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], Mohammad Aslam[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: Transcription factors (TFs), such as Heat shock transcription factors (HSFs), usually play critical regulatory functions in plant development, growth, and response to environmental cues. However, no HSFs have been characterized in pineapple thus far. Here, we identified 22 *AcHSF* genes 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). The *AcHSF* promoters contained various *cis*-elements associated with stress, hormones, and plant development processes, for instance, STRE, WRKY, and ABRE binding sites. The majority of *HSFs* were expressed in diverse 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 pineapple *HSF* gene family.

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

INTRODUCTION

The living environment of plants is faced with many challenges, including cold, heat, drought, and salinity stresses (Hu & Xiong 2014; Pereira 2016; Zhu 2016). Due to global warming, heat stress is becoming a serious agricultural threat for agricultural production and planting areas worldwide (Wahid et al. 2007). Typically, plants face heat stress when the temperature rises 10 to 15 degrees above the optimum growth environment. Heat stress affects plant development and growth and eventually leads to a decrease in crop yield. Consequently, as a defense or signaling mechanism in response to environmental stresses, plants regulate the expression of several genes through different transcription factors (TFs). The HSF transcription factor family of plants is involved in heat stress response and regulates the expression of several stress-responsive proteins, including heat shock proteins (HSPs), ascorbate peroxidase (APX), and catalase (CAT) (Ohama et al. 2017). Previously, several studies have validated the roles of HSFs in plant stress response. For example, the overexpression of *CarHsf-A2* could enhance chickpea stress tolerance without any pleiotropic effects (Chidambaranathan et al. 2018). Besides, *AtHSF-B1* and *AtHsf-B2b* act as expression repressors after heat-stress, and the *AtHSF-A1* is involved in cold acclimation in *Arabidopsis thaliana* (Ikeda et al. 2011; Olate et al. 2018). Similarly, *PeuHsf-A2* gets induced by heat stress, increasing desert poplar acclimation (Zhang et al. 2016b).

HSFs contain several evolutionarily conserved functional domains. The canonical HSF protein contains an N-terminal DNA-binding domain (DBD) that binds to HSEs; a hydrophobic amino acid residue (HR-A/B) oligomerization domain (OD) heptad repeat bound to DBD by a flexible linker. In addition, it also includes a region of nuclear localization signal (NLS), a region of nuclear export signal (NES), and a motif of activator AHA located at the C-terminal (Guo et al. 2016; Nishizawa-Yokoi et al. 2011; Singh et al. 2012; Yabuta 2016). Three types of plant HSF have been identified based on the variable linker (usually 15-80 aa) and HR-A/B domain, including nine

class A (A1 - A9), four class B (B1 - B4), and two class C (C1- C2) (Giesguth et al. 2015; Nishizawa-Yokoi et al. 2011; Shim et al. 2009; Singh et al. 2012; Yabuta 2016). Initially, HSFs were identified in yeast (Scharf et al. 1990), and the first HSF plant gene was identified in tomatoes (Sorger & Pelham 1988). Since then, several plant HSF gene families, 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) and tea plant (*Camellia sinensis*) (Liu et al. 2016), etc.

Pineapple (*Ananas comosus*) is grown in subtropical and tropical regions and is widely loved worldwide and its genome is sequenced (Bai et al. 2019; Ming et al. 2015). However, many biotic and abiotic environmental stresses, pathogen infection, and degradation of good breeds limit pineapple production (Barral et al. 2019). Therefore, it is essential to identify and characterize genes involved in response to environmental stresses and study the underlying molecular mechanism that could be used for possible genetic breeding applications. *HSFs* are well known for the co-involvement in different environmental cues, such as cold, heat, and ABA. However, *AcHSFs* and their possible role in pineapples have not been explored. Here, whole-genome identification and expression analysis of *AcHSFs* gene during flower, fruit development, and abiotic stress was conducted to expand the understanding of *AcHSFs* gene and its application in genetic breeding.

METHODS

Identification and Characterization of *AcHSF* Genes in pineapple

The amino acid sequences of pineapple HSFs were retrieved using the HSF-type DBD domain (Pfam: PF00447) as a query in *Phytozome* JGI. A total of 30 pineapple HSFs

were obtained from JGI. Only 22 *AcHSFs* left, after filtering out the non-typical HSF-type DBD and repeated sequences or canonical coiled-coil structures by SMART online tool (Letunic et al. 2012). The information of Chromosome localization, CDS, and AA length for *AcHSFs* was obtained from JGI *Phytozome v12.1*. The biophysical properties of coding AcHSFs were calculated using the Expasy ProtParam tool. The subcellular localization of *AcHSFs* was analyzed using BUSCA (http://busca.biocomp.unibo.it/).

Chromosome Localization phylogenetic relationships

The information of all pineapple *HSFs'* chromosome localization sites was acquired from *Phytozome v12.1*, including chromosome length, chromosome location, and gene start site. The MapChart v2.0 (https://mapchart.net/) was adopted to map the chromosomal location. The phylogenetic relationship of different HSF proteins was explored, and the phylogenetic tree was created using the AcHSF amino acid sequences and the other three species, *Arabidopsis thaliana* (*A. thaliana*, *Oryza sativa* (*O. sativa)*, and *Populus. trichocarpa* (*P. trichocarpa*) by the MEGAX with a bootstrap value of 1000. The HSF gene in pineapples was referred to as *AcHSF* genes and classified according to HSFs in the phylogenetic tree classes A, B, and C,

Genetic structure and *cis*-acting elements

The gene structures of *AcHSF*, including exons, introns, and UTR were displayed by the GSDS online tool, (Guo et al. 2007). The promoter sequence of *AcHSFs* found in *Phytozome* is located 2kb upstream of the translation initiation site. These sequences were analyzed using a plant *cis*-acting element database New PLACE (Higo et al. 1999), to identify *cis*-elements necessary for gene expression, development, and hormone signaling under abiotic stress.

Conserved domains and motifs analysis of AcHSFs

Clustal X 2.0 and DNAMAN software was used to align and edit the DBD domain and

HR-A/B regions (OD). Using the cNLS online tool, NLS domains were predicted and the NetNES 1.1 online server identified NES domains in the *AcHSFs*. MEME server (http://meme-suite.org/) was applied to define the conserved motifs of AcHSFs following the parameters: the number of repetitions = any, the maximum number of motifs = 10, minimum width≥10, maximum width≤50, and motifs with an E-value < 0.01 were retained.

Expression patterns analysis

The transcriptomic data generated from different organs and developmental stages of pineapple have been described previously (Ming et al. 2015; 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, and root, mixed flower, leaf, and 6 stages of fruit ~~from *Wang et al*~~ were used to generate heatmap using the pheatmap package of R software.

Stress treatments

One-month-old uniform tissue-cultured seedlings in a rooting medium were used for the stress treatment analyses (Priyadarshani et al. 2018). The stress treatments of pineapple seeding as follows: cold (4℃), heat (45℃), and ABA (100 mM). Leaves were collected from three independent plants at 12 h, 24 h and 48 h after treatment and immediately stored in liquid nitrogen before RNA extraction.Untreated pineapple seedlings of the same size were used as controls.

RNA extraction and qRT-PCR

Total RNA of pineapple leaf tissues was obtained using the RNeasy Plant Mini Kit (50) (Qiagen). For cDNA synthesis, a total of 1 μg RNA per sample was used with cDNA Synthesis SuperMix (Transgen, Beijing, China). qPCR was conducted by the Bio-Rad CFX manager machine using TransStart® Top Green qPCR SuperMix (Transgen, Beijing, China). Pineapple *Actin2* is used as the reference gene for qPCR (Wang et al.

2020). There are a total of three biological replicates for each sample, and the results are shown as the mean ± standard deviations.

RESULTS

Genome-Wide Identification of *HSF* Genes in Pineapple

HSF-type DBD domains (Pfam: PF00447) amino acid sequences were submitted into *Ananas comosus* v3 *Phytozome* database v12.1. A total of 30 putative pineapple HSF sequences were acquired. And then, checked by the *SMART* online tool and *Pfam* database, 1 pineapple HSF sequence was rejected due to the absence of typical HSF-DBD domains, and 7 HSF sequences were abandoned due to the absence of coiled-coil structure. As a result, 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 gene with the longest amino acid length is *ACHSF-A5*, which contains 601 amino acids and also has the largest molecular weight of 65.81 KDa; and the gene with the shortest amino acid length was *AcHSF-A9b*, which contains 129 amino acids and also has the smallest molecular weight of 13.77 KDa. Prediction of protein isoelectric points (pI) can aid in the purification and isolation of proteins. The predicted isoelectric points (pI) of *AcHSFs* ranged from 4.68 (*AcHSF*-B2b) to 9.63 (*AcHSF*-B4c). Detailed information on other parameters has been given 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 for each chromosome varied significantly, 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

A phylogenetic analysis of 31 *Populus trichocarpa* HSFs, 25 rice HSFs, and 21 *Arabidopsis* HSFs was performed to classify the phylogenetic relationships (Guo et al. 2016), together with those of *AcHSFs* by generating a neighbor-joining phylogenetic tree. The HSFs were grouped into three clusters, A, B, and C, according to the difference between the amino acid sequences of the DBD domain, the HR-A/B region, and the linker between them (Guo et al. 2016; Scharf et al. 2012). Class A consisted of 10 subclusters, designated A1 to A10. Class B contained B1 to B4, and Class C comprises C1 and C2 sub-clusters. 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*

Intron, exon, 5 'UTR, and 3' UTR structures were analyzed using Gene Structure Display Server (GSDS) v2.0.0 to reveal the gene structural features of AcHSFs. The number of exons for AcHSFs ranged from 1 to 5 (Figure 3), while only in the longest *AcHSF-A9a* (genomic sequence 34598 bp.) was found 5 exons. The 5’ UTR, and 3’ UTR 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 5’ UTR and 3’ UTR sequences, 3 *AcHSFs* (*AcHSF-A1b*, *AcHSF-A9a*, and *AcHSF-B4b*) have only 5’ UTR sequences, 6 *AcHSFs* (*AcHSF-A4*, *AcHSF-A5*, *AcHSF-A7*, *AcHSF-B1a*, *AcHSF-B1b*, and *AcHSF-C2*) have only 3’ UTR

sequences, and 5 *AcHSFs* (*AcHSF-A1a*, *AcHSF-A1c*, *AcHSF-A2*, *AcHSF-A6*, and *AcHSF-A9*) both have 5' UTR and 3' UTR sequences.

Stress responses elements of the dehydration-responsive element (DRE), ABA-responsive element (ABRE), low-temperature responsive element (LTRE), MYB, MYC, and WRKY elements has been reported to play important roles in drought, salt, cold, ABA, and GA responses (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-specific expression related elements (ROOTMOTIFTAPOX1) can be found. 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 cis-element studies in the promoters indicate that HSFs are highly related to the response to stress.

Conserved domains and motifs of pineapple *HSFs*

The modular structure of the HSFs 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. The intracellular distribution of HSFs varies dynamically between the nucleus and the cytoplasm, depending on nuclear import and export balance. (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). The motif composition of the same group members is similar, but there are great differences among different group members. The conserved motifs in HSFs indicated that all AcHSFs contained motif 1, motif 2, except for AcHSF-A9a and AcHSF-B4c lack of motif 1. Motif 3 only exists in class A and C HSFs, not in class B. However, motif 7 only present in class A HSFs, and motif 5 only presents in class B HSFs. Additionally, some motifs were only discovered in a certain subfamily of AcHSFs, for example, 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, fruit S1, S2, S3, S4, S5, and S7, Se1-4, Petal 1-3, Ov 1-7, St 1-6, and Gy 1-7 from *Wang et al.* (Wang et al. 2020).

The 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 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 (*AcHSF-A1a, AcHSF-A2a, AcHSF-A9a, AcHSF-B2a, AcHSF-B4a*, and *AcHSF-C1a*) 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, 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 24 h and then reduced at 48 h (Figure 7A). These may indicate that *AcHSFs* are commonly up-regulated within a short-timer by cold stresses, and then the expression is down-regulated rapidly. Heat shock transcription factors play crucial roles in response to heat shock induction. The result showed that the expression of 6 *AcHSFs* continues to increase from 0 h to 48 h in heat stress treatment (Figure 7B). After ABA treatment, the expression of most selected *AcHSF* genes increased from 0h to 12 h, and then decreased after 12 h, while the expression of *AcHSF-A2a* continued to increase (Figure 7C). This result implies that the expressions of *AcHSFs* were suppressed under the longtime ABA treatment and might play crucial roles in different stress (cold and heat, etc) response pathways.

DISCUSSION

During its growth and developmental stages, the pineapple is severely destroyed by various abiotic stresses (cold, heat, drought, etc.) and biotic stresses (especially fungal pathogen infection). HSFs are among the critical regulatory components of various abiotic and biotic stresses in plants. This research identified and characterized, for the first time, a systematic genome-wide review of the AcHSF family. Consequently, from the pineapple genome, a total of 30 AcHSF genes were identified. The widely accepted model of *HSFs* defines the necessity of HSF-type DBD and OD characterized by a coiled-coil structure. Thus, due to the absence of HSF-type DBD domains and/or coiled-coil structures, 8 of them were discarded. 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*. Some genes are unique to monocots or dicots. For example, the subclasses AcHSF-A8 and AcHSF-B3 are confined to dicots, while AcHSF-A9 and AcHSF-C2 are characteristic of monocots,

suggesting that different evolutionary events of HSF genes occurred in dicots and monocots (Figure 2, Table 1).

Research on gene expression regulation mediated by introns has made significant progress (Le Hir et al. 2003; Li et al. 2019; Rose 2008; Shaul 2017). Therefore the study of gene structure is beneficial to elucidate the gene function. Analysis of AcHSFs gene structures 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. The required DBD domain and unique protein domains (HR-A/B, NLS, NES, RD, and AHA) are found in all 22 AcHSF proteins (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 acids and is strongly preserved in various plant-to-animal species; we also found the same conserved domain in pineapple (Figure 4A). As reported previously in other plants, the transcription activator AHA motif was only located in class A AcHSFs, but AcHSF-A3 lacks the AHA motif (Table 3). The *HSFs* that lack AHA domains might contribute to the activator's function differently or form hetero-oligomers by binding to other *HSF-As* (Guo et al. 2008).

The expression patterns analysis of different *AcHSFs* showed that *AcHSF-B4b* and *AcHSF-B4c* are highly expressed in 7 ovule development stages, indicating the potential functions in pineapple ovule development. The high expression levels of *AcHSF-A7b*, *AcHSF-C2*, and *AcHSF-A1c* in fruit development stages uncovered their important roles in fruit development (Figure 6). Furthermore, we found that the expression of *AcHSF-A9a* gradually increased throughout the development stage and reached the highest expression level in the third stage of petal development. *AcHSF-A2 and AcHSF-A6* have high expression levels in leave and mixed flower tissues (Figure 6). These data suggest that *AcHSFs* may regulate several developmental processes. The stress response is very significant for plant growth and development. Previous studies

have shown that the *HSF* genes are involved in several abiotic stress response, including heat, cold, drought, and salt stress responses in different plants such as *Arabidopsis*, tomato, apple, *Populus euphratica*, and *Phyllostachys edulis* (Fragkostefanakis et al. 2015; Giorno et al. 2012; Ikeda et al. 2011; Xie et al. 2019; Xue et al. 2014; Zhang et al. 2016a). In our study, most of the selected *AcHSFs* showed similar expression patterns under the same stress conditions. Under the cold stress (4°C) treatment, the expressions of *AcHSFs* were induced from 0 h to 12 h and then inhibited after 12 h (Figure 7A). The same expression pattern was also observed in the 100 mM ABA treatment, but the difference was that the *AcHSFs* were more sensitive to ABA treatment (Figure 7C). The continuous increase in the expression pattern of *AcHSFs* was observed at 45°C treatment, indicating that heat stress-induced the expression of *AcHSFs* (Figure 7C).

Taken together, this study is the first to identify the *AcHSF* family genes, their properties as well as their expression profiles. This information could be used to utilize them as potential candidates in a breeding program of pineapple. However, gene expression and function analysis are complicated biological mechanisms, and additional studies are necessary to interpret the regulatory process.

CONCLUSIONS

In the present study, 22 *AcHSF* genes were identified in pineapple (*Ananas comosus*) and generated detailed information on the gene and protein structures. The expression profiles of various tissues and developmental stages were analyzed by the RNA-seq data, which may help to study their functions in different developmental processes or regulatory pathways. We also showed that some *AcHSF* genes participate in various biotic and abiotic stresses (heat, cold, and ABA), which may help develop new pineapple varieties with desired agronomic traits 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, Aslam, 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 665810.1038/s41598-019-43088-8

Barral B, Chillet M, Lechaudel M, Lartaud M, Verdeil JL, Conejero G, and Schorr-Galindo S. 2019. An Imaging aApproach to iIdentify mMechanisms of rResistance to pPineapple fFruitlet cCore rRot. *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 shock transcription factor families in rice and Arabidopsis. *J Genet Genomics* 35:105-118.

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

10.1016/S1673-8527(08)60016-8

Guo M, Liu JH, Ma X, Luo DX, Gong ZH, and Lu MH. 2016. The pPlant hHeat sStress tTranscription fFactors (HSFs): sStructure, rRegulation, and fFunction in rResponse to aAbiotic sStresses. *Frontiers in Plant Sci ence* 7. ARTN 11410.3389/fpls.2016.00114

Guo M, Lu JP, Zhai YF, Chai WG, Gong ZH, and Lu MH. 2015. Genome-wide analysis, expression profile of heat shock factor gene family (CaHsfs) and characterisation of CaHsfA2 in pepper (*Capsicum annuum* L.). *BMC Plant Biol* 15:151. 10.1186/s12870-015-0512-7

Guo M, Yin YX, Ji JJ, Ma BP, Lu MH, and Gong ZH. 2014. Cloning and expression analysis of heat-shock transcription factor gene CaHsfA2 from pepper (*Capsicum annuum* L.). *Genet Mol Res* 13:1865-1875. 10.4238/2014.March.17.14

Heerklotz D, Doring P, Bonzelius F, Winkelhaus S, and Nover L. 2001. The balance of nuclear import and export determines the intracellular distribution and function of tomato heat stress transcription factor HsfA2. *Mol Cell Biol* 21:1759-1768. 10.1128/MCB.21.5.1759-1768.2001

Higo K, Ugawa Y, Iwamoto M, and Korenaga T. 1999. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27:297-300. 10.1093/nar/27.1.297

Hu H, and Xiong L. 2014. Genetic engineering and breeding of drought-resistant crops. *Annu Rev Plant Biol* 65:715-741. 10.1146/annurev-arplant-050213-040000

Hu Y, Han YT, Wei W, Li YJ, Zhang K, Gao YR, Zhao FL, and Feng JY. 2015. Identification, isolation, and expression analysis of heat shock transcription factors in the diploid woodland strawberry *Fragaria vesca*. *Front Plant Sci* 6:736. 10.3389/fpls.2015.00736

Ikeda M, Mitsuda N, and Ohme-Takagi M. 2011. Arabidopsis HsfB1 and HsfB2b Act as Repressors of the Expression of Heat-Inducible Hsfs But Positively Regulate the Acquired Thermotolerance. *Plant Physiology* 157:1243-1254. 10.1104/pp.111.179036

Jin GH, Gho HJ, and Jung KH. 2013. A systematic view of rice heat shock transcription factor family using phylogenomic analysis. *J Plant Physiol* 170:321-329. 10.1016/j.jplph.2012.09.008

Le Hir H, Nott A, and Moore MJ. 2003. How introns influence and enhance eukaryotic gene expression. *Trends in Biochemical Sciences* 28:215-220. 10.1016/S0968-0004(03)00052-5

Letunic I, Doerks T, and Bork P. 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* 40:D302-305. 10.1093/nar/gkr931

Li PS, Yu TF, He GH, Chen M, Zhou YB, Chai SC, Xu ZS, and Ma YZ. 2014. Genome-wide analysis of the Hsf family in soybean and functional identification of GmHsf-34 involvement in drought and heat stresses. *BMC Genomics* 15:1009. 10.1186/1471-2164-15-1009

Li W, Cui X, Meng ZL, Huang XH, Xie Q, Wu H, Jin HL, Zhang DB, and Liang WQ. 2012. Transcriptional rRegulation of Arabidopsis MIR168a and ARGONAUTE1 hHomeostasis in aAbscisic aAcid and aAbiotic Sstress Rresponses. *Plant Physiology* 158:1279-1292. 10.1104/pp.111.188789

Li YJ, Chen DM, Luo SW, Zhu YY, Jia XJ, Duan YB, and Zhou MG. 2019. Intron-mediated regulation of beta-tubulin genes expression affects the sensitivity to carbendazim in *Fusarium graminearum*. *Current Genetics* 65:1057-1069. 10.1007/s00294-019-00960-4

Lin Y, Cheng Y, Jin J, Jin X, Jiang H, Yan H, and Cheng B. 2014. Genome duplication and gene loss affect the evolution of heat shock transcription factor genes in legumes. *PLoS One* 9:e102825. 10.1371/journal.pone.0102825

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

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

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 tTranscriptional rRegulation of HsfA2 fFunction as kKey rRegulators for the Hsf sSignaling nNetwork in rResponse to eEnvironmental sStress. *Plant and Cell Physiology* 52:933-945. 10.1093/pcp/pcr045

Ohama N, Sato H, Shinozaki K, and Yamaguchi-Shinozaki K. 2017. Transcriptional Rregulatory Nnetwork of Pplant Hheat Sstress Rresponse. *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 Cchallenges from the Cchanging eEnvironment. *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

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

Su ZX, Wang LL, Li WM, Zhao LH, Huang XY, Azam SM, and Qin Y. 2017. Genome-Wwide iidentification

Formatted: Font: Italic

of ~~A~~auxin ~~R~~response ~~F~~factor (ARF) ~~G~~genes ~~F~~family and its ~~T~~tissue-~~S~~specific ~~P~~prominent ~~E~~expression in ~~P~~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 50010.1038/s42003-020-01235-2

Xie LH, Li XY, Hou D, Cheng ZC, Liu J, Li J, Mu SH, and Gao J. 2019. Genome-~~W~~wide ~~A~~analysis and ~~E~~expression ~~P~~profiling of the ~~H~~heat ~~S~~shock ~~F~~factor ~~G~~gene ~~F~~family in ~~P~~phyllostachys edulis during ~~D~~development and in ~~R~~response to a~~A~~biotic ~~S~~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 ~~S~~signaling and ~~R~~responses in ~~P~~plants. *Cell* 167:313-324. 10.1016/j.cell.2016.08.029

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic