

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

Min Yang

Guangdong Academy of Agricultural Sciences

Hu Yang

Guangdong Academy of Agricultural Sciences

Ruibin Kuang

Guangdong Academy of Agricultural Sciences

Chenping Zhou

Guangdong Academy of Agricultural Sciences

Bingxiong Huang

Guangdong Academy of Agricultural Sciences

Yuerong Wei (✉ weid18@163.com)

Guangdong Academy of Agricultural Sciences

Research article

Keywords: papaya, genome-wide analysis, bHLH transcription factors, abiotic stress

Posted Date: November 7th, 2019

DOI: <https://doi.org/10.21203/rs.2.16982/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background : As a superfamily of transcription factors (TFs), the basic helix-loop-helix (bHLH) proteins 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. **Results:** In this study, a total of 73 CpbHLH genes were found in papaya, and these genes were classified into 18 subfamilies based on phylogenetic analysis, with one orphans. Almost all of the CpbHLH in the same subfamily shared similar gene structures and protein motifs according to an analysis of exon/intron organizations and motif compositions. The number of exons in CpbHLH genes varied from 1 to 11 with an average of 5. 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 CpbHLH genes contained cis -elements that can respond to various biotic/abiotic stress-related events. Gene ontology (GO) analysis revealed that Cp bHLH mainly functions in protein dimerization activity and DNA-binding, and most Cp bHLH proteins were predicted to localize in the nucleus. Abiotic stress treatment and quantitative real-time PCR (qRT-PCR) revealed some predicted CpbHLH genes that might be responsible for abiotic stress responses in papaya. **Conclusions :** A total of 73 bHLH transcription factors were identified from papaya, and their gene structures, conserved domains, sequence features, phylogenetic relationship, promoter cis -element, GO annotation and gene expression profiles responsible for abiotic stress were investigated. Our findings lay a foundation for further evolutionary and functional elucidation of Cp bHLHs. **Keywords :** papaya, genome-wide analysis, bHLH transcription factors, abiotic stress

Background

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. To optimize their growth and survival, plants must utilize a wide range of physiological and biochemical processes in their responses to diverse abiotic stresses by regulating the expression of a series of genes[1-3].

Transcription factors (TFs) play key roles in the stress regulation network and signaling pathways mainly by activating and repressing related downstream genes to regulate gene expression. Among them, basic/helix-loop-helix (bHLH) TFs are widely found in all eukaryotes and are the second largest TF family in plants[2-5]. Proteins of the bHLH superfamily in all eukaryotes are characterized by one highly conserved bHLH domain, which is approximately 50–60 amino acids in length and divided into two different functional regions: the basic region and the HLH region[6, 7]. The basic region is located at the N-terminal end of the domain and consists of 10–15 amino acids. It is a DNA-binding region that enables bHLH TFs to bind to E-boxes (CANNTG)[8, 9]. 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[7, 10]. Outside of the two conserved regions, the rest of the bHLH protein sequences are usually very different[11].

In animals, the bHLH TFs can be divided into six groups (A to F) based on their phylogenetic relationships, functional properties and DNA-binding specificity[8]. These bHLH groups can be divided into several small subfamilies[12, 13]. The bHLHs mainly function in sensing the external environment, cell cycle regulation and tissue differentiation[8, 14-16]. Compared to animals, the research on bHLH proteins in plants is limited: the number of groups has not been determined but is thought to cover 15–26 subfamilies[3, 17]. Phylogenetic analyses of some atypical bHLH proteins have even extended that number to 32[5]. 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*[6], peanut[18], apple[19], tomato[20], potato[21], peach[22], grapes[23], sweet orange[24], and bamboo[25]. The results from these studies have shown that bHLH TFs have versatile biological functions, such as regulating light morphogenesis[26, 27], hormone signals[28, 29], the developmental of root[30] and anther[31], regulating epidermal cell fate determination[32], participating in various biotic and abiotic stress responses[21, 33, 34], etc.

As an important and popular fruit, papaya is widely grown throughout the tropics and subtropics. However, the production and quality of papaya were often threatened by various abiotic stresses, such as salt, drought, and cold. In recent years, some studies demonstrated that bHLH transcription factors play important roles in the stress-related regulation network and signalling pathways in many species. However, no systematic analysis of bHLH TFs has been performed in papaya previously. In this study, a total of 73 predicted *bHLH* 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 *CpbHLHs* 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 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 on the bHLH family in papaya.

Results

Identification and characterization of *CpbHLHs*

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 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 *CpbHLHs*, 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 Additional file 1. In previous studies, 129/132, 188, 159, 147, 124, 95, 94 and 56 *bHLH* genes were

identified in peanut[18], apple[19], tomato[20], Arabidopsis[6], potato[21], peach[22], grapes[23] and sweet orange[24], respectively. Compared with the above dicotyledonous plants, the density of *bHLHs* superfamily genes in papaya genome was approximately 0.26%, which is lower than the density of peanut[18], apple[19], tomato[20], arabidopsis[6], potato[21] and sweet orange[24], and similar to peach[22] and wine grapes[23] (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 peanut, apple, tomato, Arabidopsis, potato and sweet orange while the plants without whole-gene duplication like papaya, peach and wine grapes.

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

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

To further characterize the bHLHs in papaya, the physicochemical properties of these putative proteins were analyzed and are shown in Additional file 1. 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 patato[21]. And all of *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*[35]. That is, the *CpbHLH* proteins showed diversities in their length, molecular weight, PI and the 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 papaya bHLH 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 Additional file 2), consistent with the earlier results showing that the bHLH subfamily in plants can be divided into 15–25 subfamilies[3]. Previous studies have named the bHLH subfamilies using English letters[7, 19], Roman numerals[20, 36], or Arabic numerals[6, 37], In this study, we adopted a nomenclature using Roman numerals. As shown in figure 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 *d*, *e*, *f*, *g*, *h* and *i* 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 [38]. The annotation features of the *CpbHLH* genes were submitted to Gene Structure Display Server (GSDS) together to show their gene structures. As described in Additional file 1 and Fig. 2A, the number of introns varied from zero to ten, representing a complex distribution pattern. Most (63 (86.3%)) of the *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 *b* had one intron and two exons, the whole members of subfamily *f* had six introns and seven exons, the *a* subfamily members showed three introns and four exons, and all members of *b* subfamily consisted only one exons. 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[3]. For example, *AtbHLH10*, *AtbHLH89* and *AtbHLH91*, corresponding rice orthologs *OsbHLH141*, *OsbHLH142* are members of *e* subfamily, they are all involved in the process of pollen development[7, 39, 40]. 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[40]. In subfamily *b*, *OsbHLH001*(*OslCE2*), *OsbHLH002* (*OslCE1*), *CpbHLH027*, *CpbHLH062*, *AtbHLH116* (*ICE1*) and *AtbHLH33* (*ICE2*) were clustered within one clade. In previous studies, *AtbHLH116*(*ICE1*) and *AtbHLH33*(*ICE2*) and corresponding orthologs (*OsbHLH001*/*OslCE2*, *OsbHLH002*/*OslCE1*) have been reported to function in the stress of chilling[41-45]. And we also found transcripts of *CpbHLH027* and *CpbHLH062* were increased under chilling stress in this study, implying *CpbHLH027* and *CpbHLH062* involved in the process of chilling stress in papaya.

To further study the sequence characteristics of the bHLH domains at the amino acid level, we carried out a multiple sequence alignment of the 73 *CpbHLH* protein sequences (Fig.3). The result showed that the 73 *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, the MEME program was used to identify the conserved motifs shared by bHLH proteins by uploading the 73 amino acid sequences of the *CpbHLH* family[46]. Almost all of the sequences (except *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 blue blocks, respectively (Fig.2B and Fig.4A). Among the two motifs, motif 1 was composed of basic residues and helix 1, and motif 2 was composed of 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 blue) are shown in Fig. 4A. The backbones of motif 1 and 2 are also conserved among plant species [10, 20, 47], which is believed to indicate that these highly conserved residues in bHLH domains are responsible for dimerization[10].

In addition to the two common conserved motifs shared by the *CpbHLH* proteins, some *CpbHLH*s that are mainly distributed into eight subfamilies (including α a, α b, α f, α a, α b, α , α b and α a subfamilies) harbor another conserved motif (motif 3) with a length of 36 amino acids. This motif is indicated by the green blocks and the sequence logo is visualized as logo3 (Fig. 2B and Fig. 4B). In previous reports, members of a given subfamily of the bHLH superfamily exhibited another conserved motif (conserved nonbHLH motif: motif 3) in plants[3]. 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 *CpbHLH*s, one protein (*CpbHLH003*) exhibited incomplete bHLH domains, whereas the remaining 72 bHLH proteins all presented complete bHLH domains. Similar observations have been made in other plant species, such as peach and blueberry[22, 36].

Promoter analysis of *bHLH* genes in papaya

To further understand their functions and regulation patterns, *cis*-elements in *CpbHLH* gene 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 Additional file 3). 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 (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 IAA (AuxRR-core, TGA-element, AuxRE, TCA-element and TGA-box), which included 158 (29.15%), 128 (23.62%) and 90 (16.79%), respectively (Fig. 5B). Category three contained plant growth and development elements, such as anaerobic induction elements (ARE), O₂-site, CAT-box and so on.

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 Additional file 4. A total of 70 *CpbHLH*s were involved in protein dimerization activity (GO:0046983), and previous studies showed that the HLH domain was necessary in both dimerization and DNA binding[48]. Especially, Leu-27 in helix 1 and Leu-73 in helix 2 are important for protein interactions[5]. In this study, 72 and 73 (out of 73) *CpbHLH* proteins were found to have Leu-27 (corresponding to Leu-27 in *AtbHLH*s) and Leu-63 (corresponding to Leu-73 in *AtbHLH*s), respectively (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 annotation of papaya bHLH proteins, and 54 of 73 *CpbHLH* proteins were

obtained with results compared to Arabidopsis. We summarized the results in Fig. 6 and Additional file 5. The majority of *CpbHLH* proteins were involved in DNA binding. Almost all of the *CpbHLH* proteins (37, 68.5%) were predicted to localize in the nucleus, whereas the remaining *CpbHLH* proteins were located in other organelles, including plastids, the cytoplasm, and chloroplasts. Additionally, some *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 *CpbHLH* proteins (47, 87.0%). Biosynthetic processes and gene expression involved the second greatest number of *CpbHLH* proteins (46, 85.2%). In addition, bHLH proteins could respond to stimulus, morphogenesis, cell differentiation and the 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 bHLHs 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 of 22 bHLH mRNAs were increased, and 3 bHLH mRNAs were reduced more than 2-fold in salt (200 mM NaCl) treated papaya seedlings. Under drought stress (25% PEG), 8 of 22 bHLH mRNAs were upregulated, and 3 bHLH mRNAs were downregulated more than 2-fold. Under ABA treatment (100 mM), 3 of 22 bHLH mRNAs were upregulated, and 5 bHLH mRNAs were downregulated. Under cold stress (4°C), there were 3 genes whose expression increased more than 1.5-fold, and 4 bHLH mRNAs were reduced more than 2-fold.

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

The name of bHLHs	CK	NaCl (200 mM)	PEG (25%)	ABA (100 mM)
bHLH006	1	0.74±0.002	0.95±0.12	0.75±0.15
bHLH011	1	2.45±0.09	4.06±0.20	1.66±0.02
bHLH019	1	0.80±0.08	0.64±0.13	0.18±0.04
bHLH020	1	0.20±0.0001	0.39±0.05	0.22±0.03
bHLH022	1	4.21±0.57	3.05±0.35	1.54±0.04
bHLH027	1	4.49±0.54	2.32±0.05	2.27±0.002
bHLH035	1	1.62±0.48	1.79±0.14	0.71±0.19
bHLH037	1	0.96±0.15	1.04±0.20	1.92±0.09
bHLH040	1	1.38±0.12	0.77±0.13	0.51±0.06
bHLH041	1	0.55±0.08	0.76±0.15	0.70±0.11
bHLH042	1	0.55±0.10	0.33±0.04	0.26±0.005
bHLH046	1	0.77±0.10	5.46±0.22	0.53±0.12
bHLH050	1	0.53±0.07	11.86±0.34	1.18±0.01
bHLH052	1	1.64±0.02	8.71±0.66	4.42±0.16
bHLH053	1	0.44±0.01	0.31±0.05	0.19±0.04
bHLH056	1	7.11±0.02	32.65±5.63	12.22±3.04
bHLH060	1	0.89±0.03	0.76±0.24	0.75±0.24
bHLH062	1	0.38±0.07	0.65±0.19	0.30±0.10
bHLH065	1	1.64±0.21	1.04±0.03	0.52±0.14
bHLH068	1	1.72±0.09	2.37±0.12	1.39±0.20
bHLH069	1	0.68±0.02	0.89±0.16	0.52±0.004
bHLH070	1	1.11±0.03	1.17±0.23	1.47±0.35

Quantitative RT-PCR was used to investigate the expression levels (shown in fold change) of the *bHLHs*. 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.

Interestingly, a few transcripts of bHLH responded to all or multiple stresses. For instance, bHLH056 was sensitive to all four stresses and was upregulated distinctly under the four stresses. The orthologue of *bHLH056* in Arabidopsis is *BEE1 (AtbHLH044)* (Fig. 1), which has been functionally characterized in previous reports. At low temperatures, BEE1 is a positive regulator of flavonoid accumulation[49], 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[29]. However, we found the transcription of bHLH056 was notably upregulated (>10-fold) under ABA treatment. More interestingly, bHLH042, it is an orthologue of BEE2 (Fig. S1), was distinctly repressed by ABA (approximately 4-fold). These results suggested that bHLH056 and bHLH042 may be 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[41]. There are two orthologues of bHLH027 in rice, one ortholog is *OslICE2/OsbHLH001*, is induced by salt stress, and its overexpression can enhanced the tolerance to freezing and salt stress[44, 45]. *OslICE1/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 biosynthesis[43]. These results implied bHLH027 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[42, 50], implying the *CpbHLH062* may involved the cold stress. 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 [51], implying *CpbHLH053* may have the similar function with the *AtbHLH129* in the process of ABA response.

We should also noticed a few bHLHs that showed distinct increases or decreases in their mRNA levels under different treatments while these bHLHs' orthologues have not been reported in previous study. For instance, *CpbHLH050* is notably upregulated (>10-fold) under PEG treatment. *CpbHLH046* is upregulated by PEG reatment, 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[6, 18-25]. However, the bHLH TF family has not previously been reported in papaya (*Carica papaya L.*). In this paper, the number of 73 predicted *bHLH* genes were identified by bioinformatics methods in papaya. This TF family seemed to be one of the moderately sized families compared with other plant species, which might be because the papaya has a relatively small reference genome, the size is only 372 Mb[52]. Gene organization plays an important role in gene evolution. In this study, we found that *bHLH* genes showed diversities in their number introns, ranging from 0 to 10 (Additional file 1). This result implied these genes may have undergone a number of genetic evolution events, and the genes in different subfamily may have different functions[25]. Most *CpbHLH*s in the same subfamily share similar gene structures and protein motifs according to the analysis of exon/intron organizations and motif compositions (Fig. 2, 3 and 4), indicating the functions of encoded proteins in each subfamily were stable. However, the conserved motif analysis showed that some bHLHs, which are mainly distributed in eight subfamilies (including α a, α b, α f, α a, α b, α , α b and α a subfamilies) from the phylogenetic tree, harbor another conserved motif (motif 3) with a length of 36 amino acids (Fig. 2 and 4), indicating these proteins may have additional functions.

Promoter *cis*-acting regulatory element analysis showed the *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 IAA (16.79%), suggesting these phytohormones may play important roles in the regulation of papaya growth and development (Fig. 5 and Additional file 3). In addition, the promoter *cis*-acting involved in the abscisic acid responsiveness analysis is consistent with the qRT-PCR results (Additional file 3 and Table

2), showing four genes (*HLH020/-027/-053/-056*) involved in abscisic acid response. Another two genes (*HLH020/-062*) were also identified that were involved in abscisic acid response by GO annotation analysis and qRT-PCR (Additional file 5 and Table 2). We also identified a large number of *cis*-acting elements in *bHLH* genes that may respond to drought (MBS, 8.44%), which is also consistent with the qRT-PCR (including seven genes: *bHLH027/-050/-056/-011/-068/-042/-053*). Other genes also had important elements, including LTR, TC-rich and WUN-motifs, which reflected plant responses to low temperatures, defense stresses and wound-responsiveness, respectively. These results implied *bHLH* genes may have a wide range of functions in papaya growth, disease resistance, and response to environmental conditions.

Many studies have shown that *bHLH* genes are involved in responses to various abiotic and biotic stresses. We also 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* (*OslCE2*) and *OsbHLH002* (*OslCE1*) were clustered within one clade. *AtbHLH116* (ICE1), *AtbHLH33* (ICE2), *OsbHLH001* (*OslCE2*) and *OsbHLH002* (*OslCE1*) have been reported function in chilling stress in Arabidopsis and rice[41-45]. And the transcripts of *CpbHLH027*, *CpbHLH062* were increased under chilling stress in this study, implying *CpbHLH027* and *CpbHLH062* may be also involved in the process of chilling stress. The orthologue of *bHLH056* in Arabidopsis is *BEE1* (AT1G18400), which is a positive regulator of flavonoid accumulation[49]. *BEE1*, *BEE2* and *BEE3* are functionally redundant positive regulators of BR signaling, and their transcription is repressed by ABA in Arabidopsis[29]. However, we found the transcription of *bHLH39* is notably upregulated (>10-fold) under ABA treatment rather than downregulated in this study. These results imply *bHLH056* may involved in the process of ABA stress but has different function compared to Arabidopsis. We should also noticed a few candidate *bHLHs* that showed distinct increases or decreases in their mRNA levels under different treatments while these *bHLHs'* orthologues have not been reported in previous study. For instance, *CpbHLH050*, *CpbHLH020*, *CpbHLH046* and *CpbHLH053*. These findings provide important information necessary for further functional research on the *bHLH* family in papaya.

Conclusions

In conclusion, our study provided comprehensive information on the *bHLH* family in papaya, including phylogenetic relationships, gene structures, protein motifs, sequence features, promoter *cis*-elements, GO annotations and gene expression profiles responsible for abiotic stress. These results of *bHLH* TFs will help to build a solid foundation for future exploration and potential improvements in abiotic stress resistance, growth and development of papaya.

Methods

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

Papaya (*Carica papaya* L.) bHLH protein sequences were downloaded from the Plant TFDB V4.0 database[53]. 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 bHLH proteins were predicted by ProPAS[54]. The genomic sequences, ID numbers and coding sequences (CDS) corresponding to each predicted *bHLH* 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 *bHLH* genes were analyzed by Gene Structure Display Server (GSDS) v2.0[55].

Phylogenetic tree building, motif identification and multiple sequence alignment

To investigate the phylogenetic relationship between bHLH proteins, protein sequences of papaya were pre-aligned using HMM align [56] 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[3], the identified bHLH domains were later aligned using MAFFT v7.305b[57] with default settings. Phylogenetic tree was constructed based on the neighbor-joining method using FastTree v2.1.11[58] 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 bHLH TFs was generated by MAFFT v7.305b[57] with default settings.

The online tool Multiple EM for Motif Elicitation (MEME, version 5.02) was used to search for conserved motifs among the bHLH proteins (<http://meme-suite.org/tools/meme>) by uploading the protein sequences of the papaya bHLH superfamily. 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

The upstream 2000 bp genomic DNA sequences of the *bHLH* genes were downloaded and submitted to PlantCARE [59] to predict the putative *cis*-elements. The full-length protein sequences of papaya bHLH were blasted against Arabidopsis proteins with default parameters. The best hits were submitted to AgriGOv2.0 (<http://systemsbiology.cau.edu.cn/agriGOv2/>) for GO annotation[60]. 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 ‘YiChiGua’ 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 salt, drought and ABA treatments, seedlings were treated

in MS liquid medium containing 200 mM NaCl, 25% PEG6000 (to mimic drought stress) and 100 μ M ABA for 2 hours respectively, and then the roots were collected. For cold treatment, seedlings were subjected to 4°C for 2 hours 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 PrimeScript™ 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 (TBP2) amplification was used as an internal control[61]. The qRT-PCR reactions used at least three biological replicates, and each biological repeat had at least three technical replicates. Gene-specific primers for qRT-PCR of the 22 *bHLH* genes were designed based on the CDSs of the *bHLH* genes using Primer Premier 5.0 (Additional file 6). The relative expression levels of each gene were calculated using the $2^{-\Delta\Delta CT}$ method.

Abbreviations

TFs: transcription factors; bHLH: the basic helix-loop-helix; GO: gene ontology; qRT-PCR: quantitative real-time PCR; ORF: opening reading frame; PI: the theoretical isoelectric points; CDS: coding sequences; GSDS: gene structure display server; MEME: multiple EM for motif elicitation; SMART: a simple modular architecture research tool; MBS: MYB binding site involved in drought-inducibility; LTR: low temperature response elements; TC-rich: defense and stress responsive elements; WUN-motifs: wound-responsive elements; GA: gibberellin; IAA: auxin; MeJA: methyl jasmonate; SA: salicylic acid; ABA: abscisic acid; ARE: anaerobic induction elements; BR: brassinosteroid.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Putative bHLH transcription factors were downloaded from the PlantTFDBv2.0 (<http://planttfdb.cbi.pku.edu.cn/>). The genomic sequences, ID numbers and coding sequences (CDS) corresponding to each predicted *bHLH* gene were obtained from the Phytozome database

(<https://phytozome.jgi.doe.gov/pz/portal.html>). The online tool Multiple EM for Motif Elicitation (MEME, version 5.02) was used to search for conserved motifs among the bHLH proteins (<http://meme-suite.org/tools/meme>) by uploading the protein sequences of the papaya bHLH superfamily.

Competing interests

The authors declare that they have no competing interest.

Funding

This work was supported by the Presidential Foundation of Guangdong Academy of Agricultural Sciences, China (Grant No.201820), the Public Welfare and Capacity Building Research Funds of Guangdong Science and Technology Program, China (Grant No.2017A030303037), and the Major Project of Guangzhou Industry-university-research Collaboration Innovation Funds, China (Grant No. 201704020031). These fundings only provided financial support for this research project, but were not involved in study design, sample collection, data analysis, and manuscript preparation.

Authors' contributions

YW perceived and planned the study and MY performed most of the experiments and all of bioinformatic analysis. HY, RK, CZ and BH helped collected the samples and extracted total RNAs for qPCR. MY wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Authors' Information

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, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, Guangdong, China.

References

1. Agarwal PK, Agarwal P, Reddy MK, Sopory SK: **Role of DREB transcription factors in abiotic and biotic stress tolerance in plants.** *Plant cell reports* 2006, **25**(12):1263-1274.
2. Feller A, Machemer K, Braun EL, Grotewold E: **Evolutionary and comparative analysis of MYB and bHLH plant transcription factors.** *The Plant journal : for cell and molecular biology* 2011, **66**(1):94-116.
3. Pires N, Dolan L: **Origin and diversification of basic-helix-loop-helix proteins in plants.** *Molecular biology and evolution* 2010, **27**(4):862-874.
4. Jones S: **An overview of the basic helix-loop-helix proteins.** *Genome Biology* 2004, **5**(6):226.

5. Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martinez-Garcia JF, Bilbao-Castro JR, Robertson DL: **Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae.** *Plant physiology* 2010, **153**(3):1398-1412.
6. Toledo-Ortiz G: **The Arabidopsis Basic/Helix-Loop-Helix Transcription Factor Family.** *The Plant Cell Online* 2003, **15**(8):1749-1770.
7. Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, Guo J, Liang W, Chen L, Yin J *et al*: **Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis.** *Plant physiology* 2006, **141**(4):1167-1184.
8. Atchley WR, Fitch WM: **A natural classification of the basic helix-loop-helix class of transcription factors.** *Proceedings of the National Academy of Sciences* 1997, **94**(10):5172-5176.
9. Atchley WR, Terhalle W, Dress A: **Positional Dependence, Cliques, and Predictive Motifs in the bHLH Protein Domain.** *Journal of Molecular Evolution* 1999, **48**(5):501-516.
10. Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC: **The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity.** *Molecular biology and evolution* 2003, **20**(5):735-747.
11. Morgenstern B, Atchley WR: **Evolution of bHLH transcription factors: modular evolution by domain shuffling?** *Molecular Biology & Evolution* 1999, **16**(12):1654-1663.
12. [Ledent V](#), [Vervoort M](#): **The Basic Helix-Loop-Helix Protein Family: Comparative Genomics and Phylogenetic Analysis.** *Genome research* 2001, **11**(5):754-770.
13. Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, Coornaert D, Degnan BM, Vervoort M: **Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics.** *BMC evolutionary biology* 2007, **7**:33.
14. Vervoort M, Ledent V: **The evolution of the neural basic Helix-Loop-Helix proteins.** *TheScientificWorldJournal* 2001, **1**:396-426.
15. Amoutzias GD, Robertson DL, Bornberg-Bauer E: **The evolution of protein interaction networks in regulatory proteins.** *Comparative and functional genomics* 2004, **5**(1):79-84.
16. Stevens JD, Roalson EH, Skinner MK: **Phylogenetic and expression analysis of the basic helix-loop-helix transcription factor gene family: genomic approach to cellular differentiation.** *Differentiation; research in biological diversity* 2008, **76**(9):1006-1022.
17. Buck MJ, Atchley WR: **Phylogenetic analysis of plant basic helix-loop-helix proteins.** *J Mol Evol* 2003, **56**(6):742-750.
18. Gao C, Sun J, Wang C, Dong Y, Xiao S, Wang X, Jiao Z: **Genome-wide analysis of basic/helix-loop-helix gene family in peanut and assessment of its roles in pod development.** *PloS one* 2017, **12**(7):e0181843.
19. Mao K, Dong Q, Li C, Liu C, Ma F: **Genome Wide Identification and Characterization of Apple bHLH Transcription Factors and Expression Analysis in Response to Drought and Salt Stress.** *Frontiers in plant science* 2017, **8**:480.

20. Sun H, Fan HJ, Ling HQ: **Genome-wide identification and characterization of the bHLH gene family in tomato.** *BMC genomics* 2015, **16**:9.
21. Wang R, Zhao P, Kong N, Lu R, Pei Y, Huang C, Ma H, Chen Q: **Genome-Wide Identification and Characterization of the Potato bHLH Transcription Factor Family.** *Genes* 2018, **9**(1).
22. Zhang C, Feng R, Ma R, Shen Z, Cai Z, Song Z, Peng B, Yu M: **Genome-wide analysis of basic helix-loop-helix superfamily members in peach.** *PloS one* 2018, **13**(4):e0195974.
23. Wang P, Su L, Gao H, Jiang X, Wu X, Li Y, Zhang Q, Wang Y, Ren F: **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* 2018, **9**:64.
24. Geng J, Liu JH: **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* 2018, **69**(10):2677-2692.
25. Cheng X, Xiong R, Liu H, Wu M, Chen F, Hanwei Y, Xiang Y: **Basic helix-loop-helix gene family: Genome wide identification, phylogeny, and expression in Moso bamboo.** *Plant physiology and biochemistry* 2018, **132**:104-119.
26. Roig-Villanova I, Bou-Torrent J, Galstyan A, Carretero-Paulet L, Portolés S, Rodríguez-Concepción M, Martínez-García JF: **Interaction of shade avoidance and auxin responses: a role for two novel atypical bHLH proteins.** *EMBO JOURNAL* 2007, **26**(22):4756-4767.
27. Leivar P, Monte E, Al-Sady B, Carle C, Storer A, Alonso JM, Ecker JR, Quail PH: **The Arabidopsis phytochrome-interacting factor PIF7, together with PIF3 and PIF4, regulates responses to prolonged red light by modulating phyB levels.** *The Plant cell* 2008, **20**(2):337-352.
28. Lee S, Lee S, Yang KY, Kim YM, Park SY, Kim SY, Soh MS: **Overexpression of PRE1 and its homologous genes activates Gibberellin-dependent responses in Arabidopsis thaliana.** *Plant & cell physiology* 2006, **47**(5):591-600.
29. Friedrichsen DM, Nemhauser J, Muramitsu T, Maloof JN, Alonso J, Ecker JR, Furuya M, Chory J: **Three redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth.** *Genetics*, 2002, **162**(3):1445.
30. Feng Y, Xu P, Li B, Li P, Wen X, An F, Gong Y, Xin Y, Zhu Z, Wang Y, Guo H: **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* 2017, **114**(52):13834.
31. Farquharson KL: **A Domain in the bHLH Transcription Factor DYT1 Is Critical for Anther Development.** *The Plant cell* 2016, **28**(5):997-998.
32. Bernhardt C, Lee MM, Gonzalez A, Zhang F, Lloyd A, Schiefelbein J: **The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root.** *Development* 2003, **130**(26):6431-6439.
33. Jiang Y, Yang B, Deyholos MK: **Functional characterization of the Arabidopsis bHLH92 transcription factor in abiotic stress.** *Molecular genetics and genomics : MGG* 2009, **282**(5):503-516.

34. Liu W, Tai H, Li S, Gao W, Zhao M, Xie C, Li WX: **bHLH122 is important for drought and osmotic stress resistance in Arabidopsis and in the repression of ABA catabolism.** *The New phytologist* 2014, **201**(4):1192-1204.
35. Niu X, Guan Y, Chen S, Li H: **Genome-wide analysis of basic helix-loop-helix (bHLH) transcription factors in Brachypodium distachyon.** *BMC genomics* 2017, **18**(1):619.
36. Song Y, Liu HD, Zhou Q, Zhang HJ, Zhang ZD, Li YD, Wang HB, Liu FZ: **High-throughput sequencing of highbush blueberry transcriptome and analysis of basic helix-loop-helix transcription factors.** *Journal of Integrative Agriculture* 2017, **16**(3):591-604.
37. Chen YY, Li MY, Wu XJ, Huang Y, Ma J, Xiong AS: **Genome-wide analysis of basic helix-loop-helix family transcription factors and their role in responses to abiotic stress in carrot.** *Molecular Breeding* 2015, **35**(5):125.
38. Xu G, Guo C, Shan H, Kong H: **Divergence of duplicate genes in exon-intron structure.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(4):1187-1192.
39. Zhu E, You C, Wang S, Cui J, Niu B, Wang Y, Qi J, Ma H, Chang F: **The DYT1-interacting proteins bHLH010, bHLH089 and bHLH091 are redundantly required for Arabidopsis anther development and transcriptome.** *The Plant Journal* 2015, **83**(6):976-990.
40. Liu Y, Li J, Wei G, Sun Y, Lu Y, Lan H, Li C, Zhang S, Cao M: **Cloning, molecular evolution and functional characterization of ZmbHLH16, the maize ortholog of OsTIP2 (OsbHLH142).** *Biology open* 2017, **6**(11):1654-1663.
41. Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, Zhu JK: **ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis.** *Genes & development* 2003, **17**(8):1043-1054.
42. Fursova OV, Pogorelko GV, Tarasov VA: **Identification of ICE2, a gene involved in cold acclimation which determines freezing tolerance in Arabidopsis thaliana.** *Gene* 2009, **429**(1-2):98-103.
43. Zhang Z, Li J, Li F, Liu H, Yang W, Chong K, Xu Y: **OsMAPK3 Phosphorylates OsbHLH002/OsICE1 and Inhibits Its Ubiquitination to Activate OsTPP1 and Enhances Rice Chilling Tolerance.** *Developmental Cell* 2017, **43**(6):731-743.e735.
44. Deng C, Ye H, Fan M, Pu T, Yan J: **The rice transcription factors OsICE confer enhanced cold tolerance in transgenic Arabidopsis.** *Plant signaling & behavior* 2017, **12**(5):e1316442.
45. Li F, Guo S, Zhao Y, Chen D, Chong K, Xu Y: **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* 2010, **29**(9):977-986.
46. Bailey TL, Elkan C: **Fitting a mixture model by expectation maximization to discover motifs in biopolymers.** In: *1994*. 28-36.
47. Guo XJ, Wang JR: **Global identification, structural analysis and expression characterization of bHLH transcription factors in wheat.** *BMC plant biology* 2017, **17**(1):90.
48. Murre C, McCaw PS, Baltimore D: **A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins.** *Cell* 1989, **56**(5):777-783.

49. Petridis A, Döll S, Nichelmann L, Bilger W, Mock HP: **Arabidopsis thaliana G2-LIKE FLAVONOID REGULATOR and BRASSINOSTEROID ENHANCED EXPRESSION1 are low-temperature regulators of flavonoid accumulation.** *New Phytologist* 2016, **211**(3):912-925.
50. Kurbidaeva A, Ezhova T, Novokreshchenova M: **Arabidopsis thaliana ICE2 gene: phylogeny, structural evolution and functional diversification from ICE1.** *Plant science : an international journal of experimental plant biology* 2014, **229**:10-22.
51. Tian H, Guo H, Dai X, Cheng Y, Zheng K, Wang X, Wang S: **An ABA down-regulated bHLH transcription repressor gene, bHLH129 regulates root elongation and ABA response when overexpressed in Arabidopsis.** *Scientific reports* 2015, **5**:17587.
52. Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL: **The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus).** *Nature* 2008, **452**(7190):991.
53. Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G: **PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants.** *Nucleic acids research* 2017, **45**(D1):D1040-D1045.
54. Wu S, Zhu Y: **ProPAS: standalone software to analyze protein properties.** *Bioinformatics* 2012, **8**(3):167-169.
55. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G: **GSDS 2.0: an upgraded gene feature visualization server.** *Bioinformatics* 2015, **31**(8):1296-1297.
56. Eddy. SR: **Profile hidden Markov models.** *Bioinformatics* 1998, **14**(9):755-763.
57. [Kato K](#), [Misawa K](#), [Kuma K](#), [Miyata T](#): **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** *Nucleic Acids Resarch* 2002, **30**(14): 3059-3066.
58. Price MN, Dehal PS, Arkin AP: **FastTree: computing large minimum evolution trees with profiles instead of a distance matrix.** *Molecular biology and evolution* 2009, **26**(7):1641-1650.
59. [Lescot M](#), [Déhais P](#), [Thijs G](#), [Marchal K](#), [Moreau Y](#), [Van de Peer Y](#), [Rouzé P](#), [Rombauts S](#): **PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences.** *Nucleic acids research* 2002, **30**(1):325-327.
60. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z: **agriGO v2.0: a GO analysis toolkit for the agricultural community.** *Nucleic acids research* 2017, **45**(W1):W122-W129.
61. Zhu X, Li X, Chen W, Chen J, Lu W, Chen L, Fu D: **Evaluation of new reference genes in papaya for accurate transcript normalization under different experimental conditions.** *PloS one* 2012, **7**(8):e44405.

Additional Files

Additional file 1: Detailed information of *CpbHLH* genes and *CpbHLH* proteins.

Additional file 2: bHLH subfamily members of papaya, rice and Arabidopsis.

Additional file 3: Promoter analysis of *bHLH* genes in papaya.

Additional file 4 and Additional file 5: GO annotation of bHLH proteins.

Additional file 6: Primers used for qRT-PCR in this study.

Table S1: The information of 32 predicted proteins from the PlantTFDBv2.0 that were excluded by Interproscan and SMART.

Table S2: Consensus sequences of bHLH domains in papaya, rice and Arabidopsis.

Figures

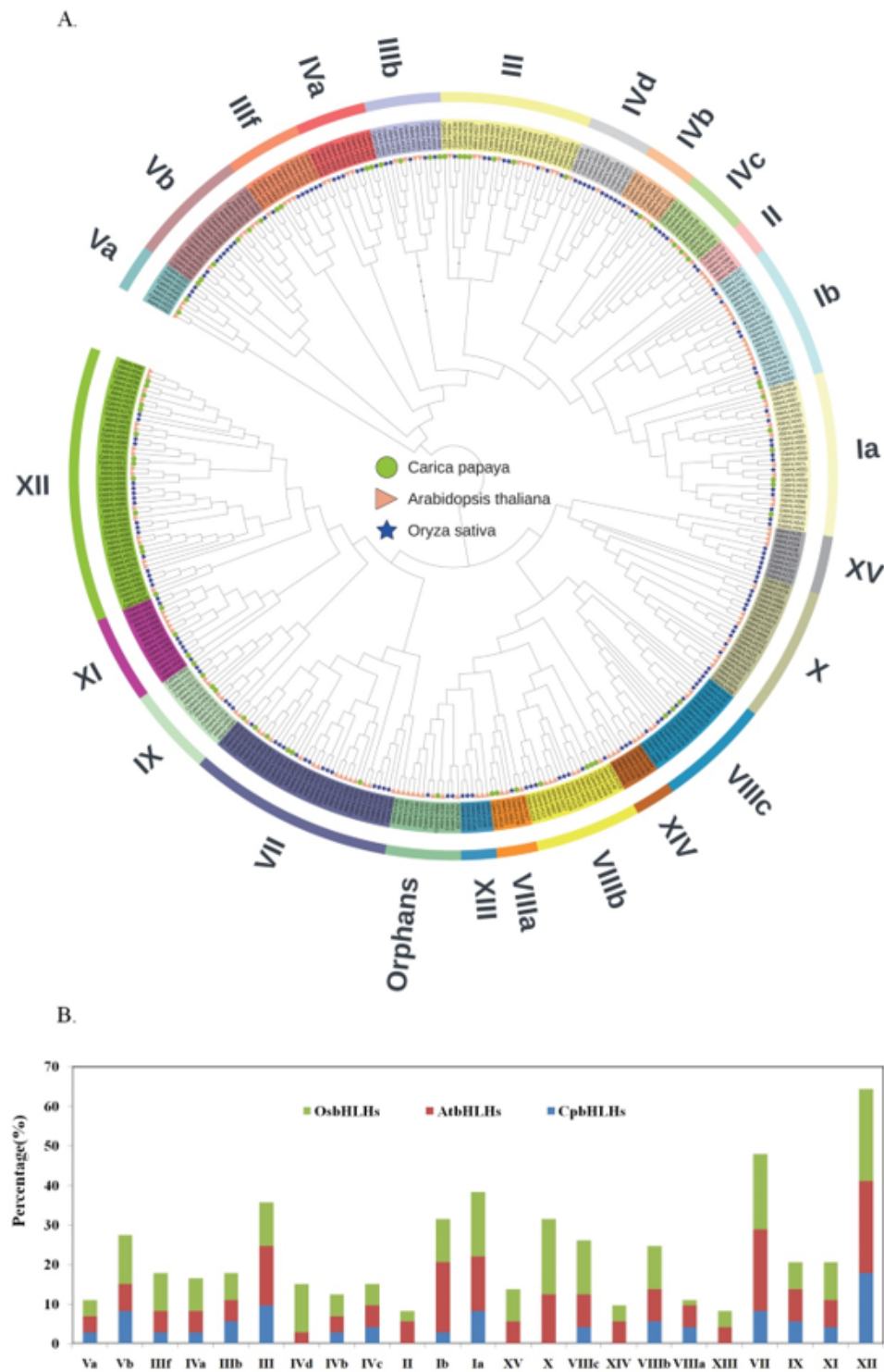


Figure 1

Phylogenetic and family members analysis of bHLHs from papaya, rice and Arabidopsis. (A) The 73 CpbHLHs are clustered into 18 subfamilies. Phylogenetic tree was constructed based on the neighbor-joining method. Bootstrapping with 1000 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. Green: OsbHLHs, red: AtbHLHs, Blue: CpbHLHs.

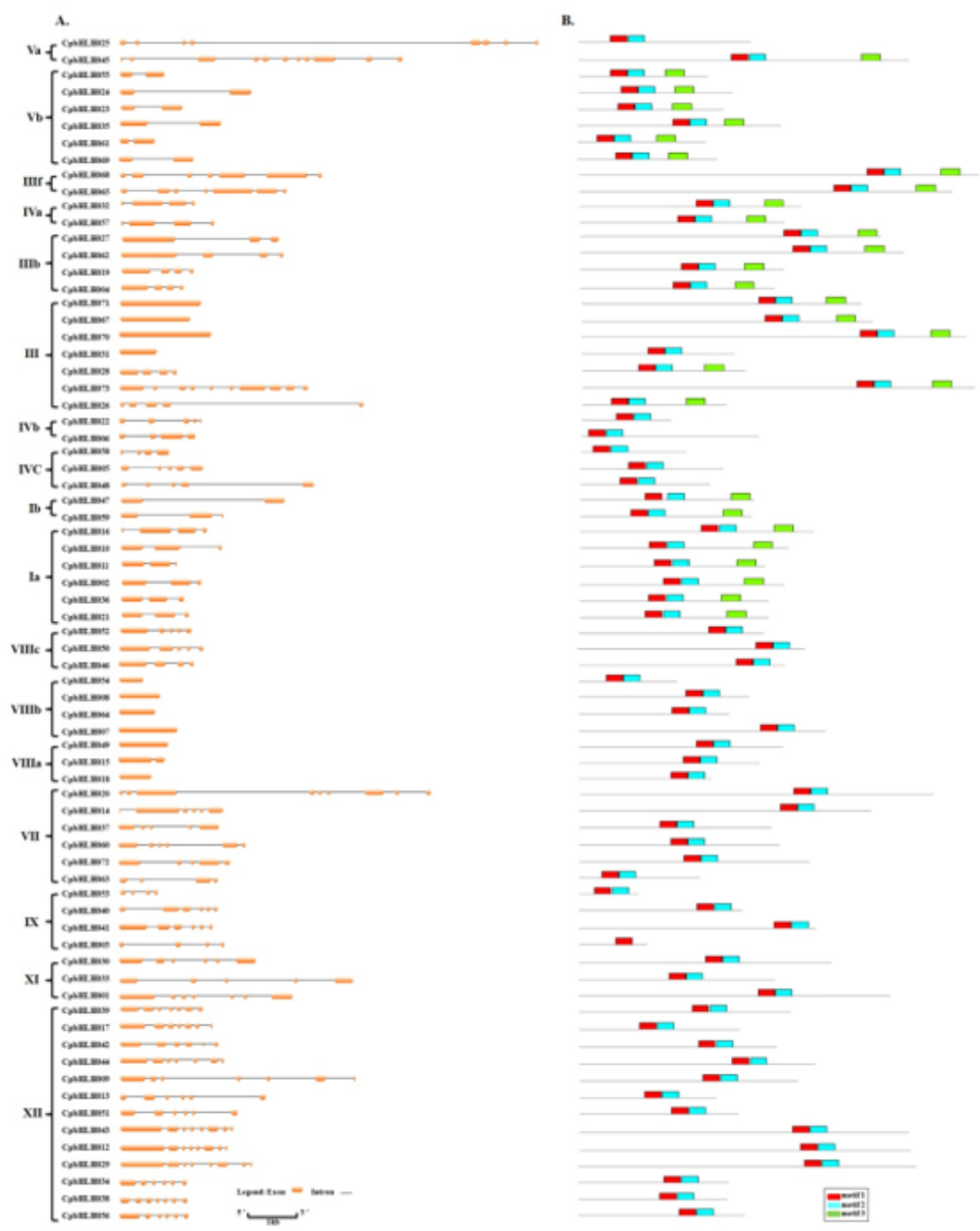


Figure 2

Gene structure and motif distribution of the papaya bHLH family. (A) Exon-intron organization of bHLH genes. Exons and introns are presented as filled orange sticks 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 (blue block) and motif 3 (green block).

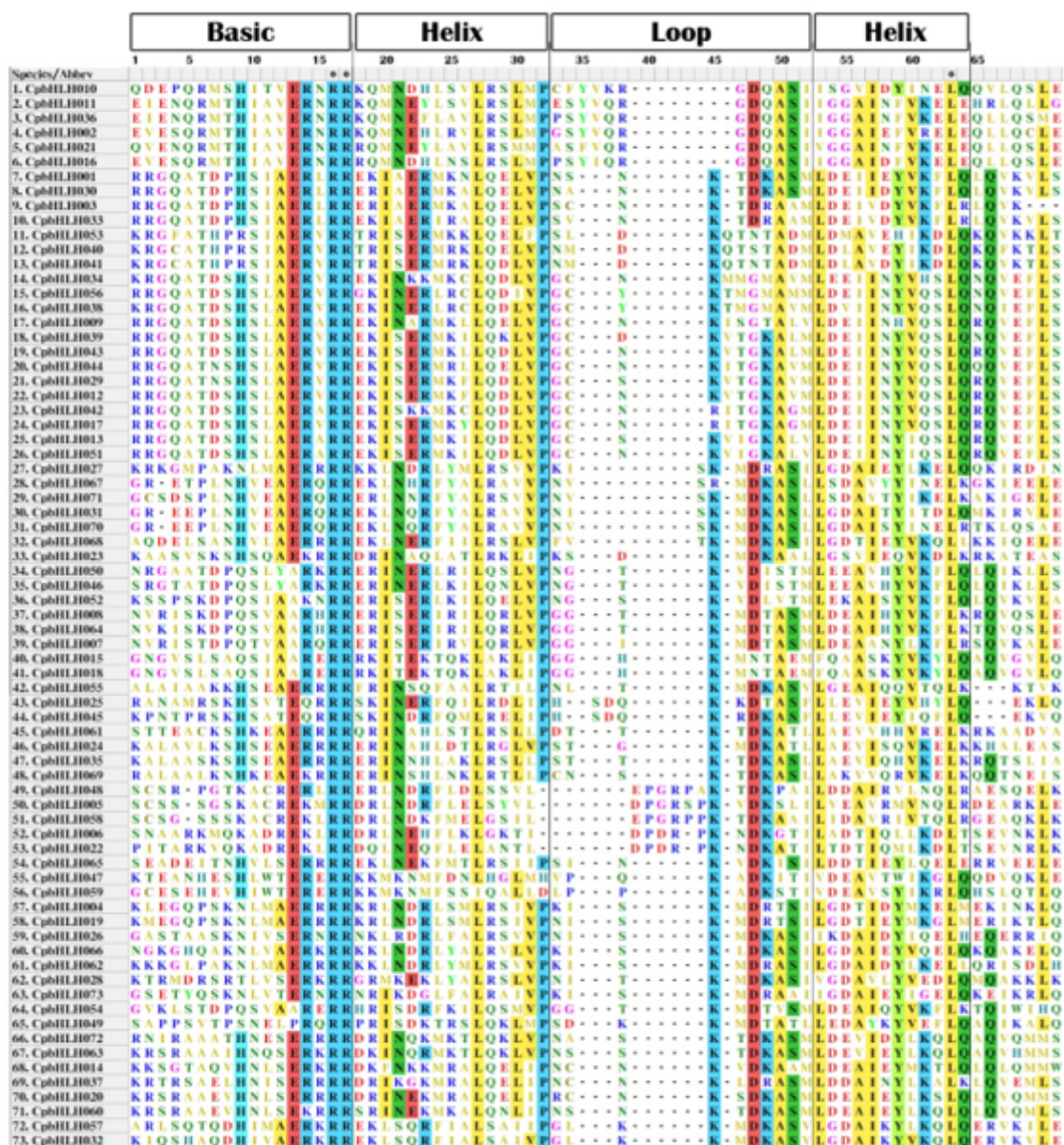


Figure 3

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

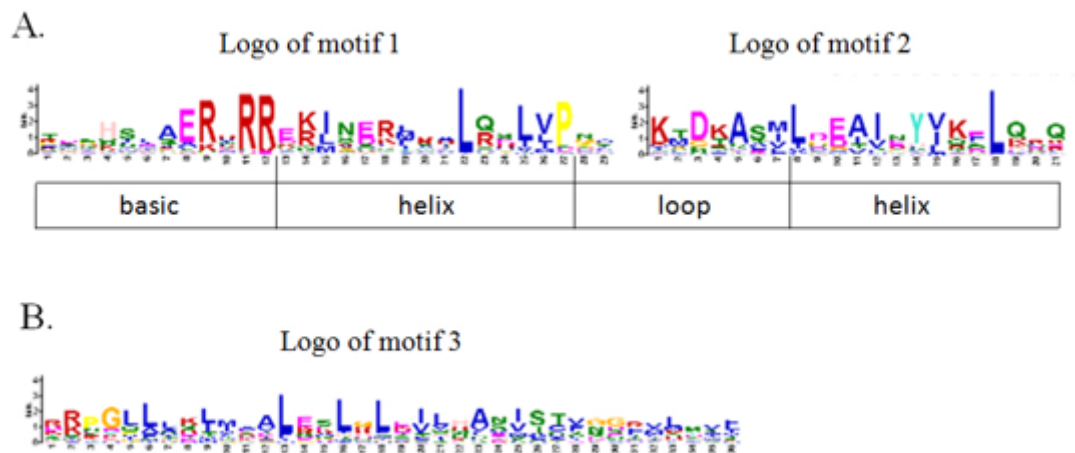


Figure 4

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.

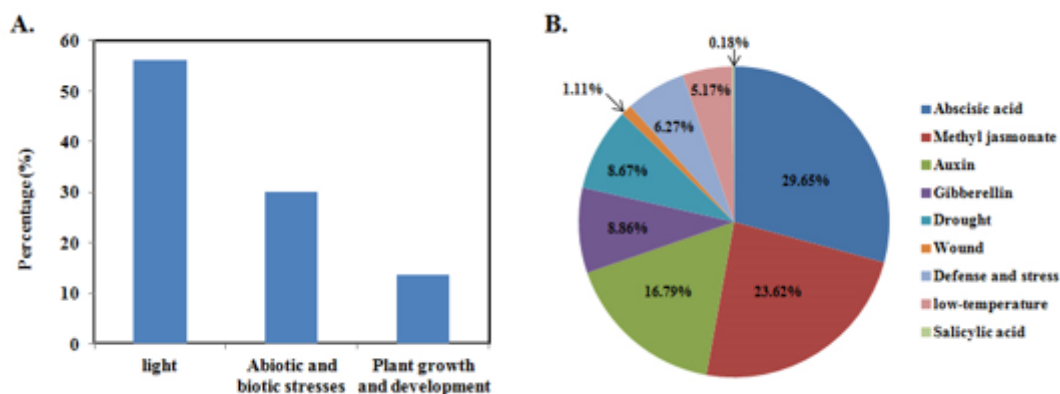


Figure 5

Cis-acting element analysis of the promoter of bHLH genes. (A) Percentage of total cis-acting elements in the promoter region of bHLH genes. (B) The percentage of each cis-acting element in the abiotic and biotic stresses categories.

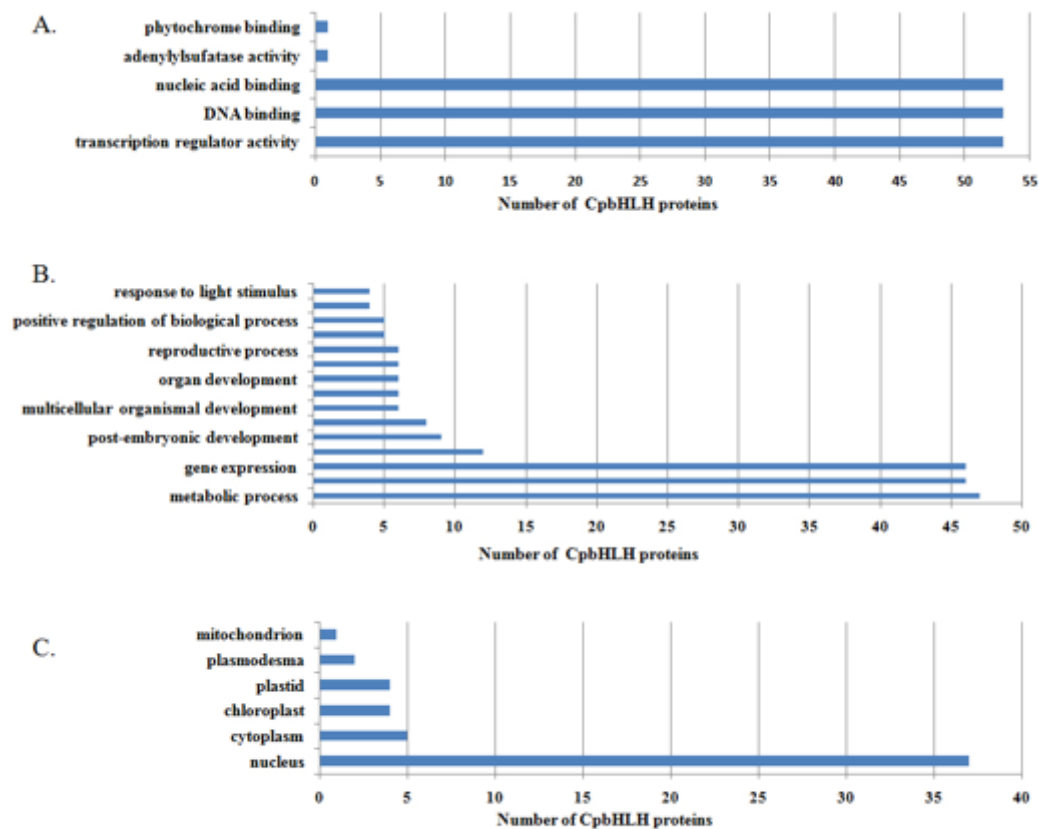


Figure 6

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

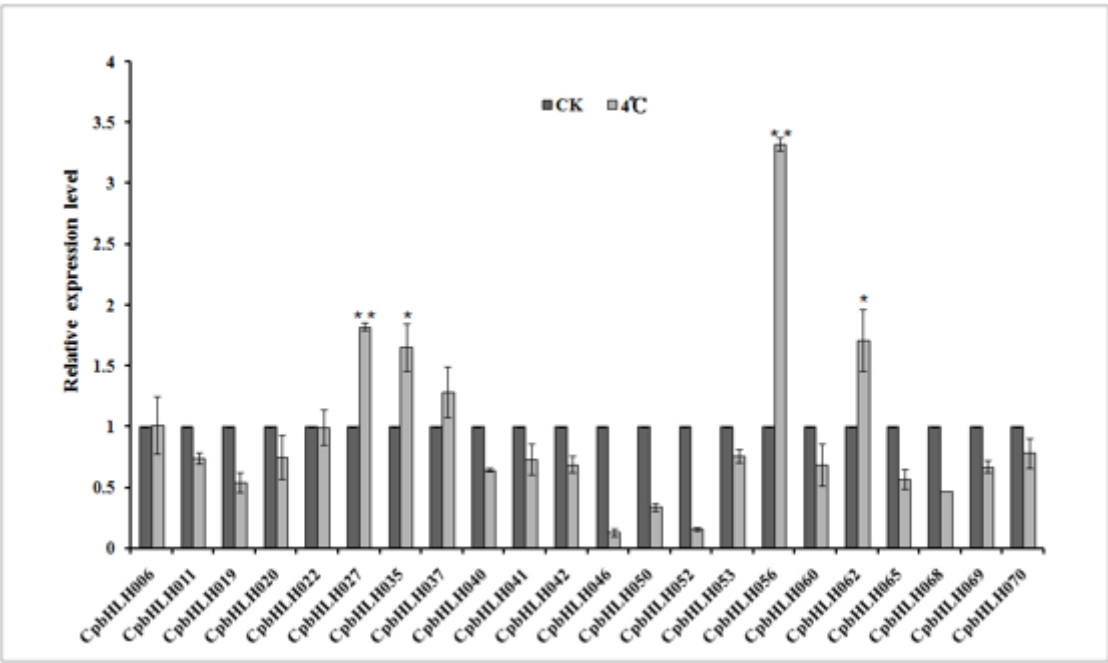


Figure 7

Quantitative RT-PCR analysis of 22 selected bHLH 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.

Supplementary Files

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

- [TableS2.xlsx](#)
- [TableS1.xlsx](#)
- [Additionalfile2.xlsx](#)
- [Additionalfile1.xlsx](#)
- [Additionalfile6.xlsx](#)
- [Additionalfile5.xlsx](#)
- [Additionalfile4.xlsx](#)
- [Additionalfile3.xlsx](#)