

Computational analysis of heat shock protein (HSP) in *Citrus X Sinensis*^a Samia Nawaz, ^a Lubna Rasool, ^a Abdur Rauf, ^b Sumaira Rasul^a Department of Chemistry, University of Sahiwal, Sahiwal, Pakistan,^b Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University (BZU), Multan, Pakistan.

Authors' Contribution	Nawaz, S. and A. Rauf completed swiss-modeling analysis, S. Rasul performed MEGA 7 analysis and L. Rasool performed Wolf-Psort analysis.
------------------------------	--

*Corresponding Author's Email Address: lubnarasool@uosahiwal.edu.pk

ABSTRACT

Review Process: Peer review

Heat shock proteins (HSPs) are molecular chaperones and one of the cell's most important regulatory proteins present in all species. HSPs are a multigene family classified into six families according to their molecular weight range between 8KDa to 110KDa: HSP100, HSP90, HSP70, HSP60, small heat shock proteins (sHSPs), and ubiquitin. *Citrus X Sinensis* is very economical, and widely grown in the world. It is one of the best sources of nutrition, vitamin C, fiber, thiamine, folate, and antioxidants. It has a variety of health benefits i.e., protects against cell damage, boosts the immune system, heals wounds, etc. For highlighting the aspect of HSPs of citrus in computing protein, this work has great significance. Present studies are based on the identification and characterization of heat shock proteins in *Citrus sinuses*. *In silico* research was performed; *Arabidopsis thaliana* protein sequences of HSP60, HSP70 and HSP90 were used as a reference to identify these molecular chaperones in *Citrus sinuses* through Blast Phytozome. Furthermore, physiochemical properties, multiple sequence alignment, Multiple Expectation Maximization for Motif Elicitation (MEME) and phylogeny were constructed for 44 sequences of HSP60, 22 sequences of HSP70 and 23 for HSP90. The best expression of HSPs was found in Endoplasmic Reticulum. The findings of the present studies provide a good understanding of the HSP60, HSP70 and HSP90 families' function and expression site in the cell of *Citrus sinuses*. It also provides the basic knowledge for further investigations of the putative structure of dynamic proteins.

Keywords: Physiochemical properties, motif analysis, multiple expectation maximization, phylogenetic analysis, swiss-Modeling, cellular localization.

INTRODUCTION: Heat shock proteins (HSPs) are one of the cell's most important regulatory proteins present in the cells of all species. These are the special type of conserved proteins produced when a cell is exposed to heat shock or other stresses. Indeed, the far more important function of HSP is to protect cells from heat shock by preventing the denaturation of cellular proteins (Kregel, 2002). Some of the members of the heat shock protein family are also called molecular chaperones. Molecular chaperones assist in the proper assembly and packing of polypeptide chains. In addition to their key function in protein folding, chaperones are also involved in protein transportation and deterioration, dissociation of protein clusters and refolding of stress-denatured proteins, and macromolecular matrix formation (Haslbeck *et al.*, 2005). HSPs elevated by several environmental conditions, metabolic stresses, physiological and pathophysiological factors such as inflammatory mediators, toxic substances, alkalinity, heavy metal ions, anoxia, ethanol, ischemia, drugs, drought, fungi, aging many more (Tutar *et al.*, 2006; Wu and Tanguay, 2006; van Noort *et al.*, 2012). Abbas *et al.* studied the SIhsp70 gene family in *Solanum lycopersicum* under Cd²⁺ stress condition. He found higher expressions of SIhsp70-11 when Cd²⁺ stress is applied to it (Abbas *et al.*, 2022). Similarly, Davoudi *et al.* also analyzed the HSP70 gene family in *Cucurbita moschata* under drought conditions (Davoudi *et al.*, 2022). The functions of HSPs in strawberry plants at the elevated temperature and found that when heat stress was applied a range of HSPs produced like 19-29 kDa in leaves, and 16-26 kDa in flowers (Ledesma *et al.*, 2004). Heat shock proteins are a multigene family based on their molecular weight with molecular sizes from 8-110 kDa. Depending upon molecular weight, the HSP families are divided into six classes which are HSP100, HSP90, HSP70 (chaperone), HSP60 (chaperone), small sHSPs, and Ubiquitin (Bakthisaran *et al.*, 2015). HSP90 are the most plentiful proteins in the cell (1-2% of overall cellular proteins) (Tutar and Tutar, 2010). The main function of HSP90 family is to control protein folding. Cell cycle control, transcriptional regulation, protein trafficking, and signal transduction are the other main contribution of HSP90 in the cell except protein folding (Wang *et al.*, 2004). This family is the most conserved protein family among all the HSP families since the process of evolution. HSP70s play crucial roles in many important biochemical functions. HSP40, HSP100, and many other nucleotide factors act as co-chaperones with HSP70 for proper protein folding (Kabani *et al.*, 2002). Other cellular processes involving HSP70 include apoptosis, targeting proteins for degradation translocation across membranes, and some other functions managed by HSP70 (Tutar, 2006; Tutar *et al.*, 2006). The folding and assembly of proteins and their subunits are aided by HSP60. It also has multiple forms in the cell. This family protein is mostly located in mitochondria and chloroplast (Efeoğlu, 2009). Bacterial and eukaryotic cells are bountiful with this family (Arsène *et al.*, 2000). A group of genera called *Citrus* is present

including many cultivated plants such as *Citrus grandis* (pummelo), *Citrus limon* (lemon), *Citrus reticulata* (tangerine and mandarin), *Citrus paradisi* (grapefruit) and *Citrus X Sinensis* (sweet orange) (Barkley *et al.*, 2006). *Citrus* belongs to the Rutaceae family (Moore, 2001). *Citrus X Sinensis* is the most extensively grown citrus group in the world, comprising over 70% of per year *Citrus* species production (Flamini *et al.*, 2003). From a nutritional point of view, sweet oranges are very economical. They contain excess amounts of vitamin C, fiber, thiamine, folate, and antioxidants. They have a variety of health benefits (Xu *et al.*, 2013). The importance of HSP in sequence study includes genomic characterization and cellular physiology to fully comprehend the *Citrus X Sinensis* evolutionary relationship, physiochemical properties, etc.

OBJECTIVES: This study aimed to provide deep insights into HSP families in *Citrus X Sinensis*. Our performed work provides physiochemical properties, phylogeny, and all the functional genomic architecture at the genomic level of HSP60, HSP70, and HSP90 families. All these provide the basis for future studies.

MATERIALS AND METHODS: Identification of HSP genes in *Citrus X Sinensis*: *Arabidopsis thaliana* was used as a reference sequence. A published sequence of HSP60 (15229559), HHSP70 (820438), and HSP90 (9758621) of *Arabidopsis thaliana* was obtained from the "National Centre for Biotechnology Information (NCBI)" (<https://www.ncbi.nlm.nih.gov/>) and domains were analyzed by the "Pfam Protein Family Database" (<http://pfam.xfam.org/>). A homology search was performed to identify the reported HSP families in *Citrus X Sinensis*. The obtained genome sequence of *Citrus X Sinensis* was analyzed from "Phytozome" (<https://phytozome.jgi.doe.gov/>). HSP family's domains of *Citrus X Sinensis* were analyzed by "Pfam" (<http://pfam.xfam.org/>) (Mistry and Finn, 2007). Moreover, conserved protein motifs of HSP were analyzed by using the "Multiple Expectation Maximization for Motif Elicitation (MEME)" (<https://meme-suite.org/meme/>) with a maximum of 8 motifs (Bailey *et al.*, 2006).

Characterization of physiochemical properties of HSP sequences: to characterize the physiochemical properties of obtained hsp sequences of *citrus x sinensis*, the "protparam" tool (<https://web.expasy.org/protparam/>) was used (Lusiana and Irfannuddin, 2022). one by one, protein sequence was submitted at protparam to identify the instability index (ii), molecular weight (mw), aliphatic index (ai), grand average of hydropathicity (gravy), number of amino acids (aa) and theoretical isoelectric point (pi). similarly, the genomic sequence, transcript sequence, strand, and cds (bp) were obtained by using "phytozome" (<https://phytozome.jgi.doe.gov/>) (Rehman *et al.*, 2020).

Cellular localization of HSP: To identify the cellular positions of obtained sequences, the Protein Subcellular Localization Prediction Tool "Wolf Psort" (<https://wolfsort.hgc.jp/>) was used (Horton *et al.*, 2007). Sequences of each family (collectively) were tested at Wolf Psort and values of each sequence for different organelles

were collected from there. Then for further analysis, the "TB tool" was used. Data obtained from Wolf Psort were further used to construct the heatmaps of respective families.

Phylogeny: Multiple sequence alignments of HSP60, HSP70 and HSP90 were performed. MEGA7 was used to conduct phylogenetic analysis using the results of multiple sequence alignment (Kumar *et al.*, 2016). Using the Maximum Likelihood technique based on the Jones-Taylor Thornton (JTT) matrix-based model with a 1000 bootstrap value, the evolutionary history between HSP60, HSP70 and HSP90 was computed.

Structure of Protein: 3D structures of proteins provide a better understanding of protein structures, binding sites and activities. A template search was conducted using Swiss-Modeling by entering amino acid sequences as input data (<https://swissmodel.expasy.org>). The templates with the highest percentage similarity, GMQE and QMEAN scoring functions were used to evaluate the global and per-residue model quality. The automated modeling procedure started with the selected template having the highest percentage of similarity (Pembroke, 2000).

RESULTS: Identification of HSP family: By using the *Arabidopsis thaliana* collected sequences as queries, 65 hits for HSP60, 44 hits for HSP70, and 15 hits for HSP90 of *Citrus X Sinensis* were collected from Phytozome BLAST analysis. Then the duplicated and unwanted sequences were removed from these hits. The remaining sequences were 44, 22, and 23 for HSP60, HSP70, and HSP90 respectively.

Physicochemical characterization of HSP families: Table 1, S1 and S2 represented the physicochemical properties of the collected HSPs sequences. The HSP isoforms have a wide diversity in their molecular weight (35000KDa to 87000KDa), number of amino acids (250 to 800), genomic sequence (1200 to 7400), transcript sequence (800 to 3500), and coding sequence (800 to 2400). An Ii of greater than 40 describes that the protein is unstable. The present study (table S1) revealed that all the members of HSP60 are stable except CSHSP60-10, CSHSP60-18, CSHSP60-21, CSHSP60-22, CSHSP60-25, CSHSP60-36, CSHSP60-37, CSHSP60-50, CSHSP60-97, and CSHSP60-108. An AI greater than 65 describes the thermostability of HSP. A lower GRAVY value indicates the hydrophilic nature of members. Thus, all the members of the HSP60, HSP70 and HSP90 families are thermostable and hydrophilic except CSHSP60-92, CSHSP60-80, CSHSP60-69, CSHSP60-52, CSHSP60-44, CSHSP70-24, CSHSP90-37, (table 1, S1 and S2). The pI value indicated the acidic and basic nature of HSP. The pI value shows that most of the members of the HSP60 family are acidic. But some exceptions are present such as CSHSP60-28, CSHSP60-52, CSHSP60-56, CSHSP60-80, CSHSP60-86, CSHSP60-97, CSHSP60-101, CSHSP60-102, CSHSP60-106, and CSHSP60-108, these are basic in nature. Out of 44 sequences, 25 are reverse strands and 19 are forward strands. Table 1 revealed that all the members of HSP70 are acidic except CSHSP70-18, CSHSP70-35, CSHSP70-38, CSHSP70-41, and CSHSP70-43 which are basic. Except for CSHSP70-7, CSHSP70-12, CSHSP70-14, CSHSP70-16, CSHSP70-35, and CSHSP70-43, all the members are stable. Out of 22 members, 11 are reverse strand while the remaining all are forward. The AI value describes that all members are thermally stable (table 1). The physicochemical properties of the HSP90 family provided in Table S2 indicate that all members are acidic except CSHSP90-28, CSHSP90-39, and CSHSP90-49. Almost 15 members are forward strands while 8 are reverse strands. Study reveals that HSP90 family is stable in nature except CSHSP90-10, CSHSP90-18, CSHSP90-21, CSHSP90-22, CSHSP90-23, CSHSP90-25, CSHSP90-33, CSHSP90-34, and CSHSP90-49.

Localization of HSP: To find the occurrence and expression of collected HSP sequences in the cell, detailed research was performed by Wolf Psort. A heat map of each HSP family was constructed using the TB tool. Heat maps are shown in figure 1(A) for HSP60, HSP70, and HSP90 respectively. Figure 1(A) confirmed that CSHSP60-10, CSHSP60-61, and CSHSP60-82 have maximum expressions in E.R. CSHSP60-21, CSHSP60-25, CSHSP60-52, CSHSP60-101, and CSHSP60-108 have the high number of expressions in the nucleus. CSHSP60-19, CSHSP60-69, CSHSP60-85, and CSHSP60-97 are present in excess amounts in the chloroplast. From all the sequences, CSHSP60-101 has the maximum value in the cytoskeleton. CSHSP60-80 has maximum expression in plastids. CSHSP60-83 shows the highest expression in the peroxisome. CSHSP60-17, CSHSP60-27, CSHSP60-42, and CSHSP60-107 show mitochondrial abundance. The least number of

expressions is shown in vacuole and extracellular. Except CSHSP60-10, CSHSP60-27, CSHSP60-42, CSHSP60-61, CSHSP60-69, CSHSP60-82, CSHSP60-85, CSHSP60-86, and CSHSP60-107 all other sequences show cytoplasmic expressions. The CSHSP70-07 and CSHSP70-30 have maximum expressions in E.R. CSHSP70-14 and CSHSP70-16 show the highest occurrence in the nucleus. CSHSP70-12, CSHSP70-21, CSHSP70-24, CSHSP70-32, CSHSP70-33, CSHSP70-34, CSHSP70-36, CSHSP70-38, and CSHSP70-39 have maximum cytoplasmic occurrence. Chloroplast consists of a high amount of CSHSP70-13 and CSHSP70-41. CSHSP70-17 and CSHSP70-42 are present in high amounts in mitochondria. Peroxisome consists of a high amount of CSHSP70-13 and CSHSP70-61. Similarly, CSHSP90-21, CSHSP90-25 and CSHSP90-63 have maximum occurrence in the nucleus. CSHSP90-47, CSHSP90-48, CSHSP90-50, and CSHSP90-52 have the highest expression in the cytoplasm. CSHSP90-19, CSHSP90-34, and CSHSP90-49 have high occurrences in the chloroplast. The maximum expressions in ER are shown by CSHSP90-10, CSHSP90-42, and CSHSP90-44. In mitochondria, CSHSP90-27 is present in abundance. Peroxisome consists of a high amount of CSHSP90-12 and CSHSP90-45.

Motif Analysis: For the sake of a better understanding of heat shock protein sequences, conserved motif analysis was predicted by the MEME suit. Through MEME suit maximum of eight motifs of HSPs were observed in *Citrus X Sinensis*. MEME motifs 1, and 5 consists of 29 amino acids, motifs 2,3 and 4 consist of 50 amino acids, while motif 7 and 8 consist of 41 amino acids. Only MEME motif 6 consists of 42 amino acids (table 2). All the motifs were annotated as HSP70 in the Pfam search. Motif analysis of HSP sequences is shown in figure. CSHSP60-19, CSHSP60-27, CSHSP60-30, CSHSP60-31, CSHSP60-56, CSHSP60-97, CSHSP60-98, CSHSP60-103, CSHSP60-106, CSHSP70-13, CSHSP70-17, CSHSP70-20, CSHSP70-21, CSHSP70-36, CSHSP70-39, CSHSP70-41, CSHSP90-19, CSHSP90-27, CSHSP90-30, CSHSP90-31, and CSHSP90-50 show maximum number of motifs while CSHSP60-93 and CSHSP70-32 show least number of motifs.

Phylogeny: To study the evolutionary history maximum likelihood method was used based on JTT matrix-based model. Bootstrap value was 1000 for taxa analysis (figure 1B). The evolutionary aspect shows the gene changes in the organism and mutation against certain environmental stress. Overall figure showed that the HSP 60 gene is more closely related to HSP-70 and is distantly related to HSP-90.

Protein structure prediction: The SWISS-MODEL template library was searched for structures that were evolutionarily related to the target sequence. Amino acids sequences of all the respective HSPs inserted as input source sequences and a template search was conducted. Those sequences having a percentage similarity of less than 50% were eliminated. Sequences having the highest percentage of similarity were selected from respective HSP families. Global Model Quality Estimation (GMQE) and QMEAN scoring functions were also evaluated for each respective model. The SWISS-MODEL template library was searched for structures that were evolutionarily related to the target sequence. Amino acids sequences of all the respective HSPs inserted as input source sequences and a template search was conducted. Those sequences having a percentage similarity of less than 50% were eliminated. Sequences having the highest percentage of similarity were selected from respective HSP families. CSHSP60-93, CSHSP70-21 and CSHSP90-30 sequence three-dimensional structure as shown in figures 1(C), S2 and S3 respectively. Global Model Quality Estimation (GMQE) and QMEAN scoring functions were also evaluated for each respective model. These putative protein structures showed an 80.55% to 82% similarity index with the templates. All three have a monomer-oligo state.

DISCUSSION: HSPs are the most conserved proteins present in all species and perform a wide variety of functions. HSPs improved tolerance to cell death caused by a variety of factors. Generally, HSP90, HSP70 and HSP60 are used to control and suppress the process of apoptosis (Kennedy *et al.*, 2014). HSPs are involved in nascent chain folding, protein disaggregation, degradation of damaged proteins, protein translocation through membranes, and cytoskeleton preservation and restructuring (Weibezahn *et al.*, 2005). The HSP90 family is a unique and highly essential protein family present in all species. The main function of this family is to control protein folding (Wang *et al.*, 2004). HSP70 has an indispensable role under non-stress conditions. Thermal sufferance and resistance enhanced to environmental stresses are

Sr No.	Gene renames	Strand	GS	AA	MW (kDa)	pI	AI	II	GRAVY
1	CSHSP70-7	R	5315	776	86299.19	5.61	91.74	42.39	-0.294
2	CSHSP70-12	R	4916	726	80485.36	5.17	81.12	43.41	-0.371
3	CSHSP70-13	R	4163	714	76618.42	5.31	85.46	27.25	-0.332
4	CSHSP70-14	F	5785	708	78205.76	5.57	84.68	42.06	-0.274
5	CSHSP70-16	R	5243	692	77370.16	5.71	86.49	41.22	-0.324
6	CSHSP70-17	F	3414	676	72557.26	5.68	88.21	37.98	-0.294
7	CSHSP70-18	F	2145	675	75062.33	9.27	79.63	36.71	-0.326
8	CSHSP70-20	F	2928	657	72773.72	5.19	89.21	33.02	-0.429
9	CSHSP70-21	F	2990	647	70889.37	5.16	83.08	34.17	-0.398
10	CSHSP70-24	R	3631	573	62139.38	5.28	98.29	37.4	0.017
11	CSHSP70-31	F	1432	428	46473.99	6.41	93.9	30.34	-0.095
12	CSHSP70-32	R	1222	330	35898.5	4.75	82.79	29.08	-0.542
13	CSHSP70-33	R	2121	315	34296.47	4.96	72.83	36.78	-0.63
14	CSHSP70-35	R	3025	493	54549.79	8.7	89.17	45.04	-0.128
15	CSHSP70-36	F	1971	624	68324.14	5.2	81.46	34.95	-0.456
16	CSHSP70-37	F	1452	412	45179.18	5.26	84.32	33.97	-0.51
17	CSHSP70-38	F	1425	475	52310.65	7.51	81.71	37.39	-0.289
18	CSHSP70-39	R	1953	650	71054.44	5.43	82.4	32.68	-0.412
19	CSHSP70-40	F	1452	443	48886.92	5.39	91.11	32.02	-0.313
20	CSHSP70-41	R	3355	593	64042.78	7.17	89.73	32.99	-0.226
21	CSHSP70-42	R	3248	651	69956.3	6.01	86.48	35.82	-0.348
22	CSHSP70-43	F	1620	539	60223.07	8.93	78.72	44.34	-0.319

Table 1: Physicochemical properties of HSP70 family members.

[R (reverse), F (forward), GS (genomic sequence), AA (number of amino acids), MW (molecular weight in kilo Dalton), pI (Isoelectric point), AI (Aliphatic Index), II (Instability Index), GRAVY (Grand Average of hydrophobicity Index)]

Motif	Protein sequence	No. of Units
1	FNGKEPCKHINPDECVAYGCAVQCAILSG	29
2	VKNCVITVPAYFNDSQRQATKDAGRIAGLNVMRIIHEPTAAAIAAYGLDKK	50
3	TRARFEELNMDLFRKCMPECKELRDAKMDKSDVHDVVLVGGSTRIPKQV	50
4	IQVYEGERLTRDNNLLGKFLMVGIPAPRGVPPQINVCFDIDANGILHVS	50
5	VKATAGDTHLGGEDFDNRMVFNHFQWQEFKR	29
6	KHKIDISGNPRALQRLRTACERAKRTLSSTHQTTEIDCLYE	42
7	ERWIGDAAKNQAMNPTNTVFDKRLIGRRFSDPSVQSDMK	41
8	IGIDLGTTYSCVGVWQHDRPEIANDQGNRTTPSYVAFTDT	41

Table 2 Eight differentially conserved motifs of heat shock protein sequences observed in *Citrus X Sinensis*.

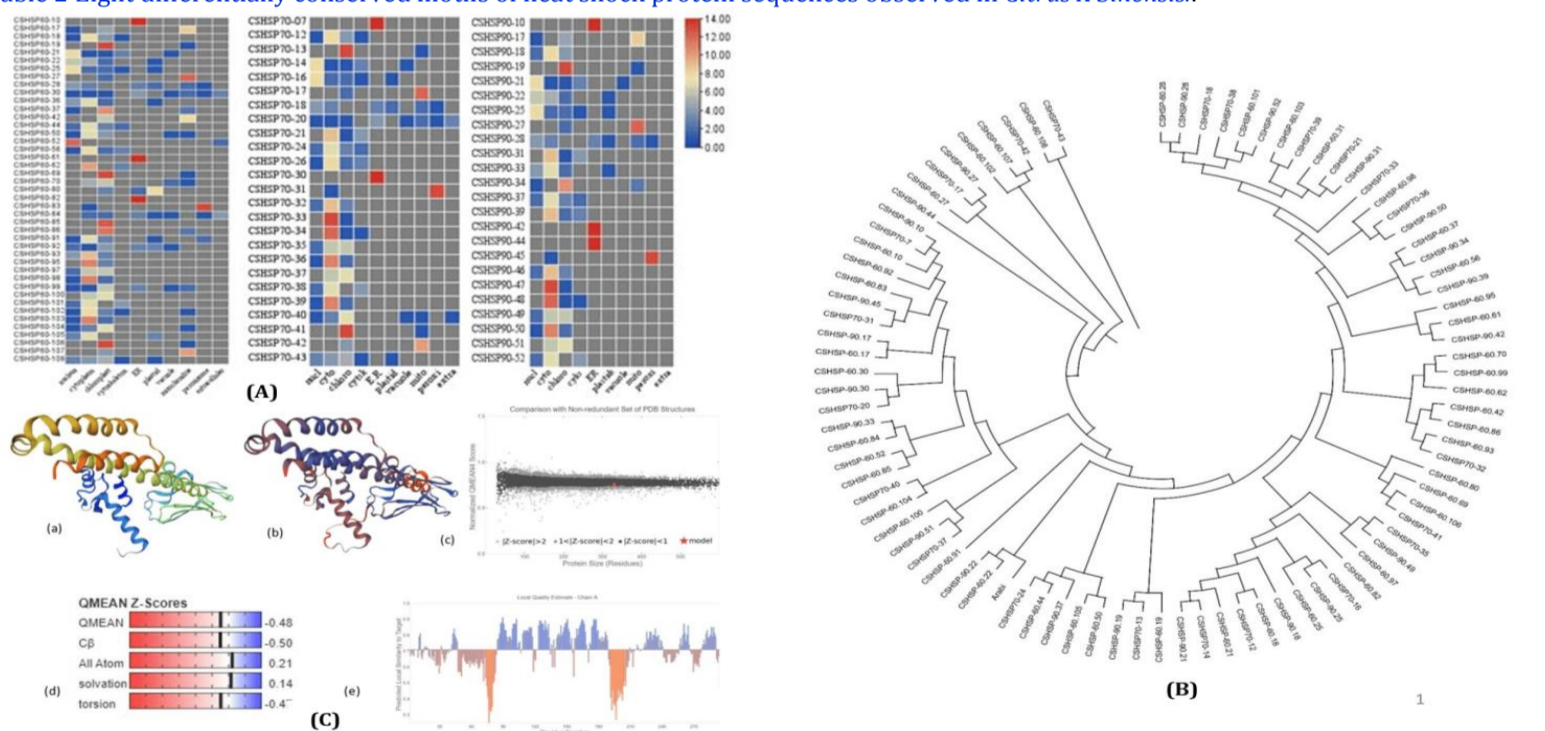


Figure 1: (A) Heat map of localization of HSP60, Hsp70 and HSP90. (B) Evolutionary relationships between HSP60, HSP70 and HSP90 families (C) (a) Template structure (b) 3D-structure of CSHSP60-93 (c) Quality Comparison (d) QMFAN Z-Scores (e) Local Quality Estimate.

stimulated by the overexpression of this gene. HSP60 is involved in the folding and assembly of proteins. Both HSP70 and HSP60 have similar characteristics (Arsène *et al.*, 2000). The two proteins work together when folding and assembling substrate proteins. Sometimes, HSPs collaborate with co-chaperones for proper and efficient functioning (Wu and Tanguay, 2006) such as HSP70 and HSP90 (molecular chaperones) assembled a multi-chaperone complex in the mammalian cell and are linked by the Hop co-chaperone, engaged in the folding and development of the main regulatory proteins (Wegele *et al.*, 2004). In this study, after removing the redundant sequences we inspected a total of 44 HSP60, 22 HSP70, and 23 HSP90 sequences of *Citrus X Sinensis*.

A crucial role in examining the stability and operation of proteins in living systems was revealed by the behavior, characterization,

and physicochemical characteristics of proteins. According to numerous researches, HSP70 is cytoprotective under conditions of thermal stress (Belhadji Slimen *et al.*, 2016). Physical and chemical properties of HSP70 protein in sorghum were analyzed by gene size, protein length and molecular weight (kDa) of the genes (Amare and Kebede, 2023). AI is a very important factor to check the thermo stability of globular proteins. An AI value greater than 65 manifested the thermo stability of protein. Higher AI means, higher protein thermo stability. Our findings revealed that all the members of respective HSP families are highly thermo stable. The isoelectric point (pI) described the acidic and basic nature of the HSP. Mostly HSPs were acidic and same results were seen with sorghum (Amare and Kebede, 2023). Mostly HSPs exhibited hydrophilic proteins; solubility of protein depends upon the

GRAVY value. Similar results have been reported for solubility of protein and localization (Tripathy *et al.*, 2021; Amare and Kebede, 2023). Molecular chaperone has recognized as poly peptide protein stabilizer during their transport into subcellular organelle mitochondria to cytoplasm of the cell. Therefore, the subcellular localization of various forms of HSPs was showed the major role in housekeeping of protein functions (Young *et al.*, 2004; Craig, 2018). Computational tool was used to find the localization of HSPs in cell organelles due to insufficient experimental data. Cytosolic, mitochondria and chloroplast HSPs each have distinct functions (Waters and Rioflorido, 2007). The current studies showed that the localization of the HSPs more in endoplasmic reticulum. It has been reported that the HSP 70 show different gene expression in different tissues (Tripathy *et al.*, 2021).

Protein sequence analysis is useful for describe the evolutionary trends of critical biological functions. The evolution of sHSPs is quite different from other HSPs. All eukaryotic organisms have numerous HSP 70s. Various studies have provided comprehensive knowledge on structure; properties and dynamic and well defined protein mechanism on plant HSPs Structural and functional understandings of HSPs were illustrated by homology modeling. Conserved motifs of heat shock proteins were predicted. The phylogenetic relationship was studied by constructing a phylogenetic tree of the collected sequences of HSP60, HSP70 and HSP90 using the Maximum Likelihood method. HSP60 is more closely resembles HSP70, which also shown in our phylogenetic analysis. Gene families have continuous gene duplication (birth) and deletion (death) process and the evolutionary history of HSPs in plants shows that stiff selection is the main cause of the divers' number and form of HSPs (Waters and Rioflorido, 2007).

3D protein structure configurations were also predicted for each HSP family having >80% sequence similarity. The template was selected with the sequence similarity $\geq 20\%$ (Dokholyan, 2012). Ramachandran plot and QMEAN has been used to analyzed the quality of the predicted model (Ashwinder *et al.*, 2016). Phylogenetic, subcellular localization, physicochemical properties and structural organization exhibited by different analysis (Amare and Kebede, 2023). Similar parameters were used to analyze the structure and function of HSPs in orange plant.

CONCLUSION: HSPs are the most important cell regulatory proteins present in all cells and every form of life. These specific types of protein families are prompted in the cell under different stress conditions involved in protein synthesis and other metabolic activities. To understand the importance of HSPs in *citrus*, a wide genome analysis was performed. This study revealed basic information regarding the nature and expression of proteins. The identified protein sequences are hydrophilic, acidic and thermo stable. Moreover, these proteins are majorly localized at the endoplasmic reticulum in the *Citrus sinuses*. Phylogenetic analysis studies indicated the evolutionary relationship which describes that these protein families are closely related to each other. Swiss modeling structural findings indicated >80% similarity and the putative protein's structure develops a monomer state.

CONFLICT OF INTEREST: Authors have no conflict of interest.

REFERENCES: Abbas, M., Y. Li, R. G. Elbaiomy, K. Yan, A. J. Ragauskas, V. Yadav, S. A. Soaud, M. M. Islam, N. Saleem and Z. Noor, 2022. Genome-wide analysis and expression profiling of slhsp70 gene family in solanum lycopersicum revealed higher expression of slhsp70-11 in roots under cd2+ stress. *Frontiers in bioscience-landmark*, 27(6): 186.

Amare, K. and M. J. J. A. H. R. Kebede, 2023. In silico analysis of stress resistance heat shock protein 70 in sorghum (*sorghum bicolor* L.). 6(1): 232-243.

Arsène, F., T. Tomoyasu and B. Bukau, 2000. The heat shock response of *escherichia coli*. *International journal of food microbiology*, 55(1-3): 3-9.

Ashwinder, K., M. T. Kho, P. M. Chee, W. Z. Lim, I. K. Yap, S. B. Choi and W. K. Yam, 2016. Targeting heat shock proteins 60 and 70 of *toxoplasma gondii* as a potential drug target: In silico approach. 8: 374-387.

Bailey, T. L., N. Williams, C. Misleh and W. W. Li, 2006. Meme: Discovering and analyzing DNA and protein sequence motifs. *Nucleic acids research*, 34(suppl_2): W369-W373.

Bakthisaran, R., R. Tangirala and C. M. Rao, 2015. Small heat shock proteins: Role in cellular functions and pathology. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1854(4): 291-319.

Barkley, N. A., M. L. Roose, R. R. Krueger and C. T. Federici, 2006. Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRS). *Theoretical and applied genetics*, 112: 1519-1531.

Belhadji Slimen, I., T. Najjar, A. Ghram, M. Abdrrabba and a. nutrition, 2016. Heat stress effects on livestock: Molecular, cellular and metabolic aspects, a review. 100(3): 401-412.

Craig, E. A. J. B. b., 2018. Hsp70 at the membrane: Driving protein translocation. 16(1): 1-11.

Davoudi, M., J. Chen and Q. Lou, 2022. Genome-wide identification and expression analysis of heat shock protein 70 (hsp70) gene family in pumpkin (*cucurbita moschata*) rootstock under drought stress suggested the potential role of these chaperones in stress tolerance. *International journal of molecular sciences*, 23(3): 1918.

Dokholyan, N. V., 2012. Computational modeling of biological systems: From molecules to pathways. *springer science & business media*.

Efeoglu, B., 2009. Heat shock proteins and heat shock response in plants. *Gazi university journal of science*, 22(2): 67-75.

Flamini, G., P. L. Cioni and I. Morelli, 2003. Use of solid-phase micro-extraction as a sampling technique in the determination of volatiles emitted by flowers, isolated flower parts and pollen. *Journal of chromatography*, 998(1-2): 229-233.

Haslbeck, M., T. Franzmann, D. Weinfurter and J. Buchner, 2005. Some like it hot: The structure and function of small heat-shock proteins. *Nature structural & molecular biology*, 12(10): 842-846.

Horton, P., K.-J. Park, T. Obayashi, N. Fujita, H. Harada, C. Adams-Collier and K. Nakai, 2007. Wolf psort: Protein localization predictor. *Nucleic acids research*, 35(suppl_2): W585-W587.

Kabani, M., J.-M. Beckerich and J. L. Brodsky, 2002. Nucleotide exchange factor for the yeast hsp70 molecular chaperone ssa1p. *Molecular and cellular biology*, 22(13): 4677-4689.

Kennedy, D., R. Jäger, D. D. Mosser and A. Samali, 2014. Regulation of apoptosis by heat shock proteins. *IUBMB life*, 66(5): 327-338.

Kregel, K. C., 2002. Invited review: Heat shock proteins: Modifying factors in physiological stress responses and acquired thermotolerance. *Journal of applied physiology*, 92(5): 2177-2186.

Kumar, S., G. Stecher and K. Tamura, 2016. Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, 33(7): 1870-1874.

Ledesma, N., S. Kawabata and N. Sugiyama, 2004. Effect of high temperature on protein expression in strawberry plants. *Biologia plantarum*, 48: 73-79.

Lusiana, E. and I. Irfannuddin, 2022. The biomolecular characteristics of angiotensin ii type 1 receptors as parameters in kidney fibrosis: A bioinformatics analysis. *Majalah kedokteran sriwijaya*, 54(2): 72-84.

Mistry, J. and R. Finn, 2007. Pfam: A domain-centric method for analyzing proteins and proteomes. *Comparative Genomics*: 43-58.

Moore, G. A., 2001. Oranges and lemons: Clues to the taxonomy of citrus from molecular markers. *Trends in genetics*, 17(9): 536-540.

Pembroke, J. T., 2000. Bio-molecular modelling utilising rasmol and pdb resources: A tutorial with hew lysozyme. *Biochemistry and Molecular biology education*, 28(6): 297-300.

Rehman, S. u., A. Nadeem, M. Javed, F.-u. Hassan, X. Luo, R. B. Khalid and Q. Liu, 2020. Genomic identification, evolution and sequence analysis of the heat-shock protein gene family in buffalo. *Genes*, 11(11): 1388.

Tripathy, K., M. Sodhi, R. Kataria, M. Chopra and M. J. B. G. Mukesh, 2021. In silico analysis of hsp70 gene family in bovine genome. 59: 134-158.

Tutar, L. and Y. Tutar, 2010. Heat shock proteins; an overview. *Current Pharmaceutical Biotechnology*, 11(2): 216-222.

Tutar, Y., 2006. Heat shock proteins, substrate specificity and modulation of function. *Protein and peptide letters*, 13(7): 699-705.

Tutar, Y., Y. Song and D. C. Masison, 2006. Primate chaperones hsc70 (constitutive) and hsp70 (induced) differ functionally in supporting growth and prion propagation in *saccharomyces cerevisiae*. *Genetics*, 172(2): 851-861.

Van Noort, J. M., M. Bsibsi, P. Nacken, W. H. Gerritsen and S. Amor, 2012. The link between small heat shock proteins and the immune system. *The international journal of biochemistry & cell biology*, 44(10): 1670-1679.

Wang, W., B. Vinocur, O. Shoseyov and A. Altman, 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in plant science*, 9(5): 244-252.

- Waters, E. R. and I. E. Rioflorido, 2007. Evolutionary analysis of the small heat shock proteins in five complete algal genomes. 65: 162-174.
- Wegele, H., L. Müller and J. Buchner, 2004. Hsp70 and hsp90—a relay team for protein folding. *Reviews of physiology, biochemistry and pharmacology*: 1-44.
- Weibezahn, J., C. Schlieker, P. Tessarz, A. Mogk and B. Bukau, 2005. Novel insights into the mechanism of chaperone-assisted protein disaggregation.
- Wu, T. and R. M. Tanguay, 2006. Antibodies against heat shock proteins in environmental stresses and diseases: Friend or foe? *Cell stress & chaperones*, 11(1): 1.
- Xu, Q., L.-L. Chen, X. Ruan, D. Chen, A. Zhu, C. Chen, D. Bertrand, W.-B. Jiao, B.-H. Hao and M. P. Lyon, 2013. The draft genome of sweet orange (*Citrus sinensis*). *Nature genetics*, 45(1): 59-66.
- Young, J. C., V. R. Agashe, K. Siegers and F. Hartl, 2004. Pathways of chaperone-mediated protein folding in the cytosol. 5(10): 781-791.



Except where otherwise noted, this item's licence is described as © **The Author(s) 2023**. Open Access. This item is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the [Creative Commons license](https://creativecommons.org/licenses/by/4.0/), and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.