Journal of Biology, Agriculture and Healthcare
ISSN 2224-3208 (Paper) ISSN 2225-093X (Online)
Vol.6, No.1, 2016

Genome Wide Analysis of Heat Shock Factors (HSF) Gene Family of Arabidopsis Thaliana

Shaiq Sultan[1*] Mohsin Ali[1] Samia Nawaz[1] Muhammad Amjad Ali[2,3] Aamar Shahzad[4]
1. Department of Bioinformatics & Biotechnology, Government College University, 38000 Faisalabad, Pakistan
2. Department of Plant Pathology
3. Centre of Agricultural Biochemistry & Biotechnology, University of Agriculture, 38040 Faisalabad, Pakistan
4. Department of Plant Breeding and Genetics, University of Agriculture, Faisalabad-38040, Pakistan

Abstract
Heat shock factors (HSF) are one of the most important regulators which control heat stress, damage and other biological processes. In HSF family, genes have been properly characterized in tomato and many other plants. In this study, the genome wide analysis of heat shock factors was performed in Arabidopsis thaliana family to understand the genomic information of HSF. Twenty-four members of HSF family were retrieved in *Arabidopsis thaliana* after structural characteristics and phylogenetic comparison. Twenty-four members of HSF divided into three subclasses according to conservation in structure. Plant Transcriptional factor database (TFDB) analysis was also used to find out location of uneven distribution of HSF five chromosomes in *Arabidopsis thaliana*. Further, conserved motifs and domain of HSF family were characterized. Gene structure analysis was used for intron and exon number and their location information of all genes of HSF. On the bases of promoter analysis, five cis-regulatory elements have been selected and then figured out on thousand base pairs of promoter sequence. Depending upon this information, one would be able to understand the genomic analysis of HSF family in *Arabidopsis thaliana* and can be further used for comparison to other species. This whole study contains the knowledge about the genome wide analysis of genes of Heat Stress factors in the *Arabidopsis thaliana,* and it also elaborate that how the HSF works and plays an important role in the heat stress conditions.
Keywords: HSF, Arabidopsis thaliana, domain, genome wide analysis

1. Introduction

Signal transduction chain activates the genes for heat stress (HS) and other chemical stress (CS). This chain has heat stress transcription factors (HSFs) as terminal segments. In all species, there must be heat shock mechanism is present in which different proteins are involved (Li et al. 2014). F. Ritossa perceived a pattern on heat shock in fruit fly Drosophila buschiiin in 1962. He found this on the polytene chromosomes of Drosophila (Ritossa 1962).In this mechanism, different reactions occur like folding, targeting, protein degradation and other biological functions. And here HSPs behave like molecular chaperones (Boston et al. 1996; Xu et al.). And these chaperones play a particular role in restoration, homeostasis and maintenance. In all eukaryotes, for a heat stress response, HSF (heat stress response) is a main module. These have basic structure and *cis* regulatory elements, and their target genes are highly conserved (Baniwal et al. 2004; Morimoto 1998; Nover et al. 2001b). The size of HSF family is large for the plants in compare to other organisms (Czarnecka et al. 2004).

HSFs have a modular structure, such as in *Arabidopsis thaliana* HsfA2, a typical illustrative of plant HSFs (Baniwal et al. 2004; Nover et al. 2001a).*Arabidopsis thaliana* have 24 genes of HSF family divided into three classes based on oligomerization domains (Kotak et al. 2007a; Kotak et al. 2004; Kotak et al. 2007b). Class B genes might be repressor or co-activators by relating with class A (Czarnecka-Verner et al. 2004; Hahn et al. ; Ikeda et al.). Class A is basically acted as transcriptional activators. For class C, functions not describe yet now in *Arabidopsis Thaliana* (Nover et al. 2001b). In a fast response for heat, four factors from class A play a central role (Busch et al. 2005; LIU et al. 2011; Panchuk et al. 2002; Yamada et al. 2007). One factor of class A (HSFA1b) have ability that when it over expresses then it recovers thermo tolerance (Galvez-Valdivieso et al. 2009; Nishizawa et al. 2006; Ogawa et al. 2007; Panchuk et al. 2002).

In class B, the HR-A/B parts are similar to non-plants HSFs. Due to addition of twenty one (class A) and seven (class C) in the hydrophobic parts of class A and B, these classes have prolonged HR-A/B regions. Sequence variation is shown by class C Hsfin; the DNA binding domain. In *Arabidopsis thaliana* just one member of class C has been found, but evidence shows that other members of class C are present in other plant's species (Nover et al. 2001b). For the activation of class A, many motifs that can be (aromatic, hydrophobic, acidic amino acids) perceived at the C terminal domain. (DA ring et al. 2000).Fifteen members of class A of HSF are unclear in *Arabidopsis thaliana* (Lohmann et al. 2004).

There is a leucine-rich export signal in the C-terminal of some plant HSFs. The balance of NLS and NES tells the actual nucleocytoplasmic dispersion of the protein, which is vital for many signaling pathways involving transcription factors (GA rlich and Kutay 1999; Heerklotz et al. 2001; Kotak et al. 2004).

HSF size in plants evolves to maintain response to other stresses also (Kotak et al. 2004). Many genes of HSF family also give a response when another environmental factor affects the plant (Miller and Mittler 2006;

Swindell et al. 2007)like during high light one gene of HSF family (HSFA2) give response and become activate(Banti et al. ; Miller and Mittler 2006; Nishizawa et al. 2006; Panchuk et al. 2002; Rizhsky et al. 2002; Schramm et al. 2006). One more gene (HSFA9) has a very important role in seed maturation (Kotak et al. 2007b).). Some HSF genes of class A have a role on growth and development (Kotak et al. 2007b; LIU et al. 2011)) their effects may have influence on other traits (Morison et al. 2008).

With the progressive access to genome Annotation of the model organisms, phylogenetic analysis of a set of homologous sequences or a gene family in various species has been a useful way to conclude the functional diversity and evolutionary relationship(FENG et al. 2004; Li and Yang 2003).In this case, vast, broad and no redundant set Arabidopsis HSF genes, figured out their presumed structures, and relatively analyzed the Arabidopsis HSFs.

In our study, we have focused on these HSF genes in Arabidopsis. These genes have given new annotations according to their position on the chromosomes. Moreover, all these HSF genes are analyzed bioinformatically depending upon their stress response individually and as a whole, to interpret the common protein features responsible for different stresses. Furthermore, the promoter and motif analysis of stress-related HFSs have also been done.

2. Materials and Methods

2.1 Database search of heat stress proteins in A. thaliana

The accession number of HSF family was collected by using of plant TFDB version 3.0 (http://planttfdb.cbi.pku.edu.cn). It gave a list of the accession number of all the genes of this family in required specie. Protein sequences of all HSFs also retrieved from this site.For promoter sequence a website Phytozome v9.1is used. Promoter of 1kb upstream to start codon retrieved for cis regulatory element analysis (http://www.phytozome.net/).To check more reliability TAIR(http://www.arabidopsis.org/)was used.

2.2 Tree building

The phylogenetic tree of the Arabidopsis HSF protein sequences was constructed using MEGA (version 6.0) (Kumar and Gadagkar, 2001). In this process, the N-J method was used with the following parameters: bootstrapped with 1,000 iterations, Gap open penalty: 10; Gap extension penalty: 0.1; Residue-specific gap penalties: on; Hydrophilic penalties: on; Gap separation distance: 0; End gap separation penalty: on; Use negative matrix: Off; Delay divergent cutoff: 30%. The Dayhoff PAM matrix in the protein distance algorithm and N-J method were applied to construct the unrooted tree, and this tree file was showed by Tree View.

2.3 Domain prediction

The conserved amino acids were highlighted using MEME suite online software Version 4.9.1 (motif based sequence analysis tool) (http://meme.nbcr.net/meme/cgi-bin/meme.cgi) . It identified the entire motif and conserved regions among all genes of HSF family. MEME analysis was run with following parameters: number of repetitions: any number; maximum number of motifs: 20; the optimum motif widths were constrained to within 20 and 90 residues.

2.4 Chromosomal mapping of HSF genes

Chromosomal mapping was Arabidopsis It based on the information that got by phytozome (http://www.phytozome.net/) Version 9.1 and NCBI (http://www.ncbi.nlm.nih.gov/) websites. And the graphical mapping of these genes on five chromosomes was done by scaling.

2.5 Multiple sequence alignment

To verify the conserved region multiple sequence alignment was done. And it performed by using UNIPROGENE software and COBALT constraint-based multiple alignment tools (http://www.stva.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi) by their by default settings.

2.6 Gene structure analysis

The genomic information of all genes of HSFs obtained from the TAIR (www.arabidopsis.org/) that identified the exon, intron and UTR exact value location on the gene. A diagrammatic map was made by the help of Microsoft Power Point to show all of these features.

2.7 Cis regulatory elements analysis

Here 1kb nucleotide sequences upstream to start codon in all genes of HSF were retrieved from Arabidopsis Thaliana specie by using the Phytozome (http://www.phytozome.net/) Version 9.1. To find out the cis regulatory elements, these promoter sequences were subjected in plant databases PLACE (Higo et al. 1999) (http://www.dna.affrc.go.jp/PLACE/). This database is already experimentally defined (Higo et al. 1999), and it

has 469 experimentally proved cis regulatory elements. Then five elements that are CAAT BOX1, GATA BOX, GTGANTG10, TATABOX5, and WBOXNTERF3 selected and scaled and mapped by using Microsoft Office power point for each gene.

3. Results

3.1 Identification of Arabidopsis thaliana HSFs

By using the database plantTFDB (plant transcriptional factor database) all 24 transcription factors of HSF family were given their accession numbers and amino acid sequences.

3.2 Phylogenetic analysis of HSF proteins

The protein sequences of all factors were then placed in MEGA 5.0 version (Tamura et al., 2011). for multiple sequence alignments by using the ClustalW and constructed the phylogenetic tree. For this purpose, neighbor joining method is used with 1,000 bootstrap. Further, it is divided into major two classes at the base of its evolution.

3.3 Mapping HSF genes on A. thaliana chromosomes

All the genes of HSF family physically localized on the chromosomes of A. Thaliana with the help of CLC Sequence Viewer and manually in excel sheet. The predicted genes are 24, and these genes are distributed along five chromosomes of A. Thaliana.. And the distribution of these genes is shown in the figure 2. Comparison with other chromosomes shows that, chromosome two has the minor number of genes that are two, while chromosome one shares four genes and chromosome three, four and five have five, seven and six genes respectively. Location of all 24 genes is presented in figure 2.

3.4 Comparison of DNA-binding HSF domains

The HSF DNA-binding domains of all the heat stress responsive HSF proteins available in the N-terminal region were compared for presence of conserved residues by alignment of complete HSF domain with the help of COBALT from NCBI server (http://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi). The HSF domain in all the WRKY proteins of A. thaliana shown a high level of conservation with 51 to 151 amino acids being percent conserved in all proteins, including the HSF signature and zinc finger motif (Figure 3).

3.5 Gene structure analysis

To check more deeply into the evolution of heat stress genes, the intron-exon structure of HSF family of A. Thaliana is analyzed. It was noticed here that all the members of heat stress responsive HSF family start from exon, and all the family had intron varying showed in figure 4. By these results, it was noticed that the average number of exon introns are two. Largest gene had three exons. Sixteen genes had two exons. Seven genes had three exons. There is one gene having only one exon .

3.6 Analysis of Cis regulatory elements

TFs give a response to be different stresses based on their transcriptional regulation. To check out the critical role of these HSF transcriptional factors, we got the 1Kb promoter sequences of all the 24 genes of HSF family upstream to the start codon that is (ATG). And then one by one placed this promoter sequence in PLACE software (Higo et al., 1998, 1999). The repeated cis elements searched during search are shown in the Figure5. The sites that are largely found are CAAT BOX1, GATA BOX, GTGANTG10, TATABOX5, and WBOXNTERF3. These genes selected and then assigned some signs shown in figure 5 and placed these on all HSF genes. Quantity of all selected regulatory elements at every gene is represented in Table2 and total amount of these elements at each gene is also shown in this table.

4. Discussion

Cells and organisms use the heat-shock response to avoid the damaging effects of stresses (heat). This is evolutionary conserved mechanism (Nover et al. 2001b). However, after extreme study of HSFs family, it has been noticed that many, if not all, HSF transcription factors from Arabidopsis thaliana and rice species are represented in 47 sequences; 22 from Arabidopsis thaliana and 25 from Rice (Guo et al. 2008).

Similarly, in previous work almost 59 heat shock factors were reported (Chung et al.) Other monocots like rice and maize also have 59 heat shock factors. In dicot plants like soybean, the number of HSF are about 38 and in *Arabidopsis thaliana*, about 21 factors are shown, but here in this work quantity of heat shock factors is about 24. According to evolution, it's all due to the double duplication in soybean (Schlueter et al. 2004) and single duplication in *Arabidopsis thaliana* (Blanc et al. 2000). Altough, the classification of HSFs in all species (either they are dicots and monocots) is same, as well as, uniqueness is also there.

However, in latest research, 24 HSFs have been found from the database plantTFDB (plant

transcriptional factor database). According to Guo's work, HSFs genes presents on all the five chromosomes of Arabidopsis thaliana and our work also concluded it. Nevertheless, Guo also worked on Rice HSFs and figured out that chromosome number 11 and 12 of Rice have fewer HSFs genes. Guo's study and our study are showing that the genes of HSF are present on all the chromosomes, and these distributed widely in the chromosomes of the common ancestor.

In this article, we studied HSF gene family in Arabidopsis thaliana. Protein sequences of 24 genes of HSF family obtained from plantTFDB. Promoter sequences also retrieved of these genes from phytozome (http://www.phytozome.net/). Phylogenetic analysis done for the 24 genes of the HSF and then constructed the unrooted tree like the work of unrooted trees of Guo and Lutz Nover (Guo et al. 2008; Nover et al. 2001a). Amino acid's alignment used and found the domain of HSF family by using of MEME suite where all motifs were present. The conserved motifs for each protein taken and then placed in against of each HSF gene member in the unrooted tree to determine the Gene's location on the chromosome. Similarly, our work indicated the distribution of genes on all five chromosomes; as Guo explained the presence of gene on all the five chromosome of Arabidopsis thaliana. However, most of these genes were located on chromosome number 4 and unevenly distributed.

The 5 cis-regulatory elements (CAATBOX1, GATABOX, GTGANTG10, TATABOX5, WBOXNTERF3) were selected and their sequences were (CAAT, GATA, GTGA, TTATTT, TGACY) located on 1kb promoter sequence of genes. And these cis-regulatory elements play an important role in stress conditions. Guo's work indicated about 22 factors but not about cis-regulatory elements, but our analysis provides information about 24 factors and also indicated the promoter analysis. This study helped a lot to understand the HSF family genome fundamental information in Arabidopsis thaliana. It also indicated that intron andexon structure makes sure to get the evolutionary relationship of HSF family.

Lastly, these findings will help to get the concepts of responses of HSF in stress condition and the whole genomic information of HSF family. This data will facilitate selecting candidate genes for stress condition and further functional and comparative characterization.

Reference

- Baniwal SK, Bharti K, Chan KY, Fauth M, Ganguli A, Kotak S, Mishra SK, Nover L, and Port M, Scharf K (2004). Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. Journal of biosciences 29: 471-487
- Banti V, Mafessoni F, Loreti E, Alpi A, Perata P (2010). The heat-inducible transcription factor HsfA2 enhances anoxia tolerance in Arabidopsis. Plant Physiology 152: 1471-1483
- Blanc G, Barakat A, Guyot R, Cooke R, Delseny M (2000). Extensive duplication and reshuffling in the Arabidopsis genome. The Plant Cell Online 12: 1093-1101
- Boston RS, Viitanen PV, Vierling E (1996). Molecular chaperones and protein folding in plants. Post-Transcriptional Control of Gene Expression in Plants. Springer, pp 191-222
- Busch W, Wunderlich M, Schaffl F (2005). Identification of novel heat shock factoral dependent genes and biochemical pathways in Arabidopsis thaliana. The Plant Journal 41: 1-14
- Chung E, Kim K, Lee J (2013).Genome-wide analysis and molecular characterization of heat shock transcription factor family in Glycine max. Journal of genetics and genomics 40: 127-135
- Czarnecka E, Pan S, Salem T, Gurley WB (2004). Plant class B HSFs inhibit transcription and exhibit affinity for TFIIB and TBP. Plant molecular biology 56: 57-75
- Feng Y, LIU P, XUE Z (2004). Comparative Phylogenetic Analysis of the Rice and Arabidopsis PHD-finger Proteins. Acta Genetica Sinica 31: 1284-1293
- Gorlich D, Kutay U (1999). Transport between the cell nucleus and the cytoplasm. Annual review of cell and developmental biology 15: 607-660
- Galvez G, Fryer MJ, Lawson T, Slattery K, Truman W, Smirnoff N, Asami T, Davies WJ, Jones AM, Baker NR (2009). The high light response in Arabidopsis involves ABA signaling between vascular and bundle sheath cells. The Plant Cell Online 21: 2143-2162
- 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. Journal of genetics and genomics 35: 105-118
- Hahn A, Bublak D, Schleiff E, Scharf D (2011). Crosstalk between Hsp90 and Hsp70 chaperones and heat stress transcription factors in tomato. The Plant Cell Online 23: 741-755
- Heerklotz D, Daring P, Bonzelius F, Winkelhaus S, Nover L (2001). The balance of nuclear import and export determines the intracellular distribution and function of tomato heat stress transcription factor HsfA2. Molecular and cellular biology 21: 1759-1768
- Higo K, Ugawa Y, Iwamoto M, and Korenaga T (1999). Plant cis-acting regulatory DNA elements

(PLACE) database. Nucleic Acids Research 27 (1): 297-300.

- Higo K, Ugawa Y, Iwamoto M and Higo H (1998). PLACE: A database of plant cis-acting regulatory DNA elements. Nucleic Acids Res. 26: 358-359.
- Ikeda M, Mitsuda N, Ohme-Takagi M (2011). Arabidopsis HsfB1 and HsfB2b act as repressors of the expression of heat-inducible Hsfs but positively regulate the acquired thermotolerance. Plant Physiology 157: 1243-1254
- Kotak S, Larkindale J, Lee U, Daring P, Vierling E, Scharf D (2007a). Complexity of the heat stress response in plants. Current Opinion in Plant Biology 10: 310-316
- Kotak S, Port M, Ganguli A, Bicker F, Daring V (2004). Characterization of Ca 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. The Plant Journal 39: 98-112
- Kotak S, Vierling E, Baoumlein H, Daring P (2007b). A novel transcriptional cascade regulating expression of heat stress proteins during seed development of Arabidopsis. The Plant Cell Online 19: 182-195
- Li C, Yang Q (2003). Phylogenetic relationships among the genera of Taxodiaceae and Cupressaceae from 28S rDNA sequences. Yi chuan Hereditas Zhongguo yi chuan xue hui bian ji 25: 177-180
- Li S, Yu F, He H, Chen M, Zhou B, Chai C, Xu S, Ma Z (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: 1009
- LIU C, LIAO T, CHARNG Y (2011). The role of class A1 heat shock factors (HSFA1s) in response to heat and other stresses in Arabidopsis. Plant, cell & environment 34: 738-751
- Lohmann C, Schumacher G, Wunderlich M, Schaffl F (2004) Two different heat shock transcription factors regulate immediate early expression of stress genes in Arabidopsis. Molecular Genetics and Genomics 271: 11-21
- Miller G, Mittler R (2006). Could heat shock transcription factors function as hydrogen peroxide sensors in plants. Annals of Botany 98: 279-288
- Morimoto RI (1998). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. Genes & development 12: 3788-3796
- Morison J, Baker N, Mullineaux P, Davies W (2008). Improving water use in crop production. Philosophical Transactions of the Royal Society B: Biological Sciences 363: 639-658
- Nishizawa A, Yabuta Y, Yoshida E, Maruta T, Yoshimura K, Shigeoka S (2006). Arabidopsis heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. The Plant Journal 48: 535-547
- Nover L, Bharti K, Da ring P, Mishra SK, Ganguli A, Scharf D (2001a). Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell stress & chaperones 6: 177
- Nover L, Bharti K, Daring P, Mishra SK, Ganguli A, Scharf D (2001b). Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell stress & chaperones 6: 177
- Ogawa D, Yamaguchi K, Nishiuchi T (2007). High-level overexpression of the Arabidopsis HsfA2 gene confers not only increased themotolerance but also salt/osmotic stress tolerance and enhanced callus growth. Journal of Experimental Botany 58: 3373-3383
- Pascal D, Treuter E, Kistner C, Lyck R, Chen A, Nover L (2000). The role of AHA motifs in the activator function of tomato heat stress transcription factors HsfA1 and HsfA2. The Plant Cell Online 12: 265-278
- Panchuk II, Volkov A, Schaffl F (2002). Heat stress-and heat shock transcription factor-dependent expression and activity of ascorbate peroxidase in Arabidopsis. Plant Physiology 129: 838-853
- Ritossa F (1962). A new puffing pattern induced by temperature shock and DNP in Drosophila. Experientia 18: 571-573
- Rizhsky L, Liang H, Mittler R (2002). The combined effect of drought stress and heat shock on gene expression in tobacco. Plant Physiology 130: 1143-1151
- Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC (2004). Mining EST databases to resolve evolutionary events in major crop species. Genome 47: 868-876
- Schramm F, Ganguli A, Kiehlmann E, Englich G, Walch D, Daring P (2006). The heat stress transcription factor HsfA2 serves as a regulatory amplifier of a subset of genes in the heat stress response in Arabidopsis. Plant molecular biology 60: 759-772

- Swindell WR, Huebner M, Weber AP (2007). Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. BMC genomics 8: 125
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol., 28(10):2731–2739.
- Xu Z-S, Li Z-Y, Chen Y, Chen M, Li L-C, Ma Y-Z Heat shock protein 90 in plants: molecular mechanisms and roles in stress responses. International journal of molecular sciences 13: 15706-15723
- Yamada K, Fukao Y, Hayashi M, Fukazawa M, Suzuki I, Nishimura M (2007) Cytosolic HSP90 regulates the heat shock response that is responsible for heat acclimation in Arabidopsis thaliana. Journal of Biological Chemistry 282: 37794-37804

Tables

Table (1): Stress responsive HSF transcription factors identified in *Arabidopsis thaliana.*

Index	Account number	Location	start	end	exons	introns
1	AT1G32330.1	ch#1	11656986	11660376	3	2
2	AT1G46264.1	ch#1	17224842	17226280	2	1
3	AT1G67970.1	ch#1	25484699	25486462	2	1
4	AT1G77570.1	ch#1	29143350	29144003	1	0
5	AT2G26150.1	ch#2	11135682	11137396	2	1
6	AT2G41690.1	ch#2	17381723	17382577	2	1
7	AT3G02990.1	ch#3	673472	676403	2	2
8	AT3G22830.1	ch#3	8078820	8081051	2	1
9	AT3G24520.1	ch#3	8941066	8942874	2	1
10	AT3G51910.1	ch#3	19265294	19266619	2	1
11	AT3G63350.1	ch#3	23399468	23400812	2	1
12	AT4G11660.1	ch#4	7042630	7044708	2	1
13	AT4G13980.1	ch#4	8076980	8079314	2	2
14	AT4G17750.1	ch#4	9869818	9871603	2	1
15	AT4G18870.1	ch#4	10346168	10347228	3	2
16	AT4G18880.1	ch#4	10347582	10349562	2	1
17	AT4G19630.1	ch#4	10683936	10684683	2	1
18	AT4G36990.1	ch#4	17440177	17442151	1	1
19	AT5G03720.1	ch#5	971786	973683	2	1
20	AT5G16820.1	ch#5	5530273	5532792	2	2
21	AT5G43840.1	ch#5	17625437	17626364	2	1
22	AT5G45710.1	ch#5	18540850	18542863	2	2
23	AT5G54070.1	ch#5	21943983	219456512	2	2
24	AT5G62020.1	ch#5	24916089	24917643	2	1

Table 2. The number of *cis*-regulatory elements (CAAT BOX1, GATA BOX, GTGANTG10, TATABOX5, and WBOXNTERF3) in the genes is shown in the table.

Accession number	CAATBOX1	GATABOX	GTGANTG10	TATABOX5	WBOXNTERF3	Total
AT1G32330.1	2	4	2	4	1	13
AT1G46264.1	5	10	3	2	1	21
AT1G67970.1	6	4	6	2	3	21
AT1G77570.1	6	3	5	1	2	17
AT2G26150.1	7	7	7	0	1	22
AT2G41690.1	4	3	1	1	0	9
AT3G02990.1	7	7	3	0	2	19
AT3G22830.1	4	3	3	1	1	12
AT3G24520.1	8	1	2	1	1	13
AT3G51910.1	6	3	4	0	2	15
AT3G63350.1	6	4	0	2	0	11
AT4G11660.1	6	8	4	4	1	23
AT4G13980.1	6	8	4	4	1	23
AT4G17750.1	8	8	3	0	0	19
AT4G18870.1	8	6	4	2	2	22
AT4G18880.1	8	6	4	1	1	20
AT4G19630.1	6	3	5	0	1	15
AT4G36990.1	6	6	2	0	1	15
AT5G03720.1	12	3	4	1	5	25
AT5G16820.1	6	8	4	0	1	19
AT5G43840.1	7	4	1	1	3	16
AT5G45710.1	6	6	5	2	1	20
AT5G54070.1	12	3	3	0	1	19
AT5G62020.1	11	5	4	1	1	22

Figures

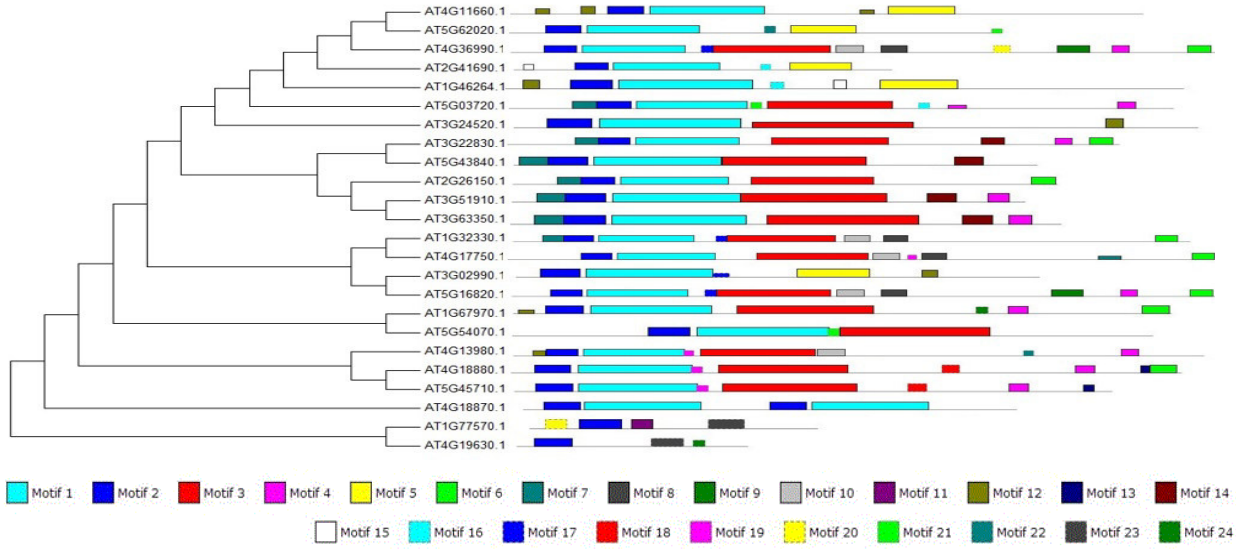

Figure 1 : Phylogenetic relationship and motif presence on the heat stress related HSF members within Arabidopsis thaliana. **1.**Multiple alignments of 24 of 25 all heat stress related amino acids of HSF genes from Arabidopsis thaliana were made by Clustal W and the phylogenetic tree was built using MEGA 6.0 by the Neighbor-Joining (NJ) method with 1,000 bootstrap replicates. While, different colourfull boxes are showing the HSF groups present in the Arabidopsis. **2**. Schematic diagram of the conserved motifs in the HSF proteins from Arabidopsis explicate from online MEME server. 24 domains of different lengths are interpreted on the protein map against phylogenetic tree. While the motif symbols are showed under the tree.

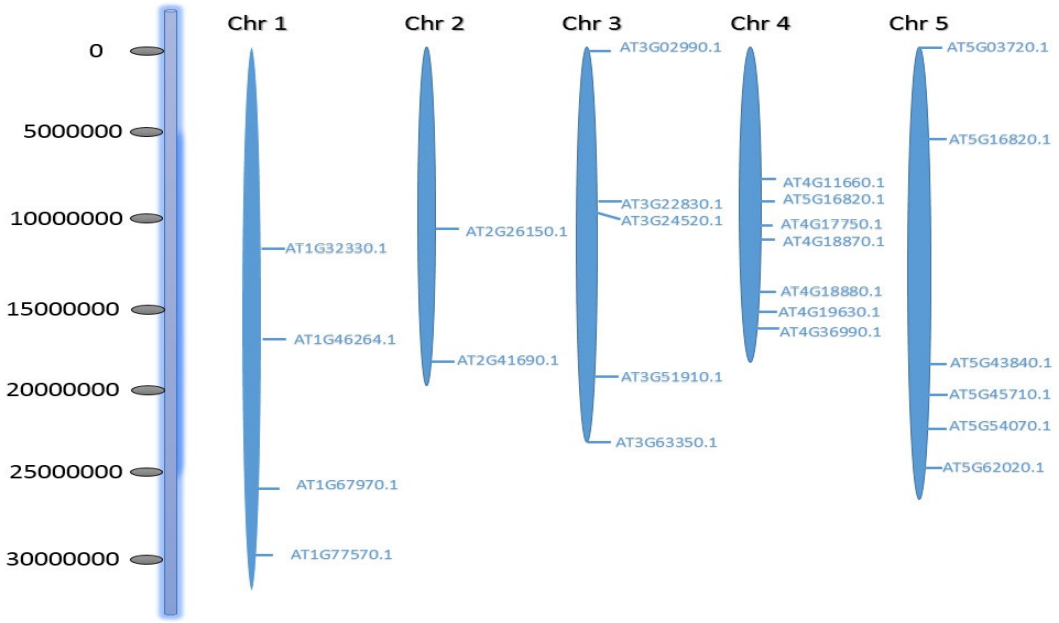

Figure 2:The localization of 24 heat Stress responsive **HSF** genes on Arabidopsis thaliana chromosomes. The chromosomes numbers are present up to the all chromosomes. HSF Genes are named according to their position and size on the chromosome and showed by their accession number and indicated by the arrow also mentioned in a map. The relative position of At HSF and size of chromosome represented using vertical scale present in left.

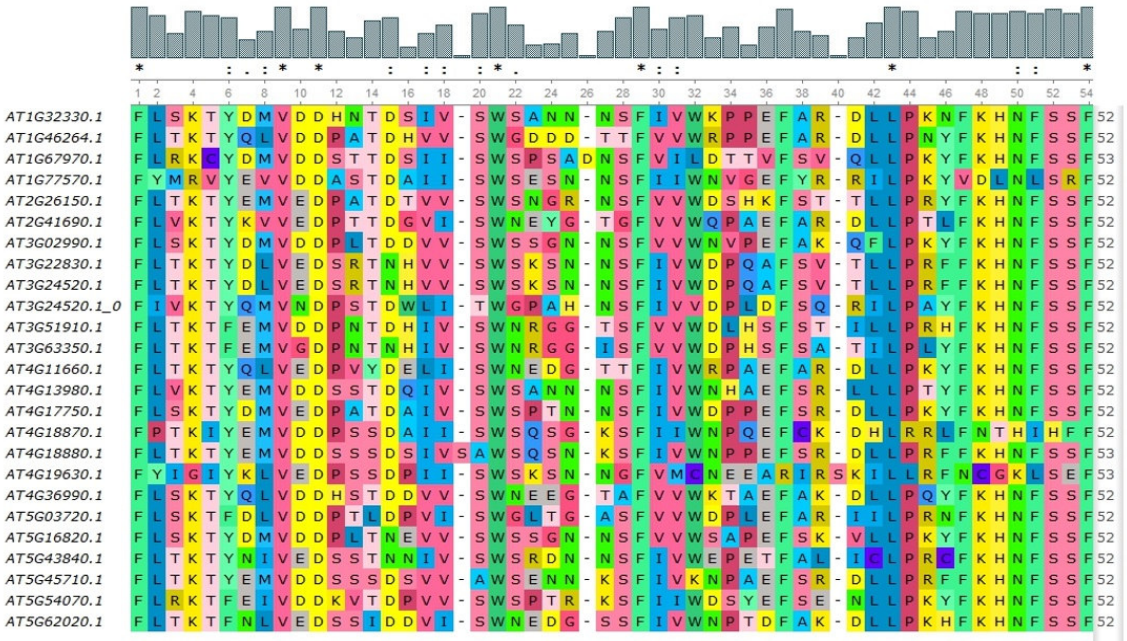

Figure 3: A comparison of HSF DNA binding domain among heat Stress responsive 24 At HSF proteins; using **COBALT**-Blast and Unipro-**UGENE** software. 1 to 54 residue HSF domain has shown in different colors scheme of UGENE software. The three signs shown as '*', ';' and '.' are use to mark conserved positions according to the Clustal X conventions.

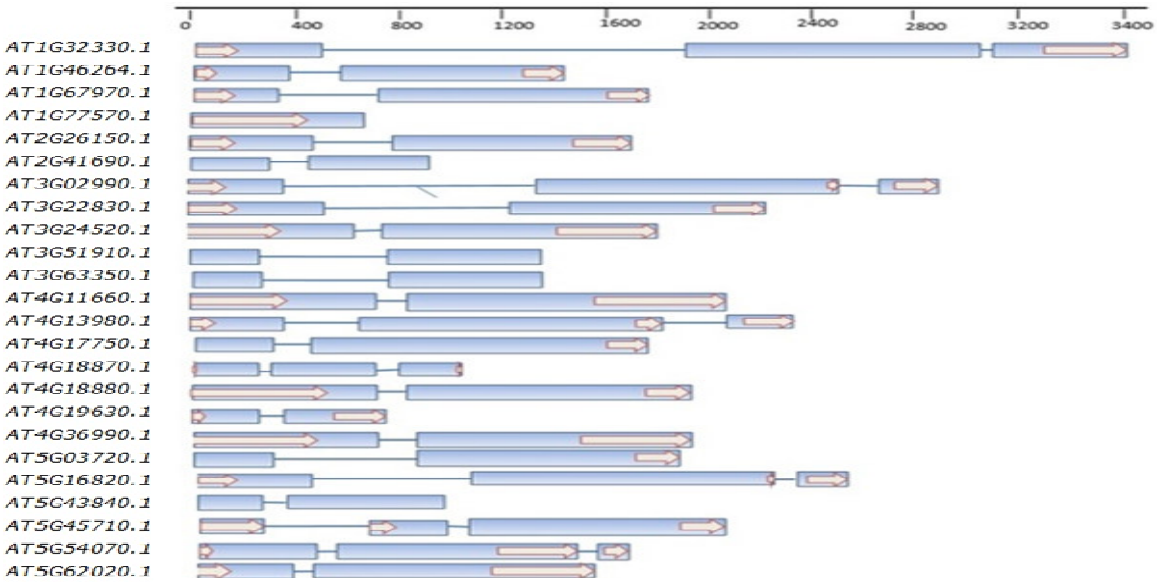

Figure 4: Exon/intron structures of HSF genes from Arabidopsis thaliana. Blue color box are showing the exon part of the gene and pink colored arrows are indicating the untranslated regions and Blue color lines are depicting introns in all the genes. The sizes of exons and introns can be estimated using the scale at top.

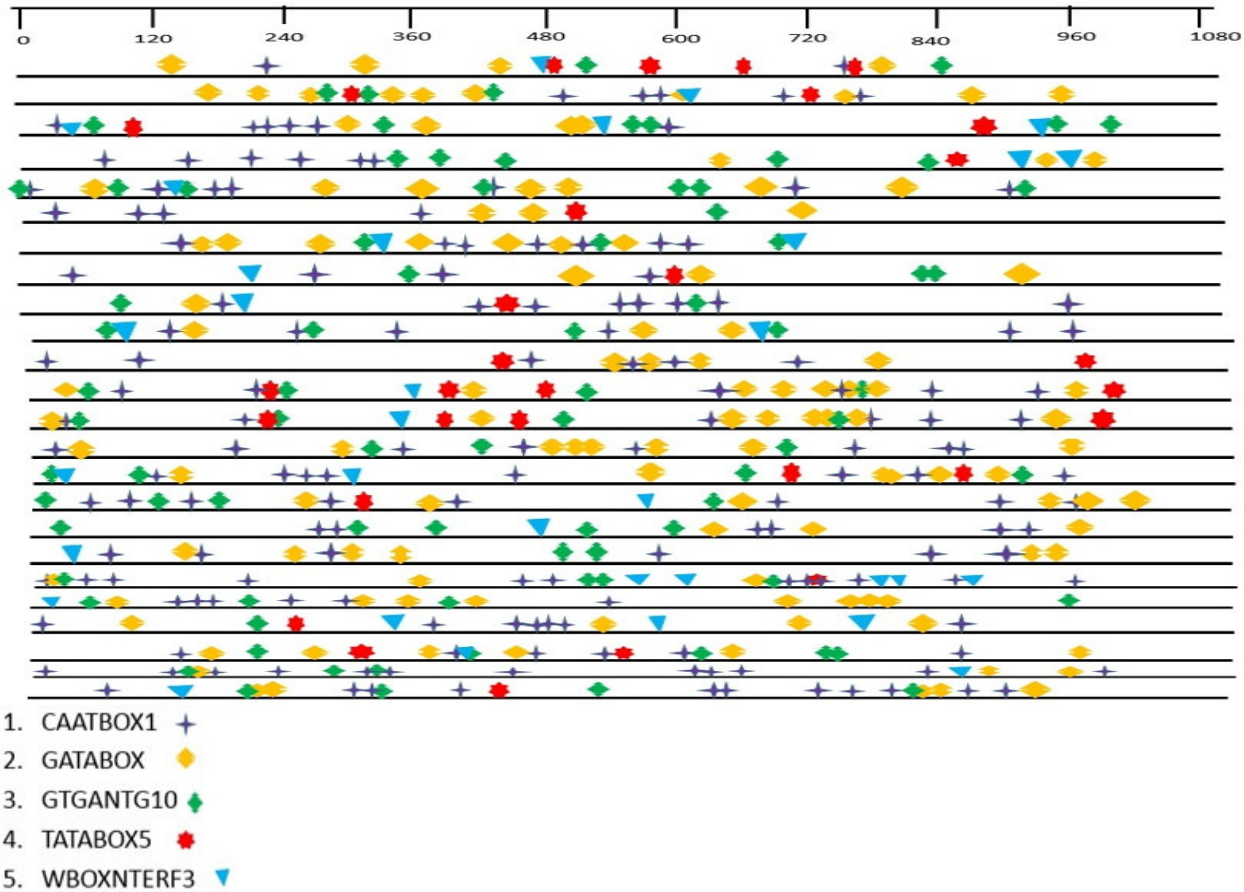

Figure 5: Schematic maps of various PLACE-based motifs in 1KB promoter region of heat stress related HSF genes. This figure has just first 5 promoters. The presence of motifs on +strand is shown in the figure. The motif codes and respective sequences are indicated below the figure.