Genome-wide identification and analysis of the heat shock transcription factor family in moso bamboo (Phyllostachys edulis)

Bin Huang  
Zhejiang A & F University

Zhinuo Huang  
Zhejiang A & F University

Ruifang Ma  
Zhejiang A & F University

Jialu Chen  
Zhejiang A & F University

Zhijun Zhang  
Zhejiang A & F University

Kim Yrjälä Yrjälä  
University of Helsinki

Research Article

Keywords: moso bamboo, heat shock factor, genome-wide, expression pattern

Posted Date: December 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-116570/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at Scientific Reports on August 13th, 2021. See the published version at https://doi.org/10.1038/s41598-021-95899-3.
Genome-wide identification and analysis of the heat shock transcription factor family in moso bamboo (Phyllostachys edulis)

Bin Huang1,2, Zhinuo Huang1,2, Ruifang Ma1,2, Jialu Chen1,2, Zhijun Zhang1,2*, Kim Yrjälä1,3

1 State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Zhejiang, China;
2 Zhejiang Provincial Collaborative Innovation Centre for Bamboo Resources and High efficiency Utilization, Zhejiang A&F University, Zhejiang, China;
3 Department of Forest Sciences, University of Helsinki, Helsinki, Finland;
* Correspondence: zjzhang@zafu.edu.cn;

Abstract

Heat shock transcription factors (Hsfs) are central elements in the regulatory network that controls plant heat stress response. They are involved in multiple transcriptional regulatory pathways and play important roles in heat stress signaling and responses to a variety of other stresses. We identified 41 members of the Hsf gene family in moso bamboo, which were distributed non-uniformly across its 19 chromosomes. Phylogenetic analysis showed that the moso bamboo Hsf genes could be divided into three major subfamilies; Hsfs from the same subfamily shared relatively conserved gene structures and sequences and encoded similar amino acids. All Hsf genes contained Hsf signature domains. Subcellular localization prediction indicated that about 80% of the Hsf proteins were located in the nucleus, consistent with the results of GO enrichment analysis. A large number of stress response–associated cis-regulatory elements were identified in the Hsf upstream promoter sequences. Synteny analysis indicated that the Hsfs in the moso bamboo genome had greater collinearity with those of rice and maize than with those of Arabidopsis and pepper. Numerous segmental duplicates were found in the moso bamboo Hsf gene family. Transcriptome data indicated that the expression of a number of PeHsfs differed in response to exogenous gibberellin (GA) and naphthalene acetic acid (NAA). A number of Hsf genes were highly expressed in the panicles and in young shoots, suggesting that they may have functions in reproductive growth and the early development of rapidly-growing shoots. This study provides fundamental information on members of the bamboo Hsf gene family and lays a foundation for further study of their biological functions in the regulation of plant responses to adversity.

Keywords: moso bamboo; heat shock factor; genome-wide; expression pattern
Plant growth is influenced by a variety of adverse environmental factors such as high temperature, salt stress, drought, and other abiotic stresses, as well as by biotic stresses such as diseases and pests. High temperature and drought are particularly important abiotic stresses that can promote chlorophyll degradation, damage chloroplast membrane mechanisms, and reduce photosynthetic efficiency [1], thereby profoundly restricting plant growth, development, and function. Because of their sessile nature, plants cannot actively escape from stress and must rely on physiological and biochemical mechanisms to survive external stresses [2]. Therefore, they have evolved a series of complex and effective strategies to maintain normal physiology, metabolism, and growth under adverse conditions [3]. Transcription factor (TF) gene expression plays an important role in plant stress resistance [4]. For example, ABRE Binding Factor (ABF) and MYC TFs are directly involved in abscisic acid (ABA) and jasmonate (JA) signaling pathways [5], and heat shock transcription factors (Hsfs) are involved in the regulation of reactive oxygen species (ROS), calcium signaling, and other signal transduction pathways [6]. Hsfs, a class of transcription factors widely found in eukaryotes, also have important roles in the receipt and transmission of signals, the recognition of heat shock elements (HSEs) and regulation of downstream genes, the stimulation of plant stress responses, and the promotion of heat tolerance [7].

Hsf transcription factors contain five basic functional domains: an N-terminal DNA-binding domain (DBD), an oligomerization domain (OD or HR-A/B), a nuclear localization signal (NLS), a nuclear export signal (NES), and in some cases, a C-terminal transcriptional activation domain (CTAD) [8]. The highly conserved DNA-binding domain (also known as the Hsf domain) consists of three α-helix bundles and four strands of reverse-parallel β-folded layers that form a tight sphere. This structure ensures the specific binding of Hsfs to HSEs in target gene promoters, thereby regulating the transcription and expression of heat shock genes [9]. The OD consists of two hydrophobic 7-peptide repeat regions, HR-A and HR-B, and Hsfs are divided into three subfamilies (HsfA–C) based on differences in the number of amino acids inserted between HR-A and HR-B. HsfAs and HsfCs have 21 and 7 amino acid residues between HR-A and HR-B, respectively, whereas HsfBs have none [10, 11]. The nuclear localization signal (NLS) is composed of basic amino acids and directs the transport of Hsfs into the nucleus [12], whereas the leucine-rich nuclear export signal (NES) facilitates their export from the nucleus to the cytoplasm [13]. The CTAD is the least conserved region and contains AHA motifs that consist of aromatic, large hydrophobic, and acidic amino acid residues. The AHA sequence is specific to the class A subfamily,
and it is generally required for transcriptional activation. HsfBs and HsfCs do not contain AHA motifs and thus lack a transcriptional activation function [14].

The first plant Hsf gene was cloned from tomato in 1990. As more completed plant genome sequences became available, Hsf family members were identified in Arabidopsis (21 Hsf genes) [10], tomato (26) [15], cabbage (52) [16], pepper (25) [17], poplar (31) [18], maize (25) [19], and rice (25) [20]. Detailed functional studies of individual Hsf subfamilies have been performed in Arabidopsis, tomato, and other model plants, and experiments have confirmed that Hsfs have important roles in plant responses to heat and other abiotic stresses. Arabidopsis HsfAs were shown to be transcriptional activators, and AtHsfA1a regulates the synthesis of heat shock proteins (Hsps) during the heat-response phase to minimize plant growth damage at high temperatures [21]. The overexpression of Arabidopsis AtHsfA2 [22] and tomato SlHsfA3 [23] significantly increased both basal and acquired heat tolerance. The expression of AtHsfA6a increased nearly 150-fold under salt stress compared with control conditions [24]. By contrast, HsfB1 and HsfB2b from subfamily B were shown to be transcriptional repressors, and rice OsHsfB2b was found to significantly reduce plant salt tolerance after NaCl treatment [20, 25]. Under drought stress, plants with upregulated expression of AtHsfA1b showed significant increases in yield and harvest index. Plants that overexpressed HsfA1b also had enhanced resistance to pathogenic bacteria, indicating that HsfA1b is involved in the regulation of both abiotic and biotic stress resistance [26].

Moso bamboo (Phyllostachys edulis) is a non-timber forestry species from the subfamily Bambusoideae of the family Poaceae and is native to China [27]. It has a wide distribution area and a high economic value, and it plays an important role in soil and water conservation, soil carbon sequestration and oxygen dynamics, and climate regulation [28]. However, water availability in moso bamboo forests is entirely dependent on natural precipitation, and its two most important water-demanding growth stages, ‘asparagus gestation’ in the autumn and ‘asparagus emergence’ in the spring, are most sensitive to heat and drought stress [28, 29]. High temperatures and drought not only damage moso bamboo growth and development, thereby reducing the yield of bamboo shoots, but also reduce the carbon stocks of moso bamboo [30]. The intensification of global climate change is predicted to significantly alter the level and quality of ecological services to forests [31], which will aggravate the effects of water stress on moso bamboo and significantly limit its yield.

To date, there have been no studies of the Hsf transcription factor family in moso bamboo. Here, we used a bioinformatics approach to identify and characterize Hsf genes and their encoded proteins in the moso bamboo genome [32, 33]; we analyzed their evolutionary relationships, conserved domains, gene structures, promoter
elements, duplication patterns, tissue-specific expression, and responses to external
hormone treatments. Our results lay the groundwork for further study of heat stress
response mechanisms in bamboo.

2. Materials and methods

2.1. Identification and sequence analysis of Hsf proteins from moso bamboo

Genomic data from moso bamboo were downloaded from the *P. edulis* genome
database (ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001_101000/100498/). A
hidden Markov model of the Hsf DBD (PF00447) was obtained from the Pfam database
(http://pfam.xfam.org/) and used as the seed model for an HMMER3 search
(http://hmmer.janelia.org/) of the local bamboo protein database (*E* ≤ 10<sup>−20</sup>) [34], and
redundant genes were removed to produce a set of preliminary Hsf candidate sequences.
To verify that these candidates were Hsfs, we used the SMART (http://smart.embl-
heidelberg.de/) [35] and Pfam [36] databases to filter out sequences that lacked
complete Hsf-type DBDs. The confirmed *Hsf* genes were renamed according to their
positions on the moso bamboo chromosomes.

Subcellular localization predictions were generated with CELLO
(http://cello.life.nctu.edu.tw/) [37], and the ExPASy ProtParam tool
(https://web.expasy.org/protparam/) [38] was used to predict protein physicochemical
parameters such as molecular weight (MW) and isoelectric point (pI).

2.2. Sequence Alignment and Phylogenetic Tree Construction

Whole genome information for Arabidopsis and rice was downloaded from the TAIR10
database (http://www.arabidopsis.org/index.jsp) and the Rice Genome Annotation
Project database (http://rice.plantbiology.msu.edu). Maize and pepper genomic data
were downloaded from the Ensembl database (http://asia.ensembl.org/index.html).
Twenty-one Arabidopsis Hsf proteins and 25 rice Hsf proteins were identified from
HMMER3 searches of the corresponding local protein databases [34]. The Arabidopsis
and rice Hsf sequences were combined with those from moso bamboo, a multiple
protein sequence alignment was produced with ClustalX, and the resulting alignment
was used to construct a maximum likelihood (ML) phylogenetic tree in MEGA 7.0 with
1000 bootstrap replicates [39]. Intraspecific phylogenetic trees were also constructed
using the Hsf protein sequences from bamboo. The amino acid sequences of conserved
domains were compared and edited using Jalview software (http://www.jalview.org/
[40], and the Jalview output was submitted to JPred
2.3. Gene Structures, Motif Identification, and Conserved Domains
The intron–exon distributions of the moso bamboo *Hsf* genes were obtained using GFF annotation files from the moso bamboo genome. Conserved amino acid sequences of Hsf proteins were analyzed using the online MEME tool ([http://meme-suite.org/](http://meme-suite.org/)) [42]. MEME analysis parameters included a minimum width ≥ 6, a maximum width of 50, and a motif number of 10; all other parameters were set to default values. Conserved domains in the Hsfs were predicted using the NCBI Conserved Domain Database ([https://www.ncbi.nlm.nih.gov/cdd/](https://www.ncbi.nlm.nih.gov/cdd/)), and DOG 2.0 ([http://dog.biocuckoo.org](http://dog.biocuckoo.org)) was used for protein structure visualization [43].

2.4. Chromosomal locations, genomic duplications, and Ka/Ks ratios
Chromosome lengths and gene locations were obtained from the moso genome annotation file, and MG2C v.2 ([http://mg2c.iask.in/mg2c_v2.0/](http://mg2c.iask.in/mg2c_v2.0/)) was used to visualize the gene locations on chromosomes [44]. The moso bamboo protein sequences were aligned to one another or to the protein sequences from Arabidopsis, rice, maize, or pepper using TBtools software. MCScanX [45] was used to identify gene duplication events and syntenic relationships among the Hsf proteins, and the results were visualized using Circos and Dual Synteny Plot in TBtools [46].

For Ka/Ks analysis, thirteen homologous gene pairs were identified by BLASTn using two criteria: (1) >75% sequence similarity and (2) an alignable region >75% of the length of the longer sequence [47]. Kaks_Calculator2.0 was used to calculate the synonymous substitution rate (Ks), nonsynonymous substitution rate (Ka), and Ka/Ks ratio between homologous gene pairs [48]. Evolutionary divergence times within the bamboo Hsf gene family were calculated using the bamboo-specific divergence time formula $T = \frac{K_s}{2\lambda}$ (where $\lambda = 6.5 \times 10^{-9}$) [32].

2.5. Identification of *cis*-acting elements
PlantCARE ([http://bioinformatics.psb.ugent.be/webtools/plantcare/html/](http://bioinformatics.psb.ugent.be/webtools/plantcare/html/)) was used to identify *cis*-acting elements in the 1,500-bp promoter region upstream of each gene’s transcription start site, and the results were visualized using TBtools [49].

2.6. Expression analysis of the *Hsf* genes
Replicated transcriptomic data from different plant organs (roots, rhizomes, panicles, and leaves), bamboo shoots at different germination stages (20, 50, and 100 cm), and seedling root tissues treated with 5 μM gibberellic acid (GA) or 5 μM naphthalene acetic acid (NAA) were obtained from the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) and the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home). In total, 31 transcriptomic datasets were downloaded (accession numbers SRP119924, SRP012682, SRP109631, ERR105067, ERR105068, ERR105069, ERR105070, ERR105071, ERR105072, ERR105073, ERR105074, ERR105075, and ERR105076.) Transcriptome data in the form of transcripts per million reads (TPM) were log_{10}-transformed and imported into TBtools, where Amazing Heatmap was used to generate expression heatmaps.

2.7. Three-dimensional (3D) structural modeling of Hsf family proteins

The PDB database (http://www.rcsb.org/) was used to retrieve protein models homologous to those of the Hsf proteins. Swiss Model (https://www.swissmodel.expasy.org/) was then used to predict the protein tertiary structures by homology modeling, and the quality of the resulting models was assessed using SAVES v.5.0 (http://servicesn.mbi.ucla.edu/SAVES/).

2.8. Protein interaction network predictions and GO enrichment analyses

The Hsf protein sequences were uploaded to the STRING database (https://string-db.org/) for node comparison, and relationships among important proteins were predicted based on rice protein interactions. Cytoscape was used to visualize the resulting network [50].

GOATools (http://github.com/tanghaibao/GOatools) [51] was used to assign GO annotations to Hsfs, and Fisher's exact test was used to identify biological functions enriched in the PeHSFs relative to the full GO database. A Bonferroni [52] multiple testing correction was used to minimize false positives, and functions were considered to be significantly enriched when their Bonferroni-corrected P-values (Padjust) were <0.05.

3. Results

3.1. Identification of Hsf genes in moso bamboo
Forty-four putative Hsf candidate genes were obtained from an HMMER3 search of the bamboo protein database using the plant Hsf-type DBD model (Pfam PF00447) with an E-value threshold of ≤10^{-20}. We removed redundant genes and verified the presence of conserved domains and motifs to arrive at a final set of 41 Hsf family members. The genes were renamed PeHsf01–PeHsf41 based on their positions on the chromosomal scaffolds (Table 1).

Proteins encoded by the 41 Hsf genes contained 159 (PeHsf03) to 512 (PeHsf30) amino acids, and their molecular weights ranged from 17.49 (PeHsf03) to 56.33 kDa (PeHsf32). Approximately 90% of the Hsf proteins had molecular weights of 30–50 kDa. Their predicted isoelectric points (pI) ranged from 4.72 (PeHsf04) to 9.31 (PeHsf39). Instability index calculations predicted that 39 (95%) of the Hsf proteins were stable in vitro, with the exceptions of PeHsf05 and PeHsf21, which had instability indices of 32.96 and 37.27, respectively. Aliphatic amino acid indices showed that the thermal stability of the proteins ranged from 57.01 to 83.25, indicating that differences in their thermal stability were relatively minor. The grand average of hydropathicity (GRAVY) scores of all Hsf proteins were negative, indicating that they were hydrophilic proteins. CELLO subcellular localization predictions suggested that about 80% of the Hsf proteins were located in the nucleus.

3.2. Phylogenetic analysis and sequence alignment of the PeHsf proteins

To clarify the evolutionary relationships among the Hsf proteins, amino acid sequences of 41 moso bamboo Hsfs, 25 rice Hsfs, and 21 Arabidopsis Hsfs were used to construct an ML phylogenetic tree [53] (Fig. 1). Based on well-established Arabidopsis and rice Hsf family classifications [54], the Hsf genes were divided into three major subfamilies, HsfA (blue), HsfB (green), and HsfC (yellow). The HsfA subfamily was the largest, with 48 members across the three species, whereas the PeHsfC subfamily was the smallest, with 12 members. There was only one Arabidopsis HsfC family member (AtHsfC1), suggesting that the C subfamily had expanded in monocots. To further investigate the evolutionary relationships among PeHsf, we constructed a second phylogenetic tree with only full-length PeHSF protein sequences. This analysis confirmed that there were 16 HsfAs, 18 HsfBs, and 7 HsfCs in the moso bamboo genome (Fig. 2).

A multiple sequence alignment of the 41 PeHSF proteins was generated to investigate the presence and locations of conserved protein domains. All Hsf family members contained a highly conserved DBD at the amino terminus that consisted of approximately 100 amino acids (Fig. 3). The protein secondary structure was predicted to contain three α-helix bundles (α1–α3) and four reverse-parallel β-folds (β1–β4).
DBD domain has the ability to specifically recognize and precisely localize heat stress elements, and similar phenomena are found in other plants.
3.3. Gene structures, motif identification, and conserved domain analysis

The number of introns in each moso bamboo Hsf gene ranged from one to three (Fig. 2). Approximately 75% of the genes contained a single intron, whereas six genes contained two introns and four contained three introns. With the exception of PeHsf31, all HsfCs contained only one intron. Structures of Hsf genes on the same branch were generally similar.

We identified up to 10 highly conserved motifs in each Hsf protein using MEME (Fig. 4). All Hsfs contained motifs 1–3, which constituted the most highly conserved portions of the DBD. The relative positions of the motifs were similar for most sequences, with the exception of the HsfB PeHsf03, which lacked motif 2. In addition, both class B and C subfamily members did not contain motif 6 associated with transcriptional activation.

The structures of the PeHsf proteins were further characterized using the NCBI Conserved Domain Database (Fig. 5). All PeHsf proteins contained highly conserved DBDs at the N terminus between amino acids 7 and 155, the presence of which served as confirmation of their identity as Hsf proteins. Some genes contained additional conserved domains. For example, PeHsf10, PeHsf27, and PeHsf41 contained a bZIP-like transcription factor domain, and PeHsf17 and PeHsf36 contained a PRK10246 domain. Therefore, multiple functions may have evolved in some members of the moso bamboo Hsf family.

3.4. Chromosomal distribution and synteny analysis of the PeHsf genes

The PeHsf genes were non-uniformly distributed across the 19 chromosome scaffolds of moso bamboo (Fig. 6). The largest number were found on scaffold 21 (5), followed by scaffolds 14, 15, and 16 (4), scaffolds 17, 18, and 20 (3), and scaffolds 3, 13, and 22 (2). All other chromosomes contained a single PeHsf gene. Small gene clusters were found on scaffolds 3 and 18, based on the definition of gene clusters [55].

Gene duplication events are prevalent in all species; they can give rise to new functional genes and drive species evolution [56]. We therefore used MCScanX genome synteny analysis to explore duplications within the moso bamboo Hsf gene family (Fig. 7). Twenty-seven gene pairs appeared to have arisen from segmental duplications, and only PeHsf07 and PeHsf31 had no duplicates in the genome. Previous studies have shown that two or more highly similar genes located in close proximity to one another are likely to be tandem duplicates [57]. As shown in Fig. 6, two such pairs, PeHsf01–PeHsf02 and PeHsf28–PeHsf29, were present on scaffolds 3 and 18, respectively.
To further investigate gene duplications in the *Hsf* gene family, we performed genome-to-genome synteny analysis between moso bamboo and four representative plants. These included two dicots, Arabidopsis and *Capsicum annuum* (pepper) (Fig. 8a), and two monocots, *O. sativa* (rice) and *Zea mays* (maize) (Fig. 8b). Four, 3, 58, and 52 moso bamboo *Hsf* genes were syntenic with those of Arabidopsis, pepper, rice, and maize, respectively. There was therefore greater collinearity between the moso bamboo and monocot genomes than between the bamboo and dicot genomes. Furthermore, the rice Hsf genes all had corresponding orthologs in moso bamboo, and most of them had more than two orthologs, suggesting that moso bamboo has undergone additional whole-genome duplication events during its evolution.

To investigate evolutionary constraints and selection pressures on the *PeHsf* genes, we calculated Ka, Ks, and Ka/Ks for 13 homologous *PeHsf* gene pairs (Table 2). The synonymous substitution rate (Ks) represents the background base substitution rate, and Ks values can therefore be used to predict the timing of whole-genome duplication events. The Ks values of the *PeHsf* gene pairs ranged from 0.1023 to 1.0833, indicating that a large-scale *PeHsf* gene duplication event occurred as early as 83.33 million years ago (MYA) and as recently as 7.86 MYA. The Ka/Ks values of the gene pairs were all less than 1.0, and these genes may have undergone strong purifying selection during evolution.

### 3.5. Promoter *cis*-element analysis of *PeHsf* genes

*cis*-acting elements are non-coding DNA sequences in gene promoters that regulate the transcription of their associated genes. We identified ten *cis*-acting elements in the 1,500 base pairs upstream of the *PeHsf* genes using PlantCARE software (Fig. 9). Most were stress-response elements, and all promoters except that of *PeHsf09* contained at least one MYB drought stress response element. The next most common element was the ABRE element, which is involved in abscisic acid responses. Other elements present in the *PeHsf* promoters included the MBS element (involved in drought, high salt, and low temperature responses), the TC-rich repeat (involved in defense and stress responses), the SPI light response element, the P-box and TGA gibberellin response elements, the TCA salicylic acid response element, and the CGTCA motif (involved in methyl jasmonate response). These results suggest that the expression of moso bamboo *Hsf* genes is regulated by *cis*-elements associated with plant developmental processes and abiotic stress responses.
3.6. Transcription profiles of the PeHsf genes

We used published transcriptome data to investigate the expression patterns of the PeHsf genes (1) in seedling roots treated with gibberellic acid (GA) or naphthalene acetic acid (NAA), (2) in roots, rhizomes, panicles, and leaves, and (3) in young shoots of different heights (20, 50, and 100 cm). In Fig. 10, blue indicates low transcript abundance and red indicates high transcript abundance. Under GA treatment (Fig. 10a), the expression of PeHsf17, PeHsf24, PeHsf26, PeHsf06, PeHsf22, PeHsf33, PeHsf20, PeHsf15, PeHsf08, and PeHsf36 was downregulated compared with the control, indicating that GA repressed the expression of these genes. By contrast, the expression of PeHsf27, PeHsf01, and PeHsf31 was elevated under GA treatment but reduced under NAA treatment (Fig. 10b), indicating that the root expression of individual PeHsf genes differed between the hormone treatments. Similarly, PeHsf expression differed markedly among plant organs (Fig. 10c). In general, more than half of the genes were highly expressed in panicles, and four genes (PeHsf27, PeHsf11, PeHsf10, and PeHsf41) had expression levels greater than 100 TPM in this organ. Seven genes (PeHsf33, PeHsf13, PeHsf03, PeHsf35, PeHsf09, PeHsf16, and PeHsf29) were highly expressed only in roots, suggesting that they may participate in the development and/or function of bamboo roots. PeHsf36 and PeHsf30 were highly expressed only in leaves, and a large proportion of PeHsf genes were either not expressed or expressed only at low levels in the rhizome. Expression analysis of bamboo shoots at different germination stages (Fig. 10d) showed that many PeHsf genes were expressed highly in 20-cm shoots, at intermediate levels in 50-cm shoots, and at the lowest level in 100-cm shoots. In general, these PeHsf genes were specifically expressed during rapid shoot growth, suggesting that they may have important functions during this time.

3.7. GO enrichment analysis of the PeHsf proteins

Plants have evolved complex mechanisms to perceive and respond to biotic and abiotic stresses, and Hsfs are important components of these defense systems [2]. To further investigate the biological functions of the moso bamboo Hsfs, we performed gene ontology (GO) annotation and enrichment analysis of the 41 PeHsf proteins (Fig. 11). Eight molecular function, one cellular component, and eleven biological process GO terms were enriched in the PeHSFs relative to the complete GO database. Thirty-six proteins were annotated with the enriched GO term heat response, which had an enrichment factor of 0.09, confirming that heat response is a primary function of the PeHsf proteins. GO enrichment results suggested that PeHsf transcription factors were also involved in macromolecule biosynthesis, RNA biosynthesis, nitrogen compound metabolism, cellular biosynthetic processes, primary metabolic processes, and the
regulation of RNA metabolic processes. The largest number of genes (37) were associated with the term “regulation of cellular macromolecule biosynthetic process.”

3.8. Construction of a PeHsf protein interaction network
We used the STRING database to predict potential interactions among the PeHsf proteins (Fig. 12). There were 19 nodes in the PeHsf protein interaction network, each of which interacted with multiple other nodes. Some proteins exhibited direct interactions, e.g. PeHsf14 and PeHsf16, and others showed more complex multi-gene interactions, e.g. PeHsf01, PeHsf16, and PeHsf23. PeHsf01, PeHsf24, PeHsf25, PeHsf27, and PeHsf40 were predicted to be central nodes, radiating numerous connections to other nodes.

3.9. Homology modeling of PeHsf tertiary structures
PeHsf15 from subfamily HsfA, PeHsf28 from subfamily HsfB, and PeHsf26 from subfamily HsfC were selected for tertiary structural homology modeling. The three-dimensional structures of the three subfamilies shared a number of similarities (Fig. 13). The highly conserved DNA binding domain (DBD), a distinctive feature of Hsfs, was visible as a spherical region formed from three α-helix bundles and four strands of reverse-parallel β-folded layers. Different sequences had different convoluted structures. In PeHsf15, ARG91 was linked to the substrate via hydrogen bonds (Fig. 13a). By contrast, PeHsf28 (Fig. 13b) and PeHsf26 (Fig. 13c), representing the HsfB and HsfC families, showed no hydrogen bonding with the substrate owing to their lack of AHA motifs.

4. Discussion
Hsfs are a specific class of transcription factors that play an important role in plant resistance to various stress injuries [58-60]. The Hsf gene family has been studied in Arabidopsis [10], tomato [15], cabbage [16], pepper [17], poplar [18], maize [19], and rice [20], and specific molecular functions of the Hsfs have been validated in the model plants Arabidopsis and rice. However, Hsfs have not previously been investigated in the economically important and widely distributed graminaceous species moso bamboo. At present, the genome draft of moso bamboo is almost complete, enabling the comprehensive characterization of important gene families [33, 34]. Here, we identified 41 PeHsf genes from moso bamboo and classified them into three subfamilies (A–C) based on phylogenetic analysis (Fig. 1). Gene structure analysis revealed that some
members within the same subfamily had structural differences. For example, PeHsf31 in the HsfC subfamily contained three introns, whereas the other HsfCs had only one. PeHsf31 may therefore have experienced the splicing or insertion of gene fragments during its evolution [61, 62]. Nonetheless, similar conserved sequences and gene structures among Hsf subfamily members suggests that genes within a subfamily may generally have similar biological functions.

The highly conserved plant Hsf DBD is located at the N terminus, where it enables the precise localization and recognition of the heat stress element (HSE) in target gene promoters [63, 64]. Here, multiple sequence alignment and secondary structure prediction showed that a DBD consisting of three α-helices and four β-folds was present in all PeHsf proteins (Fig. 2). Interestingly, other conserved domains were also present in some PeHsf proteins, and further experimental verification is needed to determine whether this is a manifestation of gene family functional diversification. Tertiary structure analysis demonstrated that the portion of the transcription factors that interacts with nucleic acids was conserved among subfamilies. HsfB and HsfC members did not contain AHA motifs and therefore probably lack transcriptional activation function, consistent with the results of previous studies [14].

Gene duplication plays an important role in evolution by facilitating the generation of new genes and gene functions. There are three main evolutionary modes of gene duplication [10]: segmental duplication, tandem duplication, and translocation events. Segmental and tandem duplication most commonly underlie the expansion of plant gene families [57, 65]. The number of Hsf genes was significantly higher in moso bamboo (41) than in the model plants Arabidopsis (21) and rice (25), suggesting that gene duplication had occurred, consistent with previous reports of whole-genome duplication in moso bamboo [32, 66]. We therefore performed intra- and inter-genomic collinearity analyses of the Hsf s. Within the moso bamboo genome, there were 29 duplicated gene pairs among the Hsf genes, including 27 segmental duplicate pairs and two tandem duplicate pairs. Segmental duplications therefore dominated the expansion of the Hsf gene family in moso bamboo. By comparison, a previous study reported nine segmental duplicate pairs in 25 rice Hsf genes [54]. Synteny analysis of the moso bamboo genome with four other sequenced plant genomes showed that there was significant collinearity of Hsf family members between bamboo and the monocots maize and rice. Only a few Hsf members were collinear between bamboo and the dicots Arabidopsis and pepper. This result is consistent with the evolutionary relationship between dicot and monocot plants.

There are four main fates of duplicated genes [67]. First, both genes may retain their original function, leading to functional redundancy. Alternatively, duplicates with the same ancestral function may gradually develop different functions in a process referred
to as subfunctionalization. In a less common scenario, one of the duplicates may acquire a wholly new function (neofunctionalization). Finally, the intermediate evolutionary stage between subfunctionalization and neofunctionalization preserves genes that are essential for plant growth. We found that many PeHsf homologous gene pairs were expressed at similar levels (Fig. 10), suggesting that duplication of the Hsf genes in moso bamboo has primarily resulted in functional redundancy.

Ka and Ks values were calculated for thirteen homologous, segmentally duplicated gene pairs. Ks values indicated that large-scale PeHsf gene duplication events had occurred between ~7.86 and 83.33 MYA. Interestingly, about 62% of the homologous gene pairs appeared to have undergone duplication far earlier than 12 MYA, consistent with the occurrence of a separate whole-genome duplication (WGD) event prior to the WGD previously reported to have occurred 7–12 MYA [32]. In addition, all the homologous gene pairs had Ka/Ks values less than 1, indicating that the PeHsf genes were subject to purifying selection during their evolutionary history.

Flowering is the most important life history trait in angiosperm plants, and bamboo is no exception. Unlike many other bamboos, moso bamboo produces a single-flowering spike approximately every fifty years, blooms sporadically, and dies after flowering. Several studies have demonstrated that Hsfs modulate the expression of stress-related proteins such as heat shock proteins (HSPs) that play important roles in plant stress response [68]. HSPs such as HSP70 and HSP90 are involved in flowering time regulation and vernalization pathways that promote flowering by inhibiting FLC expression [69]. Based on Hsf expression profiling in different bamboo tissues (Fig. 10C), we found that many PeHsf showed relatively high expression levels in the panicle. For example, orthologs of rice OsHsfB2b (PeHsf27, PeHsf10, and PeHsf38) were highly expressed in panicles, and a previous study has shown that OsHsfB2b acts as a negative regulator of plant drought response [20]. These PeHsf genes may therefore be involved not only in panicle development but also in the response of moso bamboo to drought stress. Some PeHSF genes showed little to no expression in any tissue tested, suggesting that they may act in other tissues or at other developmental stages. Interestingly, some PeHsfs were highly expressed in leaves but expressed at low levels in early 20-cm bamboo shoots, suggesting that they are mainly involved in cell differentiation and elongation, rather than in the process of lignification.

Multiple cis-acting elements located in gene promoters play a crucial role in signaling, and synergistic interactions among them can regulate complex biological processes. The drought-responsive MYB element was present in the promoter regions of almost all PeHsf genes, as were ABRE, MBS, P-box, TC-rich repeat, and SP1 elements, indicating that PeHsf expression is likely to be affected by abiotic factors such as high temperature, drought, high salinity, and light. Examination of expression
profiles of seedling roots under NAA and GA treatments revealed that most \textit{PeHsf} genes were expressed at different levels under treated and control conditions.

Some genes, (e.g., \textit{PeHsf40}, \textit{PeHsf31}, and \textit{PeHsf13}) were upregulated by GA treatment, whereas others (e.g., \textit{PeHsf33}, \textit{PeHsf06}, \textit{PeHsf03}, \textit{PeHsf35}, and \textit{PeHsf26}) were downregulated. Interestingly, the promoter analysis (Fig. 9) indicated that these genes all contained \textit{cis}-elements involved in ABA response. In rice, multiple \textit{Hsf} genes are induced by exogenous gibberellin 3 (GA3) and abscisic acid (ABA) \cite{70}, and multiple ABRE and P-box elements responsive to ABA and GA, respectively, were present in the \textit{PeHsf} promoters. Promoter analysis and transcriptomic results therefore suggest that \textit{Hsf} genes are involved not only in the regulation of heat response, but also in responses to drought, salt, and various exogenous hormones (NAA, GA, and ABA), making them important transcription factors for plant resistance to multiple stresses.

5. Conclusions

We systematically identified and analyzed 41 moso bamboo \textit{Hsf} genes and divided them into three subfamilies, each of which had similar gene structures and sequences. Evolutionary analysis indicated that segmental duplications associated with whole-genome duplication events were responsible for much of the expansion of the moso bamboo \textit{Hsf} gene family. Transcriptomic analyses confirmed that expression of some \textit{PeHSF}s responded to exogenous GA and NAA application, and the high expression of other \textit{PeHsf}s in 20-cm shoots suggested that they may function in the rapid growth and early development of bamboo shoots. The \textit{PeHsf} gene family appears to be functionally diverse, with roles in bamboo growth and development, stress response, and the functions of specific tissues and organs. These results provide a foundation for the subsequent exploration of bamboo \textit{Hsf} gene functions.


Ogawa, D., Yamaguchi, K. & Nishiuchi, T. High-level overexpression of the Arabidopsis HsfA2 gene confers not only increased thermotolerance but also


Acknowledgements

This work was supported by National Natural Science Foundation of China (NSFC, Project grant: 31770721). We are grateful to the members of the lab for their assistance and helpful discussions. The authors would like to thank TopEdit (www.topeditsci.com) for its linguistic assistance during the preparation of this manuscript.

Author contributions

B.H. performed data collection and processing, participated in study design and interpretation, performed experiments, and wrote the manuscript. Z.H. participated in some of the experiments. R.M. and J.C. assisted in the interpretation of the results. Z.Z. and K.Y. are responsible for the completeness of the data and accuracy of the data analysis. The manuscript was reviewed and edited by all authors.

Competing interests

No potential conflicts of interest relevant to this article were reported.
Phylogenetic analysis of the PeHsf proteins

Figure 1. Phylogenetic analysis of full-length Hsf protein sequences from *Phyllostachys edulis* (Pe, bamboo), *Arabidopsis thaliana* (At, Arabidopsis), and *Oryza sativa* (Os, rice). ClustalX was used to build a multiple sequence alignment, and MEGA7.0 was used to construct a maximum likelihood (ML) phylogenetic tree with 1000 bootstrap replicates. Blue, red, and green circles indicate bamboo, Arabidopsis, and rice sequences, respectively.

*Phyllostachys edulis* (Pe, bamboo), *Arabidopsis thaliana* (At, Arabidopsis), and *Oryza sativa* (Os, rice). ClustalX was used to build a multiple sequence alignment, and MEGA7.0 was used to construct a maximum likelihood (ML) phylogenetic tree with 1000 bootstrap replicates. Blue, red, and green circles indicate bamboo, Arabidopsis, and rice sequences, respectively.
Gene structures of the PeHsf proteins

Figure 2. Intron-exon structure of the Hsf genes. Green boxes represent exons (CDS), black lines represent introns, and light blue boxes represent 5’ and 3’ untranslated regions.
**Figure 3.** Multiple sequence alignment of the DNA-binding domains (DBDs) of Hsf proteins from moso bamboo. Secondary structure elements of the DBDs (α1-β1-α2-α3-β3-β4) based on JNet structure predictions are shown above the alignment (see Methods). α-helices are indicated by red bars, and β-folds are indicated by green arrows.
Figure 4. Distribution of conserved motifs in PeHsf proteins. The scale bar at the bottom indicates the protein lengths, and sequence logos for each conserved motif are shown at right.
**Figure 5.** Conserved domain predictions for 41 PeHsf proteins. The gray bars represent the length of each protein sequence, and conserved domains are represented by colored boxes.
Chromosomal distribution

Figure 6. The distribution of *PeHsf* genes on the moso bamboo chromosome scaffolds.
Intra- and inter-species synteny analysis of the PeHsf genes

Figure 7. Chromosomal distribution and interchromosomal relationships of PeHsf genes. Gray lines indicate syntenic blocks within the moso bamboo genome, and red lines represent duplicate PeHsf gene pairs.
Figure 8. Synteny analysis of the moso bamboo genome with two monocot (a) and two dicot (b) plant genomes. The gray lines represent aligned blocks between the paired genomes, and the red lines indicate syntenic Hsf gene pairs.

Promoter cis-element analysis of PeHsf genes

Figure 9. Diagram of cis-acting elements identified in the 1,500-bp upstream promoter regions of the PeHsf genes. Rectangular boxes of different colors represent individual cis-acting elements, and some elements overlap with one another.
Figure 10. Heat maps of PeHsf gene expression (log_{10}-transformed TPM values) in multiple tissues and in response to hormone treatments. (a) GA treatment. (b) NAA treatment. (c) Roots, rhizomes, panicles, and leaves. Each tissue is the average of three replicates. (d) Young bamboo shoots of different heights. CK, control group. Each tissue or treatment was replicated three times.
Table 1. Detailed information on 41 PeHsf genes and their encoded proteins.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Gene ID</th>
<th>Chromosome location</th>
<th>Size (aa)</th>
<th>MW (kDa)</th>
<th>pI</th>
<th>Stability</th>
<th>A.I.</th>
<th>GRAVY</th>
<th>Predicted Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>PeHsf01</td>
<td>PH02Gene27951.t1</td>
<td>S3:17095591–17101377</td>
<td>292</td>
<td>31.90</td>
<td>9.21</td>
<td>U</td>
<td>66.88</td>
<td>−0.601</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf02</td>
<td>PH02Gene22062.t1</td>
<td>S3:17300229–17302543</td>
<td>385</td>
<td>41.30</td>
<td>8.86</td>
<td>U</td>
<td>68.83</td>
<td>−0.359</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf03</td>
<td>PH02Gene32485.t1</td>
<td>S4:5279994–5282107</td>
<td>159</td>
<td>17.49</td>
<td>5.82</td>
<td>U</td>
<td>68.62</td>
<td>−0.484</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf04</td>
<td>PH02Gene23029.t2</td>
<td>S5:34339214–34343740</td>
<td>382</td>
<td>43.06</td>
<td>4.72</td>
<td>U</td>
<td>83.25</td>
<td>−0.539</td>
<td>Cytoplasmic</td>
</tr>
<tr>
<td>PeHsf05</td>
<td>PH02Gene21935.t1</td>
<td>S6:50181002–50183068</td>
<td>276</td>
<td>29.02</td>
<td>6.98</td>
<td>S</td>
<td>63.73</td>
<td>−0.264</td>
<td>Extracellular</td>
</tr>
<tr>
<td>PeHsf06</td>
<td>PH02Gene24187.t1</td>
<td>S7:43016455–43019591</td>
<td>445</td>
<td>49.96</td>
<td>4.95</td>
<td>U</td>
<td>67.46</td>
<td>−0.626</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf07</td>
<td>PH02Gene17129.t1</td>
<td>S8:53174946–53177158</td>
<td>348</td>
<td>37.56</td>
<td>4.87</td>
<td>U</td>
<td>81.49</td>
<td>−0.250</td>
<td>Cytoplasmic</td>
</tr>
<tr>
<td>PeHsf08</td>
<td>PH02Gene21435.t1</td>
<td>S9:57449663–57452744</td>
<td>453</td>
<td>51.38</td>
<td>5.23</td>
<td>U</td>
<td>66.29</td>
<td>−0.649</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf09</td>
<td>PH02Gene35749.t1</td>
<td>S10:3519871–3522549</td>
<td>300</td>
<td>33.58</td>
<td>6.35</td>
<td>U</td>
<td>66.33</td>
<td>−0.653</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf10</td>
<td>PH02Gene31077.t2</td>
<td>S13:91174631–91176831</td>
<td>383</td>
<td>40.63</td>
<td>5.40</td>
<td>U</td>
<td>69.16</td>
<td>−0.380</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf11</td>
<td>PH02Gene16951.t1</td>
<td>S13:117714221–117720267</td>
<td>362</td>
<td>41.14</td>
<td>4.98</td>
<td>U</td>
<td>78.37</td>
<td>−0.698</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf12</td>
<td>PH02Gene22176.t1</td>
<td>S14:33627815–33630503</td>
<td>319</td>
<td>35.90</td>
<td>5.76</td>
<td>U</td>
<td>74.92</td>
<td>−0.490</td>
<td>Cytoplasmic</td>
</tr>
<tr>
<td>PeHsf13</td>
<td>PH02Gene40343.t1</td>
<td>S14:40539424–40541046</td>
<td>311</td>
<td>34.31</td>
<td>5.73</td>
<td>U</td>
<td>71.25</td>
<td>−0.388</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf14</td>
<td>PH02Gene18810.t1</td>
<td>S14:54857540–54,858,808</td>
<td>249</td>
<td>27.71</td>
<td>9.19</td>
<td>U</td>
<td>69.92</td>
<td>−0.310</td>
<td>Cytoplasmic</td>
</tr>
<tr>
<td>PeHsf15</td>
<td>PH02Gene15090.t1</td>
<td>S14:57115384–57117986</td>
<td>443</td>
<td>49.15</td>
<td>5.24</td>
<td>U</td>
<td>64.94</td>
<td>−0.732</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf16</td>
<td>PH02Gene17948.t1</td>
<td>S15:38771524–38774540</td>
<td>279</td>
<td>31.40</td>
<td>5.69</td>
<td>U</td>
<td>66.74</td>
<td>−0.638</td>
<td>Nuclear</td>
</tr>
<tr>
<td>Gene</td>
<td>Gene ID</td>
<td>Chromosome</td>
<td>Start</td>
<td>End</td>
<td>Length</td>
<td>Score</td>
<td>Strain</td>
<td>Start (Genome)</td>
<td>End (Genome)</td>
</tr>
<tr>
<td>------------</td>
<td>---------------</td>
<td>------------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>-------</td>
<td>----------</td>
<td>----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>PeHsf17</td>
<td>PH02Gene50090.t1</td>
<td>S15:61679276–61684197</td>
<td>409</td>
<td>45.25</td>
<td>4.84</td>
<td>U</td>
<td>72.47</td>
<td>−0.631</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf18</td>
<td>PH02Gene03550.t1</td>
<td>S15:72232636–72234806</td>
<td>315</td>
<td>36.53</td>
<td>5.05</td>
<td>U</td>
<td>73.4</td>
<td>−0.696</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf19</td>
<td>PH02Gene11863.t1</td>
<td>S15:94500097–94503858</td>
<td>342</td>
<td>38.49</td>
<td>5.32</td>
<td>U</td>
<td>63.54</td>
<td>−0.798</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf20</td>
<td>PH02Gene14728.t1</td>
<td>S16:34492266–34494798</td>
<td>444</td>
<td>49.50</td>
<td>5.34</td>
<td>U</td>
<td>59.35</td>
<td>−0.856</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf21</td>
<td>PH02Gene30851.t1</td>
<td>S16:72232636–72234806</td>
<td>247</td>
<td>26.77</td>
<td>8.93</td>
<td>S</td>
<td>72.43</td>
<td>−0.264</td>
<td>Mitochondrial</td>
</tr>
<tr>
<td>PeHsf22</td>
<td>PH02Gene20958.t1</td>
<td>S16:54912146–54913593</td>
<td>310</td>
<td>34.09</td>
<td>5.86</td>
<td>U</td>
<td>74.29</td>
<td>−0.343</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf23</td>
<td>PH02Gene48407.t1</td>
<td>S16:66912172–66914576</td>
<td>319</td>
<td>35.61</td>
<td>5.67</td>
<td>U</td>
<td>71.88</td>
<td>−0.455</td>
<td>Cytoplasmic</td>
</tr>
<tr>
<td>PeHsf24</td>
<td>PH02Gene29649.t1</td>
<td>S17:49074756–49079687</td>
<td>494</td>
<td>54.90</td>
<td>4.99</td>
<td>U</td>
<td>79.17</td>
<td>−0.573</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf25</td>
<td>PH02Gene33162.t1</td>
<td>S17:54687050–54691243</td>
<td>288</td>
<td>30.99</td>
<td>5.81</td>
<td>U</td>
<td>63.33</td>
<td>−0.456</td>
<td>Chloroplast</td>
</tr>
<tr>
<td>PeHsf26</td>
<td>PH02Gene32696.t1</td>
<td>S17:80486587–80489026</td>
<td>360</td>
<td>38.81</td>
<td>5.04</td>
<td>U</td>
<td>65.69</td>
<td>−0.649</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf27</td>
<td>PH02Gene26216.t1</td>
<td>S18:26517557–26519682</td>
<td>296</td>
<td>32.43</td>
<td>8.85</td>
<td>U</td>
<td>64.63</td>
<td>−0.605</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf28</td>
<td>PH02Gene01234.t1</td>
<td>S18:37635773–37642186</td>
<td>383</td>
<td>40.93</td>
<td>8.65</td>
<td>U</td>
<td>64.63</td>
<td>−0.381</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf29</td>
<td>PH02Gene21649.t1</td>
<td>S18:38023397–38026522</td>
<td>485</td>
<td>53.62</td>
<td>5.19</td>
<td>U</td>
<td>57.01</td>
<td>−0.764</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf30</td>
<td>PH02Gene30467.t1</td>
<td>S20:5525280–5530407</td>
<td>334</td>
<td>36.32</td>
<td>6.67</td>
<td>U</td>
<td>67.99</td>
<td>−0.441</td>
<td>Chloroplast</td>
</tr>
<tr>
<td>PeHsf31</td>
<td>PH02Gene40145.t1</td>
<td>S20:31531058–31534365</td>
<td>512</td>
<td>56.33</td>
<td>5.06</td>
<td>U</td>
<td>77.95</td>
<td>−0.501</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf32</td>
<td>PH02Gene05710.t1</td>
<td>S20:52393478–52397181</td>
<td>507</td>
<td>55.29</td>
<td>4.88</td>
<td>U</td>
<td>73.16</td>
<td>−0.448</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf33</td>
<td>PH02Gene08915.t1</td>
<td>S21:1225128–1229768</td>
<td>358</td>
<td>40.28</td>
<td>5.40</td>
<td>U</td>
<td>66.15</td>
<td>−0.812</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf34</td>
<td>PH02Gene06729.t1</td>
<td>S21:9161083–9164543</td>
<td>278</td>
<td>31.78</td>
<td>5.70</td>
<td>U</td>
<td>66.22</td>
<td>−0.695</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf35</td>
<td>PH02Gene14439.t1</td>
<td>S21:64405316–64407893</td>
<td>392</td>
<td>43.64</td>
<td>5.14</td>
<td>U</td>
<td>72.37</td>
<td>−0.598</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf36</td>
<td>PH02Gene26611.t1</td>
<td>S21:96620845–96633311</td>
<td>362</td>
<td>41.44</td>
<td>5.05</td>
<td>U</td>
<td>72.49</td>
<td>−0.692</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf37</td>
<td>PH02Gene09168.t2</td>
<td>S21:105791731–105794734</td>
<td>372</td>
<td>39.87</td>
<td>4.91</td>
<td>U</td>
<td>71.75</td>
<td>−0.412</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf38</td>
<td>PH02Gene24117.t1</td>
<td>S22:2617776–2619602</td>
<td>316</td>
<td>39.36</td>
<td>9.31</td>
<td>U</td>
<td>73.32</td>
<td>−0.345</td>
<td>Nuclear</td>
</tr>
<tr>
<td>PeHsf39</td>
<td>PH02Gene46516.t1</td>
<td>S22:15546435–15548028</td>
<td>331</td>
<td>39.36</td>
<td>9.31</td>
<td>U</td>
<td>73.32</td>
<td>−0.345</td>
<td>Nuclear</td>
</tr>
</tbody>
</table>
### PeHsf40

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Transcript ID</th>
<th>Chromosome</th>
<th>Start</th>
<th>End</th>
<th>MW</th>
<th>pI</th>
<th>A.I.</th>
<th>GRAVY</th>
<th>Nuclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>PH02Gene01898.t1</td>
<td>S23:17548422–17549945</td>
<td>292</td>
<td>31.29</td>
<td>5.35</td>
<td>U</td>
<td>59.59</td>
<td>−0.771</td>
<td>Nuclear</td>
<td></td>
</tr>
</tbody>
</table>

### PeHsf41

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Transcript ID</th>
<th>Chromosome</th>
<th>Start</th>
<th>End</th>
<th>MW</th>
<th>pI</th>
<th>A.I.</th>
<th>GRAVY</th>
<th>Nuclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>PH02Gene00565.t1</td>
<td>S24:57049346–57050791</td>
<td>321</td>
<td>34.22</td>
<td>5.22</td>
<td>U</td>
<td>59.13</td>
<td>−0.743</td>
<td>Nuclear</td>
<td></td>
</tr>
</tbody>
</table>

MW, molecular weight. pI, isoelectric point. A.I., aliphatic index. GRAVY, grand average of hydropathicity score.
Table 2. Ka/Ks values of homologous *PeHsf* gene pairs.

<table>
<thead>
<tr>
<th>Gene pair</th>
<th>Ka</th>
<th>Ks</th>
<th>Ka/Ks</th>
<th>MYA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PeHsf01–PeHsf28</td>
<td>0.0315</td>
<td>0.3727</td>
<td>0.0845</td>
<td>28.6692</td>
</tr>
<tr>
<td>PeHsf04–PeHsf11</td>
<td>0.9711</td>
<td>1.0833</td>
<td>0.8964</td>
<td>83.3330</td>
</tr>
<tr>
<td>PeHsf06–PeHsf08</td>
<td>0.4283</td>
<td>0.6143</td>
<td>0.6972</td>
<td>47.2538</td>
</tr>
<tr>
<td>PeHsf15–PeHsf20</td>
<td>0.0550</td>
<td>0.1113</td>
<td>0.4936</td>
<td>8.5615</td>
</tr>
<tr>
<td>PeHsf14–PeHsf21</td>
<td>0.0342</td>
<td>0.3268</td>
<td>0.1048</td>
<td>25.1384</td>
</tr>
<tr>
<td>PeHsf12–PeHsf23</td>
<td>0.0582</td>
<td>0.1023</td>
<td>0.5690</td>
<td>7.8692</td>
</tr>
<tr>
<td>PeHsf18–PeHsf37</td>
<td>0.9907</td>
<td>1.0268</td>
<td>0.9649</td>
<td>78.9846</td>
</tr>
<tr>
<td>PeHsf19–PeHsf34</td>
<td>0.0430</td>
<td>0.1212</td>
<td>0.3546</td>
<td>9.3230</td>
</tr>
<tr>
<td>PeHsf16–PeHsf35</td>
<td>0.0350</td>
<td>0.2788</td>
<td>0.1254</td>
<td>21.4461</td>
</tr>
<tr>
<td>PeHsf17–PeHsf36</td>
<td>0.0557</td>
<td>0.1062</td>
<td>0.5248</td>
<td>8.1692</td>
</tr>
<tr>
<td>PeHsf22–PeHsf13</td>
<td>0.0577</td>
<td>0.2749</td>
<td>0.2100</td>
<td>21.1461</td>
</tr>
<tr>
<td>PeHsf24–PeHsf32</td>
<td>0.9893</td>
<td>1.0366</td>
<td>0.9544</td>
<td>79.7384</td>
</tr>
<tr>
<td>PeHsf25–PeHsf30</td>
<td>0.0420</td>
<td>0.1145</td>
<td>0.3670</td>
<td>8.8076</td>
</tr>
</tbody>
</table>

Ka, non-synonymous substitution rate. Ks, synonymous substitution rate. MYA, million years ago.
Figures

Figure 1

Phylogenetic analysis of full-length Hsf protein sequences from Phyllostachys edulis (Pe, bamboo), Arabidopsis thaliana (At, Arabidopsis), and Oryza sativa (Os, rice). ClustalX was used to build a multiple sequence alignment, and MEGA7.0 was used to construct a maximum likelihood (ML) phylogenetic tree with 1000 bootstrap replicates. Blue, red, and green circles indicate bamboo, Arabidopsis, and rice sequences, respectively.
Figure 2

Intron-exon structure of the Hsf genes. Green boxes represent exons (CDS), black lines represent introns, and light blue boxes represent 5' and 3' untranslated regions.
Figure 3

Multiple sequence alignment of the DNA-binding domains (DBDs) of Hsf proteins from moso bamboo. Secondary structure elements of the DBDs (α1-β1-β2-α2-α3-β3-β4) based on JNet structure predictions are shown above the alignment (see Methods). α-helices are indicated by red bars, and β-folds are indicated by green arrows.
Figure 4

Distribution of conserved motifs in PeHsf proteins. The scale bar at the bottom indicates the protein lengths, and sequence logos for each conserved motif are shown at right.
Figure 5

Conserved domain predictions for 41 PeHsf proteins. The gray bars represent the length of each protein sequence, and conserved domains are represented by colored boxes.
Figure 6

The distribution of PeHsf genes on the moso bamboo chromosome scaffolds.
Figure 7

Chromosomal distribution and interchromosomal relationships of PeHsf genes. Gray lines indicate syntenic blocks within the moso bamboo genome, and red lines represent duplicate PeHsf gene pairs.
Figure 8

Synteny analysis of the moso bamboo genome with two monocot (a) and two dicot (b) plant genomes. The gray lines represent aligned blocks between the paired genomes, and the red lines indicate syntenic Hsf gene pairs.
Diagram of cis-acting elements identified in the 1,500-bp upstream promoter regions of the PeHsf genes. Rectangular boxes of different colors represent individual cis-acting elements, and some elements overlap with one another.
Heat maps of PeHsf gene expression (log10-transformed TPM values) in multiple tissues and in response to hormone treatments. 
(a) GA treatment. (b) NAA treatment. (c) Roots, rhizomes, panicles, and leaves. Each tissue is the average of three replicates. (d) Young bamboo shoots of different heights. CK, control group. Each tissue or treatment was replicated three times.
Figure 11

GO enrichment analysis of the PeHsf proteins relative to the GO database. The horizontal axis indicates the enrichment factor, and the size of the circle indicates the number of genes annotated with a given GO term.
Figure 12

Interaction network of the Hsf proteins in moso bamboo. Nodes represent proteins, central nodes are indicated in red, and black lines indicate interactions between nodes.
Three-dimensional structure of the moso bamboo Hsf protein sequences. (a) PeHsf15 was selected to represent the structure of the HsfA subfamily. (b) PeHsf28 was selected to represent the structure of the HsfB subfamily. (c) PeHsf26 was selected to represent the structure of the HsfC subfamily. Models were constructed using SWISS-MODEL. ARG, arginine. GOL, substrate.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- S2Table.TPMvaluesofmosobambooHSFtranscriptionfactorsindifferentorgansgerminationstagesandhormonaltreatments.xlsx
- S1Table.Detailedinformationongeneontologyenrichmentanalysis.xlsx
- S4Table.DetailedinformationonMosobambooHSFproteinnetworknodes.xlsx
- S3Table.Detailedinformationonintraandinterspeciessyntenyofthemosobamboogenome.xlsx