Plant Omics Journal POJ

POJ 9(2): 136-141 (2016) ISSN:1836-3644
DOI: 10.21475/poj.160902.p7644x

Genome wide analysis of heat shock transcription factor (HSF) family in chickpea and its comparison with Arabidopsis

Syed Adeel Zafar[1,3], Muzammil Hussain[2], Mubashar Raza[2], Hafiz Ghulam Muhu-Din Ahmed[3], Iqrar Ahmad Rana[4], Bushra Sadia[4,5], Rana Muhammad Atif[3,5]*

[1]Institute of crop science, Chinese Academy of agricultural sciences, Beijing, China
[2]Institute of Microbiology, Chinese Academy of Sciences, Beijing, China
[3]Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad, Pakistan
[4]Center of Agricultural Biochemistry and Biotechnology, University of Agriculture, Faisalabad, Pakistan
[5]US-Pakistan centre for Advanced Studies in Agriculture and Food Security, University of Agriculture, Faisalabad, Pakistan

***Corresponding author: dratif@uaf.edu.pk, adeelzafarpbg@gmail.com**

Abstract

Plants cope with thermo-stress by increased expression of heat shock genes. These genes encode various heat shock proteins (HSPs) which rapidly accumulate and protect plants following hasty heat stress. Heat shock transcription factors (HSFs) primarily regulate expression of *HSP* genes by deciphering conserved binding motifs in promoter region. We retrieved HSF genes of Arabidopsis and chickpea from the online data bases and analyzed their structure and properties using bioinformatics tools. Here, we reported 20 non-redundant genes encoding HSF domain containing proteins in chickpea. Comparative phylogenetic analysis of *HSF* genes with Arabidopsis revealed four major groups with several paralogous and orthologous genes. Gene localization studies showed that *HSF* genes are unevenly distributed across all of the eight chromosomes. Segmental duplications were principally involved in *HSF* gene family expansion during evolution. *HSF* genes predominantly contain a single intron. However, quite a few genes also retain two introns, which suggest gain of intron during the evolutionary process. Combined conserved-domain analysis of Arabidopsis and chickpea HSF proteins revealed presence of 19 most common domains. Comparison of conserved domains with phylogenetic tree has shown that some domains were present in a clade-specific manner. The presence of multiple conserved domains in HSF proteins suggested that the respective genes originate from duplication events. Our *in-silico* work may prove helpful in understanding the evolutionary pathways of HSFs in chickpea.

Keywords: Heat stress; *Cicer arietinum*; *In-silico*; Phylogenetic analysis; Gene evolution.
Abbreviation: GSDS_gene structure display server; HSP_heat shock proteins; HSFs_heat shock transcription factors; HSBPs_heat shock factor binding proteins; HSE_heat shock elements; *Car_Cicer arietinum.*

Introduction

Climate change and global warming has become menace for crop health. Continuous increase in temperature poses significantly negative effect on crop growth and yield (Bita and Gerats, 2013). Plants activate stress responsive pathways to tolerate damaging effects of thermo-stress. Among them, rapid accumulation of heat shock proteins (HSPs) is central. These HSPs are multifunctional like helping in protecting cells against stress damage, folding; intracellular distribution; degradation of proteins, and in signal transduction chains (Hartl et al., 2002; Young et al., 2003). Heat shock proteins are encoded by several heat shock genes and heat shock factor binding proteins (HSBPs) control the expression of these genes (Chen and Zhang, 1997). However, heat shock transcription factors (HSFs) primarily regulate the expression of heat shock genes by recognizing the conserved binding motifs (heat stress element, HSE) which exist in their promoter region. HSEs have an inverted repeat region, which contains a varying number of the DNA sequence (5′-nGAAnnTTCnnGAAn-3′) (Xiao and Lis, 1988; Amin et al., 1988). HSFs utilize their oligomerization domains to form trimmers and function as sequence-specific trimeric DNA binding proteins. These are the terminal compounds of the signal transduction pathway to activate the expression of the *HSP* genes (Chen et al., 2006). It has been observed that at least three repeat HSEs are required for transcription activation in vivo when bound by HSF proteins (Drees et al., 1997). Under normal circumstances, the inactive state of a monomeric HSF is maintained by the interaction with the molecular chaperones, such as Hsp70 and Hsp90. In response to heat stress, these chaperone complexes are converted from a transcriptional inactive monomer to an active trimmer through combination of their oligomerization domains. As sequence-specific trimeric DNA binding proteins, the active HSFs are capable of recognizing and combining HSEs in the HSF-inducible gene promoters (Wang et al., 2012). HSEs are formed of repetitive palindromic binding motifs of the 5'-AGAAnnTTCT-3' sequence upstream of the TATA box in the HSF-inducible genes (Pelham, 1982; Santoro et al., 1998; Guo et al., 2008; Akerfelt, 2010). In plants, *HSF* genes were first identified in tomato (Scharf et al., 1990). Afterward, several *HSF* genes were identified in other crop species including Arabidopsis (Nover et al., 2001; Kotak et al.,

2004), rice (Guo et al., 2008), wheat (Yang et al., 2014) and soybean (Li et al., 2014). HsfA1, HsfA2 and HsfB1 from tomato play key role in heat response by regulating the expression of HSPs and other HSFs (Howarth et al., 1993; Mishra et al., 2002). The HsfA4a identified in wheat was found to be involved in the response to heavy metal stress. Over expression of HsfA4a in rice significantly increased the resistance to heavy metal stress (Shim et al., 2009). Therefore, studies on HSF gene family in chickpea is of prime importance to understand the mechanism of heat tolerance. In the present study, we scanned for and integrated all the non-redundant sets of the chickpea *HSF* genes, determined their chromosomal locations and gene structure, discovered the conserved binding motifs in their proteins and predicted their protein structures by available software and network stations. These results will help in understanding the evolutionary history and functions of HSFs, and in improving the heat tolerance of chickpea.

Results and Discussion

Identification of HSF proteins from chickpea genome

The HSF protein sequences of chickpea and Arabidopsis were obtained from Plant transcription factor database. In initial query, we have obtained 22 chickpea HSFs. However, after finding homology, two of them were removed. Currently, 20 non-redundant HSFs were extracted from the initial 22 HSF sequences of chickpea (Table 1), the polypeptide lengths of chickpea HSFs varied significantly from 156 to 500 amino acids. Previously, 25, 22, 21, 38, and 13 HSFs were identified in rice, Arabidopsis, cucumber, soybean and *Triticum urartu* with polypeptide length ranged from 249-514, 244-495, 184-560, 213-510, and 266-567 amino acids, respectively (Guo et al., 2008; Zhou et al., 2013; Li et al., 2014; Yang et al., 2014). Isoelectric points (IP) of chickpea HSF proteins were also diverse ranging from 4.42 to 9.25. Formerly, a wide range of IP was also observed in cucumber from 4.70 to 9.10 (Zhou et al., 2013), soybean from 4.35 to 9.92 (Li et al., 2014), T. *urartu* from 4.59 to 9.0 and *Aegilops tauschii* from 4.86 to 9.76 (Yang et al., 2014). The molecular weight of chickpea HSF proteins ranged from 55.18 KDa (CarHSF8.3) to 17.77 KDa (*CarHSF4.3*).

Phylogenetic tree

In order to analyze the evolutionary relationship among chickpea and *Arabidopsis thaliana* HSFs, a phylogenetic analysis was done to make a combined phylogenetic tree. The 20 non-redundant chickpea HSFs protein sequences along with 22 Arabidopsis HSFs protein sequences were used to make combined phylogenetic tree (Fig 1). Comparative phylogenetic analysis of chickpea and Arabidopsis HSF transcription factors revealed four major groups of *HSF* genes with several paralogous as well as orthologous genes. Each group contained both chickpea as well as Arabidopsis *HSFs*. Previously, the combined phylogenetic analysis of rice and Arabidopsis *HSFs* divided the phylogenetic tree in three main clusters (Guo et al., 2008). Phylogenetic analysis of cucumber and Arabidopsis HSFs also divided the HSFs in three main clusters (Zhou et al., 2013).

Chromosomal location of chickpea HSFs

The chromosomal locations of the 20 *CarHSF* genes were investigated according to genome sequencing data of chickpea. It was revealed that, 19 chickpea *HSF* genes were unevenly distributed across all of the eight chromosomes, with the exception that one *HSF* gene (XP_004515711.1) was located in scaffold as its location has not yet been assembled (Fig 2). Evolutionary studies suggested that segmental duplications were principally involved in *HSF* gene family expansion during evolution. Chromosome 1 and 2 carried only one *HSF* gene, chromosome 3, 5 and 7 each carried two *HSF* genes, chromosome 6 carried 3 *HSF* genes and chromosome 4 and 8 carried four *HSF* genes. Several reports regarding gene localization studies of various crops reported the presence of *HSF* genes on chromosomes in a very uneven manner. The identified 22 (Arabidopsis) and 25 (rice) *HSF* genes were distributed unequally on all the five and twelve chromosomes, respectively (Guo et al., 2008). Similarly, in case of cucumber, 20 *CsHSF* genes were distributed on all the seven chromosomes but in a very uneven manner. However, the remaining one *CsHSF* gene was present on scaffold chromosome (Zhou et al., 2013). In case of soybean, the 38 identified *GmHSF* genes were distributed unevenly on 15 out of 20 chromosomes (Li et al., 2014).

Discovery of conserved motifs from chickpea and Arabidopsis HSFs

We have identified 19 conserved motifs from 42 HSF sequences of chickpea (20) and Arabidopsis (22). The number of conserved motifs range between 3 and 9 in each chickpea gene. However, in case of Arabidopsis it ranged from 3 to 10 (Fig 1). *CarHSF6.3* had the least number of motifs (motif 1, motif 5 and motif 16), whereas *CarHSF6.2* and *CarHSF8.3* both have 9 motifs which was the highest number in chickpea genome. Motif 1 was found to be conserved in all *CarHSF* gene family members. Zhou et al. (2013) reported 15 conserved motifs in 21 cucumber HSF proteins and the number of motifs in each gene varied between 3 and 12. In contrast, five conserved domains were identified in soybean HSFs (Li et al., 2014).

Gene structure

Gene structure analysis has shown that *HSF* genes predominantly contain a single intron (Fig 1). However, quite a few genes also retain two introns, which suggest gain of intron during the evolutionary process. It was observed that only three out of 20 chickpea *HSF* genes i.e., CarHSF8.1, CarHSF4.3 and CarHSF3.2 have two introns. Similarly two out of 22 Arabidopsis *HSF* genes i.e., AtHSF-01 and AtHSF-14 have double introns, others have single intron. Similarly, in previous research, gene structure analysis revealed that soybean *HSF* genes contain single intron except for one gene which had two introns (Li et al., 2014). Similarly, in cucumber HSF genes predominantly contain a single intron while two genes had double introns with only one gene containing three introns (Zhou et al., 2013). In recent years, the studies on the role of introns have gain considerable success. Studies in fungi, nematodes, insects, mammals and plants suggest that introns not only play role in regulation of gene expression, but also involved in gene evolution (Rose, 2008). Interestingly, gene structure including the position of introns/exons and gene size of *AtHSF09* and *AtHSF10* is very similar which suggest that these genes may be the result of gene duplication. Furthermore, the position of introns and exons of *CarHSF8.4* and *AtHSF03* is very alike, which may suggest their co-evolution in history or it may be the result of horizontal gene flow.

Table 1. Detailed Information of the CarHSFs including gene ID, gene size (Mb), start and end position on chromosome, protein length (aa), molecular weight (KDa) and Isoelectric point (pI).

Number	Gene name	Gene ID	Chromosome	Gene Size (Mb)	Start	End	Strand +/-	MW (KDa)	pI[g]	Domain[h]	Protein size[I] (aa)
1	CarHSF1.1	XP_004485934.1	1	48.35	3522412	3525156	(-)	43.7156	4.42	10-103	382
2	CarHSF2.1	XP_004491224.1	2	36.63	33278060	33279460	(-)	30.173	7.6413	26-119	257
3	CarHSF3.1	XP_004493479.1	3	39.99	29262023	29264248	(+)	27.4287	8.1924	22-112	239
4	CarHSF3.2	XP_004493725.1	3	39.99	30858979	30861997	(+)	35.7585	7.1178	40-131	310
5	CarHSF4.1	XP_004495165.1	4	49.19	1773761	1778741	(-)	41.5982	5.8614	41-134	356
6	CarHSF4.2	XP_004496590.1	4	49.19	12653691	12655097	(+)	30.0481	8.2729	23-114	267
7	CarHSF4.3	XP_004496845.1	4	49.19	13306205	13310504	(+)	17.7732	9.2571	33-121	156
8	CarHSF4.4	XP_004497602.1	4	49.19	22123350	22126002	(+)	50.8657	5.0186	43-130	446
9	CarHSF5.1	XP_004501463.1	5	48.17	35859064	35861334	(-)	41.7825	4.5677	47-139	367
10	CarHSF5.2	XP_004501560.1	5	48.17	36710853	36712187	(+)	42.5756	8.1684	24-117	375
11	CarHSF6.1	XP_004503339.1	6	59.46	2717960	2720632	(-)	41.0083	5.3031	16-107	359
12	CarHSF6.2	XP_004504369.1	6	59.46	10612315	10615415	(-)	54.5022	5.0939	17-110	482
13	CarHSF6.3	XP_004505050.1	6	59.46	16727589	16729629	(-)	23.642	7.7672	31-122	202
14	CarHSF7.1	XP_004507947.1	7	48.96	2928586	2931061	(+)	47.1342	5.1723	102-194	415
15	CarHSF7.2	XP_004509144.1	7	48.96	12362197	12367361	(+)	29.7151	6.5241	8-100	267
16	CarHSF8.1	XP_004511596.1	8	16.47	1870504	1872871	(+)	37.1964	5.1595	24-115	337
17	CarHSF8.2	XP_004511933.1	8	16.47	4225628	4227497	(+)	31.7275	7.9589	9-100	285
18	CarHSF8.3	XP_004512501.1	8	16.47	8539086	8544194	(-)	55.1772	4.5675	32-124	500
19	CarHSF8.4	XP_004512974.1	8	16.47	15082791	15084505	(+)	37.5416	4.7741	12-103	332
20	XP_004515711.1	XP_004515711.1	Unplaced Scaffold		449400	451382	(+)	36.799		29-120	332

Information about gene ID, protein length, MW, pI, chromosome, gene size, start or end position and domain of each gene available at following websites:
http://planttfdb.cbi.pku.edu.cn/family.php?sp=Car&fam=HSF
http://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=101515265

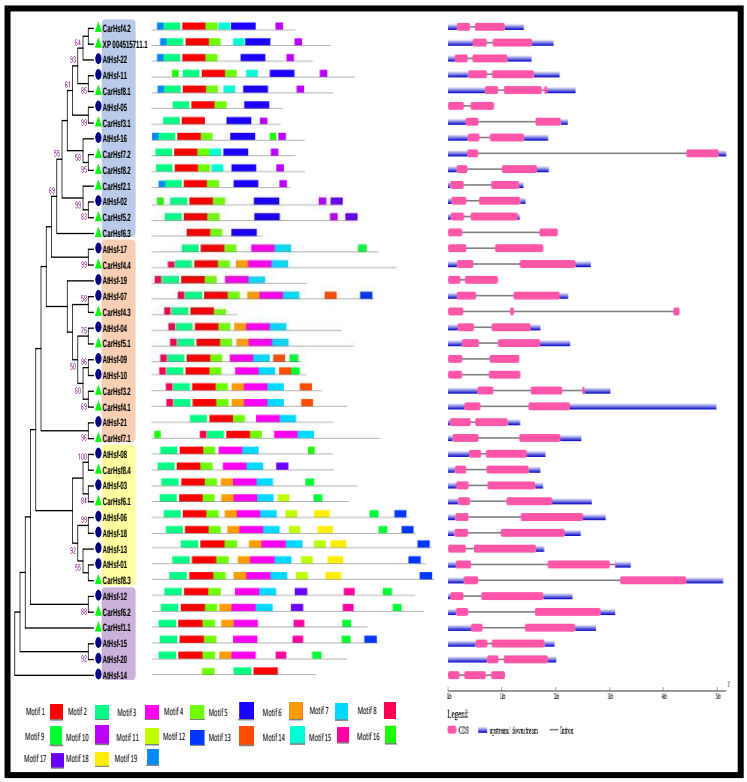

Fig 1. Comparative phylogenetic analysis, conserved domain analysis and gene structure analysis of HSFs of chickpea and Arabidopsis. (A) Phylogenetic tree showed 4 major HSF groups. Green triangles and black circles represent chickpea and Arabidopsis HSFs, respectively. (B) Various conserved domains (colored boxes) identified in Car and At HSFs are shown. (C) Gene structure analysis revealed the presence of single intron in most of the *HSF* genes of chickpea and Arabidopsis.

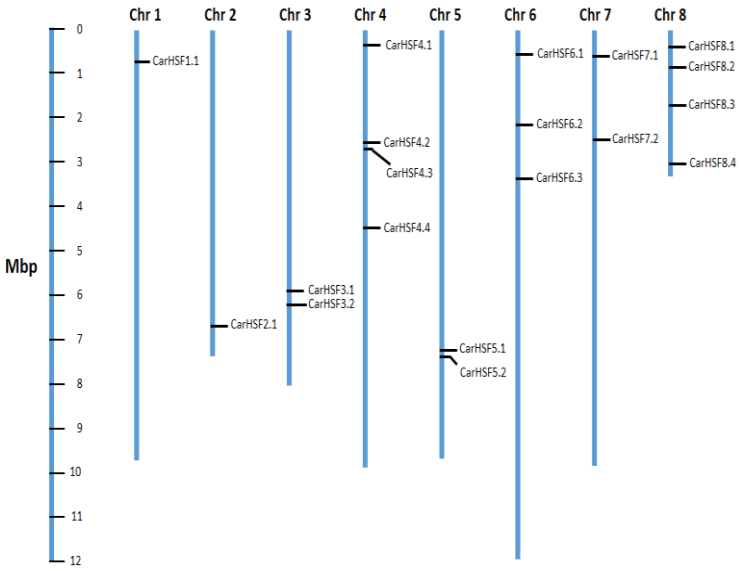

Fig 2. Distribution of Chickpea *HSF* genes on the chromosomes. The scale bar represents Mega bases (Mb) and name of each chromosome is shown at the top of blue bars. All the *HSF* genes are unevenly distributed on eight chromosomes.

Materials and Methods

Data base search and sequence retrieval

The whole genome of *Cicer arietinum* was downloaded from website of International crops research Institute for the Semi-Arid Tropics (http://www.icrisat.org/). Protein and coding sequences of HSFs of chickpea and Arabidopsis were downloaded from plant transcription factor database v3.0, Center for Bioinformatics, Peking University, China (http://planttfdb.cbi.pku.edu.cn/) while genomic sequences were obtained from National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/). The information about gene name, chromosomal location of the gene, start and end point, amino acid length, protein molecular weight, isoelectric point (PI), and Orthologous of Arabidopsis were also obtained from plant transcription factor database.

Comparative phylogenetic analysis

Amino acid sequences of all the identified non-redundant chickpea and Arabidopsis HSFs were aligned using Clustal W (Thompson et al., 1994) program of Molecular Evolutionary Genetics Analysis (MEGA version 6.0) software suite. A phylogenetic tree was constructed by using MEGA 6 software (Tamura et al., 2013) with Neighbor–Joining criteria (Saitou and Nei, 1987) and 1000 bootstrap replicates.

Gene structure analysis

Gene structure including introns and exons of chickpea and Arabidopsis *HSF* genes was investigated by using the online Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/) based on genomic and coding sequences (Hu et al., 2014).

Conserved Motifs analysis

The conserved motifs within chickpea and Arabidopsis HSF proteins were determined by using the MEME online server (Bailey et al., 2015). The parameters were set as follows: maximum numbers of different motifs, 19; minimum motif width, 12; maximum motif width, 52; other parameters retained their default settings.

Chromosomal location of chickpea HSF genes

The identified non-redundant chickpea *HSF* genes were mapped on all the eight chickpea chromosomes on the basis of the information obtained from NCBI using MapDraw software (Liu and Meng, 2003). After locating all the genes on different chromosomes, they were assigned new names on the basis of their location on the chromosomes. For example, the gene which was located on the chromosome number 1 at the start was assigned the name as *CarHSF1.1* and the gene present on the second chromosome was assigned as *CarHSF2.1*. Similarly, the first gene on chromosome number 3 was given name as *CarHSF3.1* while the next gene on the same chromosome as CarHSF3.2 and so on.

Conclusion

In this study, 20 non-redundant HSF genes were identified from the sequenced chickpea genome which was distributed across all the eight chromosomes. These chickpea HSF proteins contain 19 conserved motifs. On the basis of phylogenetic analysis, these genes can be classified into four major groups. Gene structure analysis revealed that *CarHSF* genes predominantly contain a single intron with the exception of three genes which had two introns. This study may provide new insights for the functional characterization of chickpea HSF genes.

Acknowledgements

We thank the anonymous reviewers for helpful comments to the manuscript.

Reference

Akerfelt M, Morimoto RI, Sistonen L (2010) Heat shock factors: integrators of cell stress, development and lifespan. Nat Rev Mol Cell Biol. 11: 545-555.

Amin J, Ananthan J, Voellmy R (1988) Key features of heat shock regulatory elements. Mol Cell Biol. 8: 3761-3769.

Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME Suite. Nucleic Acids Res. 37: 202-208.

Bita CE, and Gerats T (2013) Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. Front Plant Sci. 4: 273.

Chen JN, and Zhang XT (1997) New progress in research on functions of heat shock protein in human and plants. Hereditas. 19: 45-48.

Chen XJ, Ye CJ, Lu HY (2006) Cloning of GmHSFA1 gene and its overexpression leading to enhancement of heat tolerance in transgenic soybean. Acta Genet Sin. 28: 1411-1420.

Drees BL, Grotkopp EK, Nelson HC (1997) The GCN4 leucine zipper can functionally substitute for the heat shock transcription factor's trimerization domain. J Mol Biol. 273: 61-74.

Guo J, Wu J, Ji Q, Wang C, Luo L, Yuan Y, Wang Y, Wang J (2008) Genome-wide analysis of heat shock transcription factor families in rice and Arabidopsis. J Genet Genomics. 35: 105-118.

Guo L, Chen S, Liu K, Liu Y, Ni L, Zhang K, Zhang L (2008) Isolation of heat shock factor HsfA1a-binding sites in vivo revealed variations of heat shock elements in Arabidopsis thaliana. Plant Cell Physiol. 49: 1306-1315.

Hartl, FU, and Hayer-Hartl M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. Science. 295: 1852-1858.

Howarth CJ, and Ougham HJ (1993 Gene expression under temperature stress. New Phytol. 125: 1-2.

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2014) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 31(8): 1296-1297.

Kotak S, Port M, Ganguli A, Bicker F, von Koskull-Doring P (2004) Characterization of C-terminal domains of Arabidopsis heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class A Hsfs with AHA and NES motifs essential for activator function and intracellular localization. Plant J. 39: 98-112.

Li PS, Yu TF, He GH, Chen M, Zhou YB, Chai SC, Xu ZS, Ma YZ (2014) Genome-wide analysis of the Hsf family in soybean and functional identification of GmHsf-34 involvement in drought and heat stresses. BMC Genomics. 15(1): 1009.

Liu RH, and Meng, JL (2003) MapDraw: a microsoft excel macro for drawing genetic linkage maps based on given genetic linkage data. Hereditas. 25: 317-321.

Mishra S K, Tripp J, Winkelhaus S, Tschiersch B, Theres K, Nover L, Scharf KD (2002) In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermo-tolerance in tomato. Gene Dev. 16: 1555-1567.

Nover L, Bharti K, Doring P, Mishra SK, Ganguli A, Scharf, KD (2001) Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperones. 6: 177-189.

Pelham HR (1982) A regulatory upstream promoter element in the Drosophila hsp 70 heat-shock gene. Cell. 30: 517-528.

Rose A. (2008) Intron-mediated regulation of gene expression, nuclear pre-mRNA processing in plants, Springer. pp. 277-290.

Saitou N, and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4(4): 406-425.

Santoro N, Johansson N, Thiele DJ. (1998) Heat shock element architecture is an important determinant in the temperature and transactivation domain requirements for heat shock transcription factor. Mol Cell Biol. 18: 6340-6352.

Scharf KD, Rose S, Zott W, Schoffl F, Nover L. (1990) Three tomato genes code for heat stress transcription factors with a region of remarkable homology to the DNA-binding domain of the yeast Hsf. EMBO J. 9: 4495-4501.

Shim D, Hwang JU, Lee J, Lee S, Choi Y, An G, Martinoia E, Lee Y (2009) Orthologs of the class A4 heat shock transcription factor HsfA4a confer cadmium tolerance in wheat and rice. Plant Cell. 21: 4031-4043.

Tamura K, Stecher, G, Peterson, D, Filipski, A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 30(12): 2725-2729.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22(22): 4673–4680.

Wang F, Dong Q, Jiang H, Zhu S, Chen B and Xiang Y (2012) Genome-wide analysis of the heat shock transcription factors in *Populus trichocarpa* and *Medicago truncatula*. Mol Biol Rep. 39: 1877-1886

Xiao H, and Lis, JT (1988) Germ line transformation used to define key features of heat-shock response elements. Science. 239: 1139-1142.

Yang W, Li J, Liu D, Sun J, He L, Zhang A (2014) Genome-wide analysis of the heat shock transcription factor family in '*Triticum urartu*' and '*Aegilops tauschii*'. Plant Omics. 7(5): 291.

Young JC, Barral JM, Ulrich HF (2003) More than folding: localized functions of cytosolic chaperones. Trends Biochem Sci. 28: 541-547.

Zhou S, Zhang P, Jing Z, Shi J (2013) Genome-wide identification and analysis of heat shock transcription factor family in cucumber ('*Cucumis sativus*' L.). Plant Omics. 6: 449.