Genome-Wide Identification, Classification, and Expression Analysis of the *HSF* Gene Family in pineapple (*Ananas comosus*)

Lulu Wang[1], Yanhui Liu[1], Mengnan Chai[1], Huihuang Chen[1], Xiaoping Niu[2], Yuan Qin[1,2*] and Hanyang Cai[1*].

[1]State Key Lab of Ecological Pest Control for Fujian and Taiwan Crops; Key Lab of Genetics, Breeding and Multiple Utilization of Crops, Ministry of Education; Fujian Provincial Key Lab of Haixia Applied Plant Systems Biology, College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou, Fujian Province, China

[2]State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi Key Lab of Sugarcane Biology, College of Agriculture, Guangxi University, Nanning, Guangxi Province, China

[*] Correspondence: Yuan Qin (yuanqin@fafu.edu.cn), Hanyang Cai (caihanyang123@163.com)

ABSTRACT: Heat shock transcription factors (HSFs) play crucial roles in plant growth, development, and response to environmental cues. However, no HSFs have been characterized in pineapple thus far. Here, a total of 22 *AcHSF* genes were identified from the pineapple genome. Gene structure, motifs, and phylogenetic analysis showed that *AcHSF* families were distinctly grouped into three subfamilies (12 in Group A, 7 in Group B, and 4 in Group C). Promoter analysis showed that the *AcHSF* promoters contained various cis-elements related to stress, hormones, and development processes, such as STRE, MYB, and ABRE binding sites. The majority of *HSFs* were expressed in different pineapple tissues and developmental stages. The expression of *AcHSF-B4b*/*AcHSF-B4c* and *AcHSF-A7b/AcHSF-A1c* were enriched in the ovules and fruits, respectively. Six genes (*AcHSF-A1a*, *AcHSF-A2*, *AcHSF-A9a*, *AcHSF-B1a*, *AcHSF-B2a*, and *AcHSF-C1a*) were transcriptionally modified by cold, heat, and ABA. Our results provide an overview and lay the foundation for future functional characterization of the *HSF* gene family in pineapple.

Keywords: *AcHSF*, phylogenetic analysis, Pineapple, cold, heat, ABA

INTRODUCTION

Plant growth and production are affected by multiple abiotic stresses such as cold, heat, drought, and salinity (Hu & Xiong 2014; Pereira 2016; Zhu 2016). Heat stress, defined as a rise in the temperature of 10-15°C above the ambient (Wahid et al. 2007), beyond a given threshold level for ~~a period of~~some time, is an agricultural problem in many areas all over the world, affecting plant growth and development and often leading to reductions in yield. Consequently, in response to environmental stresses, plants have evolved a series of defense or signaling mechanisms. Usually, ~~these~~ this stress response process involves different types of transcription factors (TFs). These include heat shock transcription factors (HSFs), as well as WRKY, MYB, AP2/ERF, and NAC, which regulate the expression of thousands of genes under various stress conditions. In plants, the HSF family is one of the most important TF families involved in the heat stress response and regulates the expression of heat shock proteins (HSPs) as well as other stress-responsive proteins, such as ascorbate peroxidase (APX) and catalase (CAT) (Ohama et al. 2017). Besides their roles in stress responses, *HSFs* are also reported to play important roles in plant growth and development. For example, the overexpression of *CarHsfA2* could enhance chickpea stress tolerance without any pleiotropic effects (Chidambaranathan et al. 2018). *AtHSF-B1* and *AtHsf-B2b* act as repressors of the expression of heat-inducible and the *AtHSF-A1* involve in cold acclimation in *Arabidopsis (Ikeda et al. 2011; Olate et al. 2018)*. *PeuHsfA2* was induced by heat stress, which may increase the acclimation of desert poplar (Zhang et al. 2016b).

As evolutionarily conserved transcription factors, *HSFs* have some conserved domains. A typical HSF protein contains a modular structure with an N-terminal DNA-binding domain (DBD) that is responsible for binding HSEs in the promoters of several HSPs; an adjacent oligomerization domain (OD) composed of heptad repeats of hydrophobic amino acid residues (HR-A/B) that are connected to the DBD by a flexible linker; a nuclear localization signal (NLS) region essential for nuclear uptake of the protein, a nuclear export signal (NES) region, and C-terminal activator motif,

also known as AHA motif (AHA)(Guo et al. 2016; Nishizawa-Yokoi et al. 2011; Singh et al. 2012; Yabuta 2016). According to the flexible linker of variable length (about 15–80 amino acids) and the oligomerization domain (HR-A/B), plant HSFs can be divided into at least three types, including class A (A1, A2, A3, A4, A5, A6, A7, A8, A9), class B (B1, B2, B3, B4) and class C (C1, C2) (Giesguth et al. 2015; Nishizawa-Yokoi et al. 2011; Shim et al. 2009; Singh et al. 2012; Yabuta 2016). Since the initially identified in yeast (Sorger & Pelham 1988) and the first plant HSF gene identified in tomato (Scharf et al. 1990), the plant *HSF* gene family has been identified and characterized in more and more plant species, including *Arabidopsis thaliana* (Guo et al. 2008), rice (*Oryza sativa* L.) (Chauhan et al. 2011; Jin et al. 2013), maize (*Zea mays* L.) (Lin et al. 2011), *Populus trichocarpa* (Wang et al. 2012), wheat (*Triticum aestivum* L.) (Chauhan et al. 2013), soybean (*Glycine max*) (Chung et al. 2013), Chinese cabbage (*Brassica rapa ssp. pekinensis*) (Song et al. 2014), cotton (*Gossypium hirsutum*)(Wang et al. 2014), barrel medic (*Medicago truncatula*)(Lin et al. 2014), pepper (*Capsicum annuum* L.) (Guo et al. 2015; Guo et al. 2014), strawberry (*Fragaria vesca*) (Hu et al. 2015), tea plant (*Camellia sinensis*) (Liu et al. 2016), etc.

Pineapple (*Ananas comosus*) is one of the most popular fresh fruits worldwide and is cultivated in ~~the~~ subtropical and tropical areas (Bai et al. 2019). However, a lot of factors restrict the production of pineapples, such as the extreme environmental conditions (high temperature, cold temperature, drought and so on), the pathogen infection, and degradation of good breeds (Barral et al. 2019). Thus, it is very meaningful to identify candidate genes involved in pineapple response to environmental stresses and pathogen infection, as well as the molecular mechanism and possible utilization for genetic breeding. *HSFs* are widely known for their common involvement in various abiotic stresses including heat stress and plant-pathogen interaction. However, the *AcHSFs* have not been identified in pineapple, as well as their possible roles. In this study, genome-wide identification and expression analysis during flower and fruit development, as well as abiotic stresses, were performed to extend our understanding and possible utilization of *AcHSFs* in genetic breeding.

METHODS

Identification and Characterization of *AcHSF* Genes in pineapple

To obtained the protein and nucleotide sequences of pineapple HSFs, the HSF-type DBD domain (Pfam: PF00447) was submitted as a query in JGI *Ananas comosus v3* annotation. A total of 30 pineapple HSFs were obtained from JGI. Only 22 *AcHSFs* left, after manually filtering out repeated sequences and sequences without integrated HSF-type DBD domains or classic coiled-coil structures by SMART (http://smart.embl-heidelberg.de/) (Letunic et al. 2012). The information of Genomics position, Chromosome NO., CDS, and AA Length for *AcHSFs* were obtained from JGI *Phytozome v12.1* (https://phytozome.jgi.doe.gov/pz/portal.html). The biophysical properties of coding AcHSFs were calculated using the Expasy ProtParam tool (https://web.expasy.org/protparam/). The subcellular localization of *AcHSFs* was analyzed using BUSCA (http://busca.biocomp.unibo.it/).

Chromosome Localization phylogenetic relationships

The information of all pineapple HSFs' localization on chromosomes was obtained from *Phytozome v12.1*, including chromosome length, Chromosome NO., and gene start site. The MapChart 2.0 (https://mapchart.net/) software was adopted to visually map the chromosomal location. A phylogenetic tree was constructed using the adjacent method by MEGA5.0 with a 1000 bootstrap value. To further understand the phylogenetic relationship of HSF proteins, the phylogenetic tree was constructed using the AcHSF protein sequences and other three model species, i.e., *A. thaliana*, *O. sativa*, and *P. trichocarpa*. Distinctive names for each of the HSFs identified in pineapple were given according to the classification of HSFs in classes A, B, and C, referred to as *AcHSF* genes.

Genetic structure and *cis*-acting elements

The gene structures including exons and introns were displayed using Gene Structure Display Server (GSDS, http://gsds.cbi.pku.edu.cn/index.php, Beijing, China) (Guo et

al. 2007). The upstream sequences of the AcHSFs, which were 2 kb upstream from the translation start site, were retrieved from Phytozome. These sequences were analyzed for the identification of regulatory cis-elements important for gene expression under abiotic stress, development, and hormone signaling using the plant cis-acting element database New PLACE (https://www.dna.affrc.go.jp/PLACE/?action=newplace) (Higo et al. 1999).

Conserved domains and motifs analysis of *AcHSF*s

The DBD domain and HR-A/B regions (OD) aligned by Clustal X 2.0 software and edited by DNAMAN software. NLS domains were predicted using cNLS Mapper software (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi). NES domains in the *AcHSFs* were predicted by the NetNES 1.1 server software (http://www.cbs.dtudk/services/NetNES/). The conserved motifs of AcHSFs were defined by Multiple Em for Motif Elicitation (MEME, http://meme-suite.org/, U.S.A.) using the following parameters: number of repetitions = any-, the maximum number of motifs = 10, minimum width ⩾10, maximum width ⩽200, and only motifs with an E-value < 0.01 were retained for further analysis.

Expression patterns analysis

The transcriptomic data of pineapple in different organs and developmental stages have been described in the previous~~ly~~ study (Wang et al. 2020). Briefly, the different organs include 3 stages of petal tissues, 4 stages of sepal tissues, 6 stages of stamen tissues, 7 stages of ovule tissues, 7 stages of gynoecium tissues from *Wang et al* (Wang et al. 2020), and roots, flower, leaf, and 6 stages fruit tissue from *Ming et al* (Ming et al. 2015). The heatmap was then constructed using the pheatmap package of R software.

Stress treatments

One-month-old plants in rooting medium were used as the planting material for the stress treatment analyses. Uniform tissue-cultured seedlings were obtained from the Qin Lab (Priyadarshani et al. 2018). Seedlings were subjected to the following stress

treatments: low temperature (4°C), high temperature (45 °C), and ABA (100 mM). Leaves were collected from three independent lines at 12h, 24h, and 48 h after treatment. Seedlings that were not subjected to any of the stress treatments were used as controls. The collected samples were immediately stored in liquid nitrogen ~~prior to~~before total RNA extraction.

RNA extraction and qRT-PCR

Total RNA was extracted from pineapple leaf tissues following the manufacturer's protocol RNA extraction Kit (Omega Bio-Tek, Shanghai, China). cDNA was synthesized with the EasyScript® One-Step gDNA Removal and cDNA Synthesis SuperMix (Transgen, Beijing, China). ~~QPCR~~ qPCR was conducted using TransStart® Top Green qPCR SuperMix (Transgen, Beijing, China). *Actin2* was used as a reference gene (Wang et al. 2020). These assays were conducted for three biological replicates, and the results are shown as the mean ± standard deviations.

RESULTED

Genome-Wide Identification of *HSF* Genes in Pineapple

The amino acid sequences of HSF-type DBD domains (Pfam: PF00447) were submitted into *Ananas comosus* v3 Phytozome database v12.1 for BLASTP searches. A total of 30 putative pineapple HSF sequences were acquired. After checking by the *Pfam* database and *SMART* online tool, 1 pineapple HSF sequence was rejected due to the absence of typical HSF-DBD domains, and 7 were abandoned due to the absence of coiled-coil structures. Consequently, 22 non-redundant pineapple *HSFs* were identified (Table 1). The comprehensive information of these 22 *AcHSF* genes including gene name, gene ID, CDS and protein length, isoelectric points, molecular weights, predicted subcellular location, and other features are presented in Table 1. The amino acids length of *AcHSFs* ranged from 129 (*AcHSF*-A9b) to 601 (*AcHSF*-A5). The predicted isoelectric points (pI) varied from 4.68 (*AcHSF*-B2b) to 9.63 (*AcHSF*-B4c),

and the molecular weight (MW) varied from 13.77 kDa (*AcHSF*-A9b) to 65.81 kDa (*AcHSF*-A5). The detailed information about other parameters was provided in Table 1.

According to the detailed gene information, 19 *AcHSF* genes were mapped to the 11 pineapple chromosomes and 3 *AcHSF* genes located in the scaffold (Table 1). The number of pineapple *HSF* genes in each chromosome differed considerably, and there is no discernible pattern in the location of these genes on chromosomes. For example, three *AcHSF* genes were located in chromosome 5, whereas only one was present in chromosomes 2, 6, 17, and 18 respectively (Figure 1).

Phylogenetic Analysis of *AcHSFs* Gene Family

To determine the phylogenetic relationships among pineapple *HSFs*, a phylogenetic analysis of 31 *Populus trichocarpa* HSFs, 25 rice HSFs, and 21 *Arabidopsis* HSFs (Guo et al. 2016), together with those of *AcHSFs* were performed by generating a neighbor-joining phylogenetic tree. According to the difference in amino acid sequences of the DBD domain, the HR-A/B region, and the linker between them, the *HSFs* were grouped into three clusters A, B, and C (Guo et al. 2016; Scharf et al. 2012). Class A was divided into 10 sub-clusters, designated A1, A2, A3, A4, A5, A6, A7, A8, A9, and A10. Class B was divided into B1, B2, B3, and B4. And ~~the~~ class C contains sub-clusters C1 and C2. In pineapple (*Ananas comosus*), according to their phylogenetic relationship, 12 AcHSFs out of 22 proteins belong to class A, followed by 7 AcHSFs belonging to class B, and three copies of class C (Figure 2). As a monocot, the pineapple was more similar to rice, rather than the dicot *Arabidopsis* and *Populus trichocarpa*. However, none of the *AcHSFs* were found in the subclass A8 and B3, which was reported to only exist in the monocots (Li et al. 2014). It is strange that the pineapple and rice subclass A7 HSFs showed higher similarity to A2 rather than the *Arabidopsis* and *Populus trichocarpa* subclass A7, and the AtHSF-A6a also shows abnormal clustering (Figure 2).

Gene structures and *cis*-acting elements analysis of *AcHSFs*

To reveal the gene structural features of *AcHSFs*, intron/exon and upstream (5' UTR)/downstream (3' UTR) structures were analyzed using Gene Structure Display Server (GSDS) v2.0. The exon numbers of *AcHSFs* varied from 1 to 5 (Figure 3), while only in the longest *AcHSF-A9a* (genomic sequence 34598 bp.) was found 5 exons. The upstream and downstream sequence of the *AcHSF* genes are incomplete, 8 out of 22 *AcHSFs* (*AcHSF-A3*, *AcHSF-A7b*, *AcHSF-B1c*, *AcHSF-B2a*, *AcHSF-B2b*, *AcHSF-B4c*, *AcHSF-C1a*, and *AcHSF-C1b*) do not have upstream and downstream sequences, 3 *AcHSFs* (*AcHSF-A1b*, *AcHSF-A9a*, and *AcHSF-B4b*) have only upstream sequences, 6 *AcHSFs* (*AcHSF-A4*, *AcHSF-A5*, *AcHSF-A7*, *AcHSF-B1a*, *AcHSF-B1b*, and *AcHSF-C2*) have only downstream sequences, and 5 *AcHSFs* (*AcHSF-A1a*, *AcHSF-A1c*, *AcHSF-A2*, *AcHSF-A6*, and *AcHSF-A9*) both have upstream and downstream sequences.

It has been reported that the ABA-responsive element (ABRE), low-temperature responsive element (LTRE), dehydration-responsive element (DRE), MYB, MYC, and WRKY elements play different significant roles in stress responses in plants (Chai et al. 2020; Li et al. 2012; Zhang et al. 2009). The 2kb sequences upstream of *AcHSFs* gene were selected for analysis. The *cis*-acting elements analysis of *AcHSFs* promoter demonstrated that every pineapple *HSF* contains at least 2 MYB, MYC, and WRKY elements, except for *AcHSF-B4c* (Table 2). But for the *AcHSF-B4c*, only 110bp promoter sequence can be found in the upstream area, among the 110bp promoter sequence, the main core component of the promoter TATA-box and CAAT-box, the light regulatory element (RYREPEATBNNAPA), and ~~root~~ root-specific expression related elements (ROOTMOTIFTAPOX1) can be found. In addition, we also detected the ABRE, DRE, and LTRE in the *AcHSFs* promoter area. The result showed that the *AcHSF-A1a* and *AcHSF-A4* lacked ABRE, the *AcHSF-A6* and *AcHSF-C1b* lacked DRE, the *AcHSF-C1b* lacked LTRE, and the *AcHSF-A9a*, *AcHSF-B4b*, and *AcHSF-B4c* did not have these three stress response elements (Table 2). The analyses of *cis*-elements in the promoters suggest that *HSFs* are significantly related to the stress response.

Conserved domains and motifs of pineapple *HSFs*

The modular structure of the HSF family in plants has been described thoroughly in several model plants. The HSF protein contains 5 typical conserved domains: DBD, OD, NLS, NES, and AHA domains from N to C-terminal (Table 3). The most conserved DBD domain composed of approximately 100 amino acids, containing three α-helices and a four-stranded antiparallel β-sheet (α1-β1-β2-α2-α3-β3-β4) (Figure 4 A). In addition to the DBD domain, the HR-A/B next to the DBD domain is also important and plays a crucial role in HSF-HSF interaction (Scharf et al. 2012). Besides, HR-A/B also presents in all AcHSFs (Table 3, Figure 4 B). According to the previous studies, HSFs were artificially divided into A, B, and C classes by the distinction between the HR-A and HR-B motifs (Cheng et al. 2015; Giesguth et al. 2015; Singh et al. 2012). In general, the variable length of the flexible linker between parts A and B of the HR-A/B motif of classes A and C HSFs is approximately 15 to 80 amino acids, while the HR-A/B region is tightly connected without the embedded sequence in the middle in class B members. But strangely, the insert lengths between the HR-A and HR-B have almost no difference in pineapple HSFs (Figure 4 B). And the length of the total HR-A/B domain is about 42 amino acids almost the same in pineapple classes A, B, and C HSFs, while the length of classes A, and C HSFs is about 50 amino acids and 29 amino acids of class B HSFs in *Arabidopsis*, rice and soybean (Chauhan et al. 2011; Guo et al. 2008; Jin et al. 2013; Li et al. 2014).

The nuclear localization signals (NLS) and nuclear export signals (NES) are necessary for proteins to import and export the nucleus. Depending on the balance of nuclear import and export, the intracellular distribution of HSFs changes dynamically between the nucleus and cytoplasm (Heerklotz et al. 2001; Scharf et al. 1998). After detecting, almost all the HSFs contained NLS sequences rich in basic amino acid residues (K/R), except for AcHSF-B2a, AcHSF-B2b, and AcHSF-B4b. However, a total of 8 AcHSFs did not find the NES motifs. As reported in other plants, the

transcription activator AHA motif was only located in class A AcHSFs, but the difference is AcHSF-A3 lacks the AHA motif (Table 3).

In addition to the typical conserved domains of HSF, we also detected the putative motifs by Multiple Em for Motif Elicitation (MEME). A total of 10 different motifs were identified in AcHSFs with lengths ranging from 20 to 50 aa (Figure 5, Table 4). The members in the same group showed similar motif composition, but big differences between different groups were also found. The conserved motifs in HSFs indicated that all AcHSFs contained motif 1, motif 2, except for AcHSF-A9a and AcHSF-B4c lack of ~~the~~ motif 1. ~~The m~~Motif 3 only exists in class A and C HSFs, not in class B. However, the motif 7 only present in class A HSFs, and ~~the~~ motif 5 only present~~s~~ in class B HSFs. Additionally, some motifs were only discovered in a certain subfamily of AcHSFs, for example~~s~~, motif 9 was present in the B1 subclass (Figure 5).

Expression analysis of *AcHSFs* in different tissues

Gene expression profiles are related to their functions (Su et al. 2017). To better understand the functions of 22 pineapple *AcHSF* genes, the tissue-specific expression patterns were detected by 36 different tissues transcriptome sequencing, including flower (mixed stage), leaf, root, and fruit S1, S2, S3, S4, S5, and S7 from *Ming et al.* (Ming et al. 2015) and Se1-4, Petal 1-3, Ov 1-7, St 1-6, and Gy 1-7 from *Wang et al.* (Wang et al. 2020).

The ~~resulted~~ results showed some genes are highly expressed in certain tissues, while others are expressed gradually with the development of tissues (Figure 6). For example, *AcHSF-A1c* and *AcHSF-A7b* have high expression levels in 7 fruit tissues, the expression of *AcHSF-A9a* gradually increased in petal development and have the highest expression value in the P3 development stage. The *AcHSF-B4b* and *AcHSF-B4c* are highly expressed in the 7 ovule development stages, which illustrate their important roles in the pineapple ovule development process. We also found that some genes showed tissue-specific expression patterns, such as the *AcHSF-B2a* was mainly expressed in the fruit S7 stage, *AcHSF-A2* and *AcHSF-A6* are highly expressed

in leaf and flower tissues. In addition, the expression profiles of the genes in the same class are significantly different. For instance, three members of *AcHSF-A1* have ~~the~~ different expression patterns in all detected tissues and development stages.

Expression Profiles of *AcHSFs* Response to Various Stresses

To extend our understanding of *AcHSFs* in response to stresses, we performed qRT-PCR to investigate the expression patterns of 6 randomly selected *AcHSF* genes (*AcHSFA1a, AcHSFA2a, AcHSFA9a, AcHSFB2a, AcHSFB4a,* and *AcHSFC1a,*) in heat, cold and ABA stresses. The results illustrated that almost all of the selected *AcHSFs* showed similar expression patterns under the same stress conditions.

Cold stress drastically affects plant growth and development, and leads to a significant reduction in crop yield (Cai et al. 2015); therefore, ~~it is necessary for plants to~~plants must respond quickly to cold stress. As shown in the results, under the cold stress treatment, the expression of all the 6 *AcHSFs* increased rapidly from 0h to 24h~~,~~ and then reduced at 48h (Figure 7A). These may indicate that *AcHSFs* are commonly up-regulated within a ~~short~~ short-timer by cold stresses, and then the expression is down-regulated rapidly. As the Heat shock transcription factors, they play crucial roles in responding to the induction of heat shock. The result showed that the expression of 6 *AcHSFs* continues to increase from 0h to 48h in heat stress treatment (Figure 7B). After ABA treatment, the expression of most selected *AcHSF* genes increased from 0h to 12h, and then decreased after 12h, while the expression of *AcHSFA2a* continued to increase (Figure 7C). This result implies that the expressions of *AcHSFs* were suppressed under the longtime ABA treatment. In total, these findings indicate that *AcHSF* genes might play crucial roles in different stress response pathways.

DISSCUSSION

As one of the most popular fresh fruits worldwide and one of the most important commercial crops in the tropics, pineapple is severally destroyed by various abiotic stresses (cold, salt, drought, etc.) and biotic stresses (especially the fungal pathogen infection) during growth and developmental stages. HSF is one of the most key regulatory components of various abiotic and biotic stresses in plants. In this study, a comprehensive genome-wide analysis of the *AcHSF* family was identified and characterized for the first time. Consequently, a total of 30 *AcHSF* genes were identified from the pineapple genome. The widely accepted model of *HSFs* defines the necessity of HSF-type DBD and OD characterized by a coiled-coil structure, so 8 of them were discarded due to the absence of HSF-type DBD domains and/or coiled-coil structures. Meanwhile, pineapple HSF has a similar subfamily distribution compared with the monocots plant *O. sativa*, but is different from dicots plants *A. thaliana* and *P. trichocarpa*. Several genes are unique to monocots or dicots, for example, the subclasses *AcHSF-A8* and *AcHSF-B3* are restricted to dicots, while *AcHSF-A9* and *AcHSF-C2* are characteristic of monocots, which indicates that different evolutionary events of *HSF* genes occurred in dicots and monocots (Figure 2， Table 1).

In recent years, researches on intron-mediated regulation of gene expression have made significant progress (Le Hir et al. 2003; Li et al. 2019; Rose 2008; Shaul 2017), so the study of gene structure is very helpful to elucidate the gene function. Analysis of gene structures of *AcHSF* genes revealed that most of the classes A *AcHSFs* contain more than one intron, and several *AcHSFs* have 3 or 4 introns, such as *AcHSF-A1a*, *AcHSF-A1b*, and *AcHSF-A9a*. However, the genes in the class B and C only contain 1 intron, except for *AcHSF-C1b* (Figure 3). This particular intron structure may be due to the specific functions of the *AcHSF* genes. All 22 AcHSF proteins contain the necessary DBD domain and specific protein domains (HR-A/B, NLS, NES, RD, and AHA) (Table 3, Figure 4), which provide the structural basis for their conserved function (Giorno et al. 2012). The HSF DBD domain contains approximately 100 amino acid residues and is highly conserved in different organisms from plants to

350 animals; we also found the same conserved domain in pineapple (Figure 4A). And the
351 same as reported in other plants, the transcription activator AHA motif was only
352 located in class A AcHSFs, but the difference is AcHSF-A3 lacks the AHA motif
353 (Table 3). The members of *HSFs* lacking AHA domains might contribute differently to
354 the activator function or bind to other *HSFAs* to form hetero-oligomers (Guo et al.
355 2008).

356 The expression patterns analysis of different *AcHSFs* showed that *AcHSF-B4b* and
357 *AcHSF-B4c* are highly expressed in 7 ovule development stages, indicating the
358 potential functions in pineapple ovule development. The high expression levels of
359 *AcHSF-A7b*, *AcHSF-C2*, and *AcHSF-A1c* in fruit development stages uncovered their
360 important roles in fruit development (Figure 6). Furthermore, we found that the
361 expression of *AcHSF-A9a* gradually increased throughout the development stage, and
362 reached the highest expression level in the third stage of petal development. *AcHSF-A2*
363 *and AcHSF-A6* have high expression levels in leave and mixed flower tissues (Figure
364 6). These results suggest that they may participate in various developmental processes
365 or regulatory pathways. ~~Stress~~ The stress response is very important for ~~the~~ plant
366 growth and development. Previous studies have showed that the HSF genes of
367 *Arabidopsis*, *tomato, apple*, *Populus euphratica*, and *Phyllostachys edulis* are involved
368 in heat, cold, drought, and salt stress responses (Fragkostefanakis et al. 2015; Giorno et
369 al. 2012; Ikeda et al. 2011; Xie et al. 2019; Xue et al. 2014; Zhang et al. 2016a). In our
370 study, most of the selected *AcHSFs* showed ~~the~~ similar expression patterns under the
371 same stress conditions. Under the cold stress (4℃) treatment, the expressions of
372 *AcHSFs* were induced from 0h to 12h, and then inhibited after 12h (Figure 7A). The
373 same expression pattern was also observed in the 100mM ABA treatment, but the
374 difference was that the *AcHSFs* were more sensitive to ABA treatment (Figure 7C).
375 The continuous increase in expressions pattern of *AcHSFs* was observed at 45℃
376 treatment, indicating that heat ~~stress~~ stress-induced the expression of *AcHSFs* (Figure
377 7C).

378 Taken together, this study is the first to show the *AcHSF* family genes as well as
379 their specific expression profiles, which may be used as potential candidates for

genetic breeding in pineapple. However, gene expression and function analysis is a complex biological process, and more thorough studies are required to decipher the regulatory mechanisms.

CONCLUSIONS

In this study, we identified 22 *AcHSF* genes in pineapple (*Ananas comosus*), and collected ~~the~~ detailed information ~~of~~ on the gene and protein structures. The expression profiles of different tissues and development stages were ~~analysis~~ analyzed by the RNA-seq data, which may help to study their functions in various developmental processes or regulatory pathways. ~~In addition~~, W~~w~~e also showed that some *AcHSF* genes respond to a variety of biotic and abiotic stresses (heat, cold, and ABA), which may provide some information for developing new pineapple varieties with important agronomic traits, such as stress tolerance.

ACKNOWLEDGEMENTS

We would like to thank the reviewers for their helpful comments on the original manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (U1605212, 31761130074 and 31970333 to Yuan Qin; 31700279 to Hanyang Cai) and a Guangxi Distinguished Experts Fellowship to Yuan Qin. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

National Natural Science Foundation of China: U1605212, 31761130074 and 31970333.

Guangxi Distinguished Experts Fellowship.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

Lulu Wang conceived and designed the experiments, prepared figures and tables, authored and reviewed drafts of the paper, and approved the final draft.

Yanhui Liu prepared figures and/or tables, and approved the final draft.

Huihuang Chen, Mengnan Chai and Xiaoping Niu analyzed the data, performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Yuan Qin and Hanyang Cai conceived and designed the experiments, reviewed drafts of the paper, and approved the final draft.

Reference

Bai YH, Dai XZ, Li Y, Wang LL, Li WM, Liu YH, Cheng Y, and Qin Y. 2019. Identification and characterization of pineapple leaf lncRNAs in crassulacean acid metabolism (CAM) photosynthesis pathway. *Scientific Reports* 9. ARTN 6658 10.1038/s41598-019-43088-8

Barral B, Chillet M, Lechaudel M, Lartaud M, Verdeil JL, Conejero G, and Schorr-Galindo S. 2019. An Imaging Approach to Identify Mechanisms of Resistance to Pineapple Fruitlet Core Rot. *Front Plant Sci* 10:1065. 10.3389/fpls.2019.01065

Cai H, Cheng J, Yan Y, Xiao Z, Li J, Mou S, Qiu A, Lai Y, Guan D, and He S. 2015. Genome-wide identification and expression analysis of calcium-dependent protein kinase and its closely related kinase genes in Capsicum annuum. *Front Plant Sci* 6:737. 10.3389/fpls.2015.00737

Chai M, Cheng H, Yan M, Priyadarshani S, Zhang M, He Q, Huang Y, Chen F, Liu L, Huang X, Lai L, Chen H, Cai H, and Qin Y. 2020. Identification and expression analysis of the DREB transcription factor family in pineapple (Ananas comosus (L.) Merr.). *PeerJ* 8:e9006. 10.7717/peerj.9006

Chauhan H, Khurana N, Agarwal P, Khurana JP, and Khurana P. 2013. A seed preferential heat shock transcription factor from wheat provides abiotic stress tolerance and yield enhancement in transgenic Arabidopsis under heat stress environment. *PLoS One* 8:e79577. 10.1371/journal.pone.0079577

Chauhan H, Khurana N, Agarwal P, and Khurana P. 2011. Heat shock factors in rice (Oryza sativa L.): genome-wide expression analysis during reproductive development and abiotic stress. *Mol Genet Genomics* 286:171-187. 10.1007/s00438-011-0638-8

Cheng Q, Zhou Y, Liu Z, Zhang L, Song G, Guo Z, Wang W, Qu X, Zhu Y, and Yang D. 2015. An alternatively spliced heat shock transcription factor, OsHSFA2dI, functions in the heat stress-induced unfolded protein response in rice. *Plant Biology* 17:419-429. 10.1111/plb.12267

Chidambaranathan P, Jagannadham PTK, Satheesh V, Kohli D, Basavarajappa SH, Chellapilla B, Kumar J, Jain PK, and Srinivasan R. 2018. Genome-wide analysis identifies chickpea (Cicer arietinum) heat stress transcription factors (Hsfs) responsive to heat stress at the pod development stage. *Journal of Plant Research* 131:525-542. 10.1007/s10265-017-0948-y

Chung E, Kim KM, and Lee JH. 2013. Genome-wide analysis and molecular characterization of heat shock transcription factor family in Glycine max. *J Genet Genomics* 40:127-135. 10.1016/j.jgg.2012.12.002

Fragkostefanakis S, Roth S, Schleiff E, and Scharf KD. 2015. Prospects of engineering thermotolerance in crops through modulation of heat stress transcription factor and heat shock protein networks. *Plant Cell and Environment* 38:1881-1895. 10.1111/pce.12396

Giesguth M, Sahm A, Simon S, and Dietz KJ. 2015. Redox-dependent translocation of the heat shock transcription factor AtHSFA8 from the cytosol to the nucleus in Arabidopsis thaliana. *FEBS Lett* 589:718-725. 10.1016/j.febslet.2015.01.039

Giorno F, Guerriero G, Baric S, and Mariani C. 2012. Heat shock transcriptional factors in Malus domestica: identification, classification and expression analysis. *BMC Genomics* 13:639. 10.1186/1471-2164-13-639

Guo AY, Zhu QH, Chen X, and Luo JC. 2007. [GSDS: a gene structure display server]. *Yi Chuan* 29:1023-1026.

Guo J, Wu J, Ji Q, Wang C, Luo L, Yuan Y, Wang Y, and Wang J. 2008. Genome-wide analysis of heat

466 shock transcription factor families in rice and Arabidopsis. *J Genet Genomics* 35:105-118.
467 10.1016/S1673-8527(08)60016-8
468 Guo M, Liu JH, Ma X, Luo DX, Gong ZH, and Lu MH. 2016. The Plant Heat Stress Transcription Factors
469 (HSFs): Structure, Regulation, and Function in Response to Abiotic Stresses. *Frontiers in Plant*
470 *Science* 7. ARTN 11410.3389/fpls.2016.00114
471 Guo M, Lu JP, Zhai YF, Chai WG, Gong ZH, and Lu MH. 2015. Genome-wide analysis, expression profile
472 of heat shock factor gene family (CaHsfs) and characterisation of CaHsfA2 in pepper (Capsicum
473 annuum L.). *BMC Plant Biol* 15:151. 10.1186/s12870-015-0512-7
474 Guo M, Yin YX, Ji JJ, Ma BP, Lu MH, and Gong ZH. 2014. Cloning and expression analysis of heat-shock
475 transcription factor gene CaHsfA2 from pepper (Capsicum annuum L.). *Genet Mol Res*
476 13:1865-1875. 10.4238/2014.March.17.14
477 Heerklotz D, Doring P, Bonzelius F, Winkelhaus S, and Nover L. 2001. The balance of nuclear import and
478 export determines the intracellular distribution and function of tomato heat stress
479 transcription factor HsfA2. *Mol Cell Biol* 21:1759-1768. 10.1128/MCB.21.5.1759-1768.2001
480 Higo K, Ugawa Y, Iwamoto M, and Korenaga T. 1999. Plant cis-acting regulatory DNA elements (PLACE)
481 database: 1999. *Nucleic Acids Res* 27:297-300. 10.1093/nar/27.1.297
482 Hu H, and Xiong L. 2014. Genetic engineering and breeding of drought-resistant crops. *Annu Rev Plant*
483 *Biol* 65:715-741. 10.1146/annurev-arplant-050213-040000
484 Hu Y, Han YT, Wei W, Li YJ, Zhang K, Gao YR, Zhao FL, and Feng JY. 2015. Identification, isolation, and
485 expression analysis of heat shock transcription factors in the diploid woodland strawberry
486 Fragaria vesca. *Front Plant Sci* 6:736. 10.3389/fpls.2015.00736
487 Ikeda M, Mitsuda N, and Ohme-Takagi M. 2011. Arabidopsis HsfB1 and HsfB2b Act as Repressors of the
488 Expression of Heat-Inducible Hsfs But Positively Regulate the Acquired Thermotolerance. *Plant*
489 *Physiology* 157:1243-1254. 10.1104/pp.111.179036
490 Jin GH, Gho HJ, and Jung KH. 2013. A systematic view of rice heat shock transcription factor family using
491 phylogenomic analysis. *J Plant Physiol* 170:321-329. 10.1016/j.jplph.2012.09.008
492 Le Hir H, Nott A, and Moore MJ. 2003. How introns influence and enhance eukaryotic gene expression.
493 *Trends in Biochemical Sciences* 28:215-220. 10.1016/S0968-0004(03)00052-5
494 Letunic I, Doerks T, and Bork P. 2012. SMART 7: recent updates to the protein domain annotation
495 resource. *Nucleic Acids Res* 40:D302-305. 10.1093/nar/gkr931
496 Li PS, Yu TF, He GH, Chen M, Zhou YB, Chai SC, Xu ZS, and Ma YZ. 2014. Genome-wide analysis of the Hsf
497 family in soybean and functional identification of GmHsf-34 involvement in drought and heat
498 stresses. *BMC Genomics* 15:1009. 10.1186/1471-2164-15-1009
499 Li W, Cui X, Meng ZL, Huang XH, Xie Q, Wu H, Jin HL, Zhang DB, and Liang WQ. 2012. Transcriptional
500 Regulation of Arabidopsis MIR168a and ARGONAUTE1 Homeostasis in Abscisic Acid and
501 Abiotic Stress Responses. *Plant Physiology* 158:1279-1292. 10.1104/pp.111.188789
502 Li YJ, Chen DM, Luo SW, Zhu YY, Jia XJ, Duan YB, and Zhou MG. 2019. Intron-mediated regulation of
503 beta-tubulin genes expression affects the sensitivity to carbendazim in Fusarium graminearum.
504 *Current Genetics* 65:1057-1069. 10.1007/s00294-019-00960-4
505 Lin Y, Cheng Y, Jin J, Jin X, Jiang H, Yan H, and Cheng B. 2014. Genome duplication and gene loss affect
506 the evolution of heat shock transcription factor genes in legumes. *PLoS One* 9:e102825.
507 10.1371/journal.pone.0102825
508 Lin YX, Jiang HY, Chu ZX, Tang XL, Zhu SW, and Cheng BJ. 2011. Genome-wide identification,
509 classification and analysis of heat shock transcription factor family in maize. *BMC Genomics*

12:76. 10.1186/1471-2164-12-76

Liu ZW, Wu ZJ, Li XH, Huang Y, Li H, Wang YX, and Zhuang J. 2016. Identification, classification, and expression profiles of heat shock transcription factors in tea plant (Camellia sinensis) under temperature stress. *Gene* 576:52-59. 10.1016/j.gene.2015.09.076

Ming R, VanBuren R, Wai CM, Tang HB, Schatz MC, Bowers JE, Lyons E, Wang ML, Chen J, Biggers E, Zhang JS, Huang LX, Zhang LM, Miao WJ, Zhang J, Ye ZY, Miao CY, Lin ZC, Wang H, Zhou HY, Yim WC, Priest HD, Zheng CF, Woodhouse M, Edger PP, Guyot R, Guo HB, Guo H, Zheng GY, Singh R, Sharma A, Min XJ, Zheng Y, Lee H, Gurtowski J, Sedlazeck FJ, Harkess A, McKain MR, Liao ZY, Fang JP, Liu J, Zhang XD, Zhang Q, Hu WC, Qin Y, Wang K, Chen LY, Shirley N, Lin YR, Liu LY, Hernandez AG, Wright CL, Bulone V, Tuskan GA, Heath K, Zee F, Moore PH, Sunkar R, Leebens-Mack JH, Mockler T, Bennetzen JL, Freeling M, Sankoff D, Paterson AH, Zhu XG, Yang XH, Smith JAC, Cushman JC, Paull RE, and Yu QY. 2015. The pineapple genome and the evolution of CAM photosynthesis. *Nature Genetics* 47:1435-+. 10.1038/ng.3435

Nishizawa-Yokoi A, Nosaka R, Hayashi H, Tainaka H, Maruta T, Tamoi M, Ikeda M, Ohme-Takagi M, Yoshimura K, Yabuta Y, and Shigeoka S. 2011. HsfA1d and HsfA1e Involved in the Transcriptional Regulation of HsfA2 Function as Key Regulators for the Hsf Signaling Network in Response to Environmental Stress. *Plant and Cell Physiology* 52:933-945. 10.1093/pcp/pcr045

Ohama N, Sato H, Shinozaki K, and Yamaguchi-Shinozaki K. 2017. Transcriptional Regulatory Network of Plant Heat Stress Response. *Trends Plant Sci* 22:53-65. 10.1016/j.tplants.2016.08.015

Olate E, Jimenez-Gomez JM, Holuigue L, and Salinas J. 2018. NPR1 mediates a novel regulatory pathway in cold acclimation by interacting with HSFA1 factors. *Nat Plants* 4:811-823. 10.1038/s41477-018-0254-2

Pereira A. 2016. Plant Abiotic Stress Challenges from the Changing Environment. *Front Plant Sci* 7:1123. 10.3389/fpls.2016.01123

Priyadarshani S, Hu B, Li W, Ali H, Jia H, Zhao L, Ojolo SP, Azam SM, Xiong J, Yan M, Ur Rahman Z, Wu Q, and Qin Y. 2018. Simple protoplast isolation system for gene expression and protein interaction studies in pineapple (Ananas comosus L.). *Plant Methods* 14:95. 10.1186/s13007-018-0365-9

Rose AB. 2008. Intron-Mediated Regulation of Gene Expression. *Nuclear Pre-Mrna Processing in Plants* 326:277-290.

Scharf KD, Berberich T, Ebersberger I, and Nover L. 2012. The plant heat stress transcription factor (Hsf) family: structure, function and evolution. *Biochim Biophys Acta* 1819:104-119. 10.1016/j.bbagrm.2011.10.002

Scharf KD, Heider H, Hohfeld I, Lyck R, Schmidt E, and Nover L. 1998. The tomato Hsf system: HsfA2 needs interaction with HsfA1 for efficient nuclear import and may be localized in cytoplasmic heat stress granules. *Mol Cell Biol* 18:2240-2251. 10.1128/mcb.18.4.2240

Scharf KD, Rose S, Zott W, Schoffl F, and Nover L. 1990. Three tomato genes code for heat stress transcription factors with a region of remarkable homology to the DNA-binding domain of the yeast HSF. *EMBO J* 9:4495-4501.

Shaul O. 2017. How introns enhance gene expression. *International Journal of Biochemistry & Cell Biology* 91:145-155. 10.1016/j.biocel.2017.06.016

Shim D, Hwang JU, Lee J, Lee S, Choi Y, An G, Martinoia E, and Lee Y. 2009. Orthologs of the class A4 heat shock transcription factor HsfA4a confer cadmium tolerance in wheat and rice. *Plant Cell* 21:4031-4043. 10.1105/tpc.109.066902

Singh A, Mittal D, Lavania D, Agarwal M, Mishra RC, and Grover A. 2012. OsHsfA2c and OsHsfB4b are

involved in the transcriptional regulation of cytoplasmic OsClpB (Hsp100) gene in rice (Oryza sativa L.). *Cell Stress & Chaperones* 17:243-254. 10.1007/s12192-011-0303-5

Song X, Liu G, Duan W, Liu T, Huang Z, Ren J, Li Y, and Hou X. 2014. Genome-wide identification, classification and expression analysis of the heat shock transcription factor family in Chinese cabbage. *Mol Genet Genomics* 289:541-551. 10.1007/s00438-014-0833-5

Sorger PK, and Pelham HRB. 1988. Yeast heat-shock factor is an essential DNA -binding protein that exhibits temperature-dependent phosphorylation. *Cell* 54:855-864.

Su ZX, Wang LL, Li WM, Zhao LH, Huang XY, Azam SM, and Qin Y. 2017. Genome-Wide Identification of Auxin Response Factor (ARF) Genes Family and its Tissue-Specific Prominent Expression in Pineapple (Ananas comosus). *Tropical Plant Biology* 10:86-96. 10.1007/s12042-017-9187-6

Wahid A, Gelani S, Ashraf M, and Foolad MR. 2007. Heat tolerance in plants: An overview. *Environmental and Experimental Botany* 61:199-223. 10.1016/j.envexpbot.2007.05.011

Wang F, Dong Q, Jiang H, Zhu S, Chen B, and Xiang Y. 2012. Genome-wide analysis of the heat shock transcription factors in Populus trichocarpa and Medicago truncatula. *Mol Biol Rep* 39:1877-1886. 10.1007/s11033-011-0933-9

Wang J, Sun N, Deng T, Zhang L, and Zuo K. 2014. Genome-wide cloning, identification, classification and functional analysis of cotton heat shock transcription factors in cotton (Gossypium hirsutum). *BMC Genomics* 15:961. 10.1186/1471-2164-15-961

Wang LL, Li Y, Jin XY, Liu LP, Dai XZ, Liu YH, Zhao LH, Zheng P, Wang XM, Liu YQ, Lin DS, and Qin Y. 2020. Floral transcriptomes reveal gene networks in pineapple floral growth and fruit development. *Communications Biology* 3. ARTN 500
10.1038/s42003-020-01235-2

Xie LH, Li XY, Hou D, Cheng ZC, Liu J, Li J, Mu SH, and Gao J. 2019. Genome-Wide Analysis and Expression Profiling of the Heat Shock Factor Gene Family in Phyllostachys edulis during Development and in Response to Abiotic Stresses. *Forests* 10. ARTN 100
10.3390/f10020100

Xue GP, Sadat S, Drenth J, and McIntyre CL. 2014. The heat shock factor family from Triticum aestivum in response to heat and other major abiotic stresses and their role in regulation of heat shock protein genes. *J Exp Bot* 65:539-557. 10.1093/jxb/ert399

Yabuta Y. 2016. Functions of heat shock transcription factors involved in response to photooxidative stresses in Arabidopsis. *Bioscience Biotechnology and Biochemistry* 80:1254-1263. 10.1080/09168451.2016.1176515

Zhang J, Jia H, Li J, Li Y, Lu M, and Hu J. 2016a. Molecular evolution and expression divergence of the Populus euphratica Hsf genes provide insight into the stress acclimation of desert poplar. *Sci Rep* 6:30050. 10.1038/srep30050

Zhang J, Jia HX, Li JB, Li Y, Lu MZ, and Hu JJ. 2016b. Molecular evolution and expression divergence of the Populus euphratica Hsf genes provide insight into the stress acclimation of desert poplar. *Scientific Reports* 6. ARTN 3005010.1038/srep30050

Zhang Y, Chen C, Jin XF, Xiong AS, Peng RH, Hong YH, Yao QH, and Chen JM. 2009. Expression of a rice DREB1 gene, OsDREB1D, enhances cold and high-salt tolerance in transgenic Arabidopsis. *Bmb Reports* 42:486-492. DOI 10.5483/BMBRep.2009.42.8.486

Zhu JK. 2016. Abiotic Stress Signaling and Responses in Plants. *Cell* 167:313-324. 10.1016/j.cell.2016.08.029