e-ISSN 1983-4063 - www.agro.ufg.br/pat - *Pesqui. Agropecu. Trop.*, Goiânia, v. 56, e83957, 2026

Research Article

Analysis of heat shock transcription factor gene family in cocoa[1]

Waki'atil Rosida[2], Fuad Bahrul Ulum[2], Mukhamad Su'udi[2]

ABSTRACT

The heat shock factor (HSF) is crucial for plant stress regulation. This study aimed to analyze the role and characteristics of the HSF protein in *Theobroma cacao* L. The assesment was carried out bioinformatically, consisting of HSF identification in the cocoa genome, analysis of physical properties, gene structure, motifs, and subcellular, phylogenetics and chromosomal localization. A total of 26 HSF proteins were identified, presenting a complete structure consisting of exons and introns. The analysis of the *TcHSF* gene promoter revealed several cis-acting elements that act on stress, such as the responsive elements to anaerobiosis, abscisic acid, and low temperature, as well as the myeloblastosis binding sequence. Almost all *TcHSF* exhibited a strong degree of subcellular localization in the nucleus. Cocoa phylogenetics showed that Thecc1EG037635t1 is in a clade with At5G54070, with a bootstrap value of 90 %. The three best candidates (*TcHSF-14*, *TcHSF-24*, and *TcHSF-25*) are recommended for further functional analysis, due to their potential for increasing cocoa value.

KEYWORDS: *Theobroma cacao* L., abiotic stress, gene structure.

RESUMO

Análise da família de genes do fator de transcrição de choque térmico em cacau

O fator de choque térmico (HSF) é fundamental na regulação do estresse em plantas. Objetivou-se analisar o papel e as características da proteína HSF em *Theobroma cacao* L. A avaliação foi realizada bioinformaticamente, consistindo na identificação do HSF no genoma do cacau, análise de propriedades físicas, estrutura gênica, motivos e localização subcelular, filogenética e cromossômica. Um total de 26 proteínas HSF foram identificadas, apresentando estrutura completa composta por éxons e introns. A análise do promotor gênico *TcHSF* revelou vários elementos cis-regulatórios que atuam na resposta ao estresse, como os elementos responsivos à anaerobiose, ao ácido abscísico e a baixas temperaturas, bem como a sequência de ligação à mieloblastose. Quase todos os *TcHSF* apresentaram forte grau de localização subcelular no núcleo. A filogenética do cacau mostrou que Thecc1EG037635t1 está em um clado com At5G54070, com valor bootstrap de 90 %. Os três melhores candidatos (*TcHSF-14*, *TcHSF-24* e *TcHSF-25*) são recomendados para análise funcional adicional, devido ao seu potencial para aumentar o valor do cacau.

PALAVRAS-CHAVE: *Theobroma cacao* L., estresse abiótico, estrutura gênica.

INTRODUCTION

Cocoa (*Theobroma cacao* L.) is a plant from Amazonia that belongs to the Malvaceae family (Nieves-Orduña et al. 2023). Indonesia is the third largest producer in the world, after Ivory Coast and Ghana (Izzah & Damayanti 2023).

Its global demand is increasing annually and reached 7.61 million tons in 2023 (Kongor et al. 2024). However, a decline in production has occurred since 2022, when it was estimated at 706 thousand tons, and has currently reached only 667.3 thousand tons (BPS 2025). This decline in production has been caused by the impact of climate change, diseases, and the current global macroeconomic outlook (Kehinde 2022). Drought, due to climate change, can reduce yield; so, it is necessary to study the mechanisms and gene regulation underlying drought stress.

The heat shock factor (HSF) plays a crucial role in development and response of plants in dealing with biotic and abiotic stresses, particularly in plant stress tolerance and resistance through signal transmission, regulation and gene expression (Zhang et al. 2022). It is a gene transcription regulator that

---

[1] Received: Sep. 16, 2025. Accepted: Nov. 19, 2025. Published: Jan. 07, 2026. DOI: 10.1590/1983-40632026v5683957.

[2] University of Jember, Faculty of Mathematics and Natural Sciences, Department of Biology, Jember, East Java, Indonesia.
E-mail/ORCID: wakiatil@gmail.com/0009-0006-4555-6813; fuad.fmipa@unej.ac.id/0000-0002-1969-6322; msuudi.fmipa@unej.ac.id/0000-0002-5456-4085.

Editor: Luis Carlos Cunha Junior/Data Availability Statement: Research data are only made available by authors upon request.

encodes heat shock protein to maintain protein homeostasis (Akerfelt et al. 2010).

In plants, HSF has a broad spectrum that can be induced under stress, including heat shock protein, reactive-oxygen species mechanisms, chaperones, and other enzymes involved in protective metabolic reactions (Andrási et al. 2021). It also has a structure consisting of DNA-binding domain located in the N-terminal region, an oligomerization domain, a nuclear export signal, a nuclear localization signal, and a C-terminal transcription activation domain (Guo et al. 2016). The DNA-binding and oligomerization domain are the most conserved parts of the structure (Huang & Feng 2014). The DNA-binding domain will recognize and bind the heat shock element motif on the target gene promoter (Guo et al. 2016).

HSF can regulate cell homeostasis by regulating gene expression and controlling the heat stress response in the environment (Nishizawa-yokoi et al. 2011), and plays an important role in the biological and abiotic stress response in plants. In addition, it regulates the transcription of rapidly accumulated heat shock proteins in plants under biotic and abiotic stress to aid in protein folding, intracellular distribution, and protein breakdown or inactivation (Wang et al. 2024).

Plants have protection in response to environmental stresses. The regulation is molecularly adjusted by HSF with transduction signals associated with heat stress in plants (Wang et al. 2012). The first report on the role of HSF was conducted on tomato plants under heat stress (Yang et al. 2016). Some reports also show that HSF is involved in plant stress responses, including cold stress, heat stress and tolerance to salinity (Li et al. 2023).

The HSF protein not only controls plant growth and development, but can also increase tolerance to abiotic stress (Wang et al. 2024). Research on the HSF protein in cocoa has not been conducted yet; so, it is necessary to carry out genome and protein analyses to elucidate the mechanism and role of HSF in cocoa. Thus, this study aimed to analyze and identify HSF proteins in *T. cacao*, which play a role in heat stress, through in-silico analysis.

MATERIAL AND METHODS

The research was conducted at the University of Jember, Jember city, Indonesia, from July to September 2025.

The identification of the heat shock factor (*HSF*) gene in cocoa was carried out by BLASTP motif *Arabiopsis thaliana* HSF (AtHSFs: AEK67477.1) against the cocoa genome in the Phytozome 14 database (Goodstein et al. 2012), with annotation version v1.1 or *Theobroma cacao* v1.1, and recorded transcript ID, location, coding sequence length, genome and protein sequences. The location of the chromosomes in *T. cacao* was determined from the NCBI database by writing the transcript ID on each TcHSF protein.

Analysis of the physical properties of cocoa proteins, consisting of molecular weight (kDa), isoelectric point (pI) and grand average of hydropathicity (GRAVY) were analyzed through the ExPASy ProtParam tool software (Gasteiger et al. 2005).

The structure of exons and introns in the cacao *HSF* gene was analyzed using the Gene Structure Display Server 2.0 (Hu et al. 2015), based on the collection of coding sequences and genomic sequences, then the motif analysis on the TcHSF protein sequence using the MOTIF tools of MEME (Bailey et al. 2009). Analysis of cis-acting elements in the promoter region of the cocoa *HSF* gene was carried out using the plantCARE database (Lescot et al. 2002) at 2,000 bp in the upstream section before the start of the codon (ATG), and TBtools II was used for visualization (Chen et al. 2023).

Prediction of sub-celular localization in 26 HSF *T. cacao* proteins was performed using the CELLO v.2.5 software (Yu et al. 2006) and WoLF PSORT (Horton et al. 2007), and shown using Tbtools II to see the location of the proteins (Chen et al. 2023).

Analysis of phylogenetic tree construction was carried out by collecting protein sequences of *T. cacao*, *A. thaliana*, *Oryza sativa*, *Manihot esculenta*, and *Zea mays* from the phytozome database (Goodstein et al. 2012), and alignment was carried out using the ClustalW (Thomson et al. 1994) with MEGA11 software (Tamura et al. 2021). The phylogenetic tree was constructed using the maximum likelihood method, with partial deletion parameters, and the bootstrap method with 1,000 replications. The partial deletion method is used to eliminate gaps that can reduce the accuracy of phylogenetic trees, so that the analyzed information is an informative sequence for use in phylogenetic analysis (Dwivedi & Gadagkar 2009, Seo et al. 2022).

The genomic data of *T. cacao* were obtained in the phytozome database as chromosome location data, and the cocoa phylogenetic results were grouped according to their respective clades, in which the data were divided into five groups. The data were processed using Tbtools II (Chen et al. 2023) to obtain the chromosomal length of the cocoa protein. Chromosomal localization analysis was carried out using the Phenogram Ritchie Lab (Wolfe et al. 2013) by entering chromosome location and length data.

RESULTS AND DISCUSSION

Cocoa presented a total of 26 *HSF* genes collected from the database. The *HSF* cocoa gene is sorted by chromosome number (Table 1). Gene identification aims to ensure that the analyzed HSF protein is appropriate. The coding sequence size for *TcHSF* ranged from 702 bp (*TcHSF-17*) to 2,040 bp (*TcHSF-19*). The average amount of TcHSF amino acid was 388.1 aa. The amount of amino acids is greater than that of *A. Thaliana*, which has an average count of 368.0 bp (Li et al. 2023). The cocoa molecular weight ranges from 25.40 to 77.54 kDA, whereas its isoelectric point varies between 4.59 and 8.49. All GRAVY scores on TcHSF have a negative value, so it has a better stability under abiotic stress conditions, and is not easily degraded during drought stress. A negative GRAVY value indicates hydrophilic properties (Balchin et al. 2016). The hydrophilic properties of HSF proteins facilitate in the isolation process, so that all HSF proteins in cocoa have the potential to be isolated as candidates in *TcHSF* gene selection (Jaspard & Hunault 2014).

Analysis of the *TcHSF* gene structure showed that all *TcHSF* have a complete structure in the form of upstream, intron, exon, and downstream (Figure 1). The structure of genes varies in number and length. The longest gene is owned by *TcHSF-7*, *TcHSF-8*, and *TcHSF-14*, with a length of 7 kb, whereas the shortest gene belongs to *TcHSF-21*, with a length of 1.8 kb.

Table 1. List of heat shock factors gene family in *Theobroma cacao* L.

Name	Transcript ID	Nº CDS (bp)	Nº amino acids (AA)	Molecular weight (kDA)	pI	GRAVY	Protein localization prediction
TcHSF-1	Thecc1EG004275t3	996	332	38.10	5.41	-0.876	Nuclear
TcHSF-2	Thecc1EG004275t1	1,089	363	41.68	5.16	-0.901	Nuclear
TcHSF-3	Thecc1EG001784t1	1,344	448	51.15	5.96	-0.748	Nuclear
TcHSF-4	Thecc1EG000610t1	900	300	33.29	5.32	-0.845	Nuclear
TcHSF-5	Thecc1EG004275t8	849	283	32.34	4.67	-0.858	Nuclear
TcHSF-6	Thecc1EG011279t1	1,263	421	48.10	4.89	-0.640	Nuclear
TcHSF-7	Thecc1EG016563t2	1,587	529	57.81	4.91	-0.644	Nuclear
TcHSF-8	Thecc1EG016563t1	1,587	529	57.81	4.91	-0.644	Nuclear
TcHSF-9	Thecc1EG004275t4	996	332	38.10	5.41	-0.876	Nuclear
TcHSF-10	Thecc1EG016198t1	945	315	34.67	4.76	-0.629	Nuclear
TcHSF-11	Thecc1EG015119t1	972	324	36.00	5.27	-0.669	Nuclear
TcHSF-12	Thecc1EG016198t2	705	235	25.40	4.76	-0.624	Nuclear
TcHSF-13	Thecc1EG020700t1	1,083	361	41.61	5.04	-0.816	Nuclear
TcHSF-14	Thecc1EG020222t1	1,473	491	55.35	5.40	-0.888	Nuclear
TcHSF-15	Thecc1EG025482t1	1,503	501	55.34	5.22	-0.636	Nuclear
TcHSF-16	Thecc1EG022446t1	1,431	477	53.32	4.93	-0.577	Nuclear; cytoplasmic
TcHSF-17	Thecc1EG024565t1	702	234	26.76	8.27	-0.736	Nuclear
TcHSF-18	Thecc1EG029630t1	1,218	406	45.73	5.07	-0.760	Nuclear
TcHSF-19	Thecc1EG029886t1	2,040	680	77.54	8.49	-0.524	Cytoplasmic; nuclear
TcHSF-20	Thecc1EG033816t1	1,086	362	39.85	7.72	-0.419	Nuclear
TcHSF-21	Thecc1EG037888t1	1,203	341	38.14	5.82	-0.578	Nuclear
TcHSF-22	Thecc1EG037308t1	1,176	392	43.95	4.68	-0.580	Nuclear
TcHSF-23	Thecc1EG037308t4	1,200	400	44.98	4.75	-0.578	Nuclear
TcHSF-24	Thecc1EG037635t1	1,413	471	53.04	4.73	-0.754	Nuclear
TcHSF-25	Thecc1EG040889t1	840	280	32.37	8.46	-0.711	Extracellular; nuclear
TcHSF-26	Thecc1EG037308t2	852	284	32.41	4.59	-0.662	Nuclear; cytoplasmic

CDS: coding sequences; pI: isoelectric point; GRAVY: grand average of hydropathicity.

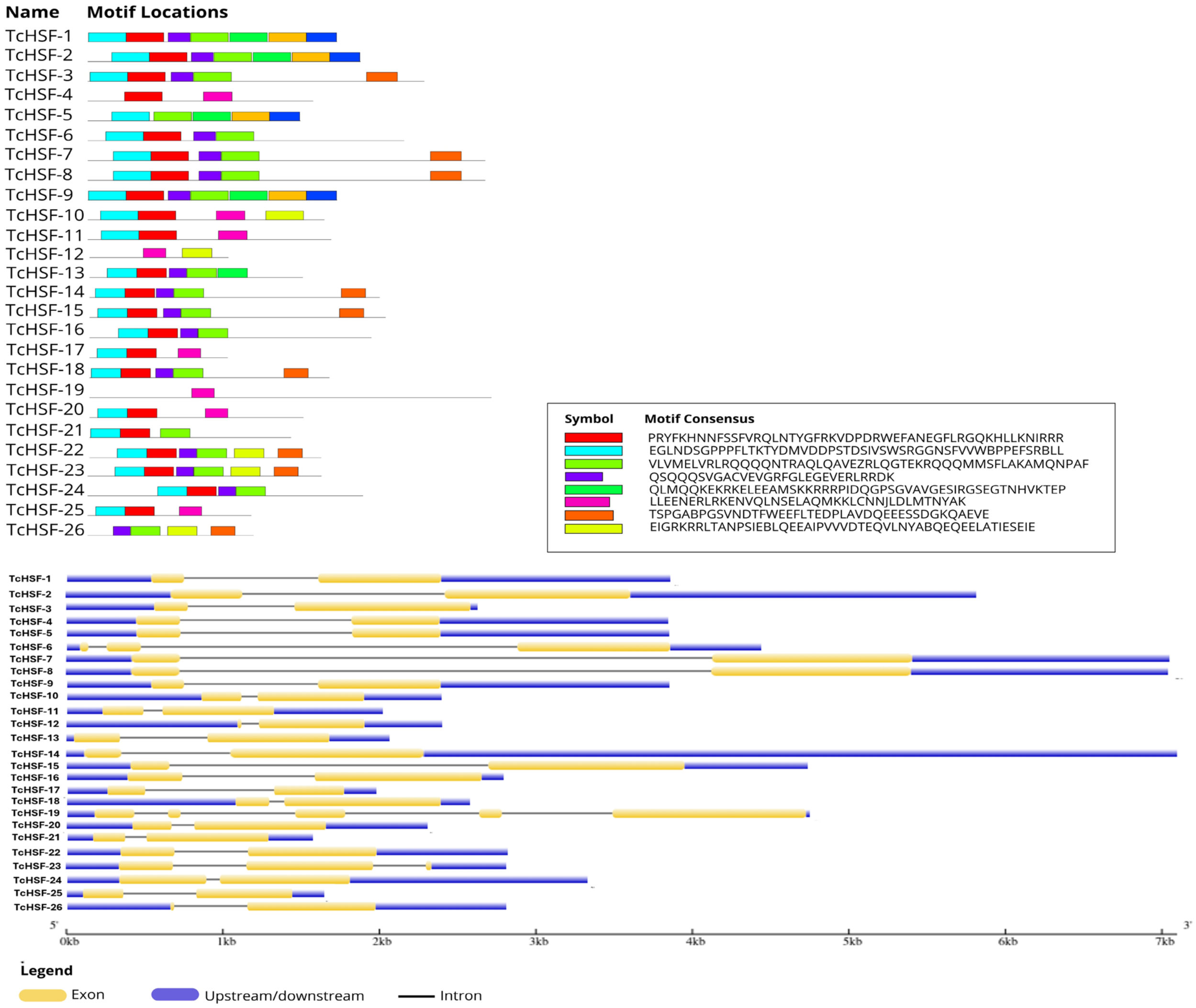

Figure 1. Motif and gene structure of HSF in *Theobroma cacao* (red and green symbols indicate the DNA-binding domain).

Introns are the part of genes that do not encode proteins (Zhang et al. 2022), and variations in the number of introns and exons affect gene expression patterns (Sajad et al. 2022). A smaller number of introns can speed up the transcription and processing process of mRNA, so that genes that respond to stress can be expressed more quickly (Jeffares et al. 2008). The analysis of the *T. cacao* motif varies where the red and green sequence motifs are present in all TcHSF proteins, so that this motif is part of the DNA-binding domain (DBD) that is very structured and located near the N-terminal of all HSFs (Scharf et al. 2011). The presence of DBD is essential for the identification of HSF in cocoa. The DNA-binding domain section plays an important role in HSF transcription regulation. This domain is a conservative region that has a role in the specific binding process on the heat shock element in DNA (Feng et al. 2021).

Promoters are DNA sequences that have several cis-acting elements and are located upstream of the coding area (Bilas et al. 2016). The abscisic acid responsive element is one of the elements that plays a role in regulating gene expression in response to abiotic stress, including heat stress, and is activated by the abscisic acid hormone, which is a hormone that accumulates in plants when experiencing extreme conditions such as drought, thus helping in plant tolerance to heat stress (Aslam et al. 2022). Abscisic acid can improve heat stress tolerance through HSF and heat shock protein regulation (Li et al. 2021). Therefore, the presence of the abscisic acid responsive element in the promoter of the *HSF* gene is important for responding to stress through the abscisic acid signaling pathway. The anaerobic responsive element functions to regulate gene expression related to anaerobic responses (Shariatipour & Heidari 2020).

The region is a specific binding site for proteins involved in transcription initiation and regulation (Hernandez-Garcia & Finer 2014). The cis-element that is in the promoter region also regulates gene expression related to metabolic pathways (Bilas et al. 2016). Some cis-acting elements have a role in heat stress. The heat shock element is the cis-acting part involved in the response to heat stress, whereas the heat shock element can be associated with other elements such as the anaerobic responsive element, abscisic acid responsive element and G-box (Figure 2), because, in the *HSF* gene, it binds to these regions (Chow et al. 2018). The abscisic acid responsive element was found in almost all TcHSF promoters, except for *TcHSF-1*, *TcHSF-2*, *TcHSF-9*, *TcHSF-13*, *TcHSF-16*, and *TcHSF-17*.

Various other elements, such as the low temperature and anaerobic responsive element, and the myeloblastosis binding sequence also will be active as a plant response to stress (Zhang et al. 2022). The anaerobic and low temperature responsive elements were found in almost all TcHSF promoters, whereas the myeloblastosis binding sequence was found only in *TcHSF-6*, *TcHSF-7*, *TcHSF-14*, *TcHSF-24*, and *TcHSF-25*. In addition, the low temperature and anaerobic responsive elements were stress-related, with the first inducing the gene when low temperature stress occured in the plant. G-box regulates cis elements involved in light responsiveness (Shariatipour & Heidari 2020). This element can regulate the adaptation of plants to high light intensity (Kaur et al. 2017). The myeloblastosis binding sequence is the binding part of the Myeloblastosis (MYB) involved in drought induction (Shariatipour & Heidari 2020). The presence of these elements is important to respond to stress from environmental stimuli.

HSF, as a transcription factor, has a subcellular localization in the nucleus. The distribution pattern and subcellular localization of HSFs in plants are determined by the presence of nuclear localization and nuclear export signals, with both signals functioning as regulators of the entry and exit of proteins to the nucleus and back to the cytoplasm (Miller & Mittler 2006).

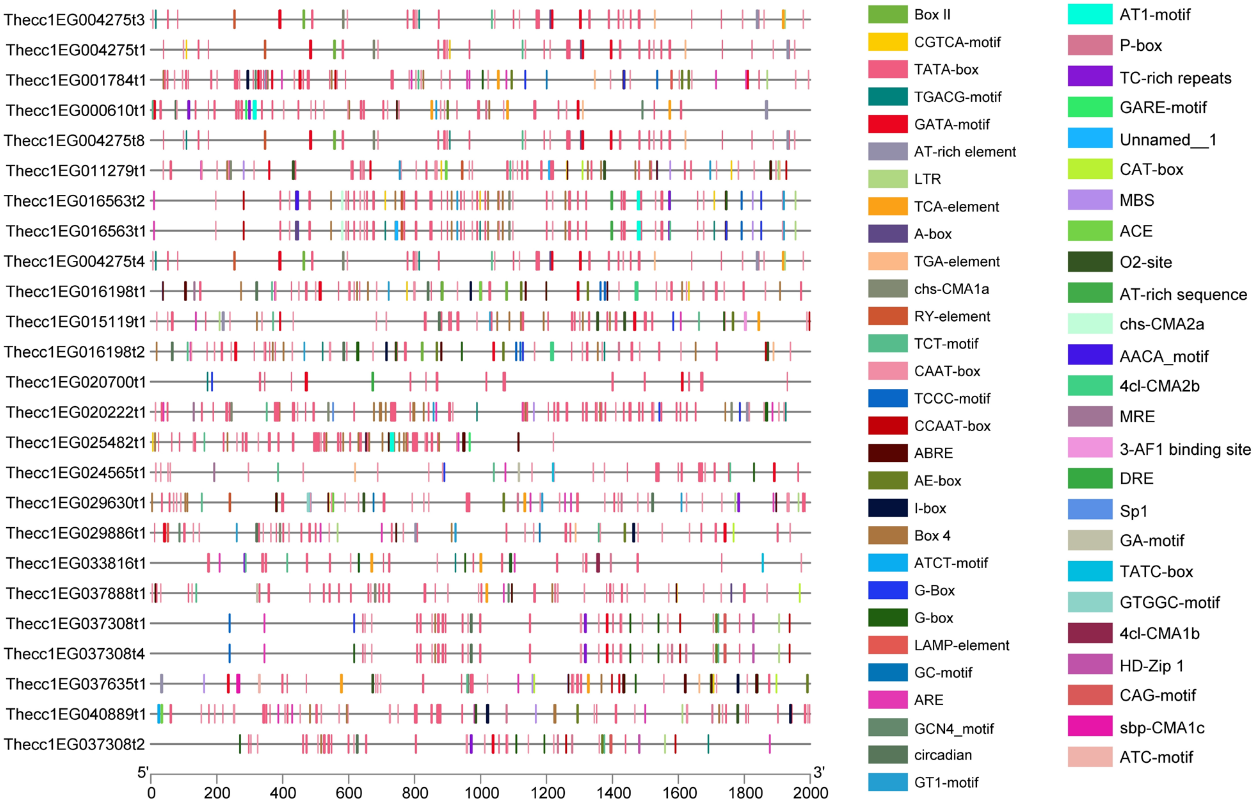

Figure 2. Cis-acting element promoter of HSF *Theobroma cacao* L.

HSFs in *T. cacao* plants showed varying subcellular localization. Most TcHSF exhibited a strong level of subcellular localization in the nucleus (Figure 3). The intensity of the darker color indicates a strong localization of the cell, and the quantitative value (0-14) indicates the number of signals or cumulative scores of the localization of each TcHSF. This is in line with a previous research on *Verbena bonariensis* plants, which stated that the subcellular localization of the HSF protein is in the nucleus (Yang et al. 2024). The results of subcellular localization in *TcHSF-12* showed a difference in strong signals in cytoplasm (cyto) and chloroplast (chlo). This can happen due to the absence of important domains such as nuclear localization signal and nuclear export signal to carry proteins into the nucleus (Zhang et al. 2020). The presence of subcellular localization in the nucleus shows that TcHSF plays a role as the main regulator during stress, so that it translocates to the nucleus and controls several proteins involved in plant stress, such as the heat shock proteins 70 and 90 (Akerfelt et al. 2010). The TcHSF located in the cytoplasm indicates an inactive and non-binding state of heat shock element (Scharf et al. 2011).

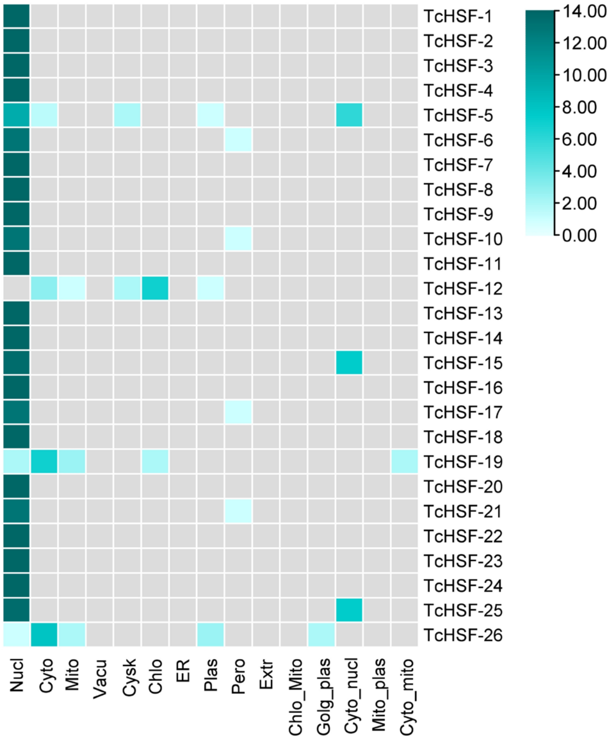

Figure 3. Subcellular localization of HSF *Theobroma cacao* (darker colors indicate a stronger localization).

The phylogenetic tree construction HSF in cocoa was formed through the alignment of HSF protein sequences that had been identified from several plants, such as *A. thaliana* with 21 sequences, *O. sativa* with 25 sequences, *M. esculenta* with 30 sequences (Zeng et al. 2020), and *Z. mays* with 30 sequences (Zhang et al. 2022) (Figure 4). HSF has been widely expressed in dry stress tolerance in *A. thaliana* and *O. sativa* (Guo et al. 2008). Arabidopsis is a model plant that is used as a comparison to analyze phylogenetics and *HSF* gene expression. HSF in *A. thaliana* has a central role in helping to maintain protein homeostasis during stressful conditions such as heat stress and drought (Li et al. 2014). Based on phylogenetic results, *T. cacao* tends to be evolutionarily closer to *Manihot esculenta*, such as Thecc1EG02022t1, which has a proximity to Manes.15G057200, with a bootstrap value of 94 %. The phylogenetic analysis of *T. cacao* also showed a close relationship with *A. Thaliana*. For example, *T. cacao* (Thecc1EG037635t1) is in the same clade as *A. thaliana* (At5G54070), with a bootstrap value of 90 %. High bootstrap results in HSF *T. cacao* and other species indicate similarity in biological function with other previously studied species such as *A. thaliana* (Li et al. 2014).

TcHSF was divided into 5 groups, based on phylogeny analysis. A total of 26 *TcHSF* genes were spread across 9 chromosomes (Figure 5), with an uneven distribution. Chromosome 7 was the only one where the *HSF* gene does not exist. The number of genes ranged from 1 to 6 on each chromosome, with 1 and 9 showing the highest number of genes with 6 *TcHSF* genes. This distribution pattern may indicate that the *TcHSF* gene has evolved structurally and functionally. The presence of a variety of different chromosomes signals the importance of the *HSF* gene in the plant's defence mechanism against abiotic stress. The *HSF* gene, that is spread across various chromosomes, shows that HSF is spread throughout the genome and has a function as the main regulator that will activate several proteins in protecting cells from various stresses such as drought, heat, and salt stress (Guo et al. 2016). Group 1 had the most spread on chromosomes. Genes belonging to group 1 on chromosome 9 (*TcHSF-21*, *TcHSF-22*, and *TcHSF-23*) are likely to experience tandem duplication (Panchy et al. 2016). This duplication is characterized by the location of chromosomes that are close to each other in a single

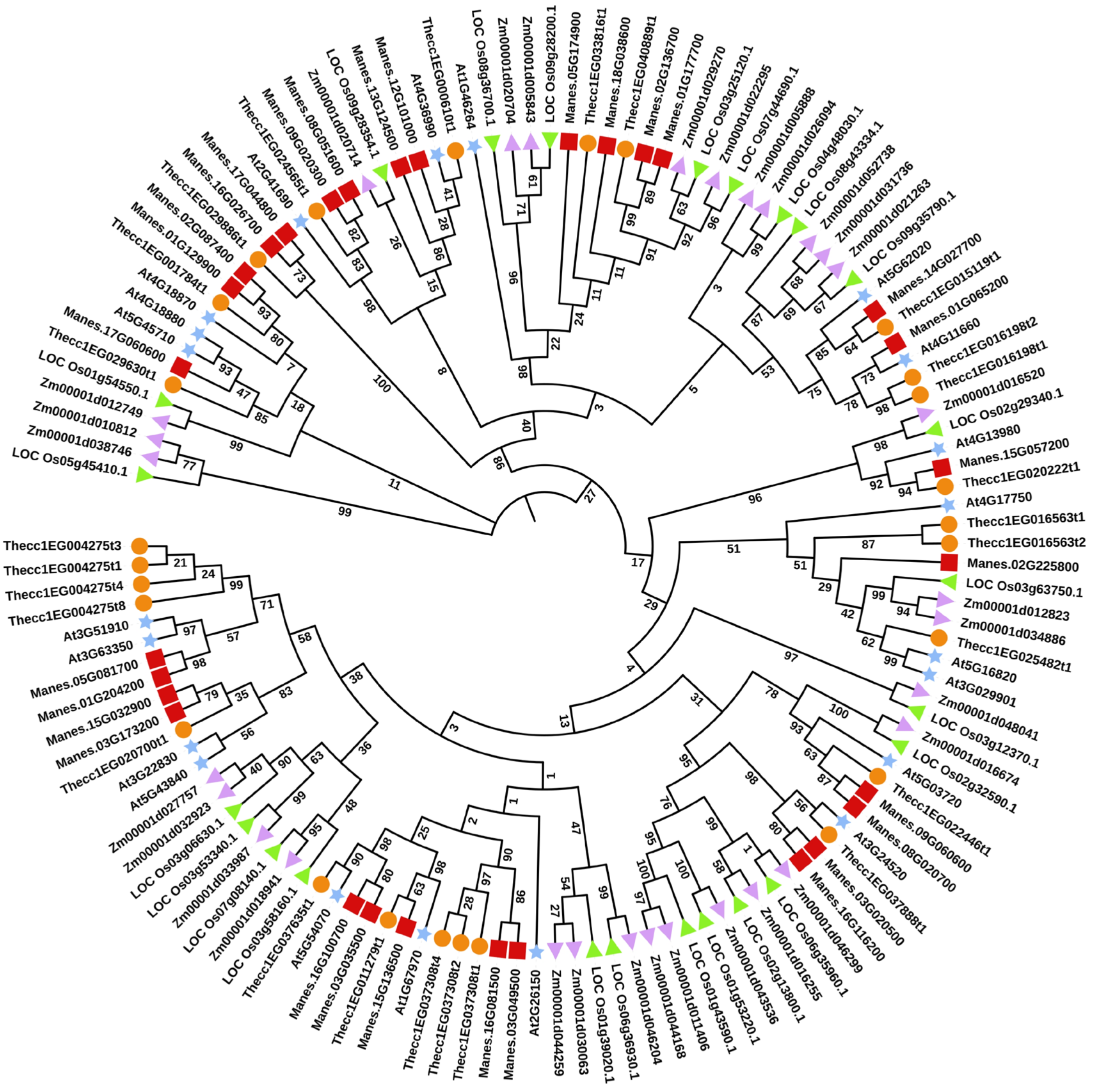

Figure 4. Phylogenetic HSF protein in *Theobroma cacao* (Thecc), *Arabidopsis thaliana* (At), *Manihot esculenta* (Manes), *Oryza sativa* (Loc), and *Zea mays* (Zm).

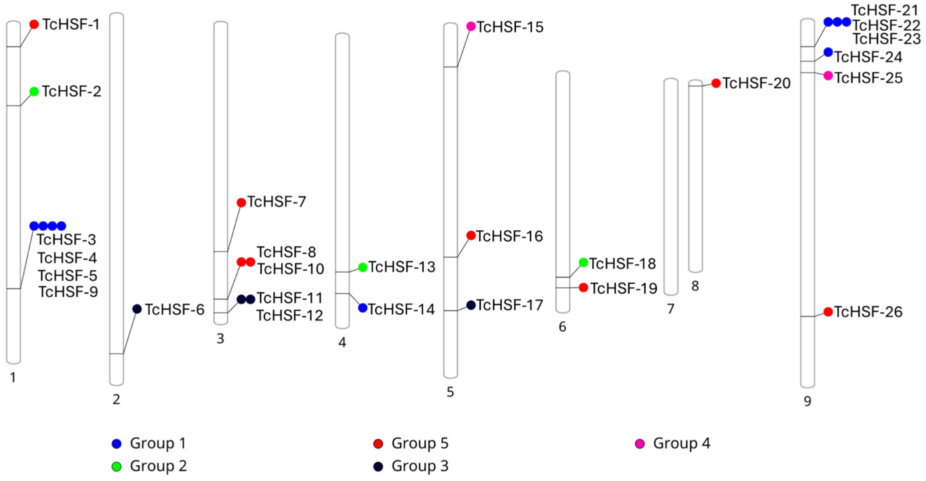

Figure 5. Location of the TcHSF gene in *Theobroma cacao* chromosomes.

chromosome, which can occur through uneven crosses (Qiao et al. 2019).

The presence of HSF plays a very important role in the regulation of plants under abiotic stress. HSF analysis has been performed on several plants such as tomato, using qPCR, which suggests that the *Solanum lycopersicum* HSF (*SlyHSF*) gene plays a role in the heat resistance pathway (Yang et al. 2024). HSF in *Phaseolus vulgaris* (*PvHSF*) plants plays a role in regulating abiotic stress (Zhang et al. 2022). In-silico analysis of HSF cocoa is needed as an analytical study for the selection of *TcHSF* gene candidates related to the adaptability of cocoa to heat stress conditions, so that it can be used as information for the development of new cocoa cultivars that are resistant to heat stress. The selection of *TcHSF* gene candidates was carried out by selecting the 26 most potential *TcHSF* genes for further analysis. Gene selection includes the presence of DNA-binding domain (DBD) motifs, the least number of intins, the presence of elements in the promoter region, subcellular localization in the nucleus and proximity to *HSF* genes of other plants such as *A. Thaliana*, as seen from the results of the phylogenetic analysis. Based on these criteria, there are three potential *TcHSF* gene candidates, namely *TcHSF-14*, *TcHSF-24*, and *TcHSF-25*. All three genes have a DNA-binding domain and a total of 1 intron. Based on the analysis of cis-elements, the gene has four elements that are responsive to stress, such as anaerobic, abscisic acid, and low temperature responsive elements, and myeloblastosis binding sequence, when compared to other genes that only have two or three elements. Analysis of the subcellular localization of gene products in the nucleus also supports the function of the *HSF* gene, and phylogenetic results show that the three genes are in the same clade as several previously studied species, such as *A. thaliana*.

CONCLUSIONS

1. A total of 26 heat shock factors (HSF) were identified for *Theobroma cacao*;
2. Almost all TcHSF proteins have subcellular localization in the nucleus, which corresponds to the function of HSF as a coder of the heat shock protein and the expression of genes that play a role in stress tolerance;
3. TcHSF has a close affinity with the heat shock factors of *Arabiopsis thaliana* and *Manihot esculenta*;
4. Analysis of cis-acting and phylogenetic elements showed that TcHSF is regulated in response to various abiotic stresses, including light, drought and low temperatures.

REFERENCES

AKERFELT, M.; MORIMOTO, R. I.; SISTONEN, L. Heat shock factors: integrators of cell stress, development and lifespan. *Nature Reviews: Molecular Cell Biology*, v. 11, n. 8, p. 545-555, 2010.

ANDRÁSI, N.; PETTKÓ-SZANDTNER, A.; SZABADOS, L. Diversity of plant heat shock factors: regulation, interactions, and functions. *Journal of Experimental Botany*, v. 72, n. 5, p. 1558-1575, 2021.

ASLAM, M. M.; WASEEM, M.; JAKADA, B. H.; OKAL, E. J.; LEI, Z. L.; SAQIB, H. S. S.; YUAN, W.; XU, W.; ZHANG, Q. Mechanisms of abscisic acid-mediated drought stress responses in plants. *International Journal of Molecular Sciences*, v. 23, n. 3, e1084, 2022.

BADAN PUSAT STATISTIK (BPS). *Statistik kakao Indonesia 2023*. Jakarta: BPS, 2025.

BAILEY, T. L.; BODEN, M.; BUSKE, F. A.; FRITH, M.; GRANT, C. E.; CLEMENTI, L.; REN, J.; LI, W. W.; NOBLE, W. S. *MEME SUITE*: tools for motif discovery and searching. 2009. Available at: https://meme-suite.org. Access on: July 21, 2025.

BALCHIN, D.; HAYER-HARTL, M.; HARTL, F. U. *In vivo* aspects of protein folding and quality control. *Science*, v. 353, e6294, 2016.

BILAS, R.; SZAFRAN, K.; HNATUSZKO-KONKA, K. Cis-regulatory elements used to control gene expression in plants. *Plant Cell, Tissue and Organ Culture*, v. 127, n. 2, p. 269-287, 2016.

CHEN, C.; WU, Y.; LI, J.; WANG, X.; ZENG, Z.; XU, J.; LIU, Y.; FENG, J.; CHEN, H.; HE, Y.; XIA, R. TBtools-II: a “one for all, all for one” bioinformatics platform for biological big-data mining. *Molecular Plant*, v. 16, n. 11, p. 1733-1742, 2023.

CHOW, C. N.; CHIANG-HSIEH, Y.; CHIEN, C.; ZHENG, H.; LEE, T.; WU, N.; TSENG, K.; HOU, P.; CHANG, W. Delineation of condition specific cis- and trans-acting elements in plant promoters under various endo- and exogenous stimuli. *BMC Genomics*, v. 19, e85, 2018.

DWIVEDI, B.; GADAGKAR, S. R. Phylogenetic inference under varying proportions of indel-induced alignment gaps. *BMC Evolutionary Biology*, v. 9, e211, 2009.

FENG, N.; FENG, H.; WANG, S.; PUNEKAR, A. S.; LADENSTEIN, R.; WANG, D.; ZHANG, Q.; DING, J.;

LIU, W. Structures of heat shock factor trimers bound to DNA. *Iscience*, v. 24, n. 9, e102951, 2021.

GASTEIGER, E.; HOOGLAND, C.; GATTIKER, A.; DUVAUD, S.; WILKINS, M. R.; APPEL, R. D.; BAIROCH, A. *Protein analysis tools on the ExPASy server*. Totowa: Humana Press, 2005.

GOODSTEIN, D. M.; SHU, S.; HOWSON, R.; NEUPANE, R.; HAYES, R. D.; FAZO, J.; MITROS, T.; DIRKS, W.; HELLSTEN, U.; PUTNAM, N.; ROKHSAR, D. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, v. 40, n. 1, p. 1178-1186, 2012.

GUO, J.; WU, J.; JI, Q.; WANG, C.; LUO, L.; YUAN, Y.; WANG, Y.; WANG, J. Genome-wide analysis of heat shock transcription factor families in rice and *Arabidopsis*. *Journal of Genetics and Genomics*, v. 35, n. 2, p. 105-118, 2008.

GUO, M.; LIU, J.; MA, X.; LUO, D.; GONG, Z.; LU, M. The plant heat stress transcription factors (HSFS): structure, regulation, and function in response to abiotic stresses. *Frontiers in Plant Science*, v. 7, e114, 2016.

HERNANDEZ-GARCIA, C. M.; FINER, J. J. Plant science identification and validation of promoters and cis -acting regulatory elements. *Plant Science*, v. 217-218, n. 1, p. 109-119, 2014.

HORTON, P.; PARK, K.; OBAYASHI, T.; FUJITA, N.; HARADA, H.; ADAMS-COLLIER, C. J.; NAKAI, K. WoLF PSORT: protein localization predictor. *Nucleic Acids Research*, v. 35, suppl. 2, p. 585-587, 2007.

HU, B.; JIN, J.; GUO, A.; ZHANG, H.; LUO, J.; GAO, G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, v. 31, n. 8, p. 1296-1297, 2015.

HUANG, Y.; FENG, M. L. Heat shock factors in carrot: genome-wide identification, classification, and expression profiles response to abiotic stress. *Molecular Biology Reports*, v. 42, n. 5, p. 893-905, 2014.

IZZAH, N.; DAMAYANTI, D. Pengaruh jumlah produksi dan harga terhadap nilai ekspor kakao Indonesia Tahun 2017-2020. *Transparansi*, v. 6, n. 1, p. 78-85, 2023.

JASPARD, E.; HUNAULT, G. Comparison of amino acids physico-chemical properties and usage of late embryogenesis abundant proteins, hydrophilins and WHy domain. *PLoS One*, v. 9, n. 10, e109570, 2014.

JEFFARES, D. C.; PENKETT, C. J.; BÄHLER, J. Rapidly regulated genes are intron poor. *Trends in Genetics*, v. 24, n. 8, p. 375-378, 2008.

KAUR, A.; PATI, P. K.; PATI, A. M.; NAGPAL, A. K. In-silico analysis of cis-acting regulatory elements of pathogenesis-related proteins of *Arabidopsis thaliana* and *Oryza sativa*. *PLoS One*, v. 12 , n. 9, e0184523, 2017.

KEHINDE, A. D. Access to trade credit and its impact on the use of European Union (EU) approved pesticides among smallholder cocoa farmers in Ondo state, Nigeria. *Heliyon*, v. 8, n. 12, e12409, 2022.

KONGOR, J. E.; OWUSU, M.; ODURO-YEBOAH, C. Cocoa production in the 2020s: challenges and solutions. *CABI Agriculture and Bioscience*, v. 5, e102, 2024.

LESCOT, M.; DÉHAIS, P.; THIJS, G.; MARCHAL, K.; MOREAU, Y.; PEER, Y. V.; ROUZÉ, P.; ROMBAUTS, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, v. 30, n. 1, p. 325-327, 2002.

LI, L.; JU, Y.; ZHANG, C.; TONG, B.; LU, Y.; XIE, X.; LI, W. Genome-wide analysis of the heat shock transcription factor family reveals saline-alkali stress responses in *Xanthoceras sorbifolium*. *PeerJ*, v. 11, e15929, 2023.

LI, N.; EURING, D.; CHA, J. Y.; LIN, Z.; LU, M.; HUANG, L.; KIM, W. Plant hormone-mediated regulation of heat tolerance in response to global climate change. *Frontiers in Plant Science*, v. 11, e627969, 2021.

LI, P. S.; YU, T.; HE, G.; CHEN, M.; ZHOU,Y.; CHAI, S.; XU, Z.; MA, Y. Genome-wide analysis of the Hsf family in soybean and functional identification of GmHsf-34 involvement in drought and heat stresses. *BMC Genomics*, v. 15, n. 1, e1009, 2014.

MILLER, G.; MITTLER, R. Could heat shock transcription factors function as hydrogen peroxide sensors in plants? *Annals of Botany*, v. 98, n. 2, p. 279-288, 2006.

NIEVES-ORDUÑA, H. E.; KRUTOVSKY, K. V.; GAILING, O. Geographic distribution, conservation, and genomic resources of cacao *Theobroma cacao* L. *Crop Science*, v. 63, n. 4, p. 1750-1778, 2023.

NISHIZAWA-YOKOI, A.; NOSAKA, R.; HAYASHI, H.; TAINAKA, H.; MARUTA, T.; TAMOI, M.; IKEDA, M.; OHME-TAKAGI, M.; YOSHIMURA, K.; YABUTA, Y.; SHIGEOKA, S. HsfA1d and HsfA1e involved in the transcriptional regulation of HsfA2 function as key regulators for the Hsf signaling network in response to environmental stress. *Plant and Cell Physiology*, v. 52, n. 5, p. 933-945, 2011.

PANCHY, N.; LEHTI-SHIU, M.; SHIU, S. Evolution of gene duplication in plants. *Plant Physiology*, v. 171, n. 4, p. 2294-2316, 2016.

QIAO, X.; LI, Q.; YIN, H.; QI, K.; LI, L.; WANG, R.; ZHANG, S.; PATERSON, A. H. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biology*, v. 20, n. 1, p. 38-60, 2019.

SAJAD, S.; JIANG, S.; ANWAR, M.; DAI, Q.; LUO, Y.; HASSAN, M. A.; TETTEH, C.; SONG, J. Genome-wide study of Hsp90 gene family in cabbage (*Brassica oleracea* var. *capitata* L.) and their imperative roles in response to cold stress. *Frontiers in Plant Science*, v. 13, e908511, 2022.

SCHARF, K.; BERBERICH, T.; EBERSBERGER, I.; NOVER, L. The plant heat stress transcription factor (Hsf) family: structure, function and evolution. *Biochimica et Biophysica Acta*, v. 1819, n. 2, p. 104-119, 2011.

SEO, T.-K.; GASCUEL, O.; THORNE, J. Measuring phylogenetic information of incomplete sequence data. *Systematic Biology*, v. 71, n. 3, p. 630-648, 2022.

SHAMSHAD, A.; RASHID, M.; ZAMAN, Q. U. In-silico analysis of heat shock transcription factor (OsHSF) gene family in rice (*Oryza sativa* L.). *BMC Plant Biology*, v. 23, n. 1, p. 395-419, 2023.

SHARIATIPOUR, N.; HEIDARI, B. Meta-analysis of expression of the stress tolerance associated genes and uncover their cis-regulatory elements in rice (*Oryza sativa* L.). *The Open Bioinformatics Journal*, v. 13, n. 1, p. 39-49, 2020.

TAMURA, K.; STECHER, G.; KUMAR, S. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, v. 38, n. 7, p. 3022-3027, 2021.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, v. 22, n. 22, p. 4673-4680, 1994.

WANG, F.; DONG, Q.; JIANG, H.; ZHU, S.; CHEN, B.; XIANG, Y. Genome-wide analysis of the heat shock transcription factors in *Populus trichocarpa* and *Medicago truncatula*. *Molecular Biology Reports*, v. 39, n. 2, p. 1877-1886, 2012.

WANG, N.; SHU, X.; ZHANG, F.; SONG, G.; WANG, Z. Characterization of the heat shock transcription factor family in *Lycoris radiata* and its potential roles in response to abiotic stresses. *Plants*, v. 13, n. 2, p. 271-290, 2024.

WOLFE, D.; DUDEK, S.; RITCHIE, M. D.; PENDEGRASS, S. A. Visualizing genomic information across chromosomes with PhenoGram. *BioData Mining*, v. 6, n. 1, e18, 2013.

YANG, X.; WANG, S.; CAI, J.; ZHANG, T.; YUAN, D.; LI, Y. Genome-wide identification and expression analysis of the HSF gene family in *Ammopiptanthus mongolicus*. *BMC Genomics*, v. 25, e729, 2024.

YANG, X.; ZHU, W.; ZHANG, H.; LIU, N.; TIAN, S. Heat shock factors in tomatoes: genome-wide identification, phylogenetic analysis and expression profiling under development and heat stress. *PeerJ*, v. 4, n. 1, e1961, 2016.

YU, C.-S.; CHEN, Y.-C.; LU, C.-H.; HWANG, J.-K. Prediction of protein subcellular localization. *Proteins: Structure, Function, and Bioinformatics*, v. 64, n. 3, p. 643-651, 2006.

ZENG, J.; WU, C.; WANG, C.; LIAO, F.; MO, J.; DING, Z.; TIE, W.; YAN, Y.; HU, W. Genomic analyses of heat stress transcription factors (HSFs) in simulated drought stress response and storage root deterioration after harvest in cassava. *Molecular Biology Reports*, v. 47, n. 8, p. 5997-6007, 2020.

ZHANG, Q.; GENG, J.; DU, Y.; ZHAO, Q.; ZHANG, W.; FANG, Q.; YIN, Z.; LI, J.; YUAN, X.; FAN, Y.; CHENG, X.; DU, J. Heat shock transcription factor (Hsf) gene family in common bean (*Phaseolus vulgaris*): genome-wide identification, phylogeny, evolutionary expansion and expression analyses at the sprout stage under abiotic stress. *BMC Plant Biology*, v. 22, n. 1, p. 33-47, 2022.

ZHANG, X.; XU, W.; NI, D.; WANG, M.; GUO, G. Genome-wide characterization of tea plant (*Camellia sinensis*) Hsf transcription factor family and role of CsHsfA2 in heat tolerance. *BMC Plant Biology*, v. 20, n. 1, p. 244-260, 2020.