Tropical Plants

ARTICLE

Open Access

https://doi.org/10.48130/tp-0026-0017

Tropical Plants **2026**, 5: e019

Heat shock protein 20 gene family involved in the temperature adaptation of typical *Euphorbiaceae*

Authors

Linling Zheng[#], Maria-Kristina Abello Ambuyoc[#], Jian Jin, Yuhua Chen, Assane Hamidou Abdoulaye, ..., Xiaohui Yu[*], Yinhua Chen[*]

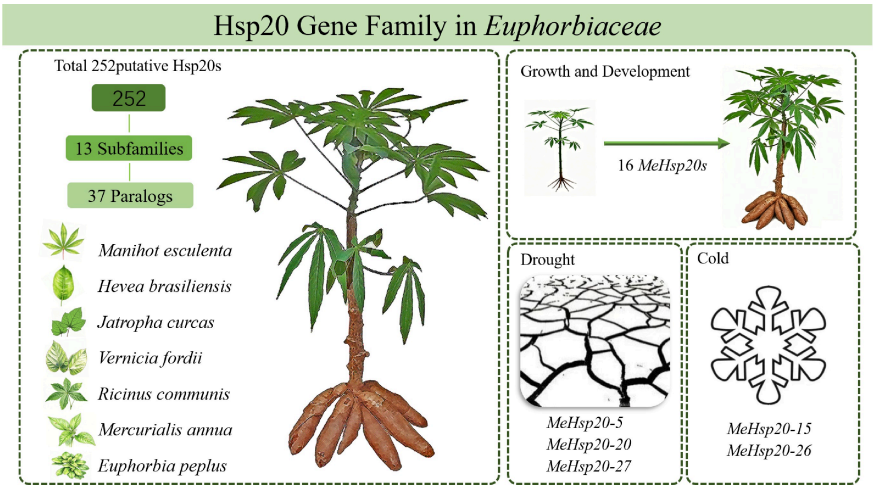
Correspondence

xiaohuiyu@hainanu.edu.cn (Yu X); yhchen@hainanu.edu.cn (Chen Y)

In Brief

To investigate the potential function of the heat shock protein 20 (Hsp20) gene family, the 252 gene family members from seven *Euphorbiaceae* species were analyzed. The origins of the important *Hsp20* genes could be explained by WGD events and homology analysis. Using cassava as a representative species, most *Hsp20s* were associated with plant growth and development, whereas several were involved in drought stress; only an extremely small number responded to cold stress.

Graphical abstract



Highlights

- A total of 252 *Hsp20* genes were identified across the genomes of seven typical *Euphorbiaceae* species.
- All syntenic gene pairs arose from segment duplication events and were under purifying selection.
- The 24 *Hsp20s* in *Euphorbiaceae* may have been derived from the ancient β WGD event.
- *MeHsp20-17*, *EpHsp20-7*, *MaHsp20-14*, and *HbHsp20-30* are involved in temperature adaptation.

Citation: Zheng L, Ambuyoc MKA, Jin J, Chen Y, Hamidou Abdoulaye A, et al. 2026. Heat shock protein 20 gene family involved in the temperature adaptation of typical *Euphorbiaceae*. *Tropical Plants* 5: e019 https://doi.org/10.48130/tp-0026-0017

Tropical Plants

ARTICLE

Open Access

https://doi.org/10.48130/tp-0026-0017

Tropical Plants **2026**, 5: e019

Heat shock protein 20 gene family involved in the temperature adaptation of typical *Euphorbiaceae*

Linling Zheng[1,2#], Maria-Kristina Abello Ambuyoc[1,3#], Jian Jin[1], Yuhua Chen[1,3], Assane Hamidou Abdoulaye[1,2,3], Zihe Zhang[1], Honggang Wang[1,3], Xiaohui Yu[1,3*] and Yinhua Chen[1,3*]

[1] *State Key Laboratory for Tropical Crop Breeding /Institute of Breeding and Multiplication, Hainan University, Sanya 572024, China*
[2] *School of Life and Health Sciences, Hainan University, Haikou 570228, China*
[3] *College of Tropical Agriculture and Forestry, Hainan University, Danzhou 571737, China*
[#] Authors contributed equally: Linling Zheng, Maria-Kristina Abello Ambuyoc
[*] Correspondence: xiaohuiyu@hainanu.edu.cn (Yu X); yhchen@hainanu.edu.cn (Chen Y)

Abstract

In higher plants, heat shock protein 20 (Hsp20) is integral to growth, development, and temperature stress adaptation. While most *Euphorbiaceae* species originated in the tropics and subtropics, they have since proliferated across both tropical and temperate regions. Investigating *Hsp20* function and evolution in *Euphorbiaceae* clarifies the role in temperature adaptation. This work performed a genome-wide analysis of 252 *Hsp20* genes across seven representative *Euphorbiaceae* species, enabling phylogenetic delineation of 13 distinct subfamilies. Among the 37 paralogous gene pairs shared with *Arabidopsis*, 24 *Hsp20s* were anchored within conserved syntenic blocks. Castor bean presented only two syntenic *Hsp20s* with *AtHsp20s*, whereas cassava and rubber tree possessed four and six *Hsp20s*, respectively. Transcriptome expression analysis of *MeHsp20s* unveiled that at least 16 *Hsp20s* across various tissues were involved in modulating cassava growth and development. Additionally, 25 *MeHsp20s* were drought-responsive, while only two *MeHsp20s* were induced by cold stress, which implies that *Hsp20s* in *Euphorbiaceae* may play a role in helping plants adapt to high temperatures rather than cold stress. These findings thereby establish a foundation for future investigations into the molecular mechanisms by which *Hsp20s* underlie growth and thermal responses in *Euphorbiaceae* plants.

Citation: Zheng L, Ambuyoc MKA, Jin J, Chen Y, Hamidou Abdoulaye A, et al. 2026. Heat shock protein 20 gene family involved in the temperature adaptation of typical *Euphorbiaceae*. *Tropical Plants* 5: e019 https://doi.org/10.48130/tp-0026-0017

Introduction

In natural environments, plants are confronted with various stresses such as drought, low temperature, and heat, which disrupt protein homeostasis and significantly impact their growth and development[1]. Recently, the problem of global warming has intensified, and high temperatures have become one of the most serious abiotic stresses affecting plants. Through evolution, plants have established robust self-defense mechanisms to maintain protein functions under stressful conditions[2]. Heat shock proteins (Hsps), universally existing in both prokaryotic and eukaryotic cells, are ubiquitous stress-induced proteins[3,4]. Plants that express *Hsps* are better able to cope with a variety of unfavorable conditions by preventing protein denaturation[4,5]. Under adversity, heat shock factor (Hsf) recruits the transcriptional machinery to the upstream promoter region of the *Hsp* by binding to the specific heat shock element (HSE: 5'nGAAnnTTCnnGAAn-3')[6].

On the basis of molecular weight and amino acid sequence homology, Hsps comprise five major families, including Hsp100, Hsp90, Hsp70/Dank, Hsp60/GroE, and small Hsp (sHsp), which are also referred to as Hsp20[7]. Plants possess numerous Hsp20s, with some plants even harbor over 40 Hsp20s[7]. Although the structures of different Hsp20s are distinctive, the α-crystalline C-terminal domain (ACD) is a ubiquitous and highly preserved domain common to Hsp20s[8,9]. Comprising 80–100 conserved amino acids, the ACD domain forms a β-strand structure and is defined by two conserved regions (CRs): CRI (β2-β5) and CRII (β7-β9, and the β6 loop)[10]. Plant Hsp20 proteins are categorized into discrete subfamilies based on key attributes such as sequence homology and function, subcellular localization, and immunological cross-reactivity[11]. Seven subfamilies (CI-CVII) of Hsp20 are localized in the cytosol or nucleus, while the CI gene family constitutes the largest group in plants. Additional Hsp20 subfamilies localize to multiple organelles, including chloroplast (CP), plastid (P), endoplasmic reticulum (ER), mitochondria (M), and peroxisomes (Po)[9,12]. The Hsps thereby function to safeguard cellular proteostasis, mitigating the damage inflicted by environmental stresses[13]. After plants are exposed to heat stress (HS), Hsp20 ensures the normal function of other proteins by safeguarding against heat-denatured aggregation and irreversible denaturation, providing a molecular basis for enhancing the thermotolerance of plant organs[14]. Overexpression of *ZmHSP16.9* in tobacco resulted in increased heat tolerance[15]. Heterologous overexpression of *TaHsp26* strengthened the thermotolerance of transgenic *Arabidopsis*[16].

Recognizing the critical role of *Hsp20* genes in mediating abiotic stress resistance has prompted genome-wide identification in multiple species, including *Arabidopsis*, rice, and grape, while the biological functions of crucial *Hsp20*s have been reported[14,17,18]. The *Euphorbiaceae* family is both large and geographically widespread, predominantly in tropical and subtropical regions, encompassing life forms such as trees, perennial shrubs, and herbs[19]. Consequently, it is more appropriate to select important species of *Euphorbiaceae* as objects to analyze the function of Hsp20 in helping plants adapt to environmental changes. Most plants in the *Euphorbiaceae* originated in the tropics. By contrast, *Euphorbia peplus*, which belongs to the *Euphorbioideae* subfamily, originates from the Mediterranean coast in the subtropics, and is now mostly distributed in the subtropics. *Ricinus communis* (castor bean) and *Mercurialis annua*, belonging to the *Acalyphoideae* subfamily, stemmed from the tropics and subtropics and are widely distributed

in both tropical and subtropical regions, with castor bean even distributed in temperate regions. *Manihot esculenta* (cassava), *Hevea brasiliensis* (rubber tree), *Jatropha curcas* (physic nut), and *Vernicia fordii* (tung tree), which belong to the *Crotonaceae* subfamily, all derive from the tropics. Tung trees, which are also found in small numbers in temperate climates, are the only ones that are widely dispersed in tropical and subtropical regions. Driven by the evolutionary trajectory and trait diversity, *Euphorbiaceae* species are able to adjust to changing environmental conditions, especially temperature and climate.

The species of *Euphorbiaceae* are also well known for their wide applications in the industrial and pharmaceutical areas. Cassava can be used to produce environmentally friendly hot-melt adhesives for textile industries and composite films; also, the wastewater generated from its processing industry can be utilized as an economical culture medium for biosurfactant production[20–22]. The rubber tree efficiently dominates the production of natural rubber latex, which can be used for the production of biomass energy, as well as in the construction and furniture industries[23,24]. Moreover, over 60% of natural rubber is utilized in the tire industry[25]. The oil extracted from the seeds of the physic nut can be used to produce biodiesel, aviation fuel, and is also applicable in the pharmaceutical industry[26,27]. The tung tree is an industrial oil tree that can be used to produce biocomposites, rigid polyurethane foam, and tung oil-based oligomer[28–30]. Castor bean is an important industrial crop worldwide, whose oil is extracted from seeds not only be used to produce traditional industrial products, but also to manufacture excellent nanocomposites, new polyurethane adhesives, and water-borne polyurethane composites[31–34]. Furthermore, *M. annua* is prized for the antimicrobial activity of its extract, and *E. peplus* has potential as a therapeutic agent for skin cancer[35,36]. Nevertheless, the *Hsp20* genes of *Euphorbiaceae* have not yet been identified at the genome-wide level.

Consequently, this study identified the *Hsp20* gene family in seven *Euphorbiaceae species* to conduct a comparative analysis of their evolution. The characteristics and functions, evolutionary relationships, and expression patterns of *Hsp20s* were analyzed. 252 *Hsp20* genes were identified, among which four genes (*MeHsp20-17*, *EpHsp20-7*, *MaHsp20-14*, and *HbHsp20-30*) may respond to temperature stress and stem from the same ancestor. The comprehensive identification of *Hsp20* genes in *Euphorbiaceae* species advanced a key hypothesis that *Hsp20s* mediate environmental adaptation. This premise establishes a conceptual framework for studying the mechanistic role of the *Hsp20* gene in influencing the distribution of *Euphorbiaceae* plants.

Methods

Genome-wide identification of the *Hsp20* gene family in *Euphorbiaceae*

The protein and CDS sequences for the seven species were sourced from NCBI, Phytozome, and GSA, respectively. The genome version and other information were provided in Supplementary Table S1. The *Hsp20s* in seven representative *Euphorbiaceae* species were screened through PF00011 of the Hidden Markov Model profile (HMMER) and BLASTp with an *E*-value threshold of 1.0×10^{-5}. Low-molecular-weight proteins containing an ACD domain, which is composed of approximately 80–100 conserved amino acids, were defined as candidate proteins. Prediction of the conserved domain within the candidate protein sequences was performed via the SMART (http://smart.embl-Heidelberg.de) online tool and CDD (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) database, facilitating the removal of redundant proteins. Finally, the Hsp20 family members with highly conserved feature information were screened, which were named Hsp20s.

The key physicochemical parameters, encompassing protein length, molecular weight (MW), isoelectric point (PI), Grand Average of Hydropathicity (GRAVY), and prediction of subcellular localization for the identified Hsp20 proteins were determined via the ExPASy Proteomics Server (https://web.expasy.org/protparam) and WoLF PSORT (https://wolfpsort.hgc.jp), respectively. The Hsp20 protein sequences were aligned with the aid of DNAMAN software.

Chromosomal mapping, conserved motifs, and *cis*-element analysis

The chromosome location information of *Hsp20* genes was extracted based on genome annotation information, and the chromosome distribution map was drawn using TBtools[37]. To identify conserved motifs, the MEME (https://meme-suite.org/meme/) online tool was employed, with a maximum number of motifs set at 10. TBtools assisted in the visual analysis[37].

The 2 kb promoter sequences of *Hsp20* from seven *Euphorbiaceae* species were interrogated for the possible regulatory elements. Location-based matrix, consensus, and specific promoter sequences on a single site of various regulatory elements were predicted through the online website PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

Phylogenetic, whole-genome duplication, and synteny analysis

The Hsp20 protein sequences of *M. esculenta*, *R. communis*, *J. curcas*, *M. annua*, *V. fordii*, *H. brasiliensis*, *E. peplus*, *O. sativa*, and *A. thaliana* were stored in FASTA format in the same file. The ClustalW command was used to perform protein sequence alignment using the MEGA7.1 tool. Guided by the multiple sequence alignment results, an interspecific phylogenetic tree was inferred via the neighbor-joining (NJ) method, employing 1,000 bootstrap replicates for support. The Evolview (https://evolgenius.info/helpsite/qst1.html) online website was utilized to edit the evolutionary tree.

To compare whole-genome duplication events and identify homologous blocks in different species, whole-genome duplication (WGD) analysis was performed using whole-genome duplication integrated (WGDI) in Python. The results of the comparison were presented in a dot plot. MCScanX inferred collinear regions within the intraspecific and interspecific genomes, and the results were visualized using TBtools.

Homology modeling and conservation position analysis

Arabidopsis sHsp (PDB ID: 7BZW) was used as a model[38]. The model was visualized and modified using PyMOL software. The protein sequences of *Arabidopsis* and 252 *Euphorbiaceae* Hsp20s were input on the ConSurf server (https://consurf.tau.ac.il/consurf_index.php), and the evolutionary conservation and variation scores of each residue were calculated.

Prediction of transcription factor binding *Hsp20s*, protein-protein interaction network, and KEGG pathway

To investigate the transcription factors that bind to the *Hsp20s* of seven *Euphorbiaceae*, the coding sequences of *Hsp20* genes were

extracted. The Plant Transcriptional Regulatory Map (PlantRegMap: Plant Regulation Data and Analysis Platform @ CBI, PKU [gao-lab.org]) was leveraged to predict transcription factors that bind to the promoter of *Hsp20s*.

The STRING online service (https://cn.string-db.org/) with a confidence parameter of 0.15 was used to detect the functional protein-protein interaction network (PPI). Protein association networks for *R. communis*, *M. esculenta*, and *J. curcas* were developed using their species-specific protein sequences, whereas the PPI for the other species were generated based on the homology to *Arabidopsis* Hsp20 proteins. The KEGG pathway database (www.kegg.jp/kegg/pathway.html) was employed to predict the functional pathway of Hsp20 proteins.

Expression pattern profiling of cassava *Hsp20s*

Gene expression data, encompassing responses to drought and cold treatments, as well as data from different tissues, were sourced from the NCBI database. Gene expression levels were quantified as FPKM (Fragments Per Kilobase of exon model per Million mapped fragments) value, followed by visualization with TBtools[37].

Results

Identification and characterization of *Hsp20* gene family in *Euphorbiaceae*

Seven typical *Euphorbiaceae* species were selected for analysis and classified into three different genera (Supplementary Fig. S1). Of these, four species belong to *Crotonaceae*, two to *Acalyphoideae*, and one to *Euphorbioideae* (Supplementary Fig. S1). Using 19 *A. thaliana* Hsp20s (AtHsp20s) and 39 *Oryza sativa* Hsp20s (OsHsp20s) as queries, the potential Hsp20 homologs were identified in seven representative *Euphorbiaceae* species, namely *M. esculenta* (cassava), *R. communis* (castor bean), *J. curcas* (physic nut), *H. brasiliensis* (rubber tree), *V. fordii* (tung tree), *M. annua,* and *E. peplus*. A total of 252 *Hsp20* genes constituted the *Hsp20* gene family identified in this study, comprising 23 in *R. communis*, 24 in *J. curcas*, 50 in *H. brasiliensis*, 50 in *M. esculenta*, 56 in *V. fordii*, 32 in *M. annua*, and 17 in *E. peplus* (Supplementary Fig. S2). The Hsp20 proteins from seven *Euphorbiaceae* species were named according to species, chromosome number, and gene position on the chromosome. The prefixes 'Me', 'Hb', 'Rc', 'Jc', 'Ep', 'Ma', and 'Vf' were used to denote cassava, rubber, castor bean, physic nut, *E. peplus*, *M. annua*, and *V. fordii*, respectively. As demonstrated in Table 1, although the physic nut, whose genome size was the smallest, contained a greater number of *Hsp20s* than *E. peplus*. Interestingly, except for physic nut, species in the *Crotonaceae* subfamily contained more *Hsp20* genes than the other two subfamilies (Supplementary Fig. S1; Table 1).

A comparative analysis of physicochemical properties, encompassing protein lengths, protein molecular weight (MW), isoelectric point (*p*I), and subcellular localization, was undertaken for the Hsp20 proteins from seven *Euphorbiaceae* species. The molecular weights of Hsp20s ranged from 9,944.53 kDa (MeHsp20-22) to 76,130.34 kDa (HbHsp20-22) within the seven species, while the peptide lengths of Hsp20s varied from 80 (MeHsp20-22) to 692 (HbHsp20-22) amino acids (Supplementary Table S2). The isoelectric points of the Hsp20 protein varied significantly, with the lowest being 4.69 (EpHsp20-14), and the highest being 9.84 (JcHsp20-24). Among them, 158 Hsp20s possessed values lower than 7, indicating that most were acidic. Except for VfHsp20-52, the other Hsp20 proteins showed negative GRAVY values, indicating their hydrophilic nature. The prediction of subcellular localization suggested that Hsp20 proteins of the *Euphorbiaceae* were widely distributed in the cytoplasm and different organelles, suggesting that they may perform multiple functions (Supplementary Table S2). Only VfHsp20-11 was distributed in the cytoskeleton, suggesting that it may play a special role.

Phylogenetic analysis of Hsp20s

To elucidate the evolutionary relationships, an unrooted phylogenetic tree of 252 *Euphorbiaceae* Hsp20s, 39 rice Hsp20s (OsHsp20s), and 19 *A. thaliana* Hsp20s (AtHsp20s) was constructed (Fig. 1). All of the *Hsp20s* in this study were categorized into 13 subfamilies, including two subfamilies in mitochondria (M), seven subfamilies in the cytoplasm or nucleus (CI-CIV), one subfamily in the endoplasmic reticulum (ER), chloroplast (P), and peroxisomal (Po), and 57 Hsp20s in the unclassified group. Most Hsp20s were classified into the CI subfamily, and they were predicted to be located in the cytoplasm or nucleus, including 21 MeHsp20s, 17 HbHsp20s, seven JcHsp20s, 10 MaHsp20s, 28 VfHsp20s, and eight RcHsp20s. Additionally, none of the EpHsp20s were assigned to the CI subfamily, indicating that the Hsp20s of *E. peplus*, a species that is extensively found in subtropical areas, varied functionally from Hsp20 proteins of other species. Subfamilies ER and M contained all seven species, while subfamily Po contained the Hsp20s from five species except *R. communis* and *E. peplus*, and subfamily P contained the Hsp20s except RcHsp20s and JcHsp20s. The results indicated similar evolutionary patterns of *Euphorbiaceae* and *Arabidopsis* (Fig. 1). Additionally, the phylogenetic reconstruction identified the ER and Po subfamilies, which diverged later in the evolution of the *Hsp20* gene family (Fig. 1), implying that these two subfamilies might have particular functions in the *Euphorbiaceae*. However, only two Hsp20s (MaHsp20-23 and JcHsp20-8) in *Euphorbiaceae* were grouped at the same end of the branch as AtHsp20s; while 84 Hsp20s in *Euphorbiaceae* were classified together, such as MeHsp20-28/HbHsp20-32, RcHsp20-21/HbHsp20-7, and JcHsp20-12/VfHsp20-48 (Fig. 1). These 84 genes were assigned to all subfamilies, excluding CII and MII. In particular, 18 MeHsp20s and HbHsp20s, respectively, clustered at the end of the same branch.

Whole-genome duplications analysis

Whole-genome duplication (WGD) is a major evolutionary mechanism that leads to the simultaneous duplication of all chromosomal

Table 1. The *Hsp20* gene family in *Euphorbiaceae*.

Family	Subfamily	Genus	Species	Gene number	Genome size
Euphorbiaceae	*Crotonaceae*	*Jatropha*	*Jatropha curcas*	24	266.8 Mb
		Hevea	*Hevea brasiliensis*	50	1.9 Gb
		Manihot	*Manihot esculenta*	50	582.28 Mb
		Vernicia	*Vernicia fordii*	56	1.12 Gb
	Acalyphoideae	*Ricinus*	*Ricinus communis*	23	315.6 Mb
		Mercurialis	*Mercurialis annua*	32	453.2 Mb
	Euphorbioideae	*Euphorbia*	*Euphorbia peplus*	17	267.2 Mb

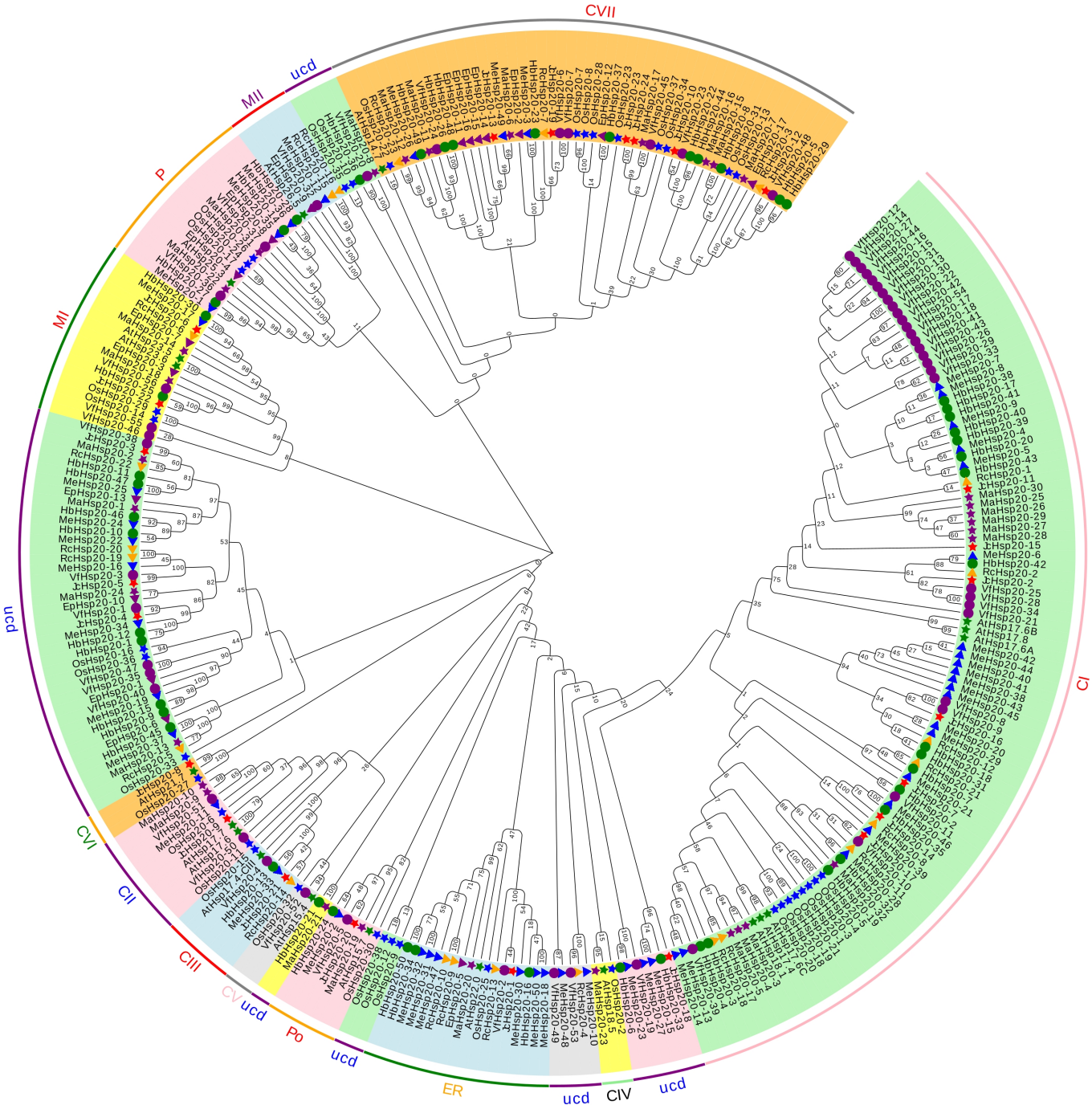


Fig. 1 Unrooted phylogenetic tree of the Hsp20s in seven *Euphorbiaceae*, rice, and *A. thaliana*. The Neighbor-joining (NJ) phylogenetic tree was inferred using MEGA 7.1. Different colors represent different clades. Me, *M. esculenta*; At, *A. thaliana*; Rc, *R. communis*; Jc, *J. curcas*; Hb, *H. brasiliensis*; Ep, *E. peplus*; Ma, *M. annua*; Vf, *V. fordii*; Os, *O. sativa*. Green star, blue star, blue triangle, yellow triangle, red star, green circle, purple triangle, purple star, and purple circle indicate *A. thaliana*, *O. sativa*, *M. esculenta*, *R. communis*, *J. curcas*, *H. brasiliensis*, *E. peplus*; *M. annua*; and *V. fordii, respectively*. CI-CIV: cytoplasm or nucleus; ER: endoplasmic reticulum; Po and P: peroxisomal; MI-MII: mitochondria; ucd: unclassified.

material. Five *Euphorbiaceae* species with genomes assembled to the chromosomal level were selected, and genome collinear dot plots were drawn with *Arabidopsis* separately (Supplementary Fig. S3). The density of dots in the dot plots of different *Euphorbiaceae* species was similar, indicating that the number of highly comparable fragments in the genomes of different *Euphorbiaceae* species and *A. thaliana* was similar. In highly similar segments, those containing *Hsp20s* were labeled. Comparative analysis of syntenic regions revealed a pronounced disparity in *Hsp20* tandem array size, which varied from two (castor bean) to nine (rubber tree) across the examined species (Fig. 2). Six *HbHsp20s* were located in collinear segments and were collinear with seven *AtHsp20s*. A total of eight *AtHsp20s* intersected with *Hsp20s* of five *Euphorbiaceae* in collinear long segments, among which the genes that intersected most with other *Hsp20s* was *AtHsp23.5* and *AtHsp23.6*, which were shared by all the other four species except castor bean (Fig. 2). However, *AtHsp25.3* only intersected *HbHsp20-27*, which was divided into the P subfamily, and *HbHsp20-27* clustered with *MeHsp20-1* at the end of the same branch (Fig. 2d, e). Significantly, *MeHsp20-1* was not located in the collinear segments. According to the above results,

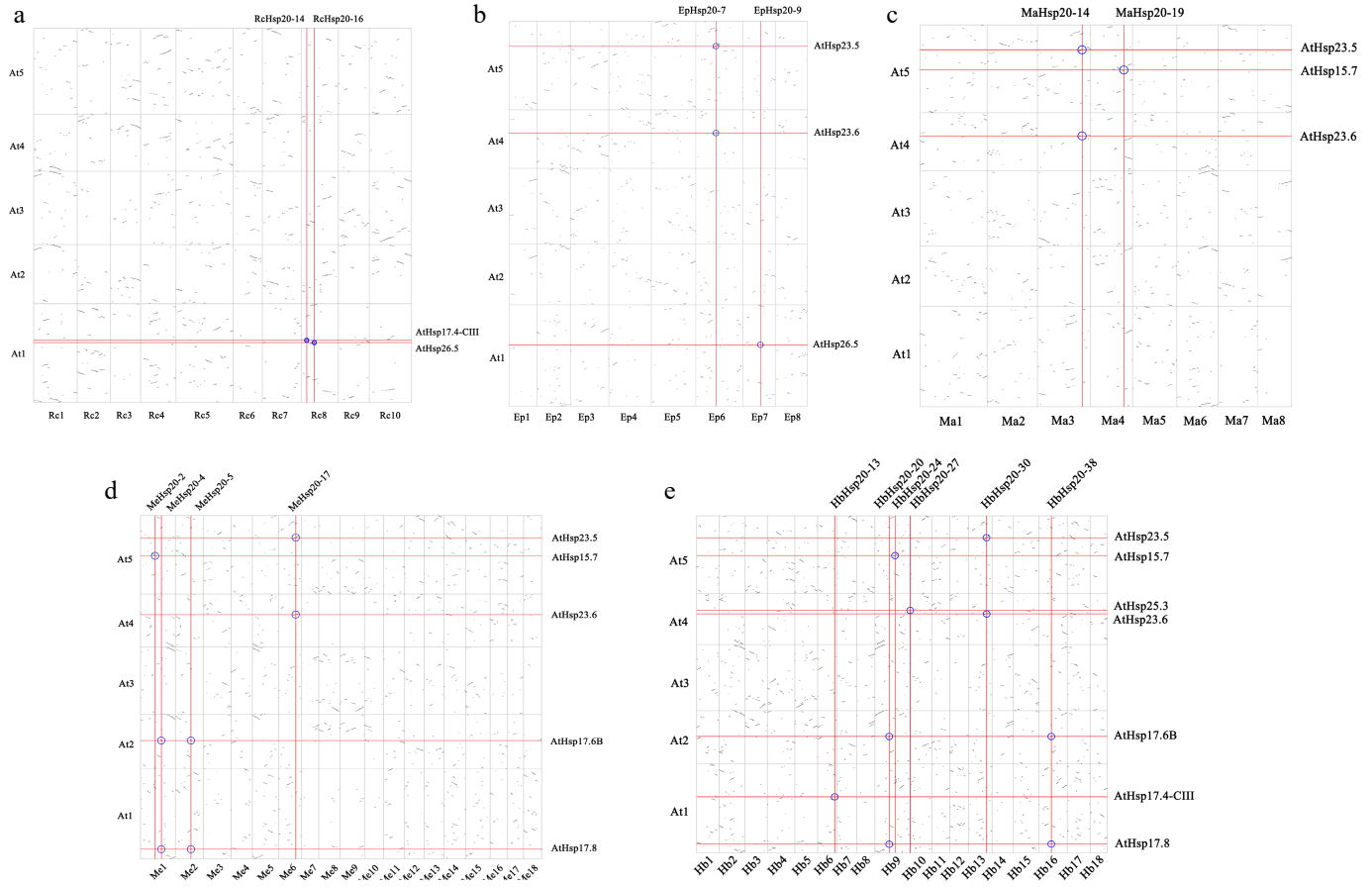


Fig. 2 Genome-wide duplication events between *Arabidopsis* and *Euphorbiaceae*. Gene dot plot between At and (a) Rc, (b) Ep, (c) Ma, (d) Me, and (e) Hb. At, *A. thaliana*; Me, *M. esculanta*; Rc, *R. communis*; Hb, *H. brasiliensis*; Ep, *E. peplus*; Ma, *M. annua*.

it was deduced that some *Hsp20s* in *Euphorbiaceae* were derived from the original *Hsp20s*, but more *Hsp20s* may have been acquired during the evolution of the species itself, and environmental adaptation.

To analyze the evolution and function of intraspecific genes, a genome collinear dot plot of five species of *Euphorbiaceae* was drawn and analyzed (Supplementary Fig. S4). Different colored dots and lines represent different degrees of similarity. The most similar segments are red, followed by yellow, and gray. Each chromosome has distinct, intense red spots that allow each fragment to be regarded as highly comparable. In addition, many unfocused yellow dots were observed, caused by ancient γ-WGD events. However, there were a few gray fragments in the five *Euphorbiaceae* species, suggesting that older duplications are not apparent. In five *Euphorbiaceae* species, a diagonal line of red dots appeared on the genome dot plot, indicating the collinearity of each gene within the genome. Particularly, different numbers of collinear long fragments can be observed in the dot plots of different species. More obvious fragments were observed in the genome dot plots of cassava and rubber trees, implying that the traces were retained after large-scale genome duplication (Supplementary Fig. S4).

Chromosomal mapping and synteny analysis of *Hsp20s*

In order to explore the replication pattern and evolutionary mechanism of the *Hsp20s* in *Euphorbiaceae*, chromosome localization analysis was conducted on the seven *Euphorbiaceae* (Supplementary Fig. S5). Notably, an uneven distribution of *Hsp20* genes was discovered throughout the chromosomes of each species. The result of *MeHsp20s* showed that Me14 contained the largest number (nine) of *MeHsp20* genes, followed by Me2, and Me10. For *RcHsp20s*, most of the *RcHsp20* genes were presented on Rc8 (eight genes), followed by Rc5, which contained four genes. The number of *HbHsp20* genes across chromosomes ranged considerably, with Hb9 and Hb16 harboring eight genes. The remaining *HbHsp20* genes were randomly distributed across the other rubber tree chromosomes and scaffolds. *EpHsp20s* were distributed on six *E. peplus* chromosomes. In contrast to the mere one to three genes found on other chromosomes, Ep8 harbored seven *EpHsp20* genes. The 32 *MaHsp20s* were unevenly distributed across eight chromosomes. Ma1 and Ma7 harbored the greatest number of genes (seven members each), followed by Ma4. Due to the incomplete genome assembly of the physic nut and tung tree, all the *JcHsp20* and *VfHsp20* genes were mapped only at the scaffold level. The *JcHsp20* and *VfHsp20* genes were randomly located on the 16 and 36 scaffolds, respectively.

A comparative analysis of genomic syntenic blocks in *A. thaliana* and seven *Euphorbiaceae* was employed to elucidate the evolutionary history of the *Hsp20* gene family (Fig. 3). *A. thaliana* had a collinear relationship with *Euphorbiaceae*, and homologous gene pairs were identified in seven studied species, including four pairs in castor bean, 26 pairs in rubber tree, 10 in *E. peplus*, 11 in *M. annua*, 15 pairs in cassava, two in physic nut and eight in tung tree (Fig. 3). Among them, castor bean, rubber tree, *E. peplus*, *M. annua*, cassava, physic nut, and tung tree respectively, had 2, 9, 3, 4, 11, 2, and 6 collinear relationships with the *sHsp* genes of *Arabidopsis* (Fig. 3; Supplementary Table S3). Among all the gene pairs, a pair of

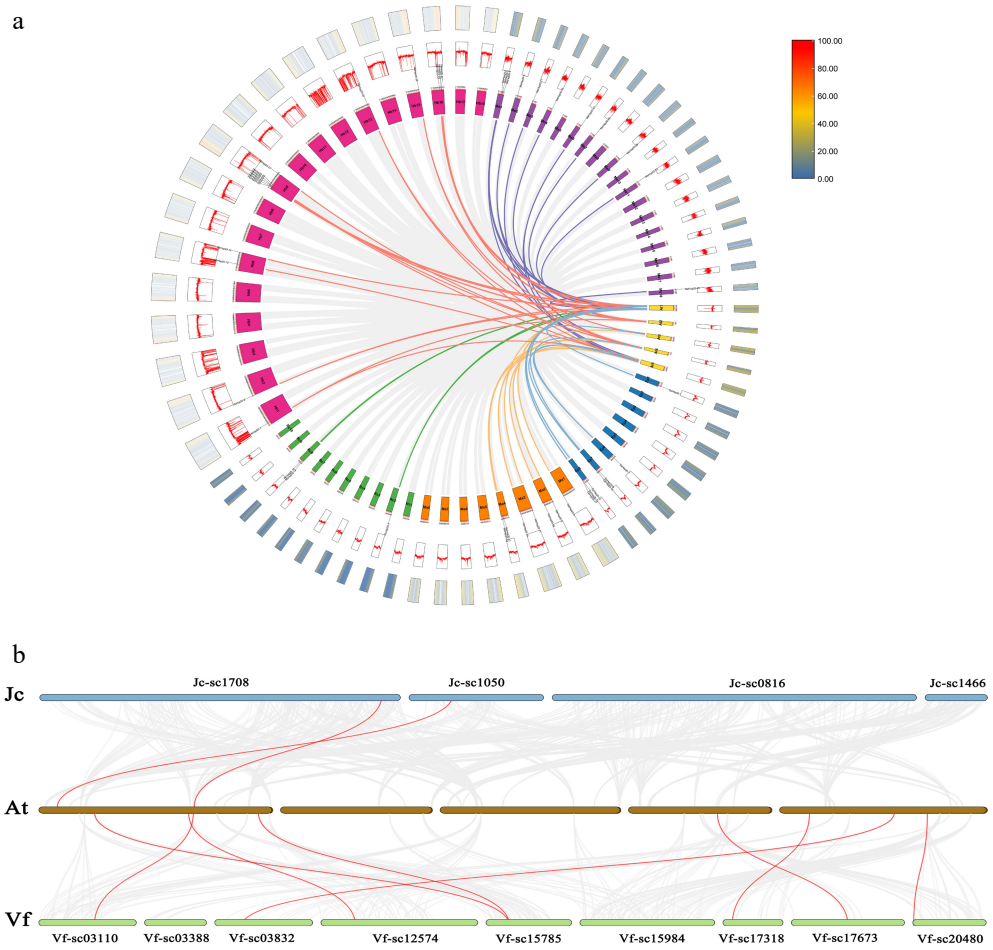


Fig. 3 Synteny relationships of *Hsp20* genes between *A. thaliana* and *Euphorbiaceae*. (a) Synteny analysis of *Hsp20* genes between Rc, Hb, Me, Ep, Ma, and *A. thaliana*. (b) Synteny analysis of *Hsp20* genes between *A. thaliana* and physic nut, tung tree. Species chromosomal-level genome assemblies were co-mapped. Jc and Vf, with genome assembly at the scaffold level, were analyzed and drawn using scaffolds containing the *Hsp20s*. The *A. thaliana*, castor bean, rubber tree, cassava, *E. peplus*, and *M. annua* are labeled At, Rc, Hb, Me, Ep, and Ma in a circle diagram. At, Rc, Hb, Me, Ep, and Ma are shown in yellow, green, pink, purple, blue, and orange. The blue, green, and brown scaffolds and chromosomes represent physic nut, tung tree, and *A. thaliana*.

collinear genes (*VfHsp20-52*, *AtHsp15.4*) was classified into the same branch end in the phylogenetic tree (Fig. 1). The results of synteny analysis pointed to a potential relationship between the number of gene pairs and both genome size and gene family size. Furthermore, *AtHsp17.4-CIII* was collinear with *Hsp20s* of the other five species except for *E. peplus* and *M. annua*. Of all the *Euphorbiaceae* collinear *Hsp20* genes, eight genes (*JcHsp20-21* and *RcHsp20-14*, *HbHsp20-13* and *MeHsp20-33*, *HbHsp20-20* and *MeHsp20-4*, *HbHsp20-30* and *MeHsp20-17*) were clustered at the branching end of the evolutionary tree (Fig. 1). Whereas, *AtHsp25.3*, *AtHsp15.4*, and *AtHsp14.7* only formed collinear gene pairs with *HbHsp20-27*, *VfHsp20-52*, and *VfHsp20-55*, respectively. The fact that Ks for 12 gene pairs was negative, however, cannot be ignored; this suggests that the sequence divergence of these gene pairs is significant and evolved over time. In addition to these genes, the remaining paralogous genes underwent purifying selection, and the proteins encoded by these genes were likely to have eliminated harmful mutations and preserved the functions of the ancient Hsp20 protein (Supplementary Table S4).

Conserved motifs, *cis*-regulatory elements, and transcription factors analysis of *Hsp20s*

Analysis of conserved motifs based on protein sequences was utilized to investigate the conservation of *Hsp20* gene family members in *Euphorbiaceae*. Among all Hsp20 members, 234 Hsp20s (92.9%) contained Motif 2, 153 (60.7%) contained Motif 1, and 111 (44.0%) contained Motif 3 (Supplementary Table S5). In particular, the Hsp20s with comparable motifs were categorized within the same subfamily (Fig. 4a, b). Additionally, Motif 2 was a universal feature of all Hsp20s. Each subfamily was defined by specific motif signatures: Subfamily I contained Motif 1, 2, 3, and 5; Subfamily Po contained Motif 1, 2, and 4; Subfamily ER contained Motif 1, 2, 4, 5, and 10; Subfamily M contained only Motif 1 and 2. Motif 2 included the β2, β3, and β4 of the ACD structure; Motif 3 included β5 to β7; Motif 4 and Motif 9 included β8 and β9, respectively (Fig. 4c).

The extraction of promoter regions from 252 genes and predictive analysis of their *cis*-acting elements were instrumental in illuminating the transcriptional regulation of *Hsp20s* in the examined species. Three categories of *cis*-elements were identified, including plant growth and development, phytohormone-related, and external stress *cis*-elements (Supplementary Fig. S6). All *Hsp20* genes had light-responsive elements whose proportion was approximately 46.63% to 58.39%. Interestingly, *E. peplus*, which originated in subtropics, has the lowest proportion of light-responsive elements compared to other species (Supplementary Table S6). Among the elements of the phytohormone-responsive category, most of the *Hsp20* genes possessed abscisic acid- and MeJA-responsive

Fig. 4 Phylogenetic tree and conserved motifs of *Hsp20* genes of *Euphorbiaceae*. (a) The phylogenetic tree of *Hsp20* genes of seven *Euphorbiaceae* species; (b) the motif compositions of Hsp20s; (c) the conserved motif contained an ACD domain. Numbers on the X-axis represent the position of the amino acids. The relative frequency of amino acids in the motif is expressed in font size. The Motif logos are Motif 2, 3, 4, and 9 from top to bottom. The color boxes in Motif 2 represent the β2, β3, and β4; Motif 3 represents the β5, β6, and β7; Motif 4 represents the β5; and Motif 9 represent the β9 of the ACD structure.

elements (Fig. 5), which implied that the function of *Hsp20s* might be regulated by abscisic acid and MeJA. Notably, the number of elements in different species varied greatly (Fig. 5; Supplementary Table S6). *MeHsp20s* contained 144 MeJA-responsiveness elements, while *RcHsp20s* had only 32. Compared with other species, *MeHsp20s* consisted of many individual response elements (Supplementary Table S6). The abundance of light response and hormone response factors demonstrated that *Hsp20* gene expression in *Euphorbiaceae* was governed by light and hormones. More diverse roles of *Hsp20* genes in stress responses were implied by the identification of diverse stress-related *cis*-acting elements (e.g., defense, drought-inducible, and anaerobic-inducible). Additionally, *cis*-elements involved in growth and development, such as circadian control, meristem expression regulatory, and endosperm expression regulatory elements, were also found in all studied species (Supplementary Table S6). The results pointed to a potential role of *Hsp20* genes in regulating the growth and development of *Euphorbiaceae* plants.

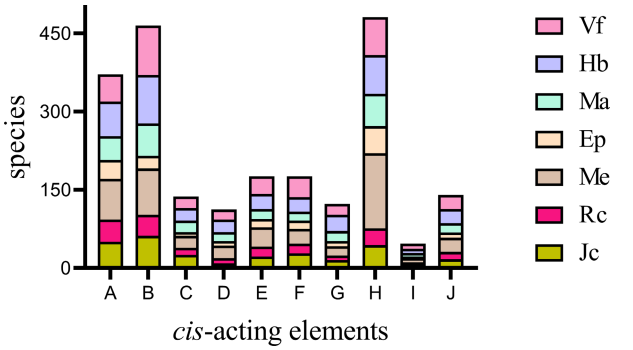


Fig. 5 *Cis*-acting element distribution in the promoter of *Hsp20s*. The number of *cis*-acting elements in different species is shown in different colors. A–J represent different *cis*-acting elements. A: abscisic acid responsiveness; B: anaerobic induction; C: auxin responsiveness; D: defense and stress responsiveness; E: drought inducibility; F: gibberellin responsiveness; G: low-temperature responsiveness; H: MeJA-responsiveness; I: MYB binding site; J: salicylic acid responsiveness.

Hsp20s of different *Euphorbiaceae* species contained different types and quantities of transcription factor binding sites (Fig. 6; Supplementary Table S7). The *Hsp20s* of the remaining species all had 43 distinct transcription factor binding sites, while *JcHsp20s* and *MeHsp20s* had 39, and 40 distinct binding sites, respectively (Supplementary Table S7). Notably, the number of the top 10 transcription factor binding sites contained by *Hsp20s* was similar across species.

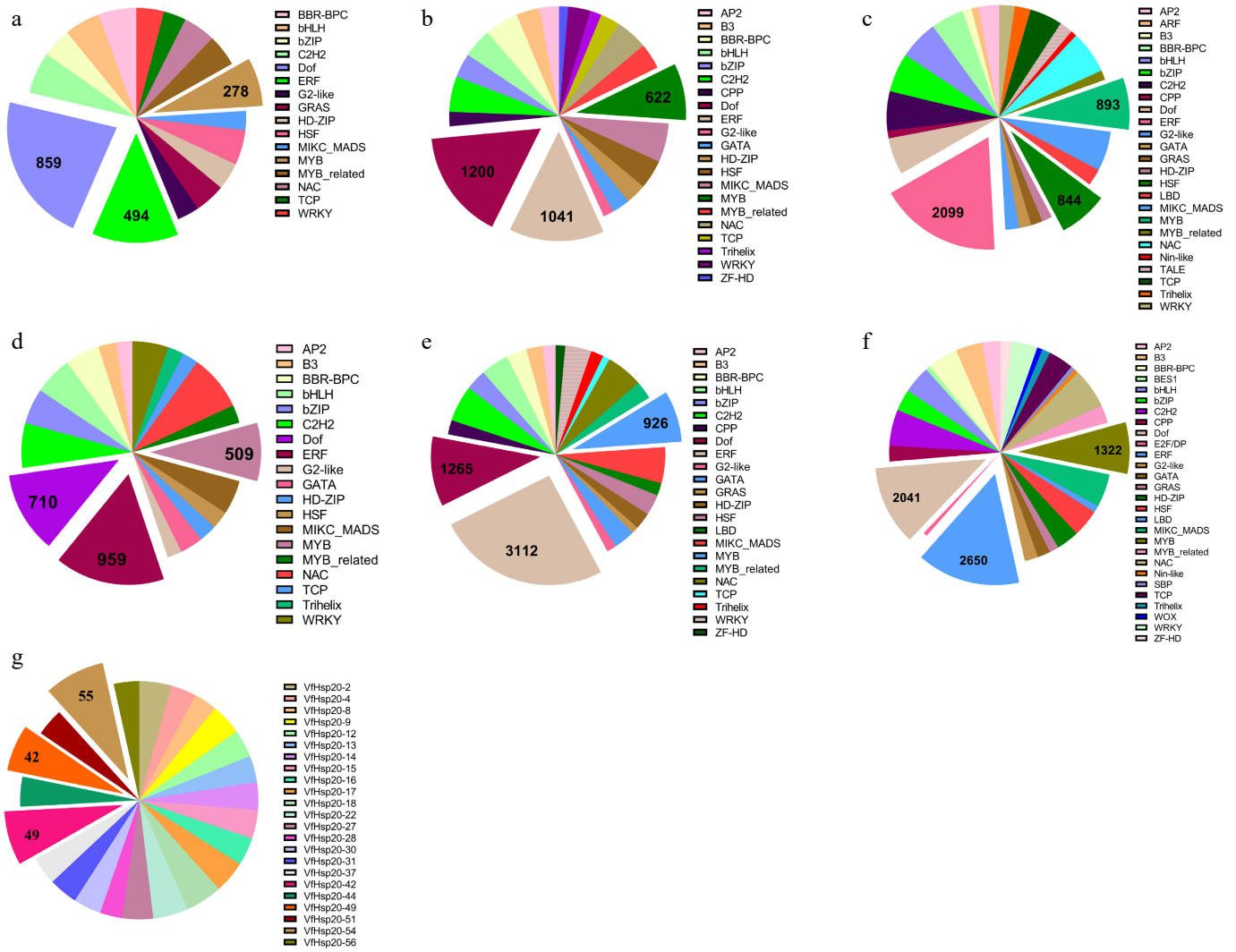


Fig. 6 Composition of predicted transcription factor binding sites in *Hsp20s*. The abundance of different transcription factors is displayed on the pie chart. (a) *J. curcas;* (b) *R. communis;* (c) *M. esculenta;* (d) *E. peplus;* (e) *M. annua;* (f) *H. brasiliensis;* and (g) *V. fordii.*

JcHsp20s and *RcHsp20s* contained the largest number of Dof (DNA binding with one finger) binding sites, while the others contained ERF (ethylene-responsive factor) binding sites. In particular, the transcription factors ERF, Dof, and MYB (v-myb avian myeloblastosis viral oncogene homolog) were the most abundant binding sites across all studied species except cassava.

It is well known that *Hsf* binds to the upstream region of *Hsp* to regulate expression[6]. All cassava *Hsp20s* contained 844 *Hsf* binding sites, the largest number of *Hsf* binding sites among all studied species (Fig. 6; Supplementary Table S7). The number of *Hsf* binding sites contained in *RcHsp20*, *JcHsp20*, *MaHsp20*, *EpHsp20*, *HbHsp20*, and *VfHsp20* was 318, 200, 378, 155, 700, and 836, respectively (Supplementary Fig. S7; Supplementary Table S7). Among the studied species, *E. peplus* had the lowest proportion of the *Hsp20* genes containing the *Hsf* binding site (Supplementary Table S8). Although Hsf binding sites were ubiquitous, being identified in 90% of *MeHsp20s*, their distribution was markedly skewed, as the three most abundant genes collectively constituted merely 14.95% of the total sites. This indicated that *Hsf* binding sites were widely distributed in *MeHsp20s* and might be involved in controlling *Hsp20* transcription (Supplementary Fig. S7; Supplementary Table S8). Of the top three *Hsp20s* containing the most *Hsf* binding sites in the same species, four genes (*JcHsp20-1*, *JcHsp20-11*, *MeHsp20-9*, *VfHsp20-49*) each phylogenetically clustered at the branch terminus with *Hsp20s* from other *Euphorbiaceae* (Fig. 1). Except for *Hsf* binding sites, the number of *ERF* binding sites was the largest (Supplementary Table S8). The proportion of *Hsp20s* containing *Hsf* and *ERF* binding sites differed in different species (Fig. 6). For the Hsf binding site, 45 *MeHsp20s* (90%) contained the *Hsf* binding sites, while only 25 of 32 *MaHsp20s* contained the *Hsf* binding sites. For the *ERF* binding site, 959 were found in 16 of 17 *EpHsp20s*, and 1,908 *ERF* binding sites were found in 27 of 50 *EpHsp20s*. Only four genes (*RcHsp20-6*, *MeHsp20-11*, *EpHsp20-3*, and *EpHsp20-8*) contained many *Hsf* and *ERF* binding sites.

Analysis of key amino acid variation sites and tertiary structure

To determine the functional sites, the conservation and scores of each site of the Hsp20 protein were calculated based on the homology model of *Arabidopsis* sHsp (PDBID: 7BZW). The results of variation sites showed that E113, D123, W156, L168, and A181 were highly conserved, and these amino acids were respectively located in β2, β3, β6, β7, and β8, which constituted the ACD domain (Supplementary Fig. S2; Supplementary Table S8). In the studied *Euphorbiaceae*, most of the L168 residues mutated into F residues. K133 and I134 were the conserved sites located in β4, while L188, P193,

and K194 were the conserved sites located in β9. I143, G145, and E146 were all highly conserved sites in β5, among which I143 was a buried structural residue. Except for I143, the other highly conserved residues were all exposed functional residues. The Hsp20 protein consisted of 12 peptide chains, with different types and numbers of secondary structures that made up each chain (Supplementary Fig. S8). The 12 peptide chains that make up the Hsp20 protein are composed of various secondary structures, including a comparatively high percentage of random coils.

Protein–protein interaction (PPI) network and KEGG pathway of Hsp20

The highest homologous STRING proteins were identified by the Hsp20 protein sequences of seven *Euphorbiaceae* species, respectively, to predict the protein–protein interaction (PPI) network. The homologous proteins were matched based on the highest bit score by default, which led to the identification of 25 MeHsp20, 14 RcHsp20, 11 JcHsp20, 15 EpHsp20, 25 MaHsp20, 32 HbHsp20, and 49 VfHsp20 proteins within the PPI networks (Fig. 7). As shown in Fig. 7, the interacting partners of Hsp20 proteins were not only proteins from the Hsp family, but also proteins from other family members. Except for Hsp gene family members, the majority of Hsp20 protein interaction partners were BAG (Bcl-2 associated athanogene) and CLPB (casein lytic proteinase B), which implied that Hsp20 proteins might be regulated by other proteins in *Euphorbiaceae*. In cassava, MeHsp20-1 had 16 interaction partners, followed by MeHsp20-17 (13), MeHsp20-12 (10), and MeHsp20-3 (10). Nevertheless, more than 10 HbHsp20s had more interaction partners. Five HbHsp20s (HbHsp20-19/24/25/30/31) had 19 interaction partners.

Heat shock proteins, as molecular chaperones, participate in the formation and degradation of proteins. In cells, proteins fold with

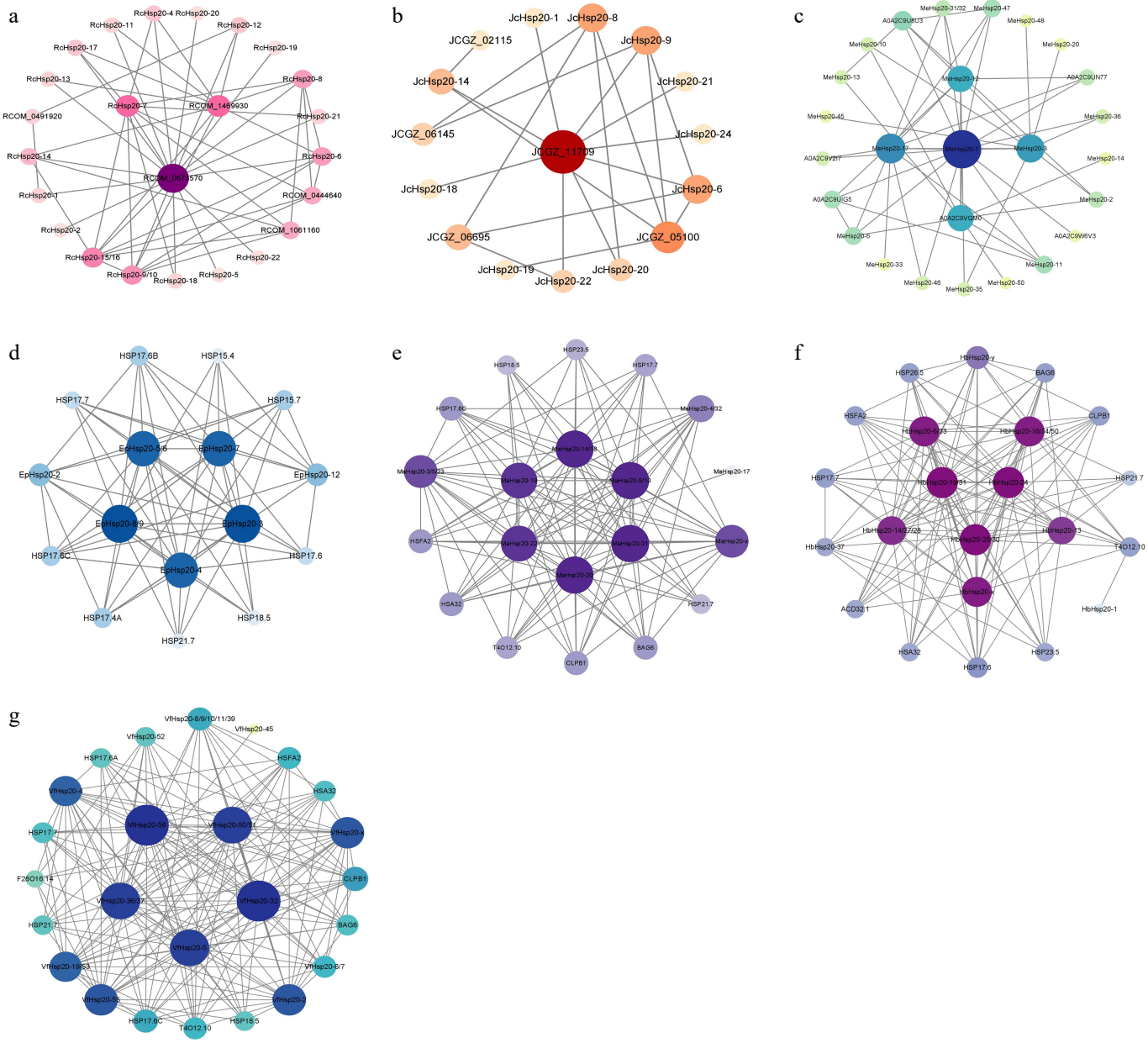


Fig. 7 PPI network of Hsp20 proteins. Circle size and color depth correspond to the number of interacting proteins. (a) *R. communis*; (b) *J. curcas*; (c) *M. esculenta*; (d) *E. peplus*; (e) *M. annua*; (f) *H. brasiliensis*; and (g) *V. fordii*.

the help of lumenal chaperones in the endoplasmic reticulum (ER). The synthesized peptide chains were transferred into the ER and were glycosylated. The correctly folded peptide chains were further transported to the Golgi complex. In contrast, the misfolded peptide chains remained in the ER and bound to the BiP, then were degraded by the proteasome during ER-associated degradation (ERAD) (Supplementary Fig. S9). The ERAD process requires the involvement of the Hsp protein.

Expression pattern profiling of *MeHsp20s*

The transcriptome expression data of *Euphorbiaceae* were downloaded from the GEO database to determine the expression patterns of *Hsp20s* in *Euphorbiaceae*. Among the seven *Euphorbiaceae* species, cassava is the main source of energy for more than 800 million people worldwide and can be used in a variety of industries, including bioenergy, industry, and medicine. Therefore, cassava was selected as a representative species for transcriptome analysis of *Hsp20* genes.

The expression profile of 50 *MeHsp20* genes was visualized by the method of hierarchical clustering (Supplementary Fig. S10). The *Hsp20* genes of cassava showed significant tissue specificity, with the mean expression values of *MeHsp20* genes maintained at high levels in root tubers (SR), leaves, and midveins. In contrast, the 50 *MeHsp20s* displayed noticeably low expression in somatic embryos (OES), with relative expression levels of only about 6. Compared with other genes, *MeHsp20-29* had the highest expression level in SR, while *MeHsp20-1* and *MeHsp20-23* were the most highly expressed genes in leaves and midveins, respectively. Notably, the *MeHsp20* genes displayed distinctive tissue-specific expression patterns. The relative expression of *MeHsp20-3* was high in leaves and midveins; however, it was barely expressed in SR. The relative expression of *MeHsp20-20* was the opposite. Other genes, such as *MeHsp20-14* showed a stable and relatively high expression in the three tissues. Among the 50 *MeHsp20* genes, *MeHsp20-29* was highly expressed in all tissues, especially in SR.

Under drought stress, the results showed that *MeHsp20* genes displayed different expression. More importantly, a clustering analysis demonstrated that the *MeHsp20* genes could be grouped into two clusters (group I and group II) (Fig. 8a). The expression of genes in group 1 was generally lower than that in group 2. The low-expression group (group I) and high-expression group (group II) included 29, and 24 genes, respectively. In group II, there were eight genes with relatively high expression, among which one gene (*MeHsp20-27*) had the highest expression in different varieties, and 2 genes (*MeHsp20-5* and *MeHsp20-20*) had more stable expression (Fig. 8a). However, under cold stress, only two genes (*MeHsp20-15* and *MeHsp20-26*) had higher expression levels than other genes. The expression levels of most genes were below 1 or even undetectable (Fig. 8b). Ultimately, *MeHsp20-26* was upregulated in response to escalating cold stress (Fig. 8b).

Discussion

Euphorbiaceae plants and their products have demonstrated high economic and practical value, and are widely used in industries such as food processing, manufacturing, aerospace, and medicine. The *Euphorbiaceae* species, such as *M. esculenta*, *E. pepluss*, and *H. brasiliensis*, are important economic plants or food sources worldwide. In this paper, a comprehensive analysis of the *Hsp20* gene family in seven representative *Euphorbiaceae* plants was conducted, encompassing biochemical characteristics, functions, and evolutionary relationships.

The genome-wide analysis established the presence of 252 *Hsp20* genes across the seven *Euphorbiaceae* genomes. A greater number of *Hsp20s* were identified in species from the *Crotonaceae* and *Acalyphoideae* subfamilies than in the *Euphorbioideae* subfamily, which evolved in the subtropics (Supplementary Fig. S1; Table 1). Within the *Euphorbiaceae* family, the species with the smallest genome size was not the one with the fewest copies of the *Hsp20s*. Except for the *Euphorbiaceae* family, *Dendrobium catenatum*, whose genome size is 1.01 Gb, contains only 18 *Hsp20s*[39,40]. The phenomenon indicated that the genome size cannot explain the variation in the *Hsp20* gene copy number. The conserved motif, chromosome localization, and phylogenetic analysis results verified the classification and conservation of the *Hsp20* gene family in *Euphorbiaceae* (Figs 2, 4). As observed in rice and tomato, the CI subfamily contained the most *Euphorbiaceae* Hsp20s, and the members within each subfamily shared similar conserved motifs (Fig. 1)[6,41]. In addition to confirming the conservation of coding sequences and conserved motifs of Hsp20s in *Euphorbiaceae*, variations were also found in their conserved motifs. Though the components of the ACD domain were distributed in four conserved motifs, mutations or deletions existed in the components of the ACD domain of several Hsp20s in *Euphorbiaceae*, suggesting potential functional divergence (Fig. 4c; Supplementary Fig. S2). Although there are variations of the *Hsp20* gene within the *Euphorbiaceae* family, the results of the phylogenetic analysis suggested that *Hsp20s* in the *Euphorbiaceae* and *A. thaliana* have gradually shown a tendency to diverge during evolution. Moreover, it is speculated that gene pairs divided into the end of the same branch and from different species may have the same function (Fig. 1). Interestingly, the distribution of the *Hsp20s* on chromosomes in *Euphorbiaceae* was similar to that in tomato and coix, being mainly located at both ends, particularly at the distal ends of the short arms (Supplementary Fig. S5)[41,42].

Whole-genome duplication (WGD) plays a pivotal role in angiosperm morphological and physiological diversity, and the evolution of plant stress resistance[43]. *Arabidopsis thaliana*, a model species for the eudicot clade, has undergone two rounds of WGDs (β WGD and α WGD)[44]. The ancient γ-whole genome triplication event approximately 117 million years ago (Mya) was shared by all core-eudicots, including *Arabidopsis*, poplar, and the four studied species (cassava, castor bean, rubber tree, and physic nut)[45]. This might explain why the *AtHsp20s* found through the WGD events were basically the same (Fig. 2). Subsequently, around 39 to 47 Mya, cassava experienced a genome-wide replication after diverging from physic nut and castor bean. A parallel WGD was also observed in the rubber tree[46]. The analysis results of the WGD events within the genome were consistent with these findings (Supplementary Fig. S4).

To develop greater awareness of the evolutionary relationship between *Hsp20s* in *Arabidopsis* and *Euphorbiaceae*, interspecific collinearity and WGD analyses were performed. Thirty seven collinear *Hsp20* gene pairs between *Euphorbiaceae* and *Arabidopsis* were observed, of which *M. esculenta* accounted for the highest number (11 pairs) (Fig. 2; Supplementary Table S3). The low prevalence of synteny (14.68%) between *Euphorbiaceae* and *Arabidopsis Hsp20s* indicated substantial evolutionary divergence, notwithstanding their notable conservation within the *Euphorbiaceae* family. Notably, the location of 24 out of the 37 gene pairs in chromosomal syntenic blocks provided evidence that they stemmed from ρ WGD (Fig. 2). Among the 24 Hsp20s, 23 Hsp20s were classified into CI, CIII, Po, and M subfamilies, while HbHsp20-27 belonged to the P subfamily (Fig. 1). Interestingly, 24 *Hsp20s* in

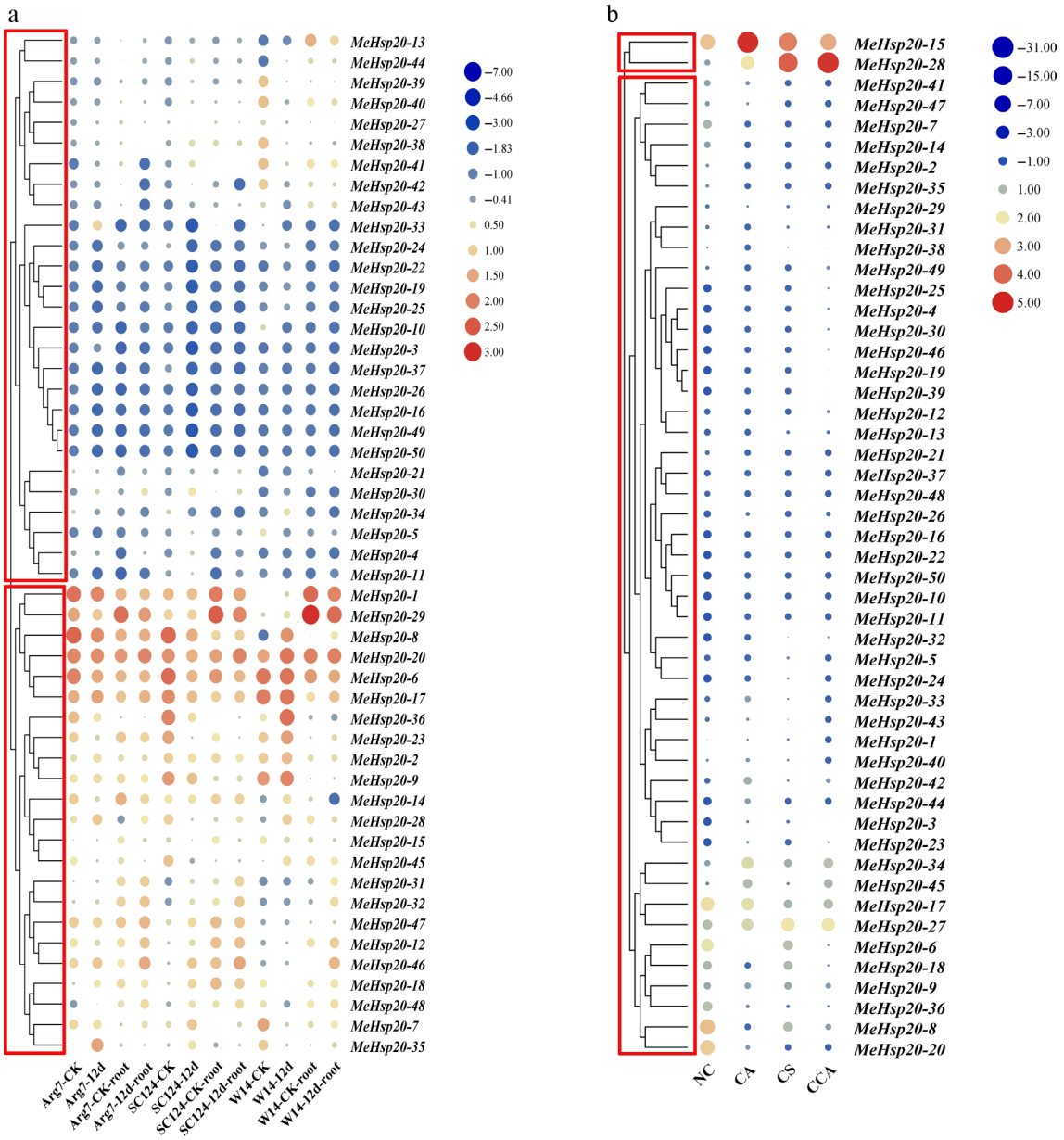


Fig. 8 The expression profile of *MeHsp20* genes under (a) drought, and (b) cold stress. NC: control group; CA: gradual chilling acclimation; CS: chilling shock; CCA: chilling stress after chilling acclimation. Control group; color scale denotes relative expression values: blue (values < 0) for downregulated genes; red (values > 0) for upregulated genes.

Euphorbiaceae were co-located with eight of the 19 *AtHsp20s* in syntenic blocks, which may illustrate that they were derived from *sHsps* in *Arabidopsis* and had similar functions. The nearly identical conserved motifs of genes in the same subfamily further supported this hypothesis (Fig. 4). With the exception of *R. communis*, most *Euphorbiaceae Hsp20s* shared collinear gene pairs with *AtHsp23.5* and *AtHsp23.6* (Fig. 3). Given the significant upregulation of *AtHsp23.5* under heat stress, it is plausible that its homologous genes are also involved in high-temperature response (Fig. 3)[47]. Four genes (*MeHsp20-17*, *EpHsp20-7*, *MaHsp20-14*, and *HbHsp20-30*) located in collinear blocks with the *AtHsp25.3-P*, a HS-responsive gene, may also be involved in responding to HS[48]. *HbHsp20-27*, as the sole gene in the *AtHsp25.3* collinear block, was speculated to be induced by HS. In addition to these 24 *Hsp20s*, the origin of other *Hsp20* genes in *Euphorbiaceae* may be ascribed to intraspecific duplication events, primarily tandem and segmental duplication.

Under various stress conditions, heat shock proteins are produced and are activated by upstream genes to perform specific functions. Analysis of the promoter regions of the *Euphorbiaceae Hsp20* genes revealed that they mainly contained 11 *cis*-acting elements, predominated by light responsiveness elements and then by MeJA-responsiveness elements (Supplementary Fig. S6; Supplementary Table S6). A parallel *cis*-acting element analysis of the genes in *Dendrobium officinale* was consistent with the result, with the analysis in Coix also yielding an abscisic acid responsiveness element[42,49]. Additionally, analysis of wheat *sHsp26* promoter elucidated that it is responsive to multiple abiotic stresses, encompassing not only heat but also salt, cold, and drought conditions[16,50]. It can be inferred that the expression of *Hsp20s* in *Euphorbiaceae* may be regulated by hormones to resist adverse environments.

Many studies have shown that *Hsp20s* respond to temperature stress. The expression of over half of the *CrHsp20s* in the *C. rosea*

seedlings increased significantly after 2 h of heat shock[51]. In *Dendrobium catenatum*, all 18 *DcHSP20* genes were upregulated under high temperature, while six genes were significantly upregulated under low-temperature conditions, with overexpression of *DcHSP20-12* conferred thermotolerance[40]. Cold stress stimulated the upregulation of *TaHSP16.9H-CI* and *TaHSP23.5B-MTI* in bread wheat[52]. Five out of 51 *GmHsp20s* in soybean responded to cold stress[53]. Transcriptome expression analysis revealed that only *MeHsp20-15* and *MeHsp20-28* displayed an upregulation tendency under the cold stress (Fig. 8b), implying that the majority of *Hsp20s* in *Euphorbiaceae* may respond to HS rather than cold stress. Commonly, *Hsf* regulates the transcription of *Hsp* genes by recognizing and binding to the heat shock elements within the *Hsp* promoters when plants experience HS. *HSP17* from wheat was identified as the target gene of *HSFA2h*, and their expression levels were positively correlated under HS in transgenic *Arabidopsis*[54]. Using gel shift and LUC-reporter experiments, Reddy et al. demonstrated that *HvHsfB2c* may regulate the expression of *HvHsp17.5-CI* under HS[55]. Likewise, out of the 252 *Hsp20s* from the *Euphorbiaceae*, 207 had Hsf binding sites, suggesting that they were engaged in reacting to HS (Supplementary Fig. S7; Supplementary Table S7).

Furthermore, it is frequently reported that *Hsp20s* react to other kinds of stress. Stronger resistance to drought and salt was conferred by overexpressing *Arabidopsis thaliana Hsp17.6A*.[56]. Pepper *CaHSP16.4* enhanced the elimination of reactive oxygen species in response to heat and drought stress[57]. A rapid and significant upregulation of multiple *MeHsp20s* was observed after drought stress across different varieties, especially *MeHsp20-35*, *MeHsp20-28* and *MeHsp20-9*, suggesting that they could be involved in drought stress (Fig. 8a). In *Dendrobium catenatum*, the germination rate, fresh weight, and root length of the *DcHSP20-12* overexpressing transgenic *Arabidopsis* lines were significantly higher than those of the wild-type (WT) plants, indicating that this *Hsp20* participated in plant growth and development[40]. Similarly, the consistently stable expression of *MeHsp20-17* across 11 tissues suggested that it may also contribute to this process (Fig. 6; Supplementary Fig. S10). The remaining *Hsp20s* might have different functional roles (Supplementary Tables S7, S8). Transcriptome analysis provides support for the inference. In cassava, some *Hsp20s* showed high expression levels in specific tissues, while others were not, which indicated that different genes play specialized roles in specific tissues (Supplementary Fig. S10).

Nonetheless, it can't be ignored that Hsp20s may function with other proteins or transcription factors to regulate plant resistance to HS. In *Arabidopsis*, *HSP21* abundance could be regulated by *FtsH6* (filamentation temperature sensitive)[58]. The enrichment of ERF transcription factor binding sites in the promoters of *Euphorbiaceae Hsp20s* implicates a functional link to the ethylene signaling pathway (Supplementary Tables S7, S8). The results of the PPI network also indicated that the Hsp20 protein of the *Euphorbiaceae* could form interaction relationships with other proteins, such as Bcl-2-associated athanogene. Bcl-2-associated athanogene (BAG), belonging to the NEF chaperone family, is known to mediate interactions between the Hsp chaperone system and its substrates[59]. In tomato, BAG9 was highly induced and interacted with Hsps in the cytoplasm, promoting the stability of Hsps under HS[59]. In the rubber tree, 12 HbHsp20s were predicted to interact with BAG, implying that they may be protected by BAG (Fig. 7f). The results strongly demonstrated that the *Hsp20* gene can play discrepant functions across different signaling pathways. Future work will further elucidate the functions of these stress-responsive *Hsp20* genes in abiotic stress resistance.

Conclusions

Hsp20 genes have been identified and characterized in several species; however, limited attention has been given to plants within the *Euphorbiaceae* family. This study conducted a genome-wide identification of 252 putative *Hsp20* genes across seven *Euphorbiaceae* species, followed by a comprehensive analysis of their functions, phylogeny, and evolutionary relationships. Moreover, the genome comparison indicated that all syntenic gene pairs underwent segmental duplication and were subjected to purifying selection, with 24 *Hsp20s* in *Euphorbiaceae* potentially originating from β WGD. Additionally, *MeHsp20-17*, *EpHsp20-7*, *MaHsp20-14*, and *HbHsp20-30* may play a role in environmental adaptation. The observed differences in tissue-specific expression and the expression levels of different genes under diverse stress conditions highlight the functional diversity acquired by *Hsp20s* during evolution. In conclusion, these results provide a foundation for further exploration of the functions of the *Hsp20* gene family in *Euphorbiaceae* plants, particularly under adverse conditions, and offer new insights into the regulatory mechanisms governing the *Hsp20* gene family.

Author contributions

The authors confirm their contributions to the paper as follows: study conception and design: Chen Yinhua, Yu X; data collection: Zheng L, Chen Yuhua, Zhang Z, Wang H; analysis and interpretation of results: Zheng L, Ambuyoc MKA; draft manuscript preparation: Zheng L, Jin J, Hamidou Abdoulaye A. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The data that support the findings of this study are available in the NCBI repository. These data were derived from the following resources available in the public domain: PRJNA324539; PRJNA227109; PRJNA246428.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (32260468) and the earmarked fund for the China Agriculture Research System (CARS-11-HNCYH).

Conflict of interest

The authors declare that they have no conflict of interest

Supplementary information accompanies this paper online at: https://doi.org/10.48130/tp-0026-0017.

Dates

Received 6 February 2026; Revised 4 April 2026; Accepted 17 April 2026; Published online 12 June 2026

References

[1] Zhang H, Zhu J, Gong Z, Zhu JK. 2022. Abiotic stress responses in plants. *Nature Reviews Genetics* 23:104–119

[2] Nakamoto H, Vígh L. 2007. The small heat shock proteins and their clients. *Cellular and Molecular Life Sciences* 64:294–306

[3] Bourgine B, Guihur A. 2021. Heat shock signaling in land plants: from plasma membrane sensing to the transcription of small heat shock proteins. *Frontiers in Plant Science* 12:710801

[4] Mogk A, Bukau B. 2017. Role of sHsps in organizing cytosolic protein aggregation and disaggregation. *Cell Stress & Chaperones* 22:493–502
[5] Haslbeck M, Vierling E. 2015. A first line of stress defense: small heat shock proteins and their function in protein homeostasis. *Journal of Molecular Biology* 427:1537–1548
[6] Sarkar NK, Kim YK, Grover A. 2009. Rice sHsp genes: genomic organization and expression profiling under stress and development. *BMC Genomics* 10:393
[7] Waters ER, Vierling E. 2020. Plant small heat shock proteins – evolutionary and functional diversity. *New Phytologist* 227:24–37
[8] Van Montfort R, Slingsby C, Vierling E. 2001. Structure and function of the small heat shock protein/alpha-crystallin family of molecular chaperones. *Advances in Protein Chemistry* 59:105–156
[9] Scharf KD, Siddique M, Vierling E. 2001. 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* 6:225–237
[10] Caspers GJ, Leunissen JAM, de Jong WW. 1995. The expanding small heat-shock protein family, and structure predictions of the conserved 'α-crystallin domain'. *Journal of Molecular Evolution* 40:238–248
[11] Waters ER. 2013. The evolution, function, structure, and expression of the plant sHSPs. *Journal of Experimental Botany* 64:391–403
[12] Helm KW, Schmeits J, Vierling E. 1995. An endomembrane-localized small heat-shock protein from *Arabidopsis thaliana*. *Plant Physiology* 107:287–288
[13] Coca MA, Almoguera C, Jordano J. 1994. Expression of sunflower low-molecular-weight heat-shock proteins during embryogenesis and persistence after germination: localization and possible functional implications. *Plant Molecular Biology* 25:479–492
[14] Ji XR, Yu YH, Ni PY, Zhang GH, Guo DL. 2019. Genome-wide identification of small heat-shock protein (*HSP20*) gene family in grape and expression profile during berry development. *BMC Plant Biology* 19:433
[15] Sun L, Liu Y, Kong X, Zhang D, Pan J, 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 Reports* 31:1473–1484
[16] Chauhan H, Khurana N, Nijhavan A, Khurana JP, Khurana P. 2012. The wheat chloroplastic small heat shock protein (sHSP26) is involved in seed maturation and germination and imparts tolerance to heat stress. *Plant, Cell & Environment* 35:1912–1931
[17] Ouyang Y, Chen J, Xie W, Wang L, Zhang Q. 2009. Comprehensive sequence and expression profile analysis of *Hsp20* gene family in rice. *Plant Molecular Biology* 70:341–357
[18] Siddique M, Gernhard S, von Koskull-Döring P, Vierling E, Scharf KD. 2008. The plant sHSP superfamily: five new members in *Arabidopsis thaliana* with unexpected properties. *Cell Stress & Chaperones* 13:183–197
[19] Xi Z, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, et al. 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proceedings of the National Academy of Sciences of the United States of America* 109:17519–17524
[20] Admase AT, Mersha DA, Kebede AY. 2024. Cassava starch-based hot melt adhesive for textile industries. *Scientific Reports* 14:20927
[21] de Oliveira Schmidt VK, de Vasconscelos GMD, Vicente R, de Souza Carvalho J, Della-Flora IK, et al. 2022. Cassava wastewater valorization for the production of biosurfactants: surfactin, rhamnolipids, and mannosileritritol lipids. *World Journal of Microbiology and Biotechnology* 39:65
[22] Huang L, Zhao H, Yi T, Qi M, Xu H, et al. 2020. Preparation and properties of cassava residue cellulose nanofibril/cassava starch composite films. *Nanomaterials* 10:755
[23] Lubura J, Kočková O, Strachota B, Bera O, Pavlova E, et al. 2023. Natural rubber composites using hydrothermally carbonized hardwood waste biomass as a partial reinforcing filler – Part II: mechanical, thermal and ageing (chemical) properties. *Polymers* 15:2397
[24] Hoy M, Tran NQ, Suddeepong A, Horpibulsuk S, Buritatum A, et al. 2023. Wetting-drying durability performance of cement-stabilized recycled materials and lateritic soil using natural rubber latex. *Construction and Building Materials* 403:133108
[25] Marín-Genescà M, García-Amorós J, Mujal-Rosas R, Massagués L, Colom X. 2020. Study and characterization of the dielectric behavior of low linear density polyethylene composites mixed with ground tire rubber particles. *Polymers* 12:1075
[26] Ewunie GA, Morken J, Lekang OI, Yigezu ZD. 2021. Factors affecting the potential of *Jatropha curcas* for sustainable biodiesel production: a critical review. *Renewable and Sustainable Energy Reviews* 137:110500
[27] Ranucci CR, Alves HJ, Monteiro MR, Kugelmeier CL, Bariccatti RA, et al. 2018. Potentialalternative aviation fuel from jatropha (*Jatropha curcas* L.), babassu (*Orbignya phalerata*) and palm kernel (*Elaeis guineensis*) as blends with Jet-A1 kerosene. *Journal of Cleaner Production* 185:860–869
[28] Yu J, Shang Q, Zhang M, Hu L, Jia P, et al. 2024. Tung oil-based waterborne UV-curable coatings via cellulose nanofibril stabilized Pickering emulsions for self-healing and anticorrosion application. *International Journal of Biological Macromolecules* 256:128114
[29] Sain S, Åkesson D, Skrifvars M, Roy S. 2020. Hydrophobic shape-memory biocomposites from tung-oil-based bioresin and onion-skin-derived nanocellulose networks. *Polymers* 12:2470
[30] Zhou W, Bo C, Jia P, Zhou Y, Zhang M. 2019. Effects of tung oil-based polyols on the thermal stability, flame retardancy, and mechanical properties of rigid polyurethane foam. *Polymers* 11:45
[31] Peres TLC, Ribeiro FV, Aramburu AB, Barbosa KT, Acosta AP, et al. 2023. Polyurethane adhesives for wood based on a simple mixture of castor oil and crude glycerin. *Materials* 16:7251
[32] Zhang W, Deng H, Xia L, Shen L, Zhang C, et al. 2021. Semi-interpenetrating polymer networks prepared from castor oil-based waterborne polyurethanes and carboxymethyl chitosan. *Carbohydrate Polymers* 256:117507
[33] Lin S, Huang J, Chang PR, Wei S, Xu Y, et al. 2013. Structure and mechanical properties of new biomass-based nanocomposite: castor oil-based polyurethane reinforced with acetylated cellulose nanocrystal. *Carbohydrate Polymers* 95:91–99
[34] Lyon CK, Garrett VH. 1973. New castor oil-based urethane elastomers. *Journal of the American Oil Chemists' Society* 50:112–114
[35] Aboukhalaf A, Lahlou Y, Kalili A, Moujabbir S, Elbiyad J, et al. 2024. Antibacterial and antifungal activities of Moroccan wild edible plants selected based on ethnobotanical evidence. *Roczniki Państwowego Zakładu Higieny [Annals of the National Institute of Hygiene]* 75:229–236
[36] Lebwohl M, Swanson N, Anderson LL, Melgaard A, Xu Z, et al. 2012. Ingenol mebutate gel for actinic keratosis. *New England Journal of Medicine* 366:1010–1019
[37] Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, et al. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* 13:1194–1202
[38] Yu C, Leung SKP, Zhang W, Lai LTF, Chan YK, et al. 2021. Structural basis of substrate recognition and thermal protection by a small heat shock protein. *Nature Communications* 12:3007
[39] Zhang GQ, Xu Q, Bian C, Tsai WC, Yeh CM, et al. 2016. The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Scientific Reports* 6:19029
[40] Wang P, Zhang T, Li Y, Zhao X, Liu W, et al. 2024. Comprehensive analysis of *Dendrobium catenatum HSP20* family genes and functional characterization of DcHSP20–12 in response to temperature stress. *International Journal of Biological Macromolecules* 258:129001
[41] Yu J, Cheng Y, Feng K, Ruan M, Ye Q, et al. 2016. Genome-wide identification and expression profiling of tomato *Hsp20* gene family in response to biotic and abiotic stresses. *Frontiers in Plant Science* 7:1215
[42] Hua Y, Liu Q, Zhai Y, Zhao L, Zhu J, et al. 2023. Genome-wide analysis of the *HSP20* gene family and its response to heat and drought stress in Coix (*Coix lacryma-jobi* L.). *BMC Genomics* 24:478
[43] Zhang L, Wu S, Chang X, Wang X, Zhao Y, et al. 2020. The ancient wave of polyploidization events in flowering plants and their facilitated adaptation to environmental stress. *Plant, Cell & Environment* 43:2847–2856
[44] Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438

[45] Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, et al. 2012. A genome triplication associated with early diversification of the core eudicots. *Genome Biology* 13:R3

[46] Zou Z, Xie G, Yang L. 2017. Papain-like cysteine protease encoding genes in rubber (*Hevea brasiliensis*): comparative genomics, phylogenetic, and transcriptional profiling analysis. *Planta* 246:999−1018

[47] Visioli G, Maestri E, Marmiroli N. 1997. Differential display-mediated isolation of a genomic sequence for a putative mitochondrial LMW HSP specifically expressed in condition of induced thermotolerance in *Arabidopsis thaliana* (L.) heynh. *Plant Molecular Biology* 34:517−527

[48] Pan C, Zhou Y, Yao L, Yu L, Qiao Z, et al. 2023. *Amomum tsaoko DRM1* regulate seed germination and improve heat tolerance in *Arabidopsis*. *Journal of Plant Physiology* 286:154007

[49] Wang H, Dong Z, Chen J, Wang M, Ding Y, et al. 2022. Genome-wide identification and expression analysis of the *Hsp20*, *Hsp70* and *Hsp90* gene family in *Dendrobium officinale*. *Frontiers in Plant Science* 13:979801

[50] Khurana N, Chauhan H, Khurana P. 2013. Wheat chloroplast targeted sHSP26 promoter confers heat and abiotic stress inducible expression in transgenic *Arabidopsis* plants. *PLoS One* 8:e54418

[51] Zhang M, Jian S, Wang Z. 2022. Comprehensive analysis of the *Hsp20* gene family in *Canavalia rosea* indicates its roles in the response to multiple abiotic stresses and adaptation to tropical coral islands. *International Journal of Molecular Sciences* 23:6405

[52] Muthusamy SK, Dalal M, Chinnusamy V, Bansal KC. 2017. Genome-wide identification and analysis of biotic and abiotic stress regulation of small heat shock protein (*HSP20*) family genes in bread wheat. *Journal of Plant Physiology* 211:100−113

[53] Lopes-Caitar VS, de Carvalho MC, Darben LM, Kuwahara MK, Nepomuceno AL, 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

[54] Kumar RR, Dubey K, Goswami S, Rai GK, Rai PK, et al. 2023. Transcriptional regulation of small heat shock protein 17 (sHSP-17) by *Triticum aestivum HSFA2h* transcription factor confers tolerance in *Arabidopsis* under heat stress. *Plants* 12:3598

[55] Reddy PS, Kavi Kishor PB, Seiler C, Kuhlmann M, Eschen-Lippold L, et al. 2014. Unraveling regulation of the small heat shock proteins by the heat shock factor *HvHsfB2c* in barley: its implications in drought stress response and seed development. *PLoS One* 9:e89125

[56] Sun W, Bernard C, Van De Cotte B, Van Montagu M, Verbruggen N. 2001. *At-HSP17.6A*, encoding a small heat-shock protein in *Arabidopsis*, can enhance osmotolerance upon overexpression. *Plant Journal* 27:407−415

[57] Huang LJ, Cheng GX, Khan A, Wei AM, Yu QH, et al. 2019. *CaHSP16.4*, a small heat shock protein gene in pepper, is involved in heat and drought tolerance. *Protoplasma* 256:39−51

[58] Sedaghatmehr M, Mueller-Roeber B, Balazadeh S. 2016. The plastid metalloprotease FtsH6 and small heat shock protein HSP21 jointly regulate thermomemory in *Arabidopsis*. *Nature Communications* 7:12439

[59] Huang H, Liu C, Yang C, Kanwar MK, Shao S, et al. 2022. BAG9 confers thermotolerance by regulating cellular redox homeostasis and the stability of heat shock proteins in *Solanum lycopersicum*. *Antioxidants* 11:1467