



E-ISSN: 2278-4136

P-ISSN: 2349-8234

JPP 2018; 7(4): 1278-1285

Received: 01-05-2018

Accepted: 05-06-2018

**Shubhendra Singh Chauhan**  
Research Associate, Department  
of Molecular and Cellular  
Engineering, JIBB, Sam  
Higginbottom University of  
Agriculture, Technology and  
Sciences, Uttar Pradesh, India

**Shikha Mittal**  
Research Associate, Centre for  
Agricultural Bioinformatics,  
Indian Agricultural Statistics  
Research Institute, Pusa, New  
Delhi, India

**Pragati Misra**  
Assistant Professor, Department  
of Molecular and Cellular  
Engineering, JIBB, Sam  
Higginbottom University of  
Agriculture, Technology and  
Sciences, Allahabad, Uttar  
Pradesh, India

## Genome-wide identification and comparative analysis of HSF genes among Solanaceae members

Shubhendra Singh Chauhan, Shikha Mittal and Pragati Misra

### Abstract

HSF family is one of the important transcription factor family involved in heat stress and several other abiotic stresses like drought, cold etc. HSF regulates expression of HSPs and controls high-temperature stress, damage, and biological process. In this study, we identified the HSF genes in different members of Solanaceae family including potato, tomato, brinjal and capsicum. We found and compared their gene structure, evolutionary relationship, motif, domains, duplication events and their time of divergence with the help of different software like MEGA, BLAST, GSDS, PAL2NAL, iTOL, and Circos. In potato, tomato, capsicum and brinjal 27, 26, 25 and 23 genes were identified from the Plant Transcription Factor database. In terms of Gene structure of potato as well as tomato, most of the genes retained only 1 intron except 2 genes i.e., one in potato and one in tomato suggesting the gain of intron during the evolutionary process. Comparative phylogenetic analysis of HSF genes revealed 2 major groups with several orthologous and paralogous genes. Time of divergence was also examined to understand evolutionary concepts. Segmental duplications were found to be involved in family expansion and evolution. Our comparative study may be helpful in the understanding of HSF genes in Solanaceae family and evolutionary pathways of HSFs in potato and tomato.

**Keywords:** potato, tomato, synteny, transcription factor

### Introduction

Plants have different defensive mechanisms to protect themselves from different biotic and abiotic stresses including salt, drought, high temperature etc. [1]. High temperature severely affects the plant growth and development resulting in reduction of crop yield [2-4]. The yield of the plants can be enhanced by tolerance to different abiotic stresses. To survive in high temperature there must be a heat shock system available in all species in which different proteins participate [5]. Tolerance against different abiotic stresses is directed by plant's ability to express a set of genes whose expression is often regulated by specific transcription factors (TFs). Among different transcription factor families, Heat shock transcription factor (HSF) is one of an important transcription factor family in plants responsible for high-temperature tolerance in plants [6-8]. In plants the size of HSF transcription factor family is large as compared to other organisms [9]. HSF regulates the expression of Heat shock proteins [10]. HSF's are present at the terminal end in signal transduction pathway and acts as transcriptional regulators which help in activation of genes of different abiotic stresses such as high temperature, drought etc [11]. Heat shock transcription factors binds with heat shock elements (HSE) to regulate the transcription of heat shock proteins (HSP). HSPs are involved in the protection of cells against stress impairment and also involved in folding of proteins [12-14]. The involvement of HSF family in high temperature stress as well as in other abiotic stresses in different crops has already been reported in previous studies [15-19].

As heat stress affects the yield of different important food crops including potato, tomato, capsicum and brinjal, so in this study we performed an *in-silico* analysis of the HSF gene family in potato and compared it with the tomato, capsicum and brinjal HSF genes. In this study, comparative analysis was performed among different species of Solanaceae family to have an idea about their expansion and evolutionary history, explore their heat stress responses as elicited by naturally increased temperature.

### Materials and Methods

#### Identification and characterization of HSF genes in Solanaceae family

The HSF family genes in *Solanum tuberosum*, *Solanum lycopersicum*, *Capsicum annum* and *Solanum melongena* were searched and obtained from Plant Transcription Factor Database [20]. The protein sequences of all the Solanaceae family members were retrieved using Phytozome Biomart (<http://www.phytozome.net>). In total, we found 27, 26, 25 and 23

#### Corresponding

**Shubhendra Singh Chauhan**  
Research Associate, Department  
of Molecular and Cellular  
Engineering, JIBB, Sam  
Higginbottom University of  
Agriculture, Technology and  
Sciences, Uttar Pradesh, India

members in potato, tomato, capsicum and brinjal respectively belonging to HSF transcription factor family. All of these significant genes belonging to different species were selected after domain search using pfam database (<http://pfam.xfam.org/>) with expectation cut-off value of 1.0.

### Physiochemical Properties

To have an idea about the nature of HSF proteins in different species of Solanaceae family, grand average of hydropathy (GRAVY), isoelectronic point (pI), Molecular weight and instability index were predicted by ProtParam tool available on Expert Protein Analysis System (ExPASy, [http://web.expasy.org/compute\\_pi](http://web.expasy.org/compute_pi)) [21]. The instability index depicts the stability of the protein. If the value of instability index is less than 40 then the protein is stable in nature while if the value is greater than 40 then it indicates unstable nature of protein. The GRAVY score indicates polar or non-polar nature of the protein. Positive GRAVY score infers hydrophobic nature of protein while negative GRAVY score suggests hydrophilic nature.

### Chromosomal distribution and duplication events analysis

Orthologs for all 4 species i.e., *S. tuberosum*, *S. lycopersicum*, *C. annuum* and *S. melongena* were identified using BLAST search. Tandem and segmental duplication events were also identified for *S. tuberosum* species using BLAST. Synonymous and non-synonymous substitutions between orthologs and paralogs were also calculated using PAL2NAL server while divergence time was calculated using  $T = Ks/2\pi$  where  $\pi = 6.5 \times 10^{-9}$ .

### Gene structure and motif prediction

To analyse intron exon structure of potato and tomato HSF genes, coding and genomic sequences of all the genes were retrieved using Ensembl genomes. Capsicum and brinjal were not used for gene structure analysis as their genomic sequences were not available. Gene Structure Display Server (GSDS server, <http://gsds.cbi.pku.edu.cn>) [22] server was used for gene structure illustration. To find the best 5 motifs among *S. tuberosum*, *S. lycopersicum*, *C. annuum* and *S. melongena*, Multiple maximization for Motif Elicitation analysis tool (MEME, <http://meme.sdsc.edu>, v4.9.0) [23] program was used. The MEME suite was analysed with the following parameters, maximum number of motifs=5, motif width=12 to 60 and E-value < 0.01.

### Phylogenetic analysis

To generate the phylogenetic trees HSF transcription factor family genes, multiple sequence alignment of the genes was done using ClustalW. Phylogenetic tree was built by using Neighbour-Joining method with 1000 rapid bootstrap replicates with the help of MEGA v6.06 [24]. In neighbour joining analysis, pairwise deletion and poisson correction options were used. The developed phylogenetic tree was visualized using iTOL (<http://itol.embl.de/>) software.

### Synteny analysis

The synteny between *Solanum tuberosum* and *Solanum lycopersicum* was identified using BLAST search. A cut-off bit score of 75 and E-value less than  $1e-05$  were considered as optimum for BLAST analysis. The synteny among potato and tomato was visualised using Circos v0.63 software. (<http://circos.ca>) [25].

## Result and Discussion

### Identification and physiochemical properties of HSF gene family

Genome wide analysis led to the identification of 27, 24, 25 and 23 significant HSF genes in potato, tomato, capsicum and brinjal. In former studies, 21, 25, 28, 16, 25 HSF genes were found in Arabidopsis, rice, populus, Medicago and maize [15, 26-28]. The physiochemical nature of these proteins was studied to understand the nature of proteins. Out of 4 species, highest molecular weight was identified in tomato protein i.e., SIHSF14 while minimum in brinjal species (SmeHSF21). Positive hydropathy score indicates insoluble nature of proteins while negative GRAVY score suggests soluble nature of protein [29]. According to our study, all the proteins of Solanaceae family belonging to HSF transcription factor have negative hydropathy score suggesting soluble nature of proteins. Isoelectronic point (pI) is the point at which net charge of the protein is zero and the protein is least soluble at this point. In terms of pI, among 4 species i.e., *S. tuberosum*, *S. lycopersicum*, *C. annuum* and *S. melongena*, minimum pI was found in SmeHSF6 and maximum in SmeHSF7. In previous studies also, wide range of isoelectronic point has been noticed in cucumber, soybean etc. [5, 30]. Protein pI is having an important role in finding out the pH dependent characteristics of a protein [31]. To analyse whether the protein is stable or unstable in nature, instability index was calculated. All the proteins were found to be unstable in nature except 8 proteins (StHSF19, SIHSF11, SIHSF5, CaHSF11, CaHSF, SmeHSF21, SmeHSF1 and SmeHSF23). Instability index greater than 40 represents unstable nature of proteins while instability index less than 40 suggests stable nature of proteins [32] (Table 1).

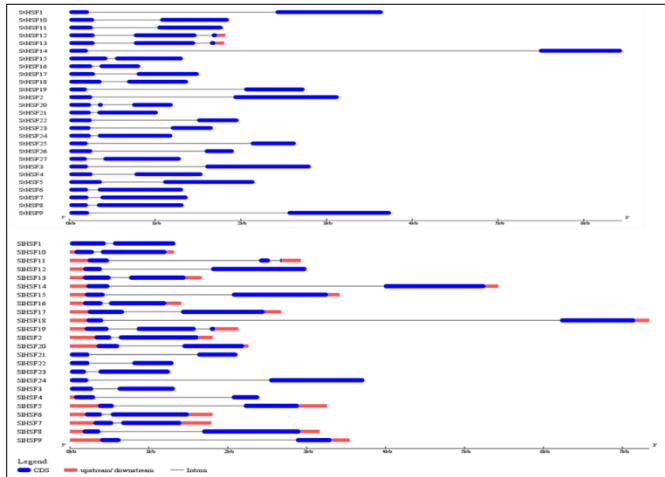
### Chromosomal distribution of HSF genes

To examine the association between genetic divergence within HSF gene family and duplication in potato and tomato, HSF genes were mapped on their chromosomes in potato and tomato species. The distribution of HSF genes was not uniform in potato as well as in tomato. Some chromosome and chromosomal regions have high density of genes while some regions were devoid of genes. The HSF family genes were unevenly mapped on 10 chromosomes of potato as well as tomato. Chromosome 5 and chromosome 1 lack the HSF gene in tomato as well as potato. Most of the HSF genes lie only on the short arm of the chromosome. In potato, chromosome 9 has highest number of genes i.e., 5 while in tomato maximum number of genes were found on chromosome 2 i.e., 5. In various previous studies it has been observed that the HSF genes were distributed unevenly on their chromosomes. In soybean also, the HSF genes were unevenly distributed on 15 chromosomes out of 20 chromosomes [5]. Gene duplication is a process in which new genes are formed which disperses in the genome [33]. In *Solanum tuberosum*, we found only 3 tandem duplication and 1 segmental duplication while in case of *Solanum lycopersicum* only 2 segmental duplication were obtained but no tandem duplication were identified.

### Gene structure

The exon and intron structure of the HSF genes of potato and tomato was analyzed to understand the structural components of the HSF genes. It was observed that most of the HSF genes in tomato and potato retain only one intron. In potato, out of 27 genes, 26 genes had only one intron while one gene (StHSF20) had 2 introns. However, in case of tomato also all

the genes except SIHSF11 (2 intron) were having only one intron. Genes having 2 introns represent gain of introns during evolutionary process. It was depicted from previous studies of cucumber and soybean that mostly HSF genes have only 1 intron (Figure 1). In soybean gene structure analysis, it was noticed that soybean HSF genes had one intron only except one gene [5] while in cucumber 2 genes had 2 introns and one gene had 3 introns [30], and in chickpea only three genes had 2 introns [34].



**Fig 1:** Gene structure of *Solanum tuberosum* and *Solanum lycopersicum* species identified using GSDS server

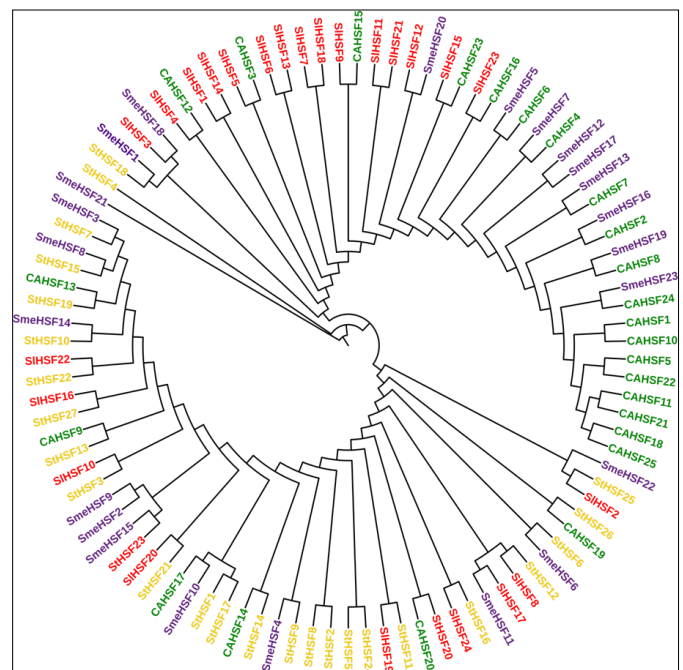
### Phylogenetic analysis

To analyze the evolutionary relationship among potato, tomato, capsicum and brinjal HSF genes, a phylogenetic tree was generated using MEGA with 1000 replicates of bootstrapping. In our phylogenetic tree, two major groups of HSF genes were found with several paralogous as well as orthologous genes. All the potato genes were distributed in one group only except two genes, StHSF4 and StHSF18. While rest of the species were evenly distributed in both the groups. Each group possesses potato as well as tomato genes (Figure 2). In previous study of chickpea, 4 groups were found in phylogenetic tree which is same as our results [34]. Similarly in arabidopsis and rice, 3 groups were found in the phylogenetic tree of HSF genes [15]. Development of phylogenetic trees would serve as the foundation which will help us to better know the evolution of cellular pathways, macromolecular machines and other emergent properties of early life [35].

Time of divergence was identified for both paralogous as well as orthologous gene pairs. While in case of potato, time of divergence of tandem and segmental duplications was also

observed. Divergence time was estimated by examining ratio of synonymous (Ks) and non-synonymous (Ka) substitution. Among duplication events, we found 9 segmentally duplicated pairs while 3 tandem duplicated pairs in potato. More duplication events may be the reason for family expansion in potato [36]. In case of tandem duplicated events, Ka/Ks varied from 0.01 to 0.39 with a mean value of 0.14 while in segmentally duplicated events, Ka/Ks varied from 0.01 to 0.24 with a mean value of 0.09. Both tandem and segmentally duplicated pairs in potato were found to be diverged 1843.91 and 1850.72 million years ago (mya).

In case of ortholog pairs of potato-tomato, potato-brinjal and potato-capsicum, Ka/Ks ranged from 0.02 to 0.67, 0.01 to 0.62 and 0.08 to 0.52 with an average value of 0.25, 0.23 and 0.27 respectively. Since Ka/Ks ratio for all the ortholog pairs were found to be less than 1, it is assumed that they had undergone natural selection [37]. Potato-tomato have been estimated to occur about 148.82 mya, Potato-brinjal about 589.03 mya whereas Potato-capsicum about 49.95mya (Supplementary Table 1). Among all these species of Solanaceae family, potato-brinjal were diverged earliest while potato-capsicum later.



**Fig 2:** Phylogenetic tree of *Solanum tuberosum*, *Solanum lycopersicum*, *Capsicum annum* and *Solanum melongena* generated using MEGA software with 1000 bootstrapping replicates. Yellow, Red, green and purple color represents *S. tuberosum*, *S. lycopersicum*, *C. annum* and *S. melongena* species

**Table 1:** List of HSF genes identified in Potato, Tomato, Capsicum and Brinjal from Plant Transcription Factor Database and their physicochemical properties.

Species	Gene Name	Gene ID	Chr	Length	MW	pI	GRAVY	Instability index
Potato ( <i>Solanum tuberosum</i> )	StHSF1	PGSC0003DMG400003160	3	492	54318.75	4.85	-0.44	59.56
	StHSF2	PGSC0003DMG400014811	8	498	55721.29	5.01	-0.671	63.7
	StHSF3	PGSC0003DMG400027036	6	481	54422.23	5.17	-0.578	54.79
	StHSF4	PGSC0003DMG400008223	8	353	40709.7	4.92	-0.659	56.69
	StHSF5	PGSC0003DMG401002683	9	480	52913.17	4.81	-0.45	55.73
	StHSF6	PGSC0003DMG401027812	3	403	46213.59	5.16	-0.77	44.3
	StHSF7	PGSC0003DMG400017334	7	412	47018.15	5.36	-0.784	54.08
	StHSF8	PGSC0003DMG400028414	2	408	46220.42	5.23	-0.764	46.07
	StHSF9	PGSC0003DMG400004662	12	478	53425.19	5.48	-0.699	57.97
	StHSF10	PGSC0003DMG400006447	9	362	42263.33	5.46	-0.931	45.46
	StHSF11	PGSC0003DMG400016270	6	344	39764.05	5.03	-0.845	58.27

	StHSF12	PGSC0003DMG402019343	9	364	42034.19	5.53	-0.784	54.59
	StHSF13	PGSC0003DMG401019343	9	360	41480.55	5.3	-0.716	48.68
	StHSF14	PGSC0003DMG400017484	9	394	45539.28	4.71	-0.593	56.07
	StHSF15	PGSC0003DMG400041361	2	411	46954.43	4.92	-0.809	47.6
	StHSF16	PGSC0003DMG400043234	11	246	28707.88	9.3	-0.651	47.63
	StHSF17	PGSC0003DMG400028416	2	340	38450.02	6.59	-0.909	48.34
	StHSF18	PGSC0003DMG400032793	7	361	41761.13	5.28	-0.825	56.45
	StHSF19	PGSC0003DMG401004023	2	302	33288.23	5.75	-0.754	35.72
	StHSF20	PGSC0003DMG400014323	3	260	29405.34	4.78	-0.893	60.4
	StHSF21	PGSC0003DMG400003053	8	317	35023.95	4.93	-0.472	54.35
	StHSF22	PGSC0003DMG400027283	4	247	28695.62	8.68	-0.754	60.24
	StHSF23	PGSC0003DMG401008167	10	244	28376.07	6.13	-0.857	48.05
	StHSF24	PGSC0003DMG400007962	4	372	42367.2	7.76	-0.714	52.8
	StHSF25	PGSC0003DMG400034428	11	245	29388.51	7.31	-0.689	48.32
	StHSF26	PGSC0003DMG400029718	2	201	23853.47	9.58	-0.845	49.15
	StHSF27	PGSC0003DMG400000380	12	368	40974.02	5.86	-0.614	66.82
<b>Tomato (<i>Solanum lycopersicum</i>)</b>	SIHSF1	Solyc02g079180.1	2	415	47140.6	4.950	-0.77	45.090
	SIHSF2	Solyc02g072000.2	2	409	46168.5	5.23	-0.726	50.950
	SIHSF3	Solyc02g072060.1	2	340	38413.1	7.63	-0.886	44.570
	SIHSF4	Solyc02g078340.2	2	206	24239.7	9.390	-0.873	46.080
	SIHSF5	Solyc02g090820.2	2	302	33257.2	5.750	-0.746	37.030
	SIHSF6	Solyc03g006000.2	3	402	46080.3	5.330	-0.798	44.550
	SIHSF7	Solyc03g026020.2	3	339	37387.6	5.340	-0.74	68.070
	SIHSF8	Solyc03g097120.2	3	492	54029.6	5.070	-0.408	55.160
	SIHSF9	Solyc04g016000.2	4	238	27566.5	8.87	-0.701	56.270
	SIHSF10	Solyc04g078770.2	4	361	40928.7	7.720	-0.65	54.990
	SIHSF11	Solyc06g053960.2	6	143	16643.6	9.440	-0.796	31.890
	SIHSF12	Solyc06g072750.2	6	483	54336	5.320	-0.615	53.810
	SIHSF13	Solyc07g040680.2	7	357	40871.2	5.330	-0.758	57.330
	SIHSF14	Solyc08g005170.2	8	528	57700.7	5.160	-0.569	57.670
	SIHSF15	Solyc08g076590.2	8	491	55016.7	5.160	-0.672	61.980
	SIHSF16	Solyc08g080540.2	8	326	35289.3	4.990	-0.433	52.330
	SIHSF17	Solyc09g009100.2	9	510	56098.4	4.82	-0.546	54.900
	SIHSF18	Solyc09g059520.2	9	390	44921.4	4.680	-0.619	52.430
	SIHSF19	Solyc09g065660.2	9	373	42442.4	5.280	-0.823	58.960
	SIHSF20	Solyc09g082670.2	9	357	41757.6	5.38	-0.963	43.540
	SIHSF21	Solyc11g064990.1	11	252	29828.8	7.290	-0.743	51.990
	SIHSF22	Solyc10g079380.1	10	256	29916.8	7.590	-0.868	50.780
	SIHSF23	Solyc12g007070.1	12	370	41173.2	6.050	-0.627	65.110
	SIHSF24	Solyc12g098520.1	12	479	53375.1	5.400	-0.688	58.510
<b>Capsicum (<i>Capsicum annuum</i>)</b>	CAHSF1	CA00g45390	-	258	30088.35	9.23	-0.763	43.78
	CAHSF2	CA00g63000	-	368	42587.3	5.52	-0.792	46.75
	CAHSF3	CA00g71530	-	326	-	-	-0.501	50.64
	CAHSF4	CA01g03100	-	281	-	-	-0.531	49.18
	CAHSF5	CA01g07540	-	496	55081.74	4.78	-0.616	67.81
	CAHSF6	CA01g30350	-	473	52892.99	4.91	-0.703	62.18
	CAHSF7	CA02g11030	-	304	33695.67	5.1	-0.722	34.69
	CAHSF8	CA02g13280	-	404	45744.76	5.65	-0.789	44.88
	CAHSF9	CA02g13320	-	328	37046.62	6.58	-0.825	45.94
	CAHSF10	CA02g16000	-	201	23373.88	9.15	-0.667	66.69
	CAHSF11	CA02g16840	-	380	43535.02	5.02	-0.62	34.64
	CAHSF12	CA03g06850	-	362	42296.21	5.28	-0.954	43.4
	CAHSF13	CA03g11650	-	359	41131.07	5.26	-0.817	55.26
	CAHSF14	CA03g16300	-	326	36370.63	5.09	-0.648	55.44
	CAHSF15	CA03g21660	-	505	56063.87	5.15	-0.529	54.9
	CAHSF16	CA04g01070	-	402	45810.7	5.21	-0.847	47.71
	CAHSF17	CA04g18550	-	379	42921.82	7.34	-0.7	52.88
	CAHSF18	CA05g00840	-	234	27416.14	9.19	-0.842	53.44
	CAHSF19	CA06g08710	-	335	38634.04	4.8	-0.79	53.66
	CAHSF20	CA07g15920	-	432	49296.72	5.3	-0.752	55.96
	CAHSF21	CA08g05000	-	362	41187.23	4.78	-0.573	55.72
	CAHSF22	CA09g01450	-	449	50282.17	4.85	-0.567	48.02
	CAHSF23	CA09g11190	-	401	46107.26	4.83	-0.537	49.09
	CAHSF24	CA10g20440	-	244	28281.86	8.22	-0.906	57.11
	CAHSF25	CA12g20590	-	453	50924.12	5.5	-0.809	55.3
<b>Brinjal (<i>Solanum melongena</i>)</b>	SmeHSF1	Sme2.5_00010.1_g00004.1	-	481	53575.07	5.24	-0.673	30.35
	SmeHSF2	Sme2.5_00023.1_g00025.1	-	410	47368.93	5.29	-0.547	50.08
	SmeHSF3	Sme2.5_00065.1_g00020.1	-	357	41699.38	5.15	-0.929	45.44
	SmeHSF4	Sme2.5_00159.1_g00006.1	-	213	24620.01	9.44	-0.791	52.04

SmeHSF5	Sme2.5_00204.1_g00007.1	-	478	53808.7	5.88	-0.614	51.96
SmeHSF6	Sme2.5_00292.1_g00007.1	-	494	55078.34	4.6	-0.58	53.86
SmeHSF7	Sme2.5_00579.1_g00002.1	-	335	37998.48	9.64	-0.506	46.41
SmeHSF8	Sme2.5_01013.1_g00005.1	-	403	46036.2	5.15	-0.78	42.59
SmeHSF9	Sme2.5_01029.1_g00008.1	-	357	40642.61	7.73	-0.604	60.84
SmeHSF10	Sme2.5_01314.1_g00005.1	-	421	48362.75	5.35	-0.74	54.15
SmeHSF11	Sme2.5_02334.1_g00004.1	-	496	54837.58	5.13	-0.482	52.58
SmeHSF12	Sme2.5_02712.1_g00007.1	-	331	36210.51	5.18	-0.437	55.45
SmeHSF13	Sme2.5_03412.1_g00012.1	-	317	36775.39	8.99	-0.743	41.31
SmeHSF14	Sme2.5_04149.1_g00004.1	-	343	39841.03	5.99	-0.793	59.73
SmeHSF15	Sme2.5_04312.1_g00005.1	-	340	38481.09	5.93	-0.856	48.87
SmeHSF16	Sme2.5_04312.1_g00009.1	-	377	43012.76	5.14	-0.76	47.45
SmeHSF17	Sme2.5_04829.1_g00004.1	-	352	39421.21	6.14	-0.649	68.54
SmeHSF18	Sme2.5_08000.1_g00008.1	-	324	37931.68	5.5	-0.827	47.19
SmeHSF19	Sme2.5_08951.1_g00003.1	-	374	43271.78	4.71	-0.62	58.33
SmeHSF20	Sme2.5_09846.1_g00002.1	-	475	53307.86	5.51	-0.777	57.57
SmeHSF21	Sme2.5_10740.1_g00006.1	-	111	12190.37	7.81	-0.007	25.68
SmeHSF22	Sme2.5_13301.1_g00001.1	-	341	38297.33	6.35	-0.624	60.33
SmeHSF23	Sme2.5_31683.1_g00001.1	-	167	19252.35	5.89	-1.141	36.49

### Motif Prediction

To predict the amino acid motifs conserved among HSF genes of Solanaceae family, we used MEME software. 5 different types of motifs were identified (Table 2). Among the 5 identified motifs, motif 1 and motif 2 were found to be conserved in equal number of proteins i.e., 97 while motif 4 was least conserved. In most of the proteins belonging to Solanaceae family, motif 1 and motif 3 were found to be conserved on N-terminal, motif 4 and motif 5 on C-terminal while motif 2 in central position. Among all the amino acids, highly frequent amino acid was serine followed by Glutamic acid and leucine. Whereas least frequent amino acid observed was cysteine. Previous study on soybean also depicted 5 conserved motifs among Soybean HSF proteins [5].

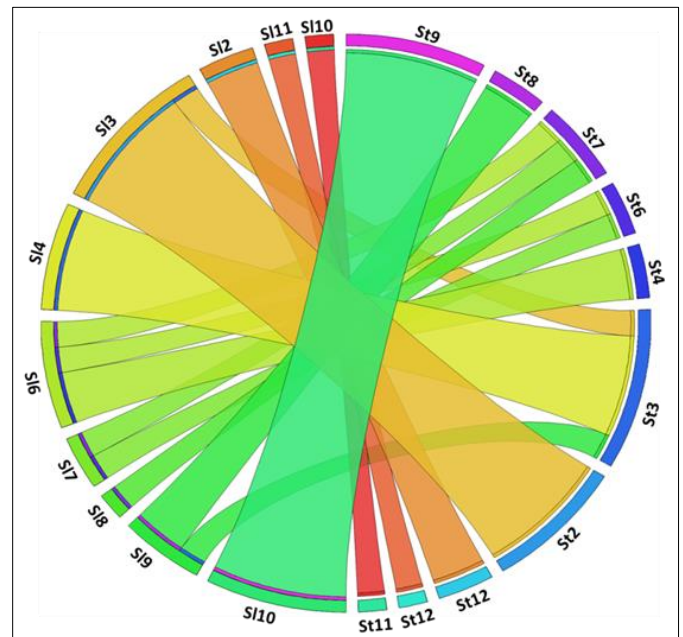
**Table 2:** List of 5 best identified motifs in HSF genes of Solanaceae family using MEME software.

Motif	Width	Best Possible Match
1	43	IVSWNRDGNFSFVWDPPEFARDLLPKYF KHNNFSSFVRQLNTY
2	16	FRKIDPDRWEFANEWF
3	22	HGNGPPPFLTKTYEMVDDPSTD
4	43	LMMELVKLRQHQQATDQMQMTMERL QAMEQRQQMMSFLAKA
5	22	LRGQKHLLCNHRRKPVHSHCH

### Synteny analysis

The synteny analysis in potato and tomato was conducted to determine whether this information might provide more functional insight. In this analysis, brinjal and capsicum were not considered because of the unavailability of their chromosomal information. To find synteny between the potato and tomato genes, potato HSF sequences were used as query against the tomato database while performing BLASTN for identification of orthologs with more than 80% similarity and E-value less than 1e-05. To confirm the potential orthologs, reciprocal BLAST was performed. A total of 16 genes of tomato showed syntenic relationship with potato HSF genes. *Solanum tuberosum* shows maximum synteny with chromosome 9 of *Solanum lycopersicum* followed by chromosome 2 and chromosome 3 (Figure 3). The comparative study of potato and tomato genomes resulted into syntenic blocks that reveal conserved features [38]. This synteny analysis revealed the evolutionary and functional links between genes in potato and tomato.

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.



**Fig 3:** Representation of syntenic relationship among different chromosomes of *Solanum tuberosum* and *Solanum lycopersicum*

### Conclusion

This study is structural characterization of the HSF genes in different members of Solanaceae family. Gene structure distribution of HSF genes revealed that the most of the genes have only one intron except few genes. Segmental duplications were found to be prominent in expansion of gene family as compared to tandem duplications. Potato-capsicum was found to be less divergent as compared to potato-tomato and potato-brinjal. 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. These genes will be helpful for production of heat tolerant potato varieties.

**Supplementary Table 1(A):** Divergence time calculated for segmentally duplicated orthologous gene pairs in potato.

Gene ID 1	Gene ID 2	dS	dN	dN/dS	Time of divergence (mya)
StHSF22	StHSF23	0.9169	0.1409	0.1537	70.53
StHSF25	StHSF9	53.5821	0.6919	0.0129	4121.7
StHSF8	StHSF24	34.5718	0.8019	0.0232	2659.37
StHSF6	StHSF8	0.6203	0.1486	0.2395	47.72
StHSF1	StHSF24	57.04	0.6988	0.0123	4387.69
StHSF25	StHSF8	57.0397	0.6989	0.0123	4387.67
StHSF19	StHSF20	4.2902	0.5811	0.1354	330.02
StHSF20	StHSF21	3.2839	0.3504	0.1067	252.61
StHSF18	StHSF22	5.1888	0.8219	0.1584	399.14
	Average	24.06	0.55	0.09	1850.72

**Supplementary Table 1(B):** Divergence time calculated for tandem duplicated pairs in potato.

Gene ID 1	Gene ID 2	dS	dN	dN/dS	Time of divergence (mya)
StHSF13	StHSF12	0.05	0.02	0.39	3.68
StHSF26	StHSF15	56.44	0.69	0.01	4341.21
StHSF24	StHSF22	15.43	0.52	0.03	1186.85
	Average	23.97	0.41	0.14	1843.91

**Supplementary Table 1(C):** Divergence time calculated for orthologous pairs between potato and tomato

Gene ID 1	Gene ID 2	dS	dN	dN/dS	Time of divergence (mya)
SIHSF1	StHSF15	0.0934	0.0196	0.2104	7.18
SIHSF10	StHSF24	0.1938	0.0305	0.1571	14.91
SIHSF11	StHSF11	0.1555	0.0583	0.375	11.96
SIHSF12	StHSF3	0.0597	0.02	0.3356	4.59
SIHSF13	StHSF18	0.1891	0.0275	0.1453	14.55
SIHSF14	StHSF7	4.8515	0.6275	0.1293	373.19
SIHSF15	StHSF2	0.1	0.0459	0.459	7.69
SIHSF16	StHSF20	2.5159	0.3503	0.1392	193.53
SIHSF16	StHSF21	0.1684	0.0555	0.3299	12.95
SIHSF17	StHSF5	0.0866	0.0332	0.383	6.66
SIHSF18	StHSF14	0.0903	0.012	0.1325	6.95
SIHSF19	StHSF12	0.1692	0.0324	0.1917	13.02
SIHSF19	StHSF13	0.1398	0.0358	0.2561	10.75
SIHSF2	StHSF8	0.1416	0.0171	0.1209	10.89
SIHSF20	StHSF10	0.0795	0.0176	0.221	6.12
SIHSF21	StHSF25	0.0548	0.0365	0.6654	4.22
SIHSF22	StHSF23	0.071	0.0291	0.4097	5.46
SIHSF23	StHSF27	0.1834	0.0156	0.0852	14.11
SIHSF24	StHSF9	0.0499	0.0145	0.2898	3.84
SIHSF3	StHSF17	0.0659	0.0179	0.2708	5.07
SIHSF4	StHSF26	0.0919	0.0229	0.2488	7.07
SIHSF5	StHSF19	0.0985	0.0106	0.1075	7.58
SIHSF5	StHSF20	3.7211	0.5652	0.1519	286.24
SIHSF6	StHSF6	0.0652	0.0125	0.1916	5.02
SIHSF7	StHSF20	0.1479	0.059	0.3991	11.38
SIHSF8	StHSF1	0.0995	0.0303	0.3047	7.65
SIHSF9	StHSF22	0.1019	0.0196	0.1928	7.84
SIHSF9	StHSF18	40.3869	0.7565	0.0187	3106.68
	Average	1.93	0.11	0.25	148.82

**Supplementary Table 1(D):** Divergence time calculated for orthologous pairs between potato and brinjal

Gene ID 1	Gene ID 2	dS	dN	dN/dS	Time of divergence (mya)
SmeHSF1	StHSF19	0.29	0.04	0.15	22.1
SmeHSF10	StHSF7	0.34	0.07	0.22	26.12
SmeHSF11	StHSF1	0.17	0.06	0.38	13.05
SmeHSF12	StHSF20	2.32	0.34	0.15	178.78
SmeHSF12	StHSF21	0.37	0.07	0.2	28.45
SmeHSF13	StHSF16	0.23	0.07	0.32	17.45
SmeHSF14	StHSF13	6.86	0.76	0.11	527.44
SmeHSF14	StHSF12	0.34	0.07	0.21	26.34
SmeHSF15	StHSF17	0.18	0.04	0.21	13.46
SmeHSF16	StHSF8	0.15	0.05	0.35	11.68
SmeHSF17	StHSF27	0.28	0.04	0.13	21.43
SmeHSF19	StHSF14	0.15	0.09	0.62	11.52
SmeHSF2	StHSF18	0.58	0.07	0.12	44.98

SmeHSF20	StHSF9	0.17	0.03	0.19	13
SmeHSF21	StHSF2	0.68	0.23	0.34	52.5
SmeHSF22	StHSF20	0.45	0.15	0.33	34.45
SmeHSF23	StHSF5	6.56	0.77	0.12	504.72
SmeHSF23	StHSF8	0.17	0.06	0.35	13.38
SmeHSF23	StHSF6	0.59	0.17	0.29	45.05
SmeHSF3	StHSF10	0.11	0.06	0.49	8.77
SmeHSF4	StHSF18	24.2	0.8	0.03	1861.69
SmeHSF4	StHSF22	0.28	0.09	0.32	21.45
SmeHSF5	StHSF3	0.14	0.06	0.44	11.13
SmeHSF6	StHSF5	0.17	0.07	0.4	12.9
SmeHSF7	StHSF25	60.46	0.75	0.01	4651.13
SmeHSF7	StHSF1	38.11	0.72	0.02	2931.32
SmeHSF7	StHSF24	6.07	0.9	0.15	467.2
SmeHSF7	StHSF3	58.76	0.66	0.01	4520.22
SmeHSF7	StHSF9	27.63	0.66	0.02	2125.44
SmeHSF8	StHSF6	0.15	0.05	0.3	11.62
SmeHSF9	StHSF24	0.41	0.03	0.08	31.28
	Average	7.66	0.26	0.23	589.03

**Supplementary Table 1(E):** Divergence time calculated for orthologous pairs between potato and capsicum

Gene ID 1	Gene ID 2	dS	dN	dN/dS	Time of divergence (mya)
CAHSF1	StHSF16	0.39	0.12	0.32	29.62
CAHSF10	StHSF26	0.24	0.1	0.43	18.52
CAHSF12	StHSF10	0.13	0.07	0.52	9.9
CAHSF13	StHSF13	0.35	0.08	0.22	27.14
CAHSF13	StHSF12	0.34	0.08	0.22	26.45
CAHSF14	StHSF8	6.48	0.76	0.12	498.53
CAHSF14	StHSF20	0.44	0.14	0.31	33.65
CAHSF15	StHSF1	0.19	0.08	0.4	14.88
CAHSF16	StHSF6	0.17	0.07	0.4	12.95
CAHSF17	StHSF24	0.7	0.05	0.08	53.81
CAHSF2	StHSF18	0.59	0.1	0.16	45.72
CAHSF20	StHSF7	0.58	0.07	0.12	44.74
CAHSF21	StHSF4	0.42	0.06	0.15	32.21
CAHSF23	StHSF14	0.2	0.08	0.38	15.61
CAHSF24	StHSF23	0.27	0.09	0.32	20.96
CAHSF25	StHSF9	0.24	0.03	0.11	18.75
CAHSF3	StHSF27	1.02	0.12	0.11	78.48
CAHSF5	StHSF2	0.24	0.11	0.48	18.25
CAHSF7	StHSF19	0.29	0.03	0.12	22.05
CAHSF8	StHSF8	0.18	0.06	0.32	13.48
CAHSF9	StHSF17	0.17	0.05	0.31	13.2
	Average	0.65	0.11	0.27	49.95

**References**

- Scharf KD *et al.* Biochimica et Biophysica Acta. 2012; 1819(2):104-19.
- Wardlaw IF and Wrigley CW. Australian Journal of Plant Physiology. 1994; 21(6):695-703.
- Skylas DJ *et al.* Journal of Cereal Science. 2002; 35(2):175-88.
- Bar-Tsur A *et al.* Journal of the American Society for Horticultural Sciences. 1985; 110:582-6.
- Li P S *et al.* BMC Genomics. 2014; 15:1009.
- Nover L *et al.* Cell Stress Chaperones. 2001; 6:177-189.
- Akerfelt M *et al.* Nature Reviews Molecular Cell Biology. 2010; 11:545-555.
- Hayashida N *et al.* Transcription. 2011; 2:91-94.
- Czarnecka E *et al.* Plant molecular biology. 2004; 56:57-75.
- Sorger PK Cell. 1991; 65:363-366.
- Baniwal SK *et al.* Journal of Biosciences. 2004; 29:471-487.
- Morimoto RI *et al.* Cold spring harbor monograph archive. 1994; 26:1-30.
- Schöffl F *et al.* Plant Physiology. 1998; 117:1135-1141.
- Hartl FU *et al.* Science. 2002; 295:1852-1858.
- Guo J *et al.* Journal of Genetics & Genomics. 2008; 35:105-118.
- Chauhan H *et al.* Molecular Genetics & Genomics. 2011; 286:171-187.
- Yoshida T *et al.* Molecular Genetics & Genomics. 2011; 286:321-332.
- Hahn A *et al.* The Plant Cell Online. 2011; 23:741-755
- Mishra S K *et al.* Genes & Development. 2002; 16:1555-1567.
- Rian ò-Pachón D M *et al.* BMC Bioinformatics. 2007; 8:42.
- Gasteiger, E. *Et al.* Nucleic acids research. 2003; 31(13): 3784-3788.
- Guo AY *et al.* Yi Chuan. 2007; 29:1023-1026.
- Bailey, T. L *et al.* Nucleic acids research. 2006; 34(2):W369-W373.
- Tamura K *et al.* Molecular Biology & Evolution. 2013; 30:2725-2729.
- Krzywinski *et al.* Genome research. 2009; 19(9):1639-1645.

26. Miller G and Mittler R *Annals of Botany*. 2006; 98:279–288.
27. Lin YX *et al.* *BMC Genomics*. 2011; 12:76.
28. Wang F *et al.* *Molecular Biology Reports*. 2012; 39:1877–1886.
29. Kyte J and Doolittle RF *Journal of Molecular Biology*. 1982; 157:105–132.
30. Zhou S *et al.* *Plant Omics*. 2013; 6: 449.
31. Talley K and Alexov E *Proteins* 2010; 78:2699–2706
32. Rogers HH *et al.* *Agriculture Ecosystems & Environment*. 1986; 16:113-128.
33. Lynch M and Conery JS *Science*. 2000; 290:1151–1155.
34. Zafar SA *et al.* *Plant Omics Journal*. 2016; 9:136-141.
35. Gaucher EA *et al.* *Cold Spring Harbor Perspectives in Biology*. 2010; 2(1):a002238.
36. Cannon SB *et al.* *BMC Plant Biology*. 2004 4:10.
37. Yang X *et al.* *Plant Physiology*. 2006; 142:820-830.
38. Ghiurcuta CG *et al.* *Bioinformatics*. 2014; 30:9-18.s