphy.2016.07.019
4. Charng YY, Liu HC, Liu NY, Chi WT, Wang CN, Chang SH, Wang TT (2007) A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermotolerance in Arabidopsis Plant Physiol 143:251-262 doi:10.1104/pp.106.091322
5. Chauhan H, Khurana N, Agarwal P, Khurana P (2011) Heat shock factors in rice (Oryza sativa L.): genome-wide expression analysis during reproductive development and abiotic stress Mol Genet Genomics 286:171-187 doi:10.1007/s00438-011-0638-8
6. Chung BY, Simons C, Firth AE, Brown CM, Hellens RPJBg (2006) Effect of 5'UTR introns on gene expression in Arabidopsis thaliana 7:1-13
7. Cui Y et al. (2019) OsMSR3, a Small Heat Shock Protein, Confers Enhanced Tolerance to Copper Stress in Arabidopsis thaliana Int J Mol Sci 20:6096 doi:10.3390/ijms20236096
8. Ding X et al. (2016) Effects of Heat Shock on Photosynthetic Properties, Antioxidant Enzyme Activity, and Downy Mildew of Cucumber (Cucumis sativus L.) PLoS One 11:e0152429 doi:10.1371/journal.pone.0152429
9. Döring P, Treuter E, Kistner C, Lyck R, Chen A, Nover L (2000) The Role of AHA Motifs in the Activator Function of Tomato Heat Stress Transcription Factors HsfA1 and HsfA2 The Plant Cell 12:265-278 doi:10.1105/tpc.12.2.265
10. Driedonks N, Rieu I, Vriezen WH (2016) Breeding for plant heat tolerance at vegetative and reproductive stages Plant Reproduction 29:67-79 doi:10.1007/s00497-016-0275-9
11. Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence J Mol Biol 300:1005-1016 doi:10.1006/jmbi.2000.3903
12. Feder ME, Hofmann GE (1999) Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology Annu Rev Physiol 61:243-282 doi:10.1146/annurev.physiol.61.1.243

13. Guo J et al. (2008) Genome-wide analysis of heat shock transcription factor families in rice and Arabidopsis J Genet Genomics 35:105-118 doi:10.1016/S1673-8527(08)60016-8
14. Guo M, Liu JH, Ma X, Luo DX, Gong ZH, Lu MH (2016) The Plant Heat Stress Transcription Factors (HSFs): Structure, Regulation, and Function in Response to Abiotic Stresses Front Plant Sci 7:114 doi:10.3389/fpls.2016.00114
15. Hahn A, Bublak D, Schleiff E, Scharf KD (2011) Crosstalk between Hsp90 and Hsp70 chaperones and heat stress transcription factors in tomato Plant Cell 23:741-755 doi:10.1105/tpc.110.076018
16. Heerklotz D, Doring P, Bonzelius F, Winkelhaus S, Nover L (2001) The balance of nuclear import and export determines the intracellular distribution and function of tomato heat stress transcription factor HsfA2 Mol Cell Biol 21:1759-1768 doi:10.1128/MCB.21.5.1759-1768.2001
17. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server Bioinformatics 31:1296-1297 doi:10.1093/bioinformatics/btu817
18. Hu W, Hu G, Han B (2009) Genome-wide survey and expression profiling of heat shock proteins and heat shock factors revealed overlapped and stress specific response under abiotic stresses in rice Plant Sci 176:583-590 doi:10.1016/j.plantsci.2009.01.016
19. Huang S et al. (2009) The genome of the cucumber, Cucumis sativus L Nat Genet 41:1275-1281 doi:10.1038/ng.475
20. Jacob P, Hirt H, Bendahmane A (2017) The heat-shock protein/chaperone network and multiple stress resistance Plant Biotechnol J 15:405-414 doi:10.1111/pbi.12659
21. Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae) Mol Biol Evol 17:1483-1498 doi:10.1093/oxfordjournals.molbev.a026248
22. Kotak S, Port M, Ganguli A, Bicker F, Von Koskull-Döring PJTPJ (2004) 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 39:98-112
23. Krishna P, Gloor G (2001) The Hsp90 family of proteins in Arabidopsis thaliana Cell Stress Chaperones 6:238-246 doi:10.1379/1466-1268(2001)006<0238:thfopi>2.0.co;2
24. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets Mol Biol Evol 33:1870-1874 doi:10.1093/molbev/msw054
25. la Cour T, Kiemer L, Molgaard A, Gupta R, Skriver K, Brunak S (2004) Analysis and prediction of leucine-rich nuclear export signals Protein Eng Des Sel 17:527-536 doi:10.1093/protein/gzh062
26. Le Hir H, Nott A, Moore MJJTibs (2003) How introns influence and enhance eukaryotic gene expression 28:215-220
27. Lee U, Rioflorido I, Hong SW, Larkindale J, Waters ER, Vierling E (2007) The Arabidopsis ClpB/Hsp100 family of proteins: chaperones for stress and chloroplast development Plant J 49:115-127 doi:10.1111/j.1365-313X.2006.02940.x

28. Lescot M et al. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences Nucleic Acids Res 30:325-327 doi:10.1093/nar/30.1.325
29. Li H et al. (2014a) Hydrogen peroxide mediates abscisic acid-induced HSP70 accumulation and heat tolerance in grafted cucumber plants Plant Cell Environ 37:2768-2780 doi:10.1111/pce.12360
30. Li M, Guo S, Xu Y, Meng Q, Li G, Yang XJPp (2014b) Glycine betaine-mediated potentiation of HSP gene expression involves calcium signaling pathways in tobacco exposed to NaCl stress 150:63-75
31. Li M, Ji L, Jia Z, Yang X, Meng Q, Guo S (2018) Constitutive expression of CaHSP22.5 enhances chilling tolerance in transgenic tobacco by promoting the activity of antioxidative enzymes Funct Plant Biol 45:575-585 doi:10.1071/FP17226
32. Li M, Ji L, Yang X, Meng Q, Guo S (2012) The protective mechanisms of CaHSP26 in transgenic tobacco to alleviate photoinhibition of PSII during chilling stress Plant Cell Rep 31:1969-1979 doi:10.1007/s00299-012-1309-x
33. Li Q et al. (2019) A chromosome-scale genome assembly of cucumber (Cucumis sativus L.) GigaScience 8:giz072 doi:10.1093/gigascience/giz072
34. Lin BL, Wang JS, Liu HC, Chen RW, Meyer Y, Barakat A, Delseny M (2001) Genomic analysis of the Hsp70 superfamily in Arabidopsis thaliana Cell Stress Chaperones 6:201-208 doi:10.1379/1466-1268(2001)006<0201:gaoths>2.0.co;2
35. Liu C et al. (2017) Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (Ananas comosus) BMC Genomics 18:503 doi:10.1186/s12864-017-3896-y
36. Liu HC, Charng YY (2012) Acquired thermotolerance independent of heat shock factor A1 (HsfA1), the master regulator of the heat stress response Plant Signal Behav 7:547-550 doi:10.4161/psb.19803
37. Liu HC, Charng YY (2013) Common and distinct functions of Arabidopsis class A1 and A2 heat shock factors in diverse abiotic stress responses and development Plant Physiol 163:276-290 doi:10.1104/pp.113.221168
38. Liu Y et al. (2014) Genome-wide identification and analysis of heat shock protein 90 in tomato Yi Chuan 36:1043-1052 doi:10.3724/SP.J.1005.2014.1043
39. Liu Y, Wang L, Zhang D, Li D (2012) Expression Analysis of Segmentally Duplicated ZmMPK3-1 and ZmMPK3-2 genes in Maize Plant Molecular Biology Reporter 31:457-463 doi:10.1007/s11105-012-0489-4
40. Ma W et al. (2019) Mitochondrial small heat shock protein mediates seed germination via thermal sensing Proc Natl Acad Sci U S A 116:4716-4721 doi:10.1073/pnas.1815790116
41. Mishra SK (2002) In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato Genes & Development 16:1555-1567 doi:10.1101/gad.228802

42. Nejat N, Mantri N (2017) Plant Immune System: Crosstalk Between Responses to Biotic and Abiotic Stresses the Missing Link in Understanding Plant Defence Curr Issues Mol Biol 23:1-16 doi:10.21775/cimb.023.001
43. Nelson CJ, Li L, Millar AH (2014) Quantitative analysis of protein turnover in plants Proteomics 14:579-592 doi:10.1002/pmic.201300240
44. Ogawa D, Yamaguchi K, Nishiuchi T (2007) High-level overexpression of the Arabidopsis HsfA2 gene confers not only increased themotolerance but also salt/osmotic stress tolerance and enhanced callus growth J Exp Bot 58:3373-3383 doi:10.1093/jxb/erm184
45. Ohama N, Sato H, Shinozaki K, Yamaguchi-Shinozaki K (2017) Transcriptional Regulatory Network of Plant Heat Stress Response Trends Plant Sci 22:53-65 doi:10.1016/j.tplants.2016.08.015
46. Pei H, Sun Q, Hao Q, Lv B, Wu J, Fu D (2015) The HSP90-RAR1-SGT1 based protein interactome in barley and stripe rust Physiological and Molecular Plant Pathology 91:11-19 doi:10.1016/j.pmpp.2015.02.004
47. Pratt WB, Toft DO (2003) Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery Exp Biol Med (Maywood) 228:111-133 doi:10.1177/153537020322800201
48. Rikhvanov EG et al. (2007) Nuclear–mitochondrial cross-talk during heat shock in Arabidopsis cell culture 52:763-778
49. Sarkar NK, Kim YK, Grover A (2009) Rice sHsp genes: genomic organization and expression profiling under stress and development BMC Genomics 10:393 doi:10.1186/1471-2164-10-393
50. Sarkar NK, Kundnani P, Grover A (2013) Functional analysis of Hsp70 superfamily proteins of rice (Oryza sativa) Cell Stress Chaperones 18:427-437 doi:10.1007/s12192-012-0395-6
51. Scharf K-D, Berberich T, Ebersberger I, Nover L (2012) The plant heat stress transcription factor (Hsf) family: Structure, function and evolution Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1819:104-119 doi:10.1016/j.bbagrm.2011.10.002
52. Scharf K-D, Siddique M, Vierling E (2001) &cestflwr; The expanding family of Arabidopsis thaliana small heat stress proteins and a new family of proteins containing α-crystallin domains (Acd proteins) Cell Stress & Chaperones 6:225 doi:10.1379/1466-1268(2001)006<0225:Tefoat>2.0.Co;2
53. Scharf KD, Rose S, Zott W, Schöffl F, Nover L, Schöff F (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 The EMBO Journal 9:4495-4501 doi:10.1002/j.1460-2075.1990.tb07900.x
54. Shao HB, Guo QJ, Chu LY, Zhao XN, Su ZL, Hu YC, Cheng JF (2007) Understanding molecular mechanism of higher plant plasticity under abiotic stress Colloids Surf B Biointerfaces 54:37-45 doi:10.1016/j.colsurfb.2006.07.002
55. Singh A, Singh U, Mittal D, Grover A (2010) Genome-wide analysis of rice ClpB/HSP100, ClpC and ClpD genes BMC Genomics 11:95 doi:10.1186/1471-2164-11-95

56. Su H, Zhang S, Yuan X, Chen C, Wang X-F, Hao Y-JJPP, Biochemistry (2013) Genome-wide analysis and identification of stress-responsive genes of the NAM–ATAF1, 2–CUC2 transcription factor family in apple 71:11-21
57. Sun L et al. (2012) ZmHSP16.9, a cytosolic class I small heat shock protein in maize (Zea mays), confers heat tolerance in transgenic tobacco Plant Cell Rep 31:1473-1484 doi:10.1007/s00299-012-1262-8
58. Swindell WR, Huebner M, Weber AP (2007) Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways BMC Genomics 8:1-15 doi:10.1186/1471-2164-8-125
59. Thompson J (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools Nucleic Acids Research 25:4876-4882 doi:10.1093/nar/25.24.4876
60. Tissiéres A, Mitchell HK, Tracy UM (1974) Protein synthesis in salivary glands of Drosophila melanogaster: Relation to chromosome puffs Journal of Molecular Biology 84:389-398 doi:10.1016/0022-2836(74)90447-1
61. Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs J Hered 93:77-78 doi:10.1093/jhered/93.1.77
62. Vu NT et al. (2019) Comparative co-expression network analysis extracts the SlHSP70 gene affecting to shoot elongation of tomato Plant Biotechnol (Tokyo) 36:143-153 doi:10.5511/plantbiotechnology.19.0603a
63. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J (2010) KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies Genomics Proteomics Bioinformatics 8:77-80 doi:10.1016/S1672-0229(10)60008-3
64. Wang R et al. (2014) Identification and phylogenetic relationships of Hsp70 genes in tomato Journal of Nuclear Agricultural Sciences 28:378-385
65. Wang W, Vinocur B, Shoseyov O, Altman A (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response Trends Plant Sci 9:244-252 doi:10.1016/j.tplants.2004.03.006
66. Wang Y et al. (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity Nucleic Acids Research 40:e49-e49 doi:10.1093/nar/gkr1293
67. Wei Y, Liu W, Hu W, Yan Y, Shi H (2020) The chaperone MeHSP90 recruits MeWRKY20 and MeCatalase1 to regulate drought stress resistance in cassava New Phytol 226:476-491 doi:10.1111/nph.16346
68. Xiong L, Ishitani M (2006) Stress Signal Transduction: components, pathways and network integration. In: Abiotic stress tolerance in plants. Springer, pp 3-29. doi:10.1007/1-4020-4389-9_1
69. Yang J-y, Sun Y, Sun A-q, Yi S-y, Qin J, Li M-h, Liu JJPMB (2006) The involvement of chloroplast HSP100/ClpB in the acquired thermotolerance in tomato 62:385-395

70. Yang S, Zhang X, Yue JX, Tian D, Chen JQ (2008) Recent duplications dominate NBS-encoding gene expansion in two woody species Mol Genet Genomics 280:187-198 doi:10.1007/s00438-008-0355-0
71. Yang X, Zhu W, Zhang H, Liu N, Tian S (2016) Heat shock factors in tomatoes: genome-wide identification, phylogenetic analysis and expression profiling under development and heat stress PeerJ 4:e1961 doi:10.7717/peerj.1961
72. Yoshida T et al. (2011) Arabidopsis HsfA1 transcription factors function as the main positive regulators in heat shock-responsive gene expression Molecular Genetics and Genomics 286:321-332 doi:10.1007/s00438-011-0647-7
73. Yu J et al. (2016) Genome-Wide Identification and Expression Profiling of Tomato Hsp20 Gene Family in Response to Biotic and Abiotic Stresses Front Plant Sci 7:1215 doi:10.3389/fpls.2016.01215
74. Zhang T, Lv W, Zhang H, Ma L, Li P, Ge L, Li G (2018) Genome-wide analysis of the basic Helix-Loop-Helix (bHLH) transcription factor family in maize BMC Plant Biol 18:235 doi:10.1186/s12870-018-1441-z

Tables

Table. 1 Ka/Ks calculation and divergent time of the duplicated cucumber *Hsf* and *Hsp* gene pairs.

Duplicated Gene Pairs	Ka	Ks	Ka/Ks	Duplicated Type	Purify selection	Time * (MYA)
CsHsp20-3/ CsHsp20-4	0.4865	2.9041	0.1675	Tandem	Yes	221.35
CsHsp20-4/ CsHsp20-5	0.0586	0.2807	0.2086	Tandem	Yes	21.39
CsHsp20-7/ CsHsp20-8	0.1454	0.2153	0.6754	Tandem	Yes	16.41
CsHsp20-9/ CsHsp20-10	0.0472	0.2244	0.2101	Tandem	Yes	17.11
CsHsp20-14/ CsHsp20-15	0.2937	0.2767	1.0616	Tandem	No	21.09
CsHsp20-21/ CsHsp20-22	0.0841	0.6262	0.1343	Tandem	Yes	47.73
CsHsp20-22/ CsHsp20-23	0.0700	0.4006	0.1746	Tandem	Yes	30.53
CsHsp20-24/ CsHsp20-25	0.0680	0.4096	0.1661	Tandem	Yes	31.22
CsHsp20-25/ CsHsp20-26	0.2610	0.9184	0.2842	Tandem	Yes	70.00
CsHsp20-26/ CsHsp20-27	0.2900	0.9398	0.3086	Tandem	Yes	71.63
CsHsp100-1/ CsHsp100-5	0.5548	0.8385	0.6616	Segmental	Yes	63.91
CsHsp100-2/ CsHsp100-3	0.2868	6.1528	0.0466	Segmental	Yes	468.96
CsHsp70-6/ CsHsp70-10	0.0308	2.5042	0.0123	Segmental	Yes	190.87
CsHsp20-21/ CsHsp20-30	0.1834	2.8788	0.0637	Segmental	Yes	219.42
CsHsp20-4/ CsHsp20-25	0.2092	4.1111	0.0509	Segmental	Yes	313.35
CsHsp20-3/ CsHsp20-24	0.4982	1.2190	0.4087	Segmental	Yes	92.91
CsHsp20-14/ CsHsp20-21	0.6852	4.9858	0.1374	Segmental	Yes	380.01
CsHsp20-18/ CsHsp20-21	0.1993	3.7675	0.0529	Segmental	Yes	287.15
CsHsf-11/ CsHsf-23	0.3684	2.6212	0.1405	Segmental	Yes	199.79

CsHsf-2/ CsHsf-22	0.3567	1.7535	0.2034	Segmental	Yes	133.65
CsHsf-8/ CsHsf-16	0.2446	4.2234	0.0579	Segmental	Yes	321.91
CsHsf-6/ CsHsf-11	0.4058	2.7181	0.1493	Segmental	Yes	207.17
CsHsf-10/ CsHsf-17	1.0568	3.0482	0.3467	Segmental	Yes	232.33

Table.2 Function of the cis-elements detected in *CsHsf* and *CsHsp* gene promoters.

Type	Site name	Function
Hormone response element	ABRE	Abscisic acid-responsive element
	TGA-element	Auxin-responsive element
	AuxRR-core	Auxin-responsive element
	TGA-box	Auxin-responsive element
	ERE	Ethylene-responsive element
	P-box	Gibberellin-responsive element
	GARE-motif	Gibberellin-responsive element
	TATC-box	Gibberellin-responsive element
	CGTCA-motif	MeJA-responsive element
	TGACG-motif	MeJA-responsive element
	TCA-element	Salicylic acid-responsive element
	SARE	Salicylic acid-responsive element
Stress-induced component	HSE	Heat stress element
	LTR	Low temperature responsive element
	DRE	Dehydration-responsive element
	MBS	Drought-responsive element
	W box	Defense-responsive element
	TC-rich repeats	Defense and stress responsive element
	WUN-motif	Wound-responsive element
	ARE	Hypoxia inducible response element
	GC-motif	Anoxic specific inducible response element

Figures

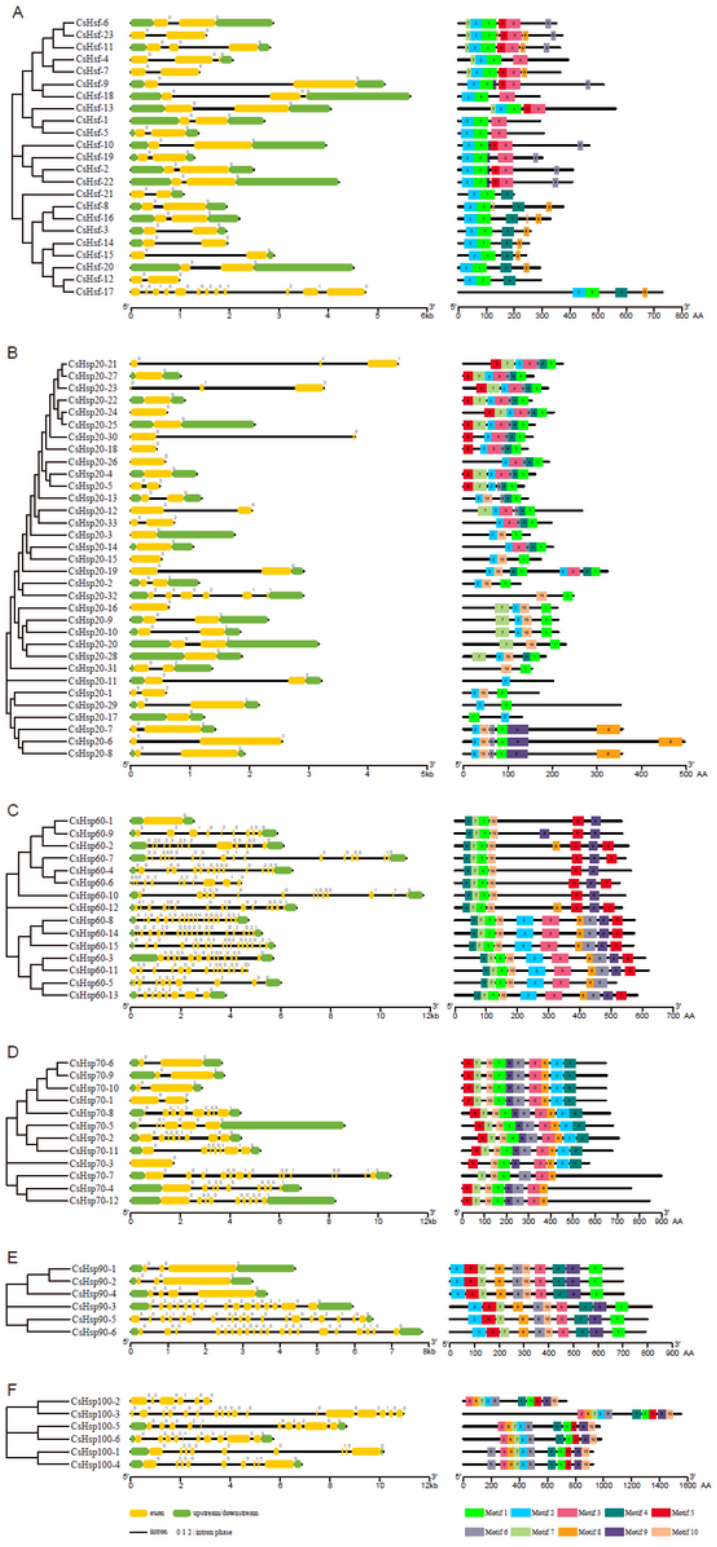

Figure 1

Phylogenetic relationship, gene structure and conserved motif analysis of CsHsf and CsHsp. Phylogenetic tree of CsHsf (A), CsHsp20 (B), CsHsp60 (C), CsHsp70 (D), CsHsp90 (E) and CsHsp100 (F) proteins are shown in the left panel. The unrooted neighbor-joining phylogenetic tree was constructed

with MEGA7.0 using full-length amino acid sequences of CsHsf and CsHsp proteins, and the bootstrap test replicate was set as 1000 times. Exon-intron organization of CsHsf (A), CsHsp20 (B), CsHsp60 (C), CsHsp70 (D), CsHsp90 (E) and CsHsp100 (F) genes are shown in the middle panel. Yellow boxes represent exons and black lines represent introns. The upstream/downstream region of CsHsf and CsHsp genes are indicated in green boxes. The numbers of 0, 1, and 2 represent the splicing phase of intron. The length of exons can be inferred by the scale at the bottom. Distributions of conserved motifs in CsHsf (A), CsHsp20 (B), CsHsp60 (C), CsHsp70 (D), CsHsp90 (E) and CsHsp100 (F) are displayed in the right panel. Ten putative motifs are indicated in different colored boxes. For details of motifs refer to Table S2.

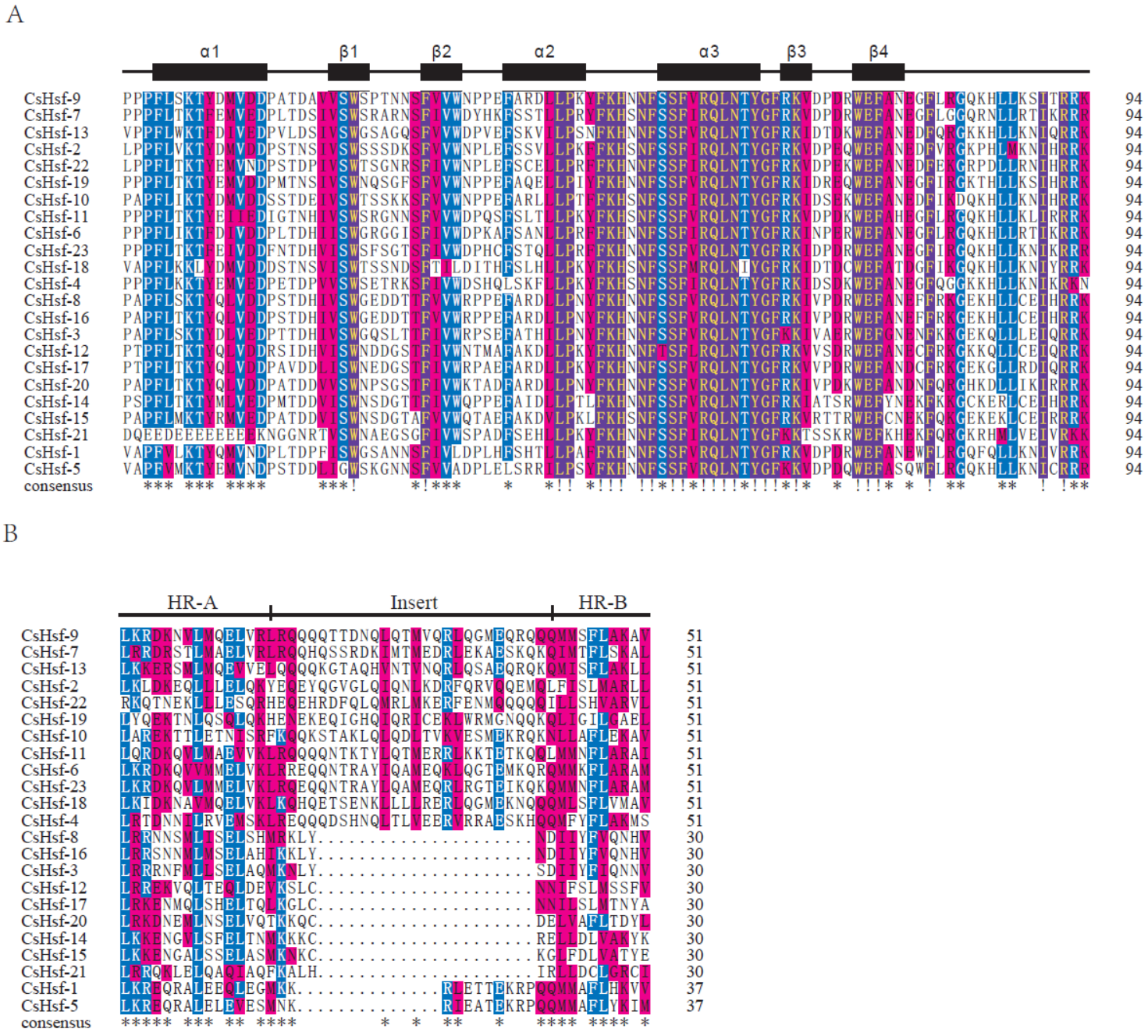

Figure 2

Multiple sequence alignment of the DNA-binding domain and HR-A/B regions (OD) of cucumber Hsfs. (A)DBD domain sequences of CsHsfs identified by Pfam database were aligned by Clustal X 2.0 software. The completely conserved residues are purple colored and highly conserved residues are red or blue colored, all visualized by Texshade (https://ctan.org/pkg/texshade). Multiple alignment clearly reveals that the DBD domains of the CsHsfs are highly conserved. The secondary structure elements of the DBD (α1-β1-β2-α2-α3-β3-β4) are shown above the alignment. (B)The HR-A/B region sequences identified by SMART online tool were aligned by Clustal X 2.0 software. The completely conserved residues are purple colored and highly conserved residues are red or blue colored, all visualized by Texshade (https://ctan.org/pkg/texshade). The three line segments at the top divide HR-A core, insert and HR-B regions orderly.

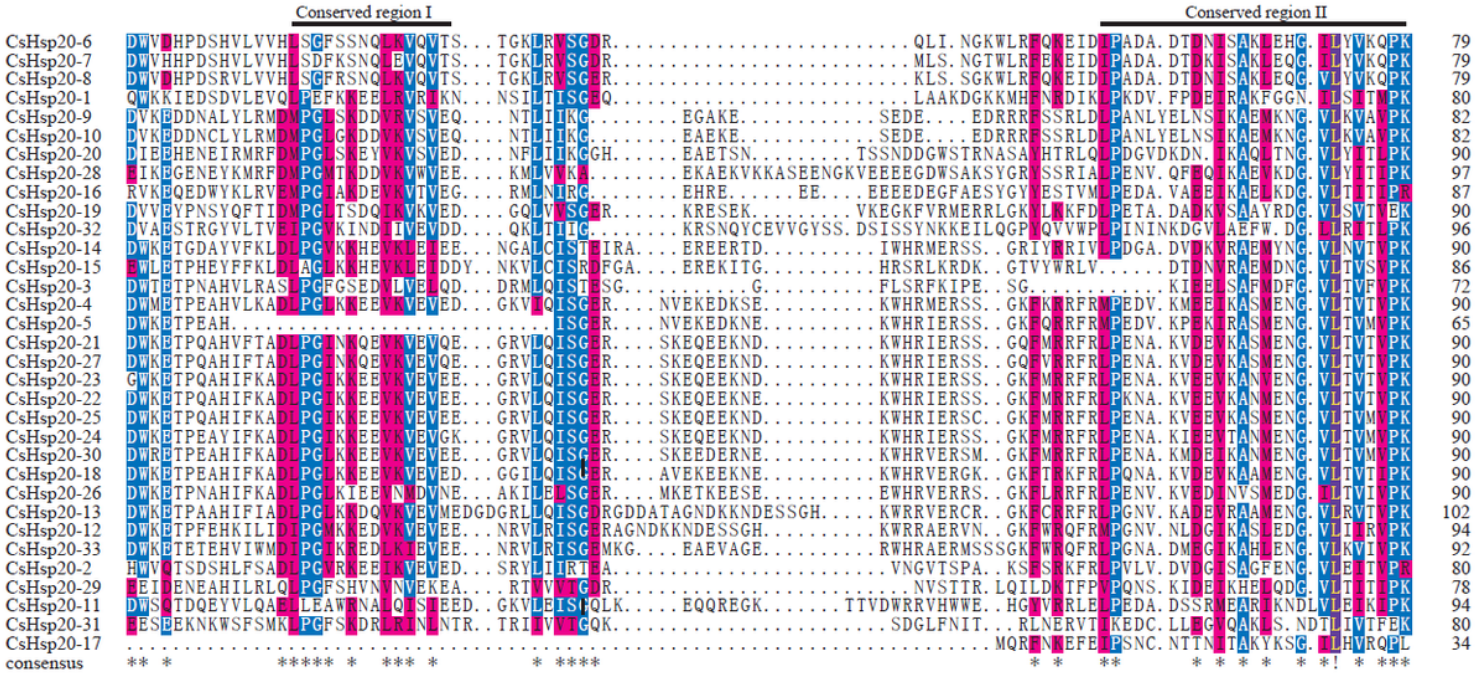

Figure 3

Multiple sequence alignment of the α-crystallin (ACD) domain of cucumber Hsp20s. The ACD domain sequences identified by SMART online tool were aligned by Clustal X 2.0 software. The completely conserved residues are purple colored and highly conserved residues are red or blue colored, all visualized by Texshade (https://ctan.org/pkg/texshade). Two consensus regions (consensus I and consensus II) are underlined at the top.

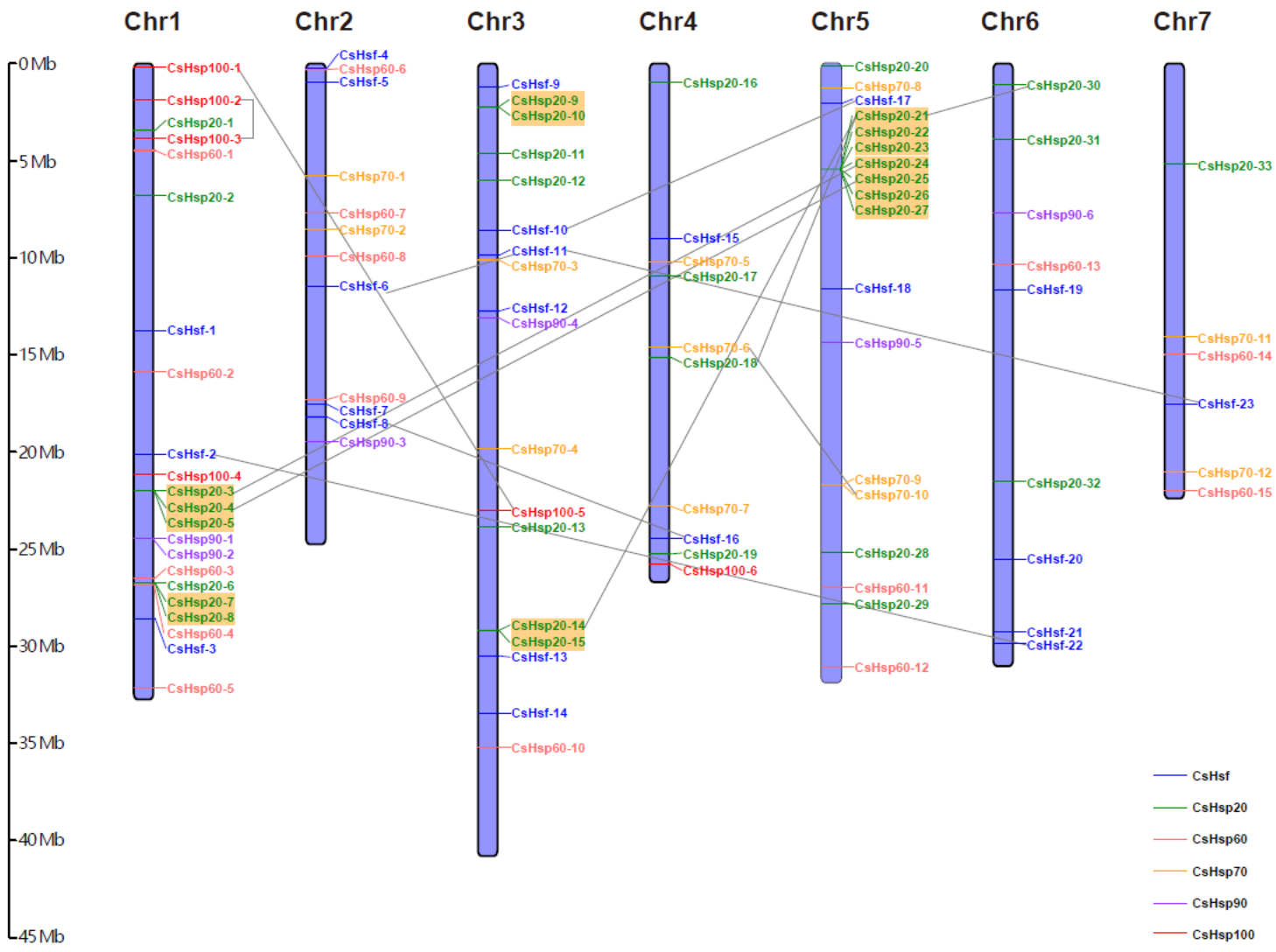

Figure 4

Chromosome distribution of CsHsf and CsHsp genes. The 95 heat stress genes were marked on 7 cucumber chromosomes (vertical bar). Six different heat shock gene families were indicated by different colored lines on chromosomes. Chromosome numbers are indicated at the top of each bar. Six tandem clustered genes of CsHsp20s were indicated by yellow boxes at chromosomes 1, 3 and 5. Gray lines represent gene Segmental duplications.

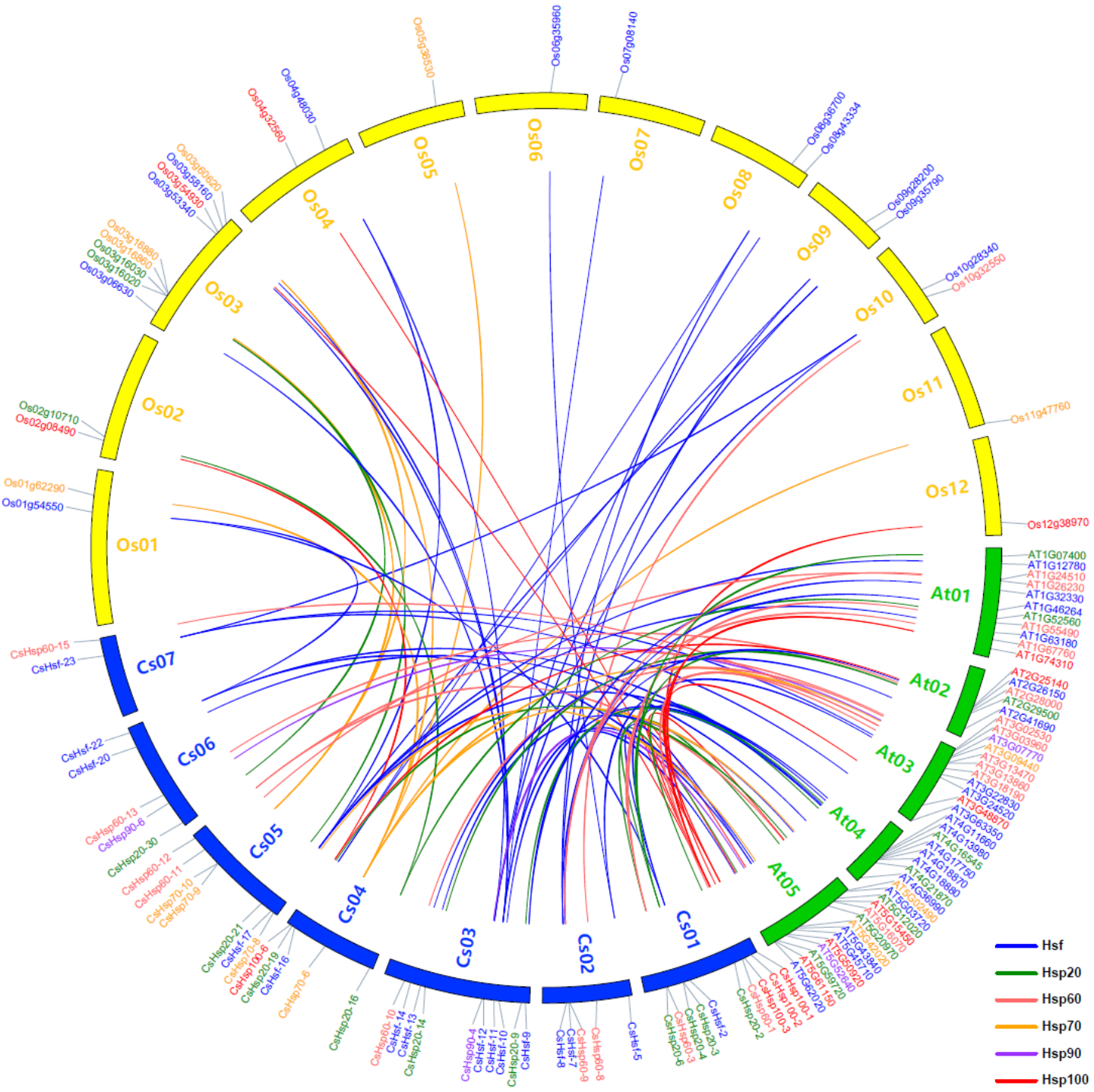

Figure 5

Synteny analysis of Hsf and Hsp genes between Arabidopsis, rice and cucumber. Different color lines in the circle highlight the syntenic gene pairs in six different heat shock gene families in Arabidopsis, rice and cucumber genomes. The yellow bars represent 12 rice chromosomes, green bars represent 5 Arabidropsis chromosomes and blue bars represent 7 cucumber chromosomes.

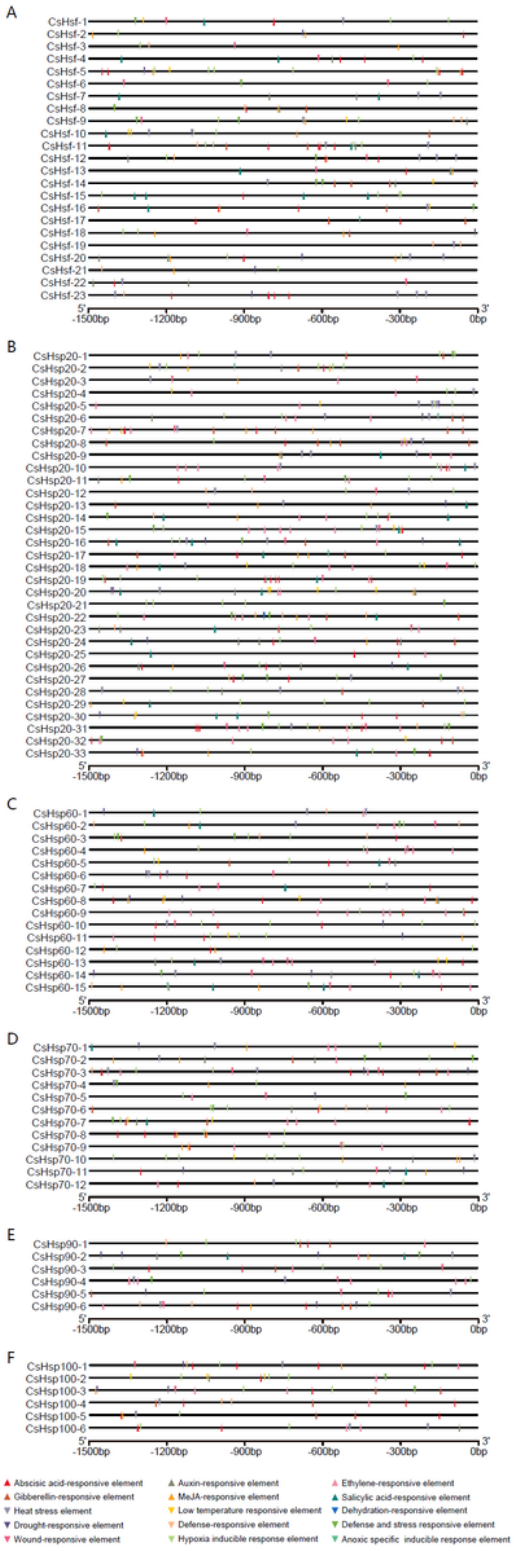

Figure 6

Predicted cis-elements in CsHsf and CsHsp promoters. Promoter sequences (1500 bp) of 23 CsHsf and 72 CsHsp genes are analyzed by PlantCARE. The upstream length to the translation start site can be inferred according to the scale at the bottom. The different color triangles stand for 6 hormone response elements, and the different color inverted triangles stand for 9 stress-induced components.

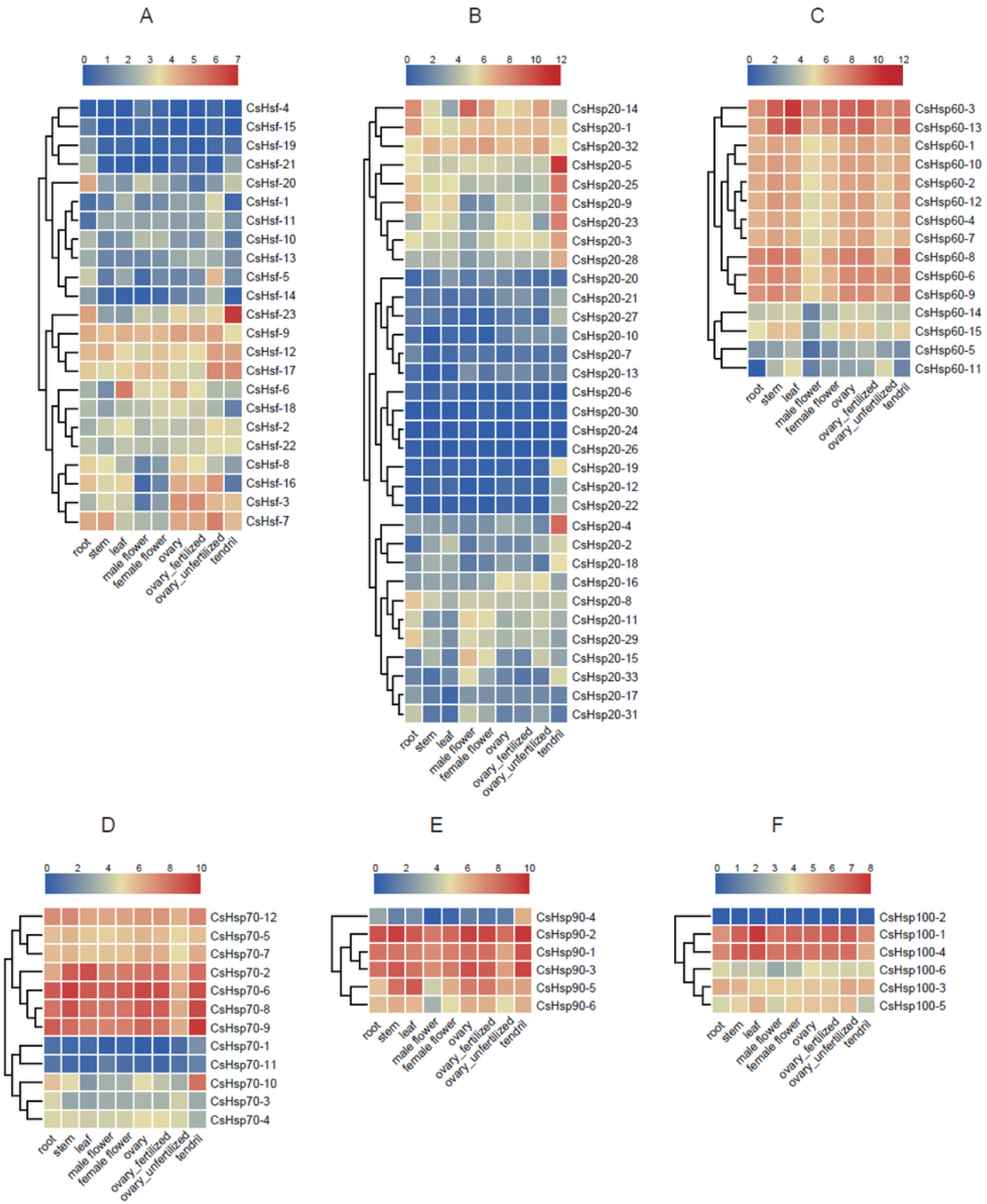

Figure 7

Heat map of the expression profiles of CsHsf and CsHsp genes in different tissues. RNA-seq relative expression data from 9 cucumber tissues including root, stem, leaf, male flower, female flower, ovary(unexpanded), expanded ovary (fertilized), expanded ovary (unfertilized), and tendril were used to reconstruct the expression patterns of CsHsf and CsHsp genes. The raw data were obtained from the Cucurbit Genomics Database (http://cucurbitgenomics.org/rnaseq/home, PRJNA80169). FPKM values of

CsHsf and CsHsp genes were transformed by log2 (FPKM+ 1) and the heat map was constructed by HemI software. Heat maps are presented in blue/yellow/red colors that represent low/medium/high expression, respectively. Color scales representing the relative expression values are shown on the upper of heat map. The normal relative expression levels of 23 CsHsf and 72 CsHsp genes in different tissues are shown in Table S4.

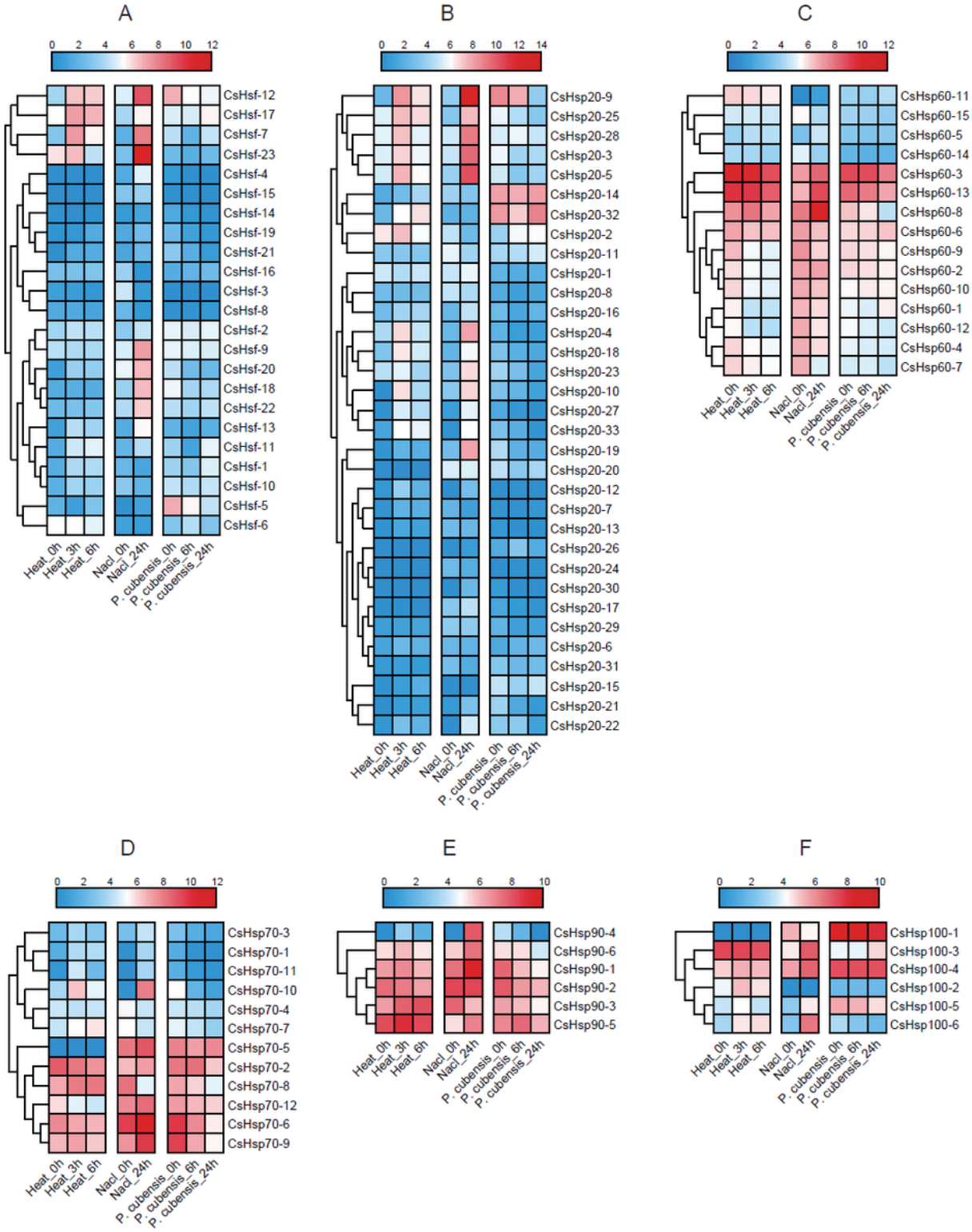

Figure 8

Expression profiles of CsHsf and CsHsp genes under heat, NaCl and downy mildew stresses. Our own RNA-seq data were used to analyze the response of CsHsf and CsHsp genes to heat stress, and the RNA-seq data for CsHsf and CsHsp genes in response to NaCl and downy mildew stresses were obtained from the Cucurbit Genomics Database (http://cucurbitgenomics.org/rnaseq/home, PRJNA437579 and PRJNA388584). FPKM values of CsHsf and CsHsp genes were transformed by log2 (FPKM+ 1) and the heat map was constructed by HemI software. Heat maps are presented in blue/white/red colors that represent low/medium/high expression, respectively. Color scales representing the relative expression values are shown on the upper of heat map. The normal relative expression levels of 23 CsHsf and 72 CsHsp genes under heat, NaCl and downy mildew stresses are shown in Table S5.

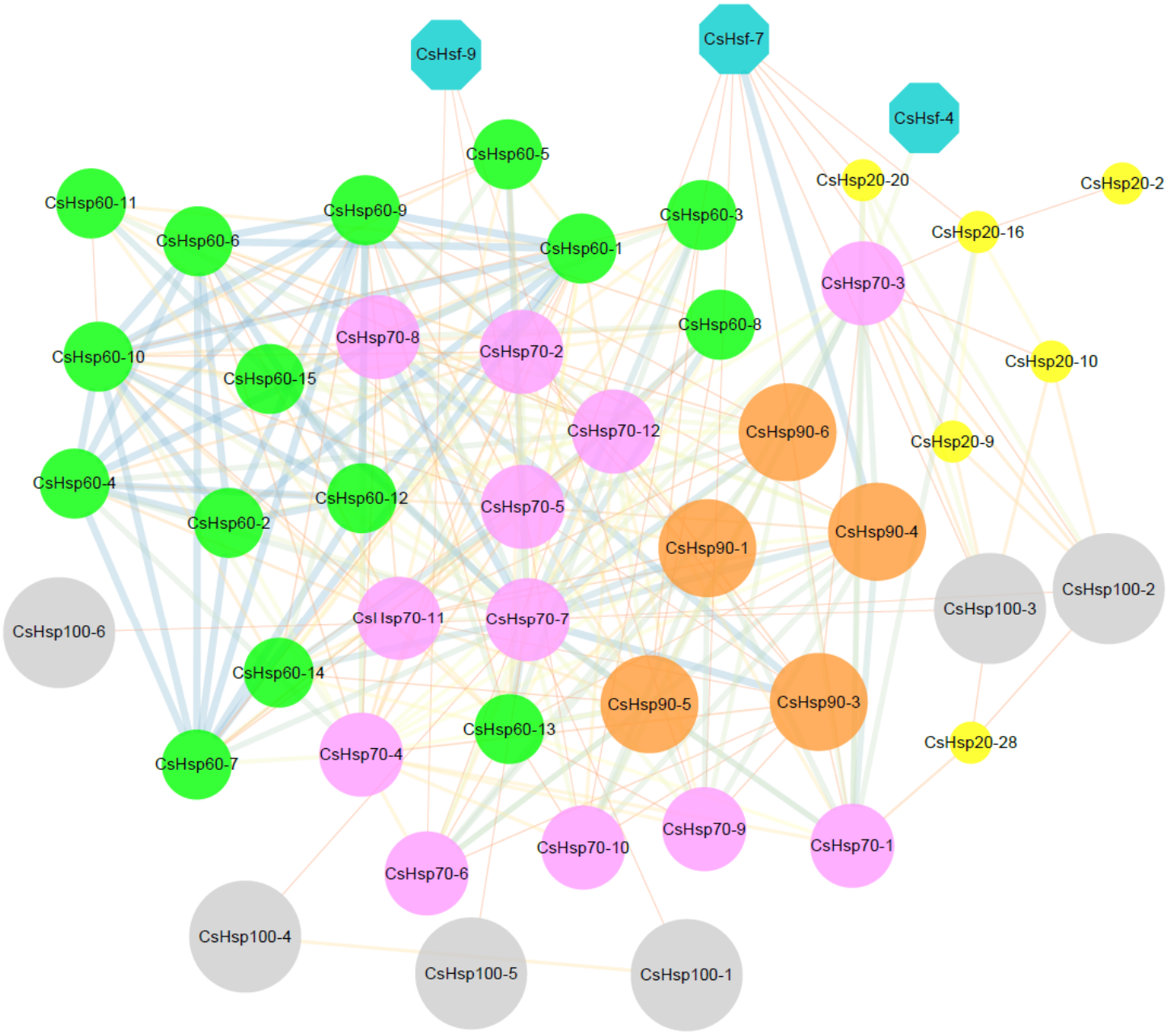

Figure 9

Interaction network analysis of CsHsp and CsHsf identified by STRING program. Nodes represent Hsfs and Hsps in cucumber, edges indicate pairwise correlation constructed by STRING. Node color codes represent different gene families. Blue octagons indicate Hsfs, solid circles with yellow, green, pink, orange, and gray indicate Hsp20s, Hsp60s, Hsp70s, Hsp90s and Hsp100s, respectively. Red thin edges indicate moderate interactions and blue thick edges indicate strong interactions between the two nodes. The network was created using Cytoscape.

Supplementary Files

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

- SupplementalFiguerS1.pdf
- SupplementalFiguerS2.pdf
- SupplementalFiguerS3.pdf
- SupplementalFiguerS4.pdf
- SupplementalFiguerS5.pdf
- SupplementalFiguerS6.pdf
- SupplementalTableS1.xlsx
- SupplementalTableS2.xlsx
- SupplementalTableS3.xlsx
- SupplementalTableS4.xlsx
- SupplementalTableS5.xlsx