Bioinformatics analysis of heat-shock proteins in pistachio compared to six other plant species

Nahid Askari (PhD) [1*], Maryam Rashidifar (PhD candidate) [2], Seyede Nasim Tabatabaeipour (PhD candidate) [3], Amin Baghizadeh (PhD) [1]

[1.] *Department of Biotechnology, Institute of Sciences and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, Iran*

[2.] *Department of Plant Sciences and Biotechnology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran*

[3.] *Department of Plant Breeding and Biotechnology, Faculty of Agriculture, Shahrekord University, Shahrekord, Iran.*

Information

Article Type:
Original Article

Article History:

Received: 20.03.2022
Accepted: 25.04.2022

Doi: 10.22123/PHJ.2022.354077.1137

Keywords:

HSPs proteins
Jones-Taylor-Thornton
MEGA
MUSCLE
MEME

Corresponding Author:
Nahid Askari

Email: n.askari@kgut.ac.ir

Tel: +98-34-33776611

Abstract

Background: Heat-shock proteins (HSPs) are ATP-dependent chaperones conserved from bacteria to eukaryotes. In plants, they are necessary to control their normal growth, development, and immune responses to environmental stimuli. The present study aimed to comprehensively analyze the HSP gene family in pistachio and six other plants.

Materials and methods: An alignment sequence of full-length proteins of Arabidopsis, pistachio, orange, corn, date, pine, and pea was prepared using MUSCLE and saved in the ClustalW format. Their phylogenetic analysis was conducted using the MEGA 6.0 program, with 1000 bootstrap replications. An unrooted Maximum-Likelihood was established by the Jones-Taylor-Thornton (JTT) model and the same alignment file. Conserved motifs in putative protein sequences were recognized by the MEME program using default parameters. In order to predict the structure of proteins, the phyre2 web server was used.

Results: According to the sequence alignment, conserved amino acid sequences showed that pistachio and pea shared the greatest similarity. Besides, pistachio and orange showed the least identity. The phylogenetic tree revealed that within the same sub-families, the most closely related StHsp70 members assigned similar exon/intron structures and intron numbers. Additionally, 18 putative conserved motifs were identified.

Conclusion: The present survey indicated the functional characteristics of the HSP genes and provided basic information on pistachio and other plants with a high commercial and economic value for further studies on their biological functions.

►Please cite this article as follows:
Askari N, Rashidifar M, Tabatabaeipour SN, Baghizadeh A. Bioinformatics analysis of heat-shock proteins in pistachio compared to six other plant species. Pistachio and Health Journal. 2022; 5 (2): 4-13.

---

1. Introduction

Environmental factors, such as radiation, drought, soil salinity, sudden changes in temperature and the presence of temperature changes, heavy metals, poisons, and chemical compounds, as well as, biological factors, such as herbivores and pathogens, may affect plants. Therefore, plants in their surrounding environment interact with many factors and environmental stresses [1]. They cann't change their environment to avoid such stresses; thus,, so they use different ways of strategies and various morphological and physiological adaptations to tolerate these such conditions. On the other hand, the response to stresses at the molecular level is also seen observed in all living organisms, which leads to a sudden change in the expression of genes, which causinges an increase in the synthesis of specific protein groups, including HSPs (Heat- Shock Proteins or stress-induced proteins); these proteins are expressed in the cell almost under the influence of many stresses [2]. For example, in Arabidopsis and some other plant species, it has been seen that low temperature, osmotic stress, salinity, oxidative stress, drought, intense radiation, as well as heavy metals have led to the synthesis of this group of protein groups [3]. Some HSPs are also present in cells under natural situation and play an essential role in cellular homeostasis. They are responsible for protein folding, assembly, or aggregation. HSPs assist in the translocation and degradation of many normal cellular processes. The molecular mechanisms, the roles of HSPs as molecular chaperones, and their molecular interactions are still unclear [4].

However, for an appropriate agricultural yield, plants' ability to response respond to stress is crucial [5]. One of the factors influencing affecting stress response is HSPs.

HSPs are one of the most important molecular elements activated when plants are under stress [6, 7]. They are implicated in growth of the plant and how it responds to stresses such as salinity, pathogen infection, and insect invasion [6]. It is a highly conserved group of stress proteins with a molecular size of 10-150 kDa and is present in all living organisms from bacteria to humans. HSPs are known from their molecular weights [8]. They are grouped into five classes based on their molecular weight: HSP100, HSP90, HSP70, HSP60, and small heat shock proteins (SHSPs) [9]. These families participate in distinct non-redundant functions at various developmental stages and are involved in cellular homeostasis. As such, HSPs are important for many plant development processes and stressful situations [10].

HSP70s, one of the most researched HSPs in mammals, are discussed along with their mostly underestimated functions in plants. HSP70 is a major chaperone that maintains protein homeostasis. It is produced as a constitutive protein or in response to stress [6]. HSP70 is a highly conserved ATP-dependent chaperone that exhibits maximal sequence similarity over large evolutionary distances. HSP70 DnaK and HSP110/SSE are two subfamilies of the plant HSP70 protein superfamily. HSP110/SSE and HSP70 are composed of 14 and 4 members, respectively, encoded in the Arabidopsis genome. HSP70 is localized to four distinct functional groups in plant cytosol, chloroplast, mitochondria, and endoplasmic reticulum. The localization of HSP70 proteins is determined by their conserved N- and C-terminal sequences [6, 11]. Its structure consists of an ATPase and a substrate-binding domain that facilitates the conformational changes required for rapid protein recruitment and release [12]. According to previous transgenic plant studies, Hsp70 promotes hypersensitivity responses in *Nicotiana benthamiana* leaves, is involved in proper leaf and development of root in *Arabidopsis thaliana*, and have a role in tolerance of heat stress in *Arabidopsis thaliana* [13].

Interdomain communication between the N-terminal nucleotide-binding domain (NBD) and the C-terminal substrate-binding domain (SBD) is facilitated by a conserved linker segment of Hsp70s, which is a less conserved 25-kDa C-terminal substrate-binding domain and a highly conserved 40-kDa N-terminal nucleotide-binding domain [14]. To reveal detailed evolutionary information on numerous gene families, HSPs have been the subject of bioinformatics studies in various plant species. The *Arabidopsis thaliana*, *Oryza sativa*, and *Populus trichocarpa* HSP70 families, respectively, consist of 18, 26, and 20 members, respectively, compared to 7, 9, and 10 members of in the HSP90 family present [15].

On the other hand, the pistachio species have great economic importance in different countries, such as Iran. The pistachio genus Pistacia from the family Anacardiaceae includes 11 species worldwide; three species of pistachio, including Atlantica, Khinjuk, and Vera, are grown in different regions. Pistachio contain numerous proteins, fatty acids (oleic, linoleic, palmitic, stearic, and linolenic acid), vitamins (E, B3, B1), iron, phosphorus, zinc, potassium, and calcium. Thus, it is crucial to study the HSPs of this plant compared to other plants that grow in different situations to producing them in different environmental conditions and increasing their resilience to environmental stress conditions [16].

In recent years, protein structural analysis has provided a clear perspective on HSP formation, maintenance, and application to plant species. Protein analyzes need to be found to explore species with specific resistance structures that are important for molecular breeding programs. However, to our knowledge, there is limited information on the population structure of the *pistachio vera* collection that has been assessed using large-scale, comprehensive protein structures. It is crucial to study the impact of plant bioinformatics on the progress of economically valuable plants and the science of their biology. This research aims to investigate the genetic diversity at the protein level and the sequence of protein-encoding amino acids based on sequence alignment, amino acid sequence conservation, and phylogenetic analyses using bioinformatics analysis in seven plant species.

This study deals with bioinformatic knowledge and investigations in pistachio vera compared to other plants. This can be useful for plant researchers to use effective tools such as new sequencing data to increase plant productivity.

2. Materials and methods

The protein sequences of 7 plants (Arabidopsis, pistachio, orange, corn, date, pine, and pea) were retrieved from the NCBI database. Their full-length protein sequence alignment was prepared using MUSCLE and saved in the ClustalW format. Meanwhile, their phylogenetic analysis was conducted using the MEGA 6.0 program with 1000 bootstrap replications. Using the same alignment file, an unrooted Maximum-Likelihood was constructed by the Jones-Taylor-Thornton (JTT) model. Conserved motifs in the putative protein sequences were identified by the MEME program (HTTP: meme-suite.org/tools/meme) with the default parameters. In order to predict the structure of proteins, the phyre2 web server was used.

3. Result

3.1. Protein sequences alignment

Protein sequences of seven plants were collected from the NCBI database, ranging from 649 to 657 amino acids. The amino acid sequences were aligned by MUSCLE. According to sequence alignment and comparison of the amino acid sequences conservation among the candidate plants, pistachio and pea shared the greatest overall identity or similarity among other members (92%), while pistachio and orange showed the least identity (85%) (Fig.1).

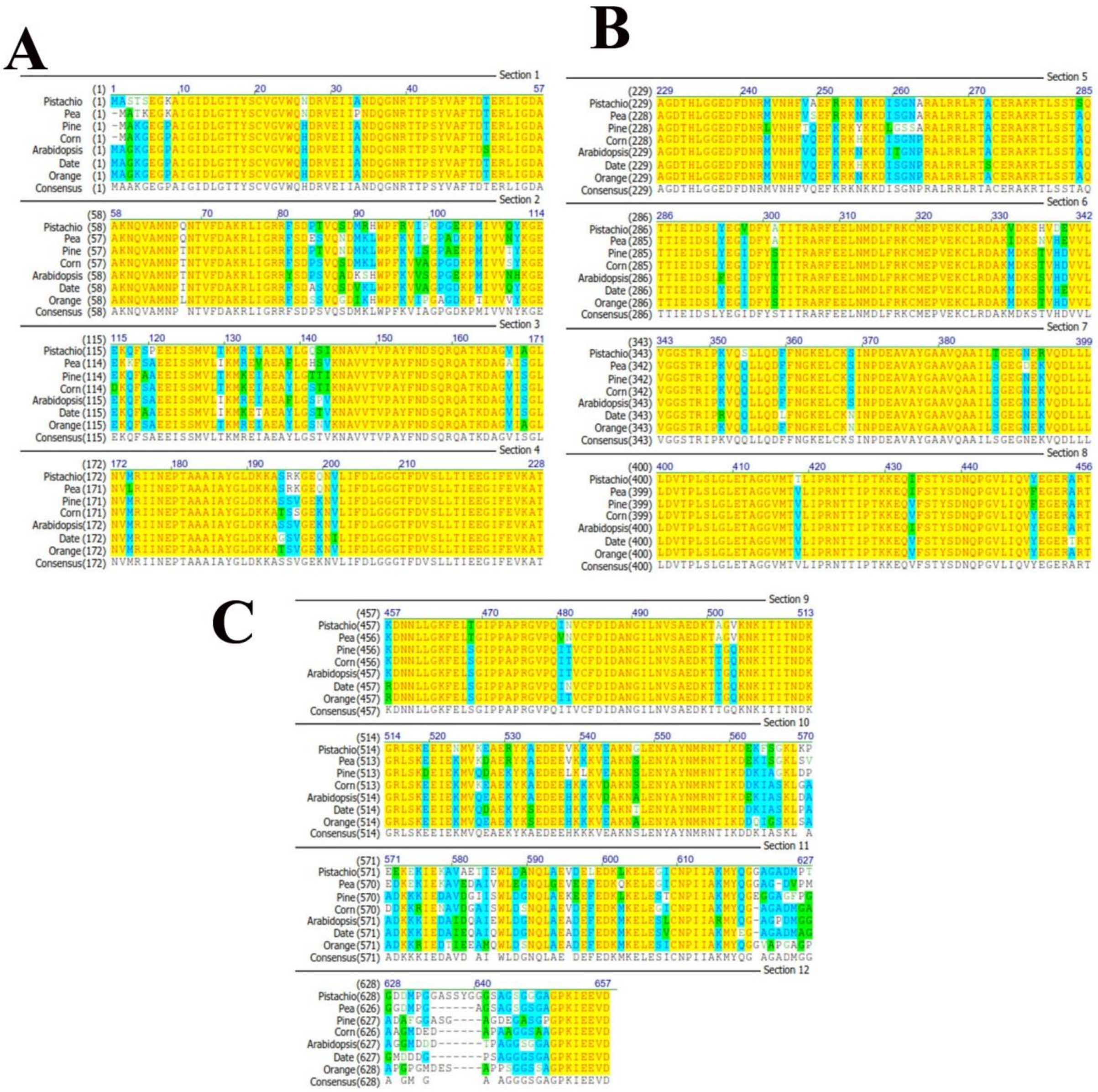

Fig. 1 Alignment of the predicted amino acid sequence of HSP70 from Arabidopsis, pistachio, orange, corn, date, pine, and pea; Additions and gaps in the sequences are indicated by dashes (–); Amino acids, which are conserved among canine babesial isolates, are marked by yellow.

Frequencies of amino acids were calculated based on all protein sequences. It was indicated that the most abundant amino acid in all plants (except pea) was Alanine (A). In contrast, Glutamic acid (E) was the most abundant amino acid in pistachio (59%), pea (58%), date (53%), and orange (54%). Alanine and Glutamic acid were the most abundant amino acids in date and orange. Alanine tends to form α-helix more than β-sheet in proteins because of it contains small side chain [17]. Glycine (G) and Leucine (L) were evolutionarily conserved amino acid residues.

The phylogenetic analysis revealed the phylogenetic relationships of HSP70 protein sequences among seven plants. In order to find the genetic distance, the HSP70 amino acid sequences were analyzed. A phylogenetic analysis was done using the full-length of HSP70 gene coding region and the entire sequence of amino acid of HSP70 protein; it indicated that all sequences were divided into 2 clusters. Cluster A was not far away from each, even with weak statistical support (Bootstrap value = 62) (Fig. 1A). It included pistachio and pea, with good statistical support (Bootstrap value = 100). Arabidopsis, date, and orange developed group B with good statistical support (Bootstrap value = 78) (Fig. 1B). Further, date and orange were closely related to each other with Bootstrap value = 68 based on the analysis of amino acid sequences (Fig. 1B). In accordance with the amino acids sequence in the polymer, proteins fold up into specific shapes, and their function is directly related to the resulting 3D structure. Based on the evolution distance of the phylogeny tree, the third structure of 4 plants was predicted in the Phyre2 database. These structures showed no significant structure for the amino acid sequences of the candidate HSP70, exhibiting a high degree of conservation to the HSP70. The analysis showed that all *Pistachio vera* was clearly separated from Arabidopsis, date, and orange. Two main clusters were detected: the first contained pea, pistachio, pine, and corn, whereas the second included Arabidopsis, date, and orange. Pea was the closest species to *P. vera*, while orange and date were the most distant (Fig. 2).

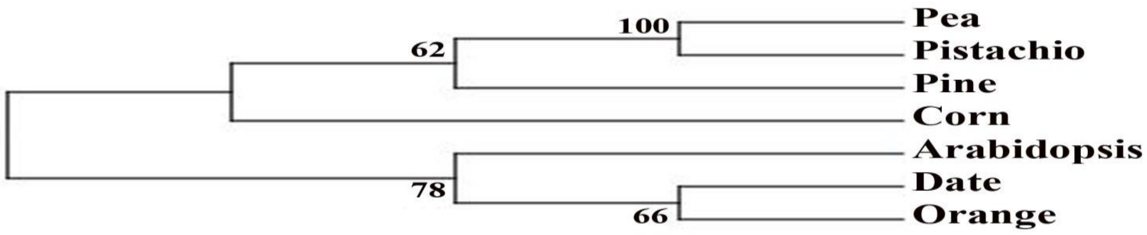

Fig. 2 Phylogenetic tree based on the heat-shock protein 70 (HSP70); The dendrogram is constructed with the Maximum-Likelihood method; Phylogenetic relationship based on the full coding region of HSP70 protein; The numbers at the nodes indicate bootstrap support from 1000 replications; The lengths of lines are proportional to the bootstrap value.

Based on the phylogenetic tree and gene, the closest family members of StHSP70 had similar exon/intron structures and the numbers of intron within the same sub-families. These findings were accordant with the characteristics defined in the phylogenetic analysis.

3.2. Protein structure

Eighteen putative conserved motifs were identified using the Meme Suite web-based version to study the protein's diversification (Table 1).

Table 1 Analysis of conserved motifs of SlHSP70 proteins in seven plants

Motif	Width	Sequences
1	50	TYSCVGVWQHDRVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVAMNP
2	50	CERAKRTLSSTAQTTIEIDSLYEGIDFYSTITRARFEELNMDLFRKCMEP
3	50	FSTYSDNQPGVLIQVYEGERARTKDNNLLGKFELSGIPPAPRGVPQITVC
4	50	LIFDLGGGTFDVSLLTIEEGIFEVKATAGDTHLGGEDFDNRMVNHFVQEF
5	50	VEKCLRDAKMDKSTVHDVVLVGGSTRIPKVQQLLQDFFNGKELCKSINPD
6	50	KNAVVTVPAYFNDSQRQATKDAGVISGLNVMRIINEPTAAAIAYGLDKKA
7	50	YGAAVQAAILSGEGNEKVQDLLLLDVTPLSLGLETAGGVMTVLIPRNTTI
8	50	TVFDAKRLIGRRFSDPTVQSDMRHWPFRVIPGPGEKPMIVVQYKGEEKQF
9	50	TITNDKGRLSKEEIEKMVQEAEKYKSEDEEHKKKVEAKNALENYAYNMRN
10	50	AADKKKIEDAIDQAIEWLDGNQLAEADEFEDKMKELESLCNPIIARMYQG
11	21	FDIDANGILNVSAEDKTTGQK
12	21	AEEISSMVLTKMREIAEAYLG
13	21	KRKHKKDISGNPRALRRLRTA
14	15	MAKGEGPAIGIDLGT
15	21	DDDGPSAGGGSGAGPKIEEVD
16	11	TIKDDKIASKL
17	6	PTKKEQ
18	8	SSVGEKNV

All HSP70 proteins showed common motif compositions and similar function among them. The length of the HSP70 motifs varied from 11 to 50. Motif 14 was presented in the N-terminal region, while motif 15 revealed in the C-terminal region. The HSP70s were made of two major domains: the N-terminal nucleotide-binding domain (NBD) (45 kDa) with a V-shaped structure consisting of two subdomains (lobes) that enclose the ATP-binding site, and the substrate-binding domain (SBD) (25 kDa) consisting of two subdomains of a Beta-sheet domain (SBDβ) or Base and an Alpha-helical subdomain (SBDα) or Lid domain. There was an allosteric relationship between the NBD and the SBD in the function of HSP70. The NBD drove conformational changes in the SBD, and substrate-induced conformational changes were transmitted from the SBD to the NBD interface [18, 19]. In this analysis, the third structure in candidate plants was similar to each other, and there were no significant differences among them (Fig. 3).

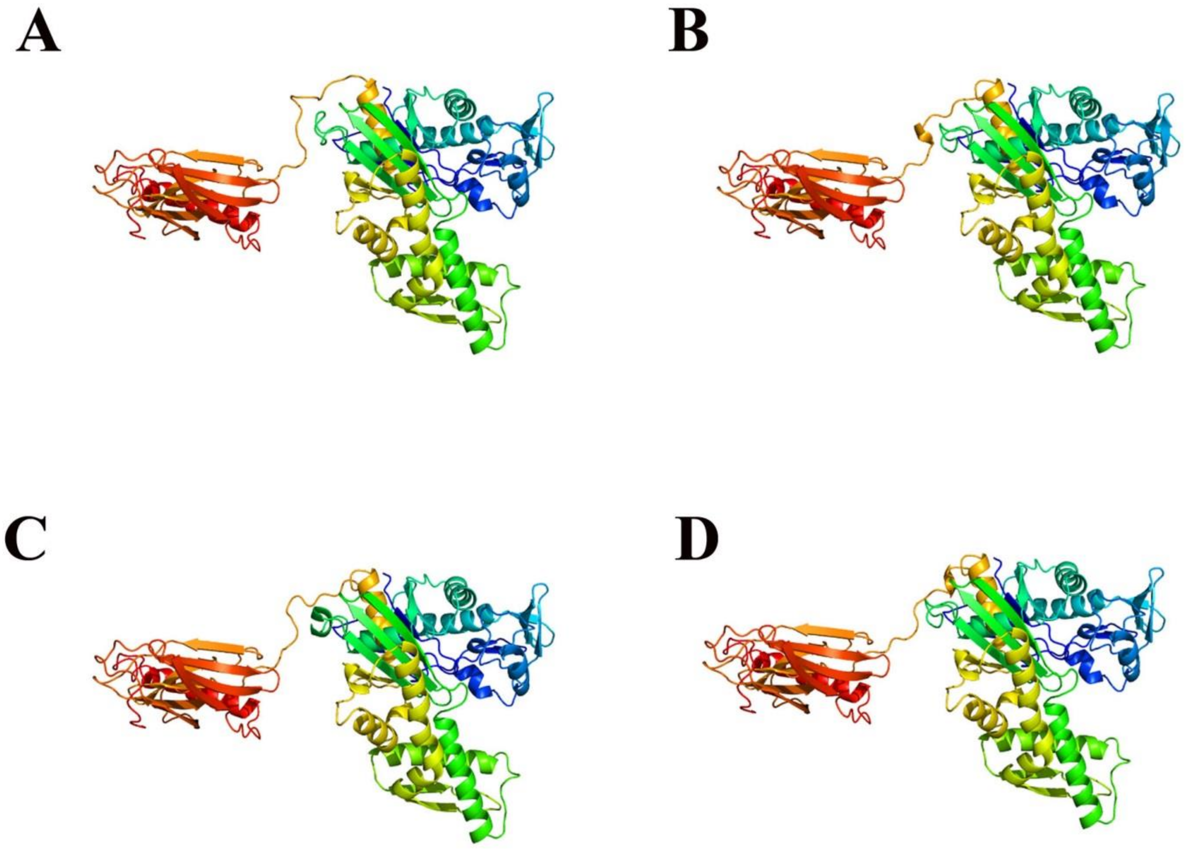

Fig. 3 Heat-shock protein 70 (HSP70)-domain organization and atomic structures of functional conformations; A: pistachio, B: Date, C: Pea, D: Orange

4. Discussion

In the past few years, bioinformatics analyzes have been conducted to find the regions of functional genes in tree species. Members of the gene families sHSP, HSP40, HSP60, HSP90, and HSP100 were found in the poplar genome. Bioinformatics studies include DNA/protein sequence identification, chromosomal localization, gene structure, genome duplication calculations, phylogenetic group determination, motif regions, genetic category identification, protein 3D structure modeling, miRNA Targeting gene determination and investigation of the members of HSP gene family in transcriptome data [20]. HSP20, HSP40, HSP60, HSP70, HSP90, and HSP100 are members of a conserved subfamily known as 'heat-inducible molecular chaperones'. Compared to non-chaperone proteins, they are 20-fold more likely to be induced by heat. Inactive cytosolic HSFA1 is hypophosphorylated at low temperatures and binds to the HSP70-HSP90 complex. According to the traditional model, when heat shock occurs, the HSP70-HSP90 complex is hijacked by excess of the level of cytosolic unfolded, misfolded or thermolabile proteins, leaving HSFA1 free to activate the HSR [21].

HSP70 family members have a critical role to save plants from environmental stresses, such as salinity, dryness, high or low temperature. The present study analyzed the HSP70 protein sequences of seven plants, including Arabidopsis, pistachio, orange, corn, date, pine, and pea. The estimated sequences of HSP70 amino acid were very similar to each other, with about 657 amino acids.

The phylogenetic tree of the *HSP70* from pistachio and other species of plant showed two subfamilies. In the phylogenic analysis based on protein sequences of HSP70 reported here, a closer relationship between pistachio and pea was strangled by the Maximum-Likelihood method. Pistachio, a member of the cashew family, showed similarities with peas from the Legumes family compared to other species. However, the orange from Rutaceae and the date from Arecaceae were closer in another clade in the phylogenetic tree, suggesting that the ortholog pair originates from a common ancestral gene that existed before the plant family diverged.

The analysis showed that all *Pistachio vera* was clearly separated from Arabidopsis, date, and orange. All seven species were grouped and separated well in the analysis of cluster and structure. The structure and dendrogram analyzes divided species into two main clusters: the first included all pea, pistachio, pine, and corn, whereas the second consisted of the Arabidopsis, date, and orange. The first cluster was divided into two subclusters: the first included pea, pistachio, and pine, while the second included corn, with corn being the most distant species. Similar results were obtained from other subclusters: the first subcluster contained Arabidopsis, while the second contained date and orange species.

Environmental stresses such as abiotic stresses increase the production of ineffective proteins. Many of the molecular chaperones are stress proteins, initially known as HSPs. Research shows that HSPs regulate the folding, aggregation, placement, and degradation of proteins in all animal and plant species [22]. Among different types of HSPs, HSP70 has a critical function in preventing aggregation and helping to refold non-native proteins under stressful conditions; in contrast, smHSPs alone cannot refold non-native proteins but form complexes with non-native proteins and other HSPs. Additionally, in some complex chaperones, Hsps help misfolded/unfold proteins to fold correctly [22]. Therefore, various chaperones classes work together in protection of cell and play a crucial role, and even they play a covering role in protecting of proteins against stress.

Research of genes and proteins related to heat shock proteins in closely related species will help to better examine the relationship of evolution of these proteins and the environmental stress that caused their evolution.

This research helps to a better understanding of the genetic polymorphisms in seven species that have undergone differential adaptations to heat stress.

5. Conclusions

Altogether, this study could help with future improvements and understanding how organisms respond to environmental stresses. This study serves as a model for future research and provides essential insights into the role of this essential gene family. Furthermore, the identified genes can be used for cloning researches in biotechnological studies to produce drought, heat and salinity tolerant plants, especially pistachios, which are economical and nutritious.

Acknowledgments

The authors are grateful to our colleagues for their contributions to the research, as well as the members of the Graduate University of Advanced Technology for their helps.

Conflicts of interest

All authors declare that they have no conflict of interest.

References

1- Jan R, Asaf S, Numan M, Kim KM. Plant Secondary Metabolite Biosynthesis and Transcriptional Regulation in Response to Biotic and Abiotic Stress Conditions. Agronomy. **2021**;11:*968*.

2- Taha EA, Ono K, Eguchi T. Roles of extracellular HSPs as biomarkers in immune surveillance and immune evasion. Int J Mol Sci. **2019;**20:4588.

3- Ghosh S, Yadav VK, Mukherjee V. Impact of environmental factors on photovoltaic performance and their mitigation strategies–A holistic review. Renew. Energy Focus. **2019**;28:*153-72*.

4- Chandler JO, Haas FB, Khan S, Bowden L, Ignatz M, Enfissi EM, et al. Rocket science: The effect of spaceflight on germination physiology, ageing, and transcriptome of Eruca sativa seeds. Life. **2020**;10:49.

5- Fahad S, Bajwa AA, Nazir U, Anjum SA, Farooq A, Zohaib A, et al. Crop production under drought and heat stress: plant responses and management options. Front. Plant Sci. **2017**:1147.

6- Berka M, Kopecká R, Berková V, Brzobohatý B, Černý M. Regulation of heat shock proteins 70 and their role in plant immunity. J Exp Bot. **2022**;73:*1894-909*.

7- Priya M, Dhanker OP, Siddique KH, HanumanthaRao B, Nair RM, Pandey S, Singh S, Varshney RK, Prasad PV, Nayyar H. Drought and heat stress-related proteins: an update about their functional relevance in imparting stress tolerance in agricultural crops. Theor Appl Genet. **2019**;132:*1607-38*.

8- Ahammed GJ, Li X, Liu A, Chen S. Brassinosteroids in plant tolerance to abiotic stress. J Plant Growth Regul. **2020**;39:*1451-64*.

9- Aghaie P, Tafreshi SA. Central role of 70-kDa heat shock protein in adaptation of plants to drought stress. Cell Stress Chaperones. **2020**;25:*1071-81*.

10- Sable A, Agarwal SK. Plant heat shock protein families: essential machinery for development and defense. J Biol Med Sci. **2018**;4:*51-64.*

11- Kampinga HH, Andreasson C, Barducci A, Cheetham ME, Cyr D, Emanuelsson C, et al. Function, evolution, and structure of J-domain proteins. Cell Stress Chaperones. **2019**;24:*7-15.*

12- Iyer K, Chand K, Mitra A, Trivedi J, Mitra D. Diversity in heat shock protein families: functional implications in virus infection with a comprehensive insight of their role in the HIV-1 life cycle. Cell Stress Chaperones. **2021**;26:*743-68.*

13- Rowarth NM, Dauphinee AN, Denbigh GL, Gunawardena AH. Hsp70 plays a role in programmed cell death during the remodelling of leaves of the lace plant (Aponogeton madagascariensis). J Exp Bot. **2020**;71:*907-18.*

14- Chen X, Lin S, Liu Q, Huang J, Zhang W, Lin J, Wang Y, Ke Y, He H. Expression and interaction of small heat shock proteins (sHsps) in rice in response to heat stress. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics. **2014**;1844(4):*818-28.*

15- Sable A, Agarwal SK. Plant heat shock protein families: essential machinery for development and defense. J Biol Med Sci. **2018**;4:*51-64.*

16- Singh D, Taspinar YS, Kursun R., Cinar I, Koklu M, Ozkan IA, Lee HN. Classification and Analysis of Pistachio Species with Pre-Trained Deep Learning Models. Electronics. **2022**;11:*981*.

17- Asakura T, Horiguchi K, Aoki A, Tasei Y, Naito A. Parallel β-Sheet Structure of Alanine Tetrapeptide in the Solid State As Studied by Solid-State NMR Spectroscopy. J. Phys. Chem. B. **2016**;120:*8932-41*.

18- Anglès F, Castanié-Cornet MP, Slama N, Dinclaux M, Cirinesi AM, Portais JC, Létisse F, Genevaux P. Multilevel interaction of the DnaK/DnaJ (HSP70/HSP40) stress-responsive chaperone machine with the central metabolism. Sci Rep. **2017**;7:*1-6.*

19- English CA, Sherman W, Meng W, Gierasch LM. The Hsp70 interdomain linker is a dynamic switch that enables allosteric communication between two structured domains. J Biol Chem. **2017**;292:*14765-74.*

20- Düzgün MB, Theofilatos K, Georgakilas AG, Pavlopoulou A. A Bioinformatic approach for the identification of molecular determinants of resistance/sensitivity to Cancer thermotherapy. Oxid. Med Cell Longev. **2019**;11:*17-19.*

21- Bourgine B, Guihur A. Heat shock signaling in land plants: From plasma membrane sensing to the transcription of small heat shock proteins. Front. Plant Sci. **2021**:1583.

22- Tichá T, Samakovli D, Kuchařová A, Vavrdová T, Šamaj J. Multifaceted roles of HEAT SHOCK PROTEIN 90 molecular chaperones in plant development. J Exp Bot. **2020**;71:*3966-85.*

23- Yu C, Leung SK, Zhang W, Lai LT, Chan YK, Wong MC, Benlekbir S, Cui Y, Jiang L, Lau WC. Structural basis of substrate recognition and thermal protection by a small heat shock protein. Nat Commun. **2021**;12:*1-11.*