ISSN: 2643-671X

Research Article

Journal of Agriculture and Horticulture Research

In Silico Analysis of Stress Resistance Heat Shock Protein 70 in Sorghum (Sorghum bicolor L.)

Kasahun Amare[1*] and Mulugeta Kebede[2]

[1]*Department of Plant Science, College of Agriculture and Natural science, Mekdela Amba University, Tulu'awlia, Ethiopia.*

[2]*Department of Applied Biology, School of Applied Natural Science, Adama Science and Technology University, Adama, Ethiopia.*

[*]Corresponding Author:
Kasahun Amare, Department of Plant Science, College of Agriculture and Natural science, Mekdela Amba University, Tulu'awlia, Ethiopia.

Submitted: 11 Jan 2023; **Accepted**: 01 Feb 2023: **Published**: 25 Feb 2023

Citation: Amare, K., Kebede, M. (2023). In Silico Analysis of Stress Resistance Heat Shock Protein 70 in Sorghum (Sorghum bicolor L.), *J Agri Horti Res*, 6(1), 232-243.

Abstract

Heat shock proteins70 (HSP70) play important roles in many biological processes. However, as is typical of sorghum bicolor, the systematic identification of the Hsp70 gene is very limited and the role of the Hsp70 gene in the evolution of sorghum bicolor has not been systematically described. In order to fill such an information gap, it was imperative to perform a insilico analysis of the HSP70 genes. The study was performed using bioinformatic methods to analyze the HSP70 gene family and identified 30 HSP70 genes from the sorghum bicolor genome sequence. A comprehensive analysis of these 30 identified genes performed analysis of gene structure, and physico-chemical properties, subcellular localization and promoter region analysis. The gene structure visualization analyzes revealed that 22 gene contains both 5 and 3 UTRS, one 5 and one 3 gene and 6 genes without UTR. The highest number of introns recorded was 12, and these genes have shown that this is not the case in any intron. In the promoter region analysis, ten protein motifs are identified and characterized and 2219 cis-acting elements are also identified. Among these, promoter-enhancer elements share the highest number (1411) and photoresponsive elements share the next value (335). Analysis of the physicochemical properties revealed that 23 families are acidic in nature, while the four families are basic and the others are somewhat neutral. The various analyses revealed their structural organization, subcellular localizations, physicochemical properties, cis-acting elements, evolutionary relationships, and under-stress situations in general.This research adds to the functional characterization of HSP70 and aids in the understanding of the processes of abiotic stress tolerance in Sorghum bicolor under a variety of stress settings.

Keywords: Gene structure, Cis-acting elements, promoter region, protein motif

Introduction

Sorghum (Sorghum bicolor L.) is a multipurpose food crop belonging to the Poaceae family, which are C4 carbon cycle plants with high photosynthetic efficiency and productivity. It ranks among the top 5 cereal crops in the world. It serves as a source of food, fodder, fodder and bioenergy [1]. It plays an important role in food security in sub-Saharan Africa, supporting approximately 500 million citizens. It is grown in semi-arid zones in drought prone and ancillary areas where other crops cannot grow consistently [2, 3].

The most critical abiotic stresses impeding the production of sorghum crops may include: inadequate supplement intake, aluminum stress, drought, high salinity, waterlogging, and temperature stress. These wonders must be adapted to the plants during development [4].

The effects of stress can lead to developmental deficiencies, crop yields, permanent damage or death if the stress exceeds the plant's capacity [5]. Plant endurance is limited to an expected thermal range of -10 to +60 °C, characterized by the freezing edge above intracellular water and the temperature of protein denaturation [6]. Such abiotic stressors have been found to cause the development of various intracellular substances including nucleic acids, amino acids, starches and proteins. After the launch of molecular biology strategies in plant science, an enormous effort was put into the disclosure of stress-inducible genes [7].

In this sense, it is currently believed that an understanding of the response to abiotic stress is perhaps the most important point in plant science. The main advance in this field of exploration is due to the use of molecular biology techniques. After this method was applied in plant science, various abiotic stress-inducible genes were isolated and their function in transgenic plants was correctly characterized. The accessibility of this information has broadened

and broadened our perspective on abiotic stress response and resilience in plants [7]. In this way, plants also respond to abiotic stresses like heat as various fears that can trigger the plant gene to protect the foreign conditions through gene articulation, which were not expressed under normal conditions. Such a response to weights at the molecular level is also found in enveloping living frameworks, microorganisms, plants and animals [8, 9].

A group of genes that are excited during heat stress and expressed proteins are termed heat shock proteins (HSPs), stress-prompt proteins or stress proteins [10]. It is known that HSPs are expressed in plants when they experience high temperature stress and in the face of a wide range of other environmental stresses, for example water stress, salt stress, cold stress and oxidative stress (Wasser, 2013) [11].

These stress-responsive proteins are classified into different groups depending on their function and expression pattern, as constitutive heat stun proteins, which are constitutively expressed, and inducible structures, which are expressed given specific components (Boone and Vijayan, 2002). Also, it has been grouped depending on their protein molecular weight, where they are divided into HSP90 (83110 kDa), HSP70 (6678 kDa), HSP60 (5865 kDa) and other small molecular weight proteins HSP20s [12]. HSP70 is described by two useful domains, the amino N-terminal ATPase domain (44 kDa), which displays ATPase activity, and a carboxyl C-terminal peptide-restricting domain (25 kDa). The peptide-restricting domain is additionally divided into a -sandwich subdomain (18 kDa), which is the substrate-restricting domain, and a -helical subdomain (Zhu et al., 1996) [13]. The 70 kDa heat shock proteins (HSP70s) are the richest and generally considered to be the conserved group of proteins. sHSPs comprise low molecular weight proteins that act as molecular chaperones, fundamental for protein collapse and avoiding irreversible protein aggregation (Wasser, 2013). Notwithstanding that sHSPs are unmistakably expressed during the heat stunning response in plants, it is currently recognized that some are also expressed in unstressed cells and are therefore involved in actions other than heat stress [11, 12].

The Problem Statement

Of the HSPs prominently expressed during the heat shock response in plants, it is now known that some are also expressed in non-stressed cells and are therefore involved in processes other than heat stress [11-12]. Several studies have shown that HSP70 is closely associated with plant abiotic stress, disease resistance, growth and development [14-16]. Heat tolerance in different plant species such as Triticum aestivum, Oryza sativa and Capsicum annuum [17-19]. The low molecular weight HSPs act as molecular chaperones that are critical for protein folding and the prevention of irreversible protein aggregation (Wasser, 2013). When the plant suffers from high temperatures, drought, high salinity, low temperatures and heavy metals, HSP70s accumulate rapidly to maintain the stability of the protein and biological macromolecules and improve the plant's resistance [20]. In addition, some studies found that the HSP protein is somewhat related to plant embryogenesis. They are also upregulated during tomato fruit ripening initiation and may also protect ripe tomato fruit from chill damage [21-22]. The relationship between heat shock treatment and embryogenesis was also studied in Brassica napus, and it was found that HSP70 and HSP90 located in the nucleus and cytoplasm are rapidly induced [23].

Despite the breakthrough in the identification and characterization of plant HSPs, in sorghum, the most widely grown and most stress-tolerant cereal crop in the world and particularly in sub-Saharan Africa the mechanism of stress tolerance through HSP70 is as well known studies published [2]. Thus, the identification and characterization of HSP70 genes in sorghum will contribute to a better understanding of the molecular mechanism of its stress tolerance. In sorghum, the existence of HSP70 was identified by immunoblotting after salt stress heat stress and Nagaraju et al., (2020) sHSP families, used as companions. However, the author could not find any literature describing the characterization of the HSP70 gene at the genome-wide level [24, 25, 13]. Due to the effects of various stressors on plants, studying different mechanisms of plant stress responses is of crucial importance. Therefore, a genome-wide analysis of the sorghum HSP70 genes will help unveil the underlying complex molecular mechanisms. The publication of sorghum genome data will enable the systematic analysis of HSP70 development and function.

In this study, the bioinformatics method was used to analyze sorghum genomic HSP70 gene family members, including the number of gene identifications, phylogenetic relationships, gene structural features (exon-intron organization), and subcellular localization of the HSP70 proteins. With general objective of stress resistance gene family analysis in sorghum.

Materials and Methods

Study Materials

The study materials in this research have consisted of the proteomic and genomic sequences of sorghum, maize, wheat, rice, sugarcane, millet, and Arabidopsis. Others like online and offline based software including, TBtool, MEGAX, Microsoft excel, and web base databases like NCBI, MEME, PlantCARE, Pfam, GSDS, Phytozome, ExPasy, and Cello life were used.

Database Mining (Hsp70 Family Genes)

The whole sorghum genome sequence was downloaded from the annotation database phytozome V3.1.1 (Cereal grass) database using a query sequence that was obtained from NCBI. To gather the probable candidates' sorghum HSP70 protein sequence, the HMM profile of the HSP70 domain was first checked in the Pfam database. The sorghum gene sequence information, including; the gene coordinate in the chromosome, genome sequence, full CDS sequence, protein sequence, and 2k bp of the nucleotide sequences upstream of the translation initiation codon was downloaded from the phytozome. The molecular weight (kDa) and iso-electric point (PI), the grand average hydropathy (GRAVY) value of each gene was calculated using compute PI/MW tool from EXPASY

proteome server (https://web.expasy.org/cgi-bin/protparam/protparam). Consequently, the genes were renamed as SbSHP70-1 to SbSHP70-30 for convenience different parameters were analyzed as depicted in Table-1.

Gene Structure Display for *Hsp70* Genes

The Gene Structure Display Server tool (http://gsds.gao-lab.org/index.php) was used to analyze the exon-intron structures [26]. Besides the exon and intron regions, the upstream and downstream UTR (un-translated) regions were also determined to show possible structures of entirely expressing mRNA. Intron phases were classified based on their positions relative to the reading frame of the translated proteins: phase 0 (located between two codons), phase 1 (splitting codons between the first and second nucleotides), or phase 2 (splitting codons between the second and third nucleotides).

Protein Motif and Cis Regulatory Elements Prediction of Hsp70 Genes

To find conserved motifs in sorghum Hsp70 gene family members, The MEME suite (version 5.3.3) (https://meme-suite.org/meme/tools/meme) were used to search for motifs in all HSP70 genes that was downloaded from phytozome [27].

To investigate the cis-acting elements, the upstream regions of all the HSP70 genes (2kbp) were extracted from the Phytozome website (https://phytozome.jgi.doe.gov/pz/portal.html). Subsequently, all of the sequences were submitted to the PlantCARE website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to identify possible cis-acting elements.

Gene Localization and Gene Duplication Analysis

The amino-acid sequences of tandemly and segmentally duplicated HSP70 genes were subjected to TBtool for the analysis of the synonymous (Ks) and non-synonymous (Ka) substitution rate determination. In light of a pace of 6.1×10^{-9} replacements for each site each year, the difference time (T) was determined as $T= Ks/(2 \times 6.1\ 10^{-9}) \times 10^{-6}$ million years prior (Mya). The gene location on chromosome were analized With the help of Phenogram - Ritchie Lab to annotated with lines in color at specific base-pair locations. PhenoGram allows for annotation of chromosomal locations and/or regions with shapes in different colors, gene identifiers, or other text.

Proteins Sub-Cellular Localization Prediction and Physico-Chemical Properties

Protein subcellular localization is crucial for genome annotation and protein function prediction. Therefore, the subcellular localization of proteins was analyzed at cello life. *For computing physicochemical features such as molecular mass, isoelectric point, instability index, aliphatic index, and average hydropathy were computed* at ProtParam expasy (https://web.expasy.org/protparam).

Result and Discussion

Identification of The Hsp70 Gene Family in Sorghum Bicolor

The systematic searching of the sorghum genome showed that a total of 39 genes were responsible for SbHSP70 protein production/ coding. However, among the identified genes eight genes were found as redundant versions of other genes and one gene was fragmented sequence and therefore was removed. After the removal of redundant sequences, domains in the proteins of these gene families were searched using the Pfam search tool for confirming the presence of specific Hsp70 domains. Under the activities, a total of nine genes were removed and 30 genes were identified. All non-redundant HSP70 genes were distributed on chromosomes 1, 2, 3, 4, 6, 8, 9, and 10 of sorghum. The largest numbers of genes (14) are distributed on chromosome one and four genes in chromosome nine (Table-2). It was understood from different kinds of literature previously various amount of HSP70 genes were identified in different plant species, for instance, 17 Hsp70 in Hordium vulgare 20 StHSP70 in *Solanum tuberosum* 21 CaHSP70 genes in *Capsicum annuum* 24 PvHSP70 *in Phaseolus vulgaris* (Buyu et al., 2016) and 61 HSP70 in Glycine max were also reported [28-31]. This variation in HSP70s families in plants may be due to the presence of extra organelles, like plastid, in the plant cell compared to other eukaryotic organisms.

Gene Structure Display for *Hsp70* Genes

The gene structure of 30 SbHSP70 gene families was established utilizing both genomic and coding sequences. During the investigation, it was discovered that 22 genes had both UTRs (5' and 3' end), two genes had only one UTR each (either 5' or 3'), and six genes had no UTR. Sobic.004G263500 gene had the most introns (12), followed by eight introns in Sobic.004G011700, Sobic.009G066900, Sobic.009G067000, and Sobic.010G230600 genes, respectively. Furthermore, in two SbHSP70s genes (Sobic.002G008000 and Sobic.003G378700), no introns were discovered (Figure-1 and Tabel-2).

Intronless SbHSP70 genes account for 6.66 percent of SbHSP70, while introned SbHSP70 genes account for 93.34 percent. The bulk of SbHSP70 proteins are found in the cytoplasm (17) and have 1-8 introns, although other SbHSP70 proteins found in other organelles have 0 or 12 introns (Table-2). Various plants with different intron counts produced similar conclusions. Some intronless genes were identified in the different TaHSP sub-families in Triticum aestivum Arabidopsis thaliana, Phaseolus vulgaris, and Popular HSP70 genes (Buyu et al., 2016; Yer et al., 2016) [32-34].

The structure of a gene determines its coding potential and can also give hints about the ancestry of genes since genes with similar structures probably evolved from a common ancestor [34]. The effect of introns on the transcription of genes is an evolutionarily conserved feature, being exhibited by such diverse organisms as yeast, plants, and mammals. Intron-containing genes are often transcribed more efficiently than non-intronic genes and therefore, the presence of introns in a gene is generally associated with an increase in protein production mediated through many different

mechanisms ranging from the increase of transcription and translation rates to the improvement of mRNA stability and folding [35].

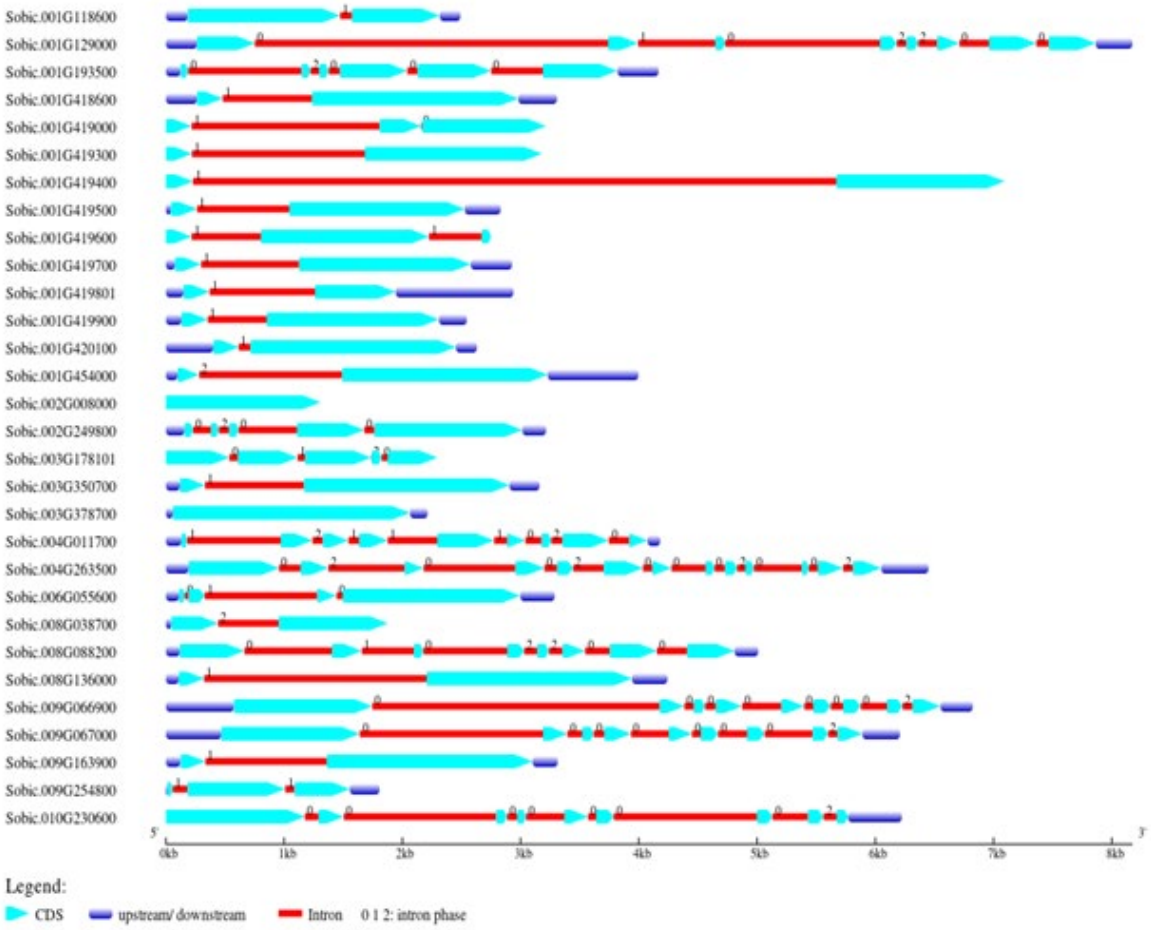

Figure 1: Exon/intron organization of 30 HSP70s genes in sorghum

Protein Motif and Cis Regulatory Element Prediction of HSP70 Genes

MEME was used to analyze the conserved motifs of different groups of protein sequences, and 10 conserved motifs were obtained (Figure-2), which were named Motif 1-10 (Table-1). In the promoter region analysis, upstream of the transcription start site of the HSP70 gene family contained a total of 2,219, different cis-acting elements which are categorized as growth hormone-related elements (286) light-responsive elements (335), promoter related elements (1,411), development/cell cycle-related elements (7), drought-related elements(20), metabolism-related elements(72) seed-specific regulation related elements(16) binding site related elements (18), temperature-related elements(19), stress defense-related elements(20) and anoxic related elements (15) were identified as computed in PLANT CARE enriched cis-acting element analysis database (Table-5). All the detail of cis-acting elements is presented on the Figure-2B.

Among the cis-acting elements discovered the promoter enhancing elements (1411) are the leading and followed by light-responsive elements (335). All SbHSP70 genes contained growth hormone-related elements, promoter enhancer-related elements and light-responsive related elements at least two and other cis-acting elements vary among those genes. Average number of cis-acting elements per SbHSP70 gene was 73.97% and the highest number (117) of cis-elements was found on SbHSP70-20 followed by SbHSP70-9 (103) SbHSP70-15 (102) SbHSP70-22 (101) while the least (48) was on SbHSP70-24 (Table-5). The current results are in line with those of the previous report of HSP70 genes in Glycine max PvHSP70 genes in *Phaseolus vulgaris* (Buyu et al., 2016) [31].

Table 1: Conserved motif result as predicted by meme

S/N	Motif sequence	Width	Description
1	VKBAVITVPAYFNBSQRQATKDAGTIAGLNVMRIINEPTAAAJAYGLDKK	50	Hsp70 protein
2	TACERAKRTLSSTAQTTIEIDSLYDGIDFSETITRARFEELNMDLFRKCM	50	Hsp70 protein
3	PAIGIDLGTTYSCVAVWRHDRVEVIANDQGNRTTPSYVAFT	41	Hsp70 protein
4	PVEKCLRDAKMDKSSIHDVVLVGGSTRIPKVQQLLQ	36	Hsp70 protein
5	EQVFSTYSDNQTGVLIQVYEGERARTKDNNLLGKFELSGIPP	42	Hsp70 protein
6	FEVKATAGDTHLGGEDFDNRLVDHFVREFKRKH	33	Hsp70 protein
7	LVGEAAKNQAALNPTNTIFDVKRLIGRRF	29	Hsp70 protein
8	FNGKELCKSINPDEAVAYGAAVQAAILSG	29	Hsp70 protein
9	GEEKQFSPEEISAMVLAKMKETAEAYLGT	29	Hsp70 protein
10	APRGVPQIEVTFDIDANGILNVSAEDKTTGQK	32	Hsp70 protein

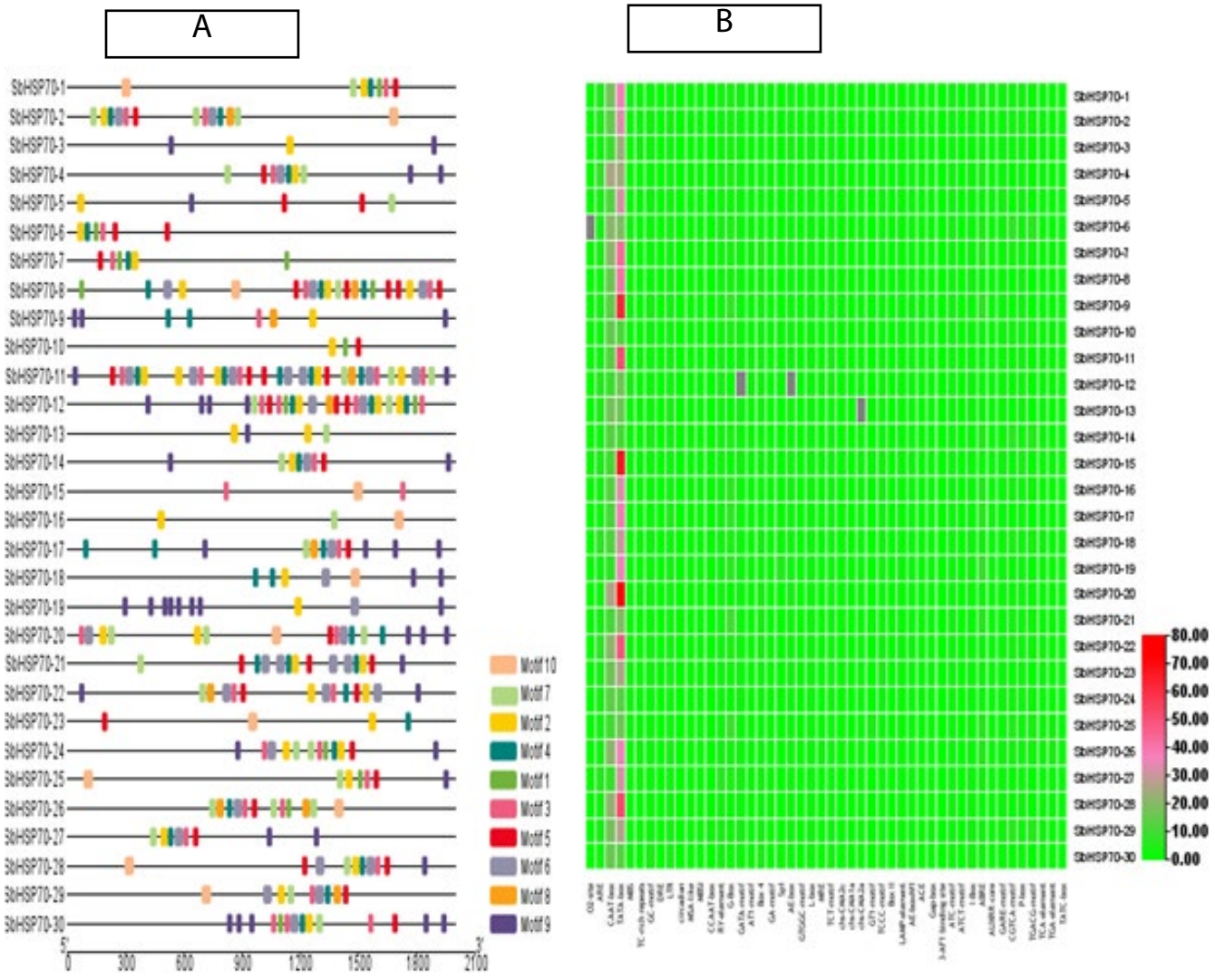

Figure 2: The Schematic representation of identified motifs and Cis-regulatory elements of HSP70 genes in sorghum

Physico-Chemical Properties Hsp70 Protein in Sorghum Bicolor

The analyzed physical and chemical properties of HSP70 protein in sorghum are presented in Table-2. The functional diversity of HSP70 isoforms could be realized from a wide scope of their MW (33.01 kDa to 103.79 kDa) and the total number of amino acids in various HSP70 proteins that went from 299 amino acid length (SbHSP70-11) to 928 amino acids length in (SbHSP70-21). The analysis of the current study uncovered that all the HSP70 protein families with record ID (SbHSP70-14, SbHSP70-17, SbHSP70-21, SbHSP70-26, SbHSP70-27, and SbHSP70-30) are unstable as determined by instability index result. The iso-electric point indicates that among the 30 HSP70 protein families recognized 23 of the families have acidic nature while the four families are basic and the rests are somewhat neutral. The aliphatic index for all HSP70 families is greater than 77.25 and therefore all of the identified HSP70 protein families have thermo-stability. Whereas, lower GRAVY values of HSP70 indicate its hydrophilic nature.

In terms of Physicochemical characteristics such as gene size, protein length and molecular weight (kDa) of the genes HSP70-1, HSP70-2, HSP70-3, HSP70-4, HSP70-13, HSP70-14, HSP70-16, HSP70-17, HSP70-18, HSP70-19, HSP70-20, HSP70-24, HSP70-25, and HSP70-28 were very similar to each other (Table-2). Similarly, HSP70-5, HSP70-6, HSP70-7, HSP70-8, HSP70-9, HSP70-10, HSP70-12, and HSP70-22 genes showed very similar physicochemical characteristics in terms of gene size, protein length, molecular weight (kDa), and PI (Table-2). Hence it could be theorized that the molecular variation inside the characterized genes will have a fundamental role in biochemical and physiological functions that gives competitor gene-based markers, which show a nearby relationship with the trait of interest.

Table 2: Physico-chemical properties and subcellular localization of HSP70 protein

Phytozome Identifier	Gene ID	Chromosome Location	Lp	CDS	A.A	MW(kDa)	PI	II	AI	GRAVY
Sobic.001G118600	SbHSP70-1	Chr001:9239221..9241712	Cyto	2031	676	74.48	5.38	29.71	84.97	-0.447
Sobic.001G129000	SbHSP70-2	Chr001:10143995..10152169	Chlo	2049	682	73.13	5.08	29.67	85.07	-0.333
Sobic.001G193500	SbHSP70-3	Chr001:17225825..17229992	Mito	2037	678	72.53	5.72	37.47	86.8	-0.285
Sobic.001G418600	SbHSP70-4	Chr001:69932394..69935701	Cyto	1959	652	71.46	5.19	34.84	81.55	-0.427
Sobic.001G419000	SbHSP70-5	Chr001:69967019..69970235	Cyto	1614	537	59.33	7.08	37.95	88.77	-0.25
Sobic.001G419300	SbHSP70-6	Chr001:70001749..70004928	Cyto	1716	571	62.19	6.46	34.39	87.27	-0.196
Sobic.001G419400	SbHSP70-7	Chr001:70008508..70015602	Chlo	1650	550	60.01	5.74	35.74	89.35	-0.169
Sobic.001G419500	SbHSP70-8	Chr001:70023901..70026732	Chlo	1701	566	61.78	8.28	33.98	92.28	-0.08
Sobic.001G419600	SbHSP70-9	Chr001:70036636..70039387	Cyto	1719	572	63.24	5.47	36.07	90.14	-0.178
Sobic.001G419700	SbHSP70-10	Chr001:70042031..70044958	Cyto	1671	556	61.69	5.98	37.18	87.14	-0.272
Sobic.001G419801	SbHSP70-11	Chr001:70045891..70048830	Cyto	900	299	33.01	6.19	38	85.42	-0.232
Sobic.001G419900	SbHSP70-12	Chr001:70053441..70055986	Cyto	1677	558	61.23	4.96	35.69	88.26	-0.259
Sobic.001G420100	SbHSP70-13	Chr001:70059899..70062529	Cyto	1950	649	71.12	5.08	34.37	81.17	-0.445
Sobic.001G454000	SbHSP70-14	Chr001:73039652..73043648	Mito	1920	639	69.73	8.55	50.5	94.05	-0.107
Sobic.002G008000	SbHSP70-15	Chr002:746212..747519	ER	1308	435	47.06	5.4	27.99	90.97	-0.037
Sobic.002G249800	SbHSP70-16	Chr002:63685205..63688420	Mito	2037	678	72.40	5.65	33.47	89.1	-0.247

Sobic.003G178101	SbHSP70-17	Chr003:45559241..45561537	Nucl	2091	696	74.65	9.81	52.41	66.28	-0.584
Sobic.003G350700	SbHSP70-18	Chr003:67021779..67024939	Cyto	1947	648	70.96	5.1	33.73	82.65	-0.397
Sobic.003G378700	SbHSP70-19	Chr003:69252194..69254407	ER	2004	667	73.29	5.01	31.59	85.37	-0.41
Sobic.004G011700	SbHSP70-20	Chr004:962795..966975	ER	2037	678	75.47	5.21	28.56	91.3	-0.37
Sobic.004G263500	SbHSP70-21	Chr004:60838256..60844705	Mito	2787	928	103.79	9.21	42.88	91.41	-0.379
Sobic.006G055600	SbHSP70-22	Chr006:39888060..39890944	Cyto	1851	616	67.50	5.08	34.73	81.88	-0.439
Sobic.008G038700	SbHSP70-23	Chr008:3664254..3666130	Cyto	1317	438	48.15	5.39	36.72	87.95	-0.171
Sobic.008G088200	SbHSP70-24	Chr008:18774401..18779411	Chlo	2109	702	74.46	5.12	29.8	86.55	-0.257
Sobic.008G136000	SbHSP70-25	Chr008:56470318..56474560	Cyto	1950	649	71.03	5.13	35.17	84.04	-0.4
Sobic.009G066900	SbHSP70-26	Chr009:7308207..7315031	Cyto	2556	851	93.94	5.19	43.14	77.25	-0.468
Sobic.009G067000	SbHSP70-27	Chr009:7341758..7347966	Cyto	2529	842	93.07	5.07	41.2	78.91	-0.427
Sobic.009G163900	SbHSP70-28	Chr009:52032825..52036139	Cyto	1950	649	71.00	5.12	34.06	81.19	-0.418
Sobic.009G254800	SbHSP70-29	Chr009:58879456..58881262	Cyto	1335	444	48.03	5.93	33.99	91.04	-0.078
Sobic.010G230600	SbHSP70-30	Chr010:57307086..57313310	Nucl	2256	751	82.83	5.6	47.13	83.02	-0.364

Table 3: Physico-chemical properties and subcellular localization of HSP70 protein (cont…)

Phytozome Identifier	Gene ID	Chromosome Location	Lp	Exon/Intron
Sobic.001G118600	SbHSP70-1	Chr001:9239221..9241712	Cyto	2:01
Sobic.001G129000	SbHSP70-2	Chr001:10143995..10152169	Chlo	8:07
Sobic.001G193500	SbHSP70-3	Chr001:17225825..17229992	Mito	6:05
Sobic.001G418600	SbHSP70-4	Chr001:69932394..69935701	Cyto	2:01
Sobic.001G419000	SbHSP70-5	Chr001:69967019..69970235	Cyto	3:02
Sobic.001G419300	SbHSP70-6	Chr001:70001749..70004928	Cyto	2:01
Sobic.001G419400	SbHSP70-7	Chr001:70008508..70015602	Chlo	2:01
Sobic.001G419500	SbHSP70-8	Chr001:70023901..70026732	Chlo	2:01
Sobic.001G419600	SbHSP70-9	Chr001:70036636..70039387	Cyto	3:02
Sobic.001G419700	SbHSP70-10	Chr001:70042031..70044958	Cyto	2:01
Sobic.001G419801	SbHSP70-11	Chr001:70045891..70048830	Cyto	2:01
Sobic.001G419900	SbHSP70-12	Chr001:70053441..70055986	Cyto	2:01
Sobic.001G420100	SbHSP70-13	Chr001:70059899..70062529	Cyto	2:01
Sobic.001G454000	SbHSP70-14	Chr001:73039652..73043648	Mito	2:01
Sobic.002G008000	SbHSP70-15	Chr002:746212..747519	ER	1:00
Sobic.002G249800	SbHSP70-16	Chr002:63685205..63688420	Mito	5:04
Sobic.003G178101	SbHSP70-17	Chr003:45559241..45561537	Nucl	5:04
Sobic.003G350700	SbHSP70-18	Chr003:67021779..67024939	Cyto	2:01

Sobic.003G378700	SbHSP70-19	Chr003:69252194..69254407	ER	1:00
Sobic.004G011700	SbHSP70-20	Chr004:962795..966975	ER	9:08
Sobic.004G263500	SbHSP70-21	Chr004:60838256..60844705	Mito	13:12
Sobic.006G055600	SbHSP70-22	Chr006:39888060..39890944	Cyto	4:03
Sobic.008G038700	SbHSP70-23	Chr008:3664254..3666130	Cyto	2:01
Sobic.008G088200	SbHSP70-24	Chr008:18774401..18779411	Chlo	8:07
Sobic.008G136000	SbHSP70-25	Chr008:56470318..56474560	Cyto	2:01
Sobic.009G066900	SbHSP70-26	Chr009:7308207..7315031	Cyto	9:08
Sobic.009G067000	SbHSP70-27	Chr009:7341758..7347966	Cyto	9:08
Sobic.009G163900	SbHSP70-28	Chr009:52032825..52036139	Cyto	2:01
Sobic.009G254800	SbHSP70-29	Chr009:58879456..58881262	Cyto	3:02
Sobic.010G230600	SbHSP70-30	Chr010:57307086..57313310	Nucl	9:08

Cyto (cytoplasmic) Chlo(Chloroplasmic), Nucl(Nuclear),Mito(mitochondrial), ER(Endoplasmic Reticulum) Lp (localization predicted), Chro(Chromosome number), CDS(coding sequence), MW (Molecular Weight in Kilo Daltons), A.A (number of amino acids), pI (Isoelectric point), AI (Aliphatic Index), II (Instability Index), and GRAVY (Grand Average of hydropathicity Index)].

Predicted Proteins Sub-Cellular Localization

The analysis undertaken by cello life for sub-cellular localization of HSP70s in sorghum uncovered that these proteins were dispersed into five areas, for example, cytoplasmic (Cyto), endoplasmic reticulum (ER), chloroplast (Chlo), nuclear (Nucl), and mitochondrial (Mito). Most extreme proteins were found in the cytoplasmic/cytosolic (17), followed by four in chloroplast and mitochondrial. The endoplasmic reticulum represents three and the nucleus offers two genes (Table-2). In the current investigation of Sorghum bicolor, an aggregate of 17 cytosolic HSP70s proteins were discovered, which is similarly higher than the previous report in Oryza sativa (11) and Arabidopsis thaliana five (5) and lower than as contrasted in Glycine max, that is, 34 [33, 18, 31].

Gene Localization and Gene Duplication Analysis

The distinguished 30 HSP70 genes were conveyed on eight chromosomes of Sorghum bicolor (Figure 3). A large portion of the HSP70 genes were available on chromosome one (14 genes) and chromosome nine (four genes), while every one of the leftover chromosomes had one or three genes.

Both pair and segmental duplication add to the creation of gene families during the course of advancement. In this way, potential duplication occasions of HSP70 genes were analized insilico. Moreover, none of the genes were recommended to be results of segmental duplication. In light of the outcomes, it could be inferred that pair duplication assumed a significant part in the extension of the HSP70 family in Sorghum bicolor.

To examine the molecular transformative paces of copied genes combines, the the non-synonymous substitution (Ka) and synonymous substitution (Ks) proportion were determined utilizing Ka/Ks calculator in TBtool. In light of a pace of 6.1 x 10^{-9} replacements for each site each year, the difference time (T) was determined as T= Ks/(2 x 6.1 10^{-9}) x 10^{-6} million years prior (Mya) (Table 4).

Gene duplications assume a significant part in advancement as duplications cause genes to create gene families [36]. Truth be told, it has been recommended that couple and segmental duplications have been the essential driving wellspring of evlution as these occasions lead to extension of gene families, and age of proteins with novel capacities [37]. Tandem duplication includes the duplication of at least two genes situated on a similar chromosome, while segmental duplication alludes to the marvel when genes having a place with a similar clade however situated on various chromosomes are duplicated [38].

In the current study, an aggregate of eihgteen (18/30; 60 %) Sorghum bicolor HSP70 genes were demonstrated to be copied (Table 4). Further, seven sets of a gene had all the earmarks of being tandemly duplicated, which was perceived on chromosome number one (Figure 3). The remainder of copied genes were all segmentally copied/duplicated.

The proportion of Ka and Ks replacement rate is a powerful strategy to research the specific imperative among copied genes sets [39]. Henceforth, in the current review, Ka, Ks, and Ka/Ks esteems for each pair of the paralogous gene were determined (Table 4). On a fundamental level, the worth of Ka/Ks < 1 connotes the decontaminating choice (negative choice), Ka/Ks > 1 implies positive determination/selection, and Ka/Ks = 1 method impartial choice [40]. Here, 18 HSP70 genes were demonstrated to be copied. The Ka/Ks proportion for copied HSP70 genes went from 0.040056013 to 0.554428423. All the HSP70 genes in the current review have Ka/Ks esteem < 1(Table 4).

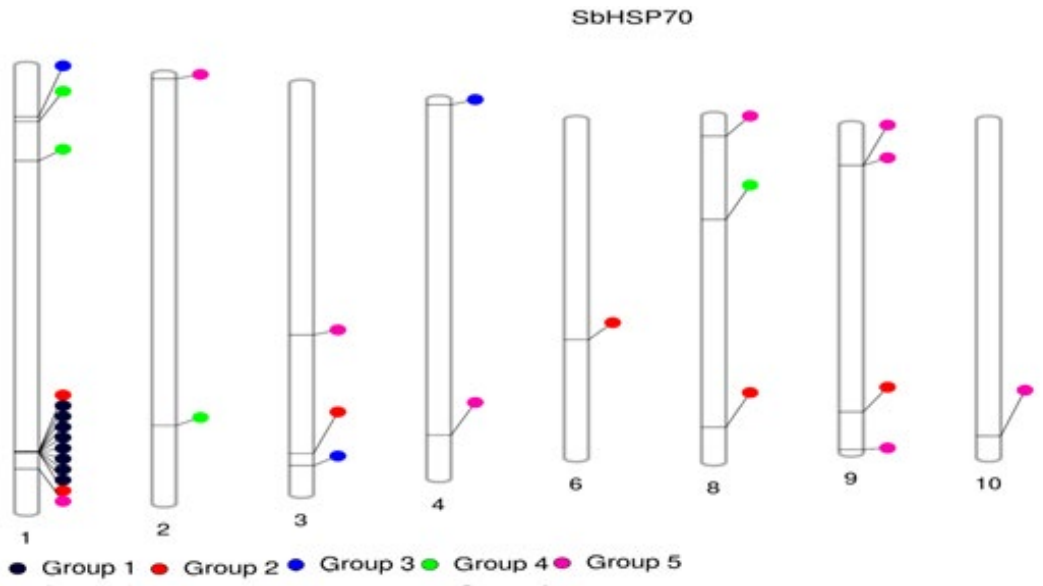

Figure 3: Distribution of HSP70 Genes on Sorghum bicolor Chromosome

Table 4: Synonymous and Non-Synonymous Substitution Rates

Paralogs Gene Pairs		Ka	Ks	Ka/Ks	Time(MYA)
Sobic.003G350700	Sobic.006G055600	0.014914382	0.372338146	0.040056013	0.0002
Sobic.001G420100	Sobic.008G136000	0.048386205	0.693205314	0.069800684	0.0003
Sobic.001G419500	Sobic.001G419600	0.105404002	0.468741704	0.22486585	0.0002
Sobic.001G419000	Sobic.001G419801	0.092001205	0.272097516	0.338118504	0.0001
Sobic.003G378700	Sobic.004G011700	0.202893273	1.076531965	0.188469344	0.0005
Sobic.001G129000	Sobic.008G088200	0.091415608	1.319369683	0.069287334	0.0007
Sobic.001G193500	Sobic.002G249800	0.15131001	2.809353502	0.05385937	0.0014
Sobic.002G008000	Sobic.008G038700	0.411522184	0.742245829	0.554428423	0.0004
Sobic.009G066900	Sobic.009G067000	0.047744901	0.488664962	0.097704777	0.0002

Table 5: Cis-acting elements analysis result as computed in PLANTCARE

Gene Name	GH R	PER	LRR	MBR	SDR	CCR	BSR	DRR	TR	SSR	ANR	Total	%
SbHSP70-1	10	54	12	3	1	0	0	2	0	0	0	82	7.45
SbHSP70-2	9	47	4	3	1	1	0	1	1	0	1	68	6.18
SbHSP70-3	8	32	11	1	3	0	0	0	0	0	1	56	5.09
SbHSP70-4	7	49	11	8	0	1	1	0	0	0	0	77	7.00
SbHSP70-5	7	44	4	3	1	0	0	0	0	0	1	60	5.45
SbHSP70-6	20	39	9	1	0	0	0	2	0	2	0	73	6.64
SbHSP70-7	12	63	15	4	0	0	1	2	0	0	0	97	8.82
SbHSP70-8	2	56	9	0	0	0	0	0	1	0	0	68	6.18
SbHSP70-9	8	75	14	2	0	0	0	1	1	0	2	103	9.36
SbHSP70-10	11	25	13	0	0	0	0	0	0	0	1	50	4.55
SbHSP70-11	5	63	9	1	2	0	0	0	1	1	0	82	7.45
SbHSP70-12	3	25	12	4	0	0	0	3	0	2	0	49	4.45
SbHSP70-13	4	33	7	3	1	0	0	0	2	0	0	50	4.55

SbHSP70-14	18	22	7	0	0	1	1	1	1	1	2	54	4.91
SbHSP70-15	6	76	12	5	1	0	1	1	0	0	0	102	9.27
SbHSP70-16	11	45	15	2	1	1	3	1	0	0	0	79	7.18
SbHSP70-17	8	48	13	0	0	1	1	0	3	1	0	75	6.82
SbHSP70-18	14	42	13	6	1	0	2	0	0	0	1	79	7.18
SbHSP70-19	10	42	13	1	0	0	1	0	1	0	0	68	6.18
SbHSP70-20	7	98	7	1	2	0	0	1	0	0	1	117	10.64
SbHSP70-21	13	30	19	3	0	1	0	0	0	0	0	66	6.00
SbHSP70-22	9	69	13	6	0	0	1	1	1	0	1	101	9.18
SbHSP70-23	3	42	9	2	0	0	1	1	1	0	1	60	5.45
SbHSP70-24	8	26	10	2	1	0	0	1	0	0	0	48	4.36
SbHSP70-25	20	26	18	1	1	0	1	1	0	0	0	68	6.18
SbHSP70-26	17	56	8	1	2	0	0	0	2	1	0	87	7.91
SbHSP70-27	6	39	11	5	2	0	0	0	0	4	0	67	6.09
SbHSP70-28	7	73	9	2	0	1	1	0	1	1	0	95	8.64
SbHSP70-29	17	44	11	2	0	0	3	0	1	2	0	80	7.27
SbHSP70-30	6	28	17	0	0	0	0	1	2	1	3	58	5.27
Total	286	1411	335	72	20	7	18	20	19	16	15	2219	73.97

GHR(Growth Hormone Related), PER(Promoter Enhancer Related), LRR (Low-Temperature Related), MBR (Metabolism Related), SDR(Stress Defense Related), CCR(Cell Cycle Related), BSR(Binding Site Related), DRR (Drought Response Related), TR (Temperature Related), SSR(Seed Specific Regulation), ANR (Anoxic Related)

Conclusions

With the increasing concerns about global warming and rising earth temperatures, it is very important to know about proteins that provide heat/stress tolerance in crop plants. The present study identified and characterized the HSP70 family in Sorghum bicolor genome. The Genome-wide assortment of Sorghum bicolor genome for the identification of SbHp70 revealed the presence of 30 genes. The different analyses performed disclosed their structural organization, subcellular localizations, physicochemical properties, cis-acting elements, phylogenetic, and under-stress conditions. This study provides further information for the functional characterization of HSP70 and helps to understand the mechanisms of abiotic stress tolerance under diverse stress conditions.

References

1. Ananda, G., Gleadow, R., Norton, S., Furtado, A., & Henry, R. (2019). Determination of Phylogenetic Relationships of the Genus Sorghum Using Nuclear and Chloroplast Genome Assembly. Multidisciplinary Digital Publishing Institute Proceedings, 36(1), 17.
2. Krishnamurthy, L., Serraj, R., Hash, C. T., Dakheel, A. J., & Reddy, B. V. (2007). Screening sorghum genotypes for salinity tolerant biomass production. Euphytica, 156, 15-24.
3. Amelework, B. A., Shimelis, H. A., Laing, M. D., Ayele, D. G., Tongoona, P., & Mengistu, F. (2016). Sorghum production systems and constraints, and coping strategies under drought-prone agro-ecologies of Ethiopia. South African Journal of Plant and Soil, 33(3), 207-217.
4. Tari, I., Laskay, G., Takács, Z., & Poór, P. (2013). Response of sorghum to abiotic stresses: A review. Journal of agronomy and crop science, 199(4), 264-274.
5. Mosa, K. A., Ismail, A., & Helmy, M. (2017). Introduction to plant stresses. In Plant stress tolerance (pp. 1-19). Springer, Cham.
6. Taiz, L., Zeiger, E., Møller, I. M., & Murphy, A. (2015). Plant physiology and development (No. Ed. 6). Sinauer Associates Incorporated.
7. Hirayama, T., & Shinozaki, K. (2010). Research on plant abiotic stress responses in the post-genome era: past, present and future. The plant journal, 61(6), 1041-1052.
8. Sarkar, N. K., Kundnani, P., & Grover, A. (2013). Functional analysis of Hsp70 superfamily proteins of rice (Oryza sativa). Cell stress and Chaperones, 18, 427-437.
9. Haag, J. (2019). Molecular and biochemical enhancement of chlorophyll in sports turf. Lulu. com.
10. Gupta, S. C., Sharma, A., Mishra, M., Mishra, R. K., & Chowdhuri, D. K. (2010). Heat shock proteins in toxicology: how close and how far?. Life sciences, 86(11-12), 377-384.
11. Tyedmers, J., Mogk, A., & Bukau, B. (2010). Cellular strategies for controlling protein aggregation. Nature reviews Molecular cell biology, 11(11), 777-788.
12. Waters, E. R. (2013). The evolution, function, structure, and expression of the plant sHSPs. Journal of experimental botany, 64(2), 391-403.
13. Mulaudzi-Masuku, T., Mutepe, R. D., Mukhoro, O. C., Faro, A., & Ndimba, B. (2015). Identification and characterization of a heat-inducible Hsp70 gene from S orghum bicolor which

confers tolerance to thermal stress. Cell Stress and Chaperones, 20, 793-804.
14. Vacchina, P., Norris-Mullins, B., Carlson, E. S., & Morales, M. A. (2016). A mitochondrial HSP70 (HSPA9B) is linked to miltefosine resistance and stress response in Leishmania donovani. Parasites & vectors, 9, 1-15.
15. Maimbo, M., Ohnishi, K., Hikichi, Y., Yoshioka, H., & Kiba, A. (2007). Induction of a small heat shock protein and its functional roles in Nicotiana plants in the defense response against Ralstonia solanacearum. Plant physiology, 145(4), 1588-1599.
16. Vega, V. L., Rodríguez-Silva, M., Frey, T., Gehrmann, M., Diaz, J. C., Steinem, C., ... & De Maio, A. (2008). Hsp70 translocates into the plasma membrane after stress and is released into the extracellular environment in a membrane-associated form that activates macrophages. The Journal of Immunology, 180(6), 4299-4307.
17. Duan, Y. H., Guo, J., Ding, K., Wang, S. J., Zhang, H., Dai, X. W., ... & Kang, Z. S. (2011). Characterization of a wheat HSP70 gene and its expression in response to stripe rust infection and abiotic stresses. Molecular biology reports, 38, 301-307.
18. Sarkar, N. K., Kundnani, P., & Grover, A. (2013). Functional analysis of Hsp70 superfamily proteins of rice (Oryza sativa). Cell stress and Chaperones, 18, 427-437.
19. Guo, M., Zhai, Y. F., Lu, J. P., Chai, L., Chai, W. G., Gong, Z. H., & Lu, M. H. (2014). Characterization of CaHsp70-1, a pepper heat-shock protein gene in response to heat stress and some regulation exogenous substances in Capsicum annuum L. International journal of molecular sciences, 15(11), 19741-19759.
20. Wang, W., Vinocur, B., Shoseyov, O., & Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends in plant science, 9(5), 244-252.
21. Shukla, V., Upadhyay, R. K., Tucker, M. L., Giovannoni, J. J., Rudrabhatla, S. V., & Mattoo, A. K. (2017). Transient regulation of three clustered tomato class-I small heat-shock chaperone genes by ethylene is mediated by SlMADS-RIN transcription factor. Scientific reports, 7(1), 6474.
22. Ré, M. D., Gonzalez, C., Escobar, M. R., Sossi, M. L., Valle, E. M., & Boggio, S. B. (2017). Small heat shock proteins and the postharvest chilling tolerance of tomato fruit. Physiologia plantarum, 159(2), 148-160.
23. Seguı-Simarro, J. M., Testillano, P. S., & Risueño, M. C. (2003). Hsp70 and Hsp90 change their expression and subcellular localization after microspore embryogenesis induction in Brassica napus L. Journal of structural biology, 142(3), 379-391.
24. Ndimba, B. K., Thomas, L. A., & Ngara, R. (2010). Sorghum 2-dimensional proteome profiles and analysis of Hsp70 expression under salinity stress. Agriculture and Natural Resources, 44(5), 768-775.
25. Ngara, R., Ndimba, R., Borch-Jensen, J., Jensen, O. N., & Ndimba, B. (2012). Identification and profiling of salinity stress-responsive proteins in Sorghum bicolor seedlings. Journal of Proteomics, 75(13), 4139-4150.
26. Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., & Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics, 31(8), 1296-1297.
27. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., ... & Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic acids research, 37(suppl_2), W202-W208.
28. Chaudhary, R., Baranwal, V. K., Kumar, R., Sircar, D., & Chauhan, H. (2019). Genome-wide identification and expression analysis of Hsp70, Hsp90, and Hsp100 heat shock protein genes in barley under stress conditions and reproductive development. Functional & integrative genomics, 19, 1007-1022.
29. Liu, J., Pang, X., Cheng, Y., Yin, Y., Zhang, Q., Su, W., ... & Wan, H. (2018). The Hsp70 gene family in Solanum tuberosum: genome-wide identification, phylogeny, and expression patterns. Scientific reports, 8(1), 1-11.
30. Guo, M., Liu, J. H., Ma, X., Zhai, Y. F., Gong, Z. H., & Lu, M. H. (2016). Genome-wide analysis of the Hsp70 family genes in pepper (Capsicum annuum L.) and functional identification of CaHsp70-2 involvement in heat stress. Plant Science, 252, 246-256.
31. Zhang, J., Liu, B., Li, J., Zhang, L., Wang, Y., Zheng, H., ... & Chen, J. (2015). Hsf and Hsp gene families in Populus: genome-wide identification, organization and correlated expression during development and in stress responses. BMC genomics, 16(1), 1-19.
32. Zhang, L., Zhao, H. K., Dong, Q. L., Zhang, Y. Y., Wang, Y. M., Li, H. Y., ... & Dong, Y. S. (2015). Genome-wide analysis and expression profiling under heat and drought treatments of HSP70 gene family in soybean (Glycine max L.). Frontiers in plant science, 6, 773.
33. Sung, D. Y., Vierling, E., & Guy, C. L. (2001). Comprehensive expression profile analysis of the Arabidopsis Hsp70 gene family. Plant physiology, 126(2), 789-800.
34. Kumar, A., Sharma, S., Chunduri, V., Kaur, A., Kaur, S., Malhotra, N., ... & Garg, M. (2020). Genome-wide identification and characterization of Heat Shock Protein Family reveals role in development and stress conditions in Triticum aestivum L. Scientific reports, 10(1), 1-12.
35. Moabbi, A. M., Agarwal, N., El Kaderi, B., & Ansari, A. (2012). Role for gene looping in intron-mediated enhancement of transcription. Proceedings of the National Academy of Sciences, 109(22), 8505-8510.
36. Jiang, M., & Chu, Z. (2018). Comparative analysis of plant MKK gene family reveals novel expansion mechanism of the members and sheds new light on functional conservation. Bmc Genomics, 19(1), 1-18.
37. Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., & May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis

thaliana. BMC plant biology, 4(1), 1-21.
38. Liu, Y., Jiang, H., Chen, W., Qian, Y., Ma, Q., Cheng, B., & Zhu, S. (2011). Genome-wide analysis of the auxin response factor (ARF) gene family in maize (Zea mays). Plant Growth Regulation, 63, 225-234.
39. Rehman, S., Jørgensen, B., Aziz, E., Batool, R., Naseer, S., & Rasmussen, S. K. (2020). Genome wide identification and comparative analysis of the serpin gene family in brachypodium and barley. Plants, 9(11), 1439.
40. Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. science, 290(5494), 1151-1155.
41. Guy, C. L., & Li, Q. B. (1998). The organization and evolution of the spinach stress 70 molecular chaperone gene family. The Plant Cell, 10(4), 539-556.
42. Hu, W., Hu, G., & Han, B. (2009). Genome-wide survey and expression profiling of heat shock proteins and heat shock factors revealed overlapped and stress specific response under abiotic stresses in rice. Plant Science, 176(4), 583-590.
43. Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., ... & Bryant, S. H. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic acids research, 45(D1), D200-D203.
44. Nagaraju, M., Reddy, P. S., Kumar, S. A., Kumar, A., Rajasheker, G., Rao, D. M., & Kishor, P. K. (2020). Genome-wide identification and transcriptional profiling of small heat shock protein gene family under diverse abiotic stress conditions in Sorghum bicolor (L.). International journal of biological macromolecules, 142, 822-834.
45. Ohama, N., Sato, H., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2017). Transcriptional regulatory network of plant heat stress response. Trends in plant science, 22(1), 53-65.
46. Rodziewicz, P., Swarcewicz, B., Chmielewska, K., Wojakowska, A., & Stobiecki, M. (2014). Influence of abiotic stresses on plant proteome and metabolome changes. Acta Physiologiae Plantarum, 36, 1-19.