

Genome-wide analysis of basic helix-loop-helix transcription factors in papaya (*Carica papaya* L.)

Min Yang, Chenping Zhou, Hu Yang, Ruibin Kuang, Bingxiong Huang and Yuerong Wei

Guangdong Academy of Agricultural Sciences, Key Laboratory of South Subtropical Fruit Biology and Genetic Resource Utilization (MOA), Guangdong Province Key Laboratory of Tropical and Subtropical Fruit Tree Research, Institute of Fruit Tree Research, Guangzhou, China

ABSTRACT

The basic helix-loop-helix (bHLH) transcription factors (TFs) have been identified and functionally characterized in many plants. However, no comprehensive analysis of the bHLH family in papaya (*Carica papaya* L.) has been reported previously. Here, a total of 73 *CpbHLHs* were identified in papaya, and these genes were classified into 18 subfamilies based on phylogenetic analysis. Almost all of the *CpbHLHs* in the same subfamily shared similar gene structures and protein motifs according to analysis of exon/intron organizations and motif compositions. The number of exons in *CpbHLHs* varied from one to 10 with an average of five. The amino acid sequences of the bHLH domains were quite conservative, especially Leu-27 and Leu-63. Promoter *cis*-element analysis revealed that most of the *CpbHLHs* contained *cis*-elements that can respond to various biotic/abiotic stress-related events. Gene ontology (GO) analysis revealed that *CpbHLHs* mainly functions in protein dimerization activity and DNA-binding, and most *CpbHLHs* were predicted to localize in the nucleus. Abiotic stress treatment and quantitative real-time PCR (qRT-PCR) revealed some important candidate *CpbHLHs* that might be responsible for abiotic stress responses in papaya. These findings would lay a foundation for further investigate of the molecular functions of *CpbHLHs*.

Submitted 14 February 2020

Accepted 17 May 2020

Published 3 July 2020

Corresponding author

Yuerong Wei, weiyuerong@gdaas.cn

Academic editor

Yuriy Orlov

Additional Information and
Declarations can be found on
page 18

DOI 10.7717/peerj.9319

© Copyright
2020 Yang et al.

Distributed under
Creative Commons CC-BY 4.0

Subjects Agricultural Science, Bioinformatics, Genomics, Plant Science

Keywords Papaya, Genome-wide analysis, bHLH transcription factors, Abiotic stress, Expression analysis

INTRODUCTION

Since plants are unable to move, plant growth and development are regularly affected by abiotic and biotic stresses, which impair yields and result in losses to farmers. For better growth and development, plants have to make use of a series of physiological and biochemical processes in their responses to multiple abiotic stresses by regulating gene expression (Agarwal et al., 2006; Feller et al., 2011; Pires & Dolan, 2010). Previous studies have demonstrated that these physiological and biochemical mechanisms are more likely to be a polygenic cooperative defense response induced by various stresses, rather than the single response of a single gene (Zhang, Creelman & Zhu, 2004). Therefore, the traditional method of obtaining the stress tolerance of plants by modification a single

OPEN ACCESS

resistance/sensitive gene is limited. Comprehensive analysis of important gene families are very important for molecular breeding.

As an important and popular fruit, papaya is famous for its high nutritional and medical values. Papaya is widely grown in southern China, the tropics and subtropics areas, and its demand is increasing every year. However, the production and quality of papaya were often threatened by various abiotic stresses, such as salt, drought, and cold. These stresses often cause severe economic losses in papaya production in China. Therefore, it is very important to study the functions of gene families that involved in abiotic stresses response in papaya. Since obtaining the whole genome sequences of papaya (Ming et al., 2008), several important gene families have been identified by the tool of genome-wide analysis in papaya, including *Aux/IAA* gene family, ARF family, *SQUAMOSA promoter binding protein-like (SPL)* gene family, NBS resistance gene family and NPR1 family. These families were essential for papaya fruit ripening, flower and fruit development, fitness and disease resistance (Liu et al., 2017a; Liu et al., 2015; Peraza-Echeverria, Santamaría & Fuentes, 2012; Porter et al., 2009; Xu et al., 2020).

In various stresses regulation network and signaling pathways, transcription factors (TFs) are important proteins that regulate gene expression by activating and repressing related downstream genes. Among them, WRKY and bHLH families are the most common TF families in higher plants (Kosugi & Ohashi, 2002). The WRKY transcription factors has been reported to be related to abiotic and biotic stresses responses in papaya (Pan & Jiang, 2014). Basic/helix-loop-helix (bHLH) TFs are widely found in almost all eukaryotes and are the second largest TFs family in plants (Carretero-Paulet et al., 2010; Feller et al., 2011; Jones, 2004; Pires & Dolan, 2010). The bHLH superfamily proteins are defined by one highly conserved bHLH domain, which comprises approximately 60 amino acids in length and contains two different functional regions: the basic region and the HLH region (Li et al., 2006; Toledo-Ortiz, Huq & Quail, 2003). The basic region is located at the N-terminal end of the bHLH domain and consists approximately 15 amino acids. It is a DNA-binding region that enables bHLH TFs to bind to a specific E-boxes (CANNTG) (Atchley & Fitch, 1997; Atchley, Terhalle & Dress, 1999). The HLH region, at the C-terminal end, is mainly composed of hydrophobic residues, containing two amphipathic α -helices linked by a loop region that has variable sequences and acts as a dimerization domain (Heim et al., 2003; Li et al., 2006). Outside of the two conserved regions, the rest of the bHLH protein sequences are usually very different (Morgenstern & Atchley, 1999).

In animals, the bHLH TFs can be divided into six main groups (designated A to F) based on phylogenetic analysis, functional properties and DNA-binding specificity (Atchley & Fitch, 1997). These bHLH groups can be divided into several small subfamilies (Ledent & Vervoort, 2001; Simionato et al., 2007). The bHLHs mainly function in sensing the external environment, cell cycle regulation and tissue differentiation (Amoutzias, Robertson & Bornberg-Bauer, 2004; Atchley & Fitch, 1997; Stevens, Roalson & Skinner, 2008; Vervoort & Ledent, 2001). Compared to animals, the research on bHLH proteins in plants is limited, even the exact number subfamilies of bHLH TFs has not been determined. Generally, the bHLH proteins is thought to cover 15–25 subfamilies (Buck & Atchley, 2003; Pires & Dolan, 2010), but some atypical bHLHs have extended the number to 32 based

on phylogenetic analysis in plants (Carretero-Paulet et al., 2010). With the availability of genome sequence data and the rapid development of molecular biology, increasing numbers of bHLH subfamily genes have been identified and characterized in a wide range of plant species, including *Arabidopsis* (Toledo-Ortiz, Huq & Quail, 2003), peanut (Gao et al., 2017), apple (Mao et al., 2017), tomato (Sun, Fan & Ling, 2015), potato (Wang et al., 2018b), peach (Zhang et al., 2018), grapes (Wang et al., 2018a), sweet orange (Geng & Liu, 2018), and bamboo (Cheng et al., 2018). The results from these research have shown that bHLH TFs have versatile biological functions, such as regulating light morphogenesis (Leivar et al., 2008; Roig-Villanova et al., 2007), hormone signals (Friedrichsen et al., 2002; Lee et al., 2006), the developmental of root (Feng et al., 2017) and anther (Farquharson, 2016), regulating epidermal cell fate determination (Bernhardt et al., 2003), participating in various biotic and abiotic stress responses (Jiang, Yang & Deyholos, 2009; Liu et al., 2014; Wang et al., 2018b), etc.

In recent years, some studies demonstrated that bHLH transcription factors play important roles in the stress-related regulation network and signaling pathways in many species. However, no systematic analysis of the bHLH TFs have previously been performed in papaya. In this study, a total of 73 *CpbHLH* genes were identified in papaya, and phylogenetic analyses were carried out to analyze the relationships among these genes. Meanwhile, gene structure, protein physicochemical properties and conserved motifs, the *cis*-element of the promoter region, and gene ontology (GO) analysis were investigated. Furthermore, to analyze the functions of *CpbHLH*s responsible for responding to abiotic stresses, the expression profiles of 22 selected genes under salt, drought, ABA and cold stresses were investigated by using quantitative real-time PCR (qRT-PCR). We identified several important candidate genes that might be responsible for abiotic stress responses. We completed the first comprehensive genome-wide analysis of the *bHLH* gene family in papaya, and our results provide information necessary for further functional research of the bHLH family in papaya.

MATERIALS & METHODS

Identification of *CpbHLH* genes, gene structure and physicochemical analysis

Papaya (*Carica papaya* L.) bHLH protein sequences were downloaded from the Plant TFDB V4.0 database (Jin et al., 2017). Furthermore, we used the SMART online software (<http://smart.embl-heidelberg.de/>) and the InterProScan tool (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>) to identify integrated bHLH domains in putative papaya bHLH proteins. The physicochemical properties of *CpbHLH* proteins were predicted by ProPAS (Wu & Zhu, 2012). The genomic sequences, ID numbers and coding sequences (CDS) corresponding to each predicted *CpbHLH* gene were obtained from the Phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>). The intron numbers, exon–intron organizations and locations of the *CpbHLH* genes were analyzed by Gene Structure Display Server (GSDS) v2.0 (Hu et al., 2015).

Phylogenetic tree building, motif identification and multiple sequence alignment

To research the phylogenetic relationship of *CpbHLH* proteins, protein sequences of papaya were pre-aligned using HMM align (Eddy, 1998) and the pHMM HLH ls.hmm from PFAM (<https://pfam.xfam.org/family/PF00010>) to identify the domains of bHLH TFs. Based on the manually aligned bHLH region of 158 bHLH proteins from *Arabidopsis* and 173 from rice (Pires & Dolan, 2010), the identified bHLH domains were later aligned using MAFFT v7.305b (Kaotoh, Kuma & Miyata, 2002) with default settings. Phylogenetic tree was constructed based on the neighbor-joining method using FastTree v2.1.11 (Price, Dehal & Arkin, 2009) with default settings. Bootstrapping with 1000 replicates was used to assess the statistical reliability of nodes in the tree. Multiple sequence alignment based on protein sequences of these 73 *CpbHLH* TFs was generated by MAFFT v7.305b (Kaotoh, Kuma & Miyata, 2002) with default settings.

To identify the conserved motifs among the *CpbHLH* proteins, we uploaded the 73 amino acid sequences of the *CpbHLH* family to the Multiple EM for Motif Elicitation (MEME, version 5.1.1) (<http://meme-suite.org/tools/meme>). The parameter settings were as follows: zero or one, occurrence of a single motif per sequence; 3, maximum number of motifs found. All other parameters were set to the default values.

Promoter *cis*-acting regulatory element analysis and gene ontology (GO) annotation

To predict and compare the putative promoter *cis*-elements of *bHLHs* in papaya and *Arabidopsis*, the upstream 2,000 bp genomic DNA sequences of 73 *CpbHLH* genes in papaya, and 47 *AtbHLH* genes in *Arabidopsis* (the putative orthologous genes corresponding to *CpbHLHs*) were downloaded and then submitted to the PlantCARE (Magali et al., 2002). The GO annotations of papaya bHLH proteins were downloaded from AgriGOv2.0 (http://systemsbiology.cau.edu.cn/agriGOv2/download/508_slimGO). For GO enrichment analysis, the full-length protein sequences of papaya bHLH were blasted against *Arabidopsis* proteins with default parameters. The best hits were submitted to AgriGOv2.0 for GO enrichment analysis (Tian et al., 2017). Fisher's exact test was used to evaluate enriched GO terms. GO terms include three aspects: biological process, cellular component and molecular function.

Plant materials, growth conditions and stress treatments

In this experiment, stems with axillary buds were selected as explants from two-year-old 'Yi Chi Gua' papaya trees grown under standard field conditions in the Institute of Fruit Tree Research, Guangdong Academy of Agriculture Science, Guangzhou, China, and cultured in vitro to obtain the complete papaya seedlings with normal leaves and roots using tissue culture techniques. Healthy and uniform papaya seedlings were used for different treatments. For the selection of stress conditions for papaya, we designed different gradients of stress conditions for pre-experiments: the concentration gradients of salt stress are 100 mM NaCl, 200 mM NaCl, and 300 mM NaCl; the concentration gradients of PEG6000 (to mimic drought stress) are 15% PEG6000, 20% PEG6000, 25% PEG6000 and 30% PEG6000; the concentration gradients of ABA are 50 μ M ABA, 100 μ M

ABA and 150 μM ABA; the temperature gradients are 0 $^{\circ}\text{C}$, 4 $^{\circ}\text{C}$ and 10 $^{\circ}\text{C}$, and finally determined the suitable stress conditions used in this manuscript. For salt, drought and ABA stresses, seedlings were treated with MS liquid medium containing 200 mM NaCl, 25% PEG6000 and 100 μM ABA for 2 h respectively, and then the roots were collected. For cold treatment, seedlings were subjected to 4 $^{\circ}\text{C}$ for 2 h and the leaves were collected. All of the collected materials were immediately frozen in liquid nitrogen and stored at -80°C for RNA isolation. Untreated seedlings were used as the control groups. Three biological replications were carried out for each treatment.

RNA extraction and quantitative real-time PCR (qRT-PCR) analysis

Total RNA from papaya after different treatments was isolated using TRIzol reagent (Invitrogen). The extracted RNA was treatment with DNase (TaKaRa), and then reverse transcribed into cDNA using the PrimeScriptTM RT Reagent Kit (TaKaRa). The qRT-PCR was conducted on the ABI StepOne Real Time PCR system using 2X SG Fast qPCR Master Mix (High Rox) (TaKaRa) according to the manufacturer's instructions. TATA binding protein 2 (*CpTBP2*) amplification was used as an internal control (Zhu *et al.*, 2012). The qRT-PCR reactions used three biological replicates, and each biological repeat had three technical replicates. Gene-specific primers for qRT-PCR of the 22 *CpbHLH* genes were designed based on the CDSs of the *CpbHLH* genes using Primer Premier 5.0 (Data S1). The relative expression levels of each gene were calculated using the $2^{-\Delta\Delta\text{CT}}$ method, the raw data was showed in File S1.

RESULTS

Identification and characterization of *CpbHLH*s

A total of 105 putative bHLH transcription factors of papaya (*C. papaya*) were downloaded from the PlantTFDBv2.0 (<http://planttfdb.cbi.pku.edu.cn/>). To verify the reliability of these results, the 105 *Cp* bHLH proteins sequences were filtered by Interproscan and SMART domain annotation, and a total of 73 predicted *CpbHLH* proteins were identified. They were named *CpbHLH001* to *CpbHLH073* at random except for 32 proteins that were explicitly excluded by Interproscan and SMART (Table S1). The detailed information on these predicted *Cp* bHLHs, including protein ID, locus ID, opening reading frame (ORF) lengths, amino acid sequences/lengths, molecular weight, isoelectric point and exon/intron numbers, are listed in Data S2. In previous studies, 129/132, 188, 159, 147, 124, 95, 94 and 56 *bHLH* genes were identified in peanut (Gao *et al.*, 2017), apple (Mao *et al.*, 2017), tomato (Sun, Fan & Ling, 2015), *Arabidopsis* (Toledo-Ortiz, Huq & Quail, 2003), potato (Wang *et al.*, 2018b), peach (Zhang *et al.*, 2018), grapes (Wang *et al.*, 2018a) and sweet orange (Geng & Liu, 2018), respectively. Compared with the above dicotyledonous plants, the density of *bHLH*s genes in papaya genome was about 0.26%, which is lower than the density of peanut (Gao *et al.*, 2017), apple (Mao *et al.*, 2017), tomato (Sun, Fan & Ling, 2015), *Arabidopsis* (Toledo-Ortiz, Huq & Quail, 2003), potato (Wang *et al.*, 2018b) and sweet orange (Geng & Liu, 2018), and similar to peach (Zhang *et al.*, 2018) and wine grapes (Wang *et al.*, 2018a) (Table 1). This is probably associated with the whole-genome duplications during evolution. Among the above plants, some plants with recent whole-gene duplication like

Table 1 Summary of TFs identified from dicotyledonous plant species with genome sequences.

Plant species	Common name	bHLH	Proteins	Ratio (%)
<i>C. papaya</i>	Papaya	73	27829	0.26
<i>A. ipaensis/A. duranensis</i>	Peanut	129/132	7243	1.78/1.82
<i>Malus x domestica</i>	Apple	188	15173	1.24
<i>S. lycopersicum</i>	Tomato	159	15722	1.01
<i>A. thaliana</i>	Arabidopsis	147	32125	0.46
<i>S. tuberosum</i>	Potato	124	17445	0.71
<i>P. persica</i>	Peach	95	28299	0.34
<i>V. vinifera</i>	Wine Grapes	94	47097	0.20
<i>C. sinensis</i>	Valencia Orange	56	13522	0.41

peanut, apple, tomato, Arabidopsis, potato and sweet orange while the plants without whole-gene duplication like papaya, peach and wine grapes.

To further characterize the bHLHs in papaya, the physicochemical properties of these putative proteins were analyzed and are shown in [Data S2](#). The size of deduced CpbHLHs ranged from 100 (CpbHLH053) to 679 (CpbHLH068) amino acids, the corresponding molecular weights from 11.525 KDa to 75.899 KDa. The predicted theoretical isoelectric points (PI) values of CpbHLHs were between 4.71(CpbHLH028) and 11.07(CpbHLH003). Similar molecular weights and isoelectric points have been made in potato ([Wang et al., 2018b](#)). And all of predicted CpbHLH proteins were hydrophilic characteristic proteins, the grand average of hydropathy values were negative, ranging from -0.2098 (CpbHLH033) to -1.0125 (CpbHLH006). Similar result has been made in *Brachypodium distachyon* ([Niu et al., 2017](#)). That is, the predicted CpbHLH proteins showed diversities in their length, molecular weight, PI and the grand average of hydropathy values.

Phylogenetic analysis, gene structure, conserved motifs analysis and multiple sequence alignment of CpbHLHs

To evaluate the evolutionary relationships of the CpbHLH proteins, a neighbor-joining phylogenetic tree was generated using conserved bHLH domains from papaya, *Arabidopsis* and rice. The phylogenetic tree showed that the 73 CpbHLH members were clustered into 18 subfamilies with one orphan ([Fig. 1A](#) and [Data S3](#)), consistent with the earlier results showing that the bHLH subfamily in plants can be divided into 15–25 subfamilies ([Pires & Dolan, 2010](#)). Previous research have named the bHLH subfamilies using English letters ([Li et al., 2006](#); [Mao et al., 2017](#)), Roman numerals ([Song et al., 2017](#); [Sun, Fan & Ling, 2015](#)), or Arabic numerals ([Chen et al., 2015](#); [Toledo-Ortiz, Huq & Quail, 2003](#)), In this study, we named CpbHLH subfamilies using Roman numerals. As shown in [Fig. 1](#), the subfamily XII was the largest subfamily among all three species, and all of subfamilies include at least two species. In papaya, none of the bHLHs were grouped into IV d, II, X V, X, XI V and XIII subfamilies compared to rice and *Arabidopsis*, which may be due to these bHLHs were lost during the process of evolution.

Exon/intron organization, as a type of structural divergence, plays an important role in the evolution of multiple gene families ([Xu et al., 2012](#)). The annotation features of the

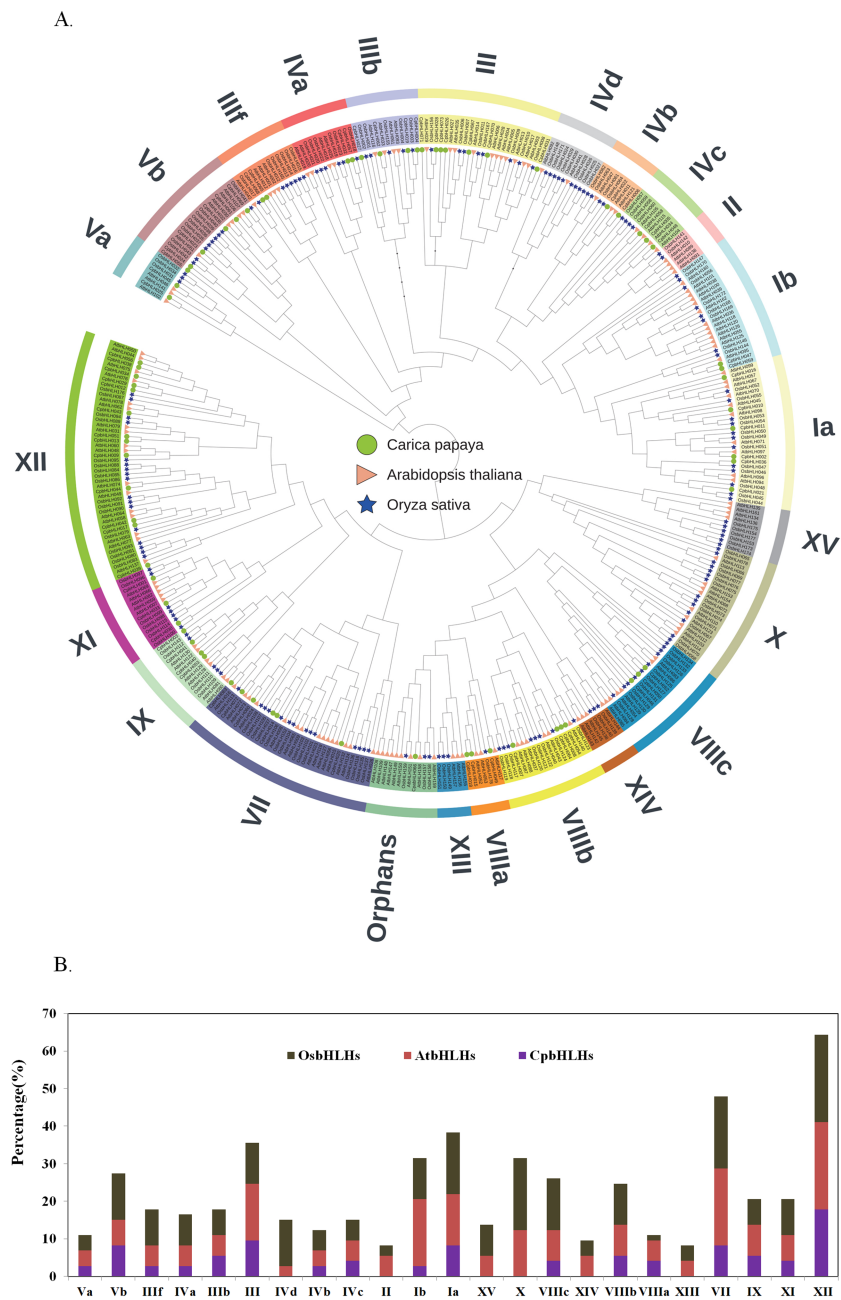


Figure 1 Phylogenetic and family members analysis of bHLHs from papaya, rice and *Arabidopsis*.

(A) The 73 *CpbHLH*s are clustered into 18 subfamilies. Phylogenetic tree was constructed based on the neighbor-joining method. Bootstrapping with 1,000 replicates was used to assess the statistical reliability of nodes in the tree. (B) Comparison of bHLH family members from papaya, rice and *Arabidopsis*. Different colors represent the different plants. Black: *OsbHLH*s, red: *ArbHLH*s, purple: *CpbHLH*s.

Full-size DOI: 10.7717/peerj.9319/fig-1

CpbHLH genes were submitted to Gene Structure Display Server (GSDS) together to show their gene structures. As described in [Data S2](#) and [Fig. 2A](#), the number of introns varied from zero to ten, representing a complex distribution pattern. Most (63 (86.3%)) of the

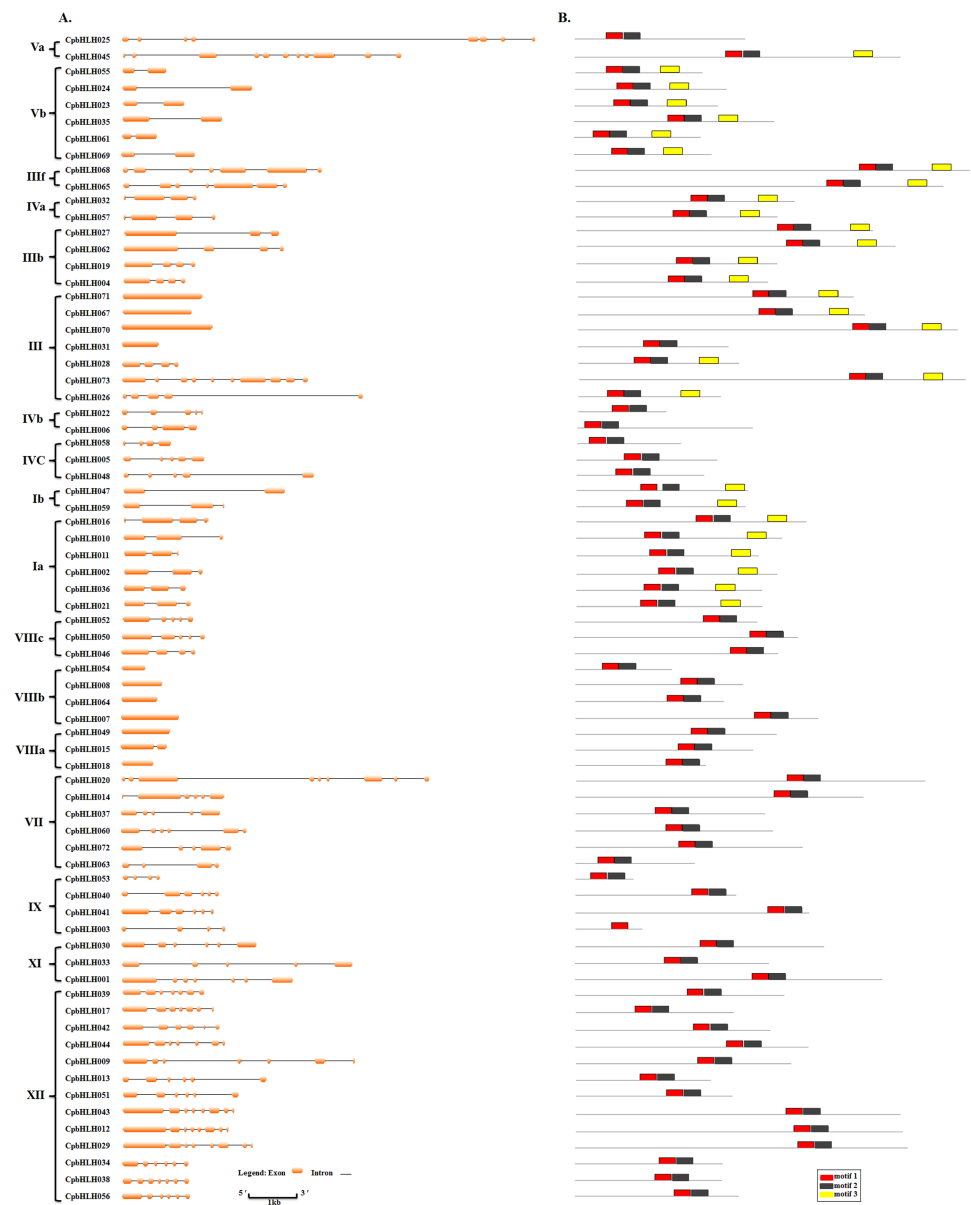


Figure 2 Gene structure and motif distribution of the papaya bHLH family. (A) Exon-intron organization of *CpbHLH* genes. Exons and introns are presented as filled orange lines and thin black single lines, respectively. The brackets and Roman numerals separate each subfamily and clearly present the member conservation of each subfamily. (B) Arrangements of conserved motifs in 73 *CpbHLH* proteins. Three predicted motifs are represented by different colored boxes, motif 1 (red block), motif 2 (black block) and motif 3 (yellow block).

Full-size DOI: 10.7717/peerj.9319/fig-2

CpbHLHs were found to possess introns among the 73 *CpbHLH* genes, while 10 (13.7%) of the genes were intron-less, 8 (11.0%) genes contained one intron, and the remaining genes had two or more introns. In addition, members of the same subfamily also displayed similar intron distribution patterns in view of the full-length genome sequences. For instance, all of the *CpbHLHs* in subfamily V b had one intron and two exons, the whole members of

subfamily III f had six introns and seven exons, the IV a subfamily members showed three introns and four exons, and all members of VIII b subfamily consisted only one exon.

Most importantly, members of the same bHLHs subfamily are usually participated in the same signaling pathway or biological process, and the functions of these members are often partially or totally redundant (*Pires & Dolan, 2010*). For example, *AtbHLH10*, *AtbHLH89* and *AtbHLH91*, corresponding rice orthologs *OsbHLH141*, *OsbHLH142* are members of subfamily II, they are all involved in the process of pollen development (*Li et al., 2006; Liu et al., 2017; Zhu et al., 2015*). Especially in *Arabidopsis*, there is no obvious phenotype in single mutant of *AtbHLH10*, *AtbHLH89* or *AtbHLH91*, only their various double or triple mutants showed the phenotype of pollen development deficiency (*Liu et al., 2017*). In subfamily III b, *OsbHLH001* (*OsICE2*), *OsbHLH002* (*OsICE1*), *CpbHLH027*, *CpbHLH062*, *AtbHLH116* (*ICE1*) and *AtbHLH33* (*ICE2*) were clustered within one clade. In previous studies, *AtbHLH116*(*ICE1*) and *AtbHLH33*(*ICE2*) and corresponding orthologs in rice (*Os bHLH001/OsICE2*, *OsbHLH002/OsICE1*) have been reported to function in the stress of chilling (*Chinnusamy et al., 2003; Deng et al., 2017; Fursova, Pogorelko & Tarasov, 2009; Li et al., 2010; Zhang et al., 2017*). And we also found transcripts of *CpbHLH027* and *CpbHLH062* were increased under chilling stress in this study, implying that *CpbHLH027* and *CpbHLH062* are involved in the process of chilling stress in papaya.

To further study the sequence characteristics of the predicted bHLH domains at the amino acid level, we carried out a multiple sequence alignment of the 73 predicted *CpbHLH* protein sequences (*Fig. 3*). The result showed that the 73 putative *CpbHLH* proteins contained two conserved regions in the bHLH domains: the basic region plus helix 1 and the loop region plus helix 2 (*Fig. 3* and *Table S2*). Additionally, we used the online MEME program to identify the conserved motifs (*Bailey & Elkan, 1994*). The result also showed that most of the sequences (exclude *CpbHLH003*) exhibited two highly conserved motifs: one is contains 29 amino acids, and the other consists of 21 amino acids, are shown in red and black blocks, respectively (*Fig. 2B*). Among the two motifs, motif 1 comprises basic residues and helix 1, and motif 2 comprises a loop and helix 2. And the space between motif 1 and 2 consists of a loop, which is variable in length in some bHLH proteins. The sequence logos of motif 1 (in red) and motif 2 (in black) are shown in *Fig. 4A*. The backbones of motif 1 and 2 are also conserved in most plant species (*Guo & Wang, 2017; Heim et al., 2003; Sun, Fan & Ling, 2015*), and these highly conserved residues in bHLH domains may be responsible for protein dimerization (*Heim et al., 2003*).

Besides these two common conserved motifs, some *CpbHLHs* that are mainly distributed into eight subfamilies (including V a, V b, III f, IV a, III b, III, I b and I a subfamilies) harbor another conserved motif (motif 3) with a length of 36 amino acids. The motif 3 is indicated by the yellow blocks and the sequence logo is visualized as logo3 (*Fig. 2B* and *Fig. 4B*). This result is accord with the previous studies that members of a given subfamily exhibited another conserved nonbHLH motif (motif 3) in plant bHLH superfamily (*Pires & Dolan, 2010*). However, in papaya, members of the bHLH proteins have the same motif that is distributed into eight subfamilies, not just one subfamily. In addition, among the 73 *CpbHLHs*, one atypical bHLH protein (*Cp bHLH003*) exhibited incomplete bHLH domains, whereas the remaining 72 *CpbHLH* proteins all presented complete bHLH

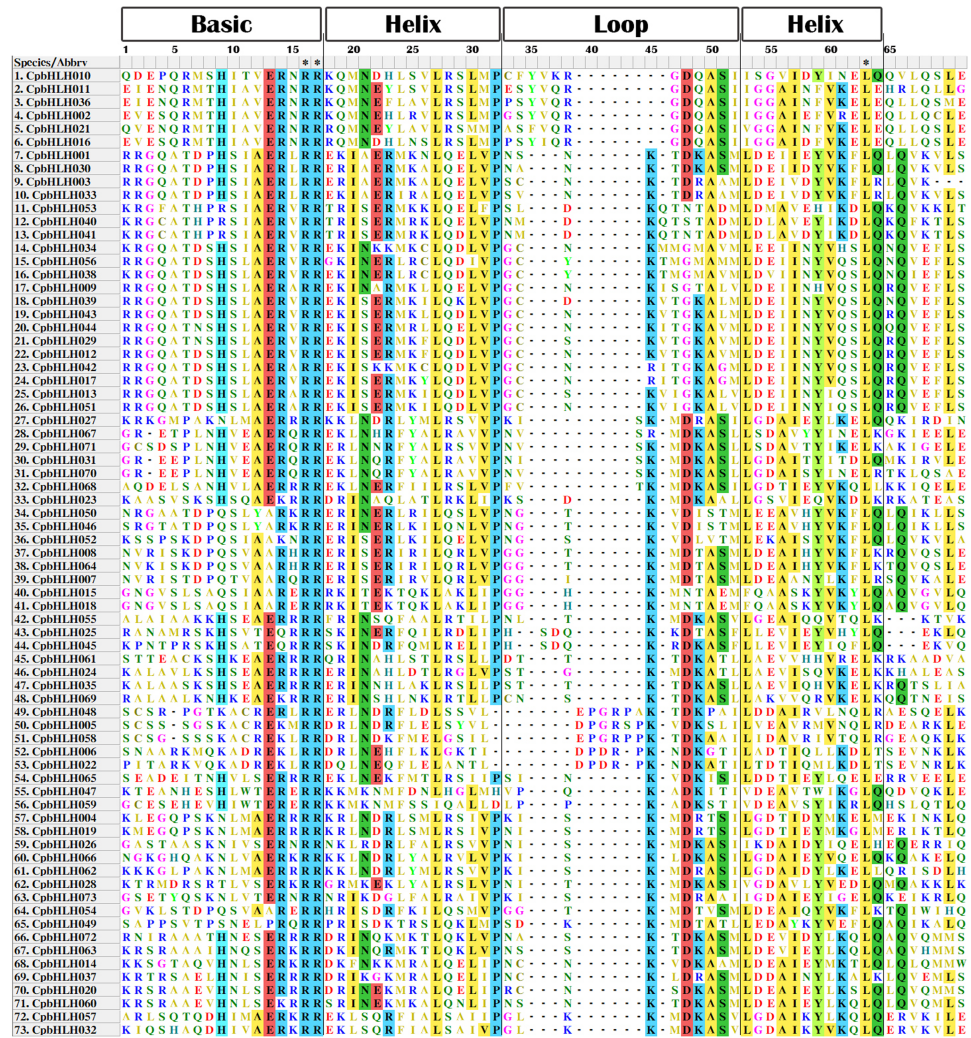


Figure 3 Multiple sequence alignment of the bHLH domains in papaya. Amino acids with more than 50% identity are labeled with colored boxes.

Full-size DOI: 10.7717/peerj.9319/fig-3

domains. Similar observations have been made in other plant species, such as peach and blueberry (Song et al., 2017; Zhang et al., 2018).

Promoter analysis of *bHLH* genes in papaya

To further understand *CpbHLHs* functions and regulation patterns, *cis*-elements in *CpbHLH* genes promoter sequences were investigated. Regions of 2,000 bp upstream from the start codons of each *CpbHLH* gene were analyzed using PlantCARE. The results showed that the *cis*-elements could be divided into three main categories (Fig. 5A and Data S4). Category one contained a ubiquitous class of plant light responsive elements among which G-Box, G-box, GT1-motif and Box 4 were common in the *CpbHLH* promoters. Category two contained important elements that were involved in the process of stress-responsiveness, including MYB binding site involved in drought-inducibility

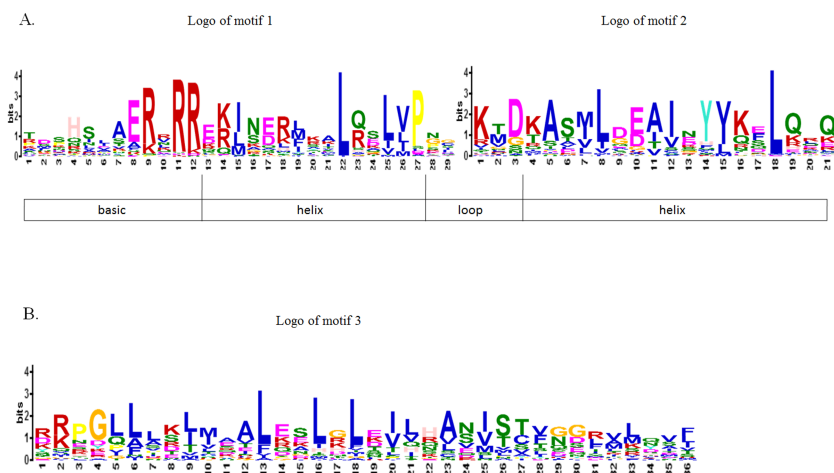


Figure 4 Motif composition and logos of papaya bHLH proteins. (A) The logos of motif 1 and 2, which together constitute the bHLH domain in papaya. The overall height of the character represents the conservation of an amino acid at the specific position. Each color of the English letters represents a type of amino acid residue. (B) The logo of motif 3, which is another conserved motif.

Full-size [DOI: 10.7717/peerj.9319/fig-4](https://doi.org/10.7717/peerj.9319/fig-4)

(MBS), low temperature response elements (LTR), defense and stress responsive elements (TC-rich) and wound-responsive elements (WUN-motifs). In addition, more than ten kinds of hormone-responsive *cis*-elements were identified (e.g., gibberellin-GA, auxin-IAA, methyl jasmonate-MeJA, salicylic acid-SA, and abscisic acid-ABA). Among them, the most common response elements were ABA (ABRE), MeJA (CGTCA-motif and TGACG-motif) and SA (TCA-element and SARE), which included 158 (29.15%), 128 (23.62%) and 53 (9.78%), respectively (Fig. 5B). Category three contained plant growth and development elements, such as anaerobic induction elements (ARE), O₂-site, CAT-box and so on. Additionally, we also analyze the *cis*-elements in the promoter regions of putative orthologous genes that corresponding to *CpbHLH* in *Arabidopsis*, and the similar result has been obtained in *Arabidopsis* (Fig. S1 and Data S4). There also existed three main categories: plant light, abiotic and biotic stresses and plant growth and development responsive elements. And the percentage of most stress-responsive elements in *Arabidopsis* were similar to papaya, including ABA responsive elements, drought-responsive elements, wound-responsive elements, low temperature-responsive elements and IAA responsive elements, implying that most of the promoter *cis*-elements of bHLH family were conserved in *Arabidopsis* and papaya.

GO annotation of *CpbHLH* proteins

To understand the functions of papaya bHLHs, we performed a GO annotation of *CpbHLH*s, and the results are shown in Data S5. A total of 70 *CpbHLH*s were involved in protein dimerization activity (GO: 0046983). The result is consistent with the earlier studies, which show that the HLH domain was necessary for protein dimerization and DNA binding (Murre, Mccaw & Baltimore, 1989). Some conserved amino acid residues are important to the function of bHLH proteins, especially the Leu-27 in helix 1 and the Leu-73 in helix

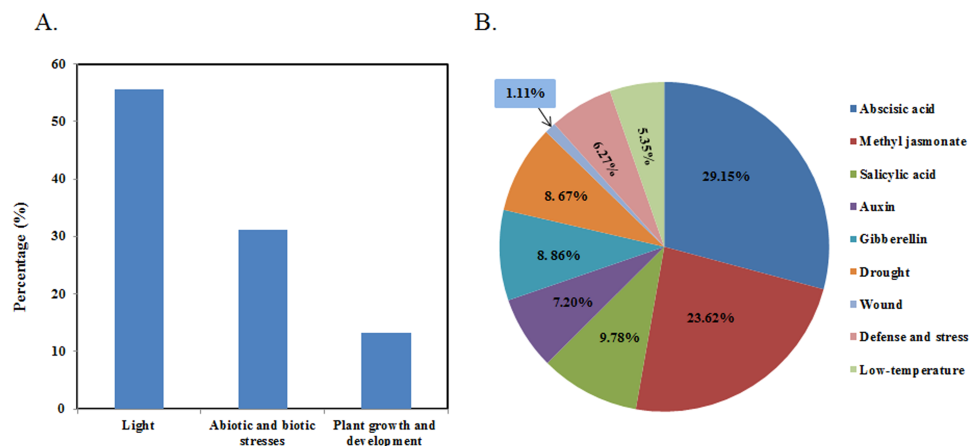


Figure 5 Cis-acting element analysis of the promoter of *bHLH* genes in papaya. (A) Percentage of total *cis*-acting elements in the promoter region of *CpbHLH* genes. (B) The percentage of each *cis*-acting element in the abiotic and biotic stresses categories. Abscisic acid (29.15%), methyl jasmonate (23.62%), salicylic acid (9.78%), auxin (7.20%), gibberellin (8.86%), drought (8.67%), wound (1.11%), defense and stress (6.27%), low temperature (5.35%).

Full-size DOI: 10.7717/peerj.9319/fig-5

2 (*Carretero-Paulet et al., 2010*). In this study, we found 72 (out of 73) *CpbHLH* proteins have Leu-27 (corresponding to Leu-27 in *AtbHLHs*), and all of the *CpbHLH* proteins have Leu-63 (corresponding to Leu-73 in *AtbHLHs*) (*Fig. 3* and *Table S2*). Because of a lack of reported experimental data and databases, we used *Arabidopsis* as the reference species to perform a GO enrichment analysis of *CpbHLH* proteins, and 54 of 73 predicted *CpbHLH* proteins were obtained with results compared to *Arabidopsis*. We summarized the results in *Fig. 6* and *Data S6*. The majority of predicted *CpbHLH* proteins were enriched in DNA binding. Almost all of the predicted *CpbHLH* proteins (37, 68.5%) were predicted to localize in the nucleus, whereas the remaining predicted *CpbHLH* proteins were located in other organelles, including plastids, the cytoplasm, and chloroplasts. Additionally, some predicted *CpbHLH* proteins existed in multiple cellular components. For example, *CpbHLH013* was located in three cellular components: chloroplasts, part of the cytoplasm, and the nucleus, which may reflect its multiple functions in various biological processes. The metabolic processes involved the greatest number of putative *CpbHLH* proteins (47, 87.0%). Biosynthetic processes and gene expression involved the second greatest number of putative *CpbHLH* proteins (46, 85.2%). In addition, *CpbHLH* proteins could respond to stimulus, morphogenesis, cell differentiation, and developmental process.

Expression analysis of bHLH superfamily genes under different abiotic stresses

The bHLH proteins have been characterized functionally in many plants with a vital role in the regulation of diverse biological processes, but little is known about their role in papaya. To analyze the functions of *CpbHLHs* responding to abiotic stresses, the expression profiles of 22 selected genes under salt, drought, ABA and cold stresses were investigated by using qRT-PCR (*Table 2* and *Fig. 7*). The results showed that 4 (*CpbHLH011*, *CpbHLH022*,

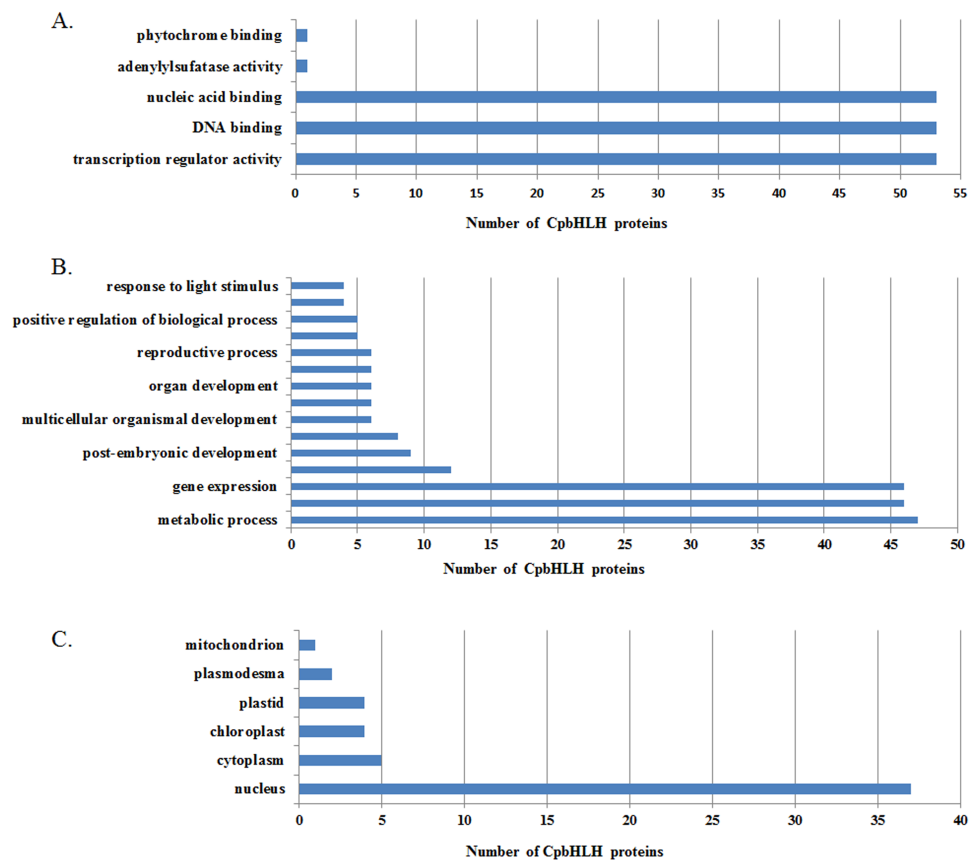


Figure 6 Gene ontology (GO) annotation of *Cp bHLH* proteins. The annotation was performed on three categories: (A) molecular function, (B) biological processes and (C) cellular components.

Full-size DOI: [10.7717/peerj.9319/fig-6](https://doi.org/10.7717/peerj.9319/fig-6)

CpbHLH027 and *CpbHLH056*) of 22 *CpbHLH* mRNAs were increased, and 3 *CpbHLH* (*CpbHLH020*, *CpbHLH053* and *CpbHLH062*) mRNAs were reduced more than 2-fold in salt (200 mM NaCl) treated papaya seedlings. Under drought stress (25% PEG), 8 (*CpbHLH011*, *CpbHLH022*, *CpbHLH027*, *CpbHLH046*, *CpbHLH050*, *CpbHLH052*, *CpbHLH056* and *CpbHLH068*) of 22 *CpbHLH* mRNAs were upregulated, and 3 *CpbHLH* (*CpbHLH020*, *CpbHLH042* and *CpbHLH053*) mRNAs were downregulated more than 2-fold. Under ABA treatment (100 μ M), 3 (*CpbHLH027*, *CpbHLH052* and *CpbHLH056*) of 22 *CpbHLH* mRNAs were upregulated, and 5 *CpbHLH* (*CpbHLH019*, *CpbHLH020*, *CpbHLH042*, *CpbHLH053* and *CpbHLH062*) mRNAs were downregulated. Under cold stress (4 $^{\circ}$ C), there were 4 *CpbHLH* genes (*CpbHLH027*, *CpbHLH035*, *CpbHLH056* and *CpbHLH062*) whose expression increased more than 1.5-fold, and 4 *CpbHLH* (*CpbHLH046*, *CpbHLH050*, *CpbHLH052* and *CpbHLH068*) mRNAs were reduced more than 2-fold.

Interestingly, a few transcripts of *CpbHLH* responded to all or multiple stresses. For instance, *CpbHLH056* was sensitive to all four stresses and was upregulated distinctly under the four stresses. The orthologue of *CpbHLH056* in Arabidopsis is *BEE1* (*AtbHLH044*) (Fig. 1), which has been functionally characterized in previous reports. At low temperatures,

Table 2 Expression levels of *CpbHLH* genes under salt, drought and ABA stresses.

The name of <i>CpbHLHs</i>	CK	NaCl (200 mM)	PEG (25%)	ABA (100 μ M)
<i>CpbHLH006</i>	1	0.74 \pm 0.13	0.95 \pm 0.15	0.75 \pm 0.07
<i>CpbHLH011</i>	1	2.45 \pm 0.04	4.06 \pm 0.12	1.66 \pm 0.12
<i>CpbHLH019</i>	1	0.80 \pm 0.03	0.64 \pm 0.02	0.18 \pm 0.02
<i>CpbHLH020</i>	1	0.20 \pm 0.003	0.39 \pm 0.01	0.22 \pm 0.01
<i>CpbHLH022</i>	1	4.21 \pm 0.03	3.05 \pm 0.03	1.54 \pm 0.08
<i>CpbHLH027</i>	1	4.49 \pm 0.09	2.32 \pm 0.06	2.27 \pm 0.12
<i>CpbHLH035</i>	1	1.62 \pm 0.08	1.79 \pm 0.07	0.71 \pm 0.02
<i>CpbHLH037</i>	1	0.96 \pm 0.02	1.04 \pm 0.01	1.92 \pm 0.07
<i>CpbHLH040</i>	1	1.38 \pm 0.04	0.77 \pm 0.03	0.51 \pm 0.03
<i>CpbHLH041</i>	1	0.55 \pm 0.01	0.76 \pm 0.03	0.70 \pm 0.03
<i>CpbHLH042</i>	1	0.55 \pm 0.03	0.33 \pm 0.04	0.26 \pm 0.03
<i>CpbHLH046</i>	1	0.78 \pm 0.06	5.46 \pm 0.10	0.53 \pm 0.03
<i>CpbHLH050</i>	1	0.53 \pm 0.04	11.86 \pm 0.10	1.18 \pm 0.02
<i>CpbHLH052</i>	1	1.64 \pm 0.06	8.71 \pm 0.06	4.42 \pm 0.13
<i>CpbHLH053</i>	1	0.44 \pm 0.04	0.31 \pm 0.02	0.19 \pm 0.01
<i>CpbHLH056</i>	1	7.11 \pm 0.04	32.65 \pm 0.45	12.22 \pm 0.21
<i>CpbHLH060</i>	1	0.89 \pm 0.07	0.76 \pm 0.05	0.75 \pm 0.11
<i>CpbHLH062</i>	1	0.38 \pm 0.02	0.65 \pm 0.02	0.30 \pm 0.02
<i>CpbHLH065</i>	1	1.64 \pm 0.06	1.04 \pm 0.03	0.52 \pm 0.05
<i>CpbHLH068</i>	1	1.72 \pm 0.05	2.37 \pm 0.04	1.39 \pm 0.07
<i>CpbHLH069</i>	1	0.68 \pm 0.07	0.89 \pm 0.18	0.52 \pm 0.08
<i>CpbHLH070</i>	1	1.11 \pm 0.02	1.17 \pm 0.07	1.47 \pm 0.04

Notes.

Quantitative RT-PCR was used to investigate the expression levels (shown in fold change) of the *CpbHLHs*. The expression level in the control (CK) was set at 1.0. The means of three replicates of qRT-PCR and standard deviations (SD) values are shown.

BEE1 is a positive regulator of flavonoid accumulation (*Petridis et al., 2016*), which is consistent with our results. In addition, BEE1, BEE2 and BEE3 are functionally redundant positive regulators of BR (brassinosteroid) signaling, but these transcripts are repressed by ABA (*Friedrichsen et al., 2002*). However, we found the transcription of *CpbHLH056* was notably upregulated (>10-fold) under ABA treatment. More interestingly, *CpbHLH042*, which is an orthologue of BEE2 (*Fig. 1*), was distinctly repressed by ABA (approximately 4-fold). These results suggested that *CpbHLH056* and *CpbHLH042* may provide different functionalities compared to *Arabidopsis*. Additionally, *CpbHLH027* was also upregulated distinctly under four stresses. In *Arabidopsis*, the orthologue of *CpbHLH027* is *AtbHLH116* (ICE1) (*Fig. 1*), which can be induced by NaCl, ABA and cold stresses, playing an important role in the cold-responsive signaling pathway via an ABA-independent pathway (*Chinnusamy et al., 2003*). There are two orthologues of *CpbHLH027* in rice, one ortholog is *OsICE2/OsbHLH001*, is induced by salt stress, and its overexpression can enhanced the tolerance to freezing and salt stress (*Deng et al., 2017; Li et al., 2010*). *OsICE1/OsbHLH002* is another ortholog in rice, which is induced by chilling stress. *OsbHLH002* can positively regulates cold signaling via targeting *OsTPP1*, which encodes a keyenzyme for trehalose

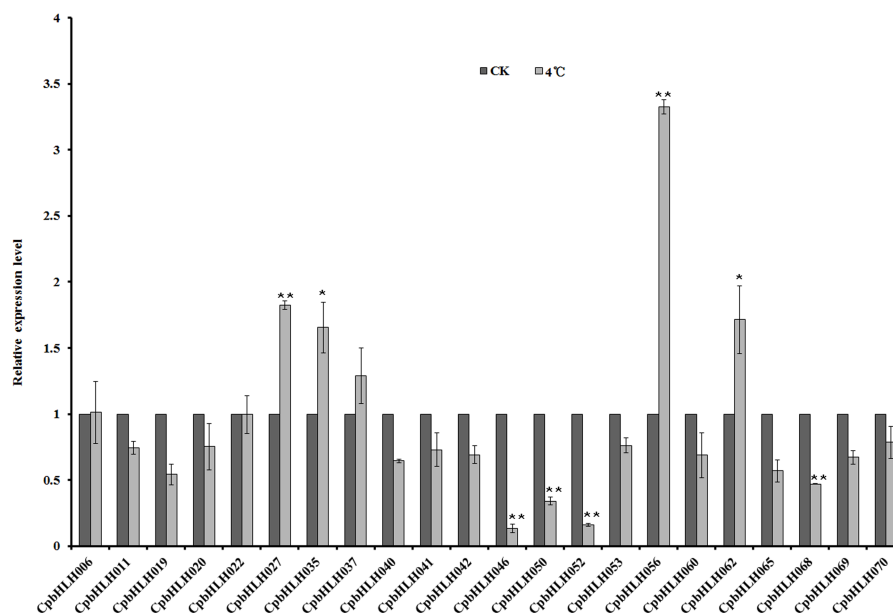


Figure 7 Quantitative RT-PCR analysis of 22 selected *CpbHLH* genes under cold stress condition (4°C). The data are expressed as means \pm SD of three independent biological determinations. Untreated seedlings were used as the control groups. * $P < 0.05$ and ** $P < 0.01$ (Student's t test) indicate significant differences between treated seedlings and control groups.

Full-size DOI: [10.7717/peerj.9319/fig-7](https://doi.org/10.7717/peerj.9319/fig-7)

biosynthesis (Zhang *et al.*, 2017). These results implied *CpbHLH027* plays essential roles in abiotic stresses in papaya. In addition, the transcript of *CpbHLH062* was also increased under cold treatment, its orthologue is *AtbHLH033/ICE2*, which involving the cold response and the ABA pathway (Fursova, Pogorelko & Tarasov, 2009; Kurbidaeva, Ezhova & Novokreshchenova, 2014), implying the *CpbHLH062* may involved in the cold stress. *Cp bHLH053* was downregulated under salt, drought and ABA stresses. The orthologue of *CpbHLH053* is *AtbHLH129* (Fig. 1), which is a transcription repressor that negatively regulates the ABA response in *Arabidopsis* (Tian *et al.*, 2015), implying *CpbHLH053* may have the similar function with the *AtbHLH129* in the process of ABA response.

We should also noticed a few *CpbHLHs* that showed distinct increases or decreases in their mRNA levels under different treatments, and these *CpbHLHs'* orthologues have not been reported in previous studies. For instance, *CpbHLH050* is notably upregulated (>10-fold) under PEG treatment, *CpbHLH046* is upregulated by PEG treatment, but sharply down regulated under ABA and cold treatments, implying these genes may have additional functions than response to drought by regulating root development. *CpbHLH020* and *CpbHLH053* were downregulated (>2-fold) by NaCl, PEG and ABA stresses distinctly. We should also pay attention to these genes in the following research.

DISCUSSION

Transcription factors (TFs) play key roles in the stress regulation network and signal pathways in plants. Basic/helix-loop-helix (bHLH) TFs are the second largest TFs family in

plants and have been identified in many species ([Cheng et al., 2018](#); [Gao et al., 2017](#); [Geng & Liu, 2018](#); [Mao et al., 2017](#); [Sun, Fan & Ling, 2015](#); [Toledo-Ortiz, Huq & Quail, 2003](#); [Wang et al., 2018a](#); [Wang et al., 2018b](#); [Zhang et al., 2018](#)). However, the bHLH TF family has not previously been reported in papaya (*Carica papaya* L.). In this paper, we found 73 *CpbHLH* genes in papaya. This TF family seemed to be one of the moderately sized families compared with other plant species, which might be because of the papaya has a relatively small reference genome, the size is only 372 Mb ([Ming et al., 2008](#)). The gene evolution changes the gene organization. In this study, we found that *CpbHLH* genes were diverse in their number introns, ranging from 0 to 10 ([Data S2](#)). This result implied these genes may have undergone numerous of genetic evolution events, and the genes in different subfamilies may have different functions ([Cheng et al., 2018](#)). Most *CpbHLHs* in the same subfamily shared similar gene structures and protein motifs according to the analysis of exon/intron organizations and motif compositions ([Figs. 2, 3 and 4](#)), indicating that the functions of encoded proteins in each subfamily are probably stable. However, the conserved motif analysis showed that some *CpbHLHs*, which are mainly distributed in eight subfamilies (including V a, V b, III f, IVa, IIIb, III, I b and I a subfamilies) from the phylogenetic tree, harbor another conserved motif (motif 3) with a length of 36 amino acids ([Figs. 2 and 4](#)), indicating that these proteins may have additional functions.

Promoter *cis*-acting regulatory element analysis showed that *cis*-elements could be divided into three main categories: light responsive, abiotic and biotic stresses and plant growth and development. Especially in abiotic and biotic stresses, the most common response elements were ABA (29.15%), MeJA (23.62%) and SA (9.78%), suggesting that these phytohormones may play important roles in the regulation of papaya growth and development ([Fig. 5 and Data S4](#)). In addition, the promoter *cis*-acting element involved in the abscisic acid responsiveness analysis is consistent with the qRT-PCR results ([Data S4 and Table 2](#)), showing four genes (*CpHLH020/-027/-053/-056*) involved in abscisic acid response. Another two genes (*CpHLH020/-062*) were also identified that were involved in abscisic acid response by GO enrichment analysis and qRT-PCR ([Data S6 and Table 2](#)). We also identified a large number of *cis*-acting elements in *CpbHLH* genes that may respond to drought (MBS, 8.67%), which is also consistent with the qRT-PCR (including seven genes: *CpbHLH027/-050/-056/-011/-068/-042/-053*). Other genes also had important elements, including LTR, TC-rich and WUN-motifs, which indicated plant responses to low temperatures, defense stresses and wound-responsiveness, respectively. These results implied *CpbHLH* genes may have a wide range of functions in papaya growth, disease resistance, and response to environmental conditions. We also analyze the *cis*-elements in the promoter regions of putative orthologous genes in *Arabidopsis* ([Fig. S1 and Data S4](#)). And we compared the promoter *cis*-elements of *bHLH* genes in *Arabidopsis*, papaya and previously reported bamboo ([Cheng et al., 2018](#)). The result showed that the promoter *cis*-elements of *bHLH* genes in these three plants were divided into three categories, and most of elements were the same. These similar results implied that most of the promoter *cis*-elements of *bHLH* family were conserved in *Arabidopsis*, papaya and bamboo. The most notable is, the percentage of MeJA responsive *cis*-elements in papaya and *Arabidopsis* (23.62%, 30.61%) were less than bamboo (43.39%). And the

percentage of SA responsive *cis*-elements in papaya and *Arabidopsis* (9.78%, 5.87%) were also less than bamboo (10.31%). SA is a phytohormone that plays important roles in plant defenses against pathogens (Pokotylo, Kravets & Ruelland, 2019). MeJA also has been identified as a vital cellular regulator that mediates defense processes (Cheong & Choi, 2003). So, the stress-responsive elements in papaya and *Arabidopsis* were corresponding less than bamboo, including drought-responsive elements, wound-responsive elements, low temperature-responsive elements, and defense and stress responsive elements. These results may help explain why papaya is more sensitive to external stresses compared to bamboo.

Many studies have shown that *bHLH* genes are involved in various abiotic and biotic stresses responses. We randomly selected 22 genes to investigate their expression profiles by using qRT-PCR under salt, drought, ABA and cold stresses (Fig. 7 and Table 2). The results revealed some candidate *CpbHLH* genes that might be responsible for abiotic stress responses in papaya. For example, *CpbHLH027*, *CpbHLH062*, *AtbHLH116* (ICE1), *AtbHLH33* (ICE2), *OsbHLH001* (OsICE2) and *OsbHLH002* (OsICE1) were clustered within one clade. Among them, *AtbHLH116* (ICE1), *AtbHLH33* (ICE2), *OsbHLH001* (OsICE2) and *OsbHLH002* (OsICE1) have been reported function in chilling stress in *Arabidopsis* and rice (Chinnusamy et al., 2003; Deng et al., 2017; Fursova, Pogorelko & Tarasov, 2009; Li et al., 2010; Zhang et al., 2017). And the transcripts of *CpbHLH027*, *CpbHLH062* were increased under chilling stress in this study, implying that *CpbHLH027* and *CpbHLH062* may be also involved in the process of chilling stress. The orthologue of *CpHLH056* in *Arabidopsis* is *BEE1* (At1G18400), which is a positive regulator of flavonoid accumulation (Petridis et al., 2016). *BEE1*, *BEE2* and *BEE3* are functionally redundant positive regulators of BR signaling, and their transcription is repressed by ABA in *Arabidopsis* (Friedrichsen et al., 2002). However, we found that the transcription of *CpbHLH056* is notably upregulated (> 10-fold) under ABA treatment rather than downregulated in this study. These results imply that *CpbHLH056* may involved in the process of ABA stress but has different function compared to *Arabidopsis*. We have also noticed a few candidate *CpbHLHs* that showed distinct increases or decreases in their mRNA levels under different treatments, and these *CpbHLHs*' orthologues have not been reported in other plants. For instance, *CpbHLH050*, *CpbHLH020*, *CpbHLH046*, *CpbHLH053*, and so on. These findings provide important candidate genes/proteins necessary for further functional research on the *bHLH* family in papaya.

CONCLUSIONS

In conclusion, the study performed a genome-wide analysis of basic helix-loop-helix (*bHLH*) transcription factors in papaya. As a result, a total of 73 *bHLH* genes were identified in papaya, and these *CpbHLHs* were classified into 18 subfamilies with one orphan, which was consistent with the earlier results showing that the *bHLH* subfamily in plants can be divided into 15–25 subfamilies. Almost all of the *CpbHLHs* in the same subfamily shared similar gene structures and protein motifs according to analysis of exon/intron organizations and motif compositions. These results further supported the

classification predicted by the phylogenetic tree. Compared to rice and *Arabidopsis*, the amino acid sequences of the *CpbHLH* domains were quite conservative, especially Leu-27 and Leu-63. Promoter *cis*-element and GO annotation analysis revealed that most of the *CpbHLHs* could respond to various biotic/abiotic stress-related events. Abiotic stress treatment and quantitative real-time PCR (qRT-PCR) assay further supported promoter *cis*-acting regulatory element and GO annotation analysis, revealed some important candidate *CpbHLHs* that might be responsible for abiotic stress responses in papaya. We completed the first comprehensive genome-wide analysis of the *bHLH* gene family in papaya, and our results provide information necessary for further functional research of the bHLH family in papaya.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Presidential Foundation of Guangdong Academy of Agricultural Sciences, China (No. 201820), the Key-Area Research and Development Program of Guangdong Province (No. 2019B020214005), and the Innovation Team Program of Modern Agricultural Industry Technology System in Guangdong Province of China (2019KJ116). 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:

Presidential Foundation of Guangdong Academy of Agricultural Sciences, China: 201820.

Key-Area Research and Development Program of Guangdong Province, China: 2019B020214005.

Innovation Team Program of Modern Agricultural Industry Technology System in Guangdong Province, China: 2019KJ116.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Min Yang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Chenping Zhou and Hu Yang performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Ruibin Kuang and Bingxiong Huang performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Yuerong Wei conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data are available in the [Supplementary Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.9319#supplemental-information>.

REFERENCES

- Agarwal PK, Agarwal P, Reddy MK, Sopory SK. 2006.** Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. *Plant Cell Reports* **25**:1263–1274 DOI [10.1007/s00299-006-0204-8](https://doi.org/10.1007/s00299-006-0204-8).
- Amoutzias GD, Robertson DL, Bornberg-Bauer E. 2004.** The evolution of protein interaction networks in regulatory proteins. *Comparative and Functional Genomics* **5**:79–84 DOI [10.1002/cfg.365](https://doi.org/10.1002/cfg.365).
- Atchley WR, Fitch WM. 1997.** A natural classification of the basic helix-loop-helix class of transcription factors. *Proceedings of the National Academy of Sciences* **94**:5172–5176 DOI [10.1073/pnas.94.10.5172](https://doi.org/10.1073/pnas.94.10.5172).
- Atchley WR, Terhalle W, Dress A. 1999.** Positional dependence, cliques, and predictive motifs in the bhlh protein domain. *Journal of Molecular Evolution* **48**:501–516 DOI [10.1007/PL00006494](https://doi.org/10.1007/PL00006494).
- Bailey TL, Elkan C. 1994.** Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the 2nd International Conference on Intelligent Systems for Molecular Biology* **2**:28–36.
- Bernhardt C, Lee MM, Gonzalez A, Zhang F, Lloyd A, Schiefelbein J. 2003.** The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root. *Development* **130**:6431–6439 DOI [10.1242/dev.00880](https://doi.org/10.1242/dev.00880).
- Buck MJ, Atchley WR. 2003.** Phylogenetic analysis of plant basic helix-loop-helix proteins. *Journal of Molecular Evolution* **56**:742–750 DOI [10.1007/s00239-002-2449-3](https://doi.org/10.1007/s00239-002-2449-3).
- Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martinez-Garcia JF, Bilbao-Castro JR, Robertson DL. 2010.** Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. *Plant Physiology* **153**:1398–1412 DOI [10.1104/pp.110.153593](https://doi.org/10.1104/pp.110.153593).
- Chen YY, Li MY, Wu XJ, Huang Y, Ma J, Xiong AS. 2015.** Genome-wide analysis of basic helix-loop-helix family transcription factors and their role in responses to abiotic stress in carrot. *Molecular Breeding* **35**:125–137.
- Cheng X, Xiong R, Liu H, Wu M, Chen F, Hanwei Y, Xiang Y. 2018.** Basic helix-loop-helix gene family: genome wide identification, phylogeny, and expression in Moso bamboo. *Plant Physiology and Biochemistry* **132**:104–119 DOI [10.1016/j.plaphy.2018.08.036](https://doi.org/10.1016/j.plaphy.2018.08.036).

- Cheong J-J, Choi YD. 2003.** Methyl jasmonate as a vital substance in plants. *Trends in Genetics* **19**:409–413 DOI [10.1016/S0168-9525\(03\)00138-0](https://doi.org/10.1016/S0168-9525(03)00138-0).
- Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, Zhu JK. 2003.** ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. *Genes and Development* **17**:1043–1054 DOI [10.1101/gad.1077503](https://doi.org/10.1101/gad.1077503).
- Deng C, Ye H, Fan M, Pu T, Yan J. 2017.** The rice transcription factors OsICE confer enhanced cold tolerance in transgenic Arabidopsis. *Plant Signaling & Behavior* **12**:e1316442.
- Eddy SR. 1998.** Profile hidden Markov models. *Bioinformatics* **14**:755–763 DOI [10.1093/bioinformatics/14.9.755](https://doi.org/10.1093/bioinformatics/14.9.755).
- Farquharson KL. 2016.** A Domain in the bHLH transcription factor DYT1 is critical for anther development. *The Plant Cell* **28**:997–998 DOI [10.1105/tpc.16.00331](https://doi.org/10.1105/tpc.16.00331).
- Feller A, Machemer K, Braun EL, Grotewold E. 2011.** Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *The Plant Journal* **66**:94–116 DOI [10.1111/j.1365-3113.2010.04459.x](https://doi.org/10.1111/j.1365-3113.2010.04459.x).
- Feng Y, Xu P, Li B, Li P, Wen X, An F, Gong Y, Xin Y, Zhu Z, Wang Y, Guo H. 2017.** Ethylene promotes root hair growth through coordinated EIN3/EIL1 and RHD6/RSL1 activity in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **114**:13834–13839 DOI [10.1073/pnas.1711723115](https://doi.org/10.1073/pnas.1711723115).
- Friedrichsen DM, Nemhauser J, Muramitsu T, Maloof JN, Alonso J, Ecker JR, Furuya M, Chory J. 2002.** Three redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth. *Genetics* **162**:1445–1456.
- Fursova OV, Pogorelko GV, Tarasov VA. 2009.** Identification of ICE2, a gene involved in cold acclimation which determines freezing tolerance in Arabidopsis thaliana. *Gene* **429**:98–103 DOI [10.1016/j.gene.2008.10.016](https://doi.org/10.1016/j.gene.2008.10.016).
- Gao C, Sun J, Wang C, Dong Y, Xiao S, Wang X, Jiao Z. 2017.** Genome-wide analysis of basic/helix-loop-helix gene family in peanut and assessment of its roles in pod development. *PLOS ONE* **12**:e0181843 DOI [10.1371/journal.pone.0181843](https://doi.org/10.1371/journal.pone.0181843).
- Geng J, Liu JH. 2018.** The transcription factor CsbHLH18 of sweet orange functions in modulation of cold tolerance and homeostasis of reactive oxygen species by regulating the antioxidant gene. *Journal of Experimental Botany* **69**:2677–2692 DOI [10.1093/jxb/ery065](https://doi.org/10.1093/jxb/ery065).
- Guo XJ, Wang JR. 2017.** Global identification, structural analysis and expression characterization of bHLH transcription factors in wheat. *BMC Plant Biology* **17**:90 DOI [10.1186/s12870-017-1038-y](https://doi.org/10.1186/s12870-017-1038-y).
- Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC. 2003.** The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Molecular Biology and Evolution* **20**:735–747 DOI [10.1093/molbev/msg088](https://doi.org/10.1093/molbev/msg088).
- Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. 2015.** GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31**:1296–1297 DOI [10.1093/bioinformatics/btu817](https://doi.org/10.1093/bioinformatics/btu817).

- Jiang Y, Yang B, Deyholos MK. 2009.** Functional characterization of the Arabidopsis bHLH92 transcription factor in abiotic stress. *Molecular Genetics and Genomics* **282**:503–516 DOI [10.1007/s00438-009-0481-3](https://doi.org/10.1007/s00438-009-0481-3).
- Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G. 2017.** PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research* **45**:D1040–D1045 DOI [10.1093/nar/gkw982](https://doi.org/10.1093/nar/gkw982).
- Jones S. 2004.** An overview of the basic helix-loop-helix proteins. *Genome Biology* **5**:226 DOI [10.1186/gb-2004-5-6-226](https://doi.org/10.1186/gb-2004-5-6-226).
- Kaotoh MK, Kuma K, Miyata T. 2002.** MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**:3059–3066 DOI [10.1093/nar/gkf436](https://doi.org/10.1093/nar/gkf436).
- Kosugi S, Ohashi Y. 2002.** DNA binding and dimerization specificity and potential targets for the TCP protein family. *The Plant Journal* **30**:337.
- Kurbidaeva A, Ezhova T, Novokreshchenova M. 2014.** *Plant Science* **229**:10–22 DOI [10.1016/j.plantsci.2014.08.011](https://doi.org/10.1016/j.plantsci.2014.08.011).
- Ledent V, Vervoort M. 2001.** The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. *Genome Research* **11**:754–770 DOI [10.1101/gr.177001](https://doi.org/10.1101/gr.177001).
- Lee S, Lee S, Yang KY, Kim YM, Park SY, Kim SY, Soh MS. 2006.** Overexpression of PRE1 and its homologous genes activates Gibberellin-dependent responses in Arabidopsis thaliana. *Plant and Cell Physiology* **47**:591–600 DOI [10.1093/pcp/pcj026](https://doi.org/10.1093/pcp/pcj026).
- Leivar P, Monte E, Al-Sady B, Carle C, Storer A, Alonso JM, Ecker JR, Quail PH. 2008.** The Arabidopsis phytochrome-interacting factor PIF7, together with PIF3 and PIF4, regulates responses to prolonged red light by modulating phyB levels. *The Plant Cell* **20**:337–352 DOI [10.1105/tpc.107.052142](https://doi.org/10.1105/tpc.107.052142).
- Li F, Guo S, Zhao Y, Chen D, Chong K, Xu Y. 2010.** Overexpression of a homopeptide repeat-containing bHLH protein gene (OrbHLH001) from Dongxiang Wild Rice confers freezing and salt tolerance in transgenic Arabidopsis. *Plant Cell Reports* **29**:977–986 DOI [10.1007/s00299-010-0883-z](https://doi.org/10.1007/s00299-010-0883-z).
- Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, Guo J, Liang W, Chen L, Yin J, Ma H, Wang J, Zhang D. 2006.** Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. *Plant Physiology* **141**:1167–1184 DOI [10.1104/pp.106.080580](https://doi.org/10.1104/pp.106.080580).
- Liu K, Yuan C, Feng S, Zhong S, Li H, Zhong J, Shen C, Liu J. 2017a.** Genome-wide analysis and characterization of Aux/IAA family genes related to fruit ripening in papaya (*Carica papaya* L.). *BMC Genomics* **18**:351 DOI [10.1186/s12864-017-3722-6](https://doi.org/10.1186/s12864-017-3722-6).
- Liu K, Yuan C, Li H, Lin W, Yang Y, Shen C, Zheng X. 2015.** Genome-wide identification and characterization of auxin response factor (ARF) family genes related to flower and fruit development in papaya (*Carica papaya* L.). *BMC Genomics* **16**:901.
- Liu W, Tai H, Li S, Gao W, Zhao M, Xie C, Li WX. 2014.** bHLH122 is important for drought and osmotic stress resistance in Arabidopsis and in the repression of ABA catabolism. *New Phytologist* **201**:1192–1204 DOI [10.1111/nph.12607](https://doi.org/10.1111/nph.12607).

- Liu Y, Li J, Wei G, Sun Y, Lu Y, Lan H, Li C, Zhang S, Cao M. 2017. Cloning, molecular evolution and functional characterization of ZmbHLH16, the maize ortholog of OsTIP2 (OsbHLH142). *Biology Open* 6:1654–1663 DOI 10.1242/bio.026393.
- Magali L, Patrice D, Gert T, Kathleen M, Yves M, VdP Yves, Pierre R, Stephane R. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research* 30:325–327 DOI 10.1093/nar/30.1.325.
- Mao K, Dong Q, Li C, Liu C, Ma F. 2017. Genome wide identification and characterization of apple bhlh transcription factors and expression analysis in response to drought and salt stress. *Frontiers in Plant Science* 8:480.
- Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang ML, Zhu YJ, Schatz M, Nagarajan N, Acob RA, Guan P, Blas A, Wai CM, Ackerman CM, Ren Y, Liu C, Wang J, Wang J, Na JK, Shakirov EV, Haas B, Thimmapuram J, Nelson D, Wang X, Bowers JE, Gschwend AR, Delcher AL, Singh R, Suzuki JY, Tripathi S, Neupane K, Wei H, Irikura B, Paidi M, Jiang N, Zhang W, Presting G, Windsor A, Navajas-Pérez R, Torres MJ, Feltus FA, Porter B, Li Y, Burroughs AM, Luo MC, Liu L, Christopher DA, Mount SM, Moore PH, Sugimura T, Jiang J, Schuler MA, Friedman V, Mitchell-Olds T, Shippen DE, dePamphilis CW, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya Linnaeus*). *Nature* 452:991–996 DOI 10.1038/nature06856.
- Morgenstern B, Atchley WR. 1999. Evolution of bHLH transcription factors: modular evolution by domain shuffling? *Molecular Biology & Evolution* 16:1654–1663 DOI 10.1093/oxfordjournals.molbev.a026079.
- Murre C, Mccaw PS, Baltimore D. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56:777–783 DOI 10.1016/0092-8674(89)90682-X.
- Niu X, Guan Y, Chen S, Li H. 2017. Genome-wide analysis of basic helix-loop-helix (bHLH) transcription factors in *Brachypodium distachyon*. *BMC Genomics* 18:619.
- Pan LJ, Jiang L. 2014. Identification and expression of the WRKY transcription factors of *Carica papaya* in response to abiotic and biotic stresses. *Molecular Biology Reports* 41:1215–1225 DOI 10.1007/s11033-013-2966-8.
- Peraza-Echeverria S, Santamaría JM, Fuentes G. 2012. The NPR1 family of transcription cofactors in papaya: insights into its structure, phylogeny and expression. *Genes & Genomics* 34:379–390 DOI 10.1007/s13258-011-0218-7.
- Petridis A, Döll S, Nichelmann L, Bilger W, Mock HP. 2016. Arabidopsis thaliana G2-LIKE FLAVONOID REGULATOR and BRASSINOSTEROID ENHANCED EXPRESSION1 are low-temperature regulators of flavonoid accumulation. *New Phytologist* 211:912–925 DOI 10.1111/nph.13986.
- Pires N, Dolan L. 2010. Origin and diversification of basic-helix-loop-helix proteins in plants. *Molecular Biology and Evolution* 27:862–874 DOI 10.1093/molbev/msp288.

- Pokotylo I, Kravets V, Ruelland E. 2019.** Salicylic acid binding proteins (sabps): the hidden forefront of salicylic acid signalling. *International Journal of Molecular Sciences* **20**:4377 DOI [10.3390/ijms20184377](https://doi.org/10.3390/ijms20184377).
- Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, Zhu YJ. 2009.** Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. *Molecular Genetics and Genomics* **281**:609–626 DOI [10.1007/s00438-009-0434-x](https://doi.org/10.1007/s00438-009-0434-x).
- Price MN, Dehal PS, Arkin AP. 2009.** FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution* **26**:1641–1650 DOI [10.1093/molbev/msp077](https://doi.org/10.1093/molbev/msp077).
- Roig-Villanova I, Bou-Torrent J, Galstyan A, Carretero-Paulet L, Portolés S. 2007.** Interaction of shade avoidance and auxin responses: a role for two novel atypical bHLH proteins. *Embo Journal* **26**:4756–4767 DOI [10.1038/sj.emboj.7601890](https://doi.org/10.1038/sj.emboj.7601890).
- Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, Coornaert D, Degnan BM, Vervoort M. 2007.** Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. *BMC Evolutionary Biology* **7**:33 DOI [10.1186/1471-2148-7-33](https://doi.org/10.1186/1471-2148-7-33).
- Song Y, Liu H-D, Zhou Q, Zhang H-J, Zhang Z-D, Li Y-D, Wang H-B, Liu F-Z. 2017.** High-throughput sequencing of highbush blueberry transcriptome and analysis of basic helix-loop-helix transcription factors. *Journal of Integrative Agriculture* **16**:591–604 DOI [10.1016/S2095-3119\(16\)61461-2](https://doi.org/10.1016/S2095-3119(16)61461-2).
- Stevens JD, Roalson EH, Skinner MK. 2008.** Phylogenetic and expression analysis of the basic helix-loop-helix transcription factor gene family: genomic approach to cellular differentiation. *Differentiation* **76**:1006–1022 DOI [10.1111/j.1432-0436.2008.00285.x](https://doi.org/10.1111/j.1432-0436.2008.00285.x).
- Sun H, Fan HJ, Ling HQ. 2015.** Genome-wide identification and characterization of the bHLH gene family in tomato. *BMC Genomics* **16**:9 DOI [10.1186/s12864-014-1209-2](https://doi.org/10.1186/s12864-014-1209-2).
- Tian H, Guo H, Dai X, Cheng Y, Zheng K, Wang X, Wang S. 2015.** An ABA down-regulated bHLH transcription repressor gene, bHLH129 regulates root elongation and ABA response when overexpressed in Arabidopsis. *Scientific Reports* **5**:17587 DOI [10.1038/srep17587](https://doi.org/10.1038/srep17587).
- Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. 2017.** agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research* **45**:W122–W129 DOI [10.1093/nar/gkx382](https://doi.org/10.1093/nar/gkx382).
- Toledo-Ortiz G, Huq E, Quail PH. 2003.** The arabidopsis basic/helix-loop-helix transcription factor family. *The Plant Cell Online* **15**:1749–1770 DOI [10.1105/tpc.013839](https://doi.org/10.1105/tpc.013839).
- Vervoort M, Ledent V. 2001.** The evolution of the neural basic Helix-Loop-Helix proteins. *Scientific World Journal* **1**:396–426 DOI [10.1100/tsw.2001.68](https://doi.org/10.1100/tsw.2001.68).
- Wang P, Su L, Gao H, Jiang X, Wu X, Li Y, Zhang Q, Wang Y, Ren F. 2018a.** Genome-Wide characterization of bhlh genes in grape and analysis of their potential relevance to abiotic stress tolerance and secondary metabolite biosynthesis. *Frontiers in Plant Science* **9**:64.
- Wang R, Zhao P, Kong N, Lu R, Pei Y, Huang C, Ma H, Chen Q. 2018b.** Genome-wide identification and characterization of the potato bHLH transcription factor family. *Genes (Basel)* **9**:54–74.

- Wu S, Zhu Y. 2012.** ProPAS: standalone software to analyze protein properties. *Bioinformatics* **8**:167–169 DOI [10.6026/97320630008167](https://doi.org/10.6026/97320630008167).
- Xu G, Guo C, Shan H, Kong H. 2012.** Divergence of duplicate genes in exon-intron structure. *Proceedings of the National Academy of Sciences of the United States of America* **109**:1187–1192 DOI [10.1073/pnas.1109047109](https://doi.org/10.1073/pnas.1109047109).
- Xu Y, Xu H, Wall MM, Yang J. 2020.** Roles of transcription factor SQUAMOSA promoter binding protein-like gene family in papaya (*Carica papaya*) development and ripening. *Genomics* **112**:2734–2747.
- Zhang C, Feng R, Ma R, Shen Z, Cai Z, Song Z, Peng B, Yu M. 2018.** Genome-wide analysis of basic helix-loop-helix superfamily members in peach. *PLOS ONE* **13**:e0195974 DOI [10.1371/journal.pone.0195974](https://doi.org/10.1371/journal.pone.0195974).
- Zhang JZ, Creelman RA, Zhu J-K. 2004.** From Laboratory to field, using information from *arabidopsis* to engineer salt, cold, and drought tolerance in crops. *Plant Physiology* **135**:615–621 DOI [10.1104/pp.104.040295](https://doi.org/10.1104/pp.104.040295).
- Zhang Z, Li J, Li F, Liu H, Yang W, Chong K, Xu Y. 2017.** OsMAPK3 phosphorylates OsbHLH002/OsICE1 and inhibits its ubiquitination to activate ostpp1 and enhances rice chilling tolerance. *Developmental Cell* **43**:731–743.e735 DOI [10.1016/j.devcel.2017.11.016](https://doi.org/10.1016/j.devcel.2017.11.016).
- Zhu E, You C, Wang S, Cui J, Niu B, Wang Y, Qi J, Ma H, Chang F. 2015.** The DYT1-interacting proteins bHLH010, bHLH089 and bHLH091 are redundantly required for Arabidopsis anther development and transcriptome. *The Plant Journal* **83**:976–990 DOI [10.1111/tpj.12942](https://doi.org/10.1111/tpj.12942).
- Zhu X, Li X, Chen W, Chen J, Lu W, Chen L, Fu D. 2012.** Evaluation of new reference genes in papaya for accurate transcript normalization under different experimental conditions. *PLOS ONE* **7**:e44405.