

Genome-wide Survey of the bHLH Super Gene Family in *Brassica napus*

Yunzhuo Ke

Southwest University

Yunwen Wu

Southwest University

Hongjun Zhou

Southwest University

Ping Chen

Southwest University

Mangmang Wang

Southwest University

Mingming Liu

Southwest University

Pengfeng Li

Southwest University

Jing Yang

Southwest University

Jiana Li

Southwest University

Hai Du (✉ haidu81@126.com)

Southwest University <https://orcid.org/0000-0003-4274-5724>

Research article

Keywords: Brassica napus, bHLH transcription factor, root, gene expression

Posted Date: February 26th, 2020

DOI: <https://doi.org/10.21203/rs.2.16021/v4>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Plant Biology on March 14th, 2020.
See the published version at <https://doi.org/10.1186/s12870-020-2315-8>.

Abstract

Background: The basic helix-loop-helix (bHLH) gene family is one of the largest transcription factor families in plants and is functionally characterized in diverse species. However, less is known about its functions in the economically important allopolyploid oil crop, *Brassica napus*.

Results: We identified 602 potential bHLHs in the *B. napus* genome (BnabHLHs) and categorized them into 35 subfamilies, including seven newly separated subfamilies, based on phylogeny, protein structure, and exon-intron organization analysis. The intron insertion patterns of this gene family were analyzed and a total of eight types were identified in the bHLH regions of BnabHLHs. Chromosome distribution and synteny analyses revealed that hybridization between *Brassica rapa* and *Brassica oleracea* was the main expansion mechanism for BnabHLHs. Expression analyses showed that BnabHLHs were widely in different plant tissues and formed seven main patterns, suggesting they may participate in various aspects of *B. napus* development. Furthermore, when roots were treated with five different hormones (IAA, auxin; GA3, gibberellin; 6-BA, cytokinin; ABA, abscisic acid and ACC, ethylene), the expression profiles of BnabHLHs changed significantly, with many showing increased expression. The induction of five candidate BnabHLHs was confirmed following the five hormone treatments via qRT-PCR. Up to 246 BnabHLHs from nine subfamilies were predicted to have potential roles relating to root development through the joint analysis of their expression profiles and homolog function.

Conclusion: The 602 BnabHLHs identified from *B. napus* were classified into 35 subfamilies, and those members from the same subfamily generally had similar sequence motifs. Overall, we found that BnabHLHs may be widely involved in root development in *B. napus*. Moreover, this study provides important insights into the potential functions of the BnabHLHs super gene family and thus will be useful in future gene function research.

Background

Transcription factor genes are widely distributed in the eukaryotic kingdom and usually contain two different functional domains involved in DNA binding and transcriptional activities [1, 2]. The basic helix-loop-helix (bHLH) transcription factor is characterized by a conserved 50–60 amino acid (aa) sequence responsible for DNA binding. This sequence consists of two main regions, namely the basic and the helix-loop-helix (HLH) regions. The basic region is a 10–15 aa region at the N-terminus of the protein, which functions as a DNA recognition motif and allows the binding of the HLH region [3]. The HLH region, which is composed of two relatively conserved amphipathic helices linked by a divergent loop, is approximately 40 aa long with many hydrophobic amino acids, which contributes to its function in the dimerization of HLH regions [4, 5].

The *bHLH* gene family existed in land plants over 400 million years ago and were highly conserved during plant evolution [6]. As one of the largest transcription factor gene families in plants, the number of *bHLH*

genes appears to have increased as plants evolved. For example, there was only one *bHLH* gene in *Cyanidioschyzon Merolae* [6], 98 in moss [7], 208 in *Zea mays* [8], 167 in *Arabidopsis* [7], 159 in tomato [9], and 230 in Chinese cabbage [10]. The substantial increase in the number of *bHLH* genes was concomitant with their increased involvement in diverse physiological and developmental processes, with the majority being involved in metabolism and development. For instance, *Arabidopsis AtbHLH045/MUTE* controls sequential cell fate [11]; *SibHLH22* in tomato promotes early flowering and accelerates fruit ripening [12]; the bHLH transcription factor SPATULA (*SPT*) homologs are required for either carpel development or are involved in endocarp margin development in *Arabidopsis* and *Prunus persica* [13, 14]; the bHLH transcription factor *MYC2* homologs regulate sesquiterpene and artemisinin biosynthesis in various species such as *Aquilaria sinensis* and *Artemisia annua* [15, 16]. Meanwhile, some *bHLHs* are related to abiotic stress response, including cold, drought, and salt stresses. For example, the bHLH transcription factor PHYTOCHROME-INTERACTING FACTOR 4 (*PIF4*) in *Arabidopsis* mediates plant stomatal in response to high temperatures [17]; *StbHLH1* in potato also responds to high temperature by regulating anthocyanin biosynthesis [18]; alternatively, *MdbHLH3* in *Malus domestica* responds to low temperature [19]. *bHLHs* also respond to various hormones. For instance, *Arabidopsis MYC2* is well known for its conserved roles in abscisic acid (ABA), jasmonic acid (JA), and light signaling pathways [20-22]. Moreover, *bHLHs*, e.g., *JAMs* and *TT8* are jasmonate-responsive transcription factors involved in secondary metabolism [23]. *bHLHs* also contribute to Fe homeostasis in *Arabidopsis* and rice [24, 25]. Notably, bHLH proteins tend to function in protein complexes. For example, the *Arabidopsis* bHLH gene, *GLABRA3 (GL3)/ENHANCER OF GLABRA3 (EGL3)*, is widely known as an epidermal cell fate specification gene and hair root regulator that functions by forming a protein complex with *TTG1* (WD40 repeat protein) and *WER/GL1 (R2R3-MYB protein)* (MBW complex) [26-28].

In this study, we identified 602 *bHLHs* in the important economy crop, *Brassica napus* (*BnabHLHs*), and mapped them to the 19 *B. napus* chromosomes. According to our phylogenetic analysis and gene functions, the *B. napus bHLH* gene family is divided into 35 subfamilies with seven subfamilies being newly identified. Conserved non-bHLH motifs along with intron insertion pattern analyses further support our classification. Chromosome localization combined with synteny analyses revealed the expansion mechanism of *BnabHLHs* in *B. napus*. Expression profile analyses revealed the potential functions of *BnabHLHs*, with a focus on their possible roles in roots. qRT-PCR analysis confirmed the features of the ortholog functional genes by hormone induction in roots.

Results

B. napus* contains a large number of *bHLHs

Referred to our previously described method [29], we performed BLASTP searches (e values of < 1.0) against the proteome data of the two sequenced *B. napus* genomes (Darmor-*bzh* [30], and Zhongshuang11, ZS11 [31] ecotypes) respectively, using the representative protein sequences of

Arabidopsis bHLH proteins [7] as queries. We found that the sequence quality (including sequence integrity, sequence number and genome annotation information) from Darmor–*bzh* ecotype was better than those from ZS11 by sequence comparative analyses between these two ecotypes.

As a result, 613 non-redundant putative bHLH encoding genes were obtained from *B. napus* genome (Darmor–*bzh*). Subsequently, the protein sequences of putative genes were examined by ExPASy to ensure that the candidates contained the bHLH domain. Consequently, 11 genes were excluded from our dataset as no bHLH domain was identified by ExPASy analysis. Meanwhile, the sequence information of 65 *BnabHLHs* from Darmor–*bzh* was corrected by the data from the ZS11 genome (Additional file 1: Table S1). Finally, a total of 602 *BnabHLHs* with relatively complete open reading frames [32] were obtained in this study, accounting for approximately 0.60% of the *B. napus* protein-coding genes. The corresponding proportions in wheat, rice, maize, and *Arabidopsis* are 0.55, 0.47, 0.59, and 0.61%, respectively [16]. The candidate *BnabHLHs* were then named according to their chromosomal distribution (Additional file 1: Table S1). Analysis of *BnabHLH* physicochemical properties showed that the *BnabHLH* proteins varied in length from 63 to 1440 aa; their molecular weight ranged from 6.9 (BnabHLH105) to 165 kDa (BnabHLH381); and their isoelectric points ranged from 4.36 (BnabHLH562) to 11.79 (BnabHLH023). Subcellular localization analysis demonstrated that all *BnabHLHs* were located in the nucleus (Additional file 1: Table S1).

For further comparative analysis across different species, we identified 245 *bHLHs* in the *B. oleracea* genome by the same method (Additional file 2: Table S2). Sequence information of the *bHLHs* in other species like *Arabidopsis*, *B. rapa*, tomato, potato, and rice was obtained from previously published studies [8-10, 16, 33].

Sequences characteristics of the bHLH domains of *BnabHLH* proteins

To investigate the aa sequence features, we performed multiple sequence alignment analyses of the 602 bHLH domains of candidate *BnabHLHs*. The results were visualized using Weblogo online software.

Our results showed that the length of the *BnabHLH* domains was approximately 55 aa, ranging from 39–57 aa. The bHLH domain was generally conserved in this gene family in *B. napus*, with ten residues being identified as having a conservation of more than 70% in the bHLH domains (Fig. 1a), including four located in the basic region, five in the two helix regions, and one in the loop region. Consistent with other studies [7-10], Leu-25 was the most conserved residue, with a conservation of almost 100% (Additional file 3: Table S3), indicating its essential role in bHLH proteins. Interestingly, Phe-30 was partly substituted

by Ser in *Arabidopsis*, rice, and tomato, among others, however, no such substitution was observed in *B. napus*, *B. oleracea*, or *B. rapa*, (Additional file 3: Table S3), suggesting a higher conservation and/or close relationship in these three species.

To further characterize the BnabHLH sequence features, the criterion given by Massari and Murre [34] was used (Fig. 1). Our results showed that the 602 BnabHLHs were separated into two major categories according to their bHLH domain sequence profiles: 132 (21.93%) atypical BnabHLHs (non-DNA-binding proteins) and 470 (78.07%) typical BnabHLHs (Fig. 1). The latter category was further divided into three categories, including 300 (49.83%) G-box binding proteins, 103 (17.11%) E-box binding proteins, and 67 (11.13%) non-E-box binding proteins (Fig. 1). The sequences of the 132 atypical BnabHLHs were divergent in their basic region but were relatively conserved in their HLH region, especially in the loop region (Fig. 1d). A similar situation was found for the non-E-box binding proteins (Fig. 1c), suggesting its close relationship to the atypical BnabHLHs. In contrast, the residues in the basic region of the E-box/G-box DNA binding proteins were more conserved than the residues in the HLH region (Fig. 1b, c).

Protein structures of BnabHLHs were conserved in each subfamily

To determine the evolutionary relationship between the *BnabHLHs* from the *Brassicaceae* species, we constructed a Neighbour-Joining (NJ) phylogenetic tree on the basis of the alignment of 769 bHLH domains from *B. napus* (602) and *Arabidopsis* (167).

The 769 bHLH proteins were divided into 35 subfamilies, which is the largest number of subfamilies found to date (Fig. 2a). Among these subfamilies, 28 were found previously [7], with seven (S33-S39) being newly identified in this study. Two previously reported subfamilies (S6 and S8) were not found in this study as they were only found in lower plants (moss and algae) [7]. Compared with the division of AtbHLHs, the S5 subfamily in *B. napus* was divided into S5 and S33; S17 was divided into S17 and S34; S21 was divided into S21 and S35; S24 was divided into S24, S36, and S37; S30 was divided into S30 and S38; whereas the orthologs of S39 in *Arabidopsis* were previously defined as orphan genes [7], which were defined as a new subfamily in this study. The distribution of BnabHLHs in the 35 subfamilies was biased, varying from two (S22 and S38) to 62 genes (S25). In addition, the BnabHLHs of different DNA binding types had a biased distribution tendency among different subfamilies as well, but the BnabHLHs in a given subfamily usually shared the same DNA binding type (Fig. 2b). A total of 11 subfamilies (S2, S3, S5, S7, S10, S11, S13, S14, S24, S25, and S26) contained G-box binding proteins; five subfamilies (S1, S9, S17, S27, and S37) contained E-box-binding proteins; three subfamilies (S20, S23, and S39) contained non-E-box-binding proteins; while seven subfamilies (S16, S21, S22, S33, S34, S35, and S38) contained non-DNA-binding proteins (Fig. 2b).

We subsequently used the MEME tool to discover the non-bHLH domains and explore their distribution patterns within each subfamily. A total of 27 conserved motifs of variable lengths (8–103 aa) was obtained from this analysis (Fig. 2c Additional file 4: Table S4). Motifs 1 and 2 were distributed in all BnabHLHs and consisted of the basic and two helix regions of the bHLH domains, respectively. The loop region was located between motif 1 and 2, indicating that this region was more variable than the basic and helix regions. Outside the bHLH domain, members of the same subfamily generally shared several of the same motifs. For example, all BnabHLHs in subfamily S9 contained motif 22; proteins in subfamily S26 all contained motif 5 (Fig 2b, Additional file 4: Table S4). Moreover, some motifs have been previously characterized and were defined as additional functional motifs. For instance, motifs 4, 8, and 20 were detected in many proteins within subfamilies S2 and S5 in various species, such as TabHLH239, AtMYC2, TabHLH184, and ZmbHLH103, which were significantly matched with an ACT domain that contributed to the recruitment of the C1 R2R3-MYB factor to the C1 binding sites located in the promoters of flavonoid biosynthetic genes [35]. Meanwhile, motif 6 in these two subfamilies was also conserved, and overlapped with the MIR and MYC_N domains that interact with jasmonate ZIM-domains (JAZs) [36]. Some motifs were subfamily-specific, however their functions are still unclear (Fig. 2c).

Intron insertion patterns of *BnabHLHs* were conserved within each subfamily

The intron and exon structure is an important clue to understand the gene evolutionary relationship and functional diversification within a gene family [37]. The intron and exon patterns of candidate *BnabHLHs* are determined by comparing their full-length CDS and DNA sequences using the GSDS web server [38].

A total of eight intron insertion patterns (pattern a to k) were observed in the bHLH domains in *B. napus*, containing 0 to 2 intron insertion sites (Fig. 3). The nomenclature of the intron insertion patterns of *BnabHLHs* is referred to in the study by Carretero-Paulet et al. [7]. In this study, the previously defined pattern a and b (Carretero-Paulet et al., 2010) were characterized as the same type because they share the same insertion sites and phase, and therefore are uniformly named as pattern a. Similarly, pattern d and f (Carretero-Paulet et al., 2010) were uniformly named as f. The intron insertion positions were distributed across the basic and/or HLH regions in the bHLH domain. Among these insertions, those observed in the basic and loop regions were more conserved, while those in the helix regions were variable across different patterns. The intron insertion sites in the basic and helix regions were located at three highly conserved residues, Arg-11 (the E-box recognition site), Phe-21, and Lys-33 (Fig. 3). Furthermore, the intron insertion sites of most patterns were conserved, except pattern j (Fig. 3). Patterns a, c, e, and f were similar, thus they are likely to be homologous, whereas pattern f lacks the first intron as compared with pattern a, pattern e lacks the second intron, and pattern c has the second intron inserted at L-50 as compared with pattern e. A similar situation was observed in patterns h and i. Meanwhile, phylogenetic

analyses showed a close relationship between intron insertion patterns a, c, e, and f, further confirming their close relationship. Moreover, patterns k and i appeared to be the ancestral types because they existed in members from algae [7]. Patterns a, f, and k were the three types that accounted for the majority of *BnabHLHs* (41.2%, 26.2%, and 20.1%, respectively). This trend is similar to the results in other species, such as *Arabidopsis*, rice, potato, poplar, and tomato [8-10, 16, 33]. Accordingly, these three patterns were obtained for many subfamilies while the remaining patterns (i.e., patterns c, e, h, i, and j) existed in only one or two subfamilies, indicating a different expansion trend.

The distributions of intron insertion patterns are generally conserved within most subfamilies. For example, members in subfamilies S12, S10, S11, S7, S9, S5, S33, S2, S1, S13, S23, and S38 contained pattern f, except for several genes that may have been absent due to incomplete genomic annotation information (Fig. 2b). The conservation of intron insertion patterns of *BnabHLHs* within each subfamily provides an independent criterion for the reliability of our phylogenetic analyses (Fig. 2b). Interestingly, the intron insertion patterns of *BnabHLHs* was almost the same as their orthologs in *Arabidopsis*. The only exception was in the S27 subfamily which contains pattern a for *B. napus* members while their homologs in *Arabidopsis* is pattern f [7]. We further compared these results to other species, e.g. rice [7], and found that it should be pattern a for the homologs in this subfamily, including *Arabidopsis* homologs (*At080*, *At081*, *At122*, *At128*, *At129*, and *At130*).

Overall, intron insertion patterns in *BnabHLHs* were conserved within most subfamilies and corresponded to *AtbHLH* orthologs as well. Moreover, the intron insertion sites in the basic and loop regions were more conserved than those in the helix regions.

Syntenic analyses revealed duplication events and the expansion mechanism of *BnabHLHs*

In this study, up to 602 *BnabHLHs* were identified, which is significantly higher than the gene number in lower plants, like *Volvox carteri* which only has three [6]. This indicates that there was a large-scale expansion of this gene family that occurred during plant evolution. To explore the expansion mechanism of this gene family in *B. napus*, the chromosomal locations and syntenic relationships of *BnabHLHs* were analyzed based on the genome information from Genoscope and CoGe databases [39].

Chromosomal location analysis showed that there are 294 and 306 *BnabHLHs* in A_n - and C_n -subgenomes, respectively, indicating no biased tendency between these two subgenomes (Fig. 4a). The *BnabHLHs* are distributed on all the 19 *B. napus* chromosomes, but the genes on each chromosome are uneven within the two subgenomes. For example, in A_n -subgenome, A10 has a minimum of 14

BnabHLHs, while A09 has a maximum of 41 genes. In C_n -subgenome, C06 contains a minimum of 19 *BnabHLHs*, whereas C03 has as many as 48 genes. The average number of *BnabHLHs* in A_n - and C_n -subgenomes were 27.4 and 29.9, respectively.

Based on synteny analyses, 475 of the 602 *BnabHLHs* have syntenic relationships, 382 of which were inherited from *B. rapus* or *B. oleracea* genomes (Additional file 5: Table S5). In contrast, only 79 genes (7.8%) of 79 syntenic pairs underwent segmental duplication in the *B. napus* genome, and 34 genes (5.8%) from 28 syntenic pairs underwent tandem duplications (Fig. 4b). These results that the majority of *BnabHLHs* were obtained from allopolyploidy between *B. rapa* and *B. oleracea*. Accordingly, we found that the genome-wide duplicated genes make up the largest number of *BnabHLHs* in the majority of subfamilies. However, segmental duplicated genes make up the biggest number *BnabHLHs* in subfamilies S10 and S25 (nine genes per subfamily), and the tandem duplicated genes have the largest number in the S12 subfamily (11 genes). Furthermore, the *BnabHLHs* with intron pattern f expanded most in *B. napus* (31 duplicates), contributing to the largest proportion of *BnabHLHs* (Additional file 1: Table S1).

Taken together, the main expansion mechanism of *BnabHLHs* is whole-genome duplication (allopolyploidization), while segmental and tandem duplication events preferentially occurred in certain subfamilies with specific intron patterns.

Expression profiles of *BnabHLHs* varied widely and were conserved within each subfamily

Gene expression patterns under different conditions can often give an indication of gene function. In order to explore the possible functions of the *BnabHLHs*, the temporal and spatial transcriptomes of the 602 *BnabHLHs* in 50 *B. napus* tissues of the root, leaf, flower, and seed at different developmental stages were characterized using RNA-seq (BioProject ID: PRJNA358784).

A total of 47 *BnabHLHs* (7.79%) were excluded from our analysis with FPKM < 1, which may be pseudogenes or expressed only at specific developmental stages or under special conditions. The remaining genes (555 genes) have relatively high confidence expression levels (FPKM \geq 1), the majority of which showed preferential expression in one or a few tissues/organs. Few genes were constitutively expressed in all tissues or organs tested, suggesting that this gene family tends to play regulatory roles at specific developmental stages or tissues. The wide expression profile of *BnabHLHs* suggests that they possess diverse roles in *B. napus*. The expression patterns of *BnabHLHs* were summarized into seven main blocks (I to VII) (Fig. 5). The *BnabHLHs* in blocks I to VII show obvious tissue-specific expression

patterns: the genes in block I were highly expressed in seed coats; genes in block II were highly expressed in the root and stem tissues; genes in block III were mainly expressed in the vegetative organs e.g., root, stem, leaf and silique pericarp; the genes in block IV were primarily expressed in organs at the seedling stage, such as in the germinating seed (24 to 72 h), root, and hypocotyl; the genes in block V were mainly expressed in the pistil, silique pericarp, seed coat, and seed tissues; the genes in block VI had higher expression levels in the pistil and inflorescence tip tissues; whereas the genes in block VII were mainly expressed in flower tissues including capillament, petal, and stamen. Overall, there were 153, 136, 129, 137, 149 and 145 *BnabHLHs* having relatively high expression levels in the stem, hypocotyl leaf, silique pericarp, flower and seed respectively, while up to 301 genes were highly expressed in the roots, indicating that *BnabHLHs* may possess some previously unknown functions in the roots.

Typically, members of a given subfamily generally exhibit the same/similar expression profile. For example, members of the S2 subfamily were mainly expressed in the root and leaf at seedling, budding, and flowering stages (Additional file 6: Table S6). Moreover, the *bHLHs* in the same subfamily likely process the same or similar expression profiles across different species, and thus may share conserved functions during evolution. For example, *AtbHLH155/CPU* and *AtbHLH156/LHW* in the S23 subfamily play an essential role in establishing vascular cells and the size of vascular initial population in the root meristem [40]. The corresponding homologs in *B. napus* were also expressed in roots (Additional file 6: Table S6); their expression being relatively low in roots after germinating for 24 to 72 h but high in mature roots, which corresponds to their possible function in root vascular cells. Moreover, the expression profile of 28 pairs of tandem duplication genes, and 79 pairs of segmental duplication genes show similar expression patterns, indicating the functional redundancy of the duplicates (Additional file 5 and 6: Table S5 and S6).

Overall, the *BnabHLHs* were widely expressed but at different levels within *B. napus*. This was especially the case in the root, which provides an important clue as to the possible roles of this gene family. Moreover, *BnabHLHs* from a particular subfamily tend to possess conserved expression patterns and high structural similarity across different species, indicating a possible conservation of function during evolution.

Many *BnabHLHs* were induced in hormone-treated roots

As discussed above, several *BnabHLHs* were highly expressed in the roots of *B. napus*, suggesting that they play a role(s) in some root-related biological process(es). To further explore their functional characteristics in roots, a comprehensive expression analysis of candidate *BnabHLHs* in roots treated

with five hormones (IAA, auxin; GA₃, gibberellin; 6-BA, cytokinin; ABA, abscisic acid and ACC, ethylene) was performed, based on previously acquired RNA-seq data (BioProject ID PRJNA608211).

Our results showed that many *BnabHLHs* (221, 36.7%) responded to more than one hormone treatment in roots, and most of these genes were clustered in ten subfamilies (S2, S16, S18, S19, S20, S21, S25, S27, S35, and S39) taking up over 50% of the genes in each subfamily (Additional file 7: Table S7). Meanwhile, the genes in 14 subfamilies (S1, S4, S5, S7, S9, S10, S12, S13, S14, S15, S24, S26, S34, and S38) were partly induced by the five hormone treatments. Interestingly, *BnabHLHs* in subfamilies S16, S18, and S21 were upregulated in roots under all five hormone treatments, while *BnabHLHs* in subfamilies S7, S23, and S34 were all downregulated. Notably, *BnabHLHs* in S12, which had low or no expression in roots (Fig. 5) were highly expressed after the five hormone treatments (Additional file 7: Table S7). Conversely, some genes that were normally highly expressed in roots were unresponsive to the hormone treatments. Indeed, *BnabHLHs* in S3, S11, S17, S22, S23, S28, S30, S31, S33, S36, and S37 had hardly any response to the hormone treatments in roots (Additional file 7: Table S7). Overall, most *BnabHLH* genes responded to hormone treatments in roots, indicating that they have important roles in hormone response in *B. napus* roots.

To further verify the RNA-Seq results, five *BnabHLHs* that had high expression levels in roots (Additional file 6: Table S6) and obviously responded to hormone inductions (Additional file 7: Table S7) were selected to analyze their expression profiles under hormone inductions by qRT-PCR. Among them, three genes (*BnabHLH033*, *BnabHLH041*, and *BnabHLH269*) are orthologs of the *Arabidopsis ILR3* gene which was demonstrated to participate in auxin-conjugate metabolism [41], and two genes (*BnabHLH453* and *BnabHLH126*) are orthologs of the *Arabidopsis MYC2* gene which is involved in ABA, JA, and light signaling pathways. As shown in Fig. 6, our results by qRT-PCR were similar to that observed in the RNA-seq analyses. These five *BnabHLHs* were positively induced by all five hormone treatments, with *BnabHLH126* generally having a higher expression level than the others. Moreover, the genes from the same clade show similar expression patterns under certain hormone treatments. For example, *BnabHLH033*, *BnabHLH041*, and *BnabHLH269* from the *ILR3* clade in the S4 subfamily have similar expression patterns under IAA, ACC, and ABA treatments (Fig. 6a–c). However, their expression profiles are different under GA₃ and 6-BA inductions, where *BnabHLH033* was significantly downregulated while *BnabHLH041* and *BnabHLH269* were significantly upregulated (Fig. 6d–e). Similarly, the expression of *BnabHLH453* and *BnabHLH126* from the *MYC2* clade in the S2 subfamily showed similar expression patterns under the five hormone treatments (Fig. 6). In addition, cis-acting elements analysis revealed that the promoter regions of these five *BnabHLHs* contain more than one cis-acting element that is related to the hormone response (Additional file 8: Table S8). This further supports our above results.

Together, our expression profile analyses revealed that a large proportion of *BnabHLHs* was induced by more than one hormone treatment in roots, indicating that this gene family may be important in root development in *B. napus*. Moreover, our qRT-PCR experiments confirmed the hormone-induced expression characteristics of *BnabHLH033*, *BnabHLH041*, *BnabHLH269*, *BnabHLH453*, and *BnabHLH126* which provided a valuable foundation for future functional research.

Discussion

Phylogenetic tree and subfamily division

As an important plant transcription factor super gene family, genome-wide analysis of *bHLHs* has already been conducted in a number of species [7-10]. However, we found that this gene family in plants still lacked a uniform subfamily classification. To date, there are two typical systematic subfamily divisions of this gene family in plants based on multiply species data, which were carried out by Pires and Dolan [6] and Carretero-Paulet et al. [7], respectively. Although both of these studies divided this gene family into 28 subfamilies, their results are in fact quite different, as only 19 subfamilies are the same between them while nine are different. The differences between these two divisions may result from the amount of *bHLHs* included in their phylogenetic analyses, as well as the criterion applied in the classification on the basis of the phylogenetic trees. In Pires and Dolan study [6], they adopted the *Arabidopsis* *bHLHs* proposed by Heim et al. [3] which included 118 *AtbHLHs*. However, there are a large number of *bHLHs* that were missed in this study due to the restriction of the *Arabidopsis* genome version used. By contrast, there was a total of 162 *AtbHLHs* in the study by Pires and Dolan [6] resulting in 12 newly identified orphan genes, such as *AtbHLH022/DYT1*, *AtbHLH159/P1r2*, and *AtbHLH102/BIM2* etc. Generally, the division of a gene family is based on the topology and bootstrap value of the phylogenetic tree [42, 43]. However, we found that subfamilies VII (a+b), IX, and IIIf in the results of Pires and Dolan [6] did not consist of a consensus node, but were across different branches/clades instead. By contrast, the situation was well defined in the results of Carretero-Pault et al. [7], indicating that these results may be more credible. Furthermore, most of the orphan genes (62 genes) identified by Pires and Dolan [6] were classified into different subfamilies by Carretero-Paulet et al. [7]. Consequently, an obvious decrease in the orphan gene proportions (total 15 genes, 2%) was observed, suggesting that their method offered a better solution for the classification of orphan genes. To date, this gene family has been characterized in various plants, such as *Z. mays* [8], tomato [9], and *B. rapa* [10], and the classification used in most of these studies referred to that of Pires and Dolan [6]. As a result, the subfamily classification in these studies was somewhat inadequate. For instance, in *Z. mays*, subfamilies VII, VIII, IX, XVI, and XVIII were not clustered in a consensus node [8]; the same situations were observed in *B. rapa* for subfamilies IIIf and Ib (2) [10]. Thus, we found that the criterion set forth by Carretero-Pault et al. [7] is more credible.

In this study, the 602 candidate *BnabHLHs* were classified into 35 subfamilies, using the method set forth by Carretero-Pault et al. [7] (Fig. 2). Among these subfamilies, one (S39) was newly identified

(*AtbHLH151* homologs), and 11 were separated from five former subfamilies that had a relatively low bootstrap value in the study by Carretero-Pault et al. [7], including four subfamilies were separated into two new subfamilies (S5 and S33; S17 and S34; S21 and S35; S30 and S38) and one subfamily was divided into three new subfamilies (S24, S36, and S37) in our study. These differences may be attributed to the greater availability of sequences from *Brassicaceae* species being applied that have a close evolutionary relationship.

Several other genetic characteristics must be considered when performing subfamily classifications, including the presence of highly conserved intron patterns and motif distribution within each subfamily [42, 43]. In this study, the bHLH sequence characteristics within each subfamily were also highly conserved, which independently supports our phylogeny analysis and classification results. For example, subfamily S39 had a relatively high bootstrap value (Fig. 2a) and shared a conserved non-E-box and k intron pattern (Fig. 2b, c); subfamily S24, which was defined by Carretero-Pault et al. [7], was separated into subfamilies S24, S36, and S37 in this study (Fig. 2a), and these three new subfamilies contained a subfamily-specific DNA binding type, intron insertion patterns, and motifs (Fig. 2b, c); the former S17 subfamily was separated into S17 and S34 subfamilies in the present study (Fig. 2a), and the *BnabHLHs* in the new S17 subfamily had an E-box in the DNA binding domain while those *BnabHLHs* in S33 were all non-DNA binding types; similarly, the former S5 subfamily in the study by Carretero-Pault et al.[7] was divided into two new subfamilies (S5 and S33) in our study as well.

Taken together, our study provides a more systematic classification of bHLH proteins in plants, which lays a good foundation for exploring the evolutionary characteristics of the bHLH gene family.

The potential roles of *bHLHs* in *B. napus* roots

Given their crucial roles in diverse plant biological processes, research into the functional characteristics of plant bHLHs have garnered much attention in the past three decades. The majority of the known functions of *bHLHs* center on their roles in controlling the transcriptional networks of a number of biological processes, including metabolism and development. However, an increasing number of studies have demonstrated that they play important roles in plant root development as well. To date, the functions of *bHLHs* in plant roots mainly focused on their role in iron-uptake, salt and drought stress response, hormone response (e.g., responding to ABA, JA, BR, and IAA), and the regulation of the size of the vascular system in roots and meristems (Table 1).

As mentioned above, the bHLHs are widely expressed in root tissues in *B. napus*. Accordingly, many members in nine subfamilies of this gene family (S5, S7, S15, S23–S28) have been identified to regulate many root processes (Table 1). For example, *AtbHLH156/LHW* from subfamily S23 plays an essential role in establishing vascular cells and the size of the vascular initial population in the root meristem [40]. Additionally, *AtbHLH024/SPT* in S24 regulates root growth by controlling the size of the root meristem [44]. Homologs of *AtbHLH002/GL3* (S5), *AtLRLs* (S26), and *AtRSLs* (S28) are also involved in root hair development [35, 45-54]. Moreover, *AtbHLH92* in S7 and *AtbHLH129* in S27 regulate root elongation [55, 56], while *AtbHLH74* of S25 regulates seedling root growth [57]. In this study, we found that the *BnabHLH* orthologs of those functionally characterized *bHLHs* were highly expressed in *B. napus* roots (Fig. 5), indicating their potential importance in root development.

In this study, *bHLHs* were also shown to be involved in hormone signaling pathways and environmental stress in plant roots. To date, many members in 11 subfamilies (S1-S4, S7, S12, S15, S16, S19, S26, and S27) are involved in plant responses to iron, salt, and drought stresses, and are involved in ABA and JA pathways (Table 1). For example, members in subfamilies S1, S4, and S12 affect iron uptake [58-60]. Studies have also found that *AtbHLH17/AIB* (S2), *AtbHLH92* (S7), *AtbHLH112* (S15), and *bHLH122* (S27) respond to salt stress [55, 61-63], while *AtbHLH68* (S15) and *ZmPTF1* (S26) respond to drought stress [47, 64]. Meanwhile, many *MYC2* homologous genes in the S2 subfamily (e.g., *Catharanthus roseus CrMYC2* [65] and *Salvia miltiorrhiza SmMYC2* [66]) are known to be involved in cellular responses to ABA and JA [20-22]. Similarly, members in subfamilies S15 and S26 have been demonstrated to respond to ABA [47, 64], while members of the S19 subfamily negatively regulate the BR signaling pathway [67]. Additionally, genes in the S4 subfamily participate in auxin-conjugate metabolism [41, 68] (Table 1). Accordingly, our RNA-seq data shows that many *BnabHLHs* (221, 36.7%) respond to more than one hormone treatment in *B. napus* roots (Additional file 7: Table S7). Our qRT-PCR analysis further confirmed that two *BnabHLHs* (*BnabHLH126* and *BnabHLH453*) in S2 and three *BnabHLHs* (*BnabHLH033*, *BnabHLH041*, and *BnabHLH269*) in S4 were induced by hormone treatments in *B. napus* roots (Fig. 6). Together, these results suggest that bHLHs have an important role in hormone signaling in plant roots.

Taken together, expression profile analyses along with previous gene function research indicates that *BnabHLHs* may be integrally involved in *B. napus* root biological processes including root growth and hormone signaling. Our study provided a valuable foundation for further *bHLH* gene function research.

Conclusion

In this study, 602 *BnabHLHs* were identified from *B. napus* genome and were classified into 35 subfamilies. Eight conserved intron insertion patterns were observed in the bHLH domains of *BnabHLHs*. Members of the same subfamily generally have conserved gene structure and protein motif composition. Allopolyploidization between *B. rapa* and *B. oleracea* is the major driving force for large gene expansion

of bHLH genes in *B. napus* genome. The *BnabHLHs* have diverse expression profiles in 50 *B. napus* tissues at different developmental stages, but members in the same subfamily generally have a similar expression pattern. Moreover, many *BnabHLHs* have relative high expression levels in *B. napus* roots. Many *BnabHLHs* (about 37%) were hormone-inducible in *B. napus* roots by RNA-Seq analyses. Among them, the hormone-induced expression characteristics of five genes, *BnabHLH033*, *BnabHLH041*, *BnabHLH269*, *BnabHLH453*, and *BnabHLH126* were confirmed by qRT-PCR assay. Overall, this study provides important clues about the potential functions of the *BnabHLHs* that will be useful for gene function research in the future.

Methods

Sequence retrieval

The sequences of 167 *Arabidopsis* bHLHs proteins (AtbHLHs) were retrieved from the TAIR (<http://www.arabidopsis.org/>). To identify bHLH genes in the *B. napus* genome, we referred to the previously described method of Guo et al. [29] with minor modifications. In brief, we performed a repeated BLASTP search (e values of < 1.0) against the proteome of *B. napus* (Darmor-*bzh* ecotype) in GENOSCOPE (<http://www.genoscope.cns.fr/brassicanapus/>) [30], using at least one representative protein sequence of the bHLH domain for each bHLH subfamily [7] as queries. Each matching sequence was then used to search the *B. napus* genome database, until no new sequences were found. The redundant sequences were discarded according to the loci in the genome. To ensure the integrity of the bHLH gene data in *B. napus*, we also searched another sequenced *B. napus* cultivar genome (ZS11 ecotype) in the NCBI database (https://www.ncbi.nlm.nih.gov/genome/annotationeuk/Brassica_napus/101/) [31]. We then confirmed the putative non-redundant sequences to ensure that the candidates contained the bHLH domain using ExPASy (<http://expasy.org/prosite/>) [69] and MEGA 5.0 [70] software. The sequence information of the candidate genes in these two ecotypes was manually compared and corrected. Finally, all candidates were named according to their chromosome locus. Similarly, the candidate bHLH genes in the *B. oleracea* genome (v1.0) were identified by the same method in the BRAD database (<http://brassicadb.org/brad/>).

Phylogenetic tree construction

Multiple sequence alignment of the bHLH domains of 167 *Arabidopsis* (AtbHLHs) and 602 *B. napus* (BnabHLHs) bHLH proteins was carried out by MAFFT online software under default parameters (<http://mafft.cbrc.jp/alignment/server/>) [71]. Based on the result of multiple sequence alignment, a NJ tree was constructed using MEGA 5.0 [70] with a bootstrap test (1000 replicates) based on p-distance model and pairwise deletion for gap treatment. Tree file was visualized using FigTree v1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Chromosome localization and gene synteny analysis

The information of chromosome length and bHLH genes locations were acquired from the BRAD (<http://brassicadb.org/brad/index.php>) and Genoscope databases (<http://jacob.cea.fr/drf/ifrancoisjacob/Pages/Departements/Genoscope.aspx>) [30,72], respectively. The chromosome map of candidate *BnabHLHs* was drawn by Mapchart software. Similarly, the chromosome locations of *bHLHs* in *Arabidopsis*, *B. rapa*, and *Brassica oleracea* genomes were analyzed by the same method. Gene synteny analysis of *bHLHs* in *Arabidopsis*, *B. napus*, *B. rapa*, and *B. oleracea* genomes was performed by CoGe online software (<https://genomevolution.org/coge/>) [39]. Accordingly, gene replication events for all candidate *bHLH* genes were analyzed based on the syntenic blocks for intra- and inter-genomic comparisons. Tandem duplication genes were identified according to their physical locations within individual chromosomes and <1 intervening gene.

Intron/exon structure analysis

The corresponding CDS and DNA sequences of candidate bHLH genes in *B. napus* were obtained from Genoscope database as well. The intron distribution, position and phase of candidate *B. napus* bHLH genes, including exon and intron numbers were analyzed by the Gene Structure Display Server (GSDS) online tool with default parameters (<http://gsds.cbi.pku.edu.cn/>) [38]. We manually located the intron insertion sites in the corresponding protein sequences as well.

Identification of conserved motifs

To identify the conserved protein motifs outside the bHLH domain, the protein sequences of candidate *BnabHLHs* were analyzed using MEME Version 5.1.1 software (<http://meme-suite.org/tools/meme>) [73]. The parameters settings are: number of motifs to find, 30; minimum width of motifs, 6; and maximum width of motifs, 250.

Expression analysis of *BnabHLHs*

The *B. napus* expression datasets were downloaded from the BioProject (NCBI database: PRJNA358784). The data were obtained from various tissues at different *B. napus* developmental stages and under the induction of five different hormones (IAA, ABA, 6-BA, ACC, and GA₃). The expression profiles of *BnabHLHs* were analyzed using the MeV v4.9 software (<https://sourceforge.net/projects/mev-tm4/files/>) [74] with the HCL method, and the heatmaps were drawn using the R package; The tissue high expression genes identified by z-score above 3 in at least one of the organization samples [74]. All genes with FPKM < 1 were

excluded from the heatmap, as they may be pseudogenes or may be expressed only under specific stresses or treatments.

For qRT-PCR analysis, seeds of *B. napus* variety ZS11 were obtained from the College of Agriculture and Biotechnology, Southwest University, and germinated on petri dishes. At the five-leaf stage, seedlings were treated in Hoagland liquid medium containing one of five phytohormones (50 μ M ABA, 120 μ M GA, 75 μ M 6-BA, 60 μ M ACC, and 10 μ M IAA). The seedlings were then grown in an artificial climate chamber at 25 °C with a 16/8 h photoperiod (day/night). The root tissues were then harvested at 0, 1, 3, 6, 12, and 24 h after the treatments and immediately frozen in liquid nitrogen and stored at -80 °C for RNA isolation. Extractions of total RNA and subsequent cDNA synthesis were performed as described by Guo et al. [29]. *B. napus Actin7 (BnActin7)* (GenBank accession no. AF024716) was used as the internal control. The SYBR-Green PrimeScript RT-PCR Kit (Takara, Dalian, China) was used for qRT-PCR amplification in a CFX Connect™ real-time PCR system (Bio-Rad, Chongqing, China). The reaction conditions for real-time PCR were as follows: initial denaturation at 95 °C for 3 min, followed by 40 cycles of denaturation at 95 °C for 10 s and annealing at 58 °C for 20 s. The relative expression levels of candidate genes were determined using the $2^{-\Delta\Delta C_t}$ method [75]. Each treatment consists of three technical replicates. Expression levels were calculated as the mean signal intensity across the three replicates. The primers used in this analysis are listed in Additional file 9: Table S9.

Availability of data and materials

The RNA-Seq datasets used in the current study are available in the Sequence Read Archive (SRA) at NCBI (SRA accession: PRJNA574049 and PRJNA608211) repository.

Abbreviations

bHLH: Basic helix-loop-helix

IAA: Auxin

GA₃: Gibberellin

6-BA: Cytokinin

ABA: Abscisic acid

ACC: Ethylene

qRT-PCR: Quantitative real time polymerase chain reaction

FPKM: Fragments per kilobase of transcript per million fragments mapped

GSDS: Gene Structure Display Server

NJ: Neighbour-Joining

Declarations

Acknowledgments

Not applicable.

Funding

This work was supported by the National Natural Science Foundation of China (31671727) and the Fundamental Research Funds for the Central Universities of China (XDJK2020B031).

Availability of data and material

The datasets supporting the conclusions of this article are included within the article and its additional files.

Author Contributions

H.D. contributed to the conception of the study; Y.Z.K and W.Y.W have equal contributions to this study; Y.Z.K., W.Y.W, and H.D drafted and revised the manuscript. Y.Z.K., W.Y.W., H.D. Z.H.J., and M.M.L contributed to the data analysis. Y.Z.K., W.Y.W., H.D., C.P, M.M.W., P.F.L., Y.J., and J.N.L conceived of and designed the experiments. All authors reviewed and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare they have no competing interests.

Reference

1. Amoutzias GD, Veron AS, Weiner J, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG *et al*: One billion years of bZIP transcription factor evolution: conservation and change in dimerization and

- DNA-binding site specificity. *Molecular Biology & Evolution* 2007, 24(3):827-835.
2. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J *et al*: Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 2000, 290(5499):2105-2110.
 3. Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC: The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Mol Biol Evol* 2003, 20(5):735-747.
 4. Toledo-Ortiz G, Huq E, Quail PH: The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell* 2003, 15(8):1749-1770.
 5. Murre C, McCaw PS, Baltimore D: A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 1989, 56(5):777-783.
 6. Pires N, Dolan L: Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol Biol Evol* 2010, 27(4):862-874.
 7. Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martinez-Garcia JF, Bilbao-Castro JR, Robertson DL: Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. *Plant Physiol* 2010, 153(3):1398-1412.
 8. Zhang T, Lv W, Zhang H, Ma L, Li P, Ge L *et al*: Genome-wide analysis of the basic Helix-Loop-Helix (bHLH) transcription factor family in maize. *BMC Plant Biology* 2018, 18(1):235.
 9. Sun H, Fan HJ, Ling HQ: Genome-wide identification and characterization of the bHLH gene family in tomato. *BMC Genomics* 2015, 16(1):9.
 10. Song XM, Huang ZN, Duan WK, Ren J, Liu TK, Li Y *et al*: Genome-wide analysis of the bHLH transcription factor family in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Mol Genet Genomics* 2014, 289(1):77-91.
 11. Pillitteri LJ, Sloan DB, Bogenschutz NL, Torii KU: Termination of asymmetric cell division and differentiation of stomata. *Nature* 2007, 445(7127):501-505.
 12. Waseem M, Li N, Su D, Chen J, Li Z: Overexpression of a basic helix-loop-helix transcription factor gene, SlbHLH22, promotes early flowering and accelerates fruit ripening in tomato (*Solanum lycopersicum* L.). *Planta* 2019, 250(1):173-185.
 13. Girin T, Paicu T, Stephenson P, Fuentes S, Korner E, O'Brien M *et al*: INDEHISCENT and SPATULA interact to specify carpel and valve margin tissue and thus promote seed dispersal in Arabidopsis. *Plant Cell* 2011, 23(10):3641-3653.
 14. Tani E, Tsaballa A, Stedel C, Kalloniati C, Papaefthimiou D, Polidoros A *et al*: The study of a SPATULA-like bHLH transcription factor expressed during peach (*Prunus persica*) fruit development. *Plant Physiol Biochem* 2011, 49(6):654-663.
 15. Xiang L, Jian D, Zhang F, Yang C, Bai G, Lan X *et al*: The cold-induced bHLH transcription factor AabHLH112 promotes artemisinin biosynthesis in *Artemisia annua*. *J Exp Bot* 2019:piv: erz220.
 16. Xu YH, Liao YC, Lv FF, Zhang Z, Sun PW, Gao ZH *et al*: Transcription Factor AsMYC2 Controls the Jasmonate-Responsive Expression of ASS1 Regulating Sesquiterpene Biosynthesis in *Aquilaria*

- sinensis (Lour.) Gilg. *Plant Cell Physiol* 2017, 58(11):1924-1933.
17. Lau OS, Song Z, Zhou Z, Davies KA, Chang J, Yang X *et al*: Direct Control of SPEECHLESS by PIF4 in the High-Temperature Response of Stomatal Development. *Curr Biol* 2018, 28(8):1273-1280 e1273.
 18. Liu Y, Lin-Wang K, Espley RV, Wang L, Li Y, Liu Z *et al*: StMYB44 negatively regulates anthocyanin biosynthesis at high temperatures in tuber flesh of potato. *J Exp Bot* 2019, 70(15):3809-3824.
 19. Xie XB, Li S, Zhang RF, Zhao J, Chen YC, Zhao Q *et al*: The bHLH transcription factor MdbHLH3 promotes anthocyanin accumulation and fruit colouration in response to low temperature in apples. *Plant Cell Environ* 2012, 35(11):1884-1897.
 20. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K: Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* 2003, 15(1):63-78.
 21. Lorenzo O, Chico JM, Sanchez-Serrano JJ, Solano R: JASMONATE-INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in Arabidopsis. *Plant Cell* 2004, 16(7):1938-1950.
 22. Yadav V, Mallappa C, Gangappa SN, Bhatia S, Chattopadhyay S: A basic helix-loop-helix transcription factor in Arabidopsis, MYC2, acts as a repressor of blue light-mediated photomorphogenic growth. *Plant Cell* 2005, 17(7):1953-1966.
 23. Zhou M, Memelink J: Jasmonate-responsive transcription factors regulating plant secondary metabolism. *Biotechnol Adv* 2016, 34(4):441-449.
 24. Ogo Y, Itai RN, Nakanishi H, Kobayashi T, Takahashi M, Mori S *et al*: The rice bHLH protein OsIRO2 is an essential regulator of the genes involved in Fe uptake under Fe-deficient conditions. *Plant J* 2007, 51(3):366-377.
 25. Bauer P, Ling HQ, Guerinot ML: FIT, the FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR in Arabidopsis. *Plant Physiol Biochem* 2007, 45(5):260-261.
 26. Bernhardt C, Zhao M, Gonzalez A, Lloyd A, Schiefelbein J: The bHLH genes GL3 and EGL3 participate in an intercellular regulatory circuit that controls cell patterning in the Arabidopsis root epidermis. *Development* 2005, 132(2):291-298.
 27. Bernhardt C, Lee MM, Gonzalez A, Zhang F, Lloyd A, Schiefelbein J: The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root. *Development* 2003, 130(26):6431-6439.
 28. Payne C, Thomas, Zhang F, Lloyd AM: GL3 encodes a bHLH protein that regulates trichome development in arabidopsis through interaction with GL1 and TTG1. *Genetics* 2000, 156(3):1349-1362.
 29. Guo P, Wen J, Yang J, Ke Y, Wang M, Liu M *et al*: Genome-wide survey and expression analyses of the GRAS gene family in Brassica napus reveals their roles in root development and stress response. *Planta* 2019:1-22.
 30. Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X *et al*: Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science* 2014, 345(6199):950-953.

31. Sun F, Fan G, Hu Q, Zhou Y, Guan M, Tong C *et al*: The high-quality genome of *Brassica napus* cultivar 'ZS11' reveals the introgression history in semi-winter morphotype. *Plant J* 2017, 92(3):452-468.
32. Chigri F, Flosdorff S, Pilz S, Kolle E, Dolze E, Gietl C *et al*: The Arabidopsis calmodulin-like proteins AtCML30 and AtCML3 are targeted to mitochondria and peroxisomes, respectively. *Plant Mol Biol* 2012, 78(3):211-222.
33. Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z *et al*: Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. *Plant Physiol* 2006, 141(4):1167-1184.
34. Massari ME, Murre C: Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Molecular Cellular Biology* 2000, 20(2):429-440.
35. Feller A, Hernandez JM, Grotewold E: An ACT-like domain participates in the dimerization of several plant basic-helix-loop-helix transcription factors. *J Biol Chem* 2006, 281(39):28964-28974.
36. Zhang F, Yao J, Ke J, Zhang L, Lam VQ, Xin XF *et al*: Structural basis of JAZ repression of MYC transcription factors in jasmonate signalling. *Nature* 2015, 525(7568):269-273.
37. Rameneni JJ, Dhandapani V, Paul P, Im S, Oh MH, Choi SR *et al*: Genome-wide identification, characterization, and comparative phylogeny analysis of MADS-box transcription factors in *Brassica rapa*. *Genes Genom* 2014, 36(4):509-525.
38. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G: GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 2015, 31(8):1296-1297.
39. Lyons E, Freeling M: How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J* 2008, 53(4):661-673.
40. Ohashi-Ito K, Bergmann DC: Regulation of the Arabidopsis root vascular initial population by LONESOME HIGHWAY. *Development* 2007, 134(16):2959-2968.
41. Rampey RA, Woodward AW, Hobbs BN, Tierney MP, Lahner B, Salt DE *et al*: An Arabidopsis basic helix-loop-helix leucine zipper protein modulates metal homeostasis and auxin conjugate responsiveness. *Genetics* 2006, 174(4):1841-1857.
42. Du H, Liang Z, Zhao S, Nan MG, Tran LS, Lu K *et al*: The Evolutionary History of R2R3-MYB Proteins Across 50 Eukaryotes: New Insights Into Subfamily Classification and Expansion. *Sci Rep* 2015, 5:11037.
43. Niu X, Guan Y, Chen S, Li H: Genome-wide analysis of basic helix-loop-helix (bHLH) transcription factors in *Brachypodium distachyon*. *BMC Genomics* 2017, 18:619.
44. Makkena S, Lamb RS: The bHLH transcription factor SPATULA regulates root growth by controlling the size of the root meristem. *BMC Plant Biol* 2013:13:11.
45. Wang CX, Qi CY, Luo JH, Liu L, He Y, Chen LQ: Characterization of LRL5 as a key regulator of root hair growth in maize. *Plant J* 2019, 98:71-82.
46. Breuninger H, Thamm A, Streubel S, Sakayama H, Nishiyama T, Dolan L: Diversification of a Transcription Factor Family Led to the Evolution of Antagonistically Acting Genetic Regulators of

Root Hair Growth. *Curr Biol* 2016, 26:1622-1628.

47. Li Z, Liu C, Zhang Y, Wang B, Ran Q, Zhang J: The bHLH family member ZmPTF1 regulates drought tolerance in maize by promoting root development and ABA synthesis. *J Exp Bot* 2019:pii: erz307.
48. Menand B, Yi K, Jouannic S, Hoffmann L, Ryan E, Linstead P *et al*: An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* 2007, 316(5830):1477-1480.
49. Zhang F, Gonzalez A, Zhao M, Payne CT, Lloyd A: A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. *Development* 2003, 130(20):4859-4869.
50. Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF: Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J* 2004, 40(1):22-34.
51. Honkanen S, Thamm A, Arteaga-Vazquez MA, Dolan L: Negative regulation of conserved RSL class I bHLH transcription factors evolved independently among land plants. *Elife* 2018, 7:pii: e38529.
52. Zhang C, Simpson RJ, Kim CM, Warthmann N, Delhaize E, Dolan L *et al*: Do longer root hairs improve phosphorus uptake? Testing the hypothesis with transgenic *Brachypodium distachyon* lines overexpressing endogenous RSL genes. *New Phytol* 2018, 217:1654-1666.
53. Tam TH, Catarino B, Dolan L: Conserved regulatory mechanism controls the development of cells with rooting functions in land plants. *Proc Natl Acad Sci U S A* 2015, 112:E3959-3968.
54. Pires ND, Yi K, Breuninger H, Catarino B, Menand B, Dolan L: Recruitment and remodeling of an ancient gene regulatory network during land plant evolution. *Proc Natl Acad Sci U S A* 2013, 110(23):9571-9576.
55. Jiang Y, Yang B, Deyholos MK: Functional characterization of the Arabidopsis bHLH92 transcription factor in abiotic stress. *Mol Genet Genomics* 2009, 282(5):503-516.
56. Tian H, Guo H, Dai X, Cheng Y, Zheng K, Wang X *et al*: An ABA down-regulated bHLH transcription repressor gene, bHLH129 regulates root elongation and ABA response when overexpressed in Arabidopsis. *Sci Rep* 2015:5:17587.
57. Bao M, Bian H, Zha Y, Li F, Sun Y, Bai B *et al*: miR396a-Mediated basic helix-loop-helix transcription factor bHLH74 repression acts as a regulator for root growth in Arabidopsis seedlings. *Plant Cell Physiol* 2014, 55:1343-1553.
58. Gratz R, Manishankar P, Ivanov R, Koster P, Mohr I, Trofimov K *et al*: CIPK11-Dependent Phosphorylation Modulates FIT Activity to Promote Arabidopsis Iron Acquisition in Response to Calcium Signaling. *Dev Cell* 2019, 48(5).
59. Tanabe N, Noshi M, Mori D, Nozawa K, Tamoi M, Shigeoka S: The basic helix-loop-helix transcription factor, bHLH11 functions in the iron-uptake system in Arabidopsis thaliana. *J Plant Res* 2019, 132:95-105.
60. Kurt F, Filiz E: Genome-wide and comparative analysis of bHLH38, bHLH39, bHLH100 and bHLH101 genes in Arabidopsis, tomato, rice, soybean and maize: insights into iron (Fe) homeostasis. *Biometals* 2018, 31(1572-8773 (Electronic)):489-504.

61. Babitha KC, Ramu SV, Pruthvi V, Mahesh P, Nataraja KN, Udayakumar M: Co-expression of AtbHLH17 and AtWRKY28 confers resistance to abiotic stress in Arabidopsis. *Transgenic Res* 2013, 22(2):327-341.
62. Chen HC, Hsieh-Feng V, Liao PC, Cheng WH, Liu LY, Yang YW *et al*: The function of OsbHLH068 is partially redundant with its homolog, AtbHLH112, in the regulation of the salt stress response but has opposite functions to control flowering in Arabidopsis. *Plant Mol Biol* 2017, 94(4-5):531-548.
63. Krishnamurthy P, Vishal B, Khoo K, Rajappa S, Loh CS, Kumar PP: Expression of AoNHX1 increases salt tolerance of rice and Arabidopsis, and bHLH transcription factors regulate AtNHX1 and AtNHX6 in Arabidopsis. *Plant Cell Rep* 2019:1-17.
64. Le Hir R, Castelain M, Chakraborti D, Moritz T, Dinant S, Bellini C: AtbHLH68 transcription factor contributes to the regulation of ABA homeostasis and drought stress tolerance in Arabidopsis thaliana. *Physiol Plant* 2017, 160:312-327.
65. Zhang H, Hedhili S, Montiel G, Zhang Y, Chatel G, Pre M *et al*: The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in Catharanthus roseus. *Plant J* 2011, 67(1):61-71.
66. Du T, Niu J, Su J, Li S, Guo X, Li L *et al*: SmbHLH37 Functions Antagonistically With SmMYC2 in Regulating Jasmonate-Mediated Biosynthesis of Phenolic Acids in Salvia miltiorrhiza. *Front Plant Sci* 2018, 9:1720.
67. Kim Y, Song JH, Park SU, Jeong YS, Kim SH: Brassinosteroid-Induced Transcriptional Repression and Dephosphorylation-Dependent Protein Degradation Negatively Regulate BIN2-Interacting AIF2 (a BR Signaling-Negative Regulator) bHLH Transcription Factor. *Plant Cell Physiol* 2017, 58:227-239.
68. Samira R, Li B, Kliebenstein D, Li C, Davis E, Gillikin JW *et al*: The bHLH transcription factor ILR3 modulates multiple stress responses in Arabidopsis. *Plant Mol Biol* 2018, 97:297-309.
69. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M *et al*: The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res* 2001, 29(1):37-40.
70. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011, 28(10):2731-2739.
71. Katoh K, Standley DMB: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology Evolution* 2013, 30(4):772-780.
72. Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B *et al*: BRAD, the genetics and genomics database for Brassica plants. *BMC Plant Biol* 2011, 11(1):136.
73. Bailey TL, Williams N, Misleh C, Li WW: MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* 2006, 34(Web Server issue):W369.
74. Yi F, Gu W, Chen J, Song N, Gao X, Zhang X *et al*: High Temporal-Resolution Transcriptome Landscape of Early Maize Seed Development. *Plant Cell* 2019, 31(5):974-992.

75. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001, 25:402-408.
76. Gupta N, Prasad VB, Chattopadhyay S: LeMYC2 acts as a negative regulator of blue light mediated photomorphogenic growth, and promotes the growth of adult tomato plants. *BMC Plant Biol* 2014, 14:38.
77. Penuelas M, Monte I, Schweizer F, Vallat A, Reymond P, Garcia-Casado G *et al*: Jasmonate-Related MYC Transcription Factors Are Functionally Conserved in *Marchantia polymorpha*. *Plant Cell* 2019, 31(10):2491-2509.
78. Chiasson DM, Loughlin PC, Mazurkiewicz D, Mohammadidehcheshmeh M, Fedorova EE, Okamoto M *et al*: Soybean SAT1 (Symbiotic Ammonium Transporter 1) encodes a bHLH transcription factor involved in nodule growth and NH₄⁺ transport. *Proc Natl Acad Sci U S A* 2014, 111(13):4814-4819.
79. Zhang B, Chopra D, Schrader A, Hulskamp M: Evolutionary comparison of competitive protein-complex formation of MYB, bHLH, and WDR proteins in plants. *J Exp Bot* 2019, 70:3197-3209.
80. Ramamurthy RK, Waters BM: Mapping and Characterization of the fefe Gene That Controls Iron Uptake in Melon (*Cucumis melo* L.). *Front Plant Sci* 2017;8:1003.
81. Li L, Gao W, Peng Q, Zhou B, Kong Q, Ying Y *et al*: Two soybean bHLH factors regulate response to iron deficiency. *J Integr Plant Biol* 2018, 60:608-622.
82. Zheng K, Wang Y, Zhang N, Jia Q, Wang X, Hou C *et al*: Involvement of PACLOBUTRAZOL RESISTANCE6/KIDARI, an Atypical bHLH Transcription Factor, in Auxin Responses in *Arabidopsis*. *Front Plant Sci* 2017;8:1813.
83. Gajewska P, Janiak A, Kwasniewski M, Kedzierski P, Szarejko I: Forward Genetics Approach Reveals a Mutation in bHLH Transcription Factor-Encoding Gene as the Best Candidate for the Root Hairless Phenotype in Barley. *Front Plant Sci* 2018, 9:1229.

Additional File Information

Additional file 1: Table S1. Features of the 602 *bHLH* genes from *Brassica napus* identified in this study.

Additional file 2: Table S2. Identification of the 245 *bHLH* genes from *Brassica oleracea* in this study

Additional file 3: Table S3. Amino acid composition of the bHLH domain across different species.

Additional file 4: Table S4. Conserved motifs identified in different bHLH subfamilies in this study.

Additional file 5: Table S5. Syntenic relationships of *bHLH* genes in *Brassica napus*, *Brassica rape*, and *Brassica oleracea*.

Additional file 6: Table S6. The expression values of *BnabHLHs* in 50 *Brassica napus* tissues across different developmental stages.

Additional file 7: Table S7. The expression values of *BnabHLHs* under different hormone treatments.

Additional file 8: Table S8. Cis-acting element analysis of *BnabHLH* promoter sequences.

Additional file 9: Table S9. List of primers used for the real-time PCR analysis.

Table

Please see the supplementary files section to access the table.

Figures

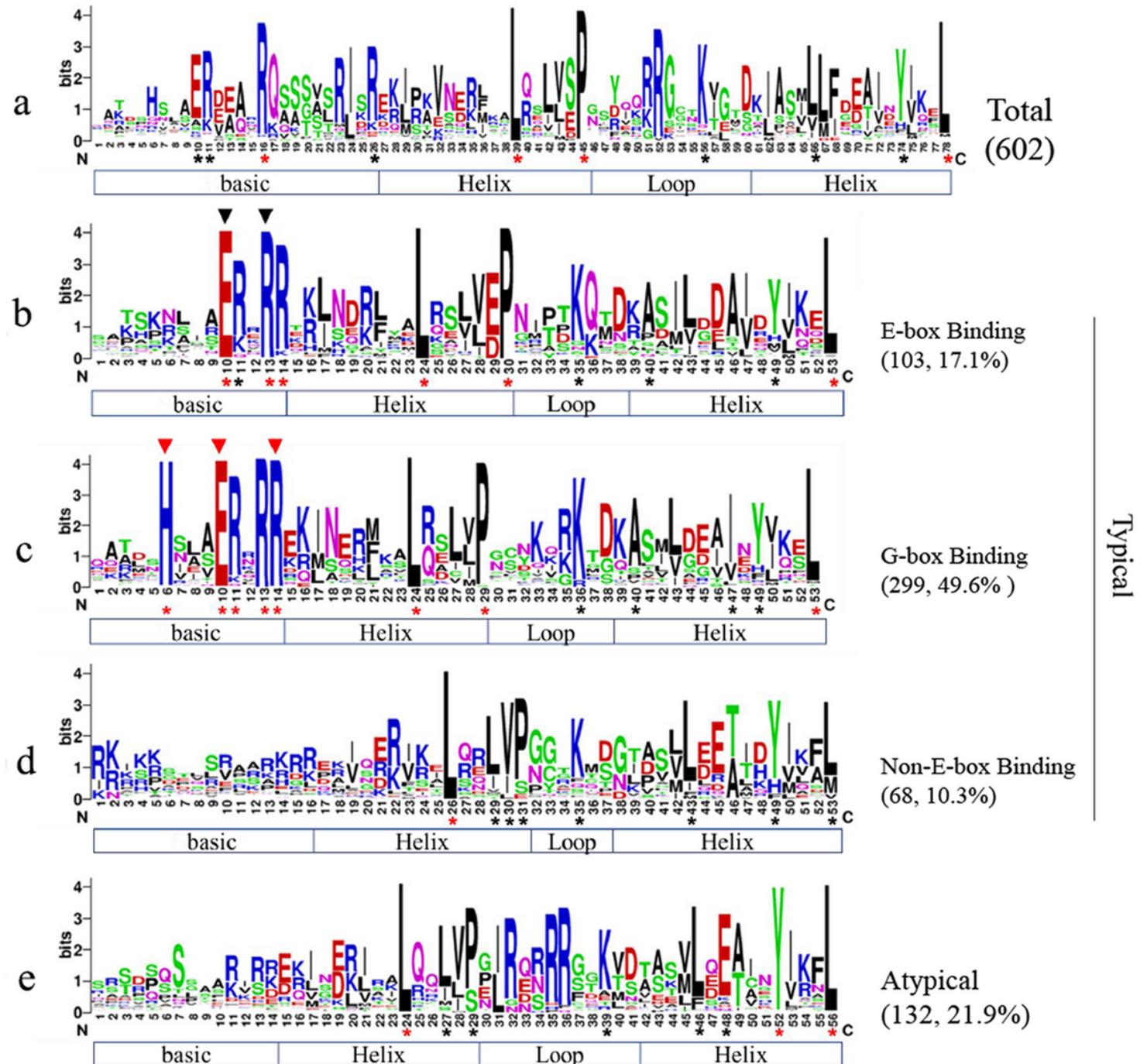


Figure 1

Sequence characteristics of the bHLH domains in different DNA-binding types. Multiple sequence alignments were conducted with the bHLH domains of all candidate proteins and then was conducted with the domains of proteins of a given DNA-binding type. The number in bracket indicates the amount of BnabHLHs in a certain category. Protein secondary structures are illustrated under the sequences. Red asterisks indicate the residues with over 90% similarity; black asterisks indicate residues with over 70% similarity. Black triangles at the top of the sequence indicates the E-box recognition sites; red triangles indicate the G-box recognition sites. The bHLH domains with at least five basic residues but no E-box/G-box binding sites are classified as non-E-box binding genes, otherwise it is a non-DNA binding gene (atypical gene) [6].

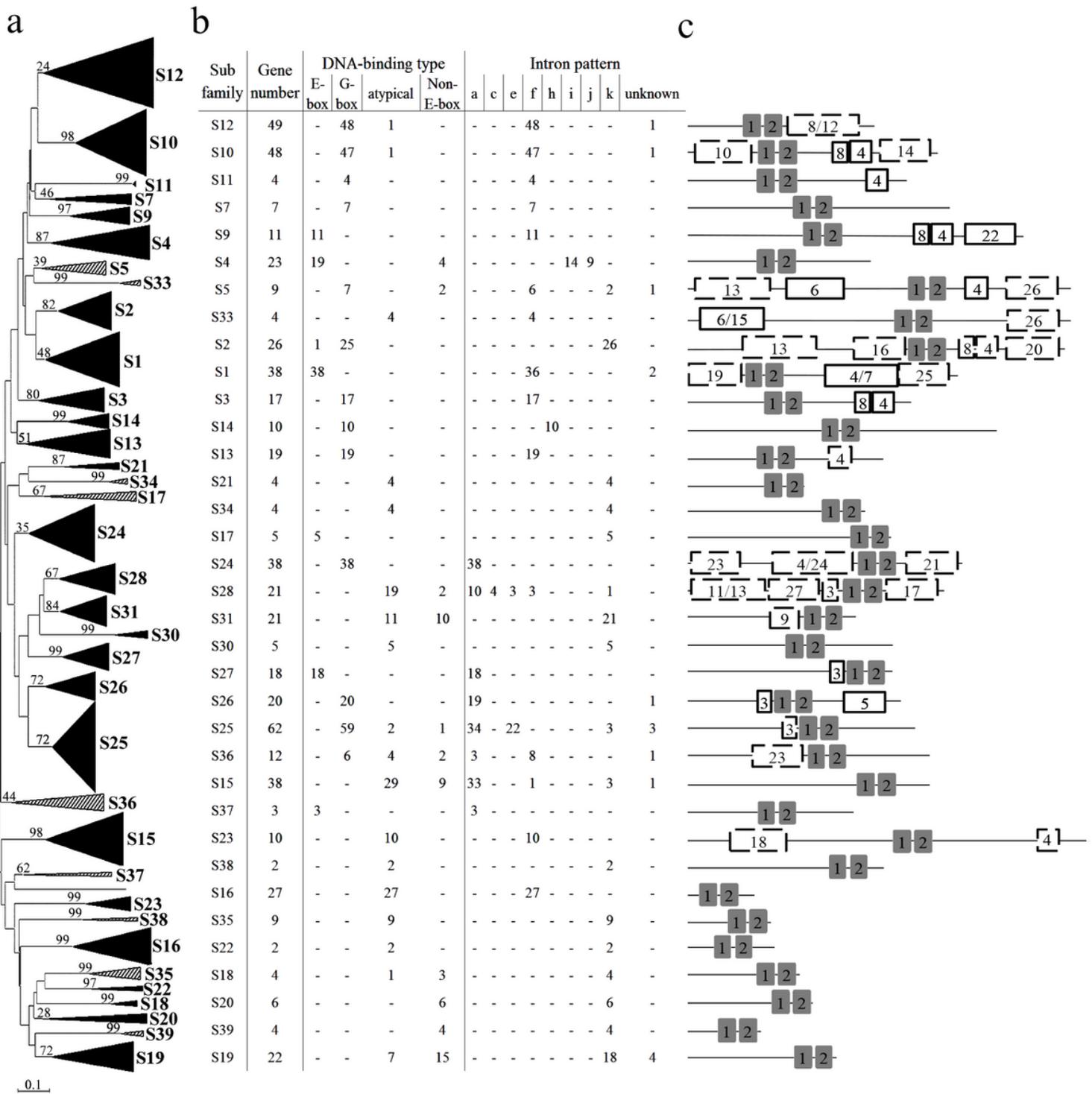


Figure 2

Phylogenetic relationships, DNA-binding types, intron insertion patterns, and architecture of conserved protein motifs in 35 bHLH subfamilies. (a) Phylogenetic relationships of 769 bHLHs from *B. napus* (602) and *Arabidopsis* (167). The phylogenetic tree is generated based on the alignment of bHLH domains of the corresponding bHLH proteins with 1000 bootstrap replicates. The subfamilies marked in grey are newly identified in this study as compared to the results in *Arabidopsis* [7]. (b) DNA-binding types and intron insertion types of BnabHLHs in each subfamily. The illustration of intron insertion patterns is

shown in Fig. 3. (c) Architecture of conserved protein motifs of BnabHLHs in each subfamily by MEME analysis. Blocks with a black background indicate the bHLH domain. Blocks with a solid line represent the motif distributed in all proteins in a certain subfamily, while those with a dotted line represent the motif distributed in a part of members in a given subfamily.

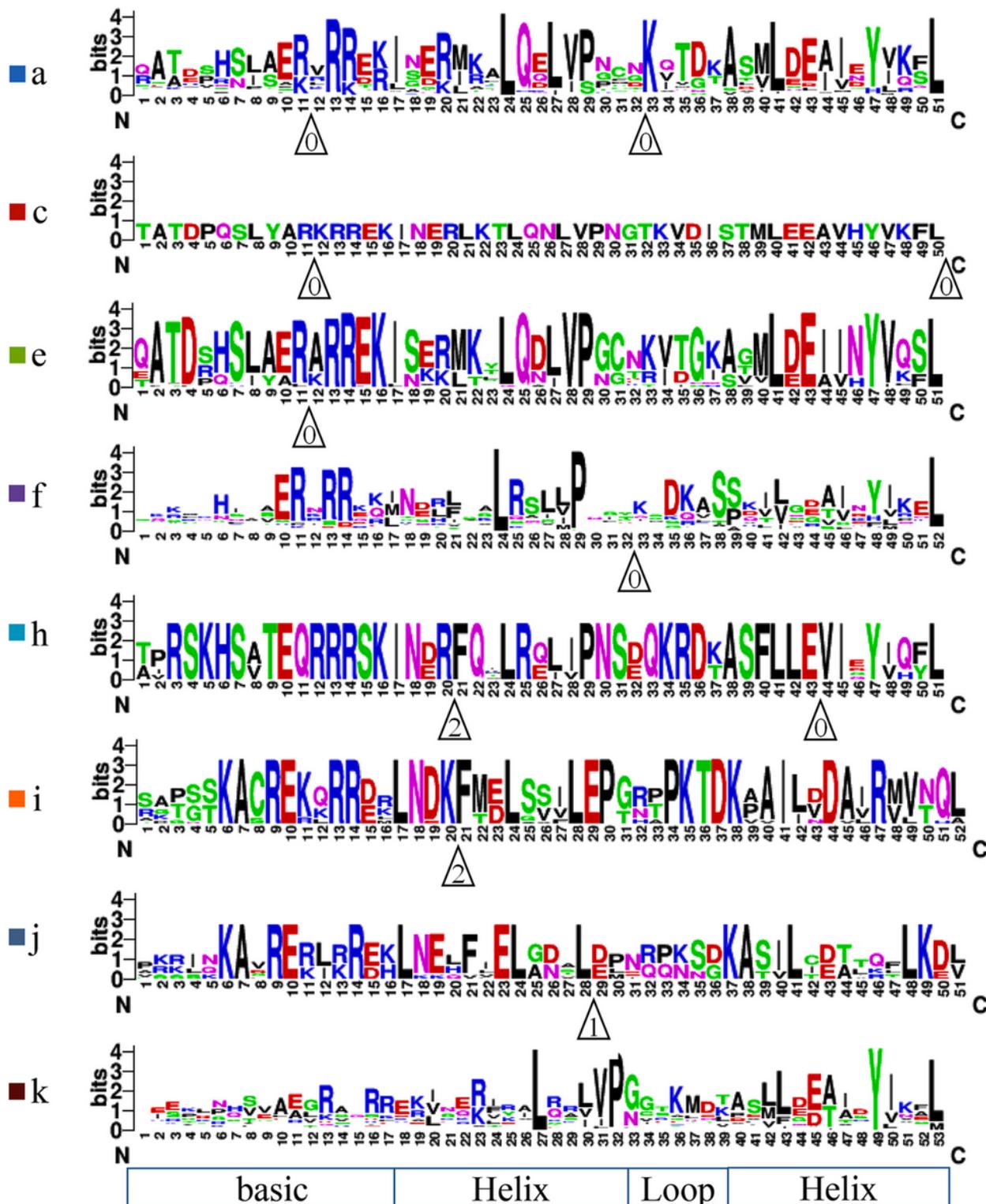


Figure 3

Schematic diagram of intron insertion patterns within the bHLH domains of BnabHLHs. The intron patterns are classified into eight intron types, namely a–k, respectively. Intron insertion sites are indicated by white triangles, and the number within each triangle indicates the splicing phases: 0 refers to phase 0, 1 to phase 1, and 2 to phase 2. A black triangle at the top of a sequence indicates a conserved E-box recognition site. The intron pattern of each subfamily is provided in Fig. 2.

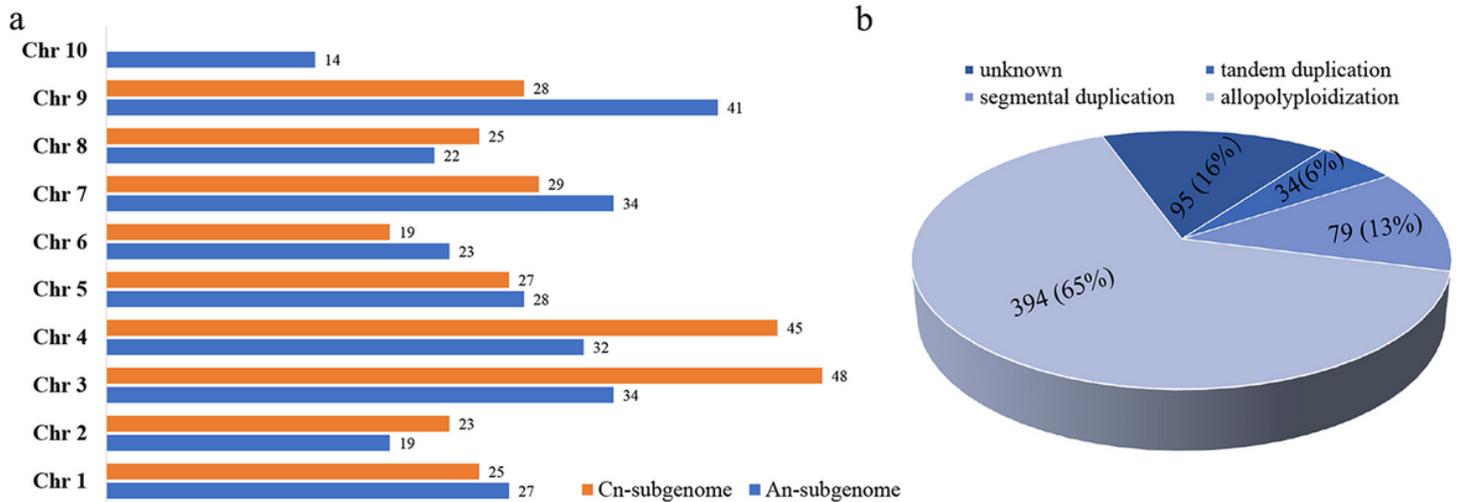


Figure 4

The distributions of candidate BnabHLHs on each chromosome and the expansion mechanism of BnabHLHs. (a) Distribution of BnabHLHs on the 19 *B. napus* chromosomes. The blue boxes indicate the chromosomes in the An-subgenome, while the orange boxes indicate those in the Cn-subgenome. The chromosome number is shown on the left and the number of BnabHLHs on a certain chromosome is listed on the right. (b) Percentage of BnabHLHs derives from the An- or Cn-subgenome.

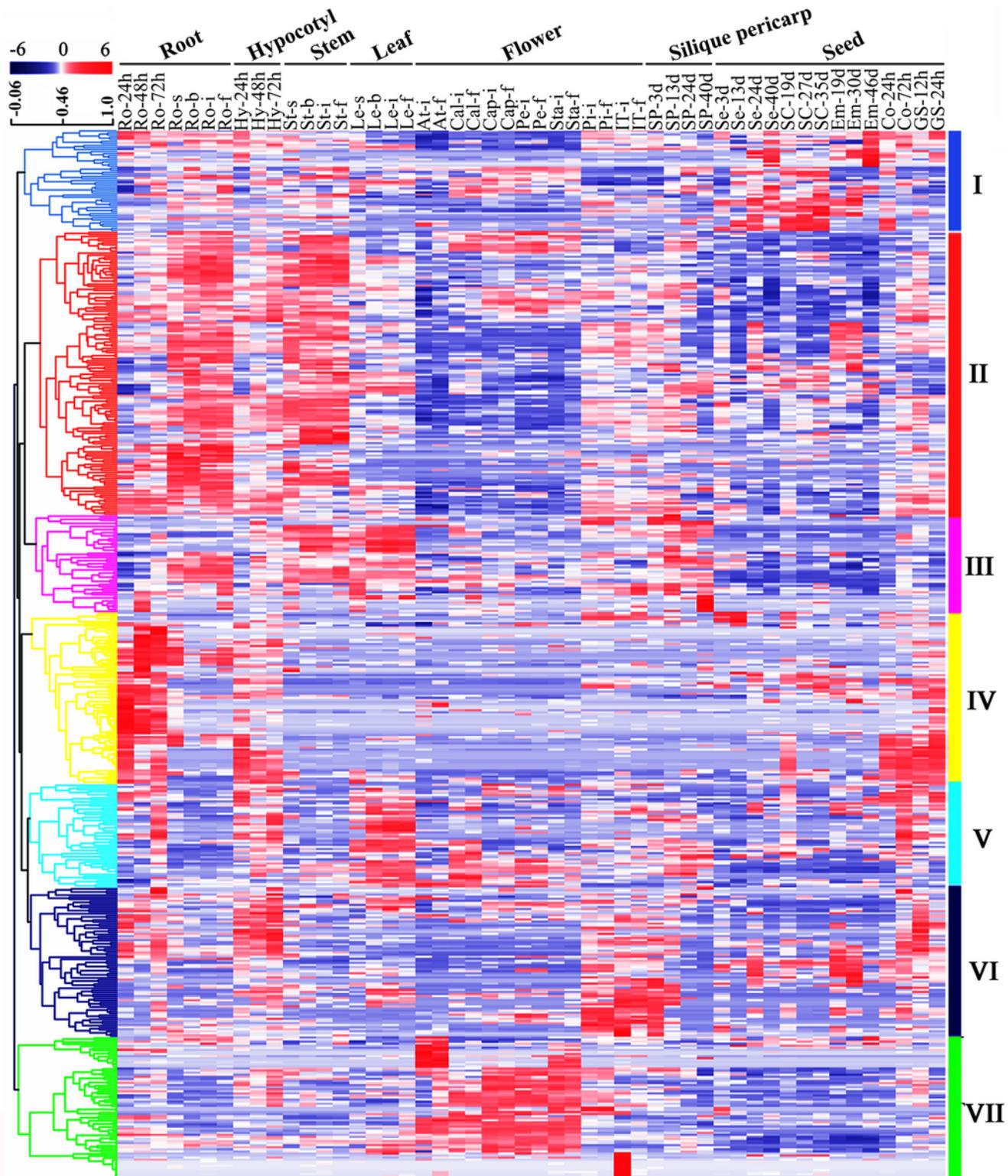


Figure 5

Expression profiles of candidate BnabHLHs in 50 *B. napus* tissues or organs across different developmental stages. Seven major blocks of different expression patterns are illustrated on the right. The tissues that were used for the expression analysis are indicated at the top of each column: GS, germinate seed; Hy, hypocotyl; Ao, anthocaulus; Ro, root; St, stem; Le, leaf; Cal, calyx; Cap, capillament; Pe, petal; Sta, stamen; Pi, pistil; IT, inflorescence tip; SP, silique; Se, seed; SC, seed coat; Em, embryo; Co,

cotyledon. The 'h', 'd', 's', 'b' 'i', 'f' indicate hour, day, seeding, budding, initial flowering, and full-bloom stages, respectively. The color bar represents log₂ (FPKM ≥ 1).

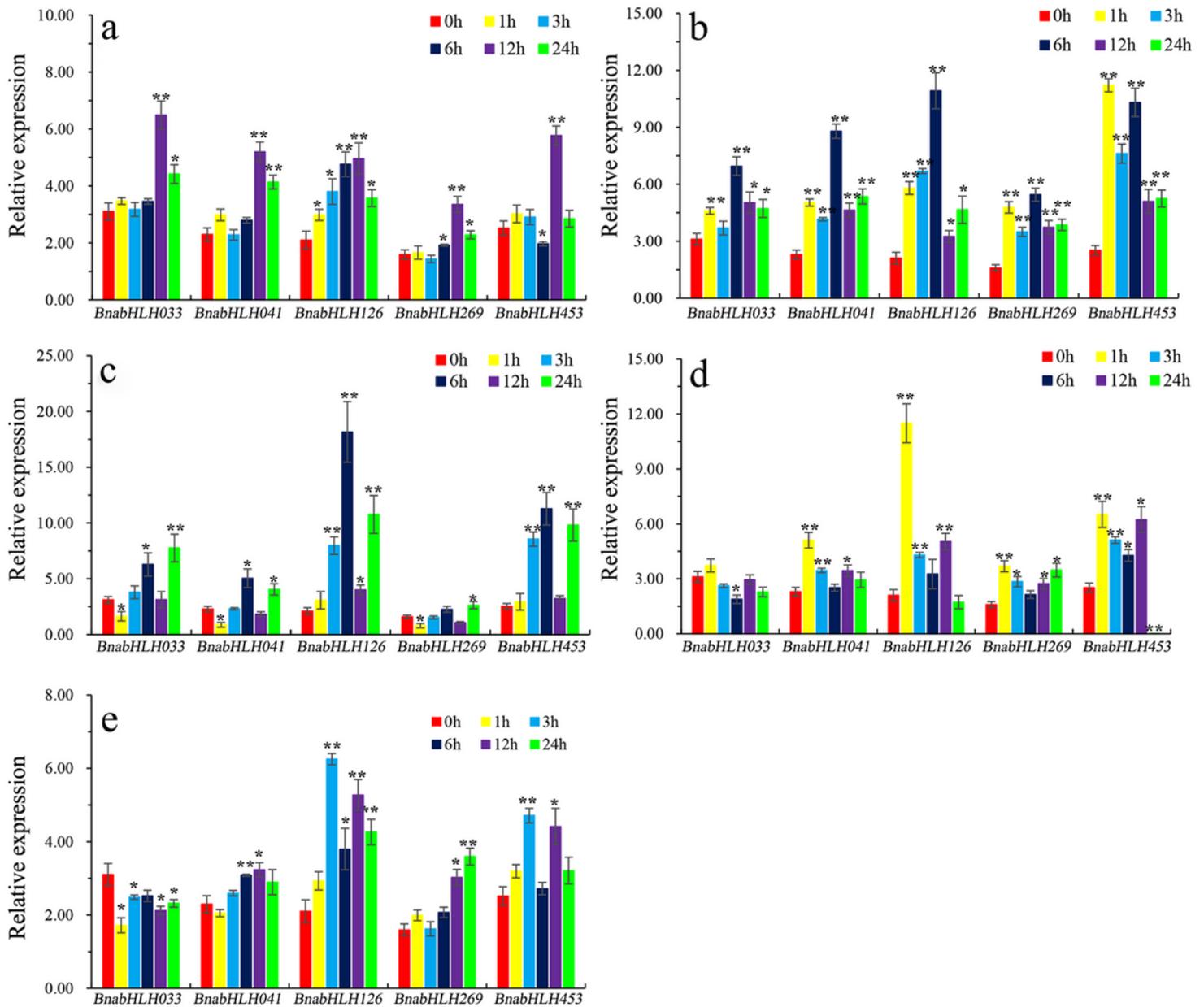


Figure 6

Expression of five BnabHLHs under hormone treatments. The transcript levels were determined in seedling roots by qRT-PCR under five different hormone treatments. (a) indole-3-acetic acid (IAA) treatment; (b) 1-aminocyclopropane-1-carboxylic acid (ACC) treatment; (c) abscisic acid (ABA) treatment; (d) gibberellic acid (GA3) treatment; (e) 6-benzyladenine (6-BA) treatment. Data are the mean ± standard deviation of three independent experiments. Error bars represent the standard errors from three independent replicates. Expression differences in BnabHLHs following hormone treatments was assessed by the Welch's t-test (*P < 0.05; **P < 0.01). Comparison between treatments and control (CK, 0 h) according to Welch's t-test (*P < 0.05; **P < 0.01).

Supplementary Files

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

- [TableS2.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [TableS7.xlsx](#)
- [TableS8.xlsx](#)
- [TableS9.xlsx](#)
- [TableS1.xlsx](#)
- [TableS4.xlsx](#)
- [Table1.xlsx](#)
- [TableS3.xlsx](#)