Journal of Genetics and Applied Biotechnology (JGAB)

Original Article

J Genet Appl Biotechnol, 2026: *e2026023*
DOI: https://doi.org/10.66432/mrq3xt08
Received: 16 Jan 2026; Revised: 08 Apr 2026; Accepted: 16 Apr 2026; Published online: 27 Apr 2026

Molecular Characterization of HSP20 and HSP40 Genes Governing Heat Stress Tolerance in Chili

Maliha Shah[1], Saman Khaliq[1*], Mahek Batool Baig[1]

[1]Department of Plant Breeding and Genetics, PMAS Arid Agriculture University Rawalpindi, Pakistan
[*]Corresponding author: khaliqsaman98@gmail.com

Abstract

Heat stress has a devastating effect on the growth and development of plants as well as on their productivity, it leads to proteins destabilization and cellular homeostasis. The mitigation of these effects depends on heat shock proteins (HSPs). This paper gives a global identification, characterization and expression profiling of Hsp20 and Hsp40 families of genes in chili (*Capsicum annuum*) to understand their functions in heat stress tolerance. There were 64 non-redundant HSP20 genes found, mostly localized in the cytoplasm, then the nucleus, chloroplast and membrane additional findings provided support the hypothesis that HSP20 has a significant role in cytoplasmic protein protection. The structure of genes showed tight folds having a limited number of introns, which allows swift transcriptional stimulation during heat stress. Promoter analysis revealed that there are five high-profile cis-acting elements with ARRIA and WRKY71OS among them, which lend evidence to the belief that stress- and hormone-responsive signals play a complex role in the process of transcription. Genome wide screening of the HSP40 genes provided several cytoplasmic members which when expressed during heat shock indicates that they are involved in the activation of protective molecular chaperone complexes and the maintenance of cellular proteostasis. Hsp20 and Hsp40 proteins were found to coordinate their functions efficiently in stabilization of proteins and aggregation inhibition and maintaining metabolic continuity in the face of thermal stress. Altogether, the results offer useful molecular understanding of the thermotolerance processes in chili and discover candidate genes to molecular breeding and genetic engineering in improving heat stress tolerance in crops.

Keywords: Genome-wide analysis; Heat Stress, Hsp20 Genes; Heat shock proteins; Capsicum annum; HSP40 Genes

How to cite this article: Shah M, Khaliq S, Baig MB. Molecular Characterization of HSP20 and HSP40 Genes Governing Heat Stress Tolerance in Chili. *J. Genet. Appl. Biotechnol.* 2026: e2026023. https://doi.org/10.66432/mrq3xt08

Introduction

Chili (*Capsicum annuum* L.) is a spice and a vegetable that is economically significant and is widely grown globally. It is a member of the Solanaceae family and one of the most widely grown species of genus Capsicum. Besides the role of culinary values, chili is also a highly nutritious food, with vitamins (A, C, and E), carotenoids, and capsaicinoids. (1,2).
Plants are constantly exposed to diverse environmental stresses, including high temperature, drought, salinity, and attack by pathogens, among others, which has dire effects on plant growth, development, and productivity. The heat stress is regarded as one of the worst among these as it interferes with the cellular homeostasis through the denaturation of proteins, destabilization of membranes, and the lack of balance in the metabolism and ultimately deteriorating the performance and yield of the plants (3). Because of their sessile nature, plants have developed complex molecular and physiological responses to environmental stresses (4). Adaptation to stress is a complex signaling network where calcium ions ($Ca2+$) play a central role as secondary messengers because they detect stimuli like temperature extremes, drought, salinity, and pathogen invasion that cause transient changes in cytosolic calcium ions (5). In fact, ABA is a key stress tolerance regulator, especially during drought and biotic stress, which regulates stomata conductance and expression of stress-responsive genes (6,7). Also, ROS are not only harmful by-products but also play a vital role as signalers, combining various stress-response pathways and, thus, organizing the cellular mechanism of defense and acclimation (8).
Complex molecular defense mechanisms have been developed over time by plants to survive under these adverse conditions, with the induction of heat shock proteins (HSPs) being one of the main aspects. The Hsp function as molecular

chaperones, which are involved in the folding, assembly, stabilization, and degradation of proteins during normal and stressful conditions (9). HSPs are divided into a few families depending on their molecular mass and functional characteristics, such as HSP100, HSP90, HSP70, HSP60, HSP40, and small heat shock proteins (sHSPs or HSP20 family) (10). HSP20 and HSP40 proteins have been of much interest among such families because of their critical role in plant stress-tolerance. The HSP20 family are small heat shock proteins with molecular weights between 15- 30 kDa and are defined by cysteine-cysteine conserved a-crystallin domain (11). These proteins are independent ATP molecular chaperones and inhibit irreversible aggregation of denatured proteins in stress situations. Past research has shown that HSP20 proteins are linked with thermotolerance, seed development, and embryogenesis as well as safeguarding against oxidative stress in plant (12).The Hsp40 gene family also known as DnaJ protein family, a significant co-chaperone of Hsp70 that regulates its activation of ATPase and capture of substrates. HSP40 proteins possess a conserved J-domain of interaction with HSP70 (13). According to the composition of their domains, the following three major types are identified; Type I, Type II, and Type III. HSP40 proteins are involved in a wide variety of cellular functions which can be categorized as protein folding, intracellular trafficking, signal transduction as well as responses to abiotic stresses (14).

Overall, biotic and abiotic stresses, especially heat and drought, have led to a significant decrease in chili production since pepper is very sensitive to high temperatures, especially at flowering and fruit development stages, when temperatures above 32 ºC interfere with pollination and fertilization resulting in losses of up to 70% of the yield and poor quality of the fruit (15). Therefore, the improvement of heat tolerance is one of the most important goals in the chili breeding programs. Genome-wide identification and characterization of essential stress-responsive gene families has now been possible with the availability of the entire genome sequence of *Capsicum annuum*. HSP gene families in model and crop plants including Arabidopsis, rice, tomatoes, and maize have been studied in several genome-wide studies and have been shown to play paramount roles in stress adaptation (16) . Nevertheless, little has been done in the comprehensive genome-wide examination of the HSP20 and HSP40 families of genes in chili as well as the expression profiling of HSP40 genes(17).

Thus, the aim of the current research was to identify and describe the HSP20 and HSP40 families of gene in the genome of chili, and subsequently to analyze their phylogenetic patterns, conserved motifs, gene structures, and distribution across chromosomes. Moreover, to clarify the possible role of this gene in heat stress tolerance, expression profiling of an identified HSP40 gene during stress conditions was done as reported in the thesis. The results of this research can be used as important information about the molecular processes of stress adaptation in chili and give possible candidate genes to the genetic improvement program.

Materials and Methods

Plant Material and Genome Data Retrieval

The complete genome sequence of *Capsicum annuum* was accessed in publicly available genome databases. The appropriate protein, coding DNA sequence (CDS), and genome annotation files were downloaded to create a local database for identifying heat shock protein families throughout the genome. The UniProt (https://www.uniprot.org) and NCBI (https://www.ncbi.nlm.nih.gov/gds) databases were searched with homology-based demonstrations of reference HSP20 and HSP40 protein sequences in Arabidopsis thaliana and other plant species.

Identification of HSP20 and HSP40 Gene Families

Putative HSP20 and HSP40 genes were identified by searching the local protein database with BLASTP for the putative HSP20 and HSP40 genes and set the E-value cutoff to < 1 x10-5 (18). Superfluous sequences were eliminated manually. To verify the authenticity of candidate genes, all the obtained sequences were analyzed by conserved domain analysis tools Pfam and SMART databases. Proteins with the conserved α-crystallin domain were selected as the members of the HSP20 family, and proteins with the characteristic J-domain were selected as the members of the HSP40 family (19,20). Proteins that lack the conserved functional domains were not subjected to any further examination.

Physicochemical Properties and Subcellular Localization Analysis

ExPASy ProtParam tool was used to calculate the molecular weights (MW), theoretical isoelectric point (pI), amino acid length, instability index, and the grand average of hydropathicity (GRAVY) values of the identified HSP20 and HSP40 proteins (21). Online prediction tools were used to predict the subcellular localization of the proteins, and their potential cellular localization and functions were found (22).

Chromosomal Distribution and Gene Structure Analysis

Genome annotation files were used to extract chromosomal positions of the HSP20 and HSP40 genes, and positioned on the respective chromosomes of *Capsicum annuum*. The genome-to-CDS sequence alignment was done on the Gene Structure Display Server (GSDS), which evaluated the structure of the genes with exon-intron organization (23). This comparison was possible to evaluate structural diversity and conservation between members of gene families in detail.

Phylogenetic Analysis

The identified HSP20 and HSP40 proteins were multiple sequence aligned with reference sequences of the Arabidopsis thaliana using ClustalW with default parameters (Thompson et al., 1994). The trees of phylogenetics were built following the neighbor-joining (NJ) algorithm in MEGA software, and 1,000 bootstrap replicates were used to test the reliability of the branch (24). The evolutionary relationships and the subgroup formation within both gene families were established by means of phylogenetic classification.

Conserved Motif Identification

The MEME suite was used to identify conserved motifs between the HSP20 and HSP40 proteins with the number of motifs being a maximum of ten and the length of motifs being between 6 and 50 amino acids (25).The Pfam and SMART databases were used to annotate identified motifs as functional and structurally conserved (26).

Plant Material and Stress Treatment for Expression Analysis

Two heat tolerant chili genotypes (24623 and AVRDC-9905) and two heat sensitive genotypes (AVRDC-1209 and 24626) were selected for the analysis of gene expression,. Surface sterilization of the seeds of all the genotypes was done by use of 0.1% mercuric chloride ($HgCl_2$) solution and germination was done under controlled conditions in an growth incubator. The seedlings were left to develop until the seven-leaf stage and thereafter three treatments were administered (1) control (normal growth conditions), (2) heat stress without recovery and (3) heat stress and recovery. The source of heat stress was exposure to high temperature conditions of 48 hours and the recovery period was by subjecting the plants to the ideal conditions after the heat exposure i.e. 24 hours. Collected samples were put into RNA extraction immediately and expression profiling of the chosen HSP20 and HSP 40 genes was carried out to determine their response to the various stress conditions. Downstream expression was done by taking leaf samples at specific time intervals. All the samples were frozen in liquid nitrogen and kept at -80ºC until RNA extraction.

RNA Isolation, cDNA Synthesis and Gene Expression Profiling

Standard RNA extraction protocol was used to extract total RNA in the collected leaf tissues (27). The standard protocol was used to extract total RNA in leaf tissues of treated and control plants. The fresh leaf samples were crushed with liquid nitrogen in order to achieve effective cell disruption and the RNA was isolated using the suggested process. Electrophoresis on 1% agarose gel was used to measure the quality and integrity of extracted RNA, whereas the concentration and purity of RNA were measured spectrophotometrically (260 and 280 nm absorbance) (28). Downstream analysis was only done using high-quality RNA at the right ratio of purity (A260/A280). A commercial reverse transcription kit was used to synthesize first-strand cDNA from the purified RNA using the instructions provided by the manufacturer (29). The resulting cDNA was then employed as a template in the analysis of quantitative gene expression.

For semi quantitative RT-PCR assays, capsicum housekeeping gene (actin) were used as internal control. PCR was performed with 25 or 30 cycles (30 s at 94°C, 30 s at 63°C, and 20s at 72°C) under the following conditions: {0.5 μL RT product was amplified in a 20-μL volume containing 2 μL 10X PCR buffer with MgCl2, 0.25 μL 10 mM dNTPs, and 0.5 μL Taqpolymerase (Tiangen, Beijing)} (30, 31).

Results

Genome-Wide Identification of Small Heat Shock Protein (HSP20) Gene Family in *Capsicum annum*

A total of ninety-six (96) members of HSP20 family genes were retrieved through BLAST searches against NCBI database. Redundant sequences were removed by sequence alignment and phylogenetic tree construction. The resulting sequences were screened in Scanprosite by the confirmation of domains and motifs. After this stringent filtering, 64 members were identified and selected for further analysis (Table 1). The subcellular location of each gene was found in Uniprot and Wolf Psort. Below are the genes and subcellular location of the genes in the HSP20 family. The analysis showed that 26 members were localized in the nucleus including A0A1U8DYF5, A0A1U8E5K0, and A0A1U8GE85. The analysis indicated that 56 members, for instance A0A1U8DWR1, A0A1U8EE11, and A0A1U8F5P6, were located in the cytoplasm. Whereas twenty (20) members, like A0A1U8ED71 and A0A1U8EV16, were found to be located in the chloroplast of the cell. Eleven (11) members, such as A0A1U8DW07 and A0A1U8ENJ0, of the HSP20 gene family were located within cell membrane.

Phylogenetic Analysis of the HSP20 Family

With the help of fig Tree phylogenic trees were constructed and in this tree there are three main clusters which are further branched into sub clusters (Figure 1). Cluster 1 contains sixty nines members which are diversified in various sub clusters. The two closely related members of the cluster one in its first branched cluster are A0A1U8FM97 and A0A1U8FY18 whereas A0A2G2YEL3 and A0A1U8FC12 are the most distant relatives. In the second branch of cluster 1 A0A2G2XZY4 and A0A2G2XXB2 are the closest relatives. In cluster 2 and cluster 3 two members each were found. This analysis shows that HSP20 gene members are highly diverse and classified into various clusters and sub clades correspondence to their evolutionary patterns in *Capsicum annuum* L.

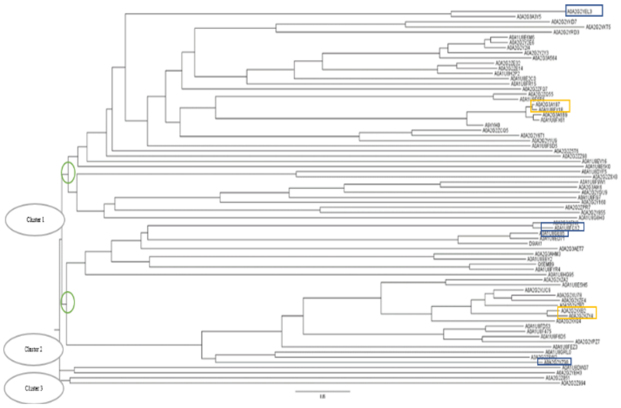

Figure 1: Phylogeny tree of HSP20 gene family

Gene Structure Prediction of HSP20 Family Members

The gene structure prediction of the HSP20 gene family members indicated the range of introns from 0-6 (Table: 2, Figure: 2). Most of the members lacked introns, with only exons. One member of HSP20 consisted of six introns named A0A1U8F6D5. The number of exons ranged from one to twelve. However, maximum number of members consisted of exons ranging from one to three. The member A0A2G2Z994 contained ten exons and A0A2G2ZSX8 consisted of twelve exons.

Promoter Analysis of HSP20 Family Members

For promoter analysis of the members of HSP20 gene family members' 2kb upstream region of coding sequence was analyzed and five (5) cis-acting elements were identified and analyzed for their expression in co expressed genes. The cis-acting elements identified are ARRIA, MYBCORE, MYCCONSENSUSAT, POLLEN1LELAT52, and WRKY71OS. Among which ARRIAT and WRKY71OS were found abundantly in HSP20 gene family members showing their significance during various stresses and developmental stages of chili plant (Table 3).

Table 1: Accession numbers and sub-cellular localization of HSP20 gene family

Accession number	Sub-cellular location	Accession number	Sub-cellular location
A0A1U8DW07	Membrane	A0A2G2Y1U3	Cytoplasm
A0A1U8DYF5	Nucleus	A0A2G2Y2E9	Cytoplasm
A0A1U8E2C0	Cytoplasm	A0A2G2Y2I4	Cytoplasm
A0A1U8E5K0	Nucleus, Chloroplast, Vacuole	A0A2G2Y2Y3	Cytoplasm
A0A1U8E6M6	Cytoplasm	A0A2G2Y6T1	Cytoplasm
A0A1U8ED71	Chloroplast, Mitochondria	A0A2G2Y160	Membrane
A0A1U8EE11	Cytoplasm	A0A2G2Y855	Membrane
A0A1U8EEY2	Cytoplasm	A0A2G2YEH3	Cytoplasm
A0A1U8EV16	Chloroplast, Plastid	A0A2G2YEL3	Cytoplasm
A0A1U8F5P6	Cytoplasm	A0A2G2YGU9	Membrane
A0A1U8F475	Chloroplast, Nucleus, Extra cellular	A0A2G2YKD7	Cytoplasm
A0A1U8FC12	Chloroplast, Mitochondria	A0A2G2YKT5	Cytoplasm
A0A1U8FD53	Extra cellular, Chloroplast, Nucleus, Cytoplasm	A0A2G2YPZ7	Nucleus
A0A1U8FEZ3	Cytoplasm, Nucleus, Extra cellular, Golgi body	A0A2G2YRD3	Cytoplasm
A0A1U8FI97	Membrane	A0A2G2Z5T6	Cytoplasm, Nucleus, Cystokinin
A0A1U8FR15	Cytoplasm	A0A2G2Z051	Cytoplasm, Chloroplast, nucleus, Extra cellular
A0A1U8FX61	Cytoplasm	A0A2G2Z994	Cytoplasm, nucleus
A0A1U8FY18	Cytoplasm	A0A2G2ZCQ5	Cytoplasm
A0A1U8FYR4	Chloroplast, Vacuole	A0A2G2ZE14	Cytoplasm
A0A1U8G8H0	Membrane	A0A2G2ZE32	Cytoplasm
A0A1U8GE85	Mitochondria, Chloroplast, Cytoplasm, Nucleus	A0A2G2ZFQ7	Cytoplasm
A0A1U8GRL0	Nucleus	A0A2G2ZPR7	Membrane
A0A1U8H2P2	Cytoplasm	A0A2G2ZQ55	Cytoplasm
A0A1U8HG95	Membrane	A0A2G2ZSW2	Chloroplast, Extra cellular, Nucleus, cytoplasm
A0A2G2XU78	Nucleus	A0A2G2ZSX8	Nucleus, Cytoplasm
A0A2G2XUC9	Nucleus	A0A2G2ZZ98	Cytoplasm
A0A2G2XXB2	Nucleus	A0A2G3A3Y5	Cytoplasm
A0A2G2XXQ4	Nucleus	A0A2G3A187	Cytoplasm
A0A2G2XZ98	Nucleus	A0A2G3AEN8	Chloroplast, Mitochondria
A0A2G2XZY4	Nucleus	A0A2G3AET7	Chloroplast, Mitochondria
A0A2G2Y1U9	Cytoplasm	A0A2G3AHM3	Cytoplasm
A0A2G2Y2E6	Cytoplasm	A0A2G3AIK6	Membrane

Table 2: Gene structure prediction of the members of the HSP20 gene family in chilli

Accession number	Gene id	Protein id	Transcript id	Domain	Exon
A0A1U8DW07	T459_22142	UP000189700	PHT63922	sHSP	1
A0A1U8DYF5	T459_34420	UP000222542	PHT72396	sHSP	1
A0A1U8E2C0	T459_23181	UP000222542	PHT72101	sHSP	1
A0A1U8E5K0	T459_22886	UP000222542	PHT63920	sHSP	1
A0A1U8E6M6	T459_32218	UP000222542	PHT63919	sHSP	1
A0A1U8ED71	T459_23727	UP000189700	PHT70210	sHSP	1
A0A1U8EE11	T459_25314	UP000189700	PHT93745	sHSP	1
A0A1U8EEY2	T459_01627	UP000189700	PHT63487	sHSP	1
A0A1U8EV16	T459_26981	UP000189700	PHT92703	sHSP	8
A0A1U8F5P6	T459_12556	UP000189700	PHT71827	sHSP	3
A0A1U8F475	T459_22612	UP000189700	PHT92702	sHSP	2
A0A1U8FC12	T459_00584	UP000189700	PHT62827	sHSP	2
A0A1U8FD53	T459_33333	UP000189700	PHT62827	sHSP	2
A0A1U8FEZ3	T459_33333	UP000189700	PHT94004	sHSP	1

A0A1U8FI97	T459_01886	UP000222542	PHT77225	sHSP	2
A0A1U8FR15	T459_07099	UP000222542	PHT90725	sHSP	2
A0A1U8FTF7	T459_11007	UP000222542	PHT87991	sHSP	2
A0A1U8FX61	T459_10097	UP000189700	PHT87992	sHSP	1
A0A1U8FY18	T459_10098	UP000189700	PHT87590	sHSP	1
A0A1U8FYR4	T459_09696	UP000189700	PHT88086	sHSP	2
A0A1U8G8H0	T459_10192	UP000222542	PHT72942	sHSP	2
A0A1U8GE85	T459_23727	UP000189700	PHT85070	sHSP	2
A0A1U8GRL0	T459_13514	UP000222542	PHT80240	sHSP	2
A0A1U8H2P2	T459_18292	UP000189700	PHT80238	sHSP	1
A0A1U8HG95	T459_20368	UP000222542	PHT60861	sHSP	6
A0A2G2XU78	T459_35288	UP000222542	PHT61070	sHSP	2
A0A2G2XUC9	T459_35080	UP000222542	PHT62100	sHSP	2
A0A2G2XXB2	T459_34050	UP000222542	PHT62101	sHSP	4
A0A2G2XXQ4	T459_34051	UP000222542	PHT62825	sHSP	4
A0A2G2XZ98	T459_33331	UP000222542	PHT62827	sHSP	2
A0A2G2XZY4	T459_33137	UP000222542	PHT63689	sHSP	5
A0A2G2Y1U9	T459_32511	UP000222542	PHT63922	sHSP	2
A0A2G2Y2E6	T459_32220	UP000189700	PHT63919	sHSP	1
A0A2G2Y2I4	T459_32217	UP000189700	PHT63920	sHSP	1
A0A2G2Y2Y3	T459_32219	UP000189700	PHT65444	sHSP	1
A0A2G2Y6T1	T459_29869	UP000189700	PHT63487	sHSP	1
A0A2G2Y160	T459_32711	UP000222542	PHT65920	sHSP	2
A0A2G2Y855	T459_30345	UP000189700	PHT66327	sHSP	2
A0A2G2Y992	T459_30752	UP000222542	PHT67494	sHSP	1
A0A2G2YEH3	T459_27622	UP000222542	PHT68001	sHSP	1
A0A2G2YEL3	T459_27488	UP000222542	PHT68950	sHSP	1
A0A2G2YGU9	T459_28437	UP000189700	PHT70210	sHSP	2
A0A2G2YKD7	T459_25314	UP000222542	PHT70211	sHSP	1
A0A2G2YKT5	T459_25315	UP000222542	PHT71827	sHSP	1
A0A2G2YPZ7	T459_22612	UP000222542	PHT72303	sHSP	7
A0A2G2YRD3	T459_23088	UP000189700	PHT72396	sHSP	2
A0A2G2Z5T6	T459_20747	UP000189700	PHT75359	sHSP	2
A0A2G2Z051	T459_18881	UP000222542	PHT78534	sHSP	1
A0A2G2Z994	T459_16586	UP000222542	PHT79776	sHSP	10
A0A2G2ZCQ5	T459_17828	UP000222542	PHT80764	sHSP	1
A0A2G2ZE14	T459_18291	UP000189700	PHT80239	sHSP	1
A0A2G2ZE32	T459_18290	UP000189700	PHT80238	sHSP	1
A0A2G2ZFQ7	T459_13779	UP000189700	PHT80764	sHSP	2
A0A2G2ZPR7	T459_12421	UP000222542	PHT83978	sHSP	2
A0A2G2ZQ55	T459_12556	UP000222542	PHT84113	sHSP	3
A0A2G2ZSW2	T459_13513	UP000189700	PHT85070	sHSP	2
A0A2G2ZSX8	T459_13522	UP000189700	PHT85079	sHSP	12
A0A2G2ZZ98	T459_09410	UP000222542	PHT87304	sHSP	2
A0A2G3A3Y5	T459_11007	UP000189700	PHT88901	sHSP	2
A0A2G3A187	T459_10098	UP000222542	PHT87992	sHSP	1
A0A2G3AEN8	T459_18290	UP000222542	PHT92702	sHSP	2
A0A2G3AET7	T459_12421	UP000222542	PHT92703	sHSP	2
A0A2G3AHM3	T459_12556	UP000189700	PHT93745	sHSP	2
A0A2G3AIK6	T459_13513	UP000222542	PHT94003	sHSP	2

Table 3: Promoter analysis of the members of HSP20 gene family

Accession number	**Arriat**	**Mybcore**	**Mycconsensusat**	**Pollen1lelat52**	**Wrky71os**
A0A1U8DW07	60	20	27	30	21
A0A1U8DWR1	39	7	22	15	12
A0A1U8DYF5	282	55	208	18	145
A0A1U8E2C0	39	6	32	37	18
A0A1U8E5K0	54	12	42	30	32

A0A1U8E6M6	53	5	28	11	19
A0A1U8E820	60	9	32	29	17
A0A1U8ED71	38	11	32	29	27
A0A1U8EE11	44	13	32	24	27
A0A1U8EEY2	161	82	156	82	90
A0A1U8EV16	96	25	70	52	63
A0A1U8F5P6	49	10	36	43	42
A0A1U8F475	101	20	48	78	58
A0A1U8FC12	52	15	42	27	26
A0A1U8FD53	23	6	24	11	8
A0A1U8FEZ3	145	34	84	99	96
A0A1U8FI97	23	3	14	24	9
A0A1U8FM97	55	11	42	23	27
A0A1U8FR15	51	15	36	32	39
A0A1U8FTF7	49	15	56	25	30
A0A1U8FX61	43	5	24	30	19
A0A1U8FY18	91	11	38	53	35
A0A1U8FYR4	51	10	32	38	37
A0A1U8G8H0	50	17	20	33	31
A0A1U8GE85	75	16	54	34	34
A0A1U8GRL0	53	26	34	40	38
A0A1U8H2P2	35	10	20	39	13
A0A1U8HG95	55	13	56	31	46
A0A2G2XU78	39	13	50	15	18
A0A2G2XUC9	40	11	58	19	24
A0A2G2XXB2	41	10	62	15	27
A0A2G2XXQ4	41	10	62	15	27
A0A2G2XZ98	38	21	58	28	36
A0A2G2XZY4	44	8	78	22	35
A0A2G2Y1U9	41	8	32	31	25
A0A2G2Y2E6	39	7	22	15	12
A0A2G2Y2I4	60	9	32	29	17
A0A2G2Y2Y3	16	2	10	6	14
A0A2G2Y6T1	16	2	16	14	8
A0A2G2Y160	23	2	10	8	7
A0A2G2Y855	14	4	22	10	19
A0A2G2Y992	15	3	20	13	10
A0A2G2YEH3	21	2	10	14	10
A0A2G2YEL3	18	6	12	10	10
A0A2G2YGU9	7	0	2	8	2
A0A2G2YKD7	19	2	12	13	14
A0A2G2YKT5	11	2	2	3	2
A0A2G2YPZ7	21	3	8	10	15
A0A2G2YRD3	27	3	14	15	10
A0A2G2Z5T6	24	5	22	11	10
A0A2G2Z051	17	3	18	15	15
A0A2G2Z994	19	4	14	13	8
A0A2G2ZCQ5	24	5	30	9	8
A0A2G2ZE14	16	8	28	13	10
A0A2G2ZE32	29	4	20	9	4
A0A2G2ZFQ7	19	2	12	13	9
A0A2G2ZPR7	29	0	14	17	13
A0A2G2ZQ55	20	1	8	8	10
A0A2G2ZSW2	18	4	22	9	7

A0A2G2ZSX8	20	5	20	13	12
A0A2G2ZZ98	29	6	16	5	9
A0A2G3A3Y5	20	3	18	12	5
A0A2G3A187	15	9	6	7	12
A0A2G3AEN8	26	4	12	14	8
A0A2G3AET7	17	4	16	10	11
A0A2G3AHM3	32	4	12	11	8
A0A2G3AIK6	17	0	6	2	1

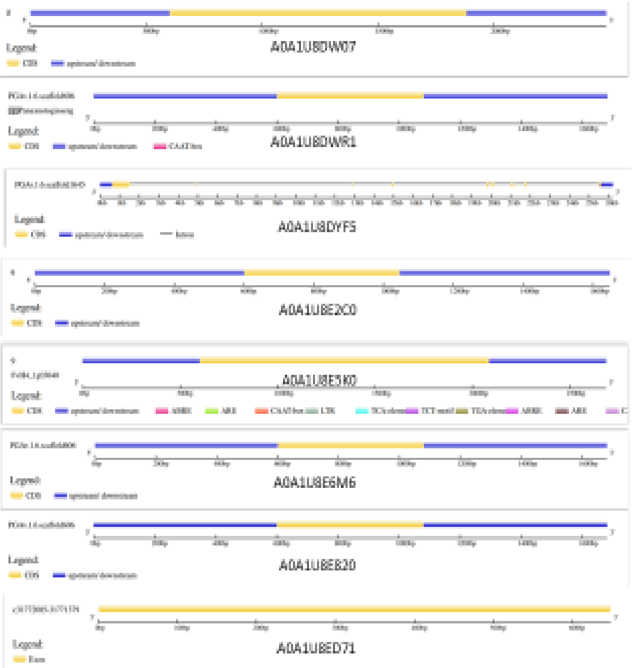

Figure 2: Gene structure of some genes of HSP20 gene family

Genome-Wide Identification, Gene Structure Prediction and Promoter Analysis of Small Heat Shock Protein(HSP40) Gene Family in *Capsicum annum*

Nine (9) members of Hsp40 family genes were identified through BLAST NCBI, and removal of redundancy was carried out by aligning and constructing phylogeny trees which were further screened in Scanprosite by the confirmation of domain and motifs. The subcellular location of each gene was found in Uniprot. Table 4 represents the sub-cellular location of the genes in the HSP40 family. Six members such as A0A2G2ZLQ5, A0A2G2Z9Y1, A0A2G2YSI7, and A0A2G2YFF4 were present in the cytosol. Whereas A0A1U8EAA3, A0A1U8F5W9, and A0A1U8H156 were found to be present in cytoplasm, nucleus and chloroplast. The gene structure of HSP 40 members indicated that the introns were present in the range of one to eighteen (Table 5, Figure 3). A0A2G2ZLQ5 gene structure consisted of 1 exon and 0 introns, A0A1U8EAA3 gene structure indicated 7 exons and 18 introns. Most of the members' gene structure consisted of 3 exons. Heat shock proteins class 40 members containing various promoters that play an active role in the coping of heat stress in chili. As well as they have developmental roles in plants. The abundance of promoters were counted for each promoter in the 2kb upstream region. The promoters include ARRIAT, MYBCORE, MYCCONSENSUSAT, WRKY71OS, and POLLEN1LELAT52 (Table 6).

Phylogenetic Analysis of HSP 40 Family

The phylogenetic tree of HSP 40 shows 3 main clusters which are further classified into sub-clusters (Figure 4). Sub-cluster 1 contains four members, sub-cluster two contains three members, and sub-cluster 3 contains two members. A0A2G2Z9Y1 and A0A1U8H156 are closely related in cluster 1 while A0A2G2Z9Y1 and A0A1U8F5W9 are distantly related. In cluster 2 A0A2G2ZLP5 and A0A2G2YSI7 are distantly related.

Table 4: accession numbers & sub-cellular localization of the members of HSP40 gene family

Accession number	Sub-cellular location
A0A1U8EAA3	Cytoplasm, Chloroplast, Vacuole, Nucleus, Peroxisome,
A0A1U8F5W9	Nucleus, Cytoplasm, Peroxisome
A0A1U8H156	Cytoplasm, Nucleus, Peroxisome,
A0A2G2YFF4	Cytosol
A0A2G2YFJ3	Cytosol
A0A2G2YSI7	Cytosol
A0A2G2Z9Y1	Cytosol
A0A2G2ZLP5	Cytosol
A0A2G2ZLQ5	Cytosol

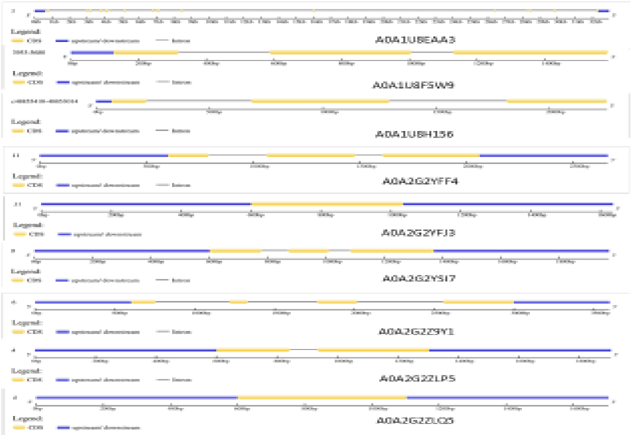

Figure 3: Gene structure of the members of HSP40 gene family

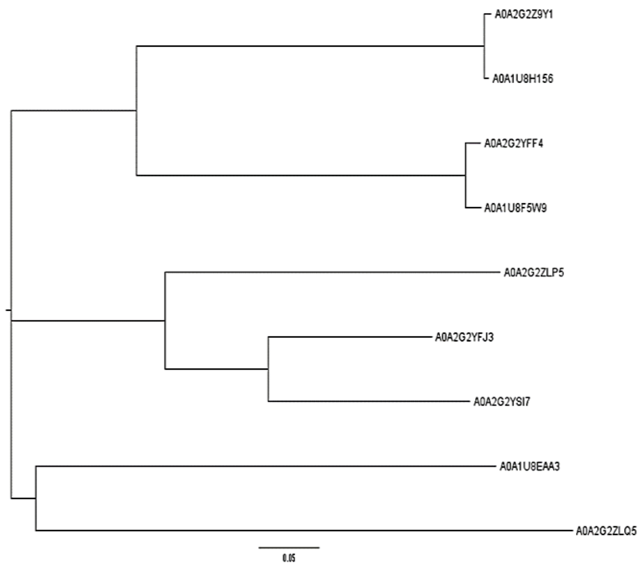

Figure 4: Phylogenetic tree of HSP40 gene family in *Capsicum annum*

Expression Profiling of selected HSP40 Gene (A0A2G2ZLQ5)

The expression pattern of the HSP40 gene A0A2G2ZLQ5 under various genotype and conditions of treatment is present in figure 5. The gene was expressed uniformly in all three samples in genotype G1, with no observed variation. The level of expression was higher than in the control sample than in the treated samples in G 2. On the contrary, G3 genotype exhibited the highest expression in treated samples that have recovered, indicating an improvement in response in the presence of recovery conditions. In G4, there was homogeneity in the levels of expression in all the control samples, which showed stable expression of the genes in the non-stress conditions.

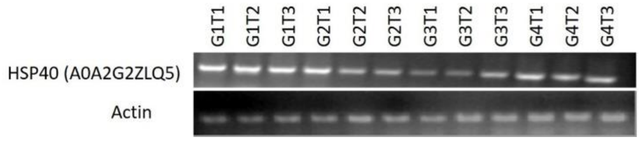

Figure 5: Expression Profiling of Hsp40 (A0A2G2ZLQ5)

Table 5: Gene structure of the members of HSP40 gene family

Accession number	Gene id	Protein id	Transcript id	Domain	Exon
A0A1U8EAA3 1	T459_03065	UP000189700	PHT95183	DnaJ	7
A0A1U8F5W9 2	T459_27942	UP000189700	PHT68455	DnaJ	3
A0A1U8H156 3	T459_16869	UP000189700	PHT78817	DnaJ	3
A0A2G2YFF4 4	T459_27942	UP000222542	PHT68455	DnaJ	3
A0A2G2YFJ3 5	T459_27982	UP000222542	PHT68495	DnaJ	1
A0A2G2YSI7 6	T459_23420	UP000222542	PHT72635	DnaJ	3
A0A2G2Z9Y1 7	T459_16869	UP000222542	PHT78817	DnaJ	4
A0A2G2ZLP5 8	T459_11363	UP000222542	PHT82920	DnaJ	2
A0A2G2ZLQ5 9	T459_11362	UP000222542	PHT82919	DnaJ	1

Table 6: Abundance of major stress related cis-acting element in HSP40 Promoter

Accession number	ARRIAT	MYBCORE	MYCCONSENSUSAT	POLLEN1LELAT52	WRKY71OS
A0A1U8EAA3 1	78	9	17	45	38
A0A1U8F5W9 2	18	6	12	20	4
A0A1U8H156	55	18	57	42	32
A0A2G2YFF4	18	7	10	15	7
A0A2G2YFJ3	14	2	16	12	10
A0A2G2YSI7	17	3	12	16	10
A0A2G2Z9Y1	12	3	6	7	13
A0A2G2ZLP5	41	10	35	24	30
A0A2G2ZLQ5	23	4	16	10	13

Discussion

The current research has found that HSP20 family proteins are localized in different subcellular sites, with most of them being found in the cytoplasm. This cytoplasmic abundance indicates that the proteins are actively involved in stress tolerance processes and other cell developmental processes. The presence of HSP20 proteins in the nucleus, chloroplast and membrane also shows that these proteins are involved in a variety of cellular processes and functional pathways.HSP20 has been reported to have a wider array of functional roles and evolutionary spread as it has been described as more diverse than other heat shock protein families in plants (32). HSP20 genes have been compared in Oryza sativa (23 genes), Populus trichocarpa (36 genes), Arabidopsis thaliana (19 genes), and Glycine max (51 genes) (33). These results indicate that there is a significant range in HSP20 gene family size across plant species and support an evolutionary expansion of the gene family.

According to (34), the absence of introns and the short length of introns may lead to the expression of HSP20 genes during heat stress. The genes that were compacted played a big part in the heat stress tolerance mechanism compared to genes in the structure that contained numerous introns (Jeffares et al., 2008). This may be inferred as the explanation behind the role played by members of the HSP20 gene family in serving as a primary defense mechanism in response to protein aggregation in heat stress to sustain the thermotolerance of plants (35).

The cis-acting factors have been known to be very important during heat stress in the process of controlling the expression of specific genes and safeguard of plants. Heat shock factors (Hsfs) are the activated factors that bind to heat shock elements (HSE) in promoters of HSP genes that control downstream-expression. The fact that CaHSP20 promoters contain multiple cis-elements, such as ARRIA and WRKY71OS, indicates that CaHSP20 genes are not specific to one HSF pathway in terms of its interactions with various functional genes and the regulation of their expression under heat stress (36).

Other plant species, which are further compared, also confirm the importance of WRKY transcription factors in stress response. In Arabidopsis, the microarray data showed that the WRKY transcripts are intensively controlled during abiotic stresses like salinity, drought, and cold (37,38). In Oryza sativa (rice) 54 WRKY genes were found to change significantly in transcription during abiotic stresses, or during the treatment with phytohormones (39). Hv-WRKY38 is a cold and drought stress-responsive gene in barley (Zheng et al., 2020), and at least nine WRKY genes are variously expressed in the abiotic stress condition cases of soybean (40).

A profiling of the expression of HSP40 genes was conducted when subjected to heat stress conditions and it was found that there was a high level of transcriptional induction especially when it was related to the cytoplasm associated members. Such an increased expression pattern during thermal stress implies that they play a very important role in the activation of protective chaperone complexes. This intense induction of HSP40 genes in heat shock indicates that it plays a role in stabilizing nascent and stress-induced proteins to avert the damage of cells and maintain their metabolic activity. This depicts the role of HSP40 genes in the signal of heat stress and adaptive defense mechanisms (41,42)

Hot temperatures have adverse effects on cell membrane integrity and cell membrane functioning. Post-secondary structural alteration of cellular membrane proteins is induced by heat stress, which causes a rise in membrane permeability. The genotypes that have high CMS are more stable to have membrane stability under acid, indicating heat tolerance. The findings are in line with the past results where CMS has been utilized indirectly by measuring electrolyte leakage to screen heat-tolerant genotypes in several crops such as tomato (43,44)

Moreover, the presence of cytoplasmic localization in the HSP20 family members is further validated to play a functional role in denaturation, aggregation, and misfolding prevention of proteins in the presence of heat stress conditions. Such a spatial distribution enables quick access to stress-damaged proteins, which promotes cellular homeostasis, preserving metabolic activity. The small gene size and low levels of introns of HSP20 genes additionally promote rapid transcriptional response, which supports the initial molecular chaperoning response in thermal stress (45). The promoter architecture, cis-regulatory divergence and cytoplasmic preponderance of CaHSP20 genes together show that they play the central role in heat stress tolerance and developmental regulation in chili. The results of these studies are very useful in understanding the molecular mechanism underlying thermotolerance and give good leads on candidate genes that can be used in genetic enhancement programs to improve stress tolerance in chili and other related crop species.

Conclusion

Genome-wide identification and functional characterization of HSP20 and HSP40 family of genes in this study suggested significant roles of these genes in heat stress tolerance in chili. The significance of the HSP20 genes was found to be rapidly responsive to stress by structural features, promoter architecture, and subcellular localization and expression profiling confirmed the importance of cytoplasmic HSP40 members in thermal stress. Collectively, these results give useful molecular clues and candidate genes to improve the thermotolerance of chili by use of molecular breeding and genetic improvement approaches.

Acknowledgement: None.
Disclaimer: None.
Conflicts of Interest: The authors declare no conflict of interest.
Source of Funding: None
Ethics approval statement: Not Applicable
Authors' Contribution Statements: MS: Execution of experiment, data collection. SK, MBB: data collection, proofreading the manuscript. MS, SK: final revision and editing.

References

1. Kim S, Park M, Yeom S-I, Kim Y-M, Lee JM, Lee H-A, et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat Genet. 2014;46(3):270–8.
2. Pickersgill B. The domestication of chili peppers. In: The domestication and exploitation of plants and animals. Routledge; 2017. p. 443–50.
3. Wahid A, Gelani S, Ashraf M, Foolad MR. Heat tolerance in plants: an overview. Environ Exp Bot. 2007;61(3):199–223.
4. Ramegowda V, Senthil-Kumar M. The interactive effects of simultaneous biotic and abiotic stresses on plants:

mechanistic understanding from drought and pathogen combination. J Plant Physiol. 2015;176:47–54.

5. Ranty B, Aldon D, Cotelle V, Galaud J-P, Thuleau P, Mazars C. Calcium sensors as key hubs in plant responses to biotic and abiotic stresses. Front Plant Sci. 2016;7:327.

6. Kim DH, Xu Z-Y, Hwang I. AtHSP17. 8 overexpression in transgenic lettuce gives rise to dehydration and salt stress resistance phenotypes through modulation of ABA-mediated signaling. Plant Cell Rep. 2013;32(12):1953–63.

7. Lim CW, Han S-W, Hwang IS, Kim DS, Hwang BK, Lee SC. The pepper lipoxygenase CaLOX1 plays a role in osmotic, drought and high salinity stress response. Plant Cell Physiol. 2015;56(5):930–42.

8. Sewelam N, Kazan K, Schenk PM. Global plant stress signaling: reactive oxygen species at the cross-road. Front Plant Sci. 2016;7:187.

9. Young JC. Mechanisms of the Hsp70 chaperone systemThis paper is one of a selection of papers published in this special issue entitled “Canadian Society of Biochemistry, Molecular & Cellular Biology 52nd Annual Meeting — Protein Folding: Principles and Diseases” and has undergone the Journal’s usual peer review process. Biochem Cell Biol. 2010;88(2):291–300. doi: 10.1139/O09-175

10. Sedaghatmehr M, Mueller-Roeber B, Balazadeh S. The plastid metalloprotease FtsH6 and small heat shock protein HSP21 jointly regulate thermomemory in Arabidopsis. Nat Commun. 2016;7(1):12439.

11. Guo M, Liu J-H, Lu J-P, Zhai Y-F, Wang H, Gong Z-H, et al. Genome-wide analysis of the CaHsp20 gene family in pepper: comprehensive sequence and expression profile analysis under heat stress. Front Plant Sci. 2015;6:806.

12. Sun W, Bernard C, Van De Cotte B, Van Montagu M, Verbruggen N. At-HSP17.6A , encoding a small heat-shock protein in Arabidopsis , can enhance osmotolerance upon overexpression. Plant J. 2001;27(5):407 – 15. doi: 10.1046/j.1365-313X.2001.01107.x

13. Fan F, Yang X, Cheng Y, Kang Y, Chai X. The DnaJ gene family in pepper (*Capsicum annuum* L.): comprehensive identification, characterization and expression profiles. Front Plant Sci. 2017;8:689.

14. Fan FF, Liu F, Yang X, Wan H, Kang Y. Global analysis of expression profile of members of DnaJ gene families involved in capsaicinoids synthesis in pepper (*Capsicum annuum* L). BMC Plant Biol. 2020;20(1):326. doi: 10.1186/s12870-020-02476-3

15. Erickson AN, Markhart AH. Flower developmental stage and organ sensitivity of bell pepper (*Capsicum annuum* L.) to elevated temperature. Plant Cell Environ. 2002;25(1):123–30. doi: 10.1046/j.0016-8025.2001.00807.x

16. Sarkar NK, Kundnani P, Grover A. Functional analysis of Hsp70 superfamily proteins of rice (Oryza sativa). Cell Stress Chaperones. 2013;18(4):427–37.

17. Sharma D, Masison DC. Hsp70 structure, function, regulation and influence on yeast prions. Protein Pept Lett. 2009;16(6):571–81.

18. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389–402.

19. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46(D1):D493–6.

20. Mandal PK, Collie GW, Srivastava SC, Kauffmann B, Huc I. Structure elucidation of the Pribnow box consensus promoter sequence by racemic DNA crystallography. Nucleic Acids Res. 2016;44(12):5936–43.

21. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. The Proteomics Protocols Handbook. Totowa, NJ: Humana Press; 2005. p. 571–607.

22. Emanuelsson O, Brunak S, Von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc. 2007;2(4):953–71.

23. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31(8):1296–7.

24. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–9.

25. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(suppl_2):W202–8.

26. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46(D1):D493–6.

27. Camacho-Villasana YM, Ochoa-Alejo N, Walling L, Bray EA. An improved method for isolating RNA from dehydrated and nondehydrated chili pepper (*Capsicum annuum* L.) plant tissues. Plant Mol Biol Report. 2002;20(4):407–14.

28. Petrov A, Wu T, Puglisi EV, Puglisi JD. RNA purification by preparative polyacrylamide gel electrophoresis. In: Methods in Enzymology. Elsevier; 2013. p. 315–30.

29. Carninci P, Shiraki T, Mizuno Y, Muramatsu M, Hayashizaki Y. Extra-Long First-Strand cDNA Synthesis. BioTechniques. 2002;32(5):984–5. doi: 10.2144/02325bm01

30. Wang F. Semi-quantitative RT-PCR: An effective method to explore the regulation of gene transcription level affected by environmental pollutants. InEnvironmental Toxicology and Toxicogenomics: Principles, Methods, and Applications 2021 Jun 8 (pp. 95-103). New York, NY: Springer US.

31. Sairam RK, Deshmukh PS, Saxena DC. Role of antioxidant systems in wheat genotypes tolerance to water stress. Biol Plant. 1998;41(3):387–94.

32. Waters ER, Aevermann BD, Sanders-Reed Z. Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. Cell Stress Chaperones. 2008;13(2):127–42.

33. Scharf K-D, Siddique M, Vierling E. 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. 2001;6(3):225.

34. Lopes-Caitar VS, De Carvalho MC, Darben LM, Kuwahara MK, Nepomuceno AL, Dias WP, et al. Genome-wide analysis of the Hsp 20 gene family in soybean: comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses. BMC Genomics. 2013;14(1):577. doi: 10.1186/1471-2164-14-577

35. Haslbeck M, Vierling E. A first line of stress defense: small heat shock proteins and their function in protein homeostasis. J Mol Biol. 2015;427(7):1537–48.

36. Sun J-T, Cheng G-X, Huang L-J, Liu S, Ali M, Khan A, et al. Modified expression of a heat shock protein gene, CaHSP22. 0, results in high sensitivity to heat and salt stress in pepper (*Capsicum annuum* L.). Sci Hortic. 2019;249:364–73.

37. Li S, Khoso MA, Xu H, Zhang C, Liu Z, Wagan S, et al. WRKY transcription factors (TFs) as key regulators of plant resilience to environmental stresses: current perspective. Agronomy. 2024;14(10):2421.

38. Li W, Pang S, Lu Z, Jin B. Function and mechanism of WRKY transcription factors in abiotic stress responses of plants. Plants. 2020;9(11):1515.

39. Ross CA, Liu Y, Shen QJ. The WRKY Gene Family in Rice (Oryza sativa). J Integr Plant Biol. 2007;49(6):827–42. doi: 10.1111/j.1744-7909.2007.00504.x

40. Win PP, Park H-H, Kuk Y-I. Integrated Approach of Using Biostimulants for Improving Growth, Physiological Traits, and Tolerance to Abiotic Stressors in Rice and Soybean. Agronomy. 2025;15(10).

41. Qiu X-B, Shao Y-M, Miao S, Wang L. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. Cell Mol Life Sci CMLS. 2006;63(22):2560–70.

42. Wang Q, Wei W, Liu Y, Zheng Z, Du X, Jiao Y. HSP40 gene family in pearl oyster Pinctada fucata martensii: genome-wide identification and function analysis. Fish Shellfish Immunol. 2019;93:904–10.

43. Alsamir MA. Genetic and physiological analysis of tomato (*Solanum lycopersicum* L.) adaption under heat and disease stress. Doctoral dissertation 2019, The University of Sydney.

44. Usman MG, Rafii MY, Ismail MR, Malek MA, Latif MA. Expression of target gene Hsp70 and membrane stability determine heat tolerance in chili pepper. J Am Soc Hortic Sci. 2015;140(2):144–50.

45. Jeffares DC, Penkett CJ, Bähler J. Rapidly regulated genes are intron poor. Trends Genet. 2008;24(8):375–8.