Genome-wide identification, classification and expression profile analysis of the HSF gene family in *Hypericum perforatum* (#52717)

First submission

Guidance from your Editor

Please submit by **4 Oct 2020** for the benefit of the authors (and your $200 publishing discount) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

8 Figure file(s)
5 Table file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Genome-wide identification, classification and expression profile analysis of the HSF gene family in *Hypericum perforatum*

Li Zhou [1] , **Xiaoding Yu** [1] , **Donghao Wang** [1] , **Lin Li** [1] , **Wen Zhou** [1] , **Qian Zhang** [1] , **Zhezhi Wang** [Corresp. 1]

[1] National Engineering Laboratory for Resource Development of Endangered Crude Drugs in Northwest China, Key Laboratory of the Ministry of Education for Medicinal Resources and Natural Pharmaceutical Chemistry, College of Life Sciences, Shaanxi Normal University, Xi'an, Shaanxi, China

Corresponding Author: Zhezhi Wang
Email address: zzwang@snnu.edu.cn

The heat shock transcription factors (HSFs) are critical regulators in plant responses to various abiotic and biotic stresses. They involve in regulating the expression of heat shock proteins (HSPs) by binding with heat stress elements (HSEs) to defense heat stress. Recently, the Hypericum perforatum genome has been fully sequenced, which provide a valuable resource for functional analysis. In this study, a total of 23 HpHSF genes were identified and divided into three groups (A, B, and C) based on their phylogeny and structural features. Gene structure and conserved motif analyses revealed that all HpHSF genes exhibit relatively conserved domains. In addition, various cis-acting elements in the promoter regions of HpHSFs are related to hormone and stress responses. And the transcriptional levels of most HpHSF genes was altered under heat stress conditions, suggesting their potential functions in heat stress resistance pathway. Our findings are helpful for further functional analysis of HpHSFs.

Genome-wide identification, classification and expression Profile analysis of the HSF gene family in Hypericum perforatum

Li Zhou, Xiaoding Yu, Donghao Wang, Lin Li, Wen Zhou, Qian Zhang, Zhezhi Wang*

National Engineering Laboratory for Resource Development of Endangered Crude Drugs in Northwest China, Key Laboratory of the Ministry of Education for Medicinal Resources and Natural Pharmaceutical Chemistry, College of Life Sciences, Shaanxi Normal University, Xi'an, China

*Corresponding Author:
Zhezhi Wang
Shaanxi Normal University, Xi'an, Shaanxi, 710100, China
Email address: zzwang@snnu.edu.cn

Abstract

The heat shock transcription factors (HSFs) are critical regulators in plant responses to various abiotic and biotic stresses. They involve in regulating the expression of heat shock proteins (HSPs) by binding with heat stress elements (HSEs) to defense heat stress. Recently, the Hypericum perforatum genome has been fully sequenced, which provide a valuable resource for functional analysis. In this study, a total of 23 HpHSF genes were identified and divided into three groups (A, B, and C) based on their phylogeny and structural features. Gene structure and conserved motif analyses revealed that all HpHSF genes exhibit relatively conserved domains. In addition, various cis-acting elements in the promoter regions of HpHSFs are related to hormone and stress responses. And the transcriptional levels of most HpHSF genes were altered under heat stress conditions, suggesting their potential functions in heat stress resistance pathway. Our findings are helpful for further functional analysis of HpHSFs.

Introduction

Plants often suffer from divergent biotic and abiotic stresses such as virus infection, vegetarian attack, drought, salt, high and low temperature and so on throughout their life cycles (Abdelrahman, et al. 2018; Zandalinas, et al. 2018). They have many complex defense mechanisms in vivo to protect themselves from stressful environment. Among various abiotic stresses, high temperature has significant impact on plant survival. Under heat stress, Heat shock transcription factors (HSFs) can activate rapid accumulation and expression of heat shock proteins (HSPs) to reduce heat-related damage. Many HSPs play a critical role in protecting the plants from stress damage, as well as function in protein folding, aggregation, degradation, and intracellular distribution (Mittler, et al. ; Lin, et al. 2011). In the process of heat shock reaction, HSFs regulate the expression of heat stress-inducible genes by recognizing the binding motifs called heat stress elements (HSEs) which present in promoter regions of the HSP genes (Treuter, et al.). Specifically, HSFs utilize their oligomerization domains to form trimmers and take effect as sequence-specific trimeric DNA binding proteins. Previous studies have shown that transcription activation in vivo requires at least three repeat HSEs when binding by HSF proteins (Drees, et al. 1997).

Recently, the Genome-wide analysis of HSF gene family in more than 20 plants were carried out and it is clear that the number of HSF gene members are varies among different species (Scharf, et al. ; Fujimoto and Nakai 2010). For instance, A total of 22 HSF genes were identified in Arabidopsis (Arabidopsis thaliana), 25 in maize (Zea mays）(Lin, et al. 2011), 25 in rice (Oryza sativa)(Guo, et al. 2008), 25 in pepper(Capsicum annuum L.)(Guo, et al. 2015), 26 in tomato (Solanumlycopersicum)(Yang, et al. 2016), 38 in soybean (Glycine max)(Li, et al. 2014), 32 in Populus euphratica(Zhang, et al. 2016), 17 in woodland strawberry (Fragaria vesca)(Hu, et al. 2015) and 24 in mungbean (Vigna radiata)(Li, et al. 2018), indicating that HSF proteins in various species may have similar but different functions in reducing stress damage, and also provide rich resources for evolutionary analysis.

In plant HSFs, the members share a typical and conserved modular structure. The highly conserved DNA-binding domain (DBD) in the N-terminus includes one three-helical bundle (α1, α2, α3) and one antiparallel four-stranded β-sheet (β1, β2, β3, β4) to form a helix-turn-helix structure, which is demanded for HSEs specific binding to regulate the expression of downstream genes(Scharf, et al. ; Guo, et al. 2016). The oligimerisation domain (OD), also known as the HR-A/B region, has the characteristic of coiled-coil structure and play a part in the transcription factor activity. It is mainly located at the C-terminal of HSF and connected to the DBD through a flexible linker comprising of a heptad pattern of hydrophobic amino acid residues (Peteranderl, et al. 1999). In addition, the nuclear localization signal (NLS) at the C-terminal of HR-A/B region consisting of a cluster of basic amino acid rich in lysine and arginine residues is essential for nuclear import, and the nuclear export signal (NES) in the C-terminal of some HSF genes, which contains many leucine residues, is crucial to regulates the nucleocytoplasmic distribution of HSF proteins(Lyck, et al. 1997; Chidambaranathan, et al. 2018). Furthermore, there are short peptide motifs (AHA motifs) closing to the C-terminal for transcriptional activator functions in some HSF proteins (Kotak, et al. 2004).

According to the characteristic of the conserved DBD domain and HR-A/B regions, HSFs in plants are classified into three main classes (class A, B, and C) (Nover, et al. 2001). The number of amino acid residues connecting DBD to HR-A/B was different among the three subgroups. Class A contains 9-39 amino acid residues, class B contains 50-78 amino acid residues, and class C contains 4-49 amino acid residues (Prändl, et al. 1998; Miller and Mittler 2006). Moreover, the number of amino acids linking HR-A and HR-B also had obvious variation in different subgroups. There are 21 and 7 amino acid residues inserted into the HR-A/B region in class A and class C, respectively, whereas this region in class B HSFs is compact without insert sequences between the heptad repeats (Baniwal, et al. 2004). Additionally, The AHA motifs, which function through binding some transcription protein complexes to activates the transcription of HSPs, are unique to class A members but not in class B or class C (Scharf, et al.).

Hypericum perforatum is a herbaceous perennial plant in the family Hypericaceae. The well-characterized secondary metabolites and pharmacological activities have attracted the attention of researchers (Galeotti 2017). The extracts of H. perforatum include acyl-phloroglucinols, naphthodianthrones, xanthones and flavonoids, and these various pharmacological compounds are related to antiviral, antitumoural, anti-inflammatory, antimicrobial, antioxidant, and other functions (Nahrstedt and Butterweck 2010). However, the production and quality of H. perforatum are challenged by various stresses from environment, such as cold, high temperature, drought etc. Therefore, it is important to characterize H. perforatum stress resistant genes. The current study identified 23 HpHSF genes and analyzed their physical and chemical characters, conserved domains, gene structures, evolutionary relationships and cis-acting elements. Moreover, we explored the expression profiles across four different tissues and under heat stress treatment. In conclusion, it provides a foundation for an improved exploration of the HpHSF gene function in H. perforatum.

Materials & Methods

Plant Material and Treatment

Seeds (2n=2x=16) of Hypericum perforatum preserved by our laboratory were germinated and grown on a seedling bed in the greenhouse (25 ± 2°C, natural lighting). Humidity was maintained at 60–80%. Two months old H. perforatum seedlings were transferred to an incubator maintained at 42°C for heat stress treatments, and five time points (0 h, 1 h, 3h, 6h and 12 h) were selected for sample collection. In addition, the different tissue samples include flower, leaf, stem and root were taken from two-year-old plants. All samples were collected in three replicates, and the samples need to be immersed in liquid nitrogen immediately and stored at -80°C for RNA isolation.

Identification of HpHSF Members

For HSF identification, the conserved amino acid sequence of DNA-binding domains (Pfam： PF00447) was used to search in the H. perforatum genome. Moreover, the HSF protein sequences of Capsicum annuum L., Vitis vinifera L. and A. thaliana. obtained from plantTFDB (http://planttfdb.cbi.pku.edu.cn) were used as BLAST queries against the H. perforatum genome. All output genes with default were searched for conserved DNA-binding domain using Interpro (http://www.ebi.ac.uk/interpro/) and SMART (http://smart.embl-heidelberg.de/). In addition, the remained genes were analyzed using MARCOIl (http://toolkit.tuebingen.mpg.de/marcoil) to remove genes without coiled-coil structure. The detected genes are listed in Supplementary Table1.

Phylogenetic Relationship Analysis and Sequence analysis

Full-length amino acid sequences of HSF from A. thaliana, Capsicum annuum L., Vitis vinifera L. and Hypericum perforatum (this study) were aligned using the Clustal X, and the phylogenetic tree was constructed using MEGA 6.0's Neighbor-Joining (NJ) method with 1000 bootstrap replicates and pairwise deletion.

The parameters including molecular weight, isoelectric point, aliphatic index, instability index, the percentage of negatively/positively charged residues, and GRAVY of HpHSF proteins were displayed using ExPASy database (https://www.expasy.org/). Furthermore, the conserved motifs of HpHSF genes were searched via Multiple Em for Motif Elicitation (MEME, http://meme-suite.org/tools/meme) and the exon/intron organization of HpHSF proteins was obtained by the Gene Structure Display Server program (GSDS, http://gsds.cbi.pku.edu.cn/). The cis-acting elements of 1.5 kb upstream sequences of the transcription initiation site in promoter region of HpHSF genes were analyzed on PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

Isolation of RNA and cDNA Synthesis

Total RNA of H. perforatum materials was isolated using the HiPure Total RNA Mini Kit following the manufacturer's protocol (Magen, China). The concentration of the isolated total RNA was detected through NanoDrop 2000c spectrophotometer (Thermo Scientific, USA), and the integrity of the RNA was directly quantified by running agarose gel (1% w/v) with 150 V, 10 min. 1 μg RNA was used for the first strand cDNA synthesis using PrimeScriptTM RT Reagent Kit (TaKaRa, China) according to the instructions. All cDNA samples should be stored at -80°C and avoid repeated freezing and thawing for RT-qPCR.

Primer Design and Quantitative RT-PCR (qRT-PCR) Analysis

The primers of the 23 HpHSF genes were designed by GenScript (https://www.genscript.com), the parameters were: PCR Amplicon Size Range:100-180; Primer Tm: Minimum, Optimum and Maximum are 59.5 ℃, 60 ℃, 60.5 ℃ respectively；Probe Tm: Minimum, Optimum and Maximum are 62 ℃, 66 ℃, 70 ℃ respectively. The specificity of the primers were detected by Bioedit through searching the primers which given by GenScript against the H. perforatum genome(Supplementary Table2). In addition, quantitative RT-PCR was performed on the LightCycler 96 system (Roche Diagnostics GmbH) using ChamQTM SYBR® qPCR Master Mix (Vazyme, Nanjing, China) following the manufacturer's procedure. The HpActin-2 was used as an internal control and the $2^{-\Delta\Delta Ct}$ method was used to analyze the relative changes in gene expression. Quantitative RT-PCR was done with three biological replicates of each sample and each sample consisted of three technical replicates. The primers of HpHSF genes used for qRT-PCR analyses are listed in Supplementary Table 1.

Results

Identification and Isolation of HSF Genes in the H. perforatum

There are 23 genes were identified as members of the HSF transcription factor family in H. perforatum based on a conserved DBD domain search and the coiled-coil structure detection. These genes were named after 'HpHSF' with the consecutive number. More detailed information about HpHSF01 to HpHSF23 are shown in Table 1, the identified HpHSFs encode 188 to 501 amino acids (average of 345 aa), and molecular weights (MW) ranged from 21.72 kDa to 54.91

kDa (average of 39.15 kDa). The isoelectric points (pI) of HpHSFs varied from 4.79 to 8.86. Among the 23 HpHSF genes, the percentages of negatively charged residues (ASP + Glu) (n.c.r.) and positively charged residues (Arg + Lys) were 11.0% - 17.6% and 8.4% - 15.8%, respectively. According to the instability index analysis, all the HpHSF proteins were unstable. In addition, the aliphatic index (A.I.) ranged from 54.52 to 76.18 and the grand average of hydropathicity (GRAVY) had a range of −0.826 to −0.523.

Conserved domains of HpHSFs

Five conserved domains were observed in the majority HpHSF genes, in order to reveal the sequence conservative regions between members of the HpHSFs, the multiple alignment of 23 HpHSFs was obtained by DNAMAN. From the Figure 1, the DBD domain being close to the N-terminal was highly conserved among all amino acids. And the secondary structure prediction showed that the majority of the DBD domains consist of a four-stranded antiparallel β-sheet and three α-helices (α1~α3). In addition, MARCOIL was used for predicting the coiled-coil structure characteristic of the HR-A/B regions which adjacent to the DBD domain in the C-terminal, it was proved that the 23 candidate HpHSF protein sequences all had coiled-coil structure, the multiple alignment results of the HR-A/B regions shows that the HpHSF protein family can be divided into three classes because of the insertion amino acid residues between the A and B parts of the HR-A/B motif (Figure2).

Phylogenetic relationship of HpHSF genes

To investigate the evolutionary relationships of the HpHSF genes, a total of 88 HSFs, comprising 21 from Arabidopsis, 25 from pepper, 19 from grape and 23 from H. perforatum were used for phylogenetic tree construction by MEGA6.0. Obviously, HSFs were classified into three main groups namely HSF A, B and C (Figure 3). HpHSF A was the largest group which represented 52.2 % of the total HpHSFs; the second was HpHSF B which represented 39.1%; and HpHSF C was the smallest group which just represented 8.7%. In addition, HpHSF A is classified into 9 subgroups (A1-A9) and includes 12 members (*HpHSF07*, *HpHSF08*, *HpHSF12*, *HpHSF11*, *HpHSF16*, *HpHSF21*, *HpHSF17*, *HpHSF02*, *HpHSF23*, *HpHSF13*, *HpHSF10*, *HpHSF20*); HpHSF B is further divided into 5 subgroups (B1-B5) consisting of 9 members (*HpHSF01*, *HpHSF03*, *HpHSF04*, *HpHSF05*, *HpHSF06*, *HpHSF14*, *HpHSF15*, *HpHSF19*, *HpHSF22*); while HpHSFC only contained *HpHSF08* and *HpHSF09* in one subgroup.

Gene Structures analysis and motifs of HpHSFs

The gene structures of PeuHSFs were investigated through an analysis of the intron/exon boundaries, as can be seen from the Figure 4a， *HpHSF20* were comprised of three exons, and *HpHSF03* were comprised of four exons, Except for the aforementioned two HpHSFs, all the other 21 HpHSFs contained two exons and one intron. The intron phases of HpHSFs were 0, except for phase 1 in *HpHSF20* and phase 2 in *HpHSF03*. In conclusion, the gene structure was conserved among the 23 HpHSF members.

In addition, we searched for motifs of the HpHSF proteins to reveal the conserved features using MEME and the results were shown in Figure 4b. Almost all of the HpHSFs contain motifs1, 2 and 3, which corresponded to the most conservative domain, the DBD domain. Motifs 4 and 5 were considered to represent the HR-A/B region and motif 11 and 13 belonged to NLS. Similarly, motifs 12 contained the AHA motif were detected in the C-terminus of some members in subclass A. Furthermore, some unknown motifs were identified in HpHSFs. (Figure 5).

Cis-acting elements analysis in the promoter regions of HpHSF genes

We searched the potential cis-acting elements in the 1.5 kb upstream sequences of the translation initiation codons of HpHSFs in the PlantCARE database, and the result showed the presence of various cis-elements in the 5‘ flanking regions associated with stress, hormone, and development(Ning, et al. 2017). In stress-related cis-acting elements, some elements related to various stresses, such as light, low/high temperature, drought, anaerobic induction and wound were found in a large number of HpHSF genes, which including heat-shock response element (HSE), TC-rich repeats, Myb-binding DNA sequence (MBS), anaerobic induction element (ARE), low temperature range (LTR) and so on (Figure 6, Supplementary Table 4). In addition, there are plenty of hormone-related cis-acting elements in the promoters, ABA-responsive element (ABRE), MeJA responsive elements (TGACG-motif/CGTCA-motif), ethylene-responsive element (ERE), auxin-responsive element (TGA-element) and salicylic acid responsive element (TCA-element) were detected in the promoters of 19, 17, 13, 13and 7 HpHSFs, respectively. The results of the cis-elements suggested that the HpHSF genes might be involved in multiple transcriptional regulation of plant growth and stress responses.

Expression profiles of HpHSFs across different tissues

To explore the transcription patterns of HpHSF genes, a heat map of the transcription patterns of the HpHSF family was generated for the H. perforatum genes against RNA-seq data of four tissues including root, stem, leaf, and flower. According to the FPKM values, the expression profiles of HpHSF gene was remarkably different in four samples. For class A members, *HpHSF12*, *HpHSF18* and *HpHSF13* were expressed at high levels, while *HpHSF02* and *HpHSF23* were expressed at relatively low levels or undetected. Moreover, the expression of *HpHSF11*, *HpHSF18*, *HpHSF13* and *HpHSF07* in leaf were higher than that in other tissues. And among class B families, *HpHSF15* were expressed significantly at high abundances in all tissues compared with other genes. The members of class B family were all expressed at higher levels in root than in other tissues except *HpHSF01*, as well as the two members of class C, implying their critical roles in roots.

Expression analysis of HpHSF genes under heat stress treatment

HSF genes were found to play an important role in thermo tolerance of plants. In our study, the expression patterns of the HpHSF gene family were determined using quantitative RT-PCR to comprehend how HSF genes respond to heat stress under 42℃ treatment. As shown in Figure 8, the expression of *HpHSF2, 11, 12, 21* had no significantly change. *HpHSF03*, *18* and *22* were repressed after heat stress treatment, the remaining HpHSFs were up-regulated in varying degree. Noticeably, the expression of *HpHSF10* increased dramatically, it has been raised more than 500 times at 3 h as compared with the control, indicating that *HpHSF10* was a very sensitive response acceptor. In addition, the expression of *HpHSF1*, *14*, *20* and *23* were also changed obviously which were able to be concerned further.

Discussion

HSF gene family play an important role in plant adaptations to various biotic or abiotic stress. In this study, the identification and characteristics of 23 HSF genes were investigated based on Hypericum perforatum genome database and the expression profiles of the 23 genes were analyzed to explore their functions in heat stress response in H. perforatum. Overall, the isolation and identification of these HSF genes are helpful for illustrating the molecular genetic basis of H. perforatum., and the expression patterns of HpHSFs in four tissues and response to heat stress obtained during 42◦C suggested that HSF gene family was ubiquitously expressed and several HpHSF genes could play important roles in adaptation to environmental stress.

The DBD domain consists of about 100 amino acid residues which is highly conserved in yeast, plants and mammals (Schultheiss, et al. 1996). Similar to the results of previous studies, our finding showed that many sequences are conserved based on phylogenetic relationships of Arabidopsis, pepper, grape and H. perforatum and coiled-coil structure of HR-A/B regions prediction. The HpHSF genes were classified into three classes (A, B, C), Classes A and B were further divided into 9 (A1-A9) and 5 (B1-B5) subclasses respectively. The number of class A HSF genes were varying in plants, such as 15 in Arabidopsis and maize, 13 in rice and Mungbean, 16 in Soybean. Similarly, there are 7 class B HSFs in H. perforatum. The number of class B HSFs identified in plants are 10 in Mungbean, 8 in rice,7 in maize, and 5 in Arabidopsis. Most of the subclasses are shared among many species but not identical. In our study, the subclasses A2, A7, and A9 had been discovered in some species such as Arabidopsis and Arachis (Wang, et al. 2017), but not found in H. perforatum. It was hypothesized that elimination of introns, exon shuffung, and generations of exons might cause altered grouping in the phylogeny(Nover, et al. 2001). Overall, these observations suggested the functional conservation and divergence of HSF genes among different plants.

HSF protein is involved in abiotic stress respones and hormone signaling in plants(Huang, et al. 2015; Zhang, et al. 2015). The cis-acting elements in promoter region can regulate the transcription activity of corresponding genes, the research of detection of cis-acting elements could help understand the function and expression profiles of genes(Fragkostefanakis, et al. 2015; Wang, et al. 2017). The promoter region of the HpHSF gene family members contains varied elements related to growth and development, hormone response, and stress response. The numbers and types of elements are variable among the HpHSF promoters, and the overlapping phenomena were existed in different genes, which imply that the members of the family may regulate a variety of abiotic stresses and plant hormone signaling pathways simultaneously, which reflected the diversity and complexity of biological functions of the HpHSF gene family.

Gene expression profiles in different tissues are usually closely correlate with their functions in organ development(Guo, et al. 2008). In this study, the expression patterns of HpHSF genes in four different tissues were investigated. Remarkably, *HSF15* was found to be expressed at the highest level in four tissues by comparison to other genes. Each gene is expressed differently in four tissues, such as *HSF10* has the highest expression in root but the lowest expression in flowers and the expression level of *HSF18* in leaf was higher than other tissues, indicating their potential function in root and leaf, respectively. All these HpHSF genes play roles in different tissues to ensure the normal development of plants. Despite low expression in certain organizations of some HSFs, it does not mean that they have no function in these organizations. Tissue-specific expression patterns of identified HpHSF genes indicated that HpHSFs are widely involved in the growth and development of various tissues, which play an important role in studying the functions of HpHSF genes in H. perforatum developmental biology.

Plant HSFs play a central role in eliciting the expression of genes encoding heat shock proteins (Hsps) or other stress-inducible genes(Scharf, et al. ; Nishizawa-Yokoi, et al. 2009), which are important for plants to be protected from heat or other stress conditions. According to previous reports, the genome-wide expression profile suggested that several HSF genes are transcribed at relatively high levels during heat stresses(Giorno, et al. 2012; Chung, et al. 2013). In this study, 23 HpHSF genes showed distinct expression patterns during heat treatment. Among these genes, 14 HpHSF genes were up-regulated (>2-fold) and 3 (*HpHSF3*, *11*, *18*) were down-regulated during the heat stress treatment. Specifically, *HpHSF10* was the most strongly induced (~300-fold) in response to heat stress; *HSF20* was more than 90 times of the control after heat treatment; *HSF14*, *HSF15* and *HSF23* were about 20 times higher than those in the control group, indicating that they were very sensitive response acceptor that responded strongly, they play an important role in regulating the response of H. perforatum to heat stress and deserved our further attention and exploration.

Conclusions

In conclusion, a comprehensive analysis of HpHSF gene family about the genomic structures, conserved motifs, phyletic evolution, cis-acting elements and expression patterns were performed in this work.Overall, these findings are helpful in providing basis for understanding HSF protein function response to stress stimuli.

Acknowledgements

I would like to express my gratitude to all the people who helped me with writing this article, especially my supervisor, Prof. Wang, the completion of this thesis is inseparable from his valuable suggestions and patient guidance. I am also grateful to Lin Li, Donghao Wang, Qian Zhang and Xiaoding Yu for their help in the design of the study, acquisition of data and writing of manuscripts. Finally, I am indebted to my beloved family for their consistent support and encouragement.

References

Genome-wide analysis of heat shock transcription factor families in rice and Arabidopsis %J Journal of Genetics & Genomics. 35:105-118.

Abdelrahman M, Jogaiah S, Burritt DJ, Tran L-SP. 2018. Legume genetic resources and transcriptome dynamics under abiotic stress conditions. 41:1972-1983.

Baniwal SK, Bharti K, Chan KY, Fauth M, Ganguli A, Kotak S, Mishra SK, Nover L, Port M, Scharf K-DJJob. 2004. Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. 29:471-487.

Chidambaranathan P, Jagannadham PTK, Satheesh V, Kohli D, Basavarajappa SH, Chellapilla B, Kumar J, Jain PK, Srinivasan RJJopr. 2018. Genome-wide analysis identifies chickpea (Cicer arietinum) heat stress transcription factors (Hsfs) responsive to heat stress at the pod development stage. 131:525-542.

Chung E, Kim K-M, Lee J-HJJog, genomics. 2013. Genome-wide analysis and molecular characterization of heat shock transcription factor family in Glycine max. 40:127-135.

Drees BL, Grotkopp EK, Nelson HCM. 1997. GCN4 leucine zipper can functionally substitute for the heat shock transcription factors' trimerization domain. 273:61-74.

Fragkostefanakis S, Roeth S, Schleiff E, SCHARF KDJP, cell, environment. 2015. Prospects of engineering thermotolerance in crops through modulation of heat stress transcription factor and heat shock protein networks. 38:1881-1895.

Fujimoto M, Nakai AJFJ. 2010. The heat shock factor family and adaptation to proteotoxic stress. 277.

Galeotti NJJoe. 2017. Hypericum perforatum (St John's wort) beyond depression: A therapeutic perspective for pain conditions. 200:136-146.

Giorno F, Guerriero G, Baric S, Mariani CJBg. 2012. Heat shock transcriptional factors in Malus domestica: identification, classification and expression analysis. 13:639.

Guo J, Wu J, Ji Q, Wang C, Luo L, Yuan Y, Wang Y, Wang JJJoG, Genomics. 2008. Genome-wide analysis of heat shock transcription factor families in rice and Arabidopsis. 35:105-118.

Guo M, Liu J-H, Ma X, Luo D-X, Gong Z-H, Lu M-HJFips. 2016. The plant heat stress transcription factors (HSFs): structure, regulation, and function in response to abiotic stresses. 7:114.

Guo M, Lu J-P, Zhai Y-F, Chai W-G, Gong Z-H, Lu M-HJBpb. 2015. Genome-wide analysis, expression profile of heat shock factor gene family (CaHsfs) and characterisation of CaHsfA2 in pepper (Capsicum annuum L.). 15:151.

Hu Y, Han Y, Wei W, Li Y, Zhang K, Gao Y, Zhao F, Feng JJFips. 2015. Identification, isolation, and expression analysis of heat shock transcription factors in the diploid woodland strawberry Fragaria vesca. 6:736.

Huang Y, Li M-Y, Wang F, Xu Z-S, Huang W, Wang G-L, Ma J, Xiong A-SJMbr. 2015. Heat shock factors in carrot: genome-wide identification, classification, and expression profiles response to abiotic stress. 42:893-905.

Kotak S, Port M, Ganguli A, Bicker F, Von Koskull‐Döring PJTPJ. 2004. Characterization of C‐terminal domains of Arabidopsis heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class A Hsfs with AHA and NES motifs essential for activator function and intracellular localization. 39:98-112.

Li P-S, Yu T-F, He G-H, Chen M, Zhou Y-B, Chai S-C, Xu Z-S, Ma Y-ZJBg. 2014. Genome-wide analysis of the Hsf family in soybean and functional identification of GmHsf-34 involvement in drought and heat stresses. 15:1009.

Li S, Wang R, Ding Y, Jin H, Cai CJFig. 2018. Molecular characterization and expression profile analysis of heat shock transcription factors in mungbean. 9:736.

Lin YX, Jiang HY, Chu ZX, Tang XL, Zhu SW, Cheng BJJBG. 2011. Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. 12:76.

Lyck R, Harmening U, Höhfeld I, Treuter E, Scharf K-D, Nover LJP. 1997. Intracellular distribution and identification of the nuclear localization signals of two plant heat-stress transcription factors. 202:117-125.

Miller G, Mittler RJAob. 2006. Could heat shock transcription factors function as hydrogen peroxide sensors in plants? 98:279-288.

Mittler R, Finka A, Sciences PGJTiB. How do plants feel the heat? 37:118-125.

Nahrstedt A, Butterweck VJJoNP. 2010. Lessons learned from herbal medicinal products: the example of St. John’s wort. 73:1015-1021.

Ning P, Liu C, Kang J, Lv JJP. 2017. Genome-wide analysis of WRKY transcription factors in wheat (Triticum aestivum L.) and differential expression under water deficit condition. 5:e3232.

Nishizawa-Yokoi A, Yoshida E, Yabuta Y, Shigeoka SJB, biotechnology,, biochemistry. 2009. Analysis of the regulation of target genes by an Arabidopsis heat shock transcription factor, HsfA2. 73:890-895.

Nover L, Bharti K, Döring P, Mishra SK, Ganguli A, Scharf K-DJCs, chaperones. 2001. Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need? 6:177.

Peteranderl R, Rabenstein M, Shin Y-K, Liu CW, Wemmer DE, King DS, Nelson HCJB. 1999. Biochemical and biophysical characterization of the trimerization domain from the heat shock transcription factor. 38:3559-3569.

Prändl R, Hinderhofer K, Eggers-Schumacher G, Schöffl FJM, MGG GG. 1998. HSF3, a new heat shock factor from Arabidopsis thaliana, derepresses the heat shock response and confers thermotolerance when overexpressed in transgenic plants. 258:269-278.

Scharf KD, Berberich T, Ebersberger I, Nover L. The plant heat stress transcription factor (Hsf) family: Structure, function and evolution. 1819:0-119.

Schultheiss J, Kunert O, Gase U, Scharf KD, Nover L, Rüterjans HJEjob. 1996. Solution structure of the DNA‐binding domain of the tomato heat‐stress transcription factor HSF24. 236:911-921.

Treuter E, Nover L, Ohme K, Genet K-DSJMG. Promoter specificity and deletion analysis of three heat stress transcription factors of tomato. 240:113-125.

Wang P, Song H, Li C, Li P, Li A, Guan H, Hou L, Wang XJFips. 2017. Genome-wide dissection of the heat shock transcription factor family genes in Arachis. 8:106.

Yang X, Zhu W, Zhang H, Liu N, Tian SJP. 2016. Heat shock factors in tomatoes: genome-wide identification, phylogenetic analysis and expression profiling under development and heat stress. 4:e1961.

Zandalinas SI, Mittler R, Balfagón D, Arbona V, Gómez-Cadenas A. 2018. Plant adaptations to the combination of drought and high temperatures. 162:2-12.

Zhang J, Jia H, Li J, Li Y, Lu M, Hu JJSr. 2016. Molecular evolution and expression divergence of the Populus euphratica Hsf genes provide insight into the stress acclimation of desert poplar. 6:30050.

Zhang J, Liu B, Li J, Zhang L, Wang Y, Zheng H, Lu M, Chen JJBg. 2015. Hsf and Hsp gene families in Populus: genome-wide identification, organization and correlated expression during development and in stress responses. 16:181.

Table 1 (on next page)

The HSF genes identifed from the *H. perforatum*.

Notes: pI, isoelectric point; n.c.r., total number of negatively charged residues (Asp +Glu); p.c.r., total number of positively charged residues (Arg +Lys); I.I., instability index; A.I., aliphatic index; GRAVY, grand average of hydropathicity.

1

Gene Name	Transcript ID	Length (aa)	MW (kDa)	pI	n.c.r. (%)	p.c.r. (%)	I.I.	Stability	A.I.	GRAVY
HpHSF01	HperS113g0097	293	32.23	5.05	43 (14.7%)	35 (11.9%)	57.40	unstable	75.26	−0.523
HpHSF02	HperS020g0043	381	43.75	5.51	59 (15.5%)	48 (12.6%)	59.80	unstable	71.55	−0.752
HpHSF03	HperS219g0006	327	37.9	7.29	36 (11.0%)	36 (11.0%)	47.69	unstable	72.14	−0.660
HpHSF04	HperS024g0021	222	25.95	7.72	34 (15.3%)	35 (15.8%)	52.96	unstable	73.24	−0.796
HpHSF05	HperS024g0048	196	22.47	6.85	31 (15.8%)	31 (15.8%)	46.57	unstable	69.08	−0.747
HpHSF06	HperS245g0169	226	25.88	6.86	34 (15.0%)	34 (15.0%)	48.26	unstable	69.38	−0.737
HpHSF07	HperS025g0041	434	48.47	5.22	64 (14.7%)	47 (10.8%)	58.65	unstable	76.18	−0.577
HpHSF08	HperS254g0338	376	42.04	5.67	45 (12.0%)	39 (10.4%)	66.18	unstable	64.84	−0.655
HpHSF09	HperS338g0001	330	36.57	5.67	43 (13.0%)	38 (11.5%)	50.96	unstable	60.24	−0.600
HpHSF10	HperS346g0011	428	48.01	4.91	66 (15.4%)	44 (10.3%)	56.52	unstable	67.64	−0.642
HpHSF11	HperS346g0247	324	37.51	5.91	45 (13.9%)	38 (11.7%)	57.46	unstable	59.85	−0.813
HpHSF12	HperS362g0014	409	46.16	5.02	65 (15.9%)	43 (10.5%)	57.83	unstable	65.99	−0.745
HpHSF13	HperS388g0082	403	46.49	4.79	71 (17.6%)	46 (11.4%)	47.52	unstable	65.56	−0.764
HpHSF14	HperS398g0019	195	22.35	8.86	26 (13.3%)	30 (15.4%)	63.69	unstable	58.10	−0.818
HpHSF15	HperS042g0257	248	27.78	5.78	40 (16.1%)	37 (14.9%)	46.02	unstable	61.33	−0.817
HpHSF16	HperS434g0151	501	54.91	4.87	64 (12.8%)	42 (8.4%)	59.15	unstable	67.60	−0.608
HpHSF17	HperS044g0424	483	53.92	4.99	63 (13.0%)	42 (8.7%)	51.71	unstable	74.66	−0.533
HpHSF18	HperS443g0073	397	44.95	4.94	63 (15.9%)	41 (10.3%)	62.46	unstable	70.48	−0.723
HpHSF19	HperS006g0172	188	21.72	8.54	25 (13.3%)	28 (14.9%)	52.32	unstable	54.52	−0.797
HpHSF20	HperS064g0032	455	51.78	5.91	64 (14.1%)	57 (12.5%)	62.11	unstable	71.12	−0.644
HpHSF21	HperS068g0017	495	54.66	4.96	65 (13.1%)	45 (9.1%)	56.75	unstable	67.39	−0.650
HpHSF22	HperS079g0626	270	31.63	6.33	33 (12.2%)	28 (10.4%)	47.61	unstable	64.59	−0.826
HpHSF23	HperS091g0277	363	41.62	5.39	57 (15.7%)	43 (11.8%)	61.47	unstable	69.53	−0.784

2

Figure 1

Multiple sequence alignment of the DBD domains of 23 members of the HSF protein family.

Three α- helices and four β- sheets were presented in the region.

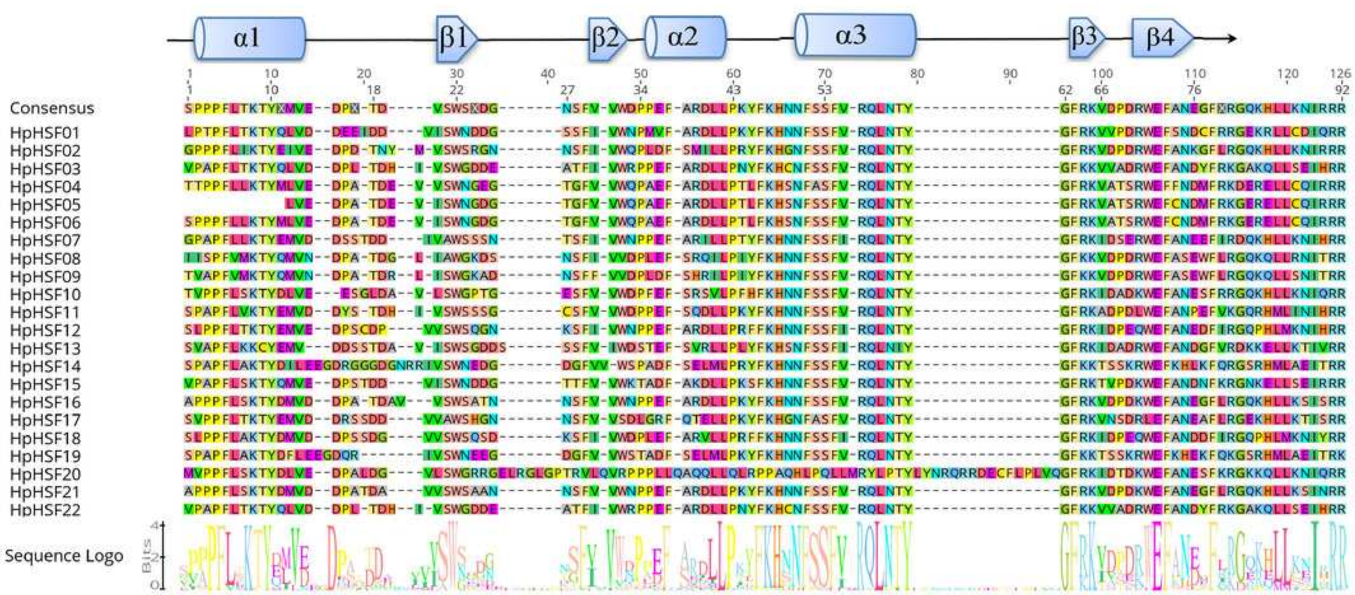


Figure 2

Multiple sequence alignment of the HR-A/B regions of 23 members of the HSF protein family.

The annotations at the top describe the location and boundaries of the HR-A core, insert, and HR-B region within the HR-A/B region.

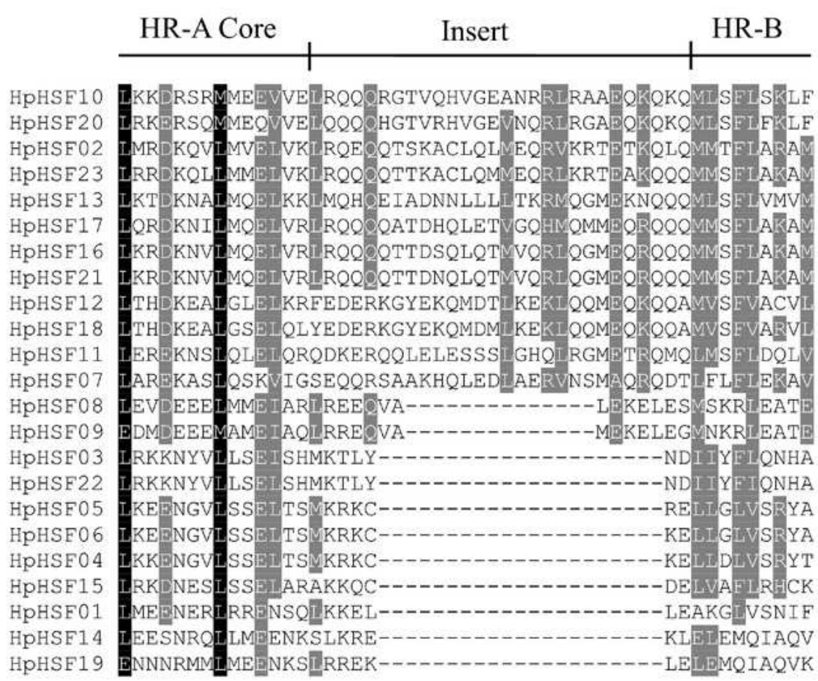


Figure 3

Neighbor-Joining phylogenetic tree of HSF proteins from *H perforatum (Hp), Capsicum annuum L. (Ca) , Vitis vinifera L. (V)* and *A. thaliana (At)*.

The full-length of amino acid sequences of HSF proteins in the four species were used to construct of the phylogenetic tree with MEGA 6 and subclass numbers of *Arabidopsis*, pepper and grape are listed.

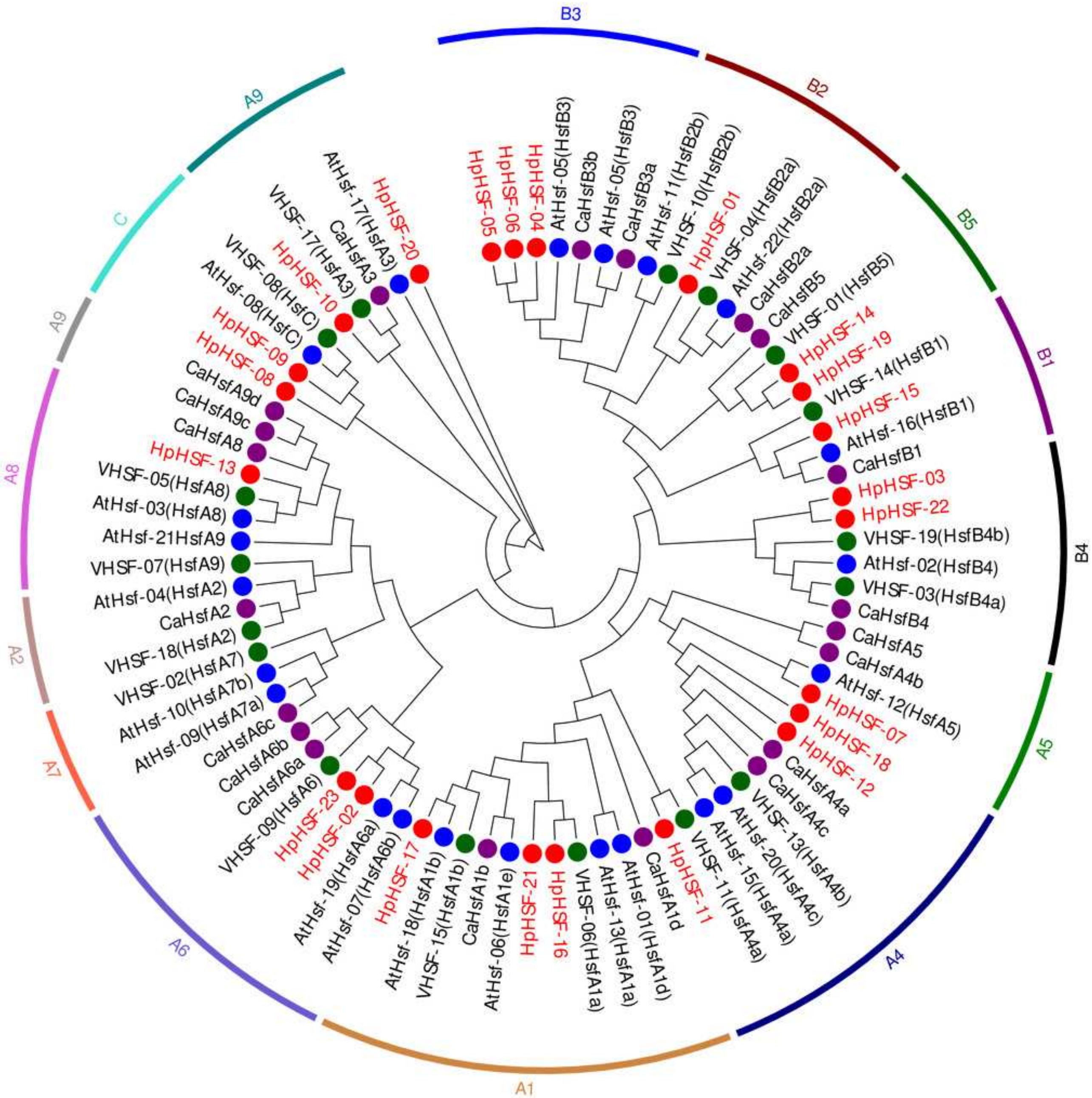

B3
B2
B5
B1
B4
A5
A4
A1
A6
A7
A2
A8
A9
C
A9
HpHSF-05
HpHSF-06
HpHSF-04
AtHsf-05(HsfB3)
CaHsfB3b
AtHsf-05(HsfB3)
CaHsfB3a
AtHsf-11(HsfB2b)
VHSF-10(HsfB2b)
HpHSF-01
VHSF-04(HsfB2a)
AtHsf-22(HsfB2a)
CaHsfB2a
CaHsfB5
VHSF-01(HsfB5)
HpHSF-14
HpHSF-19
VHSF-14(HsfB1)
HpHSF-15
AtHsf-16(HsfB1)
CaHsfB1
HpHSF-03
HpHSF-22
VHSF-19(HsfB4b)
AtHsf-02(HsfB4)
VHSF-03(HsfB4a)
CaHsfB4
CaHsfA5
CaHsfA4b
AtHsf-12(HsfA5)
HpHSF-07
HpHSF-18
HpHSF-12
CaHsfA4a
CaHsfA4c
VHSF-13(HsfA4b)
AtHsf-20(HsfA4c)
AtHsf-15(HsfA4a)
VHSF-11(HsfA4a)
HpHSF-11
CaHsfA1d
AtHsf-01(HsfA1d)
AtHsf-13(HsfA1a)
VHSF-06(HsfA1a)
HpHSF-16
HpHSF-21
AtHsf-06(HsfA1e)
CaHsfA1b
VHSF-15(HsfA1b)
AtHsf-18(HsfA1b)
HpHSF-17
AtHsf-07(HsfA6b)
AtHsf-19(HsfA6a)
HpHSF-02
HpHSF-23
VHSF-09(HsfA6)
CaHsfA6a
CaHsfA6b
CaHsfA6c
AtHsf-09(HsfA7a)
AtHsf-10(HsfA7b)
VHSF-02(HsfA7)
VHSF-18(HsfA2)
CaHsfA2
AtHsf-04(HsfA2)
VHSF-07(HsfA9)
AtHsf-21HsfA9
AtHsf-03(HsfA8)
VHSF-05(HsfA8)
HpHSF-13
CaHsfA8
CaHsfA9c
CaHsfA9d
HpHSF-08
HpHSF-09
AtHsf-08(HsfC)
VHSF-08(HsfC)
HpHSF-10
VHSF-17(HsfA3)
CaHsfA3
AtHsf-17(HsfA3)
HpHSF-20

Figure 4

Gene structure (a) and conserved motifs (b) of HpHSF family members.

(a) blank box, Grey box and black line were represented CDS, upstream/ downstream and intron, respectively. The number 0, 1, and 2 on the black line were intron phase. (b) 15 conserved motifs were identifed by MEME. The motifs which are numbered 1–15 are exhibited in different colored boxes.

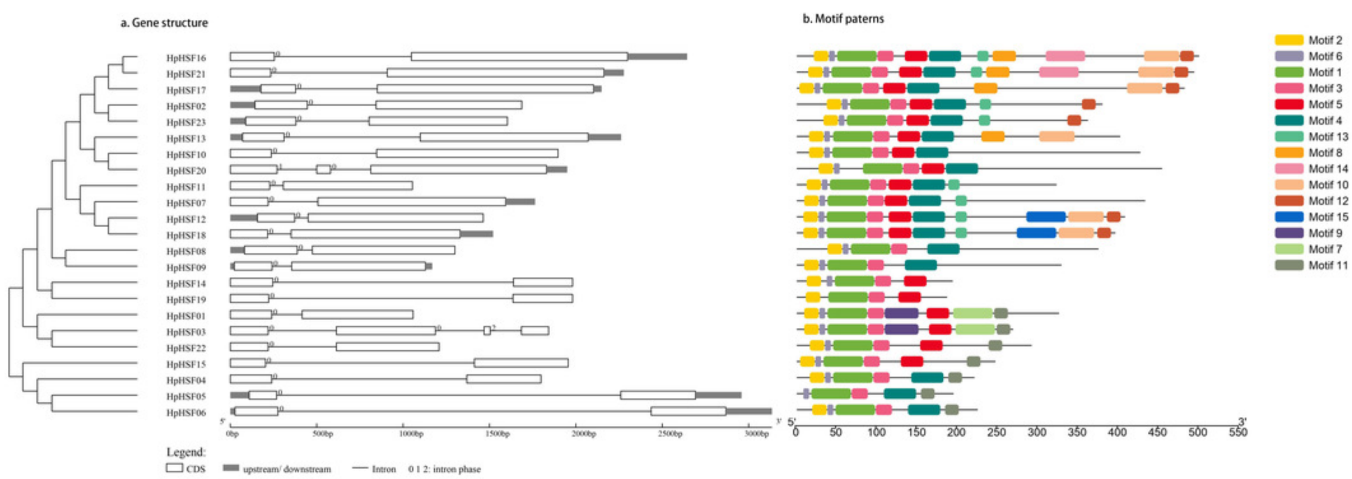
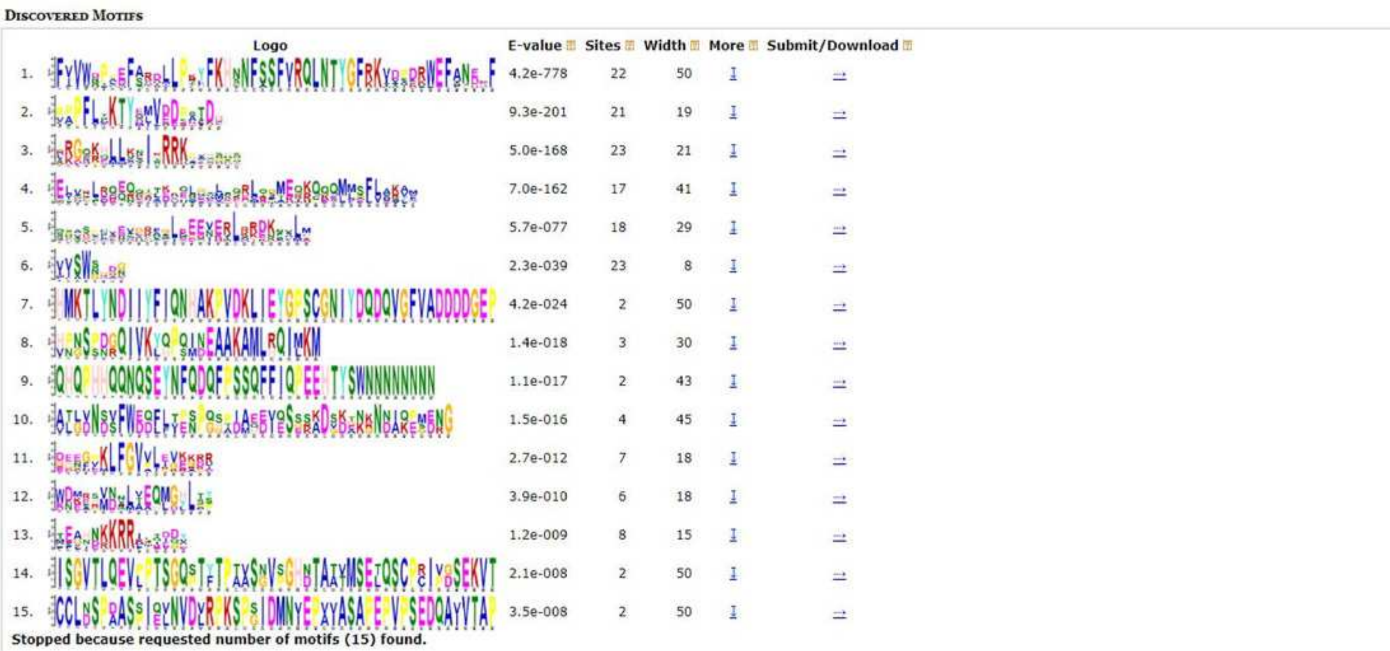


Figure 5

Sequence logos of 15 motifs in HpHSF proteins.

The “sites” indicate the number of HpHSF proteins containing each motif. The “width” indicates the amino acid number of each motif.



DISCOVERED MOTIFS

	Logo	E-value	Sites	Width	More	Submit/Download
1.		4.2e-778	22	50		
2.		9.3e-201	21	19		
3.		5.0e-168	23	21		
4.		7.0e-162	17	41		
5.		5.7e-077	18	29		
6.		2.3e-039	23	8		
7.		4.2e-024	2	50		
8.		1.4e-018	3	30		
9.		1.1e-017	2	43		
10.		1.5e-016	4	45		
11.		2.7e-012	7	18		
12.		3.9e-010	6	18		
13.		1.2e-009	8	15		
14.		2.1e-008	2	50		
15.		3.5e-008	2	50		

Stopped because requested number of motifs (15) found.

Figure 6

Number of HpHSF genes containing various cis-acting elements.

The graph was generated based on the presence of *cis*-acting elements responsive to specific processes/elicitors/conditions (x-axis) in HSF gene family members (y-axis).

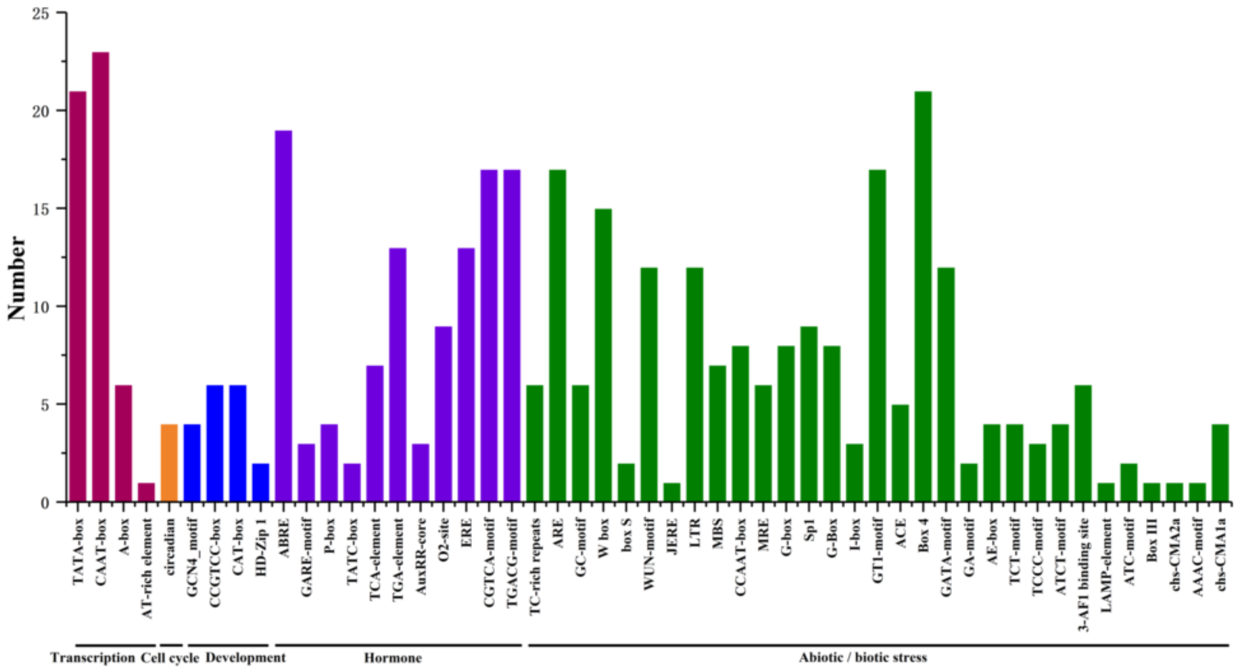


Figure 7

Heat map representation and hierarchical clustering of HpHSF genes in flower, leaf, root, stem.

The expression values were calculated by fragments per kilobase of exon model per million mapped (FPKM).

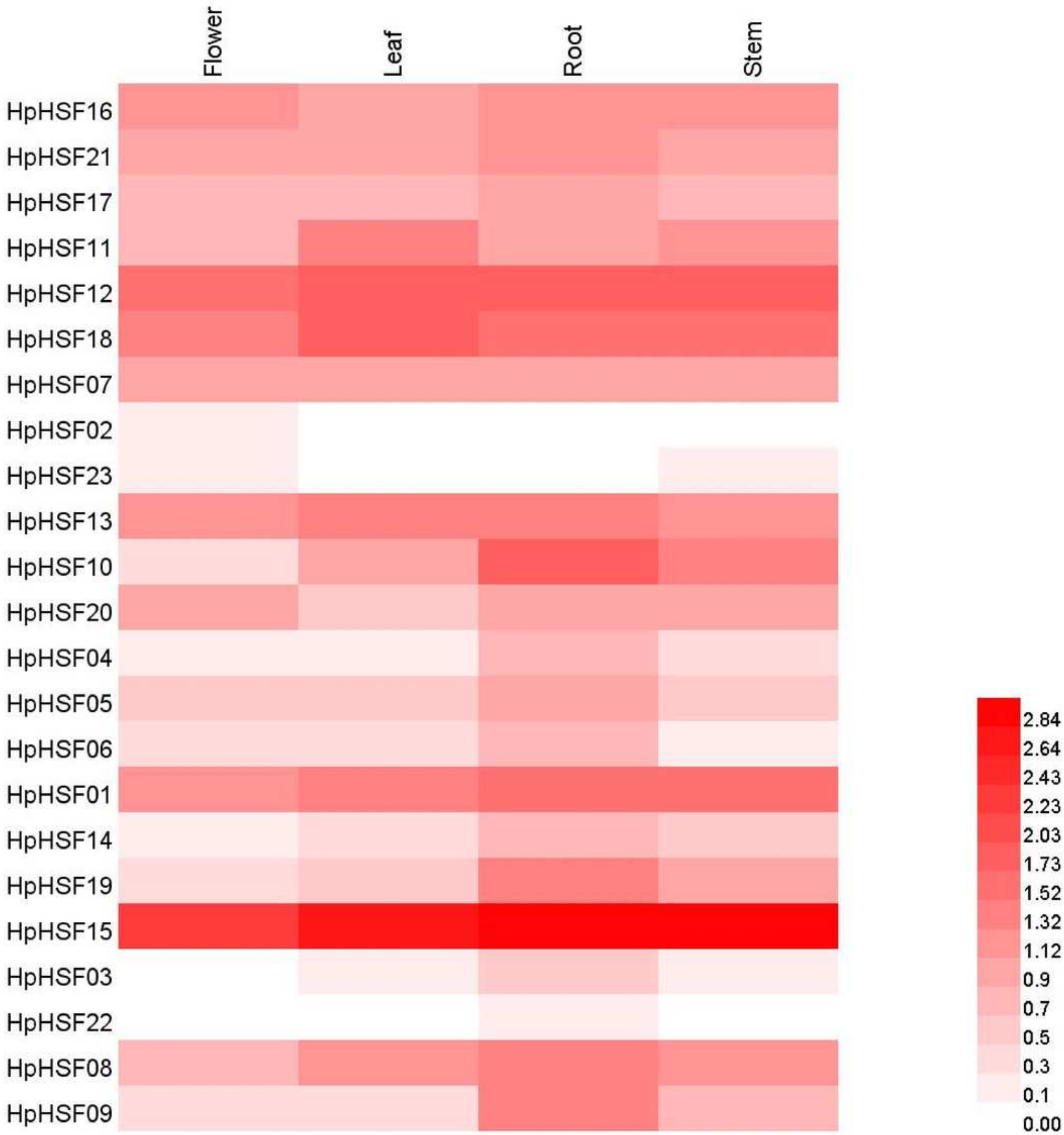
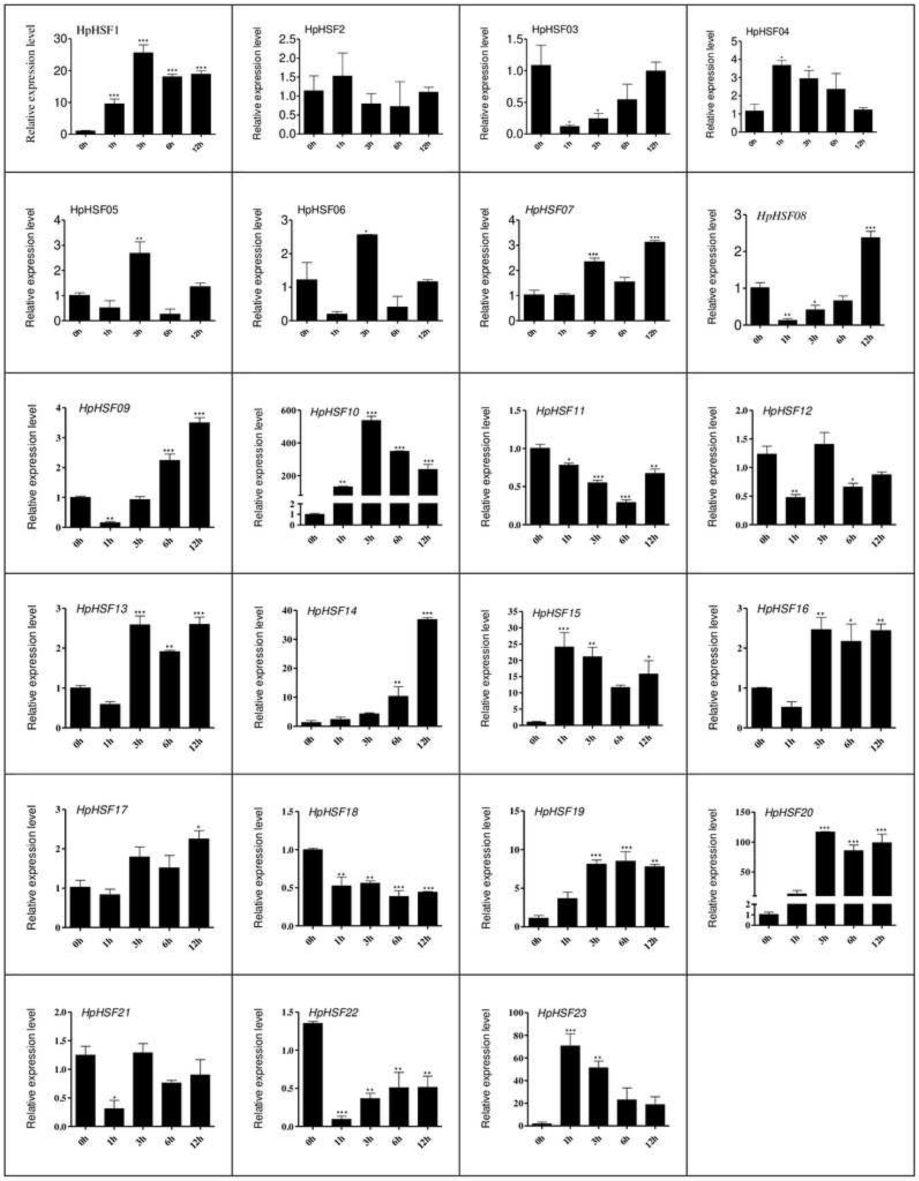

Flower
Leaf
Root
Stem
HpHSF16
HpHSF21
HpHSF17
HpHSF11
HpHSF12
HpHSF18
HpHSF07
HpHSF02
HpHSF23
HpHSF13
HpHSF10
HpHSF20
HpHSF04
HpHSF05
HpHSF06
HpHSF01
HpHSF14
HpHSF19
HpHSF15
HpHSF03
HpHSF22
HpHSF08
HpHSF09
2.84
2.64
2.43
2.23
2.03
1.73
1.52
1.32
1.12
0.9
0.7
0.5
0.3
0.1
0.00

Figure 8

Relative gene expression of HpHSFs analyzed by qRT-PCR responsed to heat stress treatment.

qRT-PCR data was normalized using *Hypericum perforatum Actin 2* gene and are shown relative to 0 h. X-axes are time course (0 h, 1 h,3 h, 6 h and 12 h) and y-axes are scales of relative expression level (error bars indicate SD).


HpHSF1
HpHSF2
HpHSF03
HpHSF04
HpHSF05
HpHSF06
HpHSF07
HpHSF08
HpHSF09
HpHSF10
HpHSF11
HpHSF12
HpHSF13
HpHSF14
HpHSF15
HpHSF16
HpHSF17
HpHSF18
HpHSF19
HpHSF20
HpHSF21
HpHSF22
HpHSF23
Relative expression level
0h
1h
3h
6h
12h