Comparative analysis of bHLH transcription factors in five Rosaceae species, and expression analysis of *PbbHLHs* in response to drought stress in pear

Xiaobing Kou

Nanjing Agricultural University

Changlong Xiong

Nanjing Agricultural University

Danqi Wang

Nanjing Agricultural University

Yangyang Sun

Nanjing Agricultural University

Peng Wang

Nanjing Agricultural University

Juyou Wu

Nanjing Agricultural University

ShaoLing Zhang (✉ Slzhang@njau.edu.cn)

Nanjing Agricultural University

Research article

Keywords: Pear, Rosaceae, bHLH transcription factors, Evolution, Expression files, Drought stress

Posted Date: September 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-78164/v1>

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

[Read Full License](#)

Abstract

Background

The bHLH (basic helix-loop-helix) transcription factor family plays important roles in regulating plant growth and development. However, informations about bHLH in rosaceous fruit species are still limited.

Results

In this study, a total of 198 *PbbHLH* genes were identified in the pear, with 188 *bHLH* genes in apple, 129 *bHLH* genes in peach, 112 in strawberry and 122 in Chinese plum. These Rosaceae *bHLH* genes, plus the 150 *Arabidopsis thaliana* *bHLH* genes, were divided into 34 groups, which included one Rosaceae-specific group. Evolutionary pattern analysis showed that whole-genome duplication (WGD) and segmental duplication played critical roles in expansion of the *PbbHLH* gene family. Ks value indicated that the two WGD duplication events (a recent WGD and an ancient WGD event) lead to the expansion of *bHLH* gene family. Tissue expression analysis shows that *PbbHLHs* may have diversity functions in different tissues. Furthermore, eight up-regulated and seven down-regulated *PbbHLH* genes were identified as the candidate genes in response to drought stress.

Conclusion

A comprehensive analysis of *bHLH* genes was performed in five Rosaceae species. The phylogenetic, evolution and expression analyses of the *bHLH* gene family in pear will be meaningful for investigating the biological roles of *PbbHLH* genes.

Background

The basic helix-loop-helix (bHLH) proteins are an important superfamily of transcription factors, which are important regulators of plant development and physiological processes [1–3]. Each bHLH transcription factor contains a bHLH domain which is involved in binding DNA, including two functional regions, the basic region and the HLH region [4, 5]. The basic region is located at the N terminus of the bHLH domain, and it consists of approximately 15 amino acids, which typically include six basic residues [6, 7]. The HLH region contains two amphipathic α helices at the C-terminal end, with a loop region of variable sequence and length; the amphipathic α helices of two bHLH proteins can interact to form homodimeric or heterodimeric complexes, to promote protein-protein interactions [4, 8–10]. Certain conserved amino acids in the basic region determine recognition. The E-box (5'-CANNTG-3') is a consensus hexanucleotide sequence. There are different types of E-boxes which depend on the identity of the two central bases, of which the G-box (5'-CACGTG-3') is the most common type [1, 11]. In addition, flanking nucleotides outside the core have also been shown to play a role in determining binding specificity [6, 12].

Identification of the *bHLH* gene family in various species could facilitate our understanding of their evolution and functions. Compared to animals, only a small number of plant bHLH proteins have been identified, while systematic studies on *bHLH* genes has been performed in only a few plants, such as *Arabidopsis* (*Arabidopsis thaliana*) and rice (*Oryza sativa*) [13, 14]. In the *Arabidopsis* and rice genomes, which are two major model species for flowering plants, 162 and 131 *bHLH* genes have been identified, respectively [13, 14]. Previous reports have demonstrated that some plant bHLH proteins play important roles in plant growth and tolerance to drought. For example, *Arabidopsis AtMYC2* (bHLH) and *AtMYB2* (MYB) function as transcriptional activators in abscisic acid signaling [15]. The *Arabidopsis myc/bHLH* gene *ALCATRAZ* controls cell separation in fruit dehiscence [16]. *Os bHLH148*, a basic helix-loop-helix protein, interacts with OsJAZ proteins in a jasmonate signaling pathway, leading to drought tolerance in rice [17].

The Pear (*Pyrus × bretschneideri* Rehder) is one of the most important fruits in the world [18]. Sequencing of the pear genome was recently completed and published, and the genomes of other rosaceous species, apple (*Malus × domestica* Borkh.), peach (*Prunus persica* (L.) Batsch), wild strawberry (*Fragaria vesca* L.) and Chinese plum (*Prunus mume* Siebold & Zucc.) have also been sequenced [19–22], but little bioinformatic analysis on the *bHLH* gene family in the Rosaceae had been reported till now. The availability of the genome sequences of five Rosaceae fruit species provides us with an unprecedented opportunity to carry out genome-wide comparative analyses of *bHLH* gene family in the Rosaceae phylogenetic trees.

In this study, we first identified the *bHLH* gene family in the pear genome before executing a detailed evolutionary and expression pattern analysis, and providing an exhaustive genome analysis of the *bHLH* genes in the Rosaceae, as well as analyzing the expression profiles of PbbHLHs in pear different tissues and response to drought stress.

Results

Identification and classification of *bHLH* genes in the Rosaceae

To define the *bHLH* gene families in the Rosaceae, we used the HMMER3 software profile and BLASTP program to identify the members of *bHLH* genes in Pear and the three other Rosaceae species. SMART analysis was used to further discriminate and confirm the candidate bHLH proteins. Finally, a total of 198 *bHLH* genes were identified in Pear (Table S1), 129 in peach, 112 in strawberry and 122 in Chinese plum (Table S2), while 188 *MdbHLHs* of apple had been identified in a previous report [23]. The *bHLH* genes were named according to the method proposed for *Arabidopsis* [24]. Among these 198 candidate *PbbHLH* genes, 169 were mapped onto all but chromosomes 9 and 11 of the pear chromosomes, while 29 *PbbHLH* genes were located on the scaffold (Table S1 and Fig. S1). Six *PbbHLHs* were located on chromosomes 1, 4 and 12, twelve on chromosomes 10 and 17, 18; fourteen located in chromosomes 6 and 14; five on chromosomes 7 and 16, 13 on chromosome 2, ten on chromosome 3, 21 on chromosome 5, nine on chromosome 13, 25 on chromosome 15, but only three genes on chromosome 8.

Multiple sequence alignments, conserved residues in the bHLH domains and predicted DNA-binding ability

To examine sequence features of these Rosaceae bHLH domains, we performed multiple sequence alignments of the Rosaceae bHLH amino acid sequences and checked each alignment by hand. In the alignment, we identified 11 and 12 residues with at least 50% conservation in pear and apple bHLH sequences, respectively, and 13, 11 and 12 in peach, strawberry and Chinese plum, respectively. Among these conserved residues, six residues (Glu-13, Arg-16, Leu-27, Lys-36, Tyr-49 and Leu-53) (Fig. 1 and Table S3) were present in more than 75% of sequences, suggesting that these residues are extremely important for the function of bHLH proteins.

It is generally acknowledged that bHLH domains with at least five basic amino acids in the basic region are considered to bind DNA. A total of 167 DNA-binding proteins and 31 DNA non-binding proteins were identified in Pear, with 164, 109, 98 and 104 DNA-binding proteins identified in apple, peach, strawberry and Chinese plum, respectively, and 24, 20, 14 and 18 DNA non-binding proteins identified in apple, peach, strawberry and Chinese plum, respectively (Table 1). According to the presence or absence of residues Glu-13 and Arg-16 in the basic region, the DNA-binding bHLH proteins were further divided into two groups: E-box-binding proteins or non-E-box-binding proteins. Twenty-nine, 22, 12, 15 and 16 non-E-box-binding proteins were found in Pear, apple, peach, strawberry and Chinese plum, respectively. According to the presence or absence of the His-9 residue, the E-box-binding proteins can be further divided into two subgroups, G-box-binding proteins and non-G-box-binding proteins. A total of 113, 116, 73, 61 and 68 G-box-binding proteins were found in Pear, apple, peach, strawberry and Chinese plum, respectively (Table 1).

Intron/exon structure in the Rosaceae bHLH domains

To better analyze the intron distribution within the bHLH domain of all of the Rosaceae *bHLH* genes, we performed a multiple alignment analysis between all the Rosaceae bHLH coding sequences and genome sequences, and eleven different distribution patterns (designated I to XI), with intron numbers ranging from 0 to 3, were found within the *bHLH* domain. Our results showed that approximately 80% of Rosaceae *bHLH* genes (159 *PbbHLHs* (from Pear), 157 *MdbHLHs* (from apple), 101 *PpbHLHs* (from peach), 90 *FvbHLHs* (from wild strawberry) and 96 *PmbHLHs* (from Chinese plum)) have introns in their *bHLH* domains (Fig. 2). The most common pattern differed among the five species. The most common pattern in pear and peach was pattern VI (including one intron), while pattern I (including three introns) was the most common pattern in apple, strawberry and Chinese plum. The most common intron pattern in pear and peach (pattern VI) was the same as in *Arabidopsis*. Patterns I and VI were found to be the most common ones, present in most *bHLH* genes.

Phylogenetic analysis of the bHLH proteins

To examine the evolutionary relationships among Rosaceae bHLH proteins, and further predict the function of *PbbHLHs*, a neighbor-joining phylogenetic tree was conducted, using the alignment

sequences of 749 Rosaceae bHLH proteins and 150 *Arabidopsis* bHLH proteins. The phylogenetic tree showed that these bHLH proteins were categorized into 34 groups (Fig. 3). The smallest group was group 19, having only six members, while the largest group was group 30, containing 65 members. Interestingly, the phylogenetic tree showed that group 20 contained 33 bHLH proteins which were from only Rosaceae species, indicating the Rosaceae-specificity of those bHLH proteins from group 20.

Evolutionary pattern analyses of the Rosaceae bHLH gene family

Different patterns of gene duplication events have driven the evolution of gene families, including singleton duplication (SD), WGD duplication, tandem duplication (TD), proximal duplication (PD) or dispersed duplication (DSD) [25]. To understand whether gene duplication events (and, if so, what type) occurred during the evolution of the Rosaceae *bHLH* gene families, we detected the origins of *bHLH* genes in the five Rosaceae genomes using the *MCScanX* package. As shown in Fig. 4 and table 2, 55.6% (110) of the *bHLH* genes in pear and 48% (108) in apple were duplicated from WGD or segmental events, compared to 41.9% (54) in peach, 62.5% (70) in strawberry and 51.6% (63) in Chinese plum being duplicated as a result of dispersed duplication events. This may be due to the recent lineage-specific WGD events (30–45 MYA) occurred in pear and apple, whereas recent WGD events did not occur in peach, strawberry or Chinese plum. In addition, genome rearrangements, gene losses, and RNA- and DNA-based transposed gene duplications may contribute to the larger proportions of dispersed duplicates in peach, strawberry or Chinese plum. These results showed that WGD or segmental duplication and dispersed gene duplication played critical roles in the expansion of the *bHLH* gene family in the Rosaceae.

To investigate the potential evolutionary mechanisms of the *bHLH* gene family, we performed the paralogous relationships across the entire pear genome, using a local synteny-based method. We observed that most *PbbHLH* genes were distributed on 16 chromosomes with an uneven distribution, and 63 paired homologous relationships from segmental duplications appeared in pear (Fig. 5A and Table S4). Interestingly, Chromosomes 15 has the most duplicated genes (Fig. 4A), indicating that it plays important roles in duplication events of the *PbbHLH* gene family. Furthermore, take two homologous gene pairs duplicated in chromosomes 15, we performed an all-vs.-all local BLASTP based on a method similar to the one used for PGDD across the whole pear genome to identify synteny blocks. Highly Conserved synteny was observed in the regions, several of which contained over 100 syntenic gene pairs (Fig. 4B).

In addition, 158 paired orthologous relationships were found in pear and apple, 173 in pear and peach, 135 in pear and strawberry, and 130 in pear and Chinese plum (Fig S2.). The numbers of the orthologous relationships among pear, apple and peach were greater than those in pear, strawberry and Chinese plum, suggesting that *bHLH* genes in pear, apple and peach may have originated from a common ancestor.

Ka/Ks ratio and Ks value drive the selective pressure on the evolution of bHLH gene family

Ka/Ks ratio is used to explore the selective pressure on duplicated genes, the ratio being an indication of negative selection (where $Ka/Ks < 1$) or positive selection (where $Ka/Ks > 1$) [26,27]. Positive selection is associated with functional divergence. Our result showed that the Ka/Ks ratio of most *bHLH* gene pairs in Rosaceae was less than one (Fig 6A), suggesting these *bHLH* genes mainly experienced negative selection. However, several *bHLH* gene pairs exhibited $Ka/Ks > 1$, including three gene pairs in pear, 12 gene pairs in apple, one gene pair in peach and four gene pairs in strawberry (Fig 6A), suggesting positive selection possibly playing an important role in the functional divergence of these genes.

The synonymous (Ks) value is used to estimate the stage of evolution for the WGD or segmental duplication events. In our study, the paralogous gene pairs were used to calculate the Ks values for dating the evolutionary time of *bHLH* genes in Rosaceae. As shown in Fig. 6A, the Ks value of most *bHLH* duplicated genes in pear and apple ranged from 0.1-0.3, this consistent with a recent WGD duplication event in pear and apple. Interesting, the Ks values for duplicated gene pairs which the ratio of Ka/Ks was more than one almost ranged from 0.1 to 0.3 (Fig 6A), suggesting that the recent WGD event in pear and apple is an active stage of evolution for *bHLH* gene family. In peach, strawberry and Chinese plum, the Ks value of *bHLH* genes was singly ranged from 0.5-3, suggesting they might have duplicated from a more ancient duplication event. In addition, we characterized the mean Ks value of the *PbbHLH* gene pairs, showing that these *PbbHLH* gene pairs were distributed at the two Ks value peaks (Fig. 6B and Table S4), further suggesting that the recent WGD event and the ancient WGD event led to the expansion of *PbbHLH* gene family.

Expression profiles of *PbbHLH* genes in pear different tissues

To further understand the potential functions of the *PbbHLH* genes, we analyzed the expression profiles of *PbbHLH* genes in different tissues, including stem, leaf, bud, sepal, petal, ovary and fruit, based on the public RNA-seq data (Fig 7 and Table S5). The result of heatmap showed that the expression pattern of each *PbbHLH* gene varied greatly in the different tissues. However, we found that 93.9% of *PbbHLHs* were expressed in at least one tissue (stage) of pear. Among these *PbbHLH* genes, 48.5% of the genes (96 genes) were expressed in all tissues. Especially, *PbbHLH195* was highly expressed in all tissues, suggesting its comprehensive functions in pear development. Furthermore, some *PbbHLH* genes are highly expressed in a specific tissue of pear, with low expression in other tissues. For example, *PbbHLH15*, *18*, *70*, *84*, *89*, *93*, *113* and *131* are accumulated mainly in leaf, suggesting these genes may play roles in leaf development or immunity. In addition, *PbbHLH138*, *167* and *176* are strongly expressed in fruit, with low expression in other tissues, suggesting the potential function in pear fruit development. On the other hand, several *PbbHLH* genes have high level of transcripts in reproductive organs. For example, *PbbHLH13*, *25* and *100* show mainly higher expression in ovary, suggesting these *PbbHLH* genes may be necessary for reproductive growth. These above results demonstrated that *PbbHLH* genes may have different functions in tissue-dependent manner.

Expression profiles of *PbbHLHs* under drought stress

Plants have formed a set of mechanisms to response for stresses during long-term evolution. Previous reports have shown that several *bHLH* genes play important roles in response to abiotic stress. However, information is limited on *PbbHLH* genes response to drought stress in pear. To investigate *PbbHLH*s response to drought stress in pear, the expression profiles of *PbbHLH* genes which under drought treatment for 0, 1h, 3h, 6h and re-watering 24h were obtained from the transcriptome data sets previously reported (Fig. 8 and Table S6). Most *PbbHLH* genes (143) showed differential gene expression. Among them, the expression of 13 *PbbHLH* genes were significantly up-regulated, while 23 *PbbHLH* genes were significantly down-regulated. Furthermore, we respectively selected 8 genes (*PbbHLH26, 30, 77, 84, 113, 131, 152* and *166*), 7 genes (*PbbHLH1, 12, 15, 52, 58, 79*, and *92*) that were up-regulated and down-regulated by at least 10-fold after drought treatment for 6h and restore original expression after re-watering 24h, suggesting these genes may participate in drought stress response.

To verify the reliability of transcriptome data, quantitative real-time PCR (qRT-PCR) was further used to analyze the relative transcript abundance of six selected genes (*PbbHLH22, 30, 39, 52, 58* and *131*) (Fig 9). As shown in Fig. 7, the result of qRT-PCR is consistent with the transcriptome data. For example, the expression of *PbbHLH30, 39* and *131* continuously rise under drought stress at certain time points and then down regulated 24 h of recovery. *PbbHLH22, 52* and *58* displayed an obvious decrease in expression under drought stress for 1, 3 and 6h, and then up or down regulated 24 h of recovery. These results further suggested that these genes might play important roles in response to drought stress in pear.

Discussion

The *bHLH* transcription factors play important roles in plant growth and defense [15, 28–30]. In previous reports, systematic and comprehensive whole-genome analyses of *bHLH* family proteins have been studied in a number of species. For example, 150 members of the *bHLH* family were identified in *Arabidopsis* [31], with 230 *BrabHLH* genes identified in the Chinese cabbage genome [32], 159 *bHLH* genes in tomato [33] and 167 *bHLH* genes in rice [24]. Although the Pear genome sequence project has been completed [18], no detailed analysis of the *bHLH* gene family has been published, and most functions of the *bHLH* genes in pear are unclear. In this study, a total of 198 *bHLH* genes were identified in Pear (Table S1). In three other members of the Rosaceae, 129 *bHLH* genes were identified in peach, 112 in strawberry and 122 in Chinese plum (Table S2). Previous reports had shown that 188 *bHLH* genes were present in apple [23]. Notably, the number of *bHLH* genes in pear and apple is largely more than that in peach, strawberry and Chinese plum, this may be due to that pear and apple belong to the subfamily Maloideae, peach and Chinese plum belong to the Prunoideae, and strawberry belongs to Rosoideae [10], and the recent WGD duplication events occurred in pear and apple, suggesting that WGD duplication contributed to the expansion of *bHLH* genes in pear and apple. Additionally, all the *bHLH*s in *Arabidopsis* and five Rosaceae species were unevenly divided into 34 subfamilies (Fig. 3). Moreover, the group 20 contained *bHLH* proteins from only rosaceous species, indicating that the *PbbHLH*s in group 20 may play a species-specific role in the Rosaceae.

The basic region of the bHLH domain has been associated with DNA-binding activity [34, 35]. The basic region of the Rosaceae bHLH protein, which functions in DNA binding, contains 17 residues. According to our analysis, the Rosaceae bHLH proteins share similar DNA-binding mechanisms (Fig. 1). According to the presence or absence of residues Glu-13 and Arg-16 in the basic region, the DNA-binding bHLH proteins were further divided into two groups, the E-box-binding proteins and the non-E-box-binding proteins, and, based on the presence or absence of the His-9 residue, the E-box-binding proteins could be further divided into two subgroups, the G-box-binding proteins and the non-G-box-binding proteins. A total of 167, 164, 109, 98 and 104 DNA-binding proteins (including 113, 116, 73, 61 and 68, respectively, G-box-binding proteins) were found in the pear, apple, peach, strawberry and Chinese plum, respectively (Table 1). These results were similar to those obtained for the Arabidopsis bHLH proteins [31]. In the further, we need to study the functions of the basic region in the bHLH domain, this will be beneficial to better understand bHLH transcription factors.

To analyze intron distribution within the coding sequence of the bHLH domain in all the Rosaceae bHLH genes, we did alignment analysis with the coding and genome sequences using BLAST software, and found 11 different intron distribution patterns (designated I to XI) (Fig. 2). The intron number ranged from 0 to 3 within the coding sequence of the bHLH domain. In our study, 80.5% of the Rosaceae bHLH genes contained introns in their coding sequence of the bHLH domain (Fig. 2). Pattern I (containing three introns) was the most common pattern in apple, strawberry and the Chinese plum, and was similar to that in the Arabidopsis bHLHs [31], but it was only the second most common one in pear and peach; this finding was consistent with that of tomato bHLHs [33]. These results showed that intron sequences and distribution in different species were different, even though their bHLH domains were conserved.

Gene duplication events are thought to have occurred frequently during plant genome evolution, and gene families have evolved via several different duplication modes, including WGD or segmental duplication, tandem duplication or dispersed duplication [36, 37]. Gene duplications can lead to expansion of gene families and can generate new functions [38, 39]. In the present work, we found that segmental or WGD duplications were found in a majority of the *bHLH* genes in pear and apple, while dispersed duplications played important roles in the expansion of *bHLH* genes in peach, strawberry and the Chinese plum (Fig. 4). A recent WGD duplication events (30–45 MYA) occurred in pear and apple, while peach, strawberry, and Chinese plum did not experience such an event, indicating that *bHLH* genes were divergent in different plant species, and the recent WGD event may have caused the different numbers of *bHLH* genes in the members of the Rosaceae investigated (Table 2). This finding was consistent with the result that the number of *bHLH* genes in pear and apple was larger than in peach, strawberry and the Chinese plum.

Ks, as a proxy for evolution time, has been widely used for estimating the evolutionary dates of segmental duplication events [40, 41]. Based on the result of Ka/Ks ratio and Ks value, the duplicated bHLH genes mainly focus on two Ks value peaks in pear and apple, Ks ~ 0.15–0.3 and Ks ~ 1.5–1.8 (Fig. 5A and 5B), this consistent with a recent WGD duplication event in pear (Ks ~ 0.15–0.3) and apple (Ks ~ 0.2) occurred at 30–45 MYA, and the ancient WGD in pear (Ks ~ 1.5–1.8) and apple (Ks ~ 2)

resulted from a paleohexaploidization (γ) event that took place ~ 140 MYA. Interestingly, the Ka/Ks ratio of several *bHLH* genes was more than one during the recent WGD duplication event (Fig. 5A), suggesting that these *bHLH* genes may have undergone positive selection and generated new functions during evolution.

Several previous reports have shown that bHLH transcription factors play important roles in plant growth and development. For example, the bHLH transcription factor *SPATULA* controls *Arabidopsis* root growth by controlling the size of the root meristem [42]. The rice bHLH transcription factor *OsBLR1* regulates leaf angle via brassinosteroid signaling [43]. However, there is limited information available on the role of bHLH transcription factors in terms of controlling growth in pear. The expression patterns are able to provide essential information for studying gene function. Therefore, in the present study, we analyzed the expression level of *PbbHLH* genes in stem, leaf, bud, sepal, petal, ovary and fruit using transcriptome data (Fig. 6 and Table S5). 96 *PbbHLH* genes were expressed in all the tissues of pear. Among these genes, *PbbHLH195* showed the highest expression level, suggesting *PbbHLH195* may play important roles in growth and development of pear. Furthermore, the *PbbHLH* genes showed the different expression in different tissues and different stages of development, suggesting the diverse functions of *PbbHLH* genes in pear. For example, *PbbHLH167* and *PbbHLH138* both highly expressed in all fruit developmental stages of pear (Fig. S3 and table S7), and, with low expression in other tissues, suggesting that *PbbHLH167* and *PbbHLH138* may involve in pear fruit development. In the further, based on the expression profiles of *PbbHLHs* in different tissues of pear, we can further study the diverse functions of *PbbHLH* genes in pear.

To date, several reports have investigated the important roles of bHLHs in response to drought stress. For instance, the bHLH family member *ZmPTF1* regulates drought tolerance in maize by promoting root development and abscisic acid synthesis [44]. *AtbHLH68* transcription factor contributes to the regulation of ABA homeostasis and drought stress tolerance in *Arabidopsis thaliana* [45]. A novel bHLH transcription factor *PebHLH35* from *Populus euphratica* confers drought tolerance through regulating stomatal development, photosynthesis and growth [46]. In the present study, we found that 72.2% *PbbHLH* genes showed the different expression in response to drought stress based on the transcriptome data and qRT-PCR results in pear (Fig. 8; Fig. 9 and Table S6). Among them, 15 *PbbHLH* genes significantly up-regulated or down-regulated by at least 10-fold after drought treatment for 6 h and then restored original expression after re-watering 24 h, suggesting that these candidate genes may play essential roles in pear responses to drought stress. However, the molecular mechanism of *PbbHLH* genes tolerance to drought still needs further investigation.

Conclusions

In the present study, we conducted a comparative analysis of *bHLH* gene family in five Rosaceae species. 749 *bHLH* genes were identified in five Rosaceae species, among them, 198 belonged to the pear. We divided these *bHLH* genes into 34 subfamilies. Evolutionary analysis suggested that the recent WGD (30–45 MYA) may drive the expansion of *bHLH* genes in pear and apple. The transcriptome data and qRT-PCR

analysis suggested that *bHLH* genes might play important roles in pear development and drought stress response. Our work provides useful information for further potential functional studies in revealing their regulation mechanism.

Materials And Methods

Identification of *bHLH* genes in the Rosaceae

The Pear (*Pyrus × bretschneideri*) genome sequence was downloaded from the pear genome project (<http://peargenome.njau.edu.cn/>). The genome sequences of peach and strawberry were downloaded from Phytozome (<http://phytozome.jgi.doe.gov/pz/portal.html#>), and the Chinese plum genome sequence was downloaded from the *Prunus mume* Genome Project (<http://prunusmumegenome.bjfu.edu.cn/index.jsp>). The sequences of *Arabidopsis* (*A. thaliana*) *bHLH* genes were retrieved from The Arabidopsis Information Resource (TAIR; <http://www.arabidopsis.org/>) [47]. To identify the members of the *bHLH* gene family in the Rosaceae, two approaches were used in this research. Firstly, the HLH domain (<http://pfam.xfam.org>, PF00010) was downloaded from the Pfam website (<http://pfam.xfam.org/family>) [48], and it was used as a query to perform BLAST searches in HMMER3 software package [49] against the genome databases of four rosaceous species (pear, peach, strawberry and Chinese plum) [18-22]. Additionally, using the sequence of *Arabidopsis* bHLH as a query [31], the pear, peach, strawberry and Chinese plum bHLH sequences were blasted. Moreover, to further confirm all candidate *bHLH* genes, the predicted bHLH domain was detected by the Simple Modular Architecture Research Tool (SMART) program (<http://smart.embl-heidelberg>).

Multiple sequence alignments

Multiple sequence alignments were aligned using ClustalX [50] and SMS programs (<http://www.bioinformatics.org/sms2/>) [51], and domains were identified using the InterProScan tool [52,53]. To obtain information on the intron/exon structure, the cDNA alignment of bHLH domain sequences was obtained, according to the amino acid sequence alignment, and the information on the intron distribution pattern and intron splicing phase were derived from the aligned cDNA sequences.

Phylogenetic analysis

Phylogenetic trees were constructed using the neighbor-joining (NJ) method with MEGA 7 [54]. The reliability of the trees obtained was tested using the bootstrap method with 1,000 replicates, and branches with fewer than 50% bootstrap values were rejected. NJ analysis was performed with the Poisson model.

Chromosomal location and duplication pattern analyses of bHLH genes in pear and four other Rosaceae species

The chromosomal location information on the *bHLH* genes was obtained from genome annotation documents. The data were then plotted using Circos software [55]. Analysis of synteny among the five

Rosaceae genomes was conducted according to duplication gene pairs, using a method similar to that developed for the Plant Genome Duplication Database (PGDD) (<http://chibba.agtec.uga.edu/duplication/>) [56]. The *MCScanX* program was used to identify the type of gene duplication in each *bHLH* gene family [57].

Calculating Ka and Ks of the *bHLH* gene family in pear

To explore whether positive selection drove the evolution of the *bHLH* gene family, the rates of nonsynonymous (Ka) and synonymous (Ks) substitutions in all paralog pairs were analyzed using the coding sequence (CDS) of *bHLH* gene paralogs in pear. The ratio of these two values (Ka/Ks) is an indication of purifying (Ka/Ks<1) or diversifying selection (Ka/Ks>1) [58]. *MCScanX* downstream analysis tools were used to annotate the Ka and Ks substitution rates of syntenic gene pairs. To date segmental duplication events, 63 consecutive *bHLH* gene pairs in pear were chosen to calculate the mean Ks. Calculating the dates of duplication events was carried out according to the previous reports [18,59].

Genome-wide expression analysis of *bHLH* gene family in pear

To investigate the expression profile of *bHLH* gene family members in pear different tissues, we utilized the transcriptome sequencing data presented in several earlier studies, while the expression data obtained from pollen was used to measure the expression level of *bHLH* genes. The RNA-seq data was downloaded from our center website (<http://peargenome.njau.edu.cn/>). The RPKM (reads per kilobase per million mapped reads) values were used to measure the gene expression levels. The log₂ transformed RPKM values were utilized to measure the expression level of the *PbbHLH* genes and to generate the heatmaps.

qRT-PCR

Total RNA was extracted using a Plant Total RNA Isolation Kit (FOREGENE, Chengdu, China) and genomic DNA contamination was removed by DNase I. Purified total RNA (3 µg) was used for reverse transcription with RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, USA). Subsequently, quantitative PCR (qPCR) was performed on 20-µL samples using SYBR Green (TaKaRa, Dalian, China) on an IQ5 real-time PCR machine (Bio-Rad, Hercules, CA, USA) according to the manufacturer's specifications. The thermal cycle was performed at 95°C for 5 min followed by 45 cycles of 95°C for 15 s each and 60°C for 15 s. The pear *TUB* gene was used as an internal control and relative expression levels were determined using the $2^{-\Delta\Delta CT}$ method as described previously [60]. Primers were designed using Primer 5 software [61] based on the target genes and are shown in Table S8.

Abbreviations

bHLH, basic helix–loop–helix; CDS, coding sequence; HMM, Hidden Markov Model; MEME, Multiple EM for Motif Elicitation; NJ, Neighbor-joining; qRT-PCR, quantitative real-time PCR; WGD, whole genome

duplication; TD, Tandem duplication; PD, Proximal duplication; DSD, Dispersed duplication; SD, Singleton duplication.

Declarations

Conflict of interest:

The authors declare no competing interests.

Authors' contributions

SLZ, JYW and XBK conceived the experimental design. XBK performed the experiments and data analyses. CLX and YYS contributed synteny analyses and the perl script, and configured some of the figures. PW revised the final manuscript. JYW and SLZ managed the experiments. All authors have read and approved the final manuscript.

Acknowledgements:

Computational resources supported by the Bioinformatics Center of Nanjing Agricultural University. This work was supported by the National Natural Science Foundation of China (31522048, 31772256, 31772276 and 31501715), the National Key Technology R&D Program of the Ministry of Science and Technology of China (2014BAD16B03-4), and the Fundamental Research Funds for the Central Universities (KYTZ201602 and KYLH201502-2).

References

1. Gabriela T-O (2003) The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell* 8.
2. Jie, Zhang, Bing, Liu, Mengshu, et al. (2015) The bHLH Transcription Factor bHLH104 Interacts with IAA-LEUCINE RESISTANT3 and Modulates Iron Homeostasis in *Arabidopsis*. *Plant Cell*.
3. Ledent V, Vervoort M (2001) The Basic Helix-Loop-Helix Protein Family: Comparative Genomics and Phylogenetic Analysis. *Genome Research* 11: 754-770.
4. Ferré-D'Amaré AR, Pognonec P, Roeder RG, Burley SK (1994) Structure and function of the b/HLH/Z domain of USF. *Embo Journal* 13: 180-189.
5. C M (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 5.
6. 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.

7. Morgenstern B, Atchley WR (1999) Evolution of bHLH transcription factors: modular evolution by domain shuffling? *Molecular Biology & Evolution*: 1654-1663.
8. Ellenberger T, Fass D, Arnaud M, Harrison SC (1994) Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes & Development* 8: 970-980.
9. C M (1989) Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* 3.
10. Bremer B, Manen JF (2000) Phylogeny and classification of the subfamily Rubioideae (Rubiaceae). *Plant Systematics and Evolution* 225: 43-72.
11. Robinson KA, Koepke JI, Murtaza K, Lopes JM (2000) A network of yeast basic helix–loop–helix interactions. *Nuclc Acids Research*: 4460-4466.
12. Zhou Q, Wang S, Anderson DJ (2000) Identification of a novel family of oligodendrocyte lineage-specific basic helix-loop-helix transcription factors. *Neuron* 25: 331-343.
13. Heim, M. A, Jakoby, M., Werber, et al. (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*.
14. Toledo-Ortiz G, Huq E, Quail PH (2003) The *Arabidopsis* Basic/Helix-Loop-Helix Transcription Factor Family. *Plant Cell* 15: 1749-1770.
15. Hiroshi A (2003) *Arabidopsis AtMYC2* (bHLH) and *AtMYB2* (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* 1.
16. Sarojam, Rajani, and, Venkatesan, Sundaresan (2001) The *Arabidopsis* myc/bHLH gene *ALCATRAZ* enables cell separation in fruit dehiscence. *Current Biology*.
17. Seo JS (2011) OsbHLH148, a basic helix-loop-helix protein, interacts with OsJAZ proteins in a jasmonate signaling pathway leading to drought tolerance in rice. *Plant J* 65.
18. Wu J, Wang Z, Shi Z, Zhang S, Zhang S (2012) The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res. Genome Research* 23.
19. Vladimir S (2011) The genome of woodland strawberry (*Fragaria vesca*). *Nature genetics* 2.
20. Riccardo V (2010) The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nature genetics* 10.
21. Initiative IPG (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature genetics* 5.
22. Zhang Q, Chen W, Sun L, Zhao F, Huang B, et al. (2012) The genome of *Prunus mume*. *Nature communications* 3: 1318.
23. Ke M, Qinglong D, Chao L, Changhai L, Fengwang M (2017) Genome Wide Identification and Characterization of Apple bHLH Transcription Factors and Expression Analysis in Response to Drought and Salt Stress. *Frontiers in Plant ence* 8.
24. Li X, Duan X, Jiang H, Sun Y, Tang Y, et al. (2006) Genome-Wide Analysis of Basic/Helix-Loop-Helix Transcription Factor Family in Rice and *Arabidopsis*. *Plant Physiology* 141: 1167.

25. Maher, C. (2006) Evolution of Arabidopsis microRNA families through duplication events. *Genome Research* 16: 510-519.
26. Hu T, Banzhaf W. Nonsynonymous to Synonymous Substitution Ratio k_a/k_s : Measurement for Rate of Evolution in Evolutionary Computation; 2008.
27. Laurence, Hurst (2002) The K_a/K_s ratio: diagnosing the form of sequence evolution. *Trends in Genetics*.
28. Friedrichsen D (2002) Three Redundant Brassinosteroid Early Response Genes Encode Putative bHLH Transcription Factors Required for Normal Growth. *Genetics* 162: 1445.
29. Pandey D, Rajendran SRCK, Gaur M, Sajeesh PK, Kumar A (2016) Plant Defense Signaling and Responses Against Necrotrophic Fungal Pathogens. *Journal of Plant Growth Regulation* 35: 1159-1174.
30. Susheng S (2013) The bHLH subgroup IIIId factors negatively regulate jasmonate-mediated plant defense and development. *PLoS genetics* 7.
31. Li X, Duan X, Jiang H, Sun Y, Tang Y, et al. (2006) Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. *Plant physiology* 141: 1167-1184.
32. Hou, Xi-Lin, Huang, Zhi-Nan, Li, et al. (2014) Genome-wide analysis of the bHLH transcription factor family in Chinese cabbage (*Brassica rapa ssp pekinensis*). *Molecular Genetics & Genomics* Mgg.
33. Sun H, Fan HJ, Ling HQ (2015) Genome-wide identification and characterization of the bHLH gene family in tomato. *BMC Genomics* 16: 9.
34. E M B (1991) Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* (New York, NY) 4998.
35. M E M (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Molecular and cellular biology* 2.
36. Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual review of plant biology* 60: 433-453.
37. Wang N, Xiang Y, Fang L, Wang... Y (2013) Patterns of Gene Duplication and Their Contribution to Expansion of Gene Families in Grapevine. *Plant Molecular Biology Reporter* 31: 852-861.
38. Austin, L., Hughes (1994) The Evolution of Functionally Novel Proteins after Gene Duplication. *Proceedings of the Royal Society B Biological Sciences*.
39. T J V (2000) The origins of genomic duplications in Arabidopsis. *Science* (New York, NY) 5499.
40. Li WH, Yang J, Gu X (2005) Expression divergence between duplicate genes. *Trends in Genetics* 21: 602-607.
41. J R (2004) DUPLICATION AND DIVERGENCE: The Evolution of New Genes and Old Ideas. *Annual Review of Genetics* 0.
42. Makkena S, Lamb RS (2013) The bHLH transcription factor *SPATULA* regulates root growth by controlling the size of the root meristem. *Bmc Plant Biology* 13: 1.

43. Wang K, Li MQ, Chang YP, Zhang B, Zhao WL (2020) The basic helix-loop-helix transcription factor *OsBLR1* regulates leaf angle in rice via brassinosteroid signalling. *Plant Molecular Biology* 102.
44. Li Z, Liu C, Zhang Y, Wang B, Ran Q, et al. (2019) The bHLH family member *ZmPTF1* regulates drought tolerance in maize by promoting root development and abscisic acid synthesis. *Journal of experimental botany* 70: 5471-5486.
45. Le Hir R, Castelain M, Chakraborti D, Moritz T, Dinant S, et al. (2017) *AtbHLH68* transcription factor contributes to the regulation of ABA homeostasis and drought stress tolerance in *Arabidopsis thaliana*. *Physiologia plantarum* 160: 312-327.
46. Dong Y, Wang C, Han X, Tang S, Liu S, et al. (2014) A novel bHLH transcription factor *PebHLH35* from *Populus euphratica* confers drought tolerance through regulating stomatal development, photosynthesis and growth in *Arabidopsis*. *Biochemical and biophysical research communications* 450: 453-458.
47. David S (2008) The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic acids research Database issue*.
48. Finn R, Coggill P, Eberhardt R, Eddy S, Mistry J, et al. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic acids research* 44: D279-285.
49. Eddy S (2011) Accelerated Profile HMM Searches. *PLoS computational biology* 7: e1002195.
50. Larkin MA, Blackshields G, Brown NP, Chenna RM, Mcgettigan PA, et al. (2007) Clustal W. Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.
51. Goran N, Togawa RC, Mancini AL, Kuser PR, Yamagishi MEB, et al. (2003) STING Millennium: A web-based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence.
52. E Q (2005) InterProScan: protein domains identifier. *Nucleic acids research Web Server issue*.
53. Zdobnov E (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 9.
54. Sudhir K, Glen S, Koichiro T (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology & Evolution*: 1870.
55. M K (2009) Circos: an information aesthetic for comparative genomics. *Genome research* 9.
56. Tae-Ho L, Haibao T, Xiyin W, Paterson AH PGDD: a database of gene and genome duplication in plants. *Nucl Acids Research*: 1152-1158.
57. Yupeng W (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic acids research* 7.
58. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J (2010) KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. *Genomics, Proteomics & Bioinformatics* 8: 77-80.
59. Xiaobing, Kou, Kaijie, Qi, Xin, et al. (2017) Evolution, expression analysis, and functional verification of *Catharanthus roseus* RLK1-like kinase (CrRLK1L) family proteins in pear (*Pyrus bretschneideri*). *Genomics*.

60. Huang X, Li K, Xu X, Yao Z, Jin C, et al. (2015) Genome-wide analysis of WRKY transcription factors in white pear (*Pyrus bretschneideri*) reveals evolution and patterns under drought stress. BMC genomics 16: 1104.
61. Lalitha S (2000) Primer Premier 5. Biotech Software & Internet Report 1: 270-272.

Tables

Due to technical limitations, table 1 and 2 is only available as a download in the Supplemental Files section.

Figures

Figure 1

Distribution of amino acids in the bHLH motif. The columns labeled a, b, c, d and e represent the bHLH proteins identified in Pear, apple, peach, strawberry and Chinese plum, respectively. The numbers below a, b, c, d and e represent the positions of the residues in the alignments of the bHLHs.

Figure 2

The distribution of introns within domains of 749 Rosaceae bHLH proteins (pear, apple, peach, strawberry and Chinese plum). All patterns are color coded and designated from I to XI. The triangles and numbers (1 to 3) are used when the position of the intron coincides with the example. The numbers of the 749 Rosaceae bHLH proteins with each pattern is given at the right.

Figure 3

Phylogenetic analysis of 749 Rosaceae bHLH proteins (pear, apple, peach, strawberry and Chinese plum) and 150 Arabidopsis bHLH proteins. The phylogenetic tree was conducted using the MEGA 6.0 software on the basis of amino acid sequences of 899 bHLHs. Bootstrap analysis was conducted with 1000 replicates. The tree shows the 34 phylogenetic subfamilies marked with numbers above a gray branch.

Figure 4

Different duplication modes of bHLH families in Rosaceae. The y-axis represents the number of duplicated gene pairs. The x-axis represents the species. WGD, whole genome duplication; TD, Tandem duplication; PD, Proximal duplication; DSD, Dispersed duplication; SD, Singleton duplication.

Figure 5

Localization and synteny of bHLH genes in pear. A. Localization and synteny of the bHLH genes in the pear genome. The bHLH genes in pear (PbbHLHs) were mapped to the 17 chromosomes. The chromosome number is indicated on the inside. Gene pairs with an orthologous relationship are joined by a line. B. Segmental duplication between members of the bHLH family in pear. Homologous gene pairs

are connected with bands. The chromosome segment is indicated by gray horizontal line, and the broad line with arrowhead represents gene and its transcriptional orientation. The bHLH genes are shown in red.

Figure 6

Ka/Ks ratio and Ks value of bHLH gene family A. Ka/Ks ratio of bHLH genes in five Rosaceae species. We analyzed the Ka/Ks values using coding sequences. The x-axis represents Ks value. The y-axis indicates

the Ka/Ks ratio. B. Distribution of mean Ks values of 63 pairs of PbrMYB genes. The x-axis represents the mean Ks value; y-axis represents the density of the distribution.

Figure 7

Expression pattern of PbbHLH genes in pear different tissues. The heat map was generated by transcriptome data of pear seven different tissues, including stem, leaf, bud, sepal, petal, ovary and fruit.

Color scale at the top represents log2 transformed RPKM (reads per kilobase per million) values. Blue indicates low expression and red indicates high expression.

Figure 8

Heat map analysis of PbbHLH genes in response to drought stress. The heat map was generated by transcriptome data of pear in response to drought stress. Color scale at the top represents log2

transformed RPKM (reads per kilobase per million) values. 0, 1, 3, 6 and D24, represents for 0, 1, 3 and 6 h under drought treatment and recovered for 24 h in water, respectively.

Figure 9

Relative expression levels of six PbbHLH genes in response to drought treatment. Quantitative RT-PCR was used to analyze the expression level of six PbbHLH genes in response to drought stress. 0, 1, 3, 6 and D24, represents for 0, 1, 3 and 6 h under drought treatment and recovered for 24 h in water, respectively.

Supplementary Files

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

- [Table1.xlsx](#)
- [Table2.xlsx](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)

- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [TableS7.xlsx](#)
- [TableS8.xlsx](#)
- [FigS1.pdf](#)
- [FigS2.pdf](#)
- [FigS3.pdf](#)