Bulletin of Biological and Allied Sciences Research
***ISSN:* 2521-0092**
www.bbasr.org
DOI: https://doi.org/10.64013/bbasr.v2026i1.135
*Bull. Biol. All. Sci. Res., Volume, **11**: 135*
Original Research Article ***Open Access***

GENOME-WIDE IDENTIFICATION AND CHARACTERIZATION OF HSF (HEAT SHOCK TRANSCRIPTION FACTOR) GENES IN COFFEE (*COFFEE ARABICA* L.) UNDER ABIOTIC STRESS

NAEEM S[1*], ALI Z[2], NAEEM Z[3], NAEEM S[4], SHAHBAZ F[2], HASSAN M[5], HAYYAT Q[2*]

[1]*Department of Agricultural Technology, Faculty of Agriculture, University Putra Malaysia*, 43400 UPM Serdang Selangor Darul Ehsan Malaysia
[2]*Department of Plant Breeding and Genetics Faculty of Agriculture Sciences, University of Punjab Lahore, Pakistan*
[3]*Department of Botany, Government College Women University, Sialkot, Pakistan*
[4]*Department of Biomedical Engineering, University of Engineering and Technology, Lahore, Pakistan*
[5]*Department of Entomology, Faculty of Agriculture Sciences, University of Punjab Lahore, Pakistan*
**Correspondence Author Email Address:* shajihanaeem78@gmail.com; *qaisarhayyat2@gmail.com*

(Received, 6th February 2026, Accepted 13th August 2026, Published 25th August 2026)

Abstract *Coffee is a livelihood source for millions of farmers and plays a significant role in the economy of many coffee-producing countries. Abiotic stresses, such as drought and high temperature, however, negatively impact coffee growth, productivity, and bean quality. Heat shock factors (HSFs) play a crucial role in the plant response to heat and other abiotic stresses. However, there is still not a lot of detailed knowledge about the HSF gene family in Coffea arabica. In this study, a genome-wide search was performed to find and characterize the genes of HSF from C.arabica. A total of 59 putative CaHSF genes have been identified and characterized according to their physicochemical properties, chromosomal distribution, gene structure, conserved motifs, phylogenetic relationship, duplication event, synteny, cis-regulatory elements, and protein–protein interactions. The genes identified for CaHSF were assigned to different scaffolds of the genome, and they had different sizes, molecular mass, isoelectric point, GRAVY value, and aliphatic index of their protein. Through the relationships induced by the phylogenetic and conserved-domain analyses, and by their conservation, it was possible to get insight into the evolution of the proteins of the CaHSF family. Abiotic stress response, hormone signaling, and light regulation were noted to have several cis-regulatory elements associated with them in the promoter analysis and hence the possible role of the CaHSF genes in stress adaptation. Duplication and synteny analyses also showed that gene duplication played a role in the diversification and expansion of the CaHSF gene family in C. arabica. Overall, this study is a thorough genomic analysis of the CaHSF gene family, which can serve as a basis for future functional studies for understanding the molecular mechanism of abiotic stress tolerance in coffee.*

[*Citation: Naeem, S., Ali, Z., Naeem, Z., Naeem, S., Shahbaz, F., Hassan, M., Hayyat, Q.* (2026). Genome-wide identification and characterization of HSF (heat shock transcription factor) genes in coffee (*Coffee arabica* L.) under abiotic stress. *Bull. Biol. All. Sci. Res.* **11***: 135.* https://doi.org/10.64013/bbasr.v2026i1.135]
Keywords*:* *Coffea arabica; Heat Shock Factors; HSF Gene Family; Abiotic Stress; Drought Stress; Thermotolerance Gene Expression*

Introduction

Abiotic stresses such as high and low temperatures, drought, and salinity continuously affect the survival rates, plant growth and development, productivity, and yield of plants(Li et al., 2025). Of these stresses, high temperature and water shortage are important environmental factors that can significantly impact productivity(Ding et al., 2026). Plants have developed sophisticated molecular and physiological pathways that allow them to sense and react to environmental stresses through transcription factors, which are stress-responsive, and protective genes. Transcription factors are important in regulating plant responses to abiotic stresses by regulating the expression of stress-responsive genes. A few families of transcription factors are known to be involved in plant stress responses, such as Heat Shock Factors (HSFs), Dehydration-Responsive Element-Binding proteins (DREBs), NAC proteins, and Multiprotein Bridging Factor proteins (Ali et al., 2026; Hayyat et al., 2026; Mustafidah And Su-Udi, 2025; Xu et al., 2020). The latter include the heat shock proteins (HSPs), which are important regulators of heat stress response through their control over the transcription of other genes that respond to stress. HSFs share several conserved structural domains, such as a DNA-binding domain (DBD), an oligomerization domain consisting of HR-A and HR-B regions, a nuclear localization signal (NLS), and a transcriptional activation domain. On the basis of the structure of the oligomerization region and other conserved properties, plant HSFs can be divided into three major groups: HSFA, HSFB, and HSFC. Within these groups, HSFA members play a special role in the control of heat stress responses. HSFs may also be involved in the response of plants

to other abiotic stresses such as drought, salinity, and cold stress(Hahn et al., 2004).

The HSF regulatory network is crucial to the regulation of cellular homeostasis under stress conditions. HSFs control the expression of heat shock proteins and other protective proteins involved in protein folding, stabilization, and recovery from stress-induced damage to the cell. HSFs can specifically modulate stress signal transduction and could play a role in the enhancement of stress tolerance in plants. *Coffea arabica* is an important economic species of coffee from the Rubiaceae family(Guo Meng et al., 2016; Ramalho et al., 2025). An allotetraploid (2n = 4x = 44) species that is thought to have evolved as a hybrid of the diploid species Coffea canephora and Coffea eugenioides. Coffee is highly sensitive to the environment, and abiotic stresses like drought and high temperatures can have negative impacts on plant growth, development, productivity, and quality of the coffee crop. With the growing availability of genome and transcriptome resources, the gene family associated with stress responses in economically important crops has been investigated. Genome-wide identification and characterization of stress-responsive gene families can provide valuable information regarding their evolutionary relationships, structural characteristics, regulatory elements, and potential biological functions. Although abiotic stress tolerance is one of the most important traits that must be developed in coffee, little information about the HSF gene family is available in *C. arabica*. So, the present study aimed to recognize and characterize the *C. arabica* genome HSF gene family(Guo et al., 2016). The study focused on the identification of CaHSF genes and the physicochemical analysis, the phylogenetic relationships, the gene structure, conserved motifs, the chromosomal distribution, the duplication events, the synteny, the cis-regulatory elements, and the predicted protein–protein interactions of these genes. The results offer a genomic context to understand the potential function and evolution of HSF genes in *C. arabica* in response to abiotic stress(Yu et al., 2022).

Materials and Methods

Sequence retrieval

The genome and annotation data for *Coffea arabica* were downloaded from the Phytozome database (https://phytozome-next.jgi.doe.gov/). The sequences of HSF proteins in Arabidopsis thaliana were extracted from the NCBI database and used as references. The conserved HSF domain (PF00447) was used for the identification of putative HSF proteins in *C. arabica*. The candidate sequences were then checked for the presence of the characteristic HSF domain. Forty-nine potential CaHSF protein sequences were then identified(Buttar et al., 2020).

Physio-chemical and Subcellular Analysis of CCO gene

ProtParam tool of the ExPASy server (https://web.expasy.org/protparam/) was used to analyze the physicochemical properties of the identified CaHSF proteins. For each CaHSF protein, the protein length, molecular weight, theoretical isoelectric point (pI), grand average of hydropathicity (GRAVY), and aliphatic index were calculated. Gene identifiers, chromosomal/scaffold locations, and corresponding protein sequences were obtained from the Phytozome database. WoLF PSORT (https://wolfpsort.hgc.jp/) was used to predict subcellular localization of CaHSF proteins(Gain et al., 2024).

Analysis of Gene Structure, Cis-Regulatory Elements and Motifs

The exon–intron structure of the CaHSF genes was studied by a Gene Structure Display Server (GSDS v2.0), available online at http://gsds.cbi.pku.edu.cn/. The conserved motifs in the CaHSF proteins were found by the MEME Suite (http://meme.nbcr.net/meme/). The identified motifs were then visualised with TBtools. To study cis-regulatory analysis, the promoter region of around 1,500 bp upstream of the identified CaHSF genes was analyzed using the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html). The identified cis-regulatory elements were grouped based on their reported associations with stress responses, hormone signaling, and other regulatory processes(Han et al., 2025).

Phylogenetic Analysis

HSF protein sequences of *Coffea arabica*, Agave tequilana, Arabidopsis thaliana, and Solanum lycopersicum were used for phylogenetic analysis. Before the phylogenetic reconstruction, the protein sequences were aligned. A maximum likelihood phylogenetic tree was built with 1,000 bootstrap replicates in MEGA 11. A phylogenetic tree was generated and visualized with the Interactive Tree of Life (iTOL; https://itol.embl.de/)(Mariyam et al., 2021).

Duplication Analysis, Chromosomal Mapping and Synteny Investigation

The genomic annotation from Phytozome was used to determine the chromosomal or scaffold distribution of the CaHSF genes. To study the gene duplication events, MCScanX was used, while synonymous (Ks) and nonsynonymous (Ka) substitution rates were used to determine evolutionary relationships between duplicated gene pairs.

The time since the duplication of gene pairs (c) was estimated as follows:

$T = Ks / 2\lambda$

The estimated divergence time, T, and λ is the substitution rate. To compare the synteny between *C. arabica* and A. thaliana, MCScanX and TBtools were used. The syntenic relationships that resulted were then presented graphically.

Protein-protein interaction studies

Protein– protein interactions between CaHSF proteins were predicted using the STRING (STRING-db.org) database. The interaction network was utilized to

predict associations between each pair of CaHSF proteins and to understand the general structure of the interaction network of HSF proteins (Sen, 2022).

Results

Identification of HSF genes in *Coffee arabica*

A genome-wide search was carried out in the genome of *Coffea arabica* to find genes of HSF. Candidate HSF proteins were identified by the conserved HSF domain (PF00447), and the sequences were subsequently checked to ensure that the HSF domain was present. 59 non-redundant CaHSF proteins were identified and then further characterized. The physicochemical properties of the identified CaHSF proteins were greatly different. The molecular weight of the CaHSF proteins varied from 16,758.74 Da for CaHSF27 to around 72,867 Da for the largest protein identified. Theoretical pI values were found to be between 4.72 for CaHSF2 and 9.72 for CaHSF35, with the mean being 6.20. GRAVY values were found to range between −0.992 and −0.407, with an average of around −0.688 for the different CaHSFs. The negative GRAVY values suggest that the proteins identified are more hydrophilic in general. The CaHSF proteins were found to have a range of aliphatic index from 58.52 for CaHSF8 to 81.77 for CaHSF50, with an average of ~69.99. These values offer information on the physicochemical properties and predicted stability of the CaHSF proteins. Gene orientation analysis revealed that 32 CaHSF genes were on the forward strand, and 27 CaHSF genes were on the reverse strand. The identified genes were spread over several different genomic scaffolds of the *C. arabica* genome. Table 1 presents the detailed information on the gene identifiers, scaffold location, gene orientation, protein length, molecular weight, pI, GRAVY values, and aliphatic indices. WoLF PSORT was used to predict the subcellular localization of the 59 CaHSF proteins. The results revealed that the predicted localization of the CaHSF proteins was mostly nuclear and cytoplasmic. Several CaHSF proteins were also predicted to be present in the chloroplasts, peroxisomes, and mitochondria – either alone or in association with other cell compartments. This nuclear localization is consistent with the known role of HSF proteins as transcription factors, since their role in transcription needs to be in the nucleus. Interestingly, the localization predicted for each CaHSF protein was different, indicating potential differences in their cellular localization and biological functions. CaHSF proteins are predicted to localize to the subcellular levels shown in Figure 1.

Table 1: Physicochemical characteristics and genomic information of 59 non-redundant CaHSF genes identified in the Coffea arabica genome

Transcription ID's	**Gene Rename**	**Chromosome No.**	**Location**		**Strand**	**Protein Length (A.A.)**	**Protein Molecular Weight (kDa)**	**pI**	**GRAVY**	**Aliphatic Index**
			start	end		(A.A)	Da			
evm.TU.Scaffold_465.702	CaHSF1	465	6118482	6119625	reverse	344	37559.72	5.04	-0.636	68.9
evm.TU.Scaffold_465.322	CaHSF2	465	3233106	3234888	forward	390	43884.9	4.72	-0.603	72.23
evm.TU.Scaffold_638.303	CaHSF3	638	3911567	3914864	reverse	382	43479.91	4.94	-0.631	79.35
evm.TU.Scaffold_624.602	CaHSF4	624	5146221	5147847	reverse	362	41553.41	4.93	-0.788	69.45
evm.TU.Scaffold_624.14	CaHSF5	624	138073	142772	forward	488	54262.34	5.46	-0.724	70.37
evm.TU.Scaffold_2295.1	CaHSF6	2295	11179	12437	forward	378	43010.24	4.83	-0.543	78.36
evm.TU.Scaffold_579.225	CaHSF7	579	1866658	1871275	forward	641	72867.37	5.95	-0.646	79.25
evm.TU.Scaffold_579.121	CaHSF8	579	960191	963168	forward	210	24858.14	9.39	-0.914	58.52
evm.TU.Scaffold_610.625	CaHSF9	**610**	6030993	6036168	forward	519	56751.14	4.93	-0.651	66.17

evm.TU.Scaffold_216.67	CaHSF10	216	624213	626854	reverse	261	30028.68	7.06	-0.801	65.4
evm.TU.Scaffold_216.293	CaHSF11	216	3510252	351185 8	reverse	262	28589.53	4.87	-0.538	67.75
evm.TU.Scaffold_216.285	CaHSF12	216	3388724	339033 1	reverse	390	43831.21	6.01	-0.683	71.54
evm.TU.Scaffold_216.292	CaHSF13	216	3505370	350670 6	reverse	343	39627.28	8.65	-0.697	73.06
evm.TU.Scaffold_216.284	CaHSF14	216	3385696	338703 3	reverse	343	39648.36	8.36	-0.696	71.08
evm.TU.Scaffold_1614.96	CaHSF15	1614	883916	885257	forward	359	41035.02	7.67	-0.992	68.97
evm.TU.Scaffold_671.1410	CaHSF16	671	11477793	1.1E+07	forward	512	56852.63	4.93	-0.562	69.88
evm.TU.Scaffold_671.719	CaHSF17	671	6159763	616134 1	reverse	408	46380.3	5.74	-0.736	71.69
evm.TU.Scaffold_671.933	CaHSF18	671	7687466	768897 2	reverse	338	37021.34	5.72	-0.522	67.28
evm.TU.Scaffold_952.220	CaHSF19	952	1549722	155359 0	forward	303	33622.3	5.2	-0.732	62.11
evm.TU.Scaffold_632.615	CaHSF20	632	9488899	949022 8	forward	340	39117.53	7.7	-0.713	67.74
evm.TU.Scaffold_629.75	CaHSF21	629	739843	742517	forward	413	46809.58	5.02	-0.548	76.92
evm.TU.Scaffold_629.461	CaHSF22	629	3752758	375386 5	reverse	332	36144.58	5.41	-0.505	75.18
evm.TU.Scaffold_449.127	CaHSF23	449	3933076	393467 1	forward	388	43607.94	6.17	-0.687	69.12
evm.TU.Scaffold_449.128	CaHSF24	**449**	3936376	393771 3	forward	343	39692.44	7.08	-0.684	74.2
evm.TU.Scaffold_214.569	CaHSF25	214	4803610	480485 9	forward	377	42476.74	7.82	-0.619	69.52
evm.TU.Scaffold_2016.268	CaHSF26	2016	1808918	181115 8	forward	362	41253.08	5.57	-0.896	59.81
evm.TU.Scaffold_1949.1	CaHSF27	1949	18901	19449	forward	142	16758.74	6.42	-0.824	64.51
evm.TU.Scaffold_2434.7	CaHSF28	2434	49076	51471	forward	414	46382.24	5.78	-0.557	70.89
evm.TU.Scaffold_557.335	CaHSF29	557	2314697	231688 1	forward	348	39805.4	5.48	-0.916	59.14
evm.TU.Scaffold_637.1083	CaHSF30	637	8059821	806484 9	reverse	385	42219.31	6.06	-0.545	69.95

evm.TU.Scaffold_637.1813	CaHSF31	637	13265098	1.3E+07	forward	380	41578.44	5.55	-0.55	69.61
evm.TU.Scaffold_618.559	CaHSF32	618	4974394	4975850	forward	374	43141.26	5.41	-0.957	64.36
evm.TU.Scaffold_618.175	CaHSF33	618	1689504	1690762	reverse	378	43010.24	4.83	-0.543	78.36
evm.TU.Scaffold_618.960	CaHSF34	618	8747975	8749109	forward	276	31870.45	9.33	-0.71	68.55
evm.TU.Scaffold_618.856	CaHSF35	618	7833986	7836765	forward	210	24870.24	9.47	-0.891	60.38
evm.TU.Scaffold_618.385	CaHSF36	618	3482511	3483848	forward	343	39572.34	8.98	-0.724	70.79
evm.TU.Scaffold_618.355	CaHSF37	618	3253756	3257303	reverse	472	54784.64	5.14	-0.604	77.67
evm.TU.Scaffold_618.959	CaHSF38	618	8744661	8746274	forward	391	44044.38	5.82	-0.661	73.32
evm.TU.Scaffold_618.996	CaHSF39	618	9025195	9026532	reverse	343	39423.88	8.38	-0.72	71.08
evm.TU.Scaffold_618.547	CaHSF40	618	4896729	4897984	forward	353	39969.57	5.08	-0.59	76.49
evm.TU.Scaffold_618.384	CaHSF41	618	3479191	3480798	forward	356	40337.27	6	-0.778	70.17
evm.TU.Scaffold_618.170	CaHSF42	618	1612794	1614130	reverse	296	34055.06	5.35	-0.896	62.84
evm.TU.Scaffold_625.92	CaHSF43	625	1562735	1564231	reverse	248	28266.08	6.46	-0.638	71.21
evm.TU.Scaffold_770.220	CaHSF44	770	1770171	1771658	reverse	443	49806.23	5.32	-0.668	66.48
evm.TU.Scaffold_571.589	CaHSF45	571	4139674	4142070	reverse	478	53394.03	5.06	-0.507	73.03
evm.TU.Scaffold_2290.5	CaHSF46	2290	54704	56159	forward	383	44254.53	5.34	-0.902	64.86
evm.TU.Scaffold_522.751	CaHSF47	522	6915768	6917017	forward	377	42453.77	7.79	-0.594	69.79
evm.TU.Scaffold_609.5	CaHSF48	609	36438	37925	forward	443	49802.31	5.45	-0.648	68.47
evm.TU.Scaffold_576.76	CaHSF49	576	529154	530758	reverse	408	46585.7	5.95	-0.719	73.36
evm.TU.Scaffold_576.281	CaHSF50	576	2041244	2042677	reverse	310	35497.75	9.2	-0.407	81.77
evm.TU.Scaffold_549.88	CaHSF51	549	814758	819716	forward	519	56765.12	4.89	-0.66	65.97

evm.TU.Scaffold_372.18	CaHSF52	372	251283	255989	reverse	488	54183.07	5.44	-0.745	68.77
evm.TU.Scaffold_2359.47	CaHSF53	2359	469945	473240	forward	382	43462.82	4.94	-0.646	79.35
evm.TU.Scaffold_315.489	CaHSF54	315	3343078	3344704	forward	362	41537.41	4.93	-0.796	69.97
evm.TU.Scaffold_2652.12	CaHSF55	2652	108730	110333	reverse	388	43466.69	5.44	-0.659	72.16
evm.TU.Scaffold_612.467	CaHSF56	612	3125033	3128308	reverse	304	33694.34	5.11	-0.721	60.63
evm.TU.Scaffold_419.391	CaHSF57	419	4377232	4378570	reverse	343	39566.21	8.36	-0.696	72.48
evm.TU.Scaffold_419.175	CaHSF58	419	1623805	1626438	reverse	260	29912.56	7.02	-0.788	64.15
evm.TU.Scaffold_419.392	CaHSF59	419	4380281	4381847	reverse	382	42769.25	6.64	-0.596	75.37

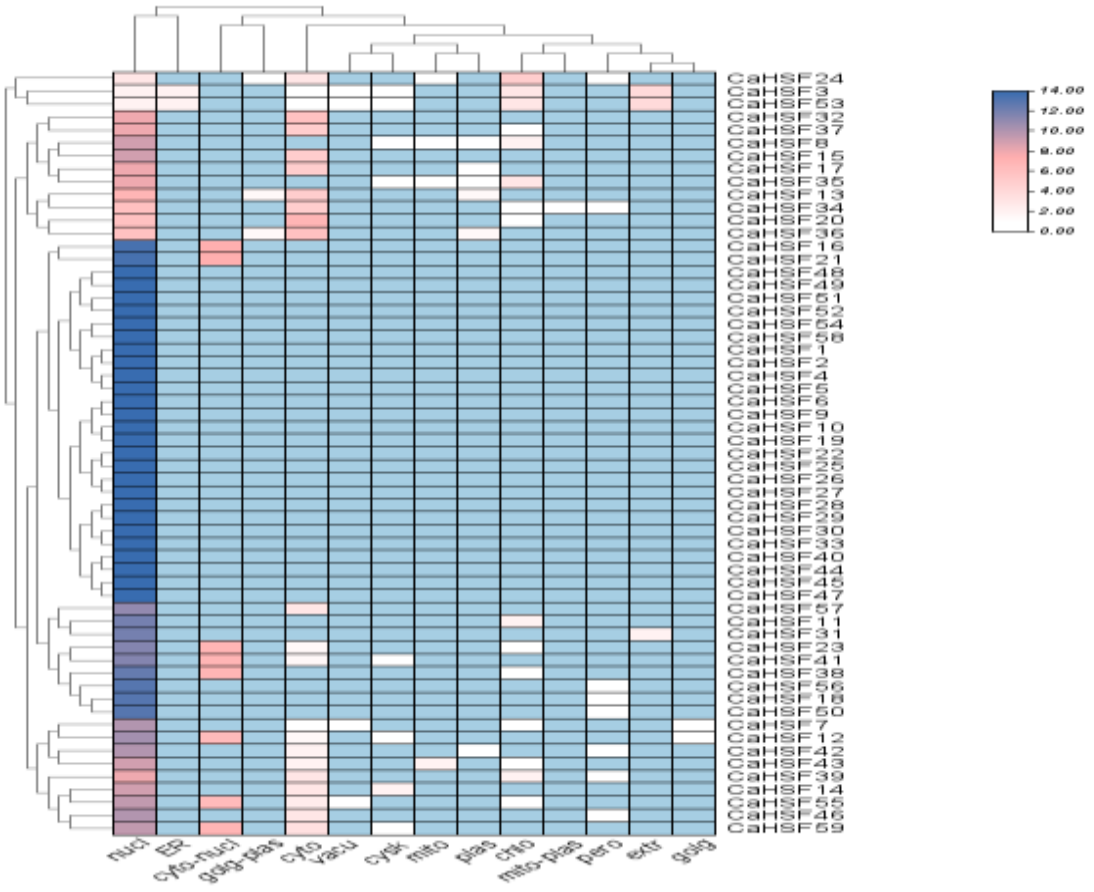

Figure 1: A heat map illustrating the subcellular localizations of all 59 CaHSF genes. The blue color indicates the highest functional importance of the gene marked in the area depicted by the diagram, whereas the white color indicates the lowest functional importance of the gene marked in the area depicted by the diagram

Phylogenetic analysis

The identified CaHSF proteins from *Coffea arabica* were then compared with the HSF protein sequences

of the selected reference plant species to address the evolutionary relationships among HSF proteins. The aligned protein sequences were used to create the phylogenetic tree used to classify the CaHSF proteins in relation to their evolutionary relationships. The phylogenetic analysis showed that the CaHSF proteins belonged to separate evolutionary groups. Close phylogenetic relationships among members generally clustered together, suggesting evolutionary conservation within the HSF gene family. The distribution of CaHSF proteins among the different phylogenetic groups also suggested that the HSF family has undergone diversification during the evolution of *C. arabica*. The relationships between the identified CaHSF proteins and the reference HSF proteins are shown in the phylogenetic tree in Figure 2 (Xi et al., 2023).

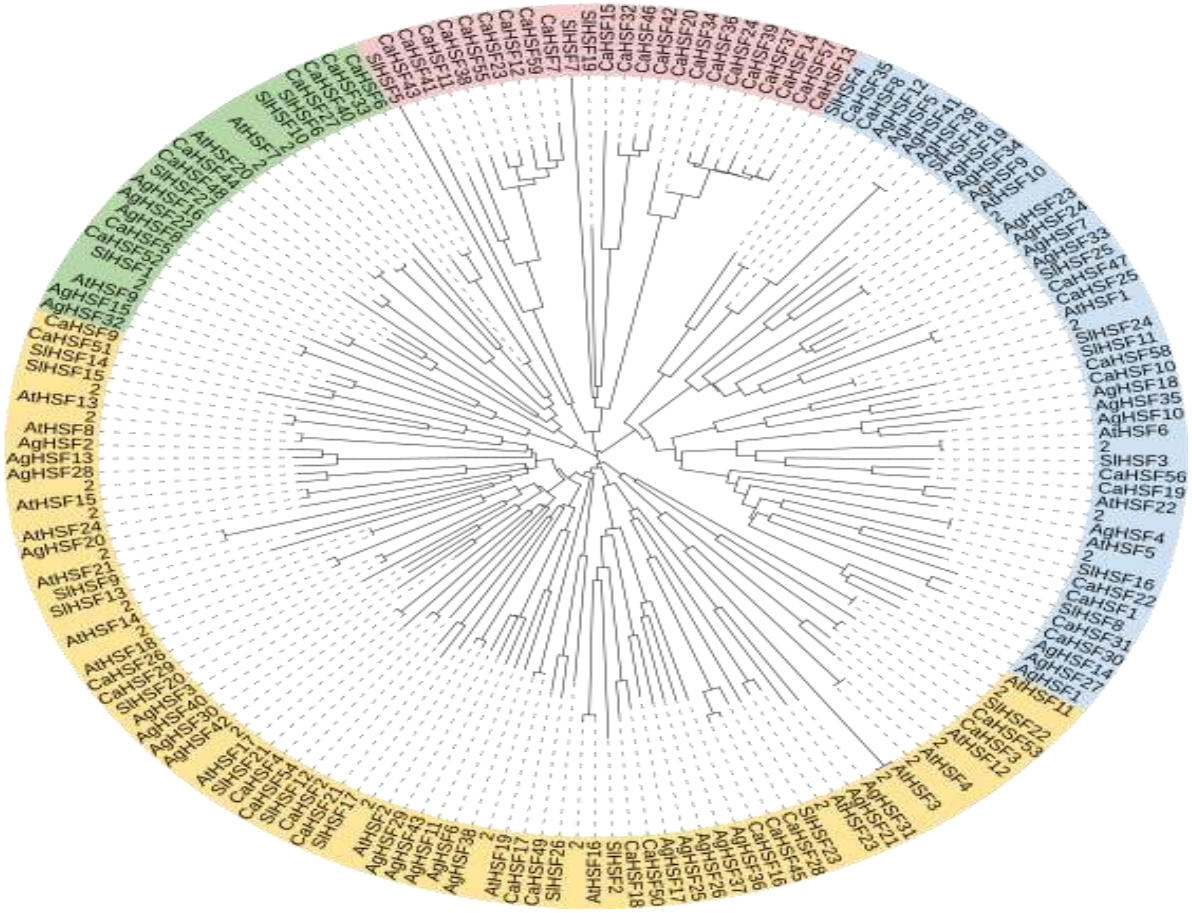

Figure 2: A phylogenetic tree method, which is commonly used in Arabidopsis research, was applied to identify the evolutionary relationships among CCO genes in four plant species - *C. arabica, A. thaliana, A. tequilauna* and *S. lycopersicum*. In the figure, the four plant species are grouped by different evolutionary traits of the CaHSF genes

Analysis of Gene Structure and Motifs

To reveal the structural features of the 59 identified CaHSF genes, their exon-intron structure was analyzed. Variation in gene structure was found amongst the CaHSF members, with genes in closely related phylogenetic groups showing similar gene structure patterns. The CaHSF gene family may have diversified in structure and function due to differences in their exonic and intronic structures. Conserved motif analysis was used to find conserved amino acid regions among the CaHSF proteins. The identified motifs exhibited differences in distribution and combination of motifs among members of the CaHSF. Several motifs were found repeatedly in proteins that are phylogenetically related, implying that such conserved regions could be related to the structural and functional properties of CaHSF proteins(Obe et al., 2002). Figure 3 shows the distribution of the conserved motifs and the exon–intron organisation of the CaHSF genes. In summary, the conservation of gene structures and protein motifs among the related CaHSF proteins is consistent with their evolutionary relationships, and differences in the motif composition and gene structure between the members of the gene family may suggest functional diversification in the members.

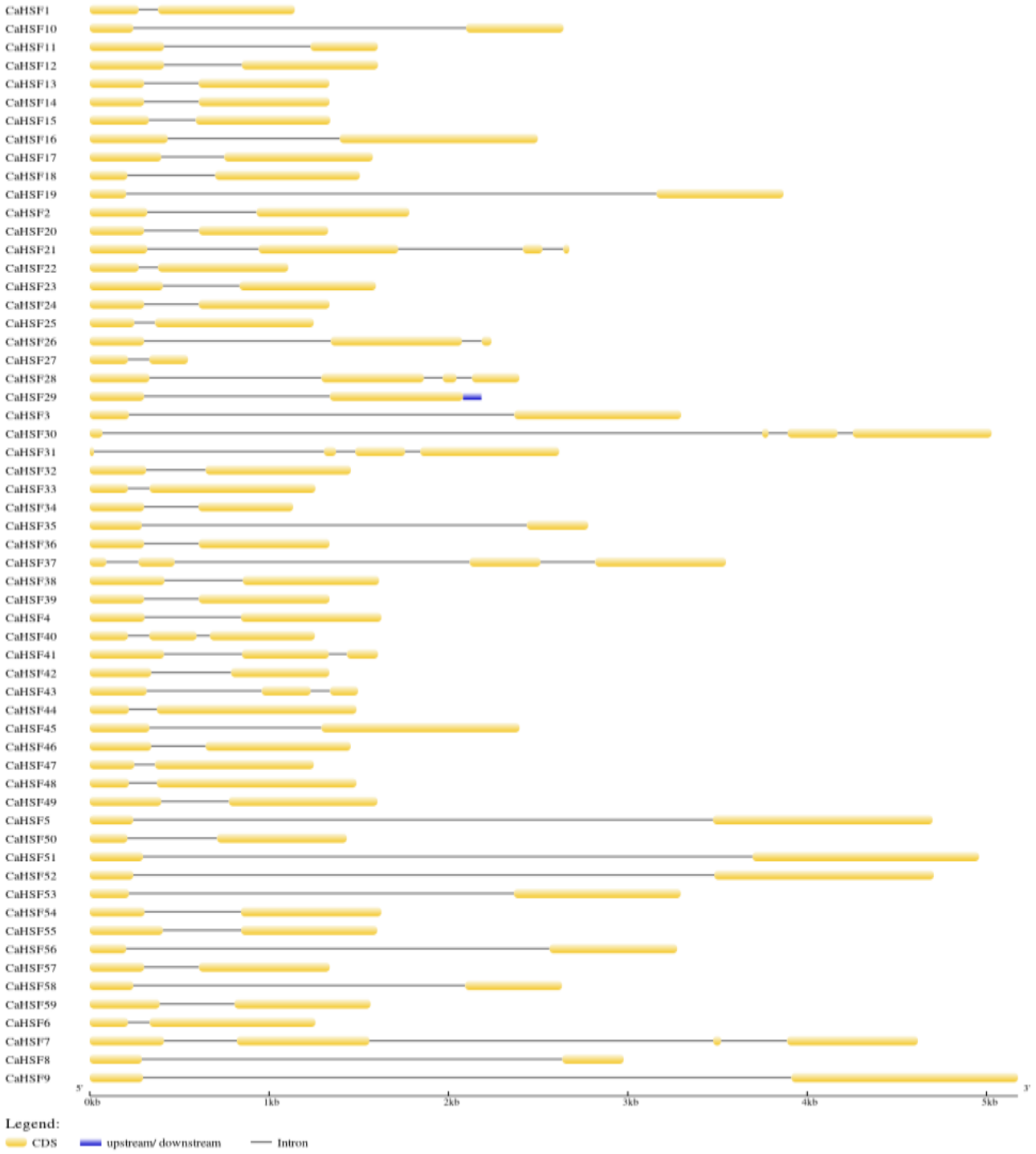

Figure 3: The intron-exon structure is phylogenetically represented, and it can be seen that the genes coding HSF have fewer coding sequences. In a few genes of the HSF gene family, the number of introns has not changed

Conserved Domain Analysis

The identified CaHSF proteins were subjected to the analysis of conserved domains to validate the typical properties of the HSF family. The results of the analysis revealed the presence of conserved regions that were characteristic of the transcription factor HSF among the proteins identified. Preservation of these domains in CaHSFs suggests that they belong to the HSF gene family. The conserved domain organization also revealed a similarity to other phylogenetically related CaHSF proteins, suggesting that these proteins might have conserved molecular functions. The variability in domain organization between members, however, may lead to diversification of function in the CaHSF family. The distribution of the CaHSF proteins by conserved domain is shown in the figure 4.

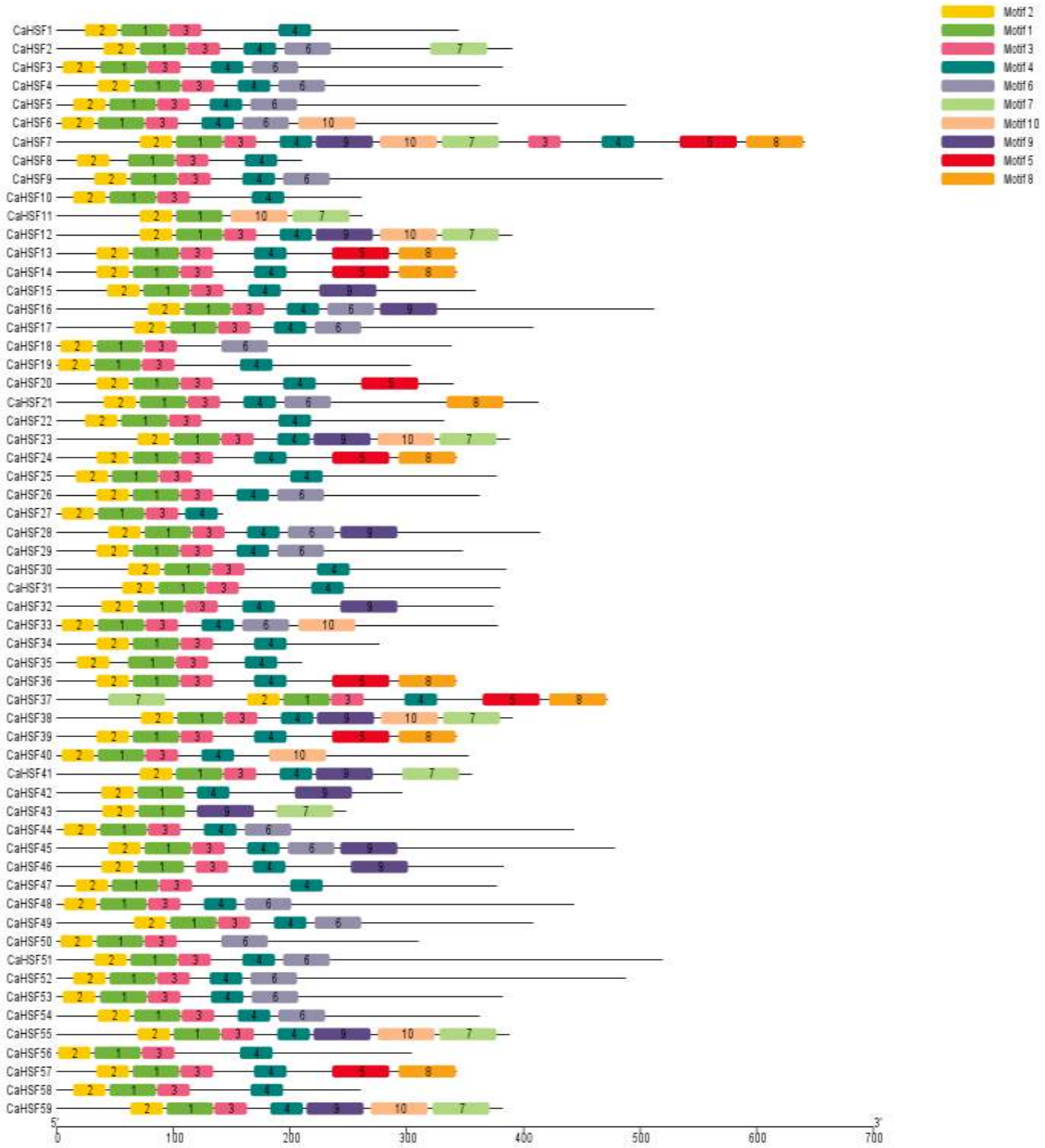

Figure 4: For each of the 10 motifs within the HSF family of C. arabica sequenced in this study, the distribution of these was plotted for the 59 members of the HSF family

Conserved Motif Analysis

To explore the conserved amino acid motifs present in the 59 CaHSF proteins, conserved motif analysis was carried out using MEME analysis. A total of 10 conserved motifs were found in the CaHSF proteins. There were differences in the distributions and arrangements of these motifs among the members of CaHSF. Multiple motifs were found to be more widely shared amongst the CaHSF proteins, such as Motifs 1, 2, 3 and 4, while others were more localized in their distribution. Conserved motifs were arranged in a similar manner in closely related CaHSF proteins, corroborating the evolutionary relationships. Some CaHSF proteins, however, had extra motifs or variations in the motifs, indicating structural diversification within the HSF gene family. Some of

the identified proteins had more conserved motifs than other proteins that were members of the CaHSF family. CaHSF7, in comparison to other family members, had a relatively complex motif arrangement. In summary, the result of the conserved motif analysis indicated that CaHSF proteins have several conserved motifs, but they also have some differences in the number and type of motifs. The 10 conserved motifs were distributed among the 59 CaHSF proteins as shown in Figure 5.

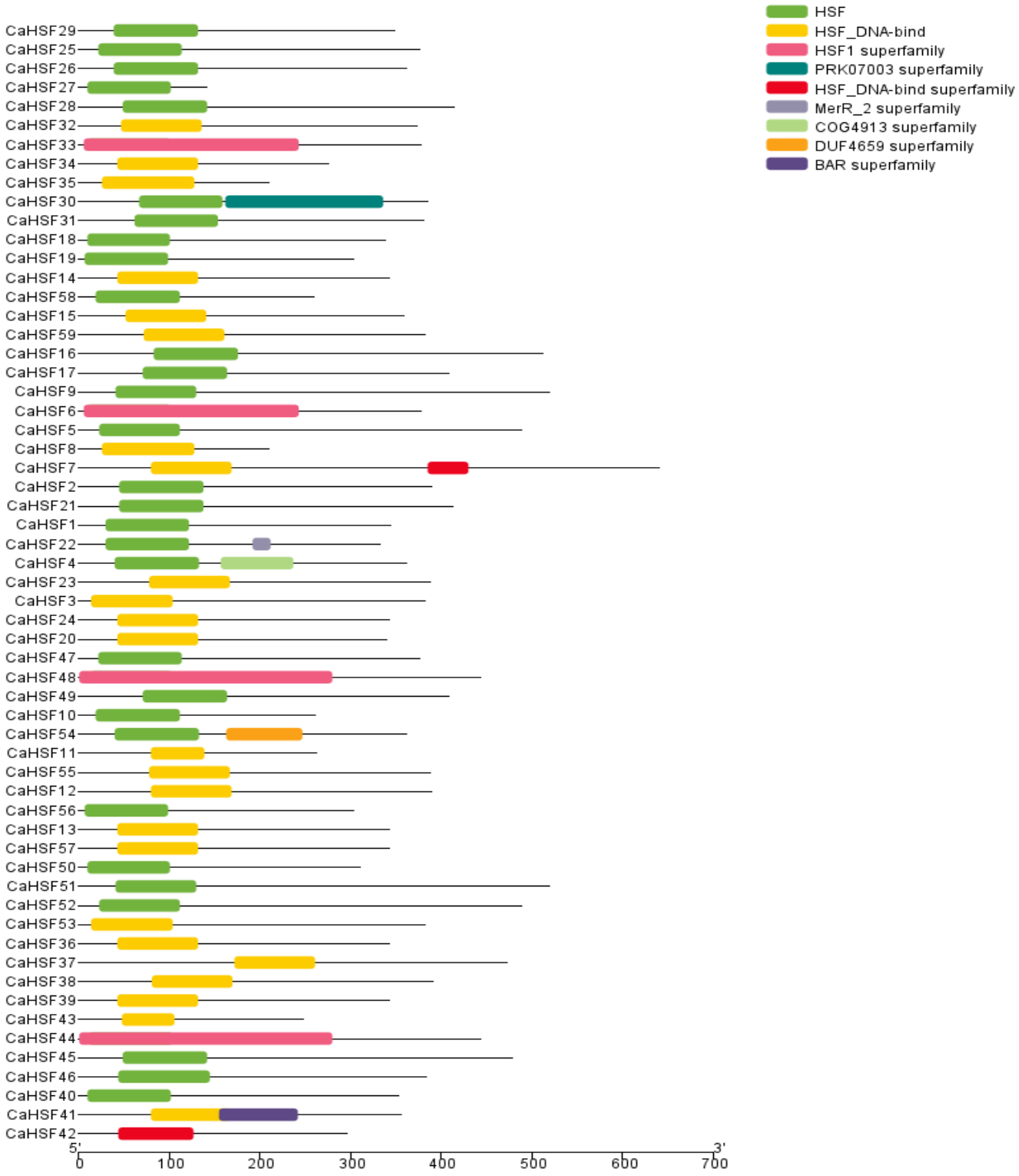

Figure 5: Conserved Domain analysis can be used to identify proteins (in this case, 59 CaHSF proteins) that contain a conserved PF00447 domain, widely implying one or more specific functions in plant growth and development

Evaluation of duplication event of *C. arabica*

To examine the evolutionary expansion and diversification of the CaHSF gene family in *Coffea arabica*, gene duplication analysis was carried out. Through Ka/Ks analysis, 595 duplicated gene pairs were detected in the coffee genome, with 492 of these linked to the CaHSF genes (Figure 6). The ratio of Ka/Ks was different for various duplicated CaHSF gene pairs. CaHSF44_CaHSF48 exhibited the highest Ka/Ks ratio (1.304804853), whereas CaHSF4_CaHSF26 showed the lowest Ka/Ks ratio (0.050473056). The relatively high Ka/Ks ratio of the gene pair CaHSF44 and CaHSF48 could suggest comparatively greater evolutionary divergence and positive selection, while the low Ka/Ks ratio of the gene pair CaHSF4 and CaHSF26 suggests a predominant purifying selection in their evolutionary pathway. The estimated divergence times of the duplicated CaHSF gene pairs also showed a wide range of variation. CaHSF44_CaHSF48 was estimated to have diverged 0.5131224 million years ago (MYA) with a relatively short estimated duplication time. CaHSF18_CaHSF44, however, had an estimated divergence time of 474.1050258 MYA, much older than the evolutionary divergence between the gene sequences in the other two organisms. These results suggest that both relatively recent and ancient duplication events have contributed to the evolution of the CaHSF gene family in *C. arabica*.

Syntenic analysis

The syntenic analysis reveals both tandem and segmental duplication events across multiple scaffolds. Tandem duplications are apparent on scaffolds 571, 522, 2290, and 624, indicating that consecutive duplications occurred within these regions, resulting in repeated gene sequences along the same chromosome. Segmental duplications, characterized by the duplication of larger genomic regions, are observed across various scaffolds, such as 770, 315, 2016, and 611, showing that these segments have been duplicated either within chromosomes. This detailed syntenic arrangement emphasizes the complex evolutionary processes and paralogous duplication mechanisms influencing genome organization (Figure 7).

Analysis of *CaCCO* Cis-regulatory elements

We analyzed the cis-regulatory elements found in the promoter region of the *Coffea arabica* HSF (CaHSF) genes to examine the change of their possible roles in stress responses, hormone signaling, and light-mediated regulation. The promoter regions of the CaHSF genes were screened for several types of cis-regulatory elements (Figure 8). ABRE (ABA-responsive element) motifs were found in the promoters of CaHSF12, CaHSF25, and CaHSF43, indicating that these genes may be responsive to abscisic acid (ABA) mediated signaling and abiotic stress. TCA element associated with salicylic acid (SA) responsiveness was identified in CaHSF15, CaHSF28, and CaHSF47, indicating their possible role in hormone-mediated stress response. Cis-regulatory elements associated with stress were also identified. CaHSF6, CaHSF20 and CaHSF33 contained MBS (MYB-binding site) associated with drought responsiveness. The LTR (low-temperature-responsive element) was detected in CaHSF9, CaHSF21 and CaHSF38, which may be involved in temperature stress responses. Moreover, several light-sensitive cis-regulatory elements were found in the promoters of CaHSF. CaHSF3, CaHSF14, and CaHSF30 were identified as containing Box 4, while the G-box motif was found in CaHSF10 and CaHSF34. The identification of these factors indicates that there are other light-responsive signaling pathways that may control some CaHSF genes (Gushchina et al., 2018). In conclusion, the presence of stress-, hormone-, and light-responsive cis-regulatory elements within the promoters of CaHSF genes suggests that these genes are likely to be regulated by various signaling pathways and play a role in *C. arabica*'s response to environmental stresses.

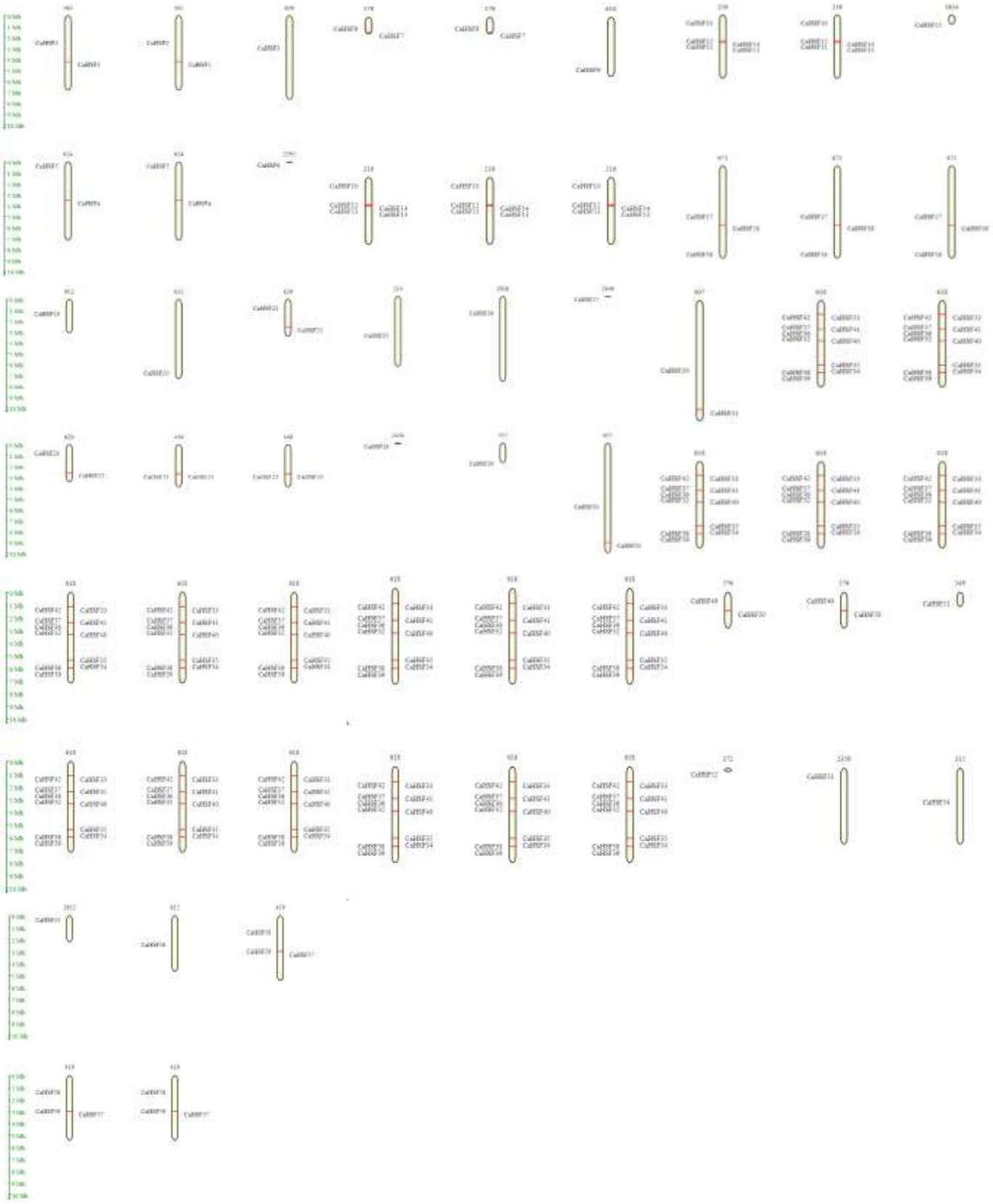

Figure 6: A genome-wide chromosomal map of the HSF genes reveals the presence of paralogous counterparts to the C. Arabica genome where they are likely to be. Although altered, the duplication of HSFgenes is retained and so is their functional roles in the C. arabica genome to enable development of positive functional feature

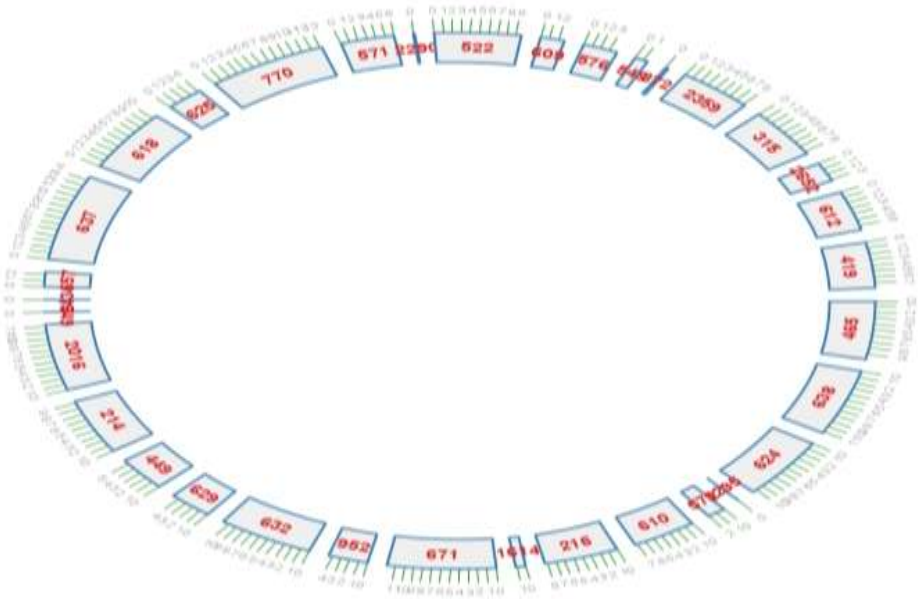

Figure 7: A thorough genome synteny study in C. arabica HSF genes shows that genes are located in structured sequences. A combination of genomic factors and dynamic processes that allow genes to remain stable and secure, adapt to their environment, change over generations, etc., gives rise to gene duplication

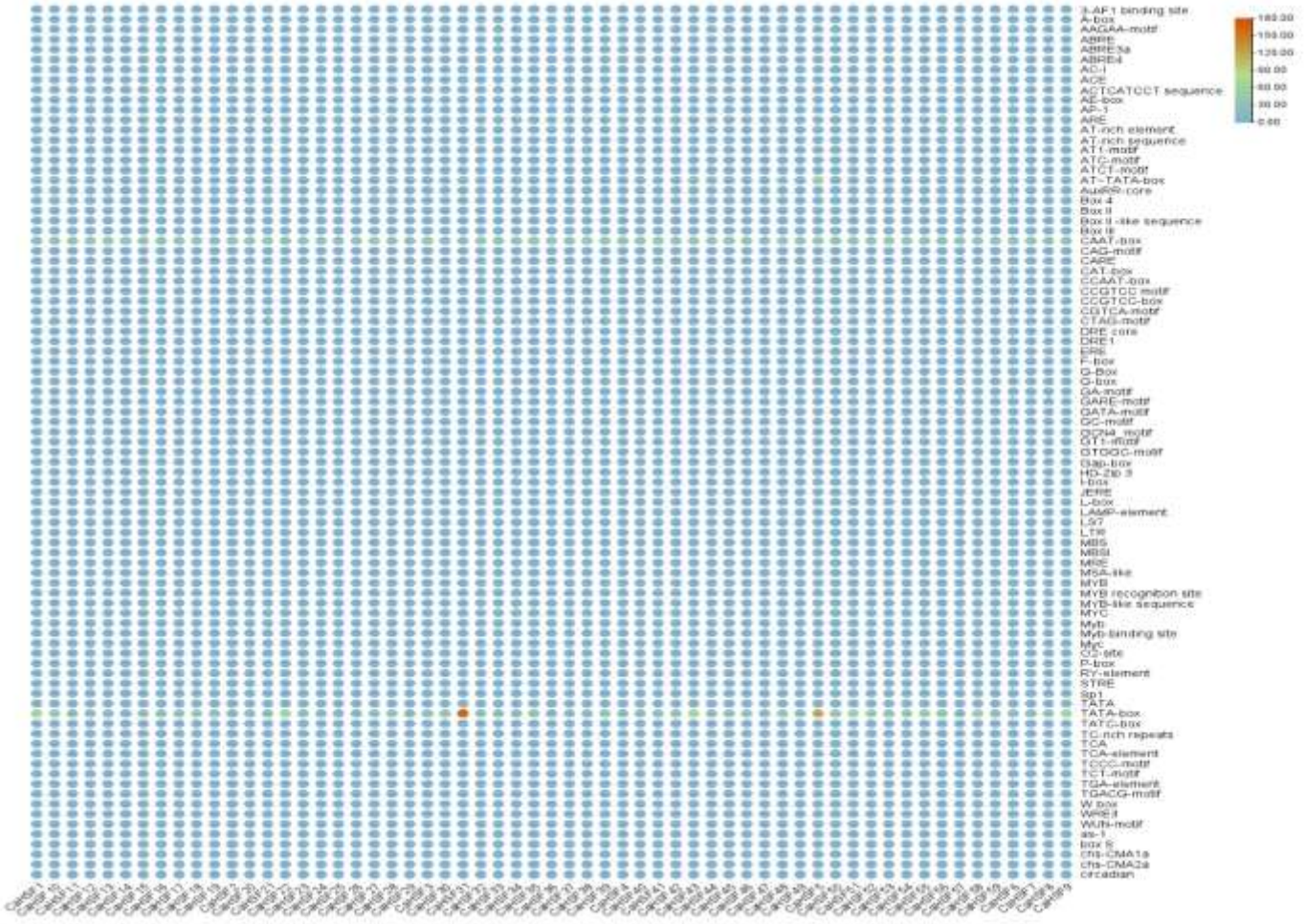

Figure 8: The cis-regulatory analysis of HSF gene and the strength of each function is represented graphically. The process intensities taking place in the plants change from red (high) to blue (low)

Analysis of protein-protein interaction network

We analyzed the cis-regulatory elements found in the promoter region of the *Coffea arabica* HSF (CaHSF) genes to examine the change in their possible roles in stress responses, hormone signaling, and light-mediated regulation. The promoter regions of the CaHSF genes were screened for several types of cis-regulatory elements. ABRE (ABA-responsive element) motifs were found in the promoters of CaHSF12, CaHSF25, and CaHSF43, indicating that these genes may be responsive to abscisic acid (ABA) mediated signaling and abiotic stress. TCA element associated with salicylic acid (SA) responsiveness was identified in CaHSF15, CaHSF28, and CaHSF47, indicating their possible role in hormone-mediated stress response. Cis-regulatory elements associated with stress were also identified. CaHSF6, CaHSF20 and CaHSF33 contained MBS (MYB-binding site) associated with drought responsiveness (Figure 9). The LTR (low-temperature-responsive element) was detected in CaHSF9, CaHSF21 and CaHSF38, which may be involved in temperature stress responses. Moreover, several light-sensitive cis-regulatory elements were found in the promoters of CaHSF. CaHSF3, CaHSF14, and CaHSF30 were identified as containing Box 4, while the G-box motif was found in CaHSF10 and CaHSF34. The identification of these factors indicates that there are other light-responsive signaling pathways that may control some CaHSF genes (Figure 9). In conclusion, the presence of stress-, hormone-, and light-responsive cis-regulatory elements within the promoters of CaHSF genes suggests that these genes are likely to be regulated by various signaling pathways and play a role in *C. arabica*'s response to environmental stresses.

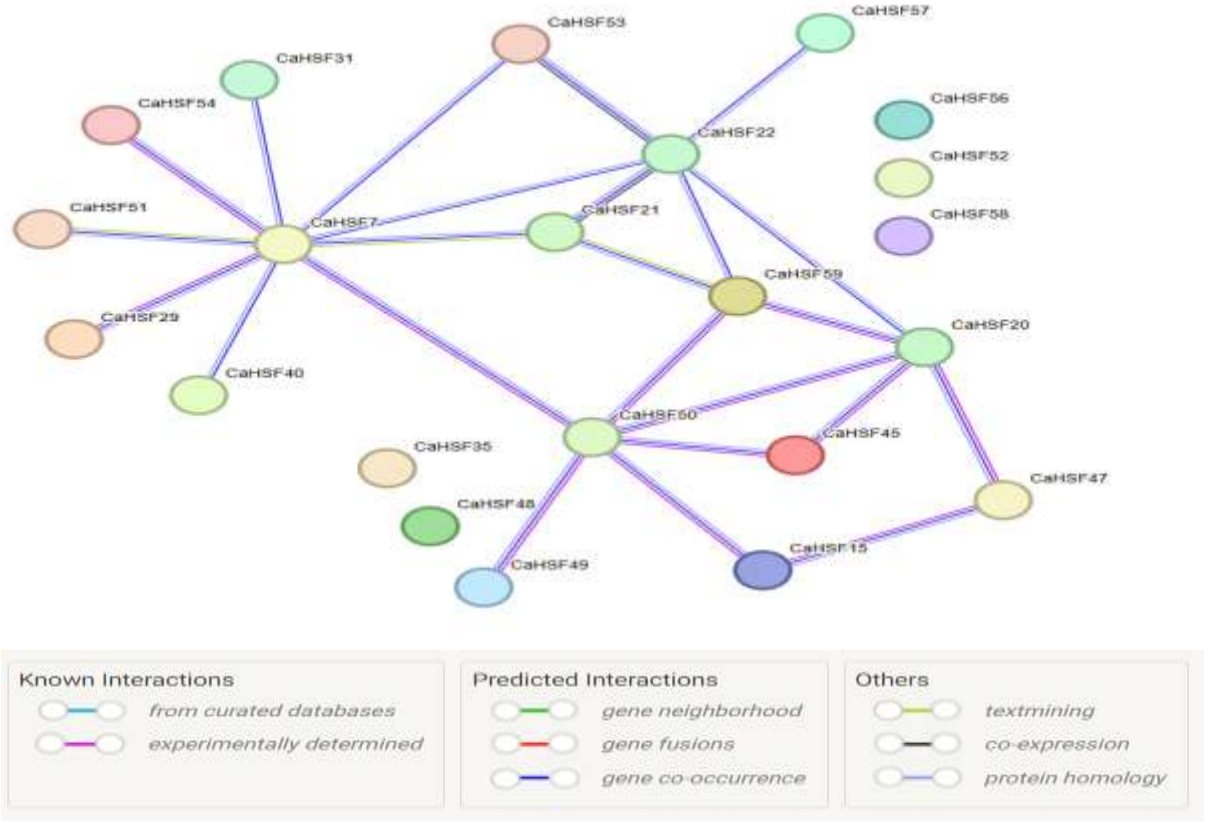

Figure 9. Based on protein-protein interactions, twenty-two genes were associated with 59 CaHSF proteins. In the dataset, CaHSF7 and CaHSF40 exhibited the highest number of interactions, being associated with 6 proteins

Discussion

Among the abiotic stresses, drought and temperature stress are key components that constrain the growth and productivity of coffee. Knowing the families of stress-induced genes in *Coffea arabica* is therefore significant in providing information on the stress adaptation molecular mechanism. In the current study, a genome-wide search was conducted to identify and characterize 49 genes for CaHSF in *C. arabica*(Gao et al., 2025). Physicochemical, phylogenetic, gene-structure, conserved-motif, duplication, synteny, cis-regulatory, and protein–protein interaction analyses were performed on the identified genes. Significant differences were found between the CaHSF proteins in the physicochemical analysis. The molecular weight of the proteins ranged between 16758.74 Da and 72867.27 Da, and the predicted pI values of the proteins ranged from 4.72 to 9.72. The GRAVY values for the CaHSF proteins are all negative, suggesting that these proteins are generally hydrophilic. The aliphatic index (AI) was between 58.52 and 81.77, with a mean value of 69.99.

The difference in physicochemical properties suggests that the CaHSF proteins vary in some way that may correlate with differences in molecular properties. But it is important to note that such computational properties are not direct evidence of heat tolerance and/or activity. The subcellular localization of the predicted CaHSF proteins revealed that most of the proteins were nuclear(Shao et al., 2022). This is in keeping with the role of HSFs as transcription factors, as most transcriptional regulation takes place in the nucleus. The differences in predicted localization for each of the individual CaHSF proteins could reflect differences in cellular localization. These localization results are, however, computational predictions; prior to the specific functions being assigned to the cells, these predictions must be experimentally verified(Marques et al., 2024).

The information from phylogenetic analysis was used to gain insight into evolutionary relationships between CaHSF proteins and HSF proteins of the selected reference species. The relatedness of the proteins grouped into clusters is a basis for understanding the evolutionary conservation and diversification of the HSF family in *C. arabica*. Proteins in the same phylogenetic cluster may share similar features and may also have similar functions, although this is not necessarily true because functional similarity cannot be inferred from phylogenetic clustering. Conserved motifs and the analysis of the exon-intron structure of CaHSF further confirmed the evolutionary relationships among CaHSF proteins. The gene structures and conserved motifs were similar in some proteins, but not in others. The MEME analysis was used to identify 10 conserved motifs among the 59 CaHSF proteins, and the positions of these motifs varied among them. This conservation and variation of motifs in the motif organization may account for the structural conservation and diversification of HSF proteins. The motif patterns observed, however, must be considered as a sign of sequence conservation and not of specific biological function. To understand the expansion of the CaHSF family, the gene duplication analysis was conducted. A total of 595 pairs of duplicated genes were identified in the coffee genome, with 492 pairs of duplicated CaHSF genes. Ka/Ks was not the same for the duplicated pairs. The pair CaHSF44-CaHSF48 gave the highest Ka/Ks value of 1.304804853, while the pair CaHSF4-CaHSF26 gave the lowest value of 0.050473056(Chen et al., 2018). The divergence-time estimates also had significant differences, with CaHSF44_CaHSF48 having a relatively recent estimated age of divergence of 0.5131224 MYA, and CaHSF18_CaHSF44 having an estimated age of divergence of 474.1050258 MYA. The results suggest that diversification of the CaHSF gene family could have been driven by both relatively recent and ancient duplication events. The cis-regulatory analysis revealed several regulatory elements associated with stress responses, hormone signaling, and light responsiveness. Three motif elements of ABRE were found in the promoters of CaHSF12, CaHSF25, and CaHSF43, while three motif elements of TCA were found in the promoters of CaHSF15, CaHSF28, and CaHSF47. CaHSF6, CaHSF20, and CaHSF33 contained MBS elements, and CaHSF9, CaHSF21 and CaHSF38 contained LTR elements. Box 4 motifs and G-box motifs were also found in some of the CaHSF promoters, which are light-responsive elements. The results indicate that the regulation of the CaHSF genes could be through several environmental and hormonal signaling pathways. However, the identification of a cis-regulatory element does not show that the associated gene is regulated in this way; it must be confirmed by experimental expression and promoter analyses(Chen JiangFei et al., 2018).

Another piece of information on possible association of CaHSF proteins was obtained from the predicted protein–protein interaction network. There are 22 nodes and 24 edges in the network, with an average node degree of 2.18 and the average local clustering coefficient of 0.496(MUSTAFIDAH and Su-udi, 2025). The interaction enrichment p-value was reported as $<1.0 \times 10^{-16}$, which means that the observed network had more interactions than a network randomly assembled from the same population. CaHSF7 and CaHSF40 had the most interactions (6), while CaHSF12 had 3 and CaHSF20 had 4.CaHSF12 had the fewest interactions (3), followed by CaHSF20 (4) and then CaHSF7 and CaHSF40 (6). These highly interconnected proteins could be interesting targets for future functional studies. The network is, however, based on computational predictions and needs experimental validation to ensure these protein interactions are accurate. In general, the present study aims to conduct a genome-wide characterization of the gene family CaHSF in *C. arabica*. The simultaneous analysis of physicochemical properties, phylogeny, gene structure, conserved motifs, duplication, synteny, cis-regulatory elements, and predicted protein interactions gives an idea of the steps that can be taken to further explore the HSF-mediated stress response in coffee. The stress-related cis-elements, recent duplications, and highly connected PPI profiles in the identified CaHSF genes suggest they could be good candidates for future experiments(Xu et al., 2020). However, additional research using gene-expression analysis, functional characterisation, and experimental validation will be needed to clarify the individual roles of the different CaHSF genes in drought and other abiotic stress responses.

Conclusion

In the present study, a genome-wide study of the Heat Shock Factor (HSF) gene family was carried out in *Coffea arabica*. 59 CaHSF genes were discovered and characterized by applying various computational techniques. Physicochemical properties of identified

CaHSF proteins were found to differ from each other, and the conserved motif, gene structure, and phylogenetic analysis of the gene family showed conserved and diversified properties. The evolutionary expansion and diversification of CaHSF genes in the *C. arabica* genome were analysed through gene duplication and synteny studies. In addition, several other stress-, hormone-, and light-responsive cis-regulatory elements were identified in the promoter regions, indicating that CaHSF genes likely play a role in several regulatory pathways related to environmental responses. Other proteins that were also highly connected in the predicted protein–protein interaction network were also candidates for further functional studies. Overall, this study offers a complete genomic resource of the CaHSF gene family in *C. arabica* and lays the groundwork for future experimental studies to determine the exact function of the CaHSF genes in response to drought, heat, and other abiotic stresses. These genes need to be confirmed with gene-expression and functional studies to validate their specific functions in stress tolerance of coffee.

References

Ali, Z., Sardar, S., & Kirbag, S. (2026). Cold stress tolerance in *Zea mays*: regulatory mechanisms. *Journal of Life and Social Sciences*, **2026**, 65. https://doi.org/10.64013/bbasrjlifess.v2026i1.65

Buttar, Z. A., Yang, Y., Sharif, R., Nan Wu, S., Xie, Y., & Wang, C. (2020). Genome Wide Identification, Characterization, and Expression Analysis of YABBY-Gene Family in WHEAT (*Triticum aestivum* L.). *Agronomy*, **10**(8), 1189. https://doi.org/10.3390/agronomy10081189 Cited by: 36

Chen, J., Gao, T., Wan, S., Zhang, Y., Yang, J., Yu, Y., & Wang, W. (2018). Genome-Wide Identification, Classification and Expression Analysis of the HSP Gene Superfamily in Tea Plant (*Camellia sinensis*). *International Journal of Molecular Sciences*, *19*(9), 2633. https://doi.org/10.3390/ijms19092633 Cited by: 102

Ding, Y., Xu, J., Liang, X., Sha, W., Hao, Z., & Li, H. (2026). Comprehensive Genome-Wide Analysis and Expression Profiling of the Heat Shock Transcription Factor (Hsf) Gene Family, and Function Detection of BpHsf11 in Heat Stress Response in Paper Mulberry (*Broussonetia papyrifera*). *Russian Journal of Plant Physiology*, **73**(1), 17.

Gain, H., De, S., & Banerjee, J. (2024). Expressional vagaries of *OsCAMTA* genes under differential abiotic stresses supported with protein–protein interaction study and prediction of miRNA target sites. *Plant Biotechnology Reports*, **18**(5), 637–658. https://doi.org/10.1007/s11816-024-00918-6

Gao, J., Zhu, J., Lan, Z., He, F., & Dong, X. (2025). Genome-Wide Analysis of NAC Gene Family and Its Cold-Responsive Transcriptional Dynamics in *Coffea arabica*. *Agronomy*, **15**(10), 2394. https://doi.org/10.3390/agronomy15102394

Guo, M., Liu, J.-H., Ma, X., Luo, D.-X., Gong, Z.-H., & Lu, M.-H. (2016). The Plant Heat Stress Transcription Factors (HSFs): Structure, Regulation, and Function in Response to Abiotic Stresses. *Frontiers in Plant Science*, **7**, 114. https://doi.org/10.3389/fpls.2016.00114

Gushchina, L. V., Kwiatkowski, T. A., Bhattacharya, S., & Weisleder, N. L. (2018). Conserved structural and functional aspects of the tripartite motif gene family point towards therapeutic applications in multiple diseases. *Pharmacology & Therapeutics*, **185**, 12–25. https://doi.org/10.1016/j.pharmthera.2017.10.020 Cited by: 92

Hahn, J.-S., Hu, Z., Thiele, D. J., & Iyer, V. R. (2004). Genome-wide Analysis of the Biology of Stress Responses through Heat Shock Transcription Factor. *Molecular and Cellular Biology*, **24**(12), 5249–5256. https://doi.org/10.1128/MCB.24.12.5249-5256.2004

Han, K., Lai, M., Zhao, T., Yang, X., An, X., & Chen, Z. (2025). Plant YABBY transcription factors: a review of gene expression, biological functions, and prospects. *Critical Reviews in Biotechnology*, **45**(2), 214–235. https://doi.org/10.1080/07388551.2024.2341238

Hayyat, Q., Ali, Q., Abrar, R., & Khan, S. (2026). Genome-wide association and identification of *Zea mays* chitinase gene family. *Journal of Life and Social Sciences*, **2026**, 63. https://doi.org/10.64013/bbasrjlifess.v2026i1.63

Li, Y., Gong, K., Wang, X., Sun, Z., & Ding, F. (2025). Heat Shock Transcription Factors as Central Integrators of Plant Stress Responses: From Thermotolerance to Multi-Stress Resilience. *Biology*, *14*(12), 1800. https://doi.org/10.3390/biology14121800

Mariyam, Shafiq, M., Haseeb, M., Atif, R. M., Naqvi, S. A. A. A. A., Ali, N., Javed, M. A., Gillani, F., & Haider, M. S. (2021). Genome-wide identification and characterization of a plant-specific Dof transcription factor gene family in olive (*Olea europaea*) and its comparison with *Arabidopsis*. *Horticulture, Environment, and Biotechnology*, **62**(6), 949–968. https://doi.org/10.1007/s13580-021-00366-2

Marques, I., Fernandes, I., Paulo, O. S., Batista, D., Lidon, F. C., Rodrigues, A. P., Partelli, F. L., DaMatta, F. M., Ribeiro-Barros, A. I., & Ramalho, J. C. (2024). Transcriptomic Analyses Reveal That *Coffea arabica* and *Coffea*

canephora Have More Complex Responses under Combined Heat and Drought Than under Individual Stressors. *International Journal of Molecular Sciences*, **25**(15), 7995. https://doi.org/10.3390/ijms25157995

Mustafidah, Q. A., & Su-udi, M. (2025). In silico study: identification and characterization of heat shock protein 90 (HSP90) in Arabica coffee (*Coffea arabica* L.). *Nusantara Bioscience*, **17**(1). https://doi.org/10.13057/nusbiosci/n170110

Obe, G., Pfeiffer, P., Savage, J. R. K., Johannes, C., Goedecke, W., Jeppesen, P., Natarajan, A. T., Martínez-López, W., Folle, G. A., & Drets, M. E. (2002). Chromosomal aberrations: formation, identification and distribution. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **504**(1–2), 17–36. https://doi.org/10.1016/S0027-5107(02)00076-3

Ramalho, J. C., Marques, I., Pais, I. P., Armengaud, J., Gouveia, D., Rodrigues, A. P., Dubberstein, D., Leitão, A. E., Rakočević, M., & Scotti-Campos, P. (2025). Stress resilience in *Coffea arabica* and *Coffea canephora* under harsh drought and/or heat conditions: selected genes, proteins, and lipid integrated responses. *Frontiers in Plant Science*, **16**, 1623156. https://doi.org/10.3389/fpls.2025.1623156

Sen, S. (2022). Genome-wide TIFY family in *Arachis hypogaea* in the perspective of legume JAZs. *Journal of Crop Science and Biotechnology*, **25**(4), 465–488. https://doi.org/10.1007/s12892-022-00147-3

Shao, K.-Z., Lyu, X.-P., Li, J.-L., Chen, J., Zhao, L.-Y., Ren, W., & Zhang, J.-L. (2022). Biological characteristics of heat shock transcription factors and their roles in abiotic stress adaptation of higher plant. *Ying Yong Sheng Tai Xue Bao*, **33**(8), 2286–2296. https://doi.org/10.13287/j.1001-9332.202208.031

Xi, R., Liu, H., Chen, Y., Zhuang, H., Han, H., Wang, H., Wang, Q., & Li, N. (2023). Genome-Wide Characterization of Tomato FAD Gene Family and Expression Analysis under Abiotic Stresses. *Plants*, **12**(22), 3818. https://doi.org/10.3390/plants12223818

Xu, P., Guo, Q., Pang, X., Zhang, P., Kong, D., & Liu, J. (2020). New Insights into Evolution of Plant Heat Shock Factors (Hsfs) and Expression Analysis of Tea Genes in Response to Abiotic Stresses. *Plants*, **9**(3), 311. https://doi.org/10.3390/plants9030311

Yu, T., Bai, Y., Liu, Z., Wang, Z., Yang, Q., Wu, T., Feng, S., Zhang, Y., Shen, S., & Li, Q. (2022). Large-scale analyses of heat shock transcription factors and database construction based on whole-genome genes in horticultural and representative plants. *Horticulture Research*, **9**, uhac035. https://doi.org/10.1093/hr/uhac035

Statements and Declarations

Funding

The authors declare that they have not received any funds, grants, or other type of financial support for the preparation of this study.

Competing Interests

The authors declare that they have no financial or none, financial competing interests related to this work.

Author Contributions

Conceptualization and study design were performed by SN, ZN, ZA and QH. Data collection, bioinformatics analysis, and interpretation were carried out by SN, MH, ZA. The writing of the first manuscript draft was done by QH has checked and approved the final version of the manuscript.

Ethics Approval

Since this research does not involve human subjects, animals, or any biological material that would require both the use of and the ethical approval, the authors confirm that no such approval was necessary.

Consent to Participate

Not applicable.

Consent to Publish

Not applicable.

Data Availability

All data generated or analyzed during this study are included in this published article.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the author(s) have not used any ChatGPT (OpenAI) tool.