RESEARCH ARTICLE

Genome-wide Identification and Evolutionary Characterization of the Heat Shock Transcription Factor (HSF) Gene Family in *Castanea Dentata* and Their Functional Implications

Kanza Rehman[1], Maryam Noor[1], Iqra Mehar[1], Bukhtawer Nasir[2] and Hadia Hussain[1,3,*]

[1] Department of Biotechnology, University of Okara, Okara 56300, Pakistan
[2] Department of Biochemistry and Biotechnology, The University of Faisalabad, Faisalabad 38000, Pakistan
[3] State Key Laboratory of Tree Genetics and Breeding, College of Life Sciences, Nanjing Forestry University, Nanjing 210037, China

Abstract

Heat shock transcription factors (HSFs) act as transcriptional regulators that mediate the defense responses against heat stress. *Castanea dentata* plays a major role in sustaining forest and landscape ecosystems and also exhibits medicinal properties. In this study, we identified 23 *CadenHSFs* genes from the genome of *C. dentata*. In silico approaches were employed to analyze gene structure, conserved domains, chromosomal locations, motif organization, comparative evolutionary analyses, synteny analyses, subcellular localization, and cis-regulatory element analysis. Cis-regulatory element analysis revealed that the promoters of *CadenHSFs* genes are associated with stress responses, light signaling, hormone pathways, and developmental processes. Phylogenetic analysis of the predicted HSF proteins from *C. dentata*, *Arabidopsis thaliana*, and *Vitis vinifera* showed that the 23 *CadenHSFs* genes were classified into seven distinct subgroups and three main classes (A, B, and C). Analyses of conserved domains, motifs, and gene structure indicated that these proteins share conserved characteristics within each group while displaying distinct features between groups. Synteny analysis identified both duplication events (including tandem repeats and segmental duplications). In silico subcellular localization results demonstrated that most *CadenHSFs* proteins were predominately localized in the nucleus. Our findings provide a robust foundation for further functional characterization of *CadenHSFs* genes and the genetic improvement of *C. dentata*.

Keywords: heat shock transcription factor (HSF), *castanea dentata*, genome-wide analysis, abiotic stress, hormone signaling.

Submitted: 18 February 2026
Accepted: 31 March 2026
Published: 07 June 2026

Vol. 1, **No.** 2, 2026.
10.62762/PIJ.2026.449234

***Corresponding author:**
✉ Hadia Hussain
hadiahussain37@outlook.com

Citation
Rehman, K., Noor, M., Mehar, I., Nasir, B., & Hussain, H. (2026). Genome-wide Identification and Evolutionary Characterization of the Heat Shock Transcription Factor (HSF) Gene Family in *Castanea Dentata* and Their Functional Implications. *Plant Innovation Journal*, 1(2), 58–71.

1 Introduction

Abiotic stresses such as cold, salt, heat, and drought cause adverse impact on the development and growth of plants, leading to significant reductions in crop yield [1]. Heat stress in particular disrupts carbohydrate metabolism and storage processes in sensitive crops, further compromising yield quality [2]. Global warming also poses a worldwide risk to various crops. Severe heat stress causes cellular damage and eventually leads to the death of the cell [3]. Heat stress can cause many reactive oxygen species to rise in plant cells and denature many proteins that are sensitive to heat [4]. Therefore, it is highly crucial to enhance the heat resistance of crops via molecular manipulation.

Heat-shock reaction is a temporary reprogramming of cellular activities, characterized by increased expression of heat shock proteins (Hsps). HSPs assemble in a dosage-dependent manner to save cells and impart an increased level of thermotolerance. The primary families of stress proteins found in nearly all species are HSP60, HSP70, HSP90, HSP100, and tiny Hsps, each of which has a few members. The coordinated induction of these chaperone families is tightly regulated at the transcriptional level in response to proteotoxic signals [5].

For the first time, the HSF gene was found in yeast [6]. Now the HSF gene family has been comprehensively categorized in a lot of species, including Chinese cabbage, *Arabidopsis*, pepper, *Triticum* aestivum, maize, rice, and grasses [4, 7]. Furthermore, HSFs have been widely analyzed in tomato [8]. HSF gene families differ notably in various plant species. In chickpea, 20 non-redundant genes encoding the HSF domain were identified [7]. There are 25 members recognized in *Oryza sativa*, 21 HSF members identified in the model plant *Arabidopsis thaliana*, 26 members identified in *Glycine max*, 56 members identified in *Triticum aestivum*, 25 members identified in *Zea mays* [8–12], and 35 members identified in carrot [13]. The largest HSF gene family till now, with 64 HSFs, has been identified in Brassica napus [14].

Their nuts or seeds provide essential nutrition for animals as well as humans. The chromosome-level genome assembly of *Castanea* species has further revealed conserved chromosomal segments and provided a genomic foundation for studying their biological traits [15]. The wood of chestnut is strong, easy to split, has a strong grain, and is highly resistant to decay [16]. Chestnuts also have healthcare medicinal attributes, such as for the development of gluten-free foods for diabetic patients or patients with high cholesterol and celiac disease. They are abundant in antioxidants and are often used as herbal remedies that help prevent chronic diseases such as cardiovascular issues and cancer [17]. They are a good source of vitamin E, lipids, protein, starch, and amino acids [18].

In this study, bioinformatics approaches were fully employed to understand the key information of the heat shock transcription gene family in Castanea, including gene structure, phylogenetic analysis, domain and motif analysis, chromosome location, synteny analysis, subcellular localization, and cis-regulatory elements. Therefore, the results of this study provide valuable insights for heat-resistant crop development (Figure 1).

2 Materials and Methods

2.1 Sequence data retrieval

All protein sequences of *Arabidopsis thaliana* were retrieved from the Plant Transcription Factor Database (PlantTFDB v5.0, http://planttfdb.gao-lab.org/). Genome files for *Castanea spp*. were downloaded from the Phytozome database (Phytozome v13, https://phytozome-next.jgi.doe.gov/). To identify candidate *CadenHSFs* proteins, protein–protein BLAST (BLASTP) searches were performed against the Castanea protein dataset in Phytozome, using *Arabidopsis* HSF protein sequences as queries with the following parameters: E-value $\leq$ 1e-5, identity $\geq$ 30%, query coverage $\geq$ 50%, and default BLOSUM62 substitution matrix. Sequences with significant matches to the corresponding Arabidopsis HSF proteins were retained as candidate *CadenHSFs* members. The physicochemical properties of all identified *CadenHSFs* proteins—including protein length, molecular weight, theoretical isoelectric point (pI), aliphatic index, and grand average of hydropathicity (GRAVY) were computed and analyzed using the ExPASy server (v2025, https://www.expasy.org/).

2.2 Gene structure and Chromosome location analysis

To characterize the gene structures of all identified *CadenHSFs* genes, the Gene Structure Display Server 2.0 (GSDS 2.0, http://gsds.cbi.pku.edu.cn/) was employed. Coding sequences (CDS) were used to generate and visualize exon–intron architectures, as previously described [19]. Basic genomic information, including chromosome length and the start and end positions of *CadenHSFs* genes, was extracted from the Castanea

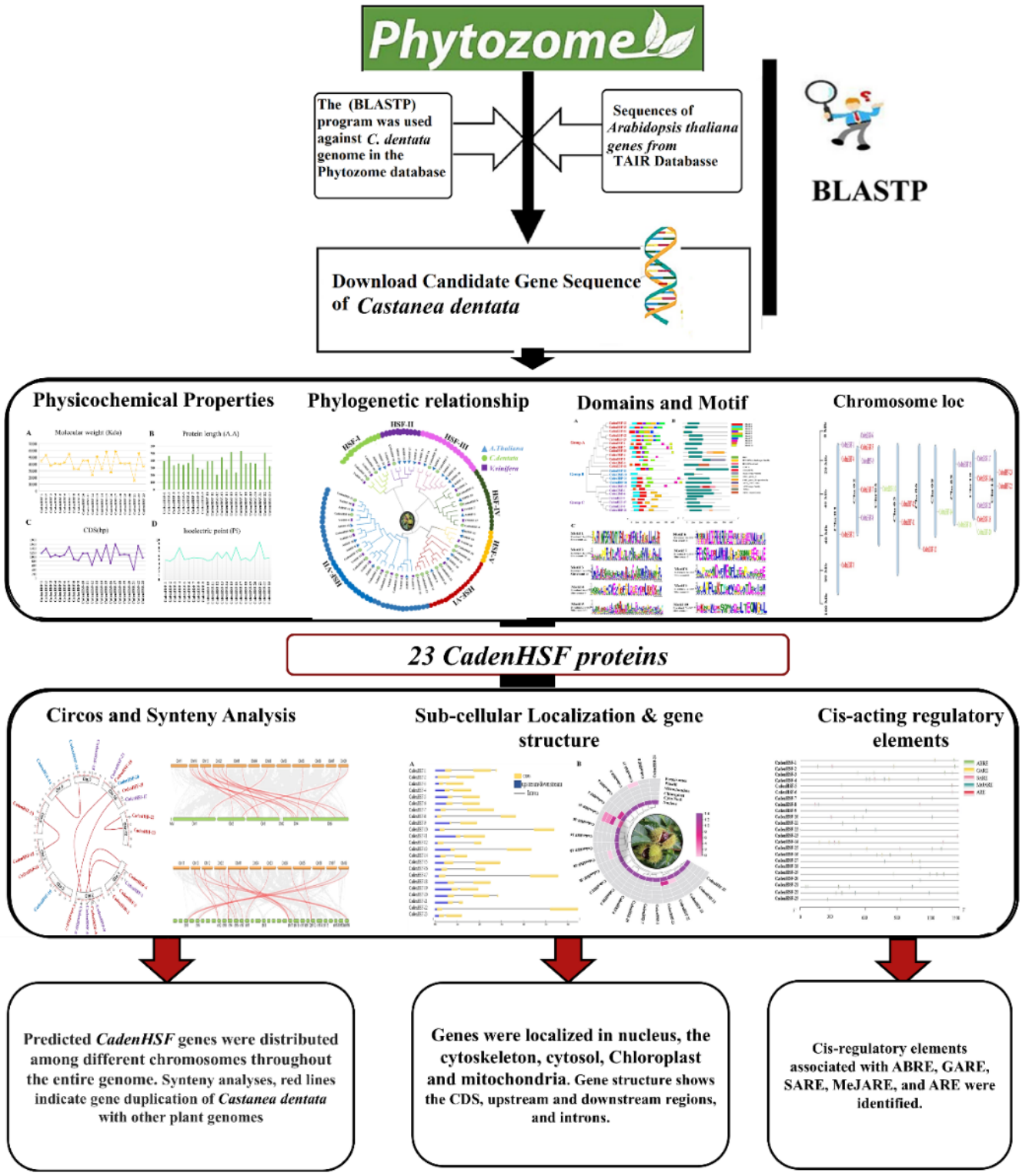

Figure 1. A graphical abstract of integrated bioinformatics approaches of our study to identify the best candidates of *CadenHSFs* genes in *Castanea dentata*. The pipeline includes gene identification using HSF domain profiles, phylogenetic analysis, conserved motif analysis, chromosomal mapping, synteny and duplication analysis, gene structure, subcellular localization prediction, and cis-regulatory element analysis.

reference genome annotation. Graphical visualization of gene distributions was performed using TBtools (v1.120, https://github.com/CJ-Chen/TBtools/releases) based on GTF/GFF annotation files [20].

2.3 Subcellular localization

The subcellular localization of the identified *CadenHSFs* proteins in *Castanea dentata* was predicted using WoLF PSORT (http://wolfpsort.hgc.jp) [21], and

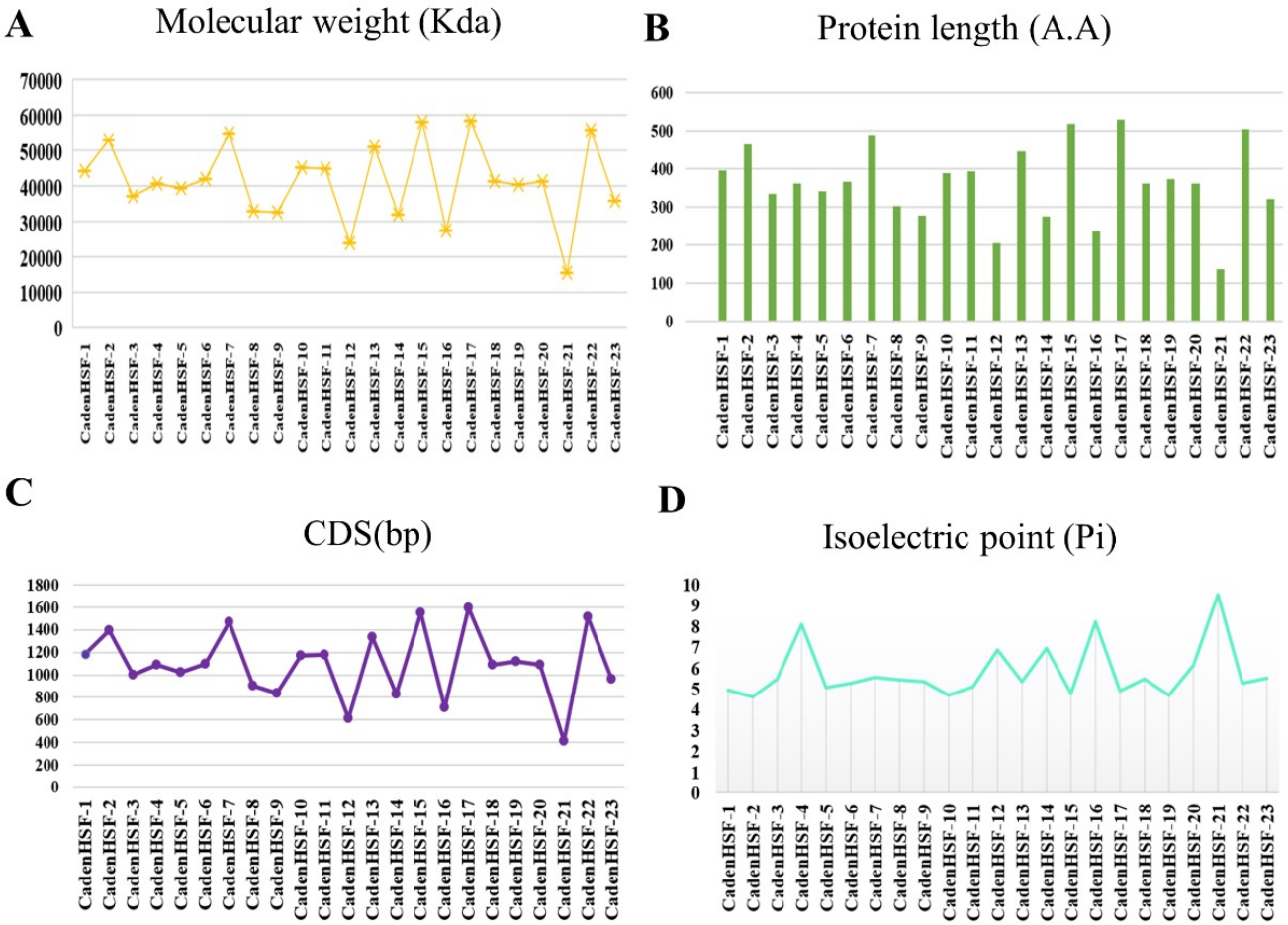

Figure 2. Physicochemical characteristics of *CadenHSFs* proteins in *Castanea dentata*. (A) Distribution of molecular weights, indicating variation in protein size among *CadenHSFs* members. (B) Amino acid length distribution, reflecting structural diversity within the gene family. (C) Coding sequence (CDS) length variation among *CadenHSFs* genes, suggesting differences in gene architecture. (D) Isoelectric point (pI) distribution, showing the range of acidic to basic proteins, which may influence protein stability and subcellular localization. Overall, these variations highlight the structural diversity and potential functional specialization of *CadenHSFs* proteins.

CELLO v.2.5 (http://cello.life.nctu.edu.tw/), with scores ≥ 0.7 indicating high confidence [22].

2.4 Conserved domain and motif analysis

To analyze the functional domains of the identified *CadenHSFs* proteins in *Castanea dentata*, redundant sequences were first removed by discarding incomplete structures, retaining only the longest complete sequence for identical/highly similar entries, excluding sequences with E-value > 1e-5 or low coverage, and removing 100% identical domain duplicates. Conserved domain searches were performed using CDD-Batch at NCBI (https://www.ncbi.nlm.nih.gov/), which supports integrated domain annotation from the Pfam and SMART databases for reliable domain verification. The domain architecture was graphically visualized using TBtools (v1.119, https://github.com/CJ-Chen/TBtools/releases). For conserved motif analysis, the MEME Suite v4.11.1 (http://meme-suite.org/tools/meme) was employed with the following parameters: maximum number of motifs = 10; optimum motif width = 1-50.

2.5 Comparative phylogenetic and synteny analysis

MEGA 12.1 (http://www.megasoftware.net/) was used to analyze the evolutionary relationships of *CadenHSFs* genes with other plant species. A phylogenetic tree was constructed using the Maximum Likelihood method with 1000 bootstrap replicates and visualized via iTOL. A dual synteny plot was generated with MCScanX integrated in TBtools to identify orthologs

of *CadenHSFs* in *Arabidopsis thaliana* and *Vitis vinifera*.

2.6 Cis-regulatory element analysis

To identify cis-regulatory elements of *CadenHSFs* genes, the 1500-bp upstream sequences from the translation start site were retrieved from Phytozome (v13, https://phytozome-next.jgi.doe.gov/). Putative cis-acting elements in the promoter regions were predicted using the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

3 Results

3.1 Identification of CadenHSF family members in *Castanea*

The amino acid sequences of the *Arabidopsis* conserved HSF domain were used as queries for BLAST searches against the *Castanea dentata* genome. In total, 23 non-redundant *CadenHSFs* genes were identified and designated as *CadenHSF-1* to *CadenHSF-23* for subsequent analysis. Their physicochemical properties were subsequently characterized (Figure 2), with molecular weights ranged from 15551.6 Da (*CadenHSF-21*) to 55739.32 Da (*CadenHSF-22*) (Figure 2(A)).

Considerable variation was observed in the amino acid sequence lengths, which ranged from 137 residues (*CadenHSF-21*) to 531 residues (*CadenHSF-17*) (Figure 2(B)). The coding sequences (CDS) ranged from 411 base pairs (*CadenHSF-21*) to 1593 base pairs (*CadenHSF-17*) (Figure 2(C)). The isoelectric point (PI), at which a molecule achieves electrical neutrality, was recorded ranged from 4.70 (*CadenHSF-19*) to 9.51 (*CadenHSF-21*) (Figure 2(D)). Detailed information of the predicted 23 *CadenHSFs* proteins is listed in Table S1 (provided in Supplementary File 1).

3.2 Evolutionary relationship of CadenHSF genes with other species

To investigate the evolutionary relationships of HSF transcription factors across representative plant species, a phylogenetic tree was constructed using HSF protein sequences from *Castanea dentata*, *Vitis vinifera*, and *Arabidopsis thaliana* (Figure 3). This phylogenetic analysis aimed to identify orthologous genes among these species and clarify their evolutionary divergence and conservation. The phylogenetic tree was divided into three major classes and further classified into seven well-supported subclasses, designated HSF-I to HSF-VII, reflecting the structural and functional diversity of the HSF family (Figure 3). The topology revealed close evolutionary relationships among HSF proteins from the three species, especially between CadenHSF, AtHSF, and VvHSF proteins. These results indicate that the HSF gene family is highly conserved during plant evolution, supporting their essential roles in plant growth, stress responses, and environmental adaptation.

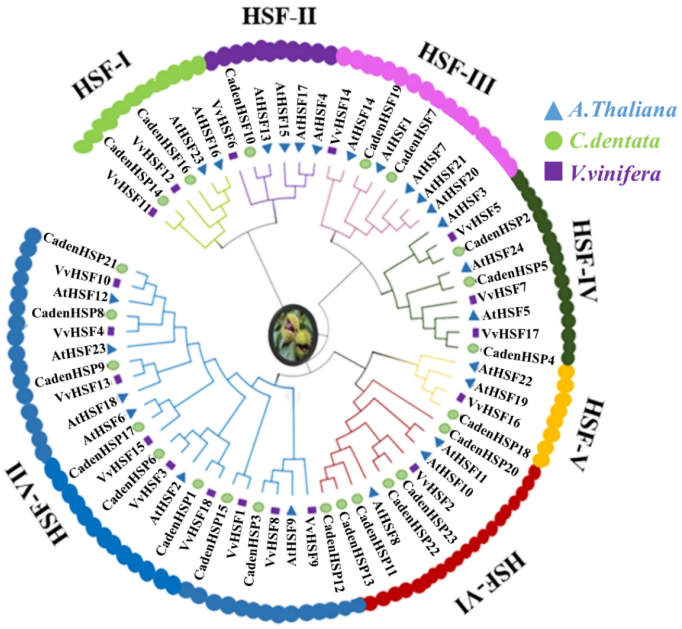

Figure 3. Comparative phylogenetic analysis of HSF proteins from *Castanea dentata*, *Vitis vinifera*, and *Arabidopsis thaliana*. The maximum likelihood tree was constructed with 1000 bootstrap replicates to infer evolutionary relationships among HSF family members. CadenHSF proteins are highlighted with green circles to distinguish them from other species. The phylogenetic clustering reveals the classification of HSF proteins into three major classes (A, B, and C) and multiple subgroups, indicating a conserved evolutionary pattern across species. The close association of *CadenHSFs* with their homologs from A. thaliana and V. vinifera suggests functional conservation and shared evolutionary origin.

3.3 Genomic Localization, Gene Duplication Events, and Syntenic Relationships of *CadenHSFs* Genes

To determine the chromosomal distribution of *CadenHSFs* genes, we analyzed genomic sequences to accurately map their locations, with each chromosome's length expressed in megabases (Mb). Our findings revealed a highly uneven distribution of *CadenHSFs* genes, which were detected on all chromosomes of *Castanea dentata* (Figure 4(A)). Notably, Chromosome 8 contained only one CadenHSF gene, whereas Chromosomes 2 and 10 exhibited the highest gene density, harboring the greatest number of CadenHSF genes. Detailed chromosomal information

for all 23 predicted *CadenHSFs* genes is provided in Table S2 (see Supplementary File 1).

Orthologous gene pairs are critical for uncovering evolutionary relationships among different plant species, as recently demonstrated through comparative genomic analyses of transcription factor families in related species [23]. To further understand the evolutionary dynamics of the predicted *CadenHSFs* genes, we examined syntenic interactions between *Castanea dentata*, Arabidopsis, and *Vitis vinifera* (grapevine). In total, 16 homologous *CadenHSFs* gene pairs were identified among the three species. These results indicate that *CadenHSFs* genes originated from a common ancestor and have undergone divergent evolution across different plant lineages (Figure 4(B)).

A comprehensive synteny analysis was performed on the entire *Castanea dentata* genome to investigate the occurrence of *CadenHSFs* gene duplication events, involving a detailed examination of the precise chromosomal locations of these genes. It was observed that genes belonging to the same subgroup are not only distributed across multiple chromosomes but also exhibit a high degree of linkage. This distinct pattern strongly suggests that these genes originated from segmental duplication events within the *Castanea dentata* genome, providing insights into their evolutionary history and genetic relationships (Figure 4(B)). Notably, Chromosomes 2 and 10 exhibited the highest abundance of *CadenHSFs* genes, each containing five copies. Chromosome 1 followed closely with four *CadenHSFs* genes, further emphasizing its important contribution to the genomic architecture of the HSF gene family.

Furthermore, our analysis revealed prominent gene duplication events, which provide valuable insights into the conserved evolutionary patterns of genes within the *Castanea dentata* genome. Two major duplication mechanisms were clearly distinguished: tandem duplication, involving adjacent gene copies on the same chromosome, and segmental duplication, characterized by the duplication and integration of larger chromosomal segments into other genomic regions (Figure 4(C)).

3.4 Gene structure and Subcellular localization analysis of *CadenHSFs* genes

The corresponding genomic DNA sequence has been compared with the full-length complementary DNA (cDNA) sequence to provide insights into the diversity of *CadenHSFs*. The structure and number of exons and introns are crucial for establishing evolutionary relationships among distinct genes and reveal the conservation of HSF genes in plants. The findings suggest that the number of introns varies from one to two, with only CadenHSF-1 and CadenHSF-20 containing two introns, as depicted in Figure 5(A).

To further examine where these genes are located in the cell. We have conducted in silico sub-cellular localization analysis. In silico sub-cellular localization results showed that most of the predicted *CadenHSFs* genes function in multiple organelles, with strong nuclear localization consistent with their primary role in transcriptional regulation. *CadenHSFs* members were also detected in mitochondria, chloroplasts, and plastids, whereas *CadenHSF-17* was uniquely localized to peroxisomes (Figure 5(B)).

3.5 Conserved motifs and functional domains of *CadenHSFs* proteins in Castanea

The *CadenHSFs* family is systematically organized into three distinct groups: Group A, Group B, and Group C. Among these, Group A stands out as the most extensive, containing a total of 15 unique HSFs proteins, while Group B and Group C comprise fewer members. To delve deeper into the characteristics of these proteins, the MEME program was employed to comprehensively identify and analyze ten conserved motifs inherent to the *CadenHSFs* proteins (Figure 6(A & C)). Notably, every protein within the CadenHSF family exhibits a uniform HSF domain (Figure 6B), underscoring a common structural feature.

Significantly, it was observed that *CadenHSFs* proteins located within the same phylogenetic group display analogous motif patterns. This similarity suggests that the entire gene family strategically utilizes these conserved motifs to fulfill specific functions pertinent to each group (Figure 6(A)). Within Group A, two notable categories are identified: the HSF 1 and the HSF-DNA-binding superfamily domains, which collectively encompass between five and seven motifs. However, CadenHSF-5 and CadenHSF-3 present a degree of complexity, featuring three motifs while also incorporating additional domains classified as the Bap31 superfamily and SPEC superfamily. This complexity highlights the evolutionary diversity and functional specialization within the CadenHSF family. CadenHSF-19 exhibits additional domains, specifically the bZIP superfamily and the SMC_pork_B superfamily (Figure 6(B)).

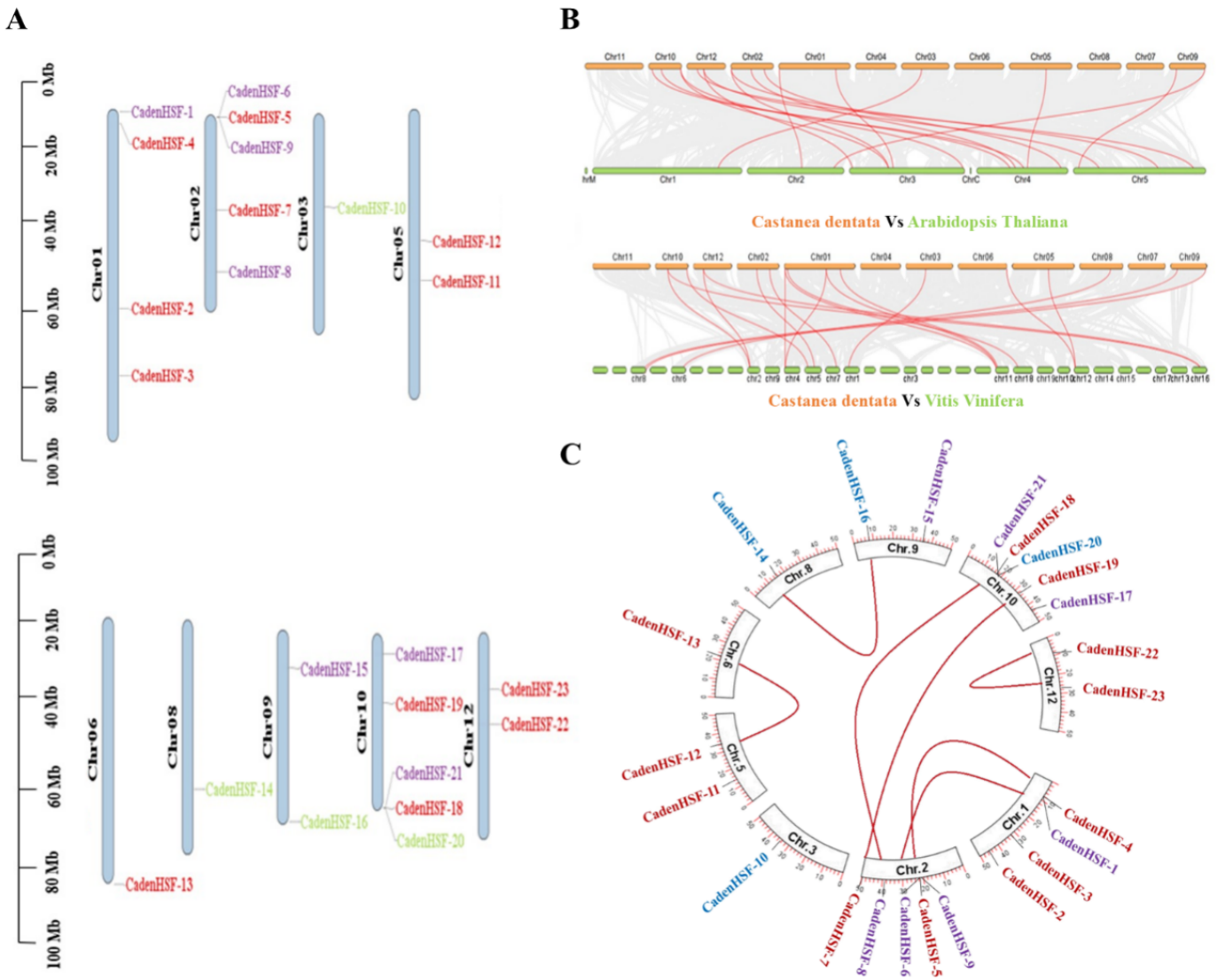

Figure 4. Genomic distribution and syntenic relationships of *CadenHSFs* genes in *Castanea dentata*. (A) Chromosomal localization of *CadenHSFs* genes across all chromosomes, showing an uneven distribution with clustering in specific regions, indicating possible duplication hotspots. Different colors represent gene groups classified based on phylogenetic relationships. (B) Comparative synteny analysis of HSFs genes among *C. dentata*, *Vitis vinifera*, and *Arabidopsis thaliana*. Orange bars represent *C. dentata* chromosomes, while green bars indicate chromosomes of A. thaliana and V. vinifera. Gray lines denote conserved collinear blocks, and red lines highlight duplicated gene pairs. The observed syntenic relationships suggest strong evolutionary conservation of HSF genes across species. (C) Intrachromosomal synteny of *CadenHSFs* genes and their orthologous relationships with A. thaliana, where red lines represent duplicated and orthologous gene pairs. These duplication patterns indicate that segmental duplication events have contributed to the expansion and evolutionary diversification of the CadenHSF gene family.

3.6 Cis-Regulatory Elements in *CadenHSFs* Promoter Regions

Cis-acting regulatory elements are key DNA motifs that regulate gene expression by controlling transcriptional activation or repression, and their systematic identification in promoter regions has proven informative for predicting stress-responsive and hormone-regulated gene functions in plant transcription factor families [24]. In the present study, a total of 3,000 cis-regulatory elements were identified across the promoter regions of *CadenHSFs* genes, representing a wide range of functional categories. These elements are primarily associated with responses to abiotic and biotic stresses, plant development, and light-responsive processes.

Among these, several hormone-responsive elements were prominently detected, including abscisic acid-responsive elements (ABRE) and salicylic acid-responsive elements (SARE), which are known to play critical roles in stress tolerance

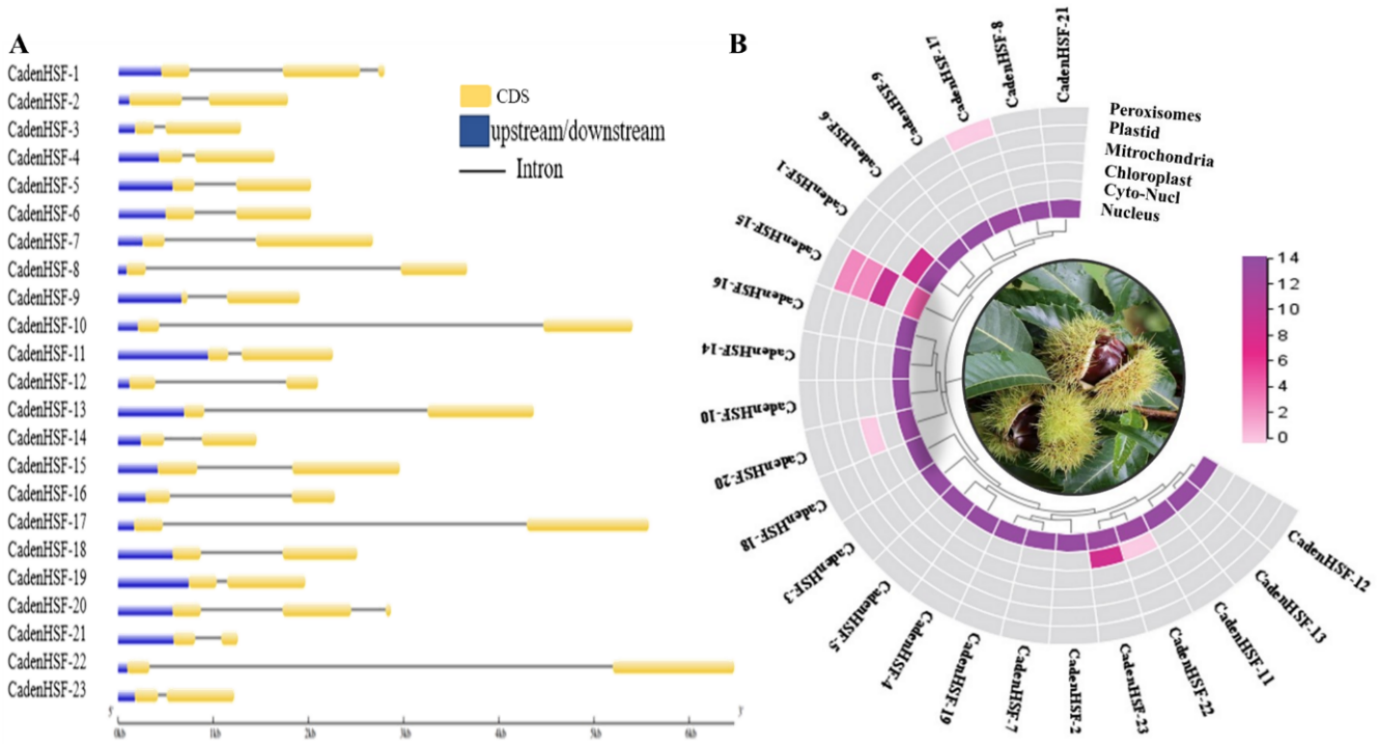

Figure 5. Gene structure and predicted subcellular localization of CadenHSF genes in *Castanea dentata*. (A) Exon–intron organization of *CadenHSFs*, where yellow boxes represent coding sequences (CDS), blue boxes denote upstream/downstream regions, and black lines indicate introns. Most *CadenHSFs* genes exhibit conserved exon–intron patterns, suggesting evolutionary conservation within the gene family, while structural variations in some members may reflect functional diversification. (B) Predicted subcellular localization of CadenHSF proteins across different cellular compartments. The heatmap indicates relative abundance, with purple representing high localization and pink low. Most *CadenHSFs* show strong nuclear localization, consistent with their roles as transcription factors, while several members also localize to mitochondria, chloroplasts, and peroxisomes, indicating potential involvement in organelle-mediated stress signaling.

and defense signaling pathways. The presence of these elements suggests that *CadenHSFs* genes may mediate adaptive responses to environmental stress through hormone-regulated pathways. In addition, gibberellin-responsive elements (GARE) were identified, indicating a potential role of *CadenHSFs* in regulating plant growth and developmental processes such as seed germination, stem elongation, and flowering. The detection of methyl jasmonate-responsive elements (MeJARE) further highlights the involvement of these genes in stress signaling and defense mechanisms. Moreover, auxin-responsive elements (ARE) were observed in several *CadenHSFs* promoters, with notable enrichment in CadenHSF-5, CadenHSF-6, and CadenHSF-13, suggesting their possible role in growth regulation and developmental adaptation. Overall, the diverse distribution of cis-regulatory elements indicates that *CadenHSFs* genes are likely regulated by multiple hormonal and environmental signals, reflecting their potential involvement in complex regulatory networks underlying plant growth, development, and stress responses (Figure 7).

4 Discussion

Heat shock transcription factors constitute a crucial regulatory network involved in plant responses to environmental stresses, particularly heat and oxidative stress. Although HSF gene families have been well characterized in crop species and model plants like *Arabidopsis thaliana*, their evolutionary dynamics and functional potential in *Castanea dentata* are still mostly unknown [25]. In this work, we present a thorough genome-wide investigation of the *CadenHSFs* gene family *C. dentata* to fill this knowledge gap. Compared to *Arabidopsis thaliana* (21 members) and *Vitis vinifera* (19 members), *Castanea dentata* showed moderate expansion of this gene family, with 23 *CadenHSFs* genes identified in this study. Such variation in gene number is indicative of species-specific evolutionary paths that are probably impacted by environmental adaptation and genome duplication events. The classification of *CadenHSFs* genes into three major classes (A, B, and

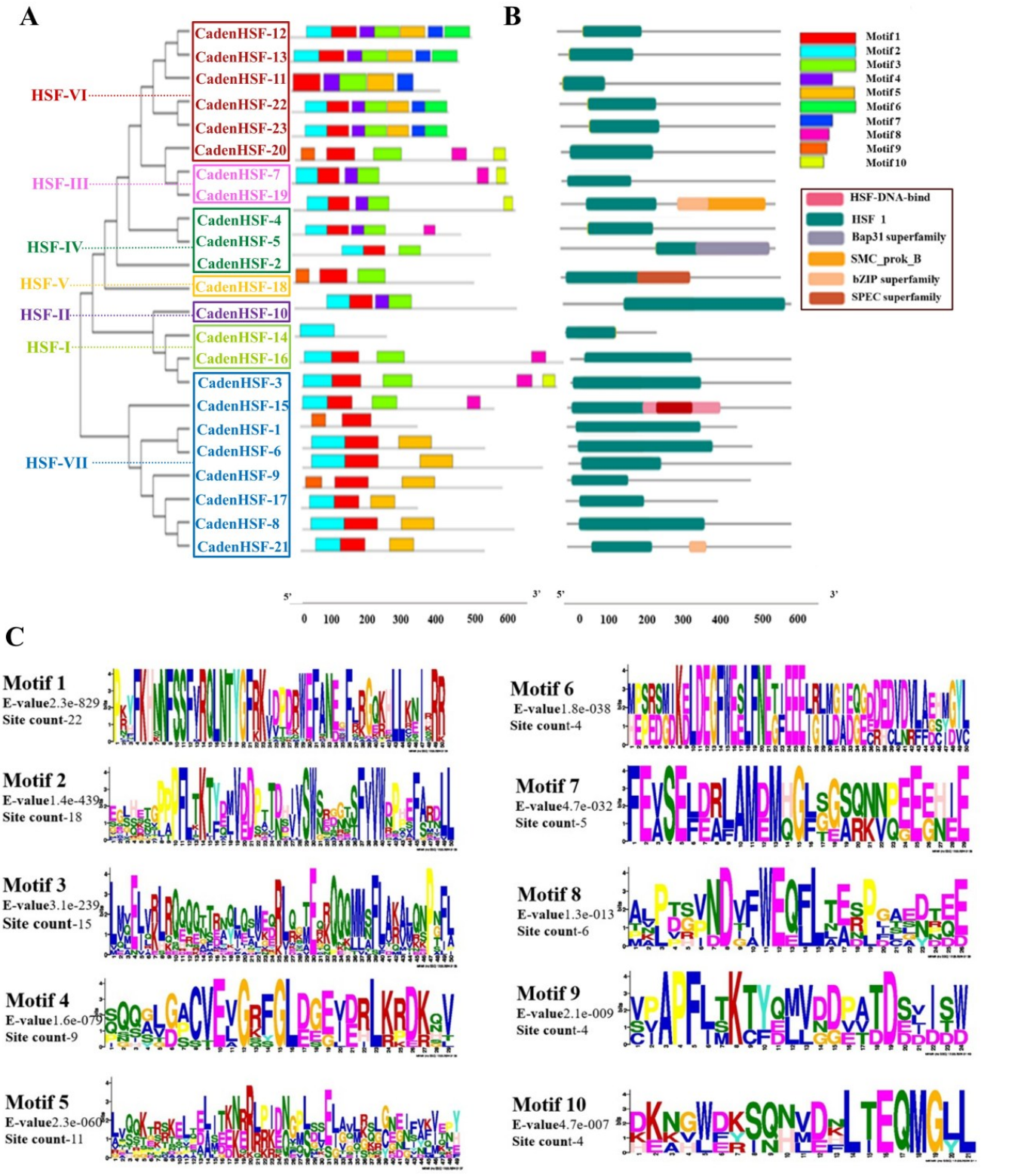

Figure 6. Functional domain and conserved motif analysis of the *CadenHSFs* gene family in *Castanea dentata*. (A) Distribution of 10 conserved motifs across 23 *CadenHSFs* proteins, organized according to phylogenetic relationships. Colored bars represent distinct motifs, highlighting the conservation and diversity within each subgroup. (B) Functional domain architecture of *CadenHSFs* proteins, where the color bar indicates conserved domains such as HSF. The presence of additional domains such as Bap31, SPEC, and bZIP indicates that some *CadenHSFs* proteins may have evolved specialized functions beyond canonical heat stress regulation. (C) Sequence logos of the 10 identified motifs, showing the conserved amino acid positions and sequence variability.

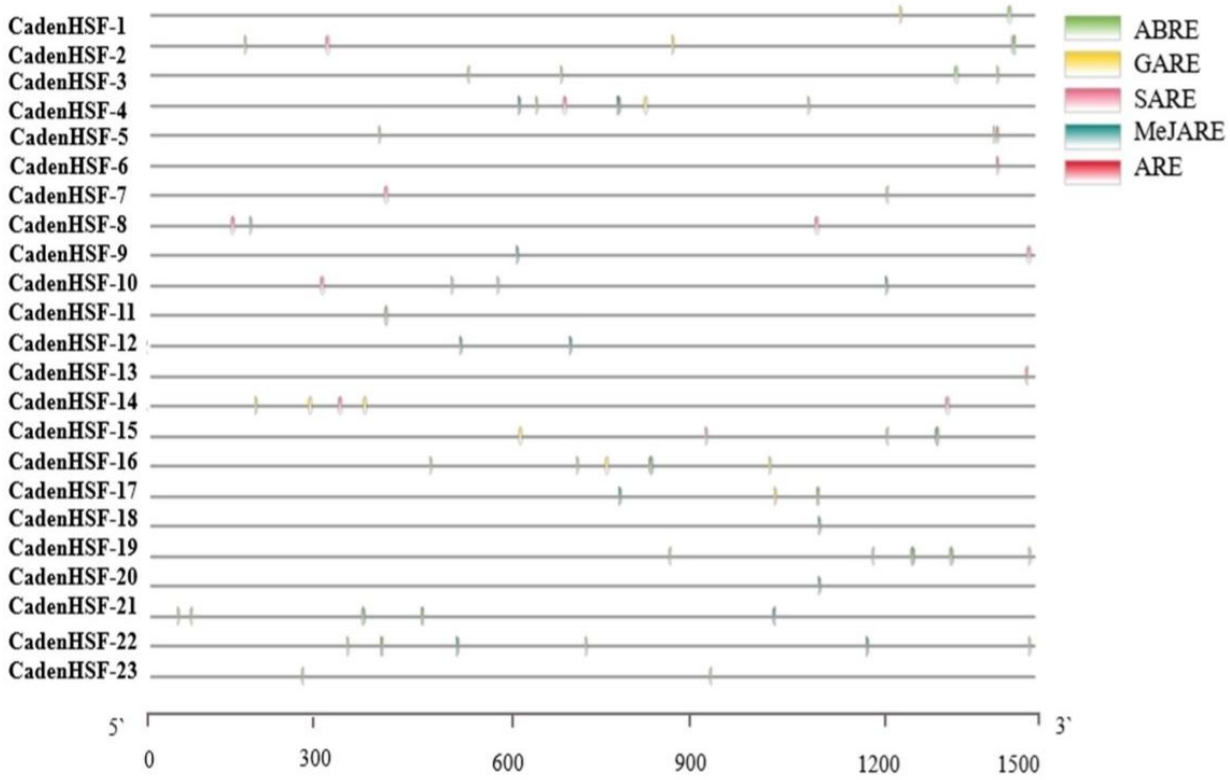

Figure 7. Cis-regulatory element analysis of *CadenHSFs* genes in *Castanea dentata*. Colored boxes indicate elements associated with hormonal and environmental stress responses: ABRE (abscisic acid), GARE (gibberellic acid), SARE (salicylic acid), MeJARE (methyl jasmonate), and ARE (auxin). The distribution of these elements across *CadenHSFs* promoters suggests that these genes are likely regulated by multiple phytohormones and may play important roles in coordinating plant responses to abiotic and biotic stresses. The diversity and abundance of cis-elements highlight potential functional specialization among *CadenHSFs* family members in stress signaling and developmental processes.

C) and seven subgroups (Figure 3) is consistent with the conserved structural and phylogenetic framework reported across plant species [26, 27].

However, the diversification into multiple subgroups suggests potential functional specialization, particularly in stress responsiveness. The uneven chromosomal distribution of *CadenHSFs* genes (Figure 4(A)), with clustering on Chromosomes 2 and 10 and minimal representation on Chromosome 8, indicates non-random genomic organization. Similar patterns have been observed in *Oryza sativa* and *Glycine max*, where gene family expansion is associated with duplication hotspots [28]. Synteny analysis further revealed conserved collinearity between *C. dentata*, *A. thaliana*, and *V. vinifera*, suggesting that HSF genes are evolutionarily conserved across dicot lineages (Figure 4(B)). Segmental duplication appears to be the primary mechanism driving HSF expansion in *C. dentata*, which aligns with previous genome-wide studies indicating that large-scale duplication events contribute significantly to transcription factor family expansion [29, 30]. In contrast, the limited role of tandem duplication highlights divergence in evolutionary mechanisms among plant species.

Gene structure analysis demonstrated conserved exon–intron organization among most *CadenHSFs*, supporting their evolutionary stability and functional conservation (Figure 5(A)). Such conservation has been widely linked to the maintenance of transcriptional regulatory roles in plants [30–32]. Nonetheless, structural variations observed in certain members may contribute to functional divergence.

The presence of the conserved HSF domain in all proteins confirms their identity, while the occurrence of additional domains, such as Bap31, SPEC, and bZIP, indicates that some *CadenHSFs* members may have evolved specialized functions beyond canonical heat stress regulation. Bap31 domains are implicated in protein folding and trafficking under stress, potentially enhancing protein stability during heat or oxidative stress [32, 33]. bZIP domains enable these HSF proteins to interact with other transcription factors and integrate hormonal signals like ABA and SA, aligning with cis-element

analyses showing hormone-responsive motifs [34–36]. SPEC domains may facilitate protein-protein interactions critical for transcriptional regulation under diverse stress conditions [33]. Collectively, these observations suggest that domain diversification contributes to the functional specialization and adaptive roles of *CadenHSFs* in *Castanea dentata*. Similar domain architectures have been reported to enhance transcriptional complexity and enable cross-talk between signaling pathways [37].

Subcellular localization analysis revealed predominant nuclear localization (Figure 5(B)), consistent with the canonical role of HSFs as transcription factors regulating stress-responsive genes [34]. However, their additional localization in mitochondria and chloroplasts suggests possible roles in organelle-to-nucleus signaling pathways, which are crucial during stress adaptation [37]. The unique peroxisomal localization of CadenHSF-17 may indicate involvement in reactive oxygen species (ROS)-mediated signaling, a key component of plant stress responses [32].

Furthermore, the identification of cis-regulatory elements (Figure 7), such as ABRE and SARE motifs, underscores the involvement of *CadenHSFs* in hormone-mediated stress responses. ABRE elements are known to regulate abscisic acid (ABA)-dependent pathways, while salicylic acid-responsive elements (SARE) are associated with defense signaling [37]. This suggests that *CadenHSFs* may function as integrators of multiple signaling pathways under abiotic and biotic stress conditions. Recent studies have also demonstrated that HSF genes play critical roles in responses to drought, heat, and pathogens, further supporting the functional significance of this gene family [33, 34].

Overall, this study provides comprehensive insights into the evolutionary dynamics, structural diversity, and regulatory potential of the HSF gene family in *C. dentata*, highlighting their crucial roles in plant adaptation to environmental challenges.

5 Conclusion

Biotic and abiotic stresses pose significant threats to *C. dentata* growth and yield at all developmental stages. This study systematically identified 23 *CadenHSFs* genes and comprehensively analyzed their functional domains, motifs, chromosomal locations, intron-exon structures, and evolutionary relationships. Phylogenetic analysis successfully identified *CadenHSFs* orthologs and paralogs, providing insights into their evolutionary origins and potential functional conservation. The expression of *CadenHSFs* genes across multiple organelles (mitochondria, plastids, chloroplasts, peroxisome) suggests their involvement in diverse cellular processes beyond stress response. Collectively, these findings establish a solid foundation for future functional research on *CadenHSFs* genes, enhance our understanding of thermotolerance mechanisms in *C. dentata*, and serve as a milestone for further investigations into the role of HSFs in mediating stress adaptation and growth regulation in chestnuts.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Kopecká, R., Kameniarová, M., Černý, M., Brzobohatý, B., & Novák, J. (2023). Abiotic Stress in Crop Production. *International Journal of Molecular Sciences*, *24*(7), 6603. [CrossRef]

[2] Herman, D. J., Knowles, L. O., & Knowles, N. R. (2017). Heat stress affects carbohydrate metabolism during cold-induced sweetening of potato (Solanum tuberosum L.). *Planta, 245*(3), 563-582. [CrossRef]

[3] Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., & Fujita, M. (2013). Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. *International journal of molecular sciences, 14*(5), 9643-9684. [CrossRef]

[4] Yang, X., Zhu, W., Zhang, H., Liu, N., & Tian, S. (2016). Heat shock factors in tomatoes: genome-wide identification, phylogenetic analysis and expression profiling under development and heat stress. *PeerJ, 4*, e1961. [CrossRef]

[5] Yura, T., & Nakahigashi, K. (1999). Regulation of the heat-shock response. *Current opinion in microbiology, 2*(2), 153-158. [CrossRef]

[6] Sorger, P. K., & Pelham, H. R. (1988). Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. *Cell, 54*(6), 855-864. [CrossRef]

[7] Chidambaranathan, P., Jagannadham, P. T. K., Satheesh, V., Kohli, D., Basavarajappa, S. H., Chellapilla, B., ... & Srinivasan, R. (2018). Genome-wide analysis identifies chickpea (Cicer arietinum) heat stress transcription factors (Hsfs) responsive to heat stress at the pod development stage. *Journal of plant research, 131*(3), 525-542. [CrossRef]

[8] Lin, Y. X., Jiang, H. Y., Chu, Z. X., Tang, X. L., Zhu, S. W., & Cheng, B. J. (2011). Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. *BMC genomics, 12*(1), 76. [CrossRef]

[9] Chung, E., Kim, K. M., & Lee, J. H. (2013). Genome-wide analysis and molecular characterization of heat shock transcription factor family in Glycine max. *Journal of genetics and genomics, 40*(3), 127-135. [CrossRef]

[10] Zhou, S., Zhang, P., Jing, Z., & Shi, J. (2013). Genome-wide identification and analysis of heat shock transcription factor family in cucumber (*Cucumis sativus* L.). *Plant Omics, 6*(6), 449-455. [Link]

[11] Zhang, J., Liu, B., Li, J., Zhang, L., Wang, Y., Zheng, H., ... & Chen, J. (2015). Hsf and Hsp gene families in Populus: genome-wide identification, organization and correlated expression during development and in stress responses. *BMC genomics, 16*(1), 181. [CrossRef]

[12] Xue, G. P., Sadat, S., Drenth, J., & McIntyre, C. L. (2014). The heat shock factor family from Triticum aestivum in response to heat and other major abiotic stresses and their role in regulation of heat shock protein genes. *Journal of experimental botany, 65*(2), 539-557. [CrossRef]

[13] Huang, Y., Li, M. Y., Wang, F., Xu, Z. S., Huang, W., Wang, G. L., ... & Xiong, A. S. (2015). Heat shock factors in carrot: genome-wide identification, classification, and expression profiles response to abiotic stress. *Molecular biology reports, 42*(5), 893-905. [CrossRef]

[14] Zhu, X., Huang, C., Zhang, L., Liu, H., Yu, J., Hu, Z., & Hua, W. (2017). Systematic analysis of Hsf family genes in the Brassica napus genome reveals novel responses to heat, drought and high CO2 stresses. *Frontiers in plant science, 8*, 1174. [CrossRef]

[15] Shirasawa, K., Nishio, S., Terakami, S., Botta, R., Marinoni, D. T., & Isobe, S. (2021). Chromosome-level genome assembly of Japanese chestnut (Castanea crenata Sieb. et Zucc.) reveals conserved chromosomal segments in woody rosids. *DNA Research, 28*(5), dsab016. [CrossRef]

[16] Jacobs, D. F., Dalgleish, H. J., & Nelson, C. D. (2016). Synthesis of American chestnut (Castanea dentata) biological, ecological, and genetic attributes with application to forest restoration. *Forest Health Initiative. Accessed online, 7*. https://www.mdsg.umd.edu/sites/default/files/files/Lesson%20Plans/american_chestnut_project/chestnut_synthesis.pdf

[17] Li, R., Sharma, A. K., Zhu, J., Zheng, B., Xiao, G., & Chen, L. (2022). Nutritional biology of chestnuts: A perspective review. *Food Chemistry, 395*, 133575. [CrossRef]

[18] Drăghici, L. R., Hădărugă, D. I., Bădoiu, B., Mitroi, C., David, I., Riviş, A., ... & Hădărugă, N. G. (2021). Nutritional values of edible chestnuts (Castanea species)—A short comparative review. *J. Agroaliment. Process. Technol, 27*, 553-564. https://journal-of-agroalimentary.ro/admin/articole/54997L80_Razvan-Draghici_27-4-S-_2021_553-564.pdf

[19] Guo, A. Y., Zhu, Q. H., Chen, X., & Luo, J. C. (2007). GSDS: a gene structure display server. *Yi chuan= Hereditas, 29*(8), 1023-1026. [CrossRef]

[20] Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular plant, 13*(8), 1194-1202. [CrossRef]

[21] Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic acids research, 35*(suppl_2), W585-W587. [CrossRef]

[22] Yu, C. S., Chen, Y. C., Lu, C. H., & Hwang, J. K. (2006). Prediction of protein subcellular localization. *Proteins: Structure, Function, and Bioinformatics, 64*(3), 643-651. [CrossRef]

[23] Noor, M., Mehar, I., Hussain, H., Gaafar, A. R. Z., Dabiellil, M., & Shah, T. A. (2025). Identification and analysis of functional regulatory elements and RNAi silencing genes in mandarin clementine (Citrus clementina). *Journal of Plant Interactions, 20*(1), 2503289. [CrossRef]

[24] Hussain, H., Alam, A., Mehar, I., Noor, M., Al-Dossary, O., Alsubaie, B., ... & Al-Khayri, J. M. (2025). Genome-wide identification and characterization of the WRKY gene family and their associated regulatory elements in fortunella hindsii. *Evolutionary Bioinformatics, 21*, 11769343241312740. [CrossRef]

[25] Jiang, L., Hu, W., Qian, Y., Ren, Q., & Zhang, J. (2021). Genome-wide identification, classification and expression analysis of the Hsf and Hsp70 gene families in maize. *Gene, 770*, 145348. [CrossRef]

[26] Duan, S., Liu, B., Zhang, Y., Li, G., & Guo, X. (2019). Genome-wide identification and abiotic stress-responsive pattern of heat shock transcription factor family in Triticum aestivum L. *Bmc Genomics, 20*(1), 257. [CrossRef]

[27] Gong, C., Pang, Q., Li, Z., Li, Z., Chen, R., Sun, G.,

& Sun, B. (2021). Genome-wide identification and characterization of Hsf and Hsp gene families and gene expression analysis under heat stress in eggplant (Solanum melongema L.). *Horticulturae, 7*(6), 149. [CrossRef]

[28] Li, P. S., Yu, T. F., He, G. H., Chen, M., Zhou, Y. B., Chai, S. C., ... & Ma, Y. Z. (2014). Genome-wide analysis of the Hsf family in soybean and functional identification of GmHsf-34 involvement in drought and heat stresses. *BMC genomics, 15*(1), 1009. [CrossRef]

[29] Qian, G., Meng, X., Wang, S., Mi, Y., Qin, Z., Liu, T., ... & Li, L. (2023). Genome-wide identification of HSF gene family and their expression analysis in vegetative tissue of young seedlings of hemp under different light treatments. *Industrial Crops and Products, 204*, 117375. [CrossRef]

[30] Movahedi, A., Hwarari, D., Dzinyela, R., Ni, S., & Yang, L. (2025). A close-up of regulatory networks and signaling pathways of MKK5 in biotic and abiotic stresses. *Critical Reviews in Biotechnology, 45*(2), 473-490. [CrossRef]

[31] Hussain, H., Fatima, N., Sajid, M., Mehar, I., Noor, M., Attia, K. A., ... & Shah, T. A. (2024). Genome-wide analysis and identification of nuclear factor Y gene family in switchgrass (Panicum virgatum L.). *BMC genomics, 25*(1), 1218. [CrossRef]

[32] 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. [CrossRef]

[33] 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. [CrossRef]

[34] Scharf, K. D., Berberich, T., Ebersberger, I., & Nover, L. (2012). The plant heat stress transcription factor (Hsf) family: structure, function and evolution. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1819*(2), 104-119. [CrossRef]

[35] Liu, G., Chai, F., Wang, Y., Jiang, J., Duan, W., Wang, Y., ... & Wang, L. (2018). Genome-wide Identification and Classification of HSF Family in Grape, and Their Transcriptional Analysis under Heat Acclimation and Heat Stress. *Horticultural Plant Journal, 4*(4), 133-143. [CrossRef]

[36] Movahedi, A., Kadkhodaei, S., & Yang, L. (2024). Transcriptional regulation and posttranslational modifications in plant growth and development under abiotic stresses. *Frontiers in Plant Science, 15*, 1454335. [CrossRef]

[37] Ohama, N., Sato, H., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2017). Transcriptional regulatory network of plant heat stress response. *Trends in plant science, 22*(1), 53-65. [CrossRef]

Appendix

Table S1. Genomic attributes of predicted *CadenHSFs* obtained from Phytozome and ExPASy server.

Transcript ID	CDS (bp)	Protein length (A.A)	Molecular weight (Kda)	Aliphatic index	GRAVY
Caden.01G004800.1	1185	395	44242.54	77.94	-0.494
Caden.01G239500.1	1392	464	52848.41	79.29	-0.628
Caden.01G337100.1	1002	334	37144.02	68.5	-0.446
Caden.01G026900.1	1089	363	40750.02	69.2	-0.559
Caden.02G003600.2	1023	341	39342.16	71.91	-0.794
Caden.02G003400.1	1098	366	41806.86	69.12	-0.78
Caden.02G122000.1	1467	489	54784.59	58.98	-0.861
Caden.02G191200.1	906	302	32787.2	61.83	-0.764
Caden.02G003700.1	837	279	32612.58	72.55	-0.888
Caden.03G117800.1	1170	390	45125.8	69.15	-0.776
Caden.05G154800.1	1179	393	44771.04	71.07	-0.698
Caden.05G122500.1	615	205	23720.97	69.36	-0.695
Caden.06G290800.1	1338	446	51170.08	69.46	-0.696
Caden.08G124200.1	828	276	32036.16	70.51	-0.641
Caden.09G053600.1	1554	518	58168.9	68.8	-0.645
Caden.09G209200.1	714	238	27531.34	71.14	-0.753
Caden.10G025500.1	1593	531	58306.8	64.91	-0.694
Caden.10G206100.1	1089	363	41151.51	63.26	-0.773
Caden.10G075500.1	1122	374	40404.52	69.2	-0.655
Caden.10G206100.3	1089	363	41430.04	62.98	-0.746
Caden.10G205900.1	411	137	15551.6	50.07	-0.735
Caden.12G113800.1	1518	506	55739.32	75.31	-0.528
Caden.12G074900.1	960	320	35941.98	68.78	-0.752

Table S2. Chromosomal information of predicted *CadenHSFs*, their start, end position obtained from Phytozome.

Gene name	Chromosome	Location-Start	Location-End	Chr. Length (bp)
CadenHSF-1	Chr01	711461	716611	50373814
CadenHSF-2	Chr01	53232978	53235464	50375985
CadenHSF-3	Chr01	71126683	71128172	50371608
CadenHSF-4	Chr01	3844373	3846368	50376771
CadenHSF-5	Chr02	494356	497001	50386094
CadenHSF-6	Chr02	494356	497001	50386093
CadenHSF-7	Chr02	25289809	25292867	50384995
CadenHSF-8	Chr02	41724736	41728791	50384639
CadenHSF-9	Chr02	514593	517248	50385195
CadenHSF-10	Chr03	24903667	24909800	50370108
CadenHSF-11	Chr05	45766043	45771132	50393179
CadenHSF-12	Chr05	35108147	35110379	50393129
CadenHSF-13	Chr06	61538180	61542798	50388276
CadenHSF-14	Chr08	38932198	38933874	50366658
CadenHSF-15	Chr09	8772506	8776363	50399631
CadenHSF-16	Chr09	44083458	44085973	50398655
CadenHSF-17	Chr10	4451623	4457978	50405774
CadenHSF-18	Chr10	40065633	40070346	50407906
CadenHSF-19	Chr10	15847216	15849747	50407142
CadenHSF-20	Chr10	40065633	40070346	50407908
CadenHSF-21	Chr10	40033293	40039590	50406536
CadenHSF-22	Chr12	20972918	20985225	50378538
CadenHSF-23	Chr12	12981747	12983158	50378857