



Genome-wide analysis of the *CaHsp20* gene family in pepper: comprehensive sequence and expression profile analysis under heat stress

OPEN ACCESS

Edited by:

Jun Yu,
Beijing Institute of Genomics, China

Reviewed by:

Xun Xu,
Beijing Genomics Institute, China
Jingfa Xiao,
Beijing Institute of Genomics, China

*Correspondence:

Zhen-Hui Gong and Ming-Hui Lu,
Department of Vegetable Science,
College of Horticulture, Northwest
A&F University, No. 3 Taicheng Road,
Yangling, Shaanxi 712100, China
zhgong@nwsuaf.edu.cn;
xnjacklu@nwsuaf.edu.cn;
Shu-Bin Wang,
Laboratory for Solanaceous Fruit
Vegetable, Institute of Vegetable
Crops, Jiangsu Academy of
Agricultural Sciences, No. 50
Zhongling Street, Nanjing,
Jiangsu 210014, China
wangsbpep@163.net

Specialty section:

This article was submitted to
Plant Genetics and Genomics,
a section of the journal
Frontiers in Plant Science

Received: 04 June 2015

Accepted: 16 September 2015

Published: 01 October 2015

Citation:

Guo M, Liu J-H, Lu J-P, Zhai Y-F,
Wang H, Gong Z-H, Wang S-B and
Lu M-H (2015) Genome-wide analysis
of the *CaHsp20* gene family in pepper:
comprehensive sequence and
expression profile analysis under heat
stress. *Front. Plant Sci.* 6:806.
doi: 10.3389/fpls.2015.00806

Meng Guo¹, Jin-Hong Liu¹, Jin-Ping Lu¹, Yu-Fei Zhai¹, Hu Wang¹, Zhen-Hui Gong^{1*},
Shu-Bin Wang^{2*} and Ming-Hui Lu^{1*}

¹ Department of Vegetable Science, College of Horticulture, Northwest A&F University, Yangling, China, ² Laboratory for Solanaceous Fruit Vegetable, Institute of Vegetable Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China

The *Hsp20* genes are present in all plant species and play important roles in alleviating heat stress and enhancing plant thermotolerance by preventing the irreversible aggregation of denaturing proteins. However, very little is known about the *CaHsp20* gene family in pepper (*Capsicum annuum* L.), an important vegetable crop with character of temperate but thermosensitive. In this study, a total of 35 putative pepper *Hsp20* genes (*CaHsp20s*) were identified and renamed on the basis of their molecular weight, and then their gene structure, genome location, gene duplication, phylogenetic relationship, and interaction network were also analyzed. The expression patterns of *CaHsp20* genes in four different tissues (root, stem, leaf, and flower) from the thermotolerant line R9 under heat stress condition were measured using semi-quantitative RT-PCR. The transcripts of most *CaHsp20* genes maintained a low level in all of the four tissues under normal temperature condition, but were highly induced by heat stress, while the expression of *CaHsp16.6b*, *16.7*, and *23.8* were only detected in specific tissues and were not so sensitive to heat stress like other *CaHsp20* genes. In addition, compared to those in thermotolerant line R9, the expression peak of most *CaHsp20* genes in thermosensitive line B6 under heat stress was hysteretic, and several *CaHsp20* genes (*CaHsp16.4*, *18.2a*, *18.7*, *21.2*, *22.0*, *25.8*, and *25.9*) showed higher expression levels in both line B6 and R9. These data suggest that the *CaHsp20* genes may be involved in heat stress and defense responses in pepper, which provides the basis for further functional analyses of *CaHsp20s* in the formation of pepper acquired thermotolerance.

Keywords: *Capsicum annuum* L., *CaHsp20* genes, heat stress, expression profile, gene family

Introduction

As the sessile organisms, plants cannot escape the extreme environment conditions, including biotic and abiotic stresses, such as pests and pathogens, drought, salinity, cold, heat, and heavy metal stresses (Wang et al., 2003; Mittler, 2006; Mittler et al., 2012). Increase in the global warming having adverse effects of heat stress (HS) on crop plants growth, development and productivity

are more serious, thus plants adaptive response to high temperature i.e., hypersensitive response (HSR) is very important for world food security (Glazebrook, 2001; Giorno et al., 2010). Under HS conditions, the heat shock transcription factors (Hsfs) bind to the heat shock elements (HSEs) of the promoters of heat-responsive genes, mainly the heat shock proteins (Hsps), to turn on or regulate their expression (Scharf et al., 2012; Waters, 2013). Plant Hsps as molecular chaperones are involved in maintaining homeostasis of protein folding, preventing or repairing the misfolding and degradation of proteins caused by HS, and then confer the acquisition of plant thermotolerance (Vierling, 1991; Sung et al., 2003; Charng et al., 2006).

The plant Hsps are grouped into five families based on the molecular weight and sequence homology, including the Hsp100s, Hsp90s, Hsp70s, Hsp60s, and Hsp20s (Sarkar et al., 2009; Waters, 2013). Among all the five conserved families, Hsp20s, a group of small heat shock proteins (sHsps) with molecular size ranging from 15 to 42 kDa, are the most prevalent and abundant family induced by HS in many higher plants. Furthermore, they can function as molecular chaperone without ATP to maintain the stability of proteins, which suggest that Hsp20s may play a vital role during the formation of plant acquired thermotolerance (Vierling, 1991; Sun et al., 2002; Wang et al., 2004; Waters, 2013; Haslbeck and Vierling, 2015).

Hsp20s share a conserved modular structure, and the most striking feature is a core conserved alpha-crystallin domain (ACD, the Hsp20 domain) containing ~90 amino acids flanked by a variable N-terminal domain and a C-terminal extension (Vierling, 1991; Scharf et al., 2001; Waters, 2013), in which the ACD is involved in substrate interactions, the N-terminus is responsible for substrate binding and the C-terminal participates in homo-oligomerization and the formation of HS granula (Kirschner et al., 2000; Giese and Vierling, 2004; Basha et al., 2006; Jaya et al., 2009). The ACD is composed of two anti-parallel sheets of four and three β -strands, i.e., conserved regions I (CRI, β 2- β 3- β 4- β 5) in N-terminal consensus and CRII (β 7- β 8- β 9) in C-terminal consensus, respectively, connected by a hydrophobic region loop (β 6-loop; Bondino et al., 2012; Haslbeck and Vierling, 2015). Besides, the N-terminus contains the transit, targeting or signal sequences, and the C-terminal may also contain organelle-specific retention amino acid motifs (Waters, 2013).

The Hsp20s form a more diverse family than other four Hsp families in plants, and are also far more than the members of Hsp20 families in animals (Wang et al., 2004; Waters, 2013). Plant Hsp20s are encoded by nuclear genes and are divided into various subfamilies based on cellular location, sequence homology and function (Vierling, 1991; Waters et al., 1996). Nineteen *Hsp20* genes in *Arabidopsis* (Scharf et al., 2001; Waters, 2013), 23 in *Oryza sativa*, 36 in *Populus trichocarpa* (Waters et al., 2008), and 51 in *Glycine max* were identified, furthermore, the *Hsp20* genes in *Arabidopsis* were grouped into 11 subfamilies (Scharf et al., 2001; Waters, 2013), while those from *G. max* were grouped into 15 subfamilies based on phylogenetic analysis, prediction of subcellular localization and prediction of protein secondary structure models (Lopes-Caitar et al., 2013). However, to our

knowledge, genome-wide analysis of *Hsp20* family in pepper is still lacking.

Pepper (*Capsicum annuum* L.) is an important vegetable crop worldwide, and is deeply loved by a large population due to its major ingredient in cuisines, essential vitamins and other healthy nutrients (Kim et al., 2014). However, pepper is sensitive to high temperature, especially in reproductive stage (Guo et al., 2014). Once suffering continuous HS, the pollination and fertilization of pepper will be severely damaged, which can greatly reduce the yield and quality of pepper fruits. Therefore, study of Hsp20s is important for understanding the mechanism of heat response in pepper, while except several *CaHsp20* genes, such as *CaHsp24* (Zhu et al., 2011), *CaHsp26* (Guo et al., 2007), little is currently known about the *C. annuum* heat shock protein 20 (*CaHsp20*) family.

In this study, we identify and analyze 35 *CaHsp20* genes from pepper genome, focusing on their heat-induced expression in various tissues, and their dynamic expression patterns in two different thermotolerant pepper lines under HS condition. Our results provide a foundation for further functional analysis of *CaHsp20* members, and will be valuable for illuminating the mechanism of pepper acquired thermotolerance, and even the adaptability of pepper to other stress conditions.

Materials and Methods

Genome-wide Identification of *CaHsp20* Genes in Pepper

The Hsp20 proteins are defined by possessing a highly conserved 80–100 amino acid residues called ACD, and a low molecular weight (15–42 kDa; Vierling, 1991; Scharf et al., 2001). To identify the Hsp20 family members in pepper, the Hidden Markov Model (HMM) profile of ACD (PF00011) downloaded from the Pfam protein family database (<http://pfam.sanger.ac.uk>) was used as queries to search the Pepper Genome Database (PGD, <http://peppergenome.snu.ac.kr/>, CM334, Zunla-1 and glabriusculum genomes; Kim et al., 2014; Qin et al., 2014). Secondly, all *O. sativa* and *G. max* Hsp20 protein sequences (Ouyang et al., 2009; Lopes-Caitar et al., 2013) were used as query sequences to search against the PGD and the NCBI database. All output putative pepper *Hsp20* genes with an *e*-value ≤ 0.001 were assembled to remove redundancy, and then submitted to Pfam and SMART (<http://smart.embl-heidelberg.de/>) to confirm the presence of the ACD domain. The Multiple Sequence Alignment (MSA) was subsequently trimmed in order to remove subsequences that were not truly homologous. The predicted genes lacking an ACD domain and/or the genes whose molecular weight out of the range from 15 to 42 kDa were rejected. All non-redundant and confident genes were assigned as pepper *Hsp20* genes (*CaHsp20s*), and annotated based on their molecular weight as with soybean (Lopes-Caitar et al., 2013). The logic flowchart of identification of pepper *Hsp20* genes was shown in **Figure S1**.

Using the same method, the putative *Hsp20* homologs in tomato, *Arabidopsis*, rice, and maize were obtained from the SOL

Genomics Network (SGN) database (http://solgenomics.net/organism/Solanum_lycopersicum/genome), the TAIR database (<https://www.arabidopsis.org/index.jsp>), the Rice Genome Annotation Project (<http://rice.plantbiology.msu.edu/>), and the TIGR Maize Database (<http://maize.jcvi.org/>), respectively.

The protein sequences of putative CaHsp20 members were analyzed with EXPASY PROTOPARAM (<http://www.expasy.org/tools/prototparam.html>) to obtain the number of amino acids, molecular weight, theoretical isoelectric point (pI) and instability index (with a value >40 considered as unstable; Lopes-Caitar et al., 2013). The chromosomal location data and intron numbers of putative *CaHsp20s* were retrieved from PGD. The WoLF PSORT program (<http://wolfsort.org/>) and TargetP (Emanuelsson et al., 2000) were used for predicting the subcellular locations.

Analysis of Conserved Motifs of Pepper *CaHsp20* Family

The conserved domains of the CaHsp20 proteins (ACD) were confirmed by Pfam and SMART. The MEME program (Version 4.10.0, <http://meme.nbcr.net/meme/tools/meme>) was used for identification of the conserved motifs, with the following parameters: number of repetitions-any, maximum number of motifs-10, the optimum motif widths set from 6 to 200 amino acid residues; the conserved motifs were annotated with InterProScan (<http://www.ebi.ac.uk/interpro/search/sequence-search>; Ouyang et al., 2009). The ACD sequences of all CaHsp20 members were aligned using the Promals3D structural alignment program (<http://prodata.swmed.edu/promals3d/promals3d.php>).

Phylogenetic Analysis and Classification of Pepper *CaHsp20* Genes

The full amino acid sequences of Hsp20 members from pepper, tomato, *Arabidopsis*, rice, and maize were aligned by CLUSTALW program. The gene IDs of *Hsp20* members from tomato, *Arabidopsis*, rice, and maize were shown in **Table S1**. An unrooted neighbor-joining phylogenetic tree was constructed using MEGA 5.10 software (Tamura et al., 2011) with the bootstrap test replicated 1000 times, pair wise deletion and a Poisson model. The pepper *CaHsp20* genes were assigned to different groups based on the multiple sequence alignment and the classification of *Hsp20* genes in the other four species.

Chromosomal Localization and Gene Duplication

Each of the *CaHsp20* genes was mapped on chromosomes using MapDraw (Liu and Meng, 2003) based on their chromosomal position and the relative distance between these genes on the same chromosome derived from PGD. The *CaHsp20* gene duplication was defined based on the following criteria: (1) the length of the sequence alignment covered $\geq 70\%$ of the longer gene; and (2) the similarity of the aligned gene regions $\geq 70\%$ (Gu et al., 2002; Yang et al., 2008). Genes separated by five or fewer genes in a 100-kb region were considered to be tandem duplicates (Wang et al., 2010); genes were regarded as segmental duplication if they were found to be coparalogs located on duplicated chromosomal blocks, as proposed by Wei et al. (2007).

Prediction of CaHsp20s Protein-Protein Interaction Network

An interaction network of CaHsp20 proteins from pepper was constructed to understand the genome-wide regulation network; the interolog from *Arabidopsis* was used for predicting protein-protein interaction network of CaHsp20 members. An interaction work of *Arabidopsis* Hsp20 proteins was constructed using the *Arabidopsis* Interactions Viewer (http://bar.utoronto.ca/interactions/cgi-bin/arabidopsis_interactions_viewer.cgi). *Arabidopsis* Hsp20 proteins were mapped to pepper CaHsp20 proteins based on their homologous relationship, and an interaction network of CaHsp20 proteins was drawn by Cytoscape_v2.8.3 (National Institute of General Medical Sciences, MD, USA).

Plant Materials and Heat Stress Treatments

The pepper thermotolerant line R9 and the thermosensitive line B6 were used in this study. Both lines were grown in a growth chamber at 26°C (day for 16 h)/20°C (night for 8 h) till reaching the age of 6–8 true leaves. R9 plants were used for the tissue specific expression analysis, the seedlings with 6–8 true leaves and the plants with flower buds were treated with HS of 40°C for 2 h by placing in a light incubator (GXZ-380C, Jiangnan Instrument Factory, Ningbo, China) and others under 26°C for 2 h were used as the control. The young leaves, stems and roots from seedling plants and the flower buds from adult plants with or without HS treatments were collected. For the dynamic expression analysis of *CaHsp20s* under HS, after being subjected to 40°C for 0, 0.5, 1, 2, 4 h, the leaves in B6 and R9 were collected, respectively. Temperature treatment was initiated during the light period and all the samples were harvested within the light period. These collected samples were frozen with liquid nitrogen for total RNA extraction and cDNA synthesis. Each treatment was conducted with three biological replicates, and samples from five plants were collected for each replicate.

RNA Isolation and Expression Analysis of Pepper *CaHsp20* Genes

Total RNA was extracted using Total RNA kit (BioTeke, Beijing, China) and reverse-transcribed using SuperScriptIII Reverse Transcriptase (Takara, Dalian, China), the operational procedure followed the manufacturer's procedure, and the cDNA was diluted to 50 ng/ μ L with ddH₂O for further examination. For expression analysis of pepper *CaHsp20s* in different tissues, semi-quantitative reverse-transcription (RT) PCR reactions were employed. Primer pairs (**Table S2**) were designed by Primer Premier 5.0 and tested by NCBI Primer BLAST. Ubiquitin binding protein gene *CaUBI-3* was used as the reference genes (Wan et al., 2011). PCR reactions for all genes were conducted with the following program: initial denaturation at 94°C for 5 min, followed with denaturation at 94°C for 30 s, annealing at 52°C for 30 s, extension at 72°C for 30 s with 30 cycles, and final extension at 72°C for 10 min. PCR products were detected by agarose gel electrophoresis with 1.5% gel concentration. The expression levels of *CaHsp20* genes in each sample were compared with the internal control gene, *CaUBI-3*.

Real-time quantitative PCR (qRT-PCR) was conducted using SYBR Green Supermix (Takara, Dalian, China) with a total volume of 20 μ L reaction system on the iQ5.0 Bio-Rad iCycler thermocycler (Bio-Rad, Hercules, CA, USA). The primers of *CaHsp20* genes were the same as those used for semi-quantitative RT-PCR. The qRT-PCR reactions were performed as described (Guo et al., 2014). The *CaUBI-3* gene was used as the internal reference gene. Three independent biological replicates were carried out and qRT-PCR of each replicate was performed in triplicate. The relative expression levels of pepper *Hsp20* genes were calculated as $2^{-\Delta\Delta CT}$ method (Schmittgen and Livak, 2008).

Search for *cis*-acting Elements in the Promoters of Pepper *CaHsp20* Genes

The upstream regions (0.5 kb) of the *CaHsp20* genes were derived from PGD, and were searched for regulatory elements, including HSEs (*cis*-acting element involved in heat stress responsiveness), TATA-box (core promoter element around -30 of transcription start), TC-rich repeats (*cis*-acting element involved in defense and stress responsiveness), W-box (binding site for the WRKY transcription factor in defense response), and CCAAT sequences (enhancer-binding protein factors) in the promoters were performed in PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>; Lescot et al., 2002) and PLACE database (Higo et al., 1999).

Results

Identification and Analysis of *CaHsp20* Gene Family Members in Pepper

A total of 41 sequences of *Hsp20* gene were gotten from PGD by HMM analysis, out of which 35 were annotated as being pepper *Hsp20* genes based on the confirmation of Pfam and SMART scans, and sequences without a complete domain (*CA09g15370*, *CA12g15460*, and *CA03g18690*) and sequences whose molecular weight out of the range of Hsp20s (*CA08g09970*, *CA04g23450*, and *CA00g46140*) were removed. The 35 *CaHsp20* genes were named according to their molecular weight (**Table 1**).

The length of *CaHsp20* proteins ranged from 141 (*CaHsp16.4*) to 313 amino acids (*CaHsp35.5*), and the predicted molecular weights of the *CaHsp20*s were between 16.33 kDa (*CaHsp16.3*) and 35.53 kDa (*CaHsp35.5*). The *CaHsp20*s shared a conserved ACD comprised about 90 amino acids, in which *CaHsp24.3b* owned the shortest ACD with 74 amino acids, while *CaHsp16.3*'s ACD (116 amino acids) was the longest. The predicted pI-values of *CaHsp20* proteins ranged from 5.07 (*CaHsp18.4a*) to 9.52 (*CaHsp26.5*), and the predicted instability indexes showed that only nine of the 35 *CaHsp20*s were stable proteins (instability index = 40). In addition, it was also found that 16 (45.71%) of the total 35 *CaHsp20* genes had no introns, 18 genes (51.43%) had a single intron, while only *CaHsp27.6* had five introns.

Characterization of the Deduced Amino Acid Sequences of *CaHsp20*s

Ten consensus motifs were found in *CaHsp20* proteins by MEME motif search tool (**Figure 1**; **Table 2**). Based on the analysis of

Pfam, SMART, and InterProScan, the full sequences of motif 1, 2, 3, and 4 formed the highly conserved ACD (**Figure 2A**), in which the combined sequence of motif 2 and 4 contained the CRI of ACD, while the combined sequence of motif 3 and 1 contained the $\beta 6$ -loop and CRII. The four motifs were detected in majority of the *CaHsp20* proteins, while only motif 2 was found in *CaHsp16.7* and *CaHsp30.1*, and *CaHsp16.6b* did not contain motif 2. Motif 6 was detected in *CaHsp22.1*, 23.8, and 26.5, and motif 8 was found behind the motif 5 in *CaHsp18.1a*, 18.1b, 18.2a, and 18.2b, while motif 10 were only found in the front of motif 2 in *CaHsp17.7a*, 17.7b, and 25.8 (**Figure 1**). Only motif 6 was found in the C-terminal regions, other motifs such as motif 5, 7, 8, 9, and 10 were found mainly in the N-terminal regions. It was worth noting that motif 9 (the signal-peptide sequence) was found in *CaHsp18.7*, 21.1, 21.2, 22.0, and 22.6, which was consistent with their prediction of secretory pathway localization (**Table 1**).

The ACD was formed by two conserved regions, CRI (motif 2 and 4) with $\beta 2$, $\beta 3$, $\beta 4$, and $\beta 5$ and CRII (part of motif 3 and 1) with $\beta 7$, $\beta 8$, and $\beta 9$, which were separated by a hydrophilic domain $\beta 6$ -loop (part of motif 3) with variable length. The multiple alignments of the conserved ACD domains, spanning ~ 90 amino acids in *CaHsp20* proteins was clearly shown in **Figure 2B**. *CaHsp16.6b*, 16.7, 24.3b, 27.6, 28.4, 30.1, and 35.5 either lacked the $\beta 6$ -loop or contained sequences with unknown function between $\beta 5$ and $\beta 7$. The shorter ACDs of *CaHsp18.4b*, 24.3b, and 28.4 lacked the $\beta 2$ (**Figure 2B**; **Table 1**), which might show the functional diversity and remains to be fully elucidated.

Phylogenetic Analysis of *CaHsp20* Family

To analyze the evolutionary relationship of the *Hsp20* gene families in pepper, tomato, *Arabidopsis*, rice, and maize, the amino acid sequences of 154 *Hsp20* proteins, including 35 from pepper, 24 from tomato, 25 from *Arabidopsis*, 35 from rice, and 35 from maize were aligned and used to construct an unrooted phylogenetic tree. As shown in **Figure 3**, 130 *Hsp20*s could be grouped into 12 distinct subfamilies, containing 47 CIs (cytosol/nucleus I), 27 CIIIs, 6 CIIIs, 1 CIVs, 5 CVs, 3 CVIs, 4 CVIIs, 5 MIs (mitochondria I), 4 MIIs, 19 Ps (plastid), 4 ERs (endoplasmic reticulum), and 5 Pos (peroxisomal), respectively. However, the left 24 proteins from pepper, *Arabidopsis*, rice and maize were not well-clustered into any subfamily. The *Hsp20* members from the five species were more closely to those in the same class in different species than to others in the same species. Many *Hsp20*s from the five species were clustered into class CI and CII, including 9 and 10 *CaHsp20* proteins, respectively, which indicated that cytosol might be the main function area for *CaHsp20*s. The CI subfamily had two subclades that consisted of monocotyledon *Hsp20*s (*OsHsp20*s and *ZmHsp20*s) and two subclades that consisted of dicotyledon *Hsp20*s (*CaHsp20*s, *SlHsp20*s, and *AtHsp20*s), suggesting a single gene duplication took place prior to monocotyledon-dicotyledon bifurcation (Bondino et al., 2012). It was interesting that the members of P and M (MI and MII) subfamilies were closely related to each other, which was consistent with the *Hsp20* of M subfamily evolved later from the P subfamily

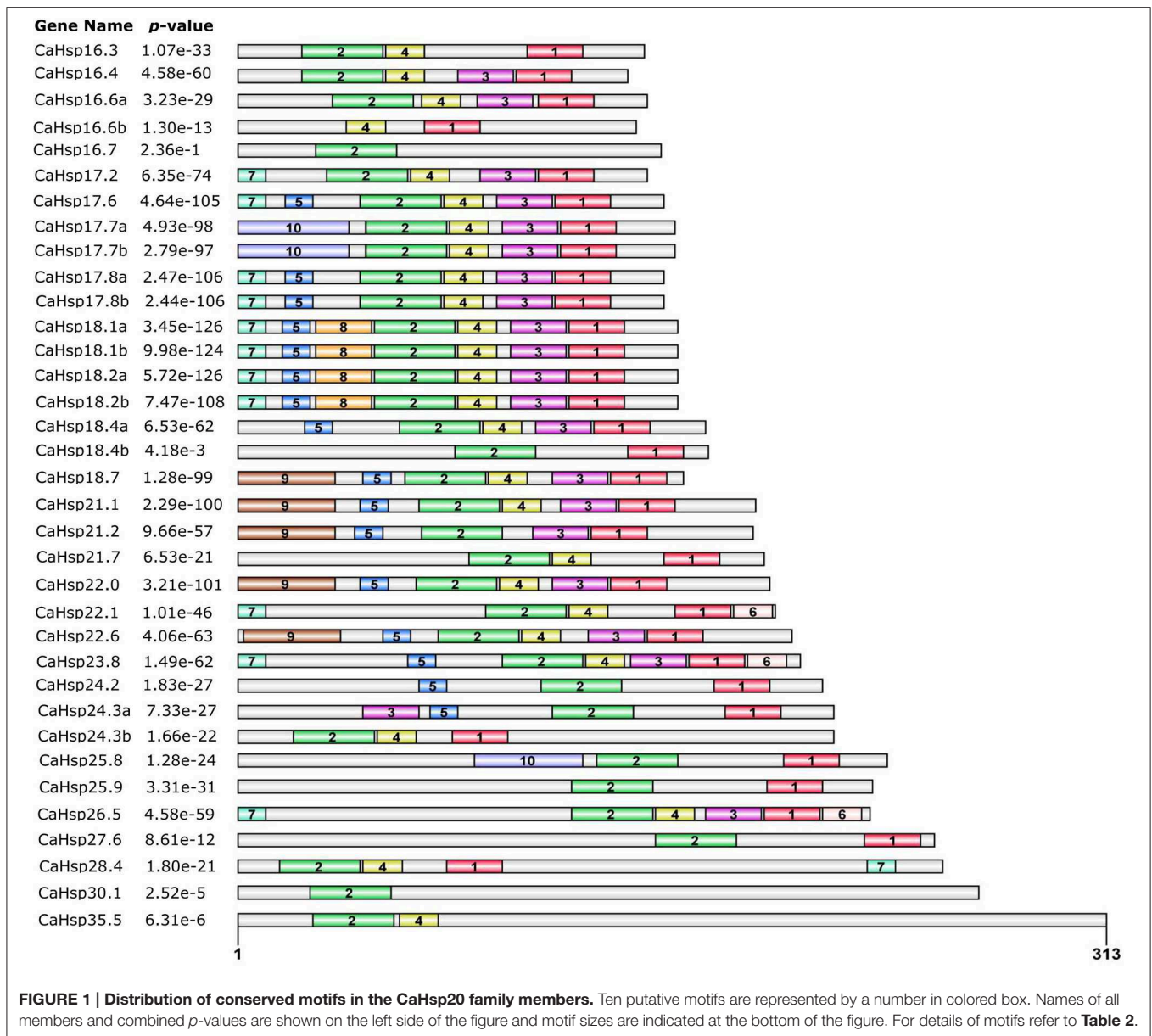
TABLE 1 | The list of *CaHsp20* members identified.

No.	Name	Gene ID	Chr.	ORF	Introns	AA	ACD	WT	pI	Localization predicted	Instability index
1	CaHsp16.3	CA05g01890	5	444	1	147	29–144	16.33	6.84	cyto, mito	47.58
2	CaHsp16.4	CA02g30320	2	426	1	141	29–138	16.35	8.47	Secretory pathway	33.80
3	CaHsp16.6a	CA03g37070	3	447	1	148	40–145	16.55	5.79	cyto, extr	65.53
4	CaHsp16.6b	CA10g22470	10	435	1	144	16–97	16.55	9.04	cyto, extr, E.R.	14.10
5	CaHsp16.7	CA04g23370	4	462	1	153	35–130	16.66	6.29	chlo, extr	47.17
6	CaHsp17.2	CA09g08950	9	447	0	148	38–147	17.22	5.66	cyto	37.04
7	CaHsp17.6	Capana06g000385▲	6	465	0	154	50–153	17.60	5.83	cyto	49.55
8	CaHsp17.7a	Capana03g001238▲	3	477	0	158	52–156	17.73	6.17	cyto, nucl	39.33
9	CaHsp17.7b	Capana03g001241▲	3	477	0	158	52–156	17.68	6.75	cyto, nucl	36.08
10	CaHsp17.8a	CA06g24940	6	465	0	154	50–153	17.76	5.83	cyto	52.77
11	CaHsp17.8b	CA06g24930	6	465	0	154	50–153	17.78	6.00	cyto	53.63
12	CaHsp18.1a	CA08g17060	8	480	0	159	55–158	18.12	5.58	cyto	68.08
13	CaHsp18.1b	CA08g17070	8	480	0	159	55–158	18.09	5.84	cyto	71.86
14	CaHsp18.2a	CA08g17040	8	480	0	159	55–158	18.22	5.39	cyto, nucl	69.52
15	CaHsp18.2b	CA08g17050	8	480	0	159	55–157	18.18	8.83	cyto	58.82
16	CaHsp18.4a	Capana00g002751▲	4	510	1	169	64–168	18.43	5.07	cyto	56.93
17	CaHsp18.4b	Capana01g000571▲	1	513	1	170	89–163	18.43	5.68	chlo	37.86
18	CaHsp18.7	CA06g20260	6	486	1	161	66–160	18.73	5.08	Secretory pathway	58.55
19	CaHsp21.1	Capana06g000831▲	6	561	0	187	71–168	21.12	6.36	Secretory pathway	42.46
20	CaHsp21.2	Capang01g004332★	8	558	0	186	72–167	21.20	5.20	Secretory pathway	46.21
21	CaHsp21.7	CA07g18820	7	573	1	190	89–185	21.67	5.19	cyto, pero	38.57
22	CaHsp22.0	CA03g27140	3	579	0	192	70–173	22.03	6.05	Secretory pathway	48.98
23	CaHsp22.1	Capang10g001114★	10	585	1	194	96–187	22.11	9.14	chlo	45.05
24	CaHsp22.6	CA05g03530	5	603	0	200	78–186	22.63	5.37	Secretory pathway	38.03
25	CaHsp23.8	CA09g06120	9	612	1	203	102–203	23.83	7.15	cyto, nucl	54.99
26	CaHsp24.2	CA01g05030	1	636	1	211	115–210	24.15	5.18	cyto, chlo, mito	61.58
27	CaHsp24.3a	CA09g08090	9	648	0	215	119–214	24.27	7.86	chlo, cyto, mito	68.39
28	CaHsp24.3b	Capana01g001839▲	1	648	1	215	35–108	24.32	9.21	cyto, golg, nucl	42.06
29	CaHsp25.8	CA03g21390	3	705	1	234	135–233	25.78	6.53	chlo	40.04
30	CaHsp25.9	CA01g11490	1	690	1	229	128–228	25.93	8.36	chlo, mito	40.47
31	CaHsp26.5	CA10g10840	10	687	0	228	128–227	26.49	9.52	mito	50.30
32	CaHsp27.6	CA07g15940	7	756	5	251	156–250	27.63	8.99	chlo, nucl	54.52
33	CaHsp28.4	CA11g15400	11	765	1	254	30–109	28.35	8.35	golg, nucl, cyto	40.24
34	CaHsp30.1	Capana12g001708▲	12	804	1	267	35–127	30.14	8.26	Secretory pathway	47.98
35	CaHsp35.5	CA04g13680	4	942	1	313	35–129	35.53	7.70	cyto, nucl	33.82

ORF, open reading frame; AA, amino acid; ACD, alpha-crystallin domain; Mol. Wt., molecular weight (kDa); pI, isoelectric point. Sequenced IDs marked with triangle are from *Zunla-1* genome, with pentagram are from *glabriusculum* genome, others are from *CM334* genome.

(Waters, 2013). It was noteworthy that no Hsp20 proteins from monocotyledons (rice and maize) were clustered into CIV cluster, which was consistent with the result suggested by Siddique et al. (2008) that CIV members of Hsp20 proteins were exclusively found in dicotyledonous. For the three species of dicotyledon, 12 pairs of orthologs (11 pairs of pepper/tomato orthologs and one pair of pepper/*Arabidopsis* ortholog) were identified, indicating that the ancestral genes of Hsp20 family might have existed before differentiation of *Arabidopsis*, tomato and pepper species. In addition, several pairs of paralog were found in all of the three Hsp20 families from pepper, *Arabidopsis* and tomato, suggesting that *CaHsp20* genes underwent many rounds of duplication events occurred after the pepper speciation.

It was found that the phylogenetic classification of Hsp20 family was consistent with the pattern of intron type (pattern 1 had no intron, pattern 2 had one intron, and pattern 3 had more than one intron) in previous report (Ouyang et al., 2009). The most *CaHsp20* members of CI and ER subfamilies had no intron, and the members of CIII, CVII, P, and MII showed one intron (Figure 3; Table 1). However, six members belonged to CII subfamily showed no intron, other four members of CII subfamily had one intron. It was noteworthy that *CaHsp16.3* belonged to Po subfamily had one intron (pattern 2), which was different from the Po subfamily from rice (pattern 1; Ouyang et al., 2009). Among the unclassified *CaHsp20s*, *CaHsp27.6* was an orphan gene with five introns (pattern 3), which indicated particular phylogenetic status.



Chromosomal Localization and Gene Duplication of *CaHsp20* Genes

The 35 *CaHsp20* genes were distributed on all 12 pepper chromosomes (**Figure 4**; **Table 1**). Among these 35 *CaHsp20* genes, five (14.3%) genes were located on each of chromosome 3, 6, and 8, four genes (11.4%) on chromosome 1, three genes on each of chromosome 4, 9, 10, two genes on each of chromosome 5 and 7, and only one gene on each of chromosome 2, 11, and 12. As gene duplication events involve in the evolution of plant genomes, the duplications of *CaHsp20s* were also analyzed. As shown in **Figure 4**, one segmental duplication event was detected between *CaHsp24.2* on chromosome 1 and *CaHsp24.3a* on chromosome 9, and three tandem duplication events (*CaHsp17.7a/17.7b*, *CaHsp17.6/17.8a/17.8b* and *CaHsp18.1a/18.1b/18.2a/18.2b*) were identified on

chromosome 3, 6, and 8, respectively. However, *CaHsp18.7*, *21.1*, and *22.0* were likely to originate from the same ancestor gene, which underwent one segmental duplication and one tandem duplication event during the pepper chromosome evolution. These results suggested that both tandem and segmental duplication events made an important contribution to pepper *CaHsp20* genes expansion.

The Interaction Network of *CaHsp20* Genes

To further analyze the relationship between *CaHsp20* genes and other genes, an interaction network of *CaHsp20* genes was built according to the orthologous to *Arabidopsis* (**Figure 5**). There were 24 (68.6%) out of 35 *CaHsp20* genes orthologously related to 13 (68.4%) out of 19 *Arabidopsis Hsp20* genes. The red, blue, and green lines stand for Pearson correlation coefficients

TABLE 2 | Motif sequences identified by MEME tools.

Motif	Length (aa)	Sequence
1	21	LPENADMDDQIKAHMENGVLTV
2	30	ANARWDWKETPEAHIFKADMPGYKKEDVKV
3	21	KKNDQWHRMERSGGKFWRRFR
4	15	QVEEGRVLQISGERK
5	11	DPFSMDLWDPF
6	15	TVPKEEVKKPDVKAI
7	11	MSLIPSFSGGR
8	21	GFPISSTIANTPSSARETSAF
9	36	MVKATVSLISFLVLAIVMAFLPSKNEALIPYTRPTW
10	41	MDFRLMGMDNTPIFHVLQHMMMDTAGDESCKSVNAPS RNYVR

Motif numbers corresponded to the motifs in **Figure 1**.

>0, <0, and not calculated, which were corresponding to positive correlation (136 pairs of interacting genes), negative correlation (99 pairs), and unclear correlation (106 pairs), respectively. The interaction network of *CaHsp20* genes showed a complex functional relationship. For example *CaHsp24.2* directly or indirectly interacted with many other *CaHsp20* genes (*CaHsp17.8a*, *17.8b*, *22.1*, *23.8*, *26.5*, *18.1a*, *18.1b*, *18.2a*, and *18.2b*), and also had more interacted genes than other *CaHsp20s*, while only one gene (*CA03g01490*) interacted with *CaHsp25.8*. In addition, *CaHsp18.1a*, *18.1b*, *18.2a*, and *18.2b* genes interacted with heat stress related genes (heat shock protein 70 family genes *CA03g30260* and *CA01g31330*), *CaHsp17.7a* and *17.7b* genes interacted with heavy metal stress related gene (heavy metal transport/detoxification superfamily protein gene *CA02g24070*), *CaHsp16.3* interacted with signal transduction related gene (calcium-binding EF-hand family protein gene *CA10g18540*), and *CaHsp25.9* with transcription factor (MADS-box transcription factor family protein gene *CA00g74550*), which might indicate that *CaHsp20* genes involved in many response mechanisms by regulating their downstream genes or being regulated by their upstream regulators in pepper.

Expression Patterns of *CaHsp20* Genes in Different Tissues

To primarily investigate the biological functions of individual *CaHsp20* members in pepper, the heat-induced expression of these *CaHsp20* genes were examined by semi-quantitative RT-PCR in four major tissues (root, stem, leaf, and flower) from thermotolerant line R9 (**Figure 6**). Under normal temperature condition, 27 of the 35 *CaHsp20* genes showed no significant tissue specificity as they were barely expressed in the tested tissues, however, the left eight *CaHsp20* genes, including *CaHsp16.3*, *16.6b*, *16.7*, *17.7a*, *17.7b*, *18.1b*, *18.2a*, and *23.8* could be detected at least in one tissue. Twenty-two (62.9%) *CaHsp20* genes were highly induced in at least one tissue after HS treatment (40°C for 2 h). For example *CaHsp16.3*, *16.4*, *17.6*, and *18.1b* were expressed at relatively high levels in all of tissues, while *CaHsp16.7* and *23.8* were highly induced by HS only in flower or leaf, respectively. Many heat-induced *CaHsp20s*, such as *CaHsp16.6a*, *17.6*, *21.1*, *21.2*, *22.0*, *24.3a*, and *25.8* were

expressed at higher level in treated pepper leaves than other three tissues, however, the highest levels of *CaHsp16.3* and *22.6* were detected in flowers with HS treatment, and the expression level of *CaHsp17.7a* was highest in treated roots. The transcripts of 12 (34.3%) genes (*CaHsp17.8a*, *17.8b*, *18.4a*, *18.4b*, *21.7*, *22.1*, *24.3b*, *26.5*, *27.6*, *28.4*, *30.1*, and *35.5*) were not visible in the both of the normal temperature and HS conditions.

Expression Profiles of *CaHsp20* Genes in Response to Heat Stress

In order to investigate the heat response profile for *CaHsp20* genes, the transcript patterns of each *CaHsp20* in leaves from thermosensitive line B6 and thermotolerant line R9 with continuous heat stress treatment (40°C for 4 h) were compared. As shown in **Figure 7**, strong response to HS in all of the 35 *CaHsp20* genes from both lines (B6 and R9) were observed, in which a majority of these genes were up-regulated by HS treatment. For example *CaHsp16.4*, *18.7*, *22.0*, *25.8*, and *25.9* were highly induced after a short-term HS (40°C for 0.5 h) and their high expression levels were maintained until the end of the HS treatment, while only *CaHsp21.7* and *28.4* were down-regulated by HS in R9. Generally, the peaks of relative expression levels of *CaHsp20* genes (compared with HS for 0 h) in thermosensitive line B6 treated with HS were higher than those of thermotolerant line R9, however the appearance of peak values were observed at 2 or 4 h in B6 but at 0.5 h in R9 after HS treatment. In addition, it was worth noting that compared with other *CaHsp20* members, *CaHsp16.4*, *18.7*, *22.0*, *25.8*, and *25.9* which were significantly up-regulated in B6 leaves treated with HS, however, no main heat-response genes was detected in line R9 after HS treatment, although many *CaHsp20* genes were induced by HS in R9.

Analysis of Stress-related *cis*-elements in the *CaHsp20* Promoters

To further explore the possible regulation mechanism of *CaHsp20* genes in the HS response of pepper, the *cis*-elements involving in the activation of defense-related genes in the promoter regions of *CaHsp20s* were also scanned, and the promoter regions (–500 bp upstream of the translation start site) of a total of 30 *CaHsp20* genes (promoter regions of *CaHsp18.4a*, *21.1*, *21.2*, *22.1*, and *28.4* were absent) from PGD were used (**Figure 8**). One to three HSE elements (heat stress responsiveness) were found in the promoter regions of 12 genes (*CaHsp16.3*, *16.4*, *17.6*, *17.8a*, *17.8b*, *18.1a*, *18.2a*, *22.0*, *25.8*, *25.9*, *26.5*, and *27.6*), respectively, in which the number of HSE elements in the promoter region of *CaHsp17.8b* was the maximum. Although HSE elements could not be found in –500 bp promoter regions of other 11 *CaHsp20* genes (*CaHsp16.7*, *17.2*, *17.7b*, *18.1b*, *18.4b*, *21.7*, *23.8*, *24.2*, *24.3b*, *30.1*, and *35.5*), 1–3 HSE elements were detected in their –1500 bp promoter regions, however, no any HSE element was present in either –1500 or –500 bp promoter regions of the rest seven *CaHsp20s* (*CaHsp16.6a*, *16.6b*, *17.7a*, *18.2b*, *18.7*, *22.6*, and *24.3a*). In addition, other stress-related *cis*-elements were detected, including TATA-boxes in all 30 selected promoter regions, CCAAT-motif in 23 genes, TC-rich repeats in seven

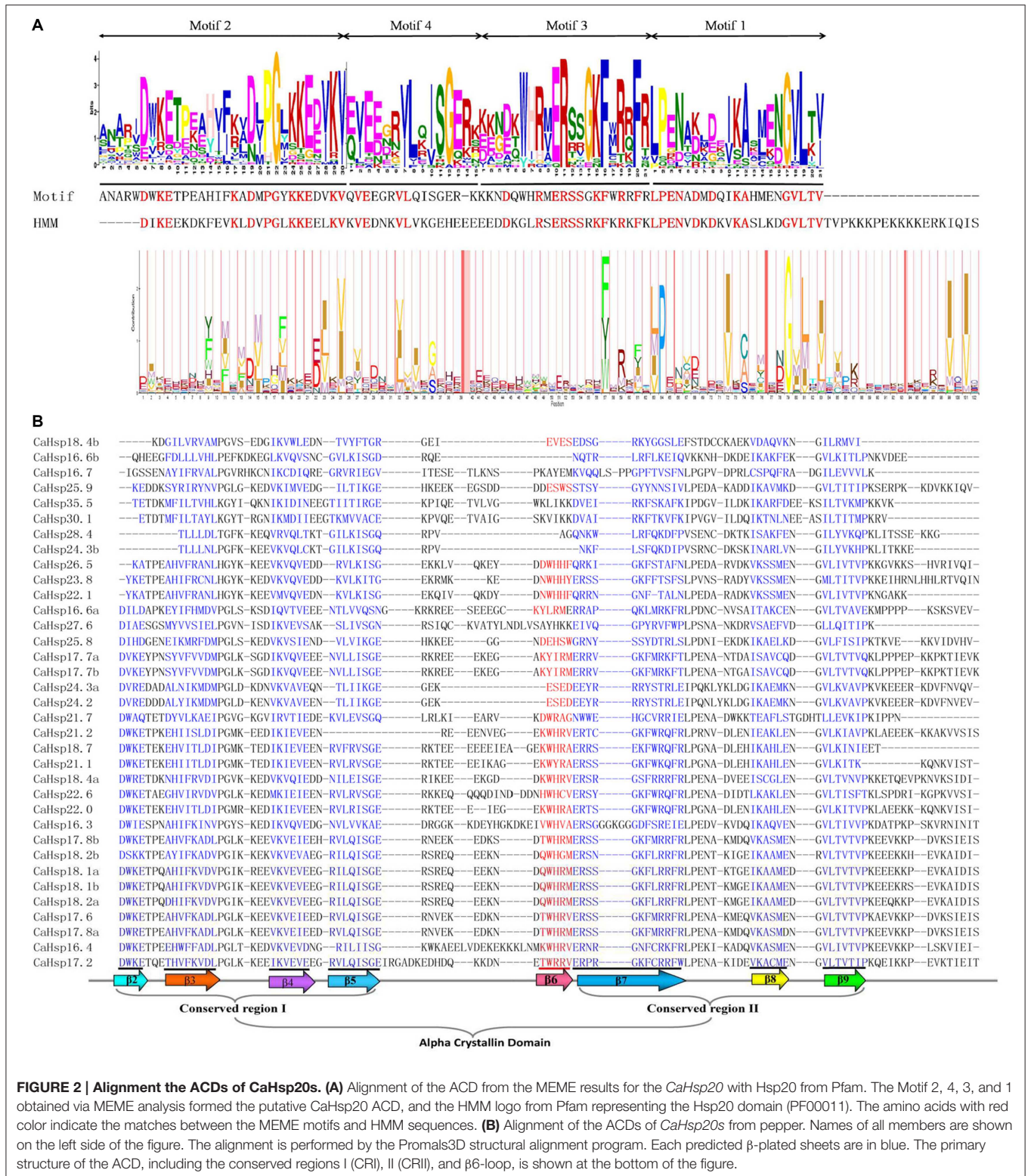


FIGURE 2 | Alignment the ACDs of CaHsp20s. (A) Alignment of the ACD from the MEME results for the *CaHsp20* with Hsp20 from Pfam. The Motif 2, 4, 3, and 1 obtained via MEME analysis formed the putative CaHsp20 ACD, and the HMM logo from Pfam representing the Hsp20 domain (PF00011). The amino acids with red color indicate the matches between the MEME motifs and HMM sequences. **(B)** Alignment of the ACDs of *CaHsp20s* from pepper. Names of all members are shown on the left side of the figure. The alignment is performed by the Promals3D structural alignment program. Each predicted β -plated sheets are in blue. The primary structure of the ACD, including the conserved regions I (CRI), II (CRII), and β 6-loop, is shown at the bottom of the figure.

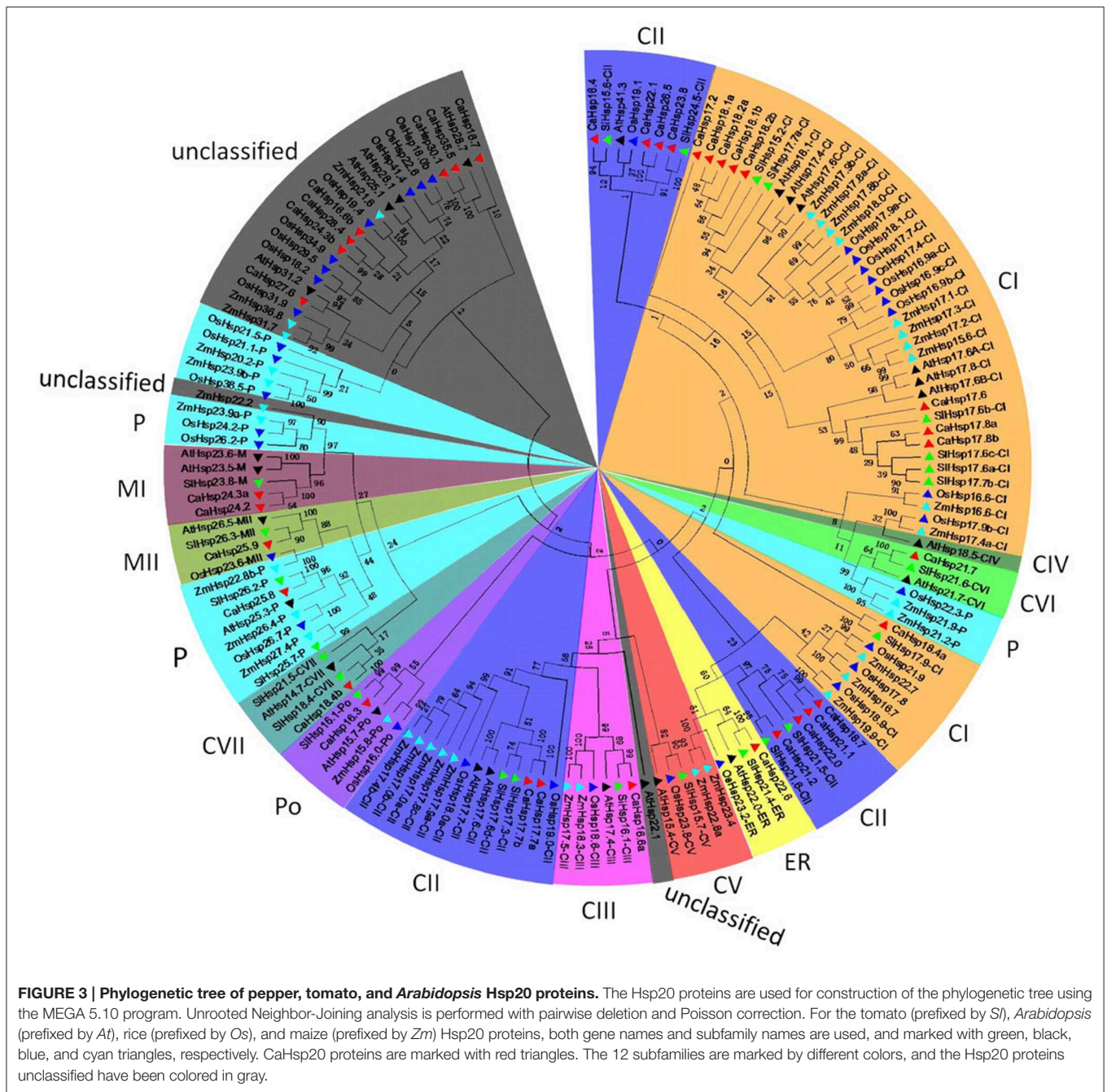


FIGURE 3 | Phylogenetic tree of pepper, tomato, and Arabidopsis Hsp20 proteins. The Hsp20 proteins are used for construction of the phylogenetic tree using the MEGA 5.10 program. Unrooted Neighbor-Joining analysis is performed with pairwise deletion and Poisson correction. For the tomato (prefixed by *Sl*), *Arabidopsis* (prefixed by *At*), rice (prefixed by *Os*), and maize (prefixed by *Zm*) Hsp20 proteins, both gene names and subfamily names are used, and marked with green, black, blue, and cyan triangles, respectively. *CaHsp20* proteins are marked with red triangles. The 12 subfamilies are marked by different colors, and the Hsp20 proteins unclassified have been colored in gray.

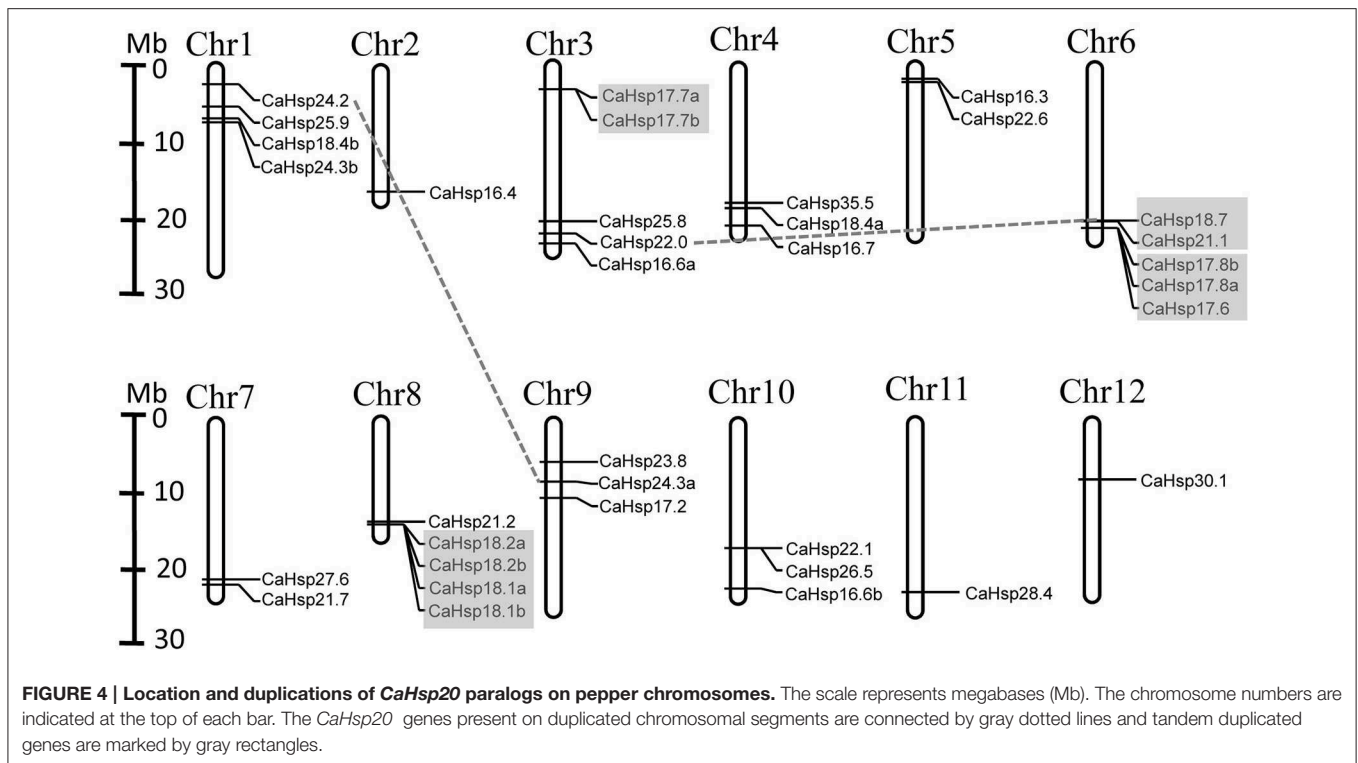
genes (*CaHsp17.8a*, *18.1b*, *18.2a*, *18.2b*, *18.7*, *30.1*, and *35.5*), and W-box in three genes (*CaHsp18.4b*, *18.7*, and *25.8*).

Discussion

It has been demonstrated that Hsp20s can prevent the irreversible aggregation of denaturing proteins and enhance plant-acquired thermotolerance (Wang et al., 2004; Haslbeck and Vierling, 2015). Benefiting from the availability of whole genome sequence in recent years, several Hsp20 families have been identified from *Arabidopsis*, rice, soybean, and *Populus trichocarpa* (Waters et al.,

2008; Lopes-Caitar et al., 2013), however, very little is known about this family in pepper.

The present study describes the identification of a total of 35 *CaHsp20* genes from pepper, investigation of their structure, chromosomal organization, evolutionary relationships, prediction of interaction network, and expression patterns across the different tissues and under heat stress condition, which provides a basis for comprehensively understanding the *CaHsp20* gene family in pepper and makes a base for future functional divergence of the *Hsp20* genes in the Solanaceae family.

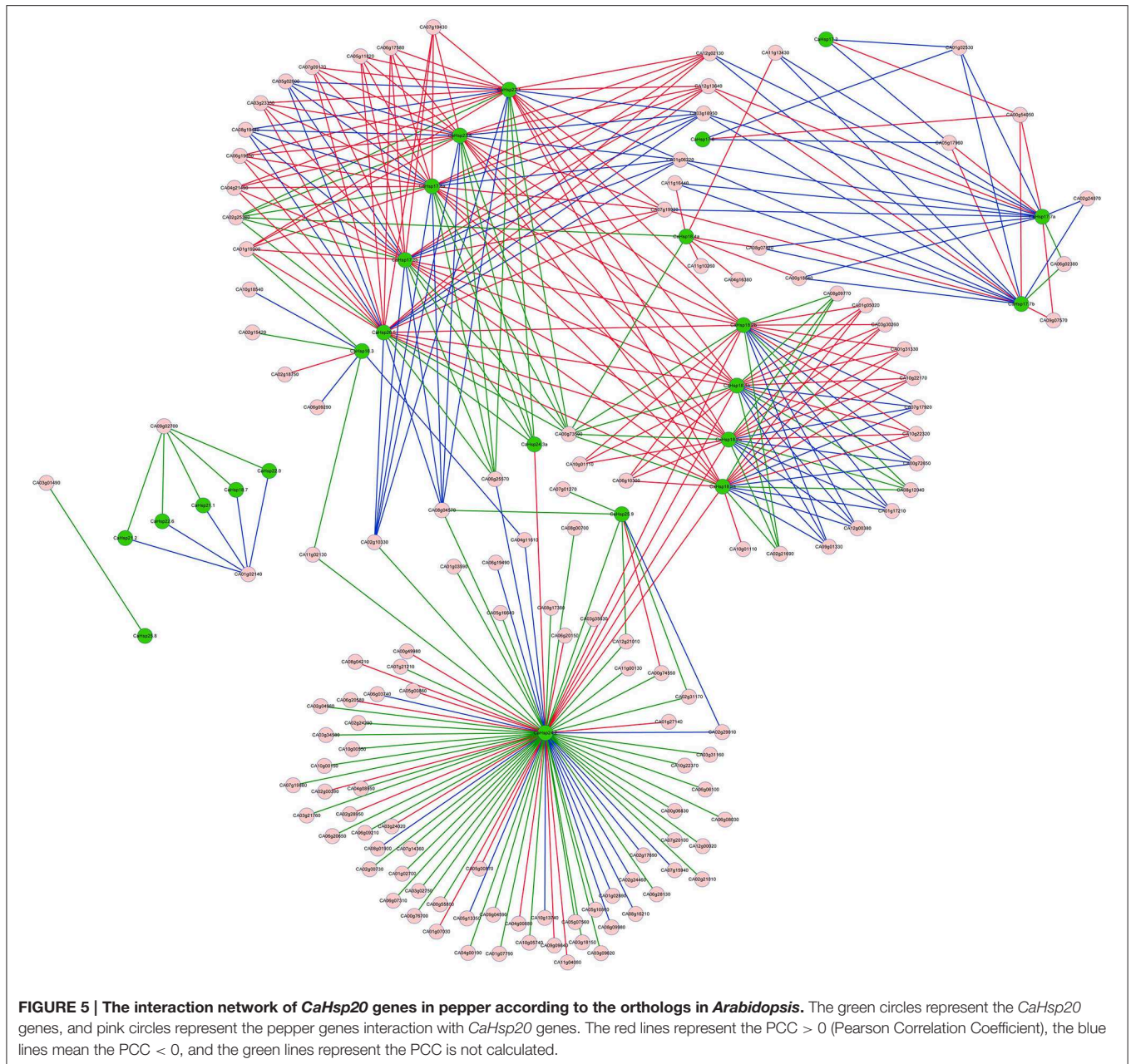


Although having the most diversity among the molecular chaperone families, the Hsp20s share a conserved structural domain, the ACD flanked by a non-conserved N-terminal sequence of variable length and a conserved short C-terminal sequence (Kriehuber et al., 2010; Bondino et al., 2012). Characterization and multiple sequence alignments divulged that all selected 35 *CaHsp20*s contained the ACD with various sequences. The CRI (motif 2 and 4) consisted of $\beta 2$, $\beta 3$, $\beta 4$, and $\beta 5$ and CRII (part of motif 3 and 1) consisted of $\beta 7$, $\beta 8$, and $\beta 9$ in ACD were structurally conserved, containing the signature typical residues Pro-X(14)-Gly-Val-Leu and Pro-X(14)-X-Val/Leu/Ile-Val/Leu/Ile, respectively (Figures 1, 2B; Caspers et al., 1995; Ouyang et al., 2009). Most of the *CaHsp20* ACDs also contain the $\beta 6$ sheet, the non-conserved motif connecting $\beta 5$ in CRI and $\beta 7$ in CRII (Bondino et al., 2012), while it is replaced by sequences of unknown function in four *CaHsp20* members (*CaHsp16.7*, 27.6, 30.1, and 35.5) or absent in three members (*CaHsp16.6b*, 24.3b, and 28.4). Plant Hsp20 proteins of CIV clusters lack the $\beta 6$ sheet (Waters, 2013) but more interesting is that none of the Hsp20 members from pepper, tomato, rice, and maize are found in the CIV cluster, while found in *Arabidopsis* Hsp20 family, in addition, the above seven *CaHsp20* genes are not clustered into any subfamily (Figure 3), it can be inferred that these genes from pepper belong to CIV cluster or form a new cluster. Because the $\beta 6$ sheet is crucial for dimer formation and oligomerization, whether these *CaHsp20* genes lacking of $\beta 6$ sheet can function as a molecular chaperone still need further comparative studies.

Gene organization also plays an important role in the evolution of multiple gene families and in response to stress

conditions (Xu et al., 2012). Sixteen *CaHsp20* genes are intronless (pattern 1), 18 genes contain one intron (pattern 2), only one gene (*CaHsp27.6*) contains five introns (pattern 3; Table 1). Most *CaHsp20* members of CI and ER subfamilies have no intron (pattern 1) and the members of MII subfamily have one intron (pattern 2; Figure 3; Table 1), which is consistent with rice *Hsp20* family, while the patterns of *CaHsp20* members and *Hsp20*s from rice belong to CII, Po and P subfamilies are not the same (Ouyang et al., 2009), indicating that although intron patterns shed light on the phylogenetic relationship, there is difference between different species. Most of these *CaHsp20* genes are highly induced by HS in different tissues (Figure 6), which may indicate that the expression levels of genes without introns or with a short length intron will be enhanced under HS in plants (Lopes-Caitar et al., 2013). The compact genes have faster response to HS and protect plants from HS damage better than those genes contain many introns, and the compact genes are gradually selected by stress conditions during the long-term evolution (Jeffares et al., 2008), which may be the reason that *Hsp20* family members form a first line of defense against protein aggregation caused by HS condition in order to maintain protein homeostasis and enhance the plant thermotolerance (Haslbeck and Vierling, 2015).

Gene duplication events play a significant role in expansions of gene family members and evolutionary mechanisms of genomes (Vision et al., 2000). Tandem duplications and segmental duplications are the major evolutionary patterns (Kong et al., 2007). In this analysis, 24 *Hsp20* genes were identified in tomato, however, in pepper with a genome size four-fold larger than that of tomato, only 35 *CaHsp20* genes were



found. The reason for this might be the whole-genome duplication (WGD) events in pepper and tomato (Qin et al., 2014). A total of 14 *CaHsp20* duplicated genes were found in pepper chromosomes, including one segmental duplication event (*CaHsp24.2/CaHsp24.3a*) and three tandem duplication events (*CaHsp17.7a/17.7b*, *CaHsp17.6/17.8a/17.8b* and *CaHsp18.1a/18.1b/18.2a/18.2b*; **Figure 4**), contributing to evolution of MI, CI and CII genes. Besides, *CaHsp18.7*, *21.1*, and *22.0* experienced one segmental duplication and one tandem duplication event, also contributing to evolution of CII genes. In previous report, the CI subfamily in plant including tomato, is always the largest (Waters, 2013), however the number of *CaHsp20* members belong to CII subfamily

is largest (**Figure 3**), which is not consistent with that of the other plants. The reason may be attributed to *CaHsp20* genes of CII subfamily appeared after speciation through multiple gene duplication. Due to the opinion that gene duplication in tandem mainly occurs in the genes encoded membrane proteins and stress-related proteins, we can suggest that the dominant tandem duplications in *CaHsp20* genes may be related to the fact that *CaHsp20s* are regulated by HS (Cannon et al., 2004; Rizzon et al., 2006; Hanada et al., 2008; Giorno et al., 2012).

The tissue-specific expression patterns of *Hsp20* genes have been reported in various species, such as *Arabidopsis*, rice, and soybean, however, there is no uniform pattern of gene expression

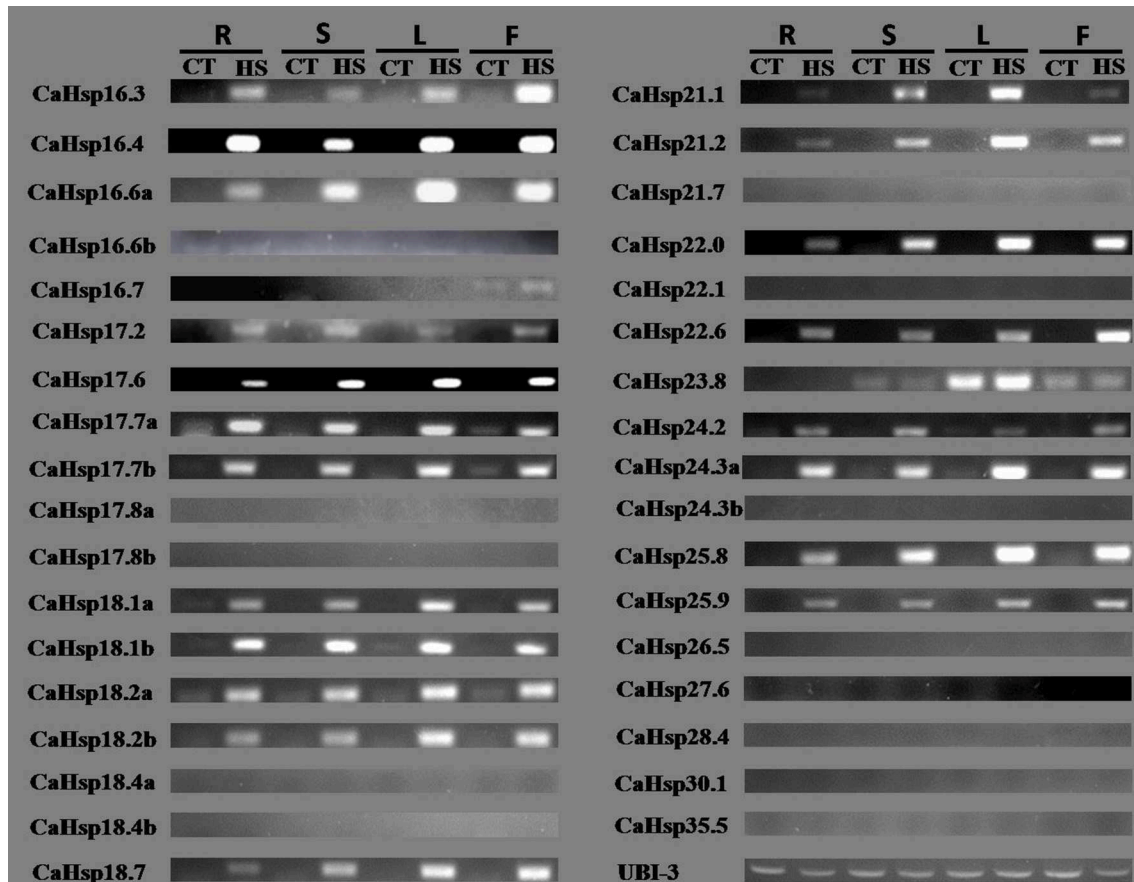
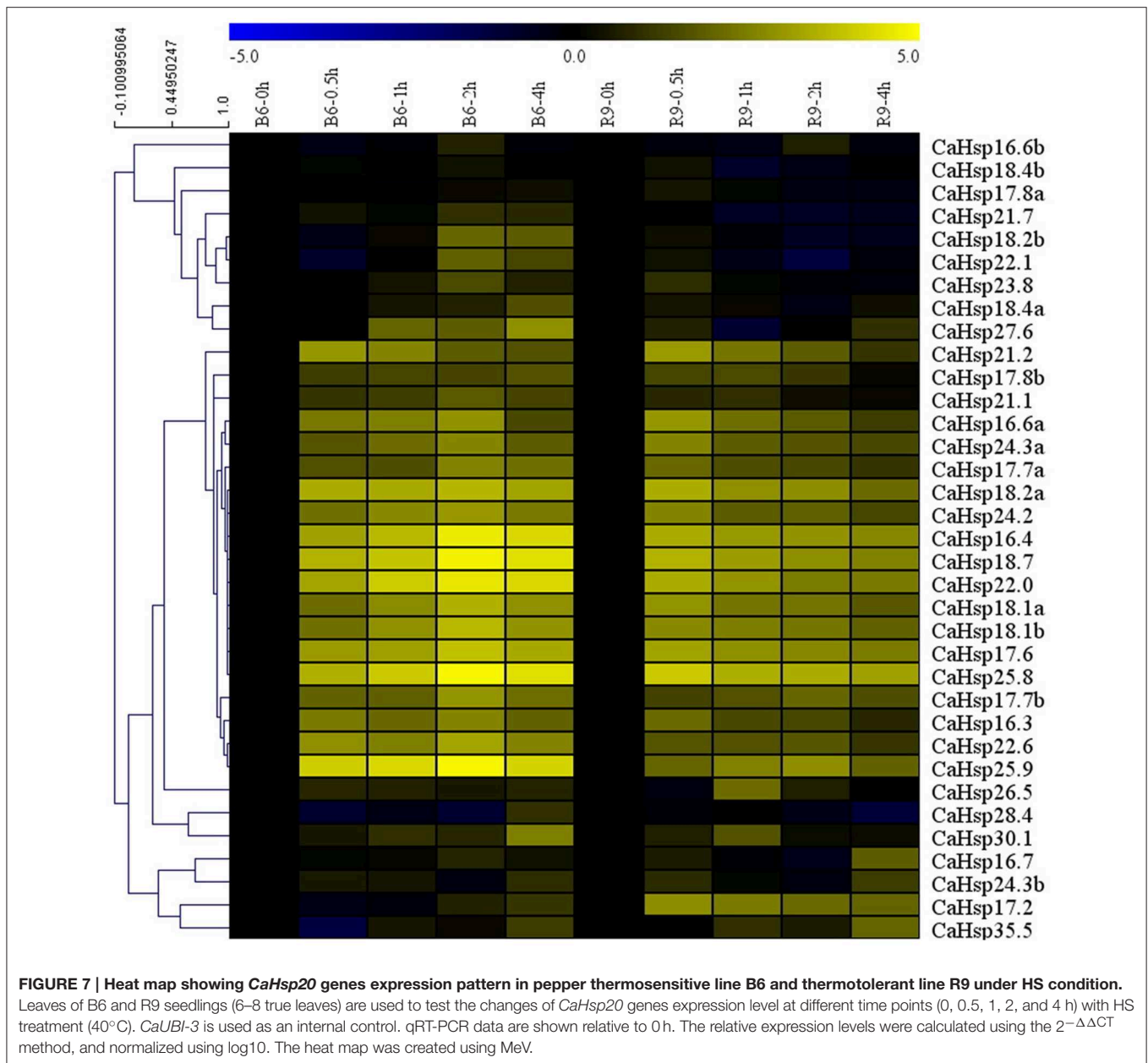


FIGURE 6 | RT-PCR analysis of expression pattern of 35 *CaHsp20* genes in four pepper tissues under normal and HS conditions. R, root; S, stem; L, leaf; F, flower. Normal condition: 26°C for 2 h; HS condition: 40°C for 2 h. The pepper ubiquitin-conjugating protein gene (*CaUBI-3*) is used as an internal control.

for all the plant *Hsp20* genes (Lopes-Caitar et al., 2013; Waters, 2013). Most *CaHsp20* genes have no or low expression in the four examined tissues (root, stem, leaf, and flower) from pepper thermotolerant line R9 under normal temperature condition, but are strongly induced by HS (Figure 6), which is largely consistent with the previous reported that almost all the plant *sHsps* are heat induced. Among the 35 *CaHsp20* genes, *CaHsp23.8* possesses the most distinctive expression pattern compared with other *CaHsp20* genes, which expresses constitutively in stems, leaves and flowers, but almost no induction by HS. Waters (2013) reported that *AtHsp21.7CV* gene lacking chaperone activity from *Arabidopsis* is constitutively expressed in shoots and not up-regulated by HS, therefore, it can be assumed that *CaHsp23.8* also may lack the chaperone activity, but it needs more scientific evidence. In addition, the expression of *CaHsp16.6b*, *16.7*, *17.7a*, *17.7b*, and *18.2a* were detected in various tissues under normal condition, we can suggest that these genes exhibit specific housekeeping activities as with some *Hsp20* genes from rice and soybean (Ouyang et al., 2009; Lopes-Caitar et al., 2013).

It has been demonstrated that *sHsps* are associated with thermotolerance in various plant species, however, many *Hsp20s* exist in plants which may lead to function redundancy of

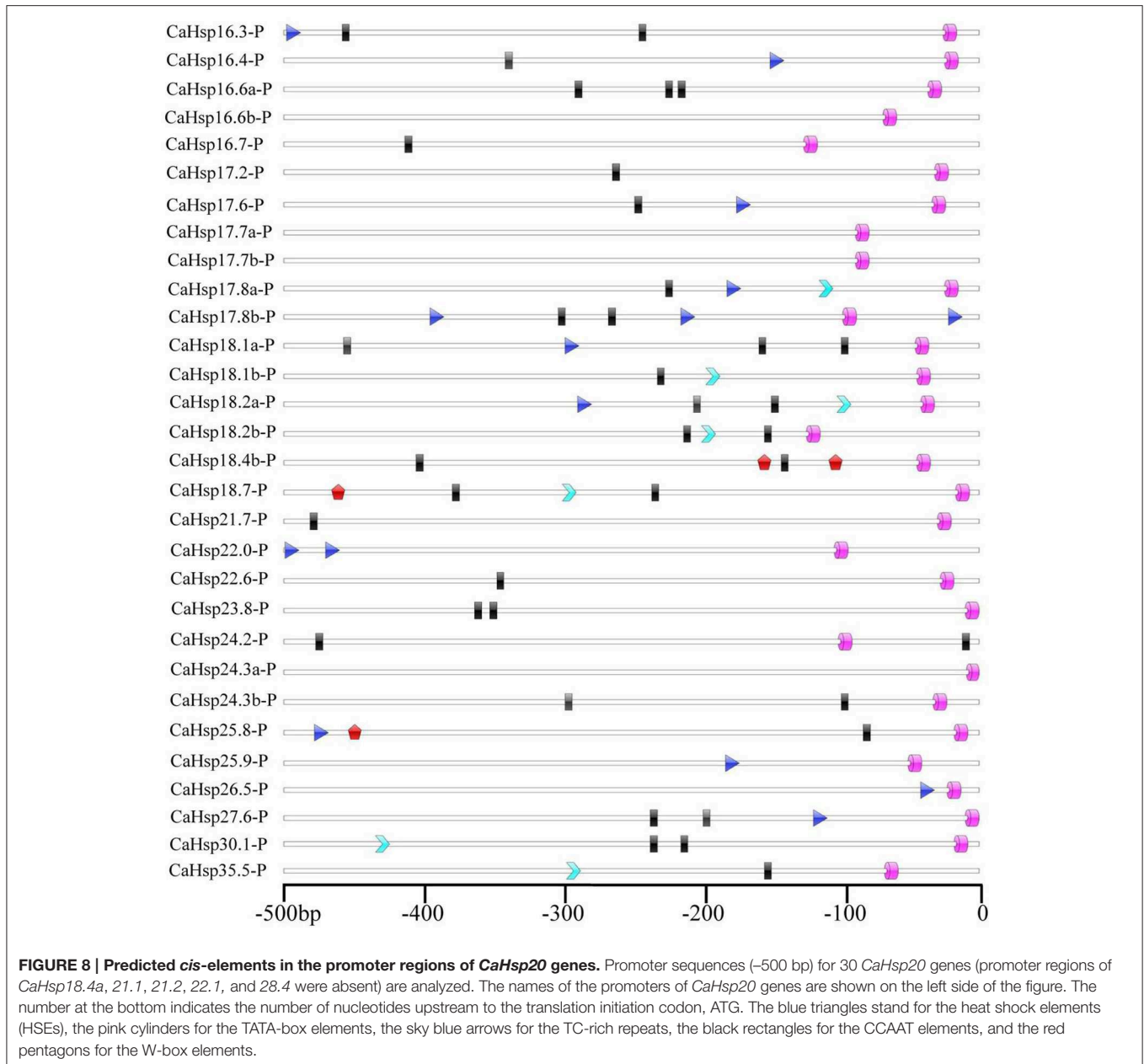
individual genes (Sun and MacRae, 2005; Waters, 2013). The comprehensive analysis of the expression for individual *Hsp20* members under HS is necessary for further in-depth functional analyses of *Hsp20* genes in plant thermotolerance. In this study, the dynamic expression of all 35 *CaHsp20s* in thermosensitive line B6 and thermotolerant line R9 with HS treatment were analyzed (Figure 7). Almost all *CaHsp20* genes are induced by HS treatment in line both B6 and R9, indicating that the accumulation of the *CaHsp20s* effectively reduce the damage from HS and enhance pepper thermotolerance by binding denaturing proteins and preventing them from irreversible aggregation (Haslbeck and Vierling, 2015). It is well-known that *Hsfs* are activated under HS condition, and then bind the HSE elements of the promoters of the *Hsp* genes to regulate the expression of downstream *Hsp* genes. The results that the promoter regions of *CaHsp20* genes contain various *cis*-elements (Figure 8) and *CaHsp20s* can interact with many functional genes (Figure 5) suggest that the expression regulation network of *CaHsp20* genes under HS condition is a complex signal transduction pathway, not just *Hsfs*. This hypothesis can be supported by the observation that some *CaHsp20* genes (*CaHsp16.6a*, *17.7a*, *18.7*, *21.2*, *22.6*, and *24.3*) whose promoter regions (–1500 bp) do not contain the HSEs can be highly



induced by HS treatment. Intriguingly, *CaHsp17.7a* and *17.7b* genes are in a tandem duplication share a similar expression pattern in both lines B6 and R9 with HS treatment, however, another tandem duplication pair *CaHsp18.2a* and *18.2b* exhibit different expression patterns in both lines B6 and R9, so it seems that not all the members of the same *sHsp* subfamily have the same expression patterns (Waters, 2013). It is worth noting that the peak of the expression of most *CaHsp20* genes in R9 (HS for 0.5 or 1 h) appeared earlier than in B6 (HS for 2 or 4 h), which may indicate that when suffering from HS, the pepper thermotolerant line R9 can start the thermal response mechanism faster than the thermosensitive line B6, however, the relationship between the function of *CaHsp20* genes and thermotolerant mechanism in pepper need to be further studied.

Conclusions

In this study, a genome-wide analysis of *CaHsp20* gene family in pepper was performed to reveal gene structure, phylogenetic relationship, evolution, and expression profiles in different tissues (root, stem, leaf, and flower) and HS conditions. Most *CaHsp20* genes maintained the low expression level under normal temperature condition and could be highly induced by HS, indicating that they played a key role in the process of pepper acquired thermotolerance. The data in this study will provide a foundation for functional characterization of *CaHsp20* family members, and make a contribution toward illuminating the mechanism of pepper acquired thermotolerance, and even the adaptability of pepper to other stress conditions.



Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 31272163, 31572114), the Shaanxi Agriculture Science and Technology Projects (Grant No. 2014K01-14-01), Jiangsu Agriculture Science and Technology Innovation Fund [CX(12)1004], the Opening Fund of Key Laboratory for Crop Biotechnology of Xinjiang Uygur Autonomous Region (Grant No. XJYS0302-2014-03) and the Tang Zhongying Fund for Breeding of Northwest A&F University.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00806>

Figure S1 | The logic flowchart of identification of pepper *Hsp20* genes.

Table S1 | The gene IDs of *Hsp20* members from *Arabidopsis*, tomato, maize and rice.

Table S2 | Primer sequences used for semi-quantitative RT-PCR and quantitative real-time PCR analysis.

References

- Basha, E., Friedrich, K. L., and Vierling, E. (2006). The N-terminal arm of small heat shock proteins is important for both chaperone activity and substrate specificity. *J. Biol. Chem.* 281, 39943–39952. doi: 10.1074/jbc.M607677200
- Bondino, H. G., Valle, E. M., and Ten Have, A. (2012). Evolution and functional diversification of the small heat shock protein/ α -crystallin family in higher plants. *Planta* 235, 1299–1313. doi: 10.1007/s00425-011-1575-9
- Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 4:10. doi:10.1186/1471-2229-4-10
- Caspers, G. J., Leunissen, J. A., and de Jong, W. W. (1995). The expanding small heat-shock protein family, and structure predictions of the conserved “alpha-crystallin domain”. *J. Mol. Evol.* 40, 238–248. doi: 10.1007/BF00163229
- Charrng, Y. Y., Liu, H. C., Liu, N. Y., Hsu, F. C., and Ko, S. S. (2006). *Arabidopsis* Hsa32, a novel heat shock protein, is essential for acquired thermotolerance during long recovery after acclimation. *Plant Physiol.* 140, 1297–1305. doi: 10.1104/pp.105.074898
- Emanuelsson, O., Nielsen, H., Brunak, S., and 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
- Giese, K. C., and Vierling, E. (2004). Mutants in a small heat shock protein that affect the oligomeric state. Analysis and allele-specific suppression. *J. Biol. Chem.* 279, 32674–32683. doi: 10.1074/jbc.M40445200
- Giorno, F., Guerriero, G., Baric, S., and Mariani, C. (2012). Heat shock transcriptional factors in *Malus domestica*: identification, classification and expression analysis. *BMC Genomics* 13:639. doi: 10.1186/1471-2164-13-639
- Giorno, F., Wolters-Arts, M., Grillo, S., Scharf, K. D., Vriezen, W. H., and Mariani, C. (2010). Developmental and heat stress-regulated expression of HsfA2 and small heat shock proteins in tomato anthers. *J. Exp. Bot.* 61, 453–462. doi: 10.1093/jxb/erp316
- Glazebrook, J. (2001). Genes controlling expression of defense responses in *Arabidopsis*–2001 status. *Curr. Opin. Plant Biol.* 4, 301–308. doi: 10.1016/S1369-5266(00)00177-1
- Gu, Z., Cavalcanti, A., Chen, F. C., Bouman, P., and Li, W. H. (2002). Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol. Biol. Evol.* 19, 256–262. doi: 10.1093/oxfordjournals.molbev.a004079
- Guo, M., Zhai, Y. F., Lu, J. P., Chai, L., Chai, W. G., Gong, Z. H., et al. (2014). Characterization of *CaHsp70-1*, a pepper heat-shock protein gene in response to heat stress and some regulation exogenous substances in *Capsicum annuum* L. *Int. J. Mol. Sci.* 15, 19741–19759. doi: 10.3390/ijms151119741
- Guo, S. J., Zhou, H. Y., Zhang, X. S., Li, X. G., and Meng, Q. W. (2007). Overexpression of *CaHSP26* in transgenic tobacco alleviates photoinhibition of PSII and PSI during chilling stress under low irradiance. *J. Plant Physiol.* 164, 126–136. doi: 10.1016/j.jplph.2006.01.004
- Hanada, K., Zou, C., Lehti-Shiu, M. D., Shinozaki, K., and Shiu, S. H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* 148, 993–1003. doi: 10.1104/pp.108.122457
- Haslbeck, M., and Vierling, E. (2015). A first line of stress defense: small heat shock proteins and their function in protein homeostasis. *J. Mol. Biol.* 427, 1537–1548. doi: 10.1016/j.jmb.2015.02.002
- Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* 27, 297–300. doi: 10.1093/nar/27.1.297
- Jaya, N., Garcia, V., and Vierling, E. (2009). Substrate binding site flexibility of the small heat shock protein molecular chaperones. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15604–15609. doi: 10.1073/pnas.0902177106
- Jeffares, D. C., Penkett, C. J., and Bähler, J. (2008). Rapidly regulated genes are intron poor. *Trends Genet.* 24, 375–378. doi: 10.1016/j.tig.2008.05.006
- Kim, S., Park, M., Yeom, S. I., Kim, Y. M., Lee, J. M., Lee, H. A., et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* 46, 270–278. doi: 10.1038/ng.2877
- Kirschner, M., Winkelhaus, S., Thierfelder, J. M., and Nover, L. (2000). Transient expression and heat-stress-induced co-aggregation of endogenous and heterologous small heat-stress proteins in tobacco protoplasts. *Plant J.* 24, 397–411. doi: 10.1046/j.1365-313x.2000.00887.x
- Kong, H., Landherr, L. L., Frohlich, M. W., Leebens-Mack, J., Ma, H., and dePamphilis, C. W. (2007). Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. *Plant J.* 50, 873–885. doi: 10.1111/j.1365-313X.2007.03097.x
- Kriehuber, T., Rattei, T., Weinmaier, T., Bepperling, A., Haslbeck, M., and Buchner, J. (2010). Independent evolution of the core domain and its flanking sequences in small heat shock proteins. *FASEB J.* 24, 3633–3642. doi: 10.1096/fj.10-156992
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., van de Peer, Y., 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
- Liu, R. H., and Meng, J. L. (2003). MapDraw: a microsoft excel macro for drawing genetic linkage maps based on given genetic linkage data. *Hereditas* 25, 317–321. doi: 10.3321/j.issn:0253-9772.2003.03.019
- Lopes-Caitar, V. S., de Carvalho, M. C., Darben, L. M., Kuwahara, M. K., Nepomuceno, A. L., Dias, W. P., et al. (2013). Genome-wide analysis of the Hsp20 gene family in soybean: comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses. *BMC Genomics* 14:577. doi: 10.1186/1471-2164-14-577
- Mittler, R. (2006). Abiotic stress, the field environment and stress combination. *Trends Plant Sci.* 11, 15–19. doi: 10.1016/j.tplants.2005.11.002
- Mittler, R., Finka, A., and Goloubinoff, P. (2012). How do plants feel the heat? *Trends Biochem. Sci.* 37, 118–125. doi: 10.1016/j.tibs.2011.11.007
- Ouyang, Y., Chen, J., Xie, W., Wang, L., and Zhang, Q. (2009). Comprehensive sequence and expression profile analysis of Hsp20 gene family in rice. *Plant Mol. Biol.* 70, 341–357. doi: 10.1007/s11103-009-9477-y
- Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5135–5140. doi: 10.1073/pnas.1400975111
- Rizzon, C., Ponger, L., and Gaut, B. S. (2006). Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLoS Comput. Biol.* 2:e115. doi: 10.1371/journal.pcbi.0020115
- Sarkar, N. K., Kim, Y. K., and 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
- Scharf, K. D., Berberich, T., Ebersberger, I., and Nover, L. (2012). The plant heat stress transcription factor (Hsf) family: structure, function and evolution. *Biochim. Biophys. Acta* 1819, 104–119. doi: 10.1016/j.bbagr.2011.10.002
- Scharf, K. D., Siddique, M., and Vierling, E. (2001). The expanding family of *Arabidopsis thaliana* small heat stress proteins and a new family of proteins containing alpha-crystallin domains (Ac proteins). *Cell Stress Chaperones* 6, 225–237. doi: 10.1379/1466-1268(2001)006<0225:TEFOAT>2.0.CO;2
- Siddique, M., Gernhard, S., von Koskull-Döring, P., Vierling, E., and Scharf, K. D. (2008). The plant sHSP superfamily: five new members in *Arabidopsis thaliana* with unexpected properties. *Cell Stress Chaperones* 13, 183–197. doi: 10.1007/s12192-008-0032-6
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C_T method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73
- Sun, W., van Montagu, M., and Verbruggen, N. (2002). Small heat shock proteins and stress tolerance in plants. *Biochim. Biophys. Acta* 1577, 1–9. doi: 10.1016/S0167-4781(02)00417-7
- Sun, Y., and MacRae, T. H. (2005). Small heat shock proteins: molecular structure and chaperone function. *Cell. Mol. Life Sci.* 62, 2460–2476. doi: 10.1007/s00018-005-5190-4
- Sung, D. Y., Kaplan, F., Lee, K. J., and Guy, C. L. (2003). Acquired tolerance to temperature extremes. *Trends Plant Sci.* 8, 179–187. doi: 10.1016/S1360-1385(03)00047-5
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Vierling, E. (1991). The roles of heat shock proteins in plants. *Annu. Rev. Plant Biol.* 42, 579–620. doi: 10.1146/annurev.pp.42.060191.003051
- Vision, T. J., Brown, D. G., and Tanksley, S. D. (2000). The origins of genomic duplications in *Arabidopsis*. *Science* 290, 2114–2117. doi: 10.1126/science.290.5499.2114

- Wan, H., Yuan, W., Ruan, M., Ye, Q., Wang, R., Li, Z., et al. (2011). Identification of reference genes for reverse transcription quantitative real-time PCR normalization in pepper (*Capsicum annuum* L.). *Biochem. Biophys. Res. Commun.* 416, 24–30. doi: 10.1016/j.bbrc.2011.10.105
- Wang, L., Guo, K., Li, Y., Tu, Y., Hu, H., Wang, B., et al. (2010). Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. *BMC Plant Biol.* 10:282. doi: 10.1186/1471-2229-10-282
- Wang, W., Vinocur, B., and Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218, 1–14. doi: 10.1007/s00425-003-1105-5
- Wang, W., Vinocur, B., Shoseyov, O., and 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
- Waters, E. R. (2013). The evolution, function, structure, and expression of the plant sHSPs. *J. Exp. Bot.* 64, 391–403. doi: 10.1093/jxb/ers355
- Waters, E. R., Aebermann, B. D., and Sanders-Reed, Z. (2008). Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. *Cell Stress Chaperones* 13, 127–142. doi: 10.1007/s12192-008-0023-7
- Waters, E. R., Lee, G. J., and Vierling, E. (1996). Evolution, structure and function of the small heat shock proteins in plants. *J. Exp. Bot.* 47, 325–338. doi: 10.1093/jxb/47.3.325
- Wei, F., Coe, E., Nelson, W., Bharti, A. K., Engler, F., Butler, E., et al. (2007). Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLoS Genet.* 3:e123. doi: 10.1371/journal.pgen.0030123
- Xu, G., Guo, C., Shan, H., and Kong, H. (2012). Divergence of duplicate genes in exon-intron structure. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1187–1192. doi: 10.1073/pnas.1109047109
- Yang, S., Zhang, X., Yue, J. X., Tian, D., and Chen, J. Q. (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
- Zhu, W., Lu, M. H., Gong, Z. H., and Chen, R. G. (2011). Cloning and expression of a small heat shock protein gene CaHSP24 from pepper under abiotic stress. *Afr. J. Biotechnol.* 10, 4968–4976. doi: 10.5897/AJB11.118

Conflict of Interest Statement: 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.

Copyright © 2015 Guo, Liu, Lu, Zhai, Wang, Gong, Wang and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.