


# SCIENTIFIC REPORTS



OPEN

## Genome-wide Identification, Classification, and Expression Pattern of Homeobox Gene Family in *Brassica rapa* under Various Stresses

Nadeem Khan<sup>1</sup>, Chun-mei Hu<sup>1,2</sup>, Waleed Amjad Khan<sup>1</sup>, Wenli Wang<sup>1</sup>, Han Ke<sup>1</sup>, Dong Huijie<sup>1</sup>, Zhang Zhishuo<sup>1</sup> & Xilin Hou<sup>1</sup> 

Homeobox (HB) genes are crucial for plant growth and development processes. They encode transcription factors and responses to various stresses, as reported by recent emerging evidence. In this study, a total of 113 *BraHB* genes were identified in *Brassica rapa*. On the basis of domain organization and phylogenetic analysis, the BraHBs were grouped into nine subclasses, in which homeobox leucine-zipper (HB LZIP-III) showed the highest number of genes (28) compared to other subclasses. The BraHBs exhibited similarities in exon–intron organization and motif composition among the members of the same subclasses. The analysis revealed that HB-Knotted was more preferentially retained than any other subclass of BraHB. Furthermore, we evaluated the impact of whole-genome triplication on the evolution of BraHBs. In order to analyze the subgenomes of *B. rapa*, we identified 39 paralogous pairs for which synonymous substitution values were lower than 1.00 for further purifying selection. Finally, the expression patterns of BraHBs across six tissues expressed dynamic variations combined with their responses against multiple stresses. The current study provides brief information on the homeobox gene family in *B. rapa*. Our findings can serve as a reference for further functional analysis of BraHBs.

Homeobox genes are known to play important roles in the body plan specification of relatively higher organisms at earlier stages of embryogenesis. Plant homeobox genes have been implicated in various processes, such as embryo patterning, development of root, shoot, and floral meristems, vascular development, and various stress responses<sup>1–6</sup>. These genes are considered to be key regulators of plant morphogenesis. Homeobox genes encode a lengthy conserve domain consisting of 60 DNA-binding amino acids, known as homeodomain (HD). Homeobox genes were first isolated in fruit fly (*Drosophila melanogaster*) and were subsequently found to be involved in many aspects of development<sup>7,8</sup>. The characteristic three-dimensional structure of HD comprises three alpha-helices, of which the second and third form a helix–turn–helix DNA-binding motif<sup>9–11</sup>. On the basis of their sequence similarity of homeodomains and co-domain characteristics, BraHB proteins have been classified into seven groups: KNOX, BEL, ZM-HOX, HAT1, HAT2, ATHB8, and GL2<sup>12</sup>. KNOX and BEL belong to the TALE superclass<sup>13</sup>. HAT1, HAT2, ATHB8, and GL2 genes are all characterized by a leucine-zipper motif downstream of the homeodomain<sup>14</sup> and have been successfully renamed as HD-ZIP I, HD-ZIP II, HD-ZIP III, and HD-ZIP IV, respectively<sup>12,15</sup>, although an alternative classification of these genes into five groups was also proposed (HD-ZIP, GLABRA, KNOTTED, PHD, and BEL)<sup>2</sup>. Furthermore, a comprehensive study on plant homeobox genes was conducted that categorized them into 14 classes, including some new classes, such as NDX, DDT, PHD, LD, SAWADEE, and PINTOX<sup>16</sup>.

Members of the plant homeobox gene family are involved in several development processes. The majority of HD-ZIP I proteins are involved in the regulation of cotyledon development, leaf cell fate determination, and blue

<sup>1</sup>State Key Laboratory of Crop Genetics and Germplasm Enhancement, Ministry of Science and Technology/College of Horticulture, Nanjing Agricultural University, Nanjing, 210095, P. R. China. <sup>2</sup>New Rural Research Institute in Lianyungang, Nanjing Agricultural University, Nanjing, P. R. China. Correspondence and requests for materials should be addressed to C.-m.H. (email: [jjjhcmm@njau.edu.cn](mailto:jjjhcmm@njau.edu.cn))

Subclasses	Identified Genes
WOX	16
HB LZP-I	26
HB LZP-II	17
HB LZP-III	28
HB-Knotted	7
BEL	14
PHD	2
HB-DDT	2
Unclassified	1
<b>Total</b>	<b>113</b>

**Table 1.** Homeobox identified genes of *Brassica rapa* classified based on their domain and phylogenetic relationships.

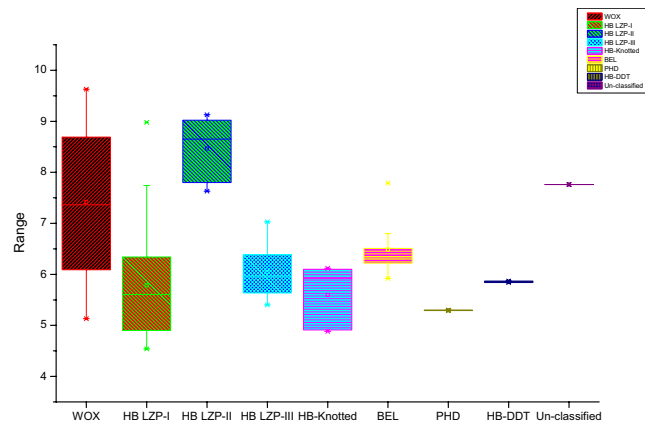
light signaling<sup>17,18</sup>. HD-ZIP II proteins participate in shade avoidance responses<sup>19</sup>. Some members of HD-ZIP III modulate the apical meristem formation, vascular development, and maintenance of adaxial or abaxial polarity of leaves and embryos<sup>20</sup>. HD-ZIP IV proteins stimulate the outer cell layer formation of plant organs and monitor anthocyanin pigmentation and epidermal layer maintenance<sup>21,22</sup>. KNOX family members provide support to apical meristem shoot growth maintenance and engage leaf form diversity<sup>23</sup>. They have been also reported to interact with BEL family members for the regulation of hormone homeostasis<sup>3</sup>. WUSCHEL (WOX) family members in *A. thaliana* mark cell fate during early embryonic patterning, while some members are also involved in stem cell maintenance and organogenesis<sup>24,25</sup>. These proteins may also be involved in cell differentiation during anther development<sup>26</sup>. Meanwhile, ZF-HD family members are involved with floral development processes in *A. thaliana*<sup>27</sup>. The above research emphasized exploring the homeobox proteins and their complex nature.

Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*) is an important Asian vegetable that is widely cultivated in China, Korea, and Japan<sup>28,29</sup>. The Chinese cabbage genome (Chifu-401-42) was sequenced and assembled recently<sup>30</sup>. It is known to play an important role in global agriculture and horticulture, and exhibits a close relationship with the model plant *A. thaliana*. Moreover, it experienced a whole-genome triplication (WGT) event since its divergence from *A. thaliana* about 13 to 17 million years ago (MYA)<sup>31</sup>. It also serves as an excellent model system for the study of genome evolution. The genome of Chinese cabbage is based on three subgenomes: least fractionated (LF), medium fractionated (MF1), and most fractionated (MF2). Intriguingly, at the gene density and expression levels, LF is dominant over the other two subgenomes<sup>30,32</sup>. The availability of genome databases for Chinese cabbage, rice, and *Arabidopsis* has enabled us to search the comparative genomics of the homeobox transcription factor gene family. Characterization of homeobox genes in *B. rapa* will be focused in such a way as to provide information about the molecular changes that occur in mechanisms against various stresses such as cold, heat, salt, drought, ABA, and GA. Subsequently, gene modification can proceed further by using resistant *Brassica* varieties.

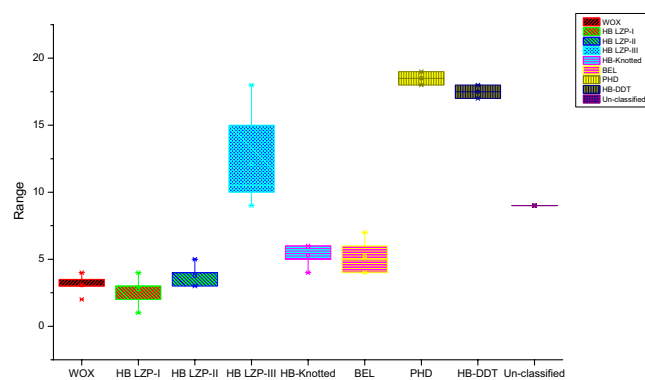
In this study, we identified 113 homeobox genes in Chinese cabbage based on their genome sequences and categorized them into nine subclasses. Gene duplications and chromosomal locations were also investigated to support our findings. Expression profiles under six different stress treatments (ABA, GA, drought, salt, heat, and cold) were evaluated to determine the responses of *BraHB* genes in Chinese cabbage (Chifu-401-42). Moreover, cis-element promoter functions were also predicted. Our results provide novel insights into the stress responses of *BraHB* genes and convey a clear image to understand the construction and function of homeobox genes in Chinese cabbage.

## Results

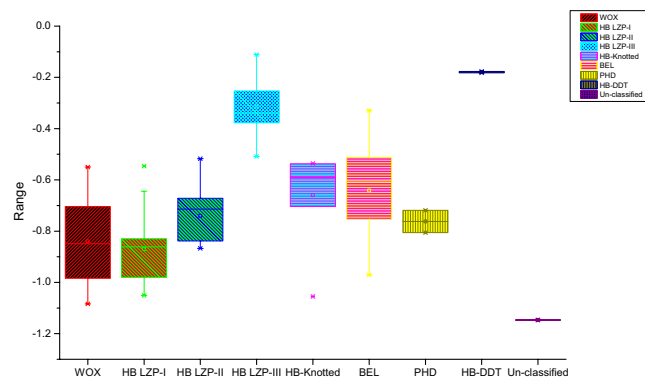
**Identification, Classification, and Comparative Analysis of BraHBs in *B. rapa*.** Based on our putative *B. rapa* genome studies<sup>33</sup>, we used sequences of rice and *A. thaliana* as queries and hidden Markov model (HMM) profile to confirm the HD in the *B. rapa* genome. According to the HDs, we identified 113 BraHB proteins (Table 1). The distribution pattern of homeobox in *B. rapa* was divided into nine subclasses, including two major classes, HD ZIP (HD ZIP-I, HD ZIP-II, HD ZIP-III) and TALE (KNOX and BEL), and three subcategorizations of WOX, HB-DDT, and PHD based on previous reports in plants<sup>16</sup>. The complete information related to BraHBs is listed in Supplementary Table S1, including gene identifier, protein length, cDNA length, genomic location, and subcellular prediction. The theoretical isoelectric point (pI) for all BraHB ranges from 4.54 to 9.63, with a median of 6.56, is shown in Fig. 1. The data of the number of exons (Fig. 2) show complex ranges for different subclasses, but the normal range was calculated as 1–19, with a median of 6.33. The subcellular predictions for all the BraHBs (Supplementary Table S1) show that the majority of proteins were localized in the nucleus, cytoplasm, and mitochondria, although some proteins were located in different organelles, i.e., Golgi body, plasma membrane, chloroplast, and others. In order to further characterize the BraHBs, we analyzed the physicochemical properties of the putative proteins (Supplementary Table S1). The value for the grand average of hydropathicity ranges from -1.147 to -0.111, which appeared to be negative and represent hydrophilic behavior (Fig. 3). The 113 members of BraHB were designated as BraHB 1 to BraHB 113 (Supplementary Table S1). In this study, the largest grouping of proteins was identified in subclass HB LZP-III (28 proteins), followed by HB LZP-I (26), with HB ZIP proven as the major class of the homeobox family. The lowest number of proteins was observed in PHD and HB-DDT, with two each. So the expansion pattern of homeobox in the *B. rapa* genome was similar to that of rice



**Figure 1.** Indication of isoelectric point (pI) values among different subclasses of BraHB.



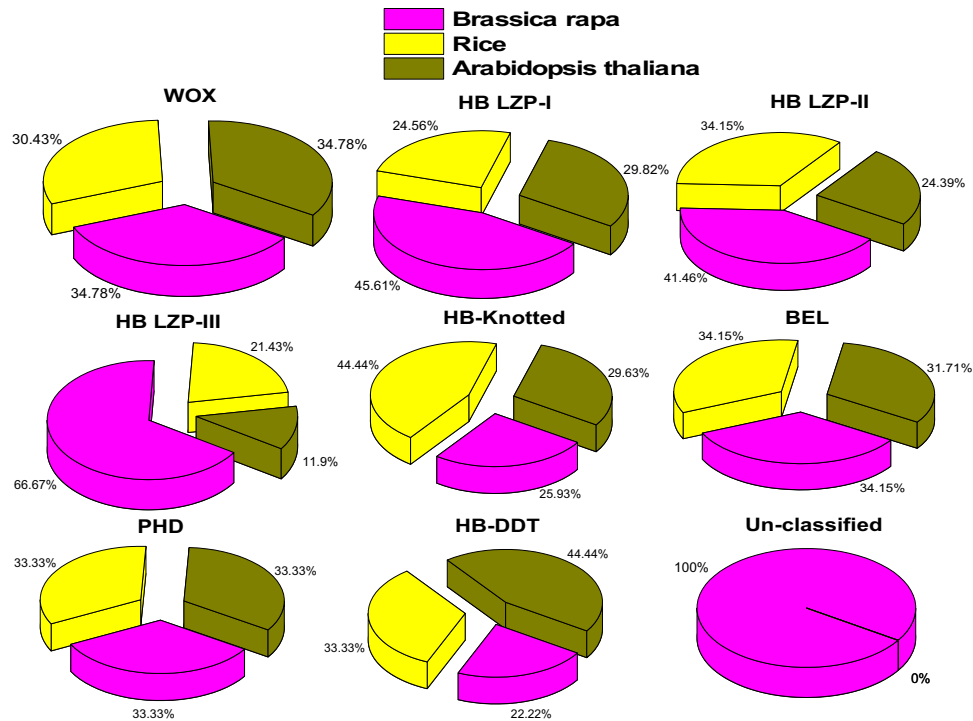
**Figure 2.** Number of exon among subclasses of BraHB.



**Figure 3.** Grand average of hydropathicity (GRAVY) among the subclasses of BraHB.

and *A. thaliana*, i.e., the number of genes was 113. We also analyzed the relative shares of different subclasses of BraHBs based on the number of genes in them (Fig. 4). Consistently, HB LZP-III showed the highest number of genes in *B. rapa* (66.77%) compared to the other two species, while the PHD subclass shared the same number of genes (33.33%).

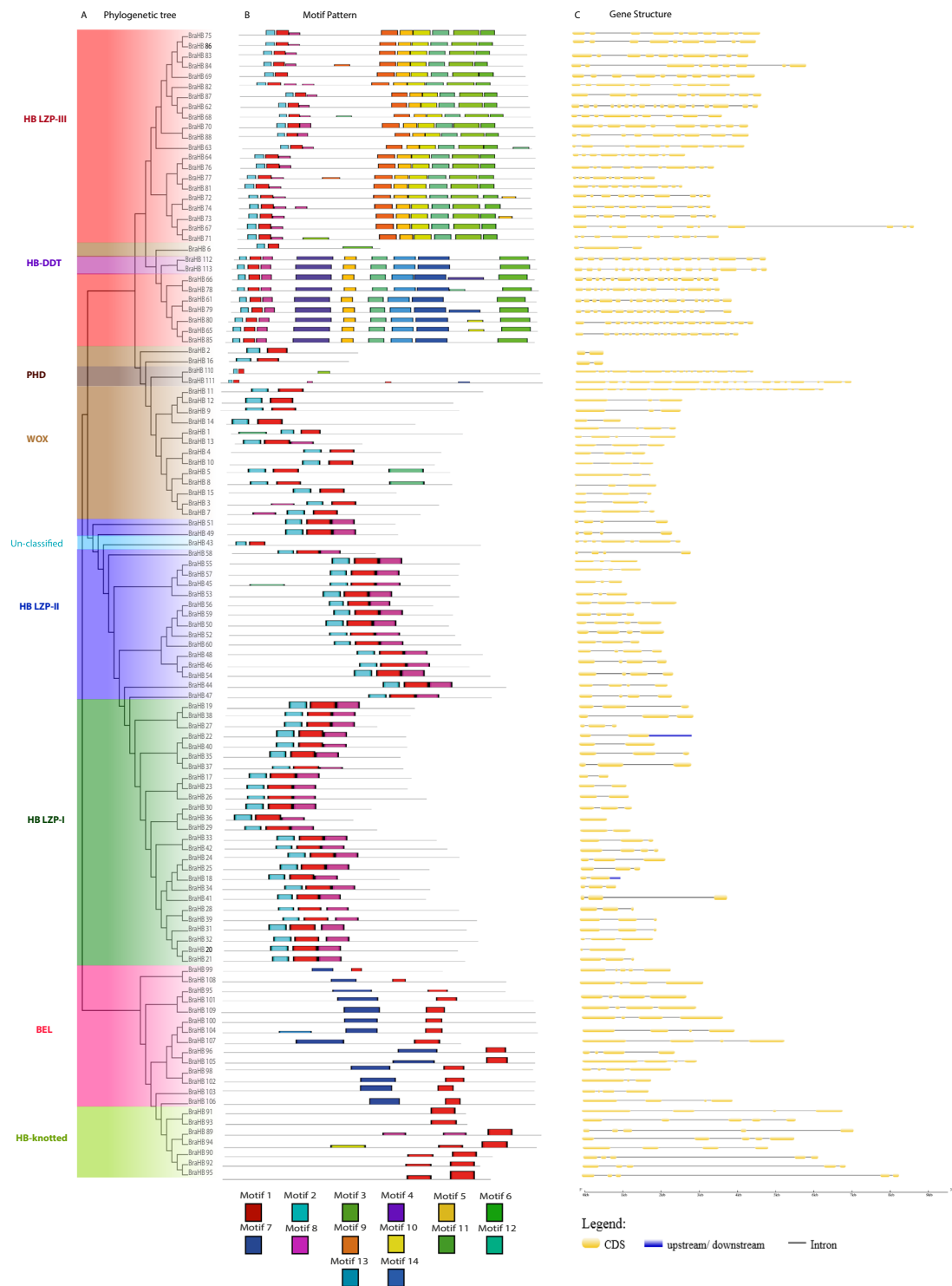
**Expansion Analysis and Characteristic Structure of BraHBs.** To examine the expansion and evolutionary relationships among different subclasses of BraHBs, a phylogenetic tree was constructed using MEGA 7.0 software. The maximum likelihood method with 1000 bootstrap replications was used (Fig. 5A). The phylogenetic results showed that the distribution of BraHBs was subcategorized into further divisions: HB ZIP contained three subclasses with different members of BraHB (HB ZIP-I (26), HB LZP-II (17), and HB LZP-III (28), HB-Knotted (7) and BEL (14)). WOX (16), HB-DDT (2), PHD (2), and uncharacterized (1) proteins were identified, which were consistent with reports on *A. thaliana* and rice<sup>33</sup>. Interestingly, all of the subclasses of BraHBs were grouped



**Figure 4.** Relative classification patterns of genes among three species.

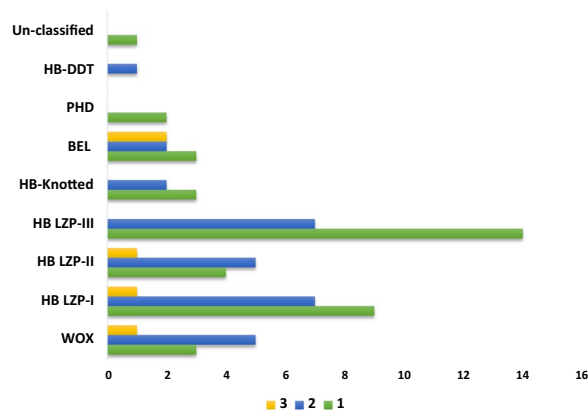
together in a phylogenetic tree except WOX, which was considered to be diverse in nature. To better understand motif composition in BraHBs, all of the putative protein sequences were analyzed by Multiple EM for Motif Elicitation (MEME) software (version 4.12) for the conserved motifs<sup>34</sup>. In total, 14 motifs were identified, designated as motifs 1 to 14, by calibrating their width ranges from 12–100 (Fig. 5B). At the same time, the LOGO of BraHB proteins was also procured with the help of MEME (Supplementary Figure S1). The number of conserved motifs varied in the range of 2–10 in all proteins. We noticed that motif 1 was the most dominant, as it was found in all BraHBs, followed by motif 2. The other motifs were specific to one or two subclasses; for example, motifs 5, 6, 9, and 12 were specific to HB LZIP-I, whereas motifs 1 and 2 were specified to WOX, which might be due to the functional similarities within subfamilies. The varied nature of different motifs within BraHBs suggests inconsistency in their functions. The exon–intron organization was employed using Gene Structure Display Server (GSDS) 2.0 software, in order to gain additional information with respect to conservation and diversification of BraHBs, as the gene structure is more closely related to the function of the gene and, together with phylogenetic tree construction, it reflects the close relationship of different subclasses of BraHBs (Fig. 5C). Most of the BraHBs within the same subclasses exhibited similar patterns of structure with respect to the same number of exons and introns, ranging from 4–19 and 2–17. On the other hand, a majority of the genes showed similar gene structure; however, significant diversification was also detected among BraHB subfamilies. We also calculated the genetic distance to estimate the relationships among the nine subclasses using box plots (Supplementary Figure S2a–c). Notably, the genetic distance between BEL and HB-Knotted was shorter than the others (Supplementary Figure S2a). Most of the groups show close distance except HB-LZIP III vs. HB-DDT (Supplementary Figure S2b). Also, most of the groups show similarity while belonging to different subclasses of BraHBs (Supplementary Figure S2c). The similar patterns among subclasses suggest that most of them share a common evolutionary origin.

**Different Retention of BraHB Genes Following the Whole Genome Triplication Event.** In order to explore the influence of WGT on the evolution of BraHBs, we studied the gene retention of BraHB after WGT. The *B. rapa* genome contains ~42,000 genes after its divergence from the *A. thaliana* genome, which was consistent with ~30,000 genes. For this occurrence, a considerable number of genes were lost after triploidization in *B. rapa*<sup>30</sup>. The results of different retention of genes in the syntenic region with respect to subclasses were demonstrated as follows: WOX (16/20), HB LZIP-I (26/31), HB LZIP-II (17/19), HB LZIP-III (28/36), HB-Knotted (7/7), BEL (14/19), PHD (2/4), HB-DDT (2/2), and unclassified (1/1) (Supplementary Table S2). In order to compare the different retention of genes among subclasses of BraHBs, we also counted the copy number of genes and analyzed their distribution patterns across the three subgenomes of *B. rapa*: least fractionated (LF), medium fractionated (MF1), and most fractionated (MF2). As shown in Fig. 6, the most copies of genes (14) were found in HB LZIP-III, followed by HB LZIP-I (9), and HB LZIP-II (4). We also counted the number of genes in the subgenomes. All nine subclasses of BraHBs showed various numbers of genes. In summary, HB LZIP-I and HB LZIP-III displayed similar numbers of genes, 26 and 28, respectively. Additionally, both of them carried a high number of genes, 13 and 14, in the LF genome compared to other subclasses of BraHBs (Fig. 7). Overall, a majority of genes (47.79%) were located in the LF genome, as described in Fig. 8. As for the other subgenomes, MF2 contained 28.32%, while the least number of genes was found in MF1, 23.89%.

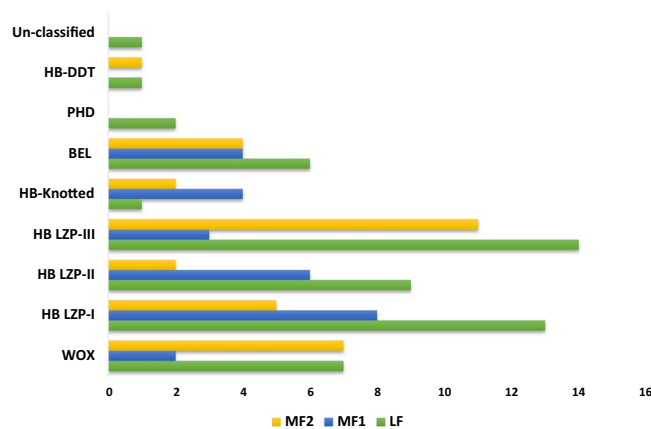


**Figure 5.** Phylogenetic tree, protein motif, and gene structure of BraHBs. **(A)** The phylogenetic tree was constructed by MEGA 7 using the maximum likelihood method (1000 bootstrap). **(B)** The conserved motifs of BraHBs were elucidated by Multiple EM for Motif Elicitation (MEME). Different motifs and their positions are represented by different colors, numbered 1–14 at the bottom. **(C)** The exon–intron and upstream/downstream regions are represented by yellow boxes, gray lines, and blue box, respectively. At the bottom of the figure the relative position is proportionally displayed based on the kilobase scale.

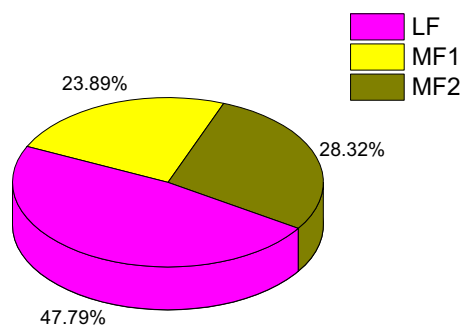
**Chromosomal Localization and Synteny Gene Analysis of BraHB.** All *BraHB* genes were mapped to the 10 chromosomes of Chinese cabbage, which were distributed in a random manner (Fig. 9). On every chromosome, the proportion of genes was found to be random. Chromosome A09 contained the most genes (19),



**Figure 6.** Copy number of variation and gene retention in different protein groupings.



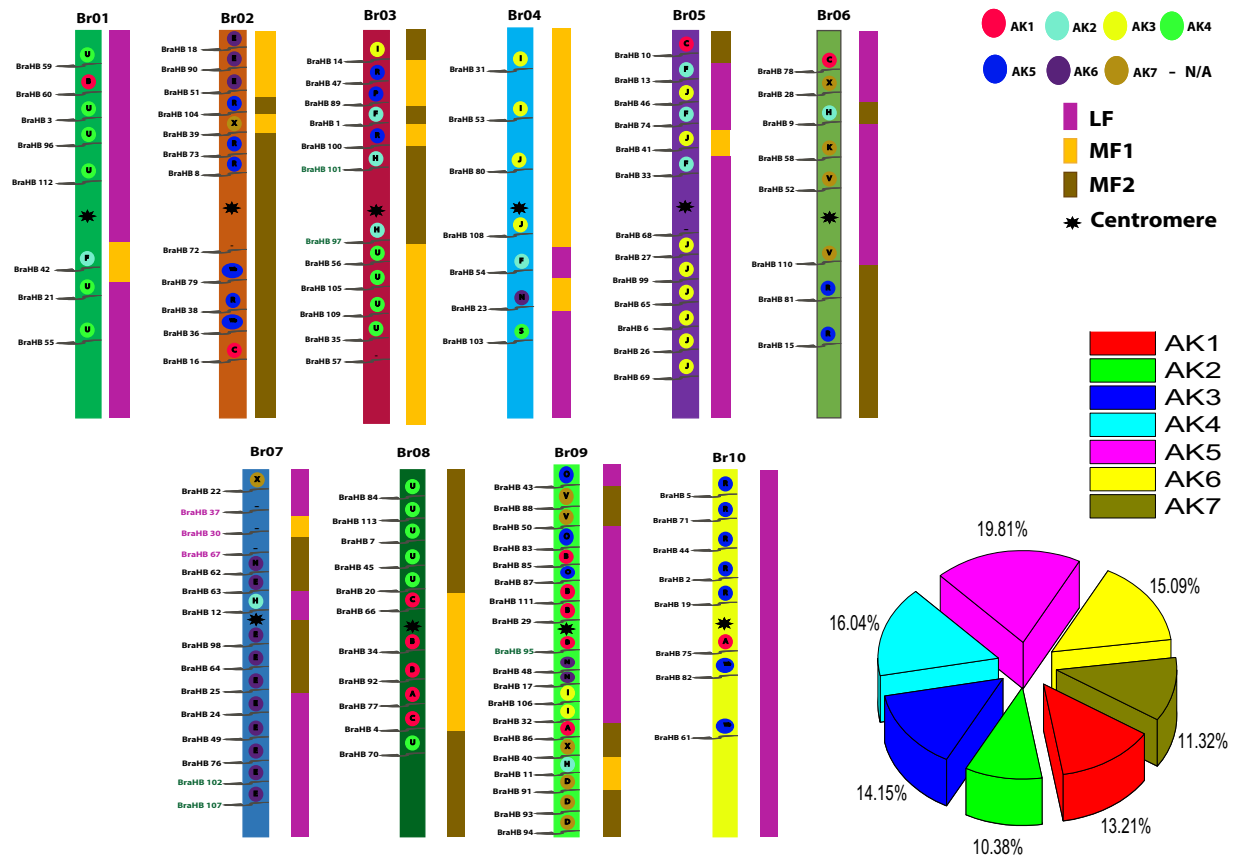
**Figure 7.** Representation of number of genes between the subclasses of BraHB within the three subgenomes of *B. rapa*.



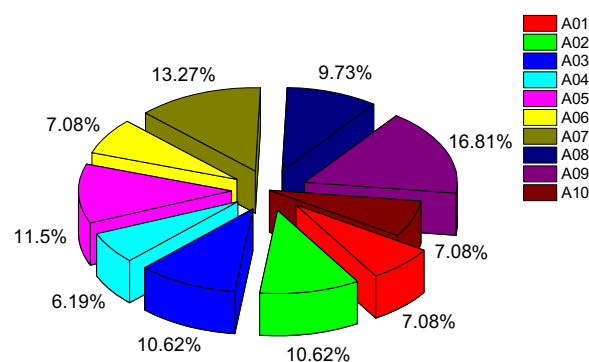
**Figure 8.** Relative percent of the number of genes within three subgenomes of *B. rapa*.

whereas chromosome A04 had the least (7). The other chromosomes, A01, A02, A03, A05, A06, A07, A08, and A10, contained 8, 12, 12, 13, 8, 15, 11, and 8 genes, respectively. Meanwhile, according to a previous study<sup>11</sup>, we also demonstrated in *B. rapa* genome the 24 conserved ancestral genomic blocks (labeled A–X). The colors were arranged according to the position of these blocks in a proposed ancestral karyotype (AK1–7). We observed that most of the *BraHB* genes clustered together in a region of AK5 (19.81%), followed by AK4 (16.04%), and AK (15.09%), whereas the least amount of genes were located in AK2 (10.38%). Based on shares among subclasses of BraHBs, chromosome A09 had the highest share (16.81%), followed by A07 (13.27%), while A01, A06, and A10 had equal shares of 7.08% (Fig. 10).

Concurrently, the types of duplication genes were also identified and classified by using the MCScanX program. In total, we identified five tandem genes that were located on three different chromosomes, of which four were on A03 and A07 (two on each), with single duplication on A09. We also identified three dispersed genes that



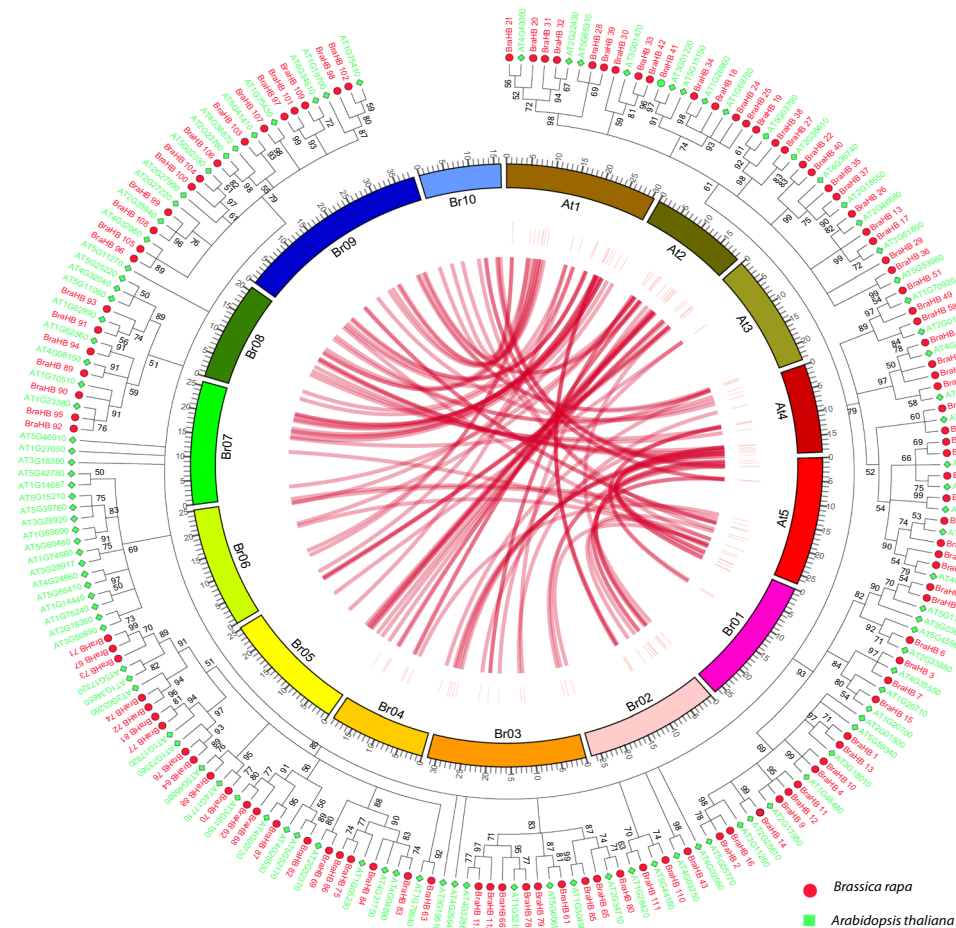
**Figure 9.** Chromosome locations of BraHBs were obtained from the generic file format (GFF) file and displayed by using MapChart (modified by Adobe Illustrator). The ancestral karyotypes are marked in different colors. Duplication types are marked as follows: segmental in black, tandem array in green, and dispersed in purple.



**Figure 10.** Relative shares of different subclasses of BraHB from A01-A10 chromosomes of *B. rapa*.

were only located on chromosome A07. The number of genes between the subgenomes and non-synteny ortholog are also presented (Supplementary Figure S3). The synteny gene relationship between chromosomes of *B. rapa* and *A. thaliana* was summarized by using Circos software along with the phylogenetic tree (Fig. 11).

**Analysis of Putative Regulatory Cis-Element in BraHBs.** Transcription factor is one of the key factors involved in the regulation and expression of genes by either promotion or suppression. However, transcription factor is also under the control of other regulators upstream during various biological processes in plants. Therefore, homeobox genes may be assumed to regulate binding of the promoter regions and control the cascade reaction that mainly occur in plants on certain occasions. For this reason, we carried out cis-regulatory element analysis for the identification of transcriptional regulation in the promoter region (2 kb upstream) of BraHBs in *B. rapa*. We utilized the PlantCARE database to identify the various cis-elements. Consequently, we figured

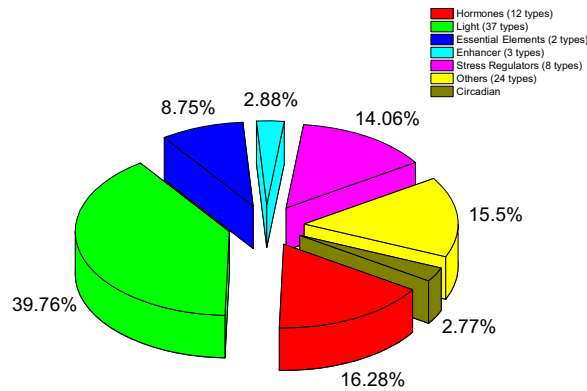


**Figure 11.** Collinear correlation for all genes of *BraHB* displayed between *B. rapa* and *A. thaliana* along with phylogenetic tree. The 10 Chinese cabbage chromosomes (Br01–Br10) and five *A. thaliana* chromosomes (At1–At5) are shown in different random colors. The illustration was drawn using Circos software. The phylogenetic tree was constructed by MEGA 7 using the maximum likelihood method (1000 bootstrap).

out a number of different elements involved in various physiological processes of plants. For example, for the processes related to hormonal pathways, we identified 11 types of cis-elements including light (37), essential elements (2), enhancer (3), stress-related (8), and others (25), which are presented in Supplementary Table S3. There were many hormone-related elements, which were found in most of the promoter sequences, i.e., CGTCA, ERE, GARE, ABRE, and AuxRR-core. They are mainly involved in hormone signaling, such as methyl jasmonate, ethylene, gibberellin, abscisic acid, and auxin, suggesting that they may control the regulatory expression of *BraHB*s. Interestingly, for light regulation factor, we found many different elements (37) that were common in most of the genes of *BraHB*. We speculated that they may be involved in plant energy metabolism. For stress-related genes, common cis-elements were HSE, MBS, and *skn-1*-motif, which indicated that they may help in the stress mechanism and other developmental pathways. Other important cis-elements like Box-W1 are responsible for fungal elicitor. Several transcriptional regulation cis-elements, 5UTR Py-rich stretch, CAAT, and TATA, were also found among *BraHB*s. Figure 12 describes the analysis based on the number of *BraHB* genes involved under various cis-elements. Most of the genes were involved in light-responsive types of promoters (39.76%), followed by hormones (16.28%) and other stress-related genes (15.5%). Notably, the promoter sequences of *BraHB*s involved in plant circadian cis-elements might be involved in controlling the plant environment, such as periodicity and temperature compensation.

**Expression Pattern Analysis in Different Tissues of *BraHB*s.** To characterize the expression patterns of *BraHB*s, we utilized publicly available resource data<sup>35</sup> for six tissues: root, stem, leaf, flower, silique, and callus. A heatmap for these tissues was generated to display the expression patterns and clustering of *BraHB* genes. Most of the genes highlighted different expression patterns, while some exhibited similar expression (Fig. 13 and Supplementary Table S4). According to the results of expression of *BraHB* genes in different tissues, high expression was recorded in roots (81.4%), followed by stem (88.49%), leaf (80.3%), flower (85.84%), silique (84.95%), and callus (83.18%). Noticeably, our results show approximately higher expression in all the tissues. Among all the *BraHB* genes, the relative expression pattern in stem was higher (88.49%). Overall, the relative expression exhibited a consistent level of about 83.95%, which was expressed across all six tissues of *BraHB* genes. Likewise,





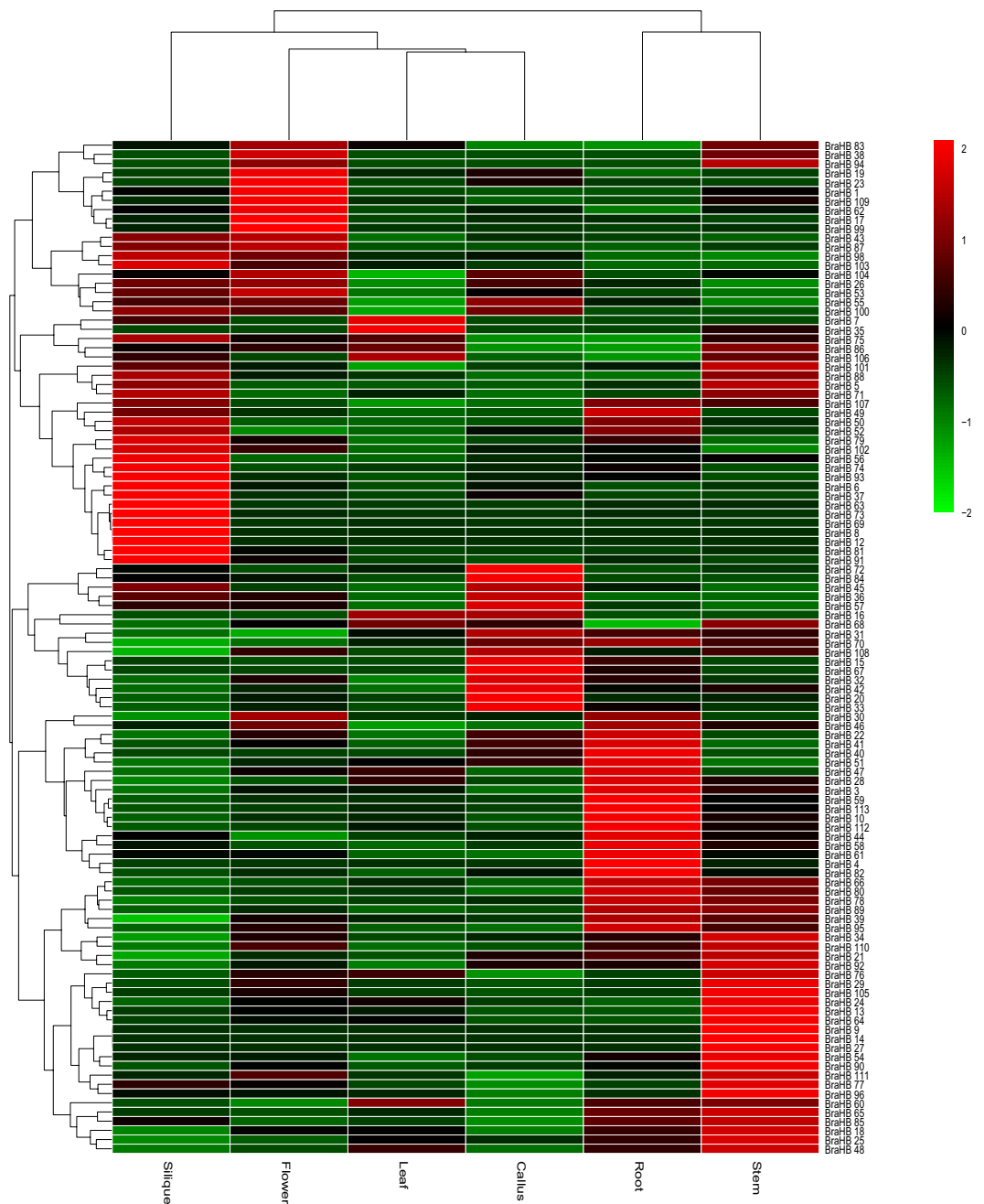
**Figure 12.** Relative percent of different genes of BraHB according to the distribution patterns of cis-elements.

clustering of different tissue-specific genes was also demonstrated (Fig. 14). Interestingly, two novel genes were found in silique, which might suggest that these genes have a tissue-specific role.

**Syntenic Paralog Pairs and Prediction of the Evolutionary Fate of BraHB.** Based on a literature review and outcomes from our study, we made an attempt to use the publicly available resource data of RNA-seq<sup>35</sup>. The information based on gene expression provided an opportunity to understand the mode and tempo of duplicated genes. We used 39 paralogous pairs of BraHB for gene expression profiling across six tissues (root, stem, leaf, flower, silique, and callus). The results, presented in Fig. 15 and Supplementary Table S5, show high relative expression among all the tissues: root (87.17%), stem (97.43%), leaf (89.74%), flower (89.74%), silique (94.87%), and callus (84.61%). The clustering image of the paralogous pairs of BraHB showed that only one specific gene was found in silique for a tissue-specific role among the six tissues (Fig. 16). Additionally, we also calculated the correlations among the 39 paralog pairs. Based on our results, the Pearson correlation coefficient (PCC) value was  $>0.6$  in about 20 pairs. The higher PCC values between these paralog pairs indicated their close relationship and the involvement of functional conservation or subfunctionalization after the event of duplication. In our results, three pairs were negatively co-related among all the BraHB genes, as these paralog pairs may indicate neofunctionalization. Interestingly, two pairs of genes from WOX and BEL showed no expression, because one copy of each did not exhibit any expression pattern. So the PCC values for these paralog pairs were not recorded (NA), which might be due to pseudogenization during the process of evolution.

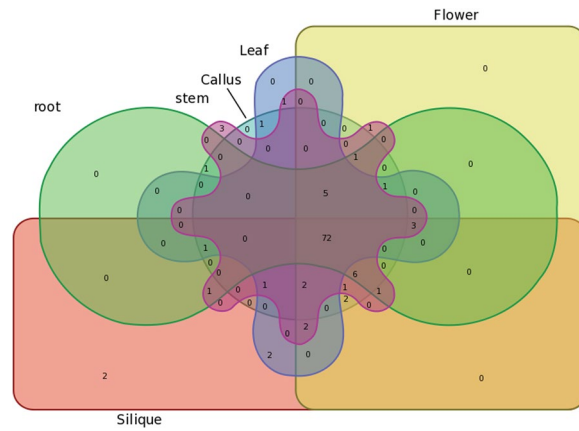
The three subgenomes of *B. rapa* are evolutionary products of WGT and contain many syntenic blocks between them. Syntenic paralogs are genes that are located in the syntenic fragments. Syntenic BraHB paralog pairs among LF, MF1, and MF2 were identified by searching “syntenic gene” in the Brassica Database (BRAD). In total, we also observed 39 pairs of paralogous genes. Synonymous ( $K_s$ ) and nonsynonymous ( $K_a$ ) values were calculated (Supplementary Table S6) to monitor the selective pressure on these paralog pairs. Interestingly, we found that the  $\omega = K_a/K_s$  ratios of 39 syntenic paralogs were estimated below 1.00, therefore they may lie in the purifying selection. In addition, the duplication time of these paralogs pairs was calculated by using a relative  $K_s$  measure as a proxy for time. The highest estimated duplication time of BraHB paralogous pairs was calculated as 16.87 MYA, with an average of 10.63 MYA. Our results suggest that stronger selective pressure on BraHBs forces them to duplicate earlier for their survival and leads to the speculation of BraHBs with varied and complex function.

**Differential Expression of BraHBs in Response to Abiotic Factors and Their Correlation.** Plant growth and development can be adversely affected by the uncertainty of climate change, such as fluctuations in temperature, salinity, and drought conditions that may limit crop productivity<sup>36–38</sup>. Since homeobox genes are known for their important role in the regulation of gene expression patterns under various abiotic stress factors, we selected 12 paralogous pairs of genes and analyzed them under multiple treatments: ABA, GA, PEG, NaCl, heat, and cold (Fig. 17 and Supplementary Table S7). A range of different expression profiles of selected pairs of BraHBs was observed as a result of these stresses. In general, in the case of ABA and cold stresses, the results of qRT-PCR showed that about 73% of genes were upregulated and 27% were downregulated (Fig. 18) on the basis of their different levels of treatment, i.e., 0, 1, 6, and 12 h. For GA, there was a slight decrease, with about 56% of genes upregulated and 44% downregulated, while NaCl had a slight increase, with 61% upregulated and 39% downregulated genes. Intriguingly, the drought and heat treatments show identical results, with 53% upregulated and 47% downregulated genes. For better understanding, a correlation was made among 12 paralogous gene pairs with respect to multiple treatments and an estimation of PCC was done. The correlation was designated as highly positive, mildly positive, or negative on the basis of PCC values<sup>39</sup>:  $>0.6$ , highly positive;  $0.5–0$ , mildly positive, and  $<0$ , negative. For both GA and PEG treatments, seven pairs of paralogs showed high correlation ( $PCC > 0.6$ ), followed by ABA with six pairs, while the most negative correlation was observed with cold stress (Fig. 19 and Supplementary Table S8). Moreover, these correlations were synchronized with the expression of syntenic pairs across the six tissues.

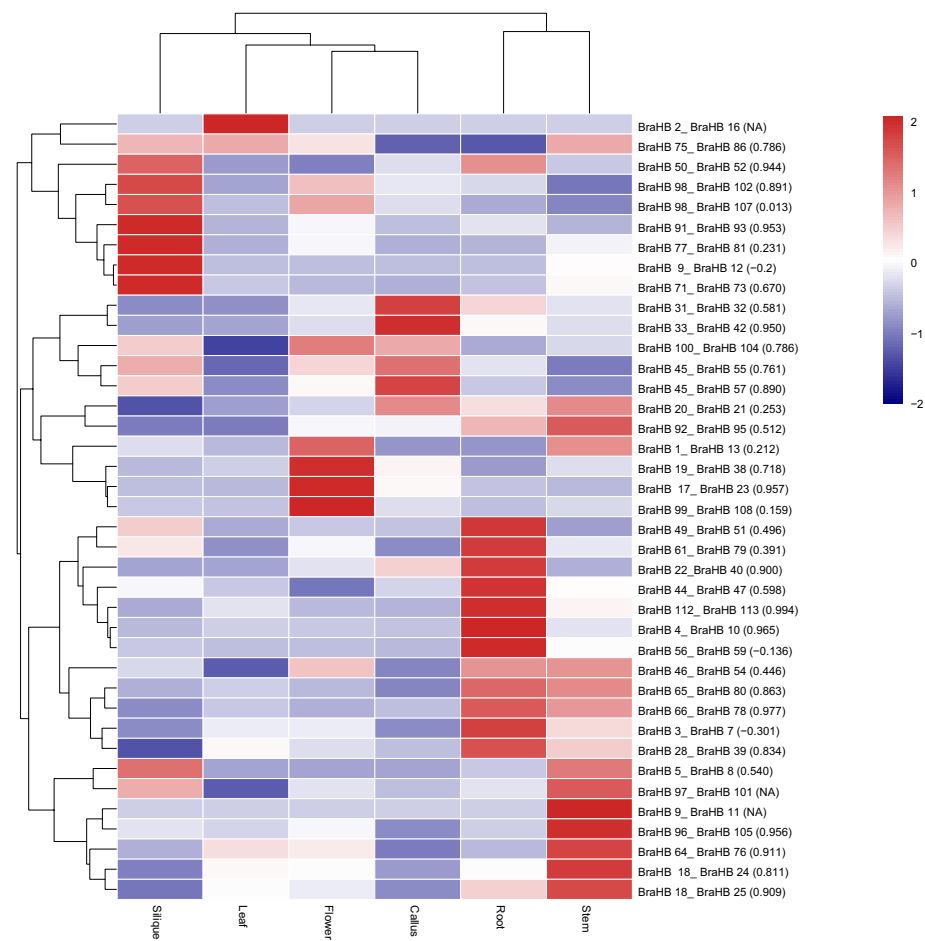


**Figure 13.** Heatmap of expression profiles (in log<sub>2</sub>-based fragments per kilobase of transcript per million mapped reads (FPKM)) of BraHBs in six tissues (stem, root, callus, leaf, flower, and silique). The expression levels are exhibited by the color bar.

**Interactions among Orthologous and Nonorthologous Pairs of BraHB Proteins.** In the present study, orthologous and nonorthologous pairs were explored between the three subgenomes of *B. rapa*, as these were the evolutionary products of WGT. For identification in the *B. rapa* database (BRAD)<sup>40</sup>, we searched the pairs by “syntenic gene” among least fractionated (LF), medium fractionated (MF1), and most fractionated (MF2) subgenomes. The regulatory network was presented using Cytoscape between orthologous and nonorthologous pairs of *BraHB* genes. All the orthologous pairs of BraHB that showed more or less similarity in their relationship strength were marked in red (Supplementary Figure S4). Similarly, the 22 nonorthologous pairs that showed a close relationship were marked in green.



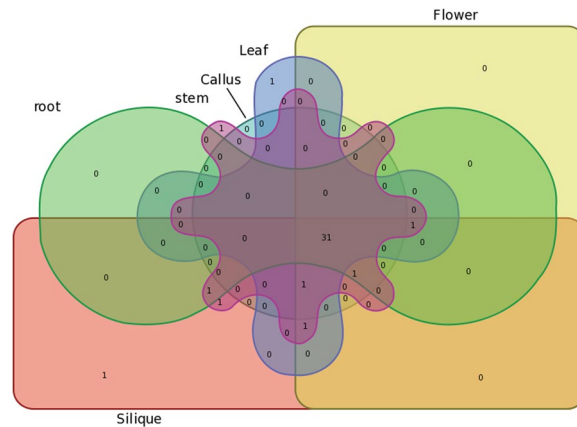
**Figure 14.** Venn diagram of the tissue expression of BraHBs.



**Figure 15.** Heatmap of expression profiles for BraHBs 39 paralogous pairs in six tissues: stem, root, callus, leaf, flower, and silique. Pearson correlation coefficients (PCCs) are shown in brackets. NA indicates no available results for PCC.

## Discussion

The plant kingdom contains a huge diversity of species. While they are sessile in nature, they still possess advanced characteristics of living activities. Homeobox transcription factor plays an important role in the morphogenesis of living organisms, and it begins from the very first stage of embryogenesis. To date, most research has been limited to certain plant species of *A. thaliana* and rice. In our research, we performed comprehensive studies of HBs in *B. rapa*, which includes whole genome-wide identification, chromosomal locations, phylogenetic analysis,



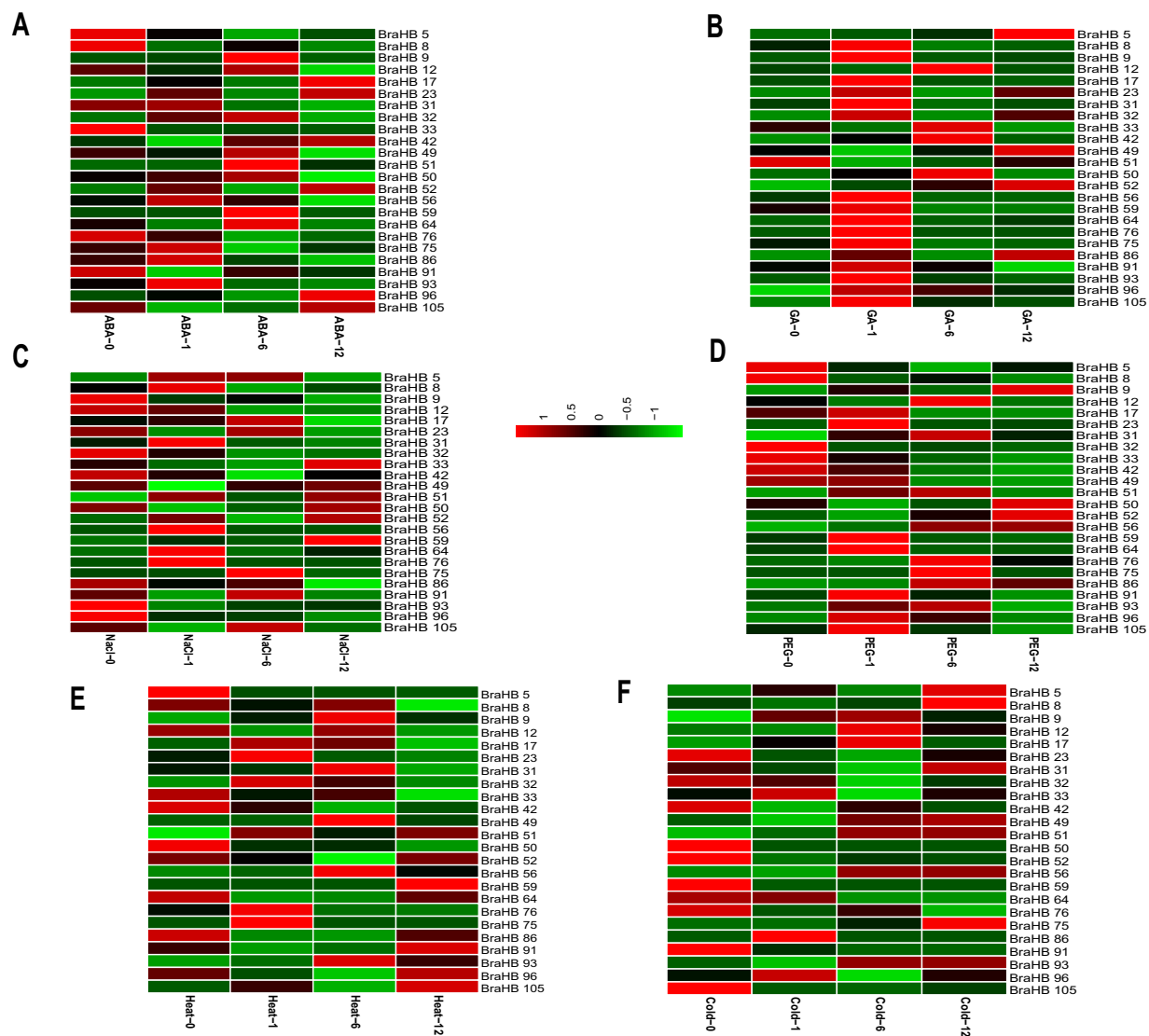
**Figure 16.** Venn diagram of the tissue expression of BraHBs.

gene duplication analysis, structural investigation, cis-elements, and interaction network. Finally, we also demonstrated the expression patterns of genes in different tissues and the response of HBs to various stress conditions.

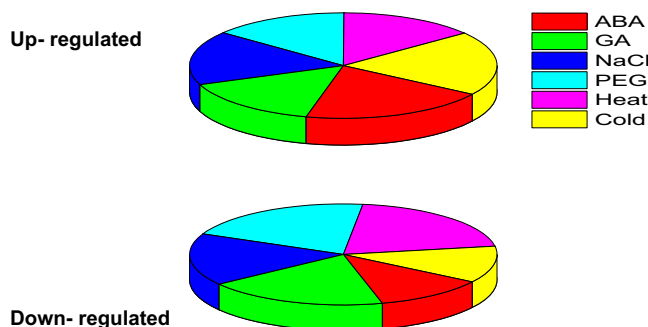
Phylogenetic analysis revealed that *B. rapa* exhibits a close evolutionary relationship with model plant *A. thaliana*<sup>11</sup>. Comparative genomic studies suggest that more than 60% of the genome assemblies between *A. thaliana* and *B. rapa* are highly conserved<sup>32,40</sup>. Around 93% of the total predicted *B. rapa* gene families also appear in *A. thaliana*<sup>32</sup>. Moreover, *A. thaliana* genes associated with regulatory networks for environmental stimuli, such as salt, cold, light, or hormonal responses, are also highly retained in *B. rapa*. In Arabidopsis, approximately 100 homeobox genes were identified initially, and then classification was done depending on their domain composition and phylogenetic relationships<sup>2</sup>. In our study, a total of 113 BraHBs were identified and verified by using various public databases.

Most land plants have undergone polyploidization that led to whole genome duplication (WGD). This provided an opportunity for duplicated genes to diverge in different evolutionary ways. Each of these genes subsequently experienced one of three fates: subfunctionalization, neofunctionalization, or nonfunctionalization (deletion or pseudogenization)<sup>41</sup>. These provided opportunities for duplicated genes to gain functional diversification, resulting in more complex organisms. In addition, segmental and tandem duplications are also known to contribute mainly in duplication modes during gene family expansion<sup>42</sup>. In previous reports, we noticed that *B. rapa* not only shared three paleo-polyploidy events with *A. thaliana*, but also underwent a further WGT event since its divergence from *A. thaliana* 13 to 17 MYA<sup>43</sup>. For the evolution of plants, gene duplication is not the only motivation, but it has a prominent role in the expansion of the gene family along with the succession of genomic rearrangements<sup>44</sup>. In our study, a majority of genes (> 92%) were segmental based on duplicated analysis, whereas only five and three genes belonged to tandem and dispersed type, respectively. These results suggest that segmental duplications contributed to the expansion of the BraHB family. Meanwhile, we calculated the rate of divergence of 39 paralogous pairs. The *Ka/Ks* ratio is an indicator of the selection history of genes or gene regions. Commonly, if the value of *Ka/Ks* is lower than 1, the duplicated gene pairs may have evolved from purifying selection (also called negative selection); *Ka/Ks* = 1 means neutral selection, while *Ka/Ks* > 1 means positive selection. In this study, we further noticed that the ratio of 39 syntenic paralogous pairs was below 1.00, which may predict that they evolved from purifying selection. These results demonstrate that during subsequent evolution, the syntenic pairs of BraHB did not diverge and suggest that the purifying selection might contribute to the prevalence of the *BraHB* gene family in *B. rapa*. The survival of plants depends on a number of environmental cues, such as extreme high/low temperature, salinity, and the disruption of water, that could adversely affect optimal plant survival and result in low productivity. In general, both biotic and abiotic stress responses are mainly governed by hormone signaling. ABA is a phytohormone that is involved in a number of different abiotic stresses such as cold, osmotic, and drought stress<sup>45</sup>. The responses of different genes to multiple treatments (ABA, GA, PEG, NaCl, heat, and cold) showed significant variation in the expression profile. The results, particularly for ABA, cold stress, and other stresses, provide a valuable clue to understanding gene function and robust candidate genes to improve the abiotic stress mechanism in *B. rapa*. In order to understand the regulatory functions of BraHBs, overexpression techniques will shed further light on the importance of these candidate genes in response to abiotic stresses.

In conclusion, the 113 HB genes in Chinese cabbage are comprehensively described, including their gene structures, phylogenetic profiles, gene duplications, subcellular localizations, conserved protein motifs, and expression patterns. To date, few genes of the Chinese cabbage transcription factor superfamily have been characterized in detail (AP2/ERF, Trihelix, and bHLH)<sup>46–48</sup>. Therefore, this is the first comprehensive and systematic research focused on Chinese cabbage. *In silico* analysis may assist in elucidating homeobox gene family function in protein interactions, signaling pathway regulations, and defense responses under different stress conditions. Altogether, it may also provide new opportunities to discover Chinese cabbage tolerance mechanisms under stress conditions. The outcome of our bioinformatics analysis provides basic resources to examine the molecular regulation of the homeobox transcription family during development and stress conditions in Chinese cabbage. In addition, the comparative study between Chinese cabbage and other species generates valuable information to study the function of homeobox transcription factor that may result in economic, agronomic, and ecological benefit for this vegetable crop.



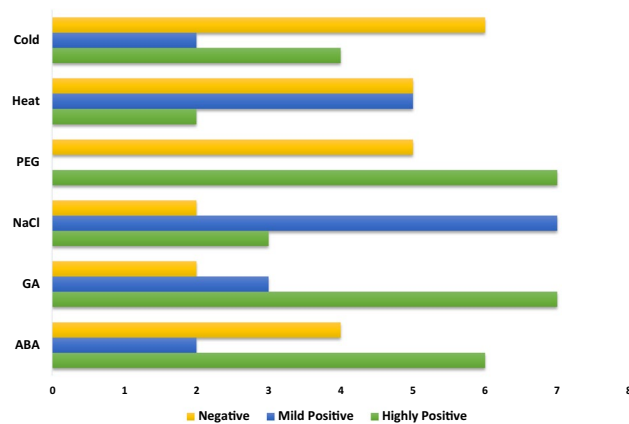
**Figure 17.** Expression analysis of *BraHB* genes under six abiotic stress treatments in *B. rapa* (A–F). Heatmap representation the *BraHB* genes under six stress treatments: ABA, GA, NaCl, PEG, heat, and cold.



**Figure 18.** Expression patterns of *BraHB* genes in response to multiple treatments.

### Materials and Methods

**Retrieval of *BraHB* Sequences.** The *B. rapa* genome sequences were downloaded from BRAD (<http://brassicadb.org/brad/>)<sup>30</sup>. The *A. thaliana* sequences were retrieved from The Arabidopsis Information Resource (TAIR) (<http://www.arabidopsis.org/>), and the sequences of rice were extracted from the Rice Genome



**Figure 19.** Pearson correlation coefficient values among the paralogous pairs of genes.

Annotation Project (<http://rice.plantbiology.msu.edu/>)<sup>49</sup>. Based on previously reported studies<sup>33</sup>, their sequences (236) were used as queries at a threshold of  $E < 1E-4$  against the *B. rapa* database. The hidden Markov model (HMM) profile was also used as a query in our study based on domain PF00046, downloaded from the Pfam 31.0 database (<https://pfam.sanger.ac.uk/>)<sup>50</sup>. Then we manually analyzed these potential sequences of the candidate *BraHB* genes by using the Simple Modular Architecture Research Tool (SMART) (<http://smart.embl-heidelberg.de/>)<sup>51</sup> and National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>) databases. Sequences that were found with obvious errors in their gene length and other complications were eliminated.

**Multiple Sequence Alignment and Phylogenetic Analysis.** For the multiple sequence alignment of *BraHB* candidate genes, we performed ClustalW, using MEGA 7 software with the default options<sup>52,53</sup>. The phylogenetic trees were constructed using the maximum likelihood (ML) method. In order to get the reliability of resulting trees, bootstrap values of 1000 replications were performed with the Jones, Taylor, and Thornton amino acid substitution model (JTT model) using MEGA 7, while keeping the other parameters as a default. We also analyzed the nucleotide divergence among different subclasses of *BraHB* with the help of MEGA 7, using the Jukes–Cantor model, and its estimation was performed with 1000 bootstrap replication.

**Calculation of  $K_a/K_s$  and Dating of the Duplication Events.** The  $K_a/K_s$  ratio was calculated among the paralog of *BraHBs*, using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) and PAL2NAL (<http://www.bork.embl.de/pal2nal/>) alignment to find synonymous and nonsynonymous substitutions<sup>54</sup>. The divergence time among the paralog pairs was calculated with the following formula:  $T = Ks/2r$ , where  $Ks$  represents the synonymous substitutions per site and  $r$  is the rate of divergence. For dicotyledonous plants, the assumption is 1.5 synonymous substitutions per site of  $10^8$  years as far as *B. rapa* is concerned<sup>55</sup>.

**Conserved Motifs, Exon-Intron Structure Analysis, and Physicochemical Parameters of *BraHB* Proteins.** To identify conserved motif in *BraHB* proteins, we used Multiple EM for Motif Elicitation (MEME) software version 4.12 with the following parameters: maximum motifs 14, minimum width 12, maximum width 100; the other parameters were set as default<sup>34</sup>. For exon–intron structure, we used Gene Structure Display Server (GSDS) 2.0 (<http://gsds.cbi.pku.edu.cn>)<sup>56</sup>. The protein property parameters, including molecular weight (MW), isoelectric points (pI), and grand average of hydropathy (GRAVY) values for each *BraHB* gene were calculated using the ProtParam tool (<http://web.expasy.org/protparam/>). The subcellular localization for *BraHB* proteins was conducted using the WoLF PSORT server (<https://wolfpsort.hgc.jp/>).

**Cis-Element Analysis and Protein Interaction Network Prediction.** The promoter sequences of *BraHB* selected as 2000 upstream bp were retrieved from the *B. rapa* genome according to generic file format (GFF). Then cis-acting regulatory elements were identified for some of the specific genes using PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>)<sup>57</sup>. The interaction network among orthologous and nonorthologous genes of *BraHB* was carried out using Cytoscape version 3.4<sup>58</sup>.

**Gene Chromosome Location, Gene Synteny Analysis, and Syntenic *BraHB* Paralogous Pair Identification.** The chromosome locations of *BraHBs* were illustrated accordingly from top to bottom with respect to their position in genome annotation by using MapChart<sup>59</sup>. For synteny gene analysis, the relationships were verified between the homologues of *A. thaliana* and subgenomes of *B. rapa* (LF, MF1, and MF2) obtained from BRAD (<http://brassicadb.org/brad/searchSynteny.php>). The Circos program was applied to demonstrate the syntenic relationships in the chromosomes of *B. rapa* and *A. thaliana*<sup>60</sup>.

**Pearson Correlation Analysis.** The Pearson correlation coefficient (PCC) analysis was performed using Excel 2013 to evaluate the PC values of the RNA-seq and paralogous genes that were used for qRT-PCR.

**Plant Material and Treatment.** In the present study, germinated seeds of Chinese cabbage (Chiifu-401-42) were grown in plastic pots containing a mixture of soil and vermiculite (3:1). The pots were then placed in an

artificial growth chamber for 5 weeks. The growth conditions were as follows: temperature was maintained at 24/16 °C, with a photoperiod of 16/8 h and relative humidity at 65–70%. The specific treatments were applied to the seedlings as follows: for heat and cold treatments, seedlings were exposed to 38 °C and 4 °C, respectively. For other stress-related treatments, plants were cultured in a nutrient medium along with a control in the following manner: (1) 100 µM ABA, (2) 100 µM GA, (3) 6000 PEG (w/v), and (4) 250 mM NaCl. Treatments were carried out in a continuous time interval of 0, 1, 6, and 12 h, respectively, with biological triplicates. Finally, all samples were frozen immediately and stored at -70 °C for further analysis.

**RNA Isolation and Expression Pattern Analysis.** Total RNA was isolated from the treated frozen leaves with Trizol (Invitrogen), following the manufacturer's instructions. RNA was reverse transcribed into cDNA using Primer Script RT reagent kit (Takara, Dalian, China) according to the manufacturer's instructions. Specific primers were used for qRT-PCR analysis using Beacon Designer 8.1, shown in Supplementary Table S9. In order to check the specificity of the primers, we used the BLAST tool against the Brassica genome to verify them. RT-PCR was performed accordingly by following the previous report<sup>46</sup>. Relative fold expression was calculated with comparative Ct methods. The expression patterns of all *BraHB* genes were analyzed based on a previous study<sup>35</sup>. Further, gene expression levels were quantified by fragments per kilobase of transcript per million mapped reads (FPKM) values and a heatmap was generated by using OmicShare Tools (<http://www.omicshare.com/>).

## References

- Ito, M., Sato, Y. & Matsuoka, M. Involvement of homeobox genes in early body plan of monocot. *International review of cytology* **218**, 1–35 (2002).
- Chan, R. L., Gago, G. M., Palena, C. M. & Gonzalez, D. H. Homeoboxes in plant development. *Biochimica et biophysica acta* **1442**, 1–19 (1998).
- Hake, S. *et al.* The role of knox genes in plant development. *Annual review of cell and developmental biology* **20**, 125–151, <https://doi.org/10.1146/annurev.cellbio.20.031803.093824> (2004).
- Hay, A., Craft, J. & Tsiantis, M. Plant hormones and homeoboxes: bridging the gap? *BioEssays: news and reviews in molecular, cellular and developmental biology* **26**, 395–404, <https://doi.org/10.1002/bies.20016> (2004).
- Byrne, M. E. Shoot meristem function and leaf polarity: the role of class III HD-ZIP genes. *PLoS genetics* **2**, e89, <https://doi.org/10.1371/journal.pgen.0020089> (2006).
- Chandler, J., Nardmann, J. & Werr, W. Plant development revolves around axes. *Trends in plant science* **13**, 78–84, <https://doi.org/10.1016/j.tplants.2007.11.010> (2008).
- Gehring, W. J., Affolter, M. & Burglin, T. Homeodomain proteins. *Annual review of biochemistry* **63**, 487–526, <https://doi.org/10.1146/annurev.bi.63.070194.002415> (1994).
- Bürglin, T. R. In *Reviews in Cell Biology and Molecular Medicine* (Wiley-VCH Verlag GmbH & Co. KGaA 2006).
- Desplan, C., Theis, J. & O'Farrell, P. H. The sequence specificity of homeodomain-DNA interaction. *Cell* **54**, 1081–1090 (1988).
- Otting, G. *et al.* Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *The EMBO journal* **9**, 3085–3092 (1990).
- Cheng, F. *et al.* Deciphering the diploid ancestral genome of the Mesoheptaploid *Brassica rapa*. *The Plant cell* **25**, 1541–1554, <https://doi.org/10.1105/tpc.113.110486> (2013).
- Bharathan, G., Janssen, B. J., Kellogg, E. A. & Sinha, N. Did homeodomain proteins duplicate before the origin of angiosperms, fungi, and metazoa? *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13749–13753 (1997).
- Bürglin, T. R. Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic acids research* **25**, 4173–4180 (1997).
- Ruberti, I., Sessa, G., Lucchetti, S. & Morelli, G. A novel class of plant proteins containing a homeodomain with a closely linked leucine zipper motif. *The EMBO journal* **10**, 1787–1791 (1991).
- Meijer, A. H. *et al.* Transcriptional repression by Oshox1, a novel homeodomain leucine zipper protein from rice. *The Plant journal: for cell and molecular biology* **11**, 263–276 (1997).
- Mukherjee, K., Brocchieri, L. & Bürglin, T. R. A Comprehensive Classification and Evolutionary Analysis of Plant Homeobox Genes. *Molecular Biology and Evolution* **26**, 2775–2794, <https://doi.org/10.1093/molbev/msp201> (2009).
- Henriksson, E. *et al.* Homeodomain leucine zipper class I genes in Arabidopsis. Expression patterns and phylogenetic relationships. *Plant physiology* **139**, 509–518, <https://doi.org/10.1104/pp.105.063461> (2005).
- Wang, Y. *et al.* The Arabidopsis homeobox gene, ATHB16, regulates leaf development and the sensitivity to photoperiod in Arabidopsis. *Developmental biology* **264**, 228–239 (2003).
- Steindler, C. *et al.* Shade avoidance responses are mediated by the ATHB-2 HD-zip protein, a negative regulator of gene expression. *Development (Cambridge, England)* **126**, 4235–4245 (1999).
- Prigge, M. J. *et al.* Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development. *The Plant cell* **17**, 61–76, <https://doi.org/10.1105/tpc.104.026161> (2005).
- Nakamura, M. *et al.* Characterization of the class IV homeodomain-Leucine Zipper gene family in Arabidopsis. *Plant physiology* **141**, 1363–1375, <https://doi.org/10.1104/pp.106.077388> (2006).
- Chew, W., Hrmova, M. & Lopato, S. Role of Homeodomain leucine zipper (HD-Zip) IV transcription factors in plant development and plant protection from deleterious environmental factors. *International journal of molecular sciences* **14**, 8122–8147, <https://doi.org/10.3390/ijms14048122> (2013).
- Hay, A. & Tsiantis, M. KNOX genes: versatile regulators of plant development and diversity. *Development (Cambridge, England)* **137**, 3153–3165, <https://doi.org/10.1242/dev.030049> (2010).
- van der Graaff, E., Laux, T. & Rensing, S. A. The WUS homeobox-containing (WOX) protein family. *Genome biology* **10**, 248, <https://doi.org/10.1186/gb-2009-10-12-248> (2009).
- Haecker, A. *et al.* Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in Arabidopsis thaliana. *Development (Cambridge, England)* **131**, 657–668, <https://doi.org/10.1242/dev.00963> (2004).
- Deyhle, F., Sarkar, A. K., Tucker, E. J. & Laux, T. WUSCHEL regulates cell differentiation during anther development. *Developmental biology* **302**, 154–159, <https://doi.org/10.1016/j.ydbio.2006.09.013> (2007).
- Tan, Q. K. & Irish, V. F. The Arabidopsis zinc finger-homeodomain genes encode proteins with unique biochemical properties that are coordinately expressed during floral development. *Plant physiology* **140**, 1095–1108, <https://doi.org/10.1104/pp.105.070565> (2006).
- Qi, J. *et al.* Reference Gene Selection for Real-Time Quantitative Polymerase Chain Reaction of mRNA Transcript Levels in Chinese Cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Plant Molecular Biology Reporter* **28**, 597–604, <https://doi.org/10.1007/s11105-010-0185-1> (2010).

29. Zhang, F. L., Takahata, Y., Watanabe, M. & Xu, J. B. Agrobacterium-mediated transformation of cotyledonary explants of Chinese cabbage (*Brassica campestris* L. ssp. *pekinensis*). *Plant Cell Reports* **19**, 569–575, <https://doi.org/10.1007/s002990050775> (2000).
30. Wang, X. *et al.* The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* **43**, 1035–1039, <https://doi.org/10.1038/ng.919> (2011).
31. Town, C. D. *et al.* Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *The Plant cell* **18**, 1348–1359, <https://doi.org/10.1105/tpc.106.041665> (2006).
32. Liu, S. *et al.* The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature communications* **5**, 3930, <https://doi.org/10.1038/ncomms4930> (2014).
33. Jain, M., Tyagi, A. K. & Khurana, J. P. Genome-wide identification, classification, evolutionary expansion and expression analyses of homeobox genes in rice. *FEBS Journal* **275**, 2845–2861, <https://doi.org/10.1111/j.1742-4658.2008.06424.x> (2008).
34. Bailey, T. L. *et al.* MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* **37**, W202–W208, <https://doi.org/10.1093/nar/gkp335> (2009).
35. Tong, C. *et al.* Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa*. *BMC Genomics* **14**, 689, <https://doi.org/10.1186/1471-2164-14-689> (2013).
36. Llorca, C. M., Potschin, M. & Zentgraf, U. bZIPs and WRKYs: two large transcription factor families executing two different functional strategies. *Front Plant Sci* **5** Art 169 <https://doi.org/10.3389/fpls.2014.00169> (2014).
37. Knight, H. & Knight, M. R. Abiotic stress signalling pathways: specificity and cross-talk. *Trends in Plant Science* **6**, 262–267, [https://doi.org/10.1016/S1360-1385\(01\)01946-X](https://doi.org/10.1016/S1360-1385(01)01946-X) (2001).
38. Wang, W., Vinocur, B. & Altman, A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* **218**, 1–14, <https://doi.org/10.1007/s00425-003-1105-5> (2003).
39. Khan, N., Hu, C.-m., Amjad Khan, W. & Hou, X. Genome-Wide Identification, Classification, and Expression Divergence of Glutathione-Transferase Family in *Brassica rapa* under Multiple Hormone Treatments. *BioMed Research International* **2018**, 19, <https://doi.org/10.1155/2018/6023457> (2018).
40. Cheng, F. *et al.* BRAD, the genetics and genomics database for Brassica plants. *BMC Plant Biology* **11**, 136, <https://doi.org/10.1186/1471-2229-11-136> (2011).
41. Lynch, M. & Conery, J. S. The Evolutionary Fate and Consequences of Duplicate Genes. *Science* **290**, 1151–1155, <https://doi.org/10.1126/science.290.5494.1151> (2000).
42. Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D. & May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol* **4**, 10, <https://doi.org/10.1186/1471-2229-4-10> (2004).
43. Town, C. D. *et al.* Comparative Genomics of *Brassica oleracea* and *Arabidopsis thaliana* Reveal Gene Loss, Fragmentation, and Dispersal after Polyploidy. *The Plant cell* **18**, 1348–1359, <https://doi.org/10.1105/tpc.106.041665> (2006).
44. Vision, T. J., Brown, D. G. & Tanksley, S. D. The Origins of Genomic Duplications in *Arabidopsis*. *Science* **290**, 2114–2117, <https://doi.org/10.1126/science.290.5499.2114> (2000).
45. Sah, S. K., Reddy, K. R. & Li, J. Abscisic Acid and Abiotic Stress Tolerance in Crop Plants. *Frontiers in Plant Science* **7** <https://doi.org/10.3389/fpls.2016.00571> (2016).
46. Song, X., Li, Y. & Hou, X. Genome-wide analysis of the AP2/ERF transcription factor superfamily in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *BMC genomics* **14**, 573, <https://doi.org/10.1186/1471-2164-14-573> (2013).
47. Wang, W. *et al.* Genome-wide Analysis and Expression Divergence of the Trihelix family in *Brassica rapa*: Insight into the Evolutionary Patterns in Plants. *Scientific Reports* **7**, 6463, <https://doi.org/10.1038/s41598-017-06935-0> (2017).
48. Song, X.-M. *et al.* Genome-wide analysis of the bHLH transcription factor family in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Molecular Genetics and Genomics* **289**, 77–91, <https://doi.org/10.1007/s00438-013-0791-3> (2014).
49. Kawahara, Y. *et al.* Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 4, <https://doi.org/10.1186/1939-8433-6-4> (2013).
50. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic acids research* **42**, D222–D230, <https://doi.org/10.1093/nar/gkt1223> (2014).
51. Letunic, I., Doerks, T. & Bork, P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic acids research* **40**, D302–D305, <https://doi.org/10.1093/nar/gkr931> (2012).
52. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL\_X Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools. *Nucleic acids research* **25**, 4876–4882, <https://doi.org/10.1093/nar/25.24.4876> (1997).
53. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **33**, 1870–1874, <https://doi.org/10.1093/molbev/msw054> (2016).
54. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research* **34**, W609–W612, <https://doi.org/10.1093/nar/gkl315> (2006).
55. Koch, M. A., Haubold, B. & Mitchell-Olds, T. Comparative Evolutionary Analysis of Chalcone Synthase and Alcohol Dehydrogenase Loci in *Arabidopsis*, *Arabis*, and Related Genera (Brassicaceae). *Molecular biology and evolution* **17**, 1483–1498, <https://doi.org/10.1093/oxfordjournals.molbev.a026248> (2000).
56. Hu, B. *et al.* GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31**, 1296–1297, <https://doi.org/10.1093/bioinformatics/btu817> (2015).
57. Lescot, M. *et al.* PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic acids research* **30**, 325–327, <https://doi.org/10.1093/nar/30.1.325> (2002).
58. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504, <https://doi.org/10.1101/gr.1239303> (2003).
59. Voorrips, R. E. M. C. Software for the Graphical Presentation of Linkage Maps and QTLs. *Journal of Heredity* **93**, 77–78, <https://doi.org/10.1093/jhered/93.1.77> (2002).
60. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Research* **19**, 1639–1645, <https://doi.org/10.1101/gr.092759.109> (2009).

## Acknowledgements

This study was financially supported by the Independent Innovation Fund for Agricultural Science and Technology of Jiangsu Province (CX (15) 1015) and North Jiangsu Special Project of Jiangsu Province (SZ-LYG2017027). The author authors express their gratitude to Dr. Rashid Latif for helpful comments for improving the quality of English.

## Author Contributions

N.K., H.C., W.A.K., H.K., Z.Z. and D.H. conceived and designed the experiments. N.K. and W.A.K. performed the experiments. N.K., H.C., and W.W. analyzed data. N.K. and X.H. wrote the manuscript. N.K., H.C. and W.A.K. revised the manuscript. All authors read and approved the manuscript.



## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-34448-x>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018