Genome-wide Survey of bHLH Super Gene Family in Brassica napus and Their Roles in Roots

CURRENT STATUS: UNDER REVIEW

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

Corresponding Author

ORCiD: 0000-0003-4274-5724

DOI:
10.21203/rs.2.16021/v1

SUBJECT AREAS
Plant Physiology and Morphology
Plant Molecular Biology and Genetics

KEYWORDS
Brassica napus; bHLH transcription factor; root; gene expression; BMW.
Abstract

Background: Basic helix-loop-helix (bHLH) gene family is one of the largest transcription factors in plants and are functionally characterized in diverse species. However, less is known about their functions in the economically important allopolyploid oil crop, Brassica napus. Results: We identified 602 potential bHLHs in B. napus genome (BnbHLHs) and categorized them into 36 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 corrected and a total of eight types were identified in the bHLH regions of BnbHLHs. Chromosome distribution and synteny analyses revealed that hybridization between Brassica rapa and Brassica oleracea was the main expansion mechanism for BnbHLHs. Expression analyses showed that BnbHLHs were wide and formed six main patterns, suggesting they may participate in various aspects in B. napus during the development. The expression profiles under five hormone treatments (IAA, ABA, ACC, GA3, 6-BA) in roots further revealed the active response of BnbHLHs with a large proportion of which being induced. qRT-PCR analysis confirmed the expression profiles of five candidate BnbHLHs under five hormone inductions. Up to 246 BnbHLHs from nine subfamilies were predicted to have potential roles relating to root development by joint analysis of expression profile and homolog function. Further, the MYB/bHLH/WD40 (MBW) protein complex regulating root hair development were verified in B. napus by yeast two-hybrid experiment. Conclusion: The 602 BnbHLHs identified from B. napus could be classed into 36 subfamilies, and those members from the same subfamily generally have similar sequence motifs. BnbHLHs may widely involve in root biological process in B. napus. Overall, this study provides important insights into the characterization and potential functions of B. napus bHLH super gene family and thus will be useful in future gene function research.
Background

Transcription factors (TF) are an important kind of genes that were widely distributed in eukaryote kingdom and usually contain two different functional domains involved in DNA binding and transcript activities [1, 2]. The bHLH transcription factor was characterized by an approximately 50-60aa conserved DNA binding domain consisted of two main regions, namely the basic and the helix-loop-helix (HLH) regions. The basic region is a 10-15aa region at the N-terminal, which functions as DNA recognition and allows the binding of HLH region [3]. The HLH region, composing of two relatively conserved amphipathic helices linked by a divergent loop, is approximately 40aa with more hydrophobic amino acids, which contributes to its function in the dimerization between 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 (bHLHs) were largely increased during the evolution. For example, there was only one bHLH gene in *Cyanidioschyzon Merolae* [6], 98 bHLHs in moss [7], 208 in *Zea mays* [8], 167 in *Arabidopsis* [7], 159 in tomato [9] while 230 in Chinese cabbage [10]. The substantial increase of bHLHs resulted in their important roles in diverse physiological and developmental processes, with majority are involved in metabolic and developmental processes. For instance, *AtbHLH045/MUTE* controls sequential cell fate; *SibHLH22* in tomato promotes early flowering and accelerates fruit ripening; *SPT* is required either for carpel development or is involved in endocarp margin development in *Arabidopsis or Prunus persica* [10, 11]; *MYC2* regulates sesquiterpene and artemisinin biosynthesis in various species like *Arabidopsis, Aquilaria sinensis* and *Artemisia annua* [12, 13]. Meanwhile, some bHLHs are related to abiotic stress response, including cold, drought, and salt stresses etc, e.g. *PIF4* in *Arabidopsis* mediates plant
architecture to response to high temperatures [14]; StbHLH1 in potato also responses to high temperature by regulating anthocyanin biosynthesis [15]; while MdbHLH3 in Malus domestica responds to low temperature [16]. Moreover, bHLHs also respond to various hormone inductions. For instance, Arabidopsis MYC2 is well known for its conserved roles in abscisic acid (ABA), jasmonic acid (JA) and light signaling pathways [17-19]; AP2/ERF is a jasmonate-responsive transcription factor involved in secondary metabolism [20].

Besides, bHLHs also contributes to Fe homeostasis in Arabidopsis and rice [21, 22]. Notably, bHLH proteins trend to function as protein complex. For example, Arabidopsis GL3/EGL3 gene is widely known as epidermal cell fate specification and hair root regulator by forming protein interaction complex with TTG1 (WD40 repeat protein) and WER/GL1 (R2R3-MYB protein) (MBW complex) [23-25].

In this study, we identified 602 bHLHs in the important economy crop, Brassica napus (BnbHLHs), and mapped them to the 19 B. napus chromosomes. According to the phylogenetic analysis and gene functions, the B. napus bHLH gene family is divided into 43 subfamilies with seven subfamilies are 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 BnbHLHs in B. napus. Expression profile analysis revealed potential functions of BnbHLHs, with focus on their possible roles in roots. qRT-PCR analysis confirmed the features of the ortholog functional genes by hormone induction in roots. Yeast two-hybrid experiment further verified the interaction relationships of BnbHLH544 (BnGL3) protein in the MBW complex, providing strong support to their roles in hair root development.

Methods
Sequence retrieval

A preliminary search for *B. napus* bHLH proteins in Genoscope (http://www.genoscope.cns.fr/brassicanapus/) is performed using BLASTP with at least one representative sequence of the bHLH domain for each bHLH subfamily. The redundant sequences are discarded to ensure the candidate genes mapped to unique loci in their respective genome. We then confirm the putative non-redundant sequences to ensure that the candidates contain the bHLH domain using ExPASy (http://expasy.org/prosite/) [26] and MEGA 5.0 [27] software. Finally, all candidates are named according to the chromosome locus.

Phylogenetic tree construction

Multiple sequence alignment is performed using bHLH protein sequences of *Arabidopsis* and *B. napus* by MAFFT software (http://mafft.cbrc.jp/alignment/server/) [28]. The phylogenetic tree is constructed based on the alignment of the bHLH domains using MEGA5.0 [27]. The neighbor-joining (NJ) method is applied, with 1000 iterations for the bootstrap values, p-distance model, and pairwise deletion for gap treatment. Tree files are viewed and edited using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Chromosome localization and gene synteny analysis

Information regarding chromosome length and gene locations are obtained from the BRAD (http://brassicadb.org/brad/index.php) and Genoscope databases (http://jacob.cea.fr/drf/ifranc/oisjacob/Pages/Departements/Genoscope.aspx), respectively. Mapchart software is used to draw the chromosome map of candidate *BnbHLHs*. Locations of *bHLHs* in *Arabidopsis*, *B. rapa*, and *Brassica oleracea* are also determined using the same method. We then use CoGe software (https://genomevolution.org/coge/) to conduct a
gene synteny analysis of bHLHs in *Arabidopsis*, *B. napus*, *B. rapa*, and *B. oleracea.*

**Intron/exon structure analysis**

To find the intron distribution and splicing phase in the candidate *BnbHLHs*, we compare and view the coding DNA (CDS) and DNA sequences of candidates using GSDS software (http://gsds.cbi.pku.edu.cn/). Then manually locates the intron insertion sites in the corresponding protein sequences.

**Identification of conserved motifs**

Full-length protein sequences of candidate BnbHLHs are analysed using MEME software (http://meme-suite.org/tools/meme) [29], with the following parameters: optimum motif width ≥ 6 and ≤ 250 and maximum number of motifs to identify = 30.

**Expression analysis of BnbHLHs**

The *B. napus* expression datasets are downloaded from the BioProject (NCBI database: PRJNA358784). The data were obtained from various tissues at different *B. napus* developmental stages and under five hormones induction (IAA, GA3, 6-BA, ABA, and ACC) were used to analyse the expression profiles of candidate *BnbHLHs* respectively. The expression profiles are analysed using Cluster 3.0 software [30] and the heatmaps are drawn using the R package. All genes with FPKM <1 are excluded from the heatmap, as they may be pseudogene or may be expressed only under specific stresses or inductions.

For qRT-PCR analysis, seeds of *B. napus* variety Zhongshuang 11 (ZS11) are obtained from the College of Agriculture and Biotechnology, Southwest University, and germinated on petri dishes. At the five-leaf stage, seedlings are treated in Hoagland liquid medium
containing five phytohormone (50 µM ABA, 120 µM GA, 75 µM 6-BA, 60 µM ACC, and 10 µM IAA) respectively. The seedlings are grown in an artificial climatic chamber at 25°C with a 16/8 h photoperiod (day/night). The root tissues are 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.

Total RNA is extracted using EASYspin total RNA Extraction kit (Biomed, Beijing). The total RNA sample is treated with DNase I (Promega, USA) before use. First-strand cDNA synthesis is performed using the M-MuLV RT kit (Takara Biotechnology, Japan). The fluorescence is measured after the extension step by using the CFX Connect™ Real-Time System (Bio-Rad, Chongqing, China) and the SYBR-Green PrimeScript RT-PCR Kit (Takara Biotechnology, Japan). The *B. napus Actin7 (BnActin7)* (GenBank accession no. AF024716) is used as the reference gene. The primers used in this analysis are listed in Additional file 9: Table S9. Three biological replicates are included for each treatment, and each consists of three technical replicates. The reaction conditions for real-time PCR are 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 are determined using the 2^(-ΔΔCt) method [31].

**Yeast two-hybrid assays**

The full-length ORF of *BnGL3* (BnaC09g12820D, *BnbHLH544*), *BnWER* (BnaA02g02300D), *BnTTG* (XP_013643414) and *BnCPC* (BnaA05g01400D) genes are amplified from ZS11 seedling root cDNA. For yeast two-hybrid assay, the coding cDNA sequences of *BnGL3*, *BnWER*, *BnTTG* and *BnCPC* are recombined into pGADT7 vector, while their BD domains
(BnGL3_{1231-1818}, BnWER_{1-360}, BnTTG_{1-1014} and BnCPC_{51-225}) are recombined into pGBK7 vector respectively. The primers used in this analysis are listed in Additional file 9: Table S9. Empty vectors of pGADT7 and pGBK7 are co-transformed as a negative control. The corresponding constructs are co-transformed into yeast strain AH109 and then test for protein–protein interaction relationship respectively, following the manufacturer’s protocols (Takara Biosciences, Clontech, Japan).

Results

A large number of bHLHs were identified in Brassica napus

To identify bHLH encoding genes in B. napus genome (Darmor-bzh), a preliminary repeated BLASTP search (e values of < 1.0) is performed using the representative sequences of Arabidopsis bHLH proteins as queries (Additional file 1: Table S1). To ensure the integrity of the bHLH protein data in B. napus, we refer to the method of Guo et al. [32] and searched the two sequenced B. napus cultivar genomes in GENOSCOPE (Darmor-bzh, http://www.genoscope.cns.fr/brassicanapus/) and NCBI database (ZS11, https://www.ncbi.nlm.nih.gov/annotationeuk/Brassica_napus/101/). The sequence information of candidate genes in these two cultivars are manually compared and corrected.

Originally, a large number of deduced amino acid sequences containing bHLH domains are obtained. Then, redundant sequences are discarded, and the bHLH domains are verified in the remaining sequences by ExPASy. Consequently, 11 genes are excluded from our dataset as no bHLH domain is identified by ExPASy analysis. Meanwhile, the sequence information of 65 BnbHLHs from Darmor-bzh are corrected by the data from ZS11 genome (Additional file 1: Table S1). Finally, a total of 602 BnbHLHs with relatively complete open reading frame (ORF) are obtained in this study, account for approximately 0.60% of the B.
napus protein-coding genes. The corresponding proportion in wheat, rice, maize and Arabidopsis are 0.55, 0.47, 0.59, and 0.61% respectively [13]. The candidate BnbHLHs are then named according to their chromosomal distribution orders (Additional file 1: Table S1). Physicochemical property analysis showed that the BnbHLH proteins (BnbHLHs) varied in length from 63 to 1440 amino acids (aa); their molecular weight ranged from 6.9 (BnbHLH105) to 165 kDa (BnbHLH381); and the isoelectric points are from 4.36 (BnbHLH562) to 11.79 (BnbHLH023). Subcellular localization analysis demonstrates that all BnbHLHs are located in nucleus (Additional file 1: Table S1).

For further comparative analysis across different species, we identified 255 bHLHs in B. oleracea genome by the same method (Additional file 2: Table S2). Sequence information of the bHLHs in other species like Arabidopsis, Brassica rapa, tomato, potato and rice are obtained from published researches [8-11, 13].

**Sequences characteristics of the bHLH domains of BnbHLH proteins**

To investigate the sequence features, we perform multiple alignment analysis of the 602 bHLH domains of candidate BnbHLHs. The result is visualized using Weblogo online software.

Our result shows that the length of the bHLH domains of BnHLHs is approximately 55 amino acids, ranging from 39–57 aa. The bHLH domain is generally conserved in this gene family in B. napus, in which two typical conserved regions are included, namely the basic and HLH regions. Ten residues are identified with conservation of more than 70% consensus ratio in the bHLH domains (Fig. 1a), including four are located in the basic region, five in the two helix regions and one in the loop region. Consistent with other
studies [7-10, 33], Leu-25 is the most conserved residue, with a conservation of almost 100% (Additional file 3: Table S3), indicating its essential role for bHLH proteins. Interestingly, Phe-30 is partly substituted by Ser in Arabidopsis, rice, tomato etc, however, no such situation is observed in B. napus, B. oleracea and B. rapa, (Additional file 3: Table S3), suggesting a higher conservation and/or close relationship in these three species.

To further characterize the BnbHLHs sequence features, the criterion given by Massari and Murre [34] is used (Fig. 1). Our result shows that the 602 BnbHLHs are separated into two major categories according to the sequence profiles of the bHLH domains: 132 (21.93%) atypical BnbHLHs (non-DNA-binding proteins), and 470 (78.07%) typical BnbHLHs (Fig. 1). And the latter is further consisted of three categories, including 300 (49.83%) G-box binding proteins, 103 (17.11%) E-box binding and 67 (11.13%) non-E-box binding proteins (Fig. 1). The sequences of the 132 atypical BnbHLHs are divergent in the basic region but are relatively conserved in the HLH region, especially in the loop region (Fig. 1d). Similar situation is found for the non-E-box binding proteins (Fig. 1c), suggesting its close relationship to the atypical BnbHLHs. In contrary, the residues in the basic region of the E-box/G-box DNA binding types are more conserved than the HLH region (Fig. 1b, c).

**Protein structures of BnbHLHs were conserved in each subfamily**

To determine the evolutionary relationship of BnbHLHs among Brassicease species, we construct a NJ phylogenetic tree on the basis of the alignment of 1243 bHLH domains from B. napus (602), B. rapa (230), B. oleracea (244) and Arabidopsis (170).

The 1243 bHLH proteins are divided into 35 subfamilies, comprising the largest number to date (Fig. 2a). Among which, 28 subfamilies are consistent with the previous research thus
their names keep the same [7]; seven subfamilies (S33-S39) containing bHLHs from these four species with a higher bootstrap value are newly identified in this study; while two previous reported subfamilies (S6 and S8) are not existed in this study as they were only found in lower plants (moss and algae) [7]. Compared with the division of AtbHLHs, S5 subfamily in B. napus is divided into S5 and S33; S17 is divided into S17 and S34; S21 is divided into S21 and S35; S24 is divided into S24, S36 and S37; S30 is divided into S30 and S38; whereas the orthologs of S39 in Arabidopsis are previously defined as orphan genes [7] which are defined as a new subfamily in this study. The distributions of BnbHLHs in the 36 subfamilies are biased, varying from two (S22 and S38) to 62 genes (S25). In addition, the BnbHLHs of different DNA binding types have a biased distribution tendency among different subfamilies as well, but the BnbHLHs in a given subfamily usually share the same DNA binding type (Fig. 2b). A total of 11 subfamilies (S2, S3, S5, S7, S10, S11, S13, S14, S24, S25 and S26) contain 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).

To further discovery the non-bHLH domains and explore their distribution patterns within each subfamily, the MEME tool is applied. A total of 27 conserved motifs with variable length (8-103 amino acids) are obtained (Fig. 2c Additional file 4: Table S4). Among which, motif 1 and 2 are distributed in all BnbHLHs, and made up the basic and the two helix regions of the bHLH domains, respectively. The loop region is located between motif 1 and 2, indicating that this region is variable than the basic and helix regions. Outside the bHLH domain, members of the same subfamily generally share several same motifs. For example, all BnbHLHs in subfamily S9 contain motif 22; proteins in subfamily S26 all
contain motif 5 (Fig2b, Additional file 4: Table S4). Moreover, some motifs have been characterized and were defined as additional functional properties, e.g. motifs 4, 8 and 20 were detected in many proteins in subfamilies S2 and S5 in various species, such as TabHLH239, AtMYC2, TabHLH184 and ZmbHLH103, which were found to be 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 are also conserved, which overlapped with the MIR and MYC_N domains that can interact with JAZs (jasmonate ZIM-domain) [36]. Besides, some motifs are demonstrated to be subfamily-specific, yet their functions are still unclear (Fig. 2c).

**Intron insertion patterns of BnbHLHs 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 BnbHLHs are determined by comparing their full length CDS and DNA sequences using GSDS web server.

A total of eight intron insertion patterns (pattern a to k) are 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 BnbHLHs is referred to that of Carretero-Paulet et al. 2010 [7]. In this study, the previous defined pattern a and b (Carretero-Paulet et al., 2010) are characterized as the same type because they share the same insertion sites and phase, therefore are uniformly named as pattern a. Similarly, pattern d and f (Carretero-Paulet et al., 2010) are uniformly named as f. The intron insertion positions distribute across the
basic and/or HLH regions in the bHLH domain. Among which, these in the basic and loop region are more conserved, while those in the helix regions are variable across different patterns. The intron insertion sites in the basic and helix regions are located at three highly conserved residues, Arg-11 (the E-box recognition site), Phe-21 and Lys-33 (Fig. 3). Further, the intron insertion sites of most patterns are conserved, excepting pattern j (Fig. 3). Among them, pattern a, c, e and f are similar, thus are likely to be homologous, where pattern f lacks the first intron as compared with pattern a, pattern e lacks the second intron while pattern c has the second intron inserted at L-50 as compares with pattern e. Similar situation is observed in pattern h and i. Meanwhile, phylogenetic analyses show a close relationship within intron insertion pattern a, c, e and f, further confirms their close relationship as well. Moreover, pattern k and i are indicated as the ancestral types because they are existed in members from algae [7]. Patterns f, a and k are the three types accounting for the majority of BnbHLHs (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 [8-11, 13]. Accordingly, these three patterns are obtained by many subfamilies while the remainings (pattern c, e, h, i and j) are 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 contain pattern f, excepting several genes that may attribute to incomplete genomic annotation information (Fig. 2b). The conservation of intron insertion pattern of BnbHLHs within each subfamily provides an independent criterion for the reliability of our phylogenetic analyses (Fig. 2b). Meanwhile, the intron insertion patterns of BnbHLHs is almost the same with their orthologs in Arabidopsis. The only exception is
S27 subfamily which contains pattern a for *B. napus* members while their homologs in *Arabidopsis* is pattern f [7]. To confirm this result, we further compare the corresponding results in other species, e.g. rice [33]. And we confirm 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 of *BnbHLHs* are conserved within most subfamilies and coordinated with these of *AtbHLH* orthologs as well. And the intron insertion sites in the basic and loop regions are more conserved than those in the helix regions.

**Syntenic analyses revealed duplication events and expansion mechanism of *BnbHLHs***

In this study, up to 602 *BnbHLHs* are identified, while the gene number in lower plants are much lower, like *Volvox carteri* has only three [6]. This indicates a large scale expansion for this gene family has occurred during the evolution. To explore the expansion mechanism of this gene family in *B. napus*, the chromosomal locations and syntenic relationships of *BnbHLHs* are analysed based on the genome information from Genoscope and CoGe databases.

Chromosomal location analysis showed that there are 294 and 306 *BnbHLHs* in *A_n*- and *C_n*- subgenomes, respectively, indicating no biased tendency between these two sub-genomes (Fig. 4a). The *BnbHLHs* are distributed on all of the 19 chromosomes. The *A_n*-subgenome has an average of 27.4 *BnbHLHs* on its 10 chromosomes with A10 has a minimum number
of 14 and A09 has a maximum of 41 genes. The average number of \textit{BnbHLHs} in $C_n$-subgenome is 29.9, with C06 contains a minimum of 19 genes and the C03 has as many as 48 genes. Thus, the genes on each chromosome is uneven within both subgenomes.

Among the 602 \textit{BnbHLHs}, 475 genes have syntenic relationships, and 382 of which are inherited from \textit{B. rapa} or \textit{B. oleracea} genomes (Additional file 5: Table S5). In contrast, only 79 genes (7.8\%) of 79 syntenic pairs are originated from segmental duplication in \textit{B. napus} genome, and 34 genes (5.8\%) from 28 syntenic pairs are from tandem duplication (Fig. 4b). These results provide an outstanding example of genome-wide allopolyplodization between \textit{B. rapa} and \textit{B. oleracea} that mainly contributes to the large \textit{BnbHLHs} expansion in \textit{B. napus} (63.3\%). Moreover, we find that the genome-wide duplicated genes take the largest number of \textit{BnbHLHs} in majority subfamilies, while the segmental duplicated genes take the biggest amount in subfamilies S10 and S25 (nine genes each subfamily), and the tandem duplicated genes in S12 is the most (11 genes). Furthermore, the \textit{BnbHLHs} with intron pattern f expands most in \textit{B. napus} (31 duplicats), contributing to the largest proportion of \textit{BnbHLHs} (Additional file 1: Table S1).

Taken together, the main expansion mechanism of \textit{BnbHLHs} is whole-genome duplication (allopolyplodization), while segmental and tandem duplication events preferentially occurred in certain subfamilies with relative specific intron patterns.

\textbf{Expression profiles of \textit{BnbHLHs} are wide and conserved within each subfamily}

Gene expression patterns are closely related to their functions. In order to explore important clues for gene functions, the temporal and spatial transcriptome of the 602
BnbHLHs in 50 B. napus tissues of root, leaf, flower, and seed at different development stage are characterized using the RNA-seq data (BioProject ID: PRJNA358784).

A total of 47 BnbHLHs (7.79%) are excluded from our analysis with FPKM<1, which may be pseudogenes or are expressed only at specific developmental stages or under special conditions. The remaining genes (555 genes) have relative high confidence expression levels (FPKM>1), and the majority of which showed preferential expressions in one or a few tissues/organs. Few genes are constitutively expressed in all tissues or organs tested, suggesting this gene family tends to play regulatory roles at specific developmental stages or tissues. The wide expression profile of BnbHLHs suggests their diverse roles in B. napus. The expression patterns of BnbHLHs can be summarized into six main blocks (I to VI) (Fig. 5). The BnbHLHs in blocks I, II and IV show obvious tissue-specific expression pattern; the genes in block I are specifically highly expressed in seed coat; these in block II are highly expressed in root and stem; the genes in block IV are mainly expressed in organs at the seedling stage, such as germinating seed (24 to 72h), root and hypocotyl; the genes in block VI are widely expressed in various tissues; while the genes in block III and V are mainly expressed in the tissues related to female reproduction organs, such as pistil, silique pericarp, seed coat and seed tissues. Overall, there are 180, 63, 119 and 106 BnbHLHs having relatively high expression levels in stem, leaf, seed and flower respectively, while up to 310 genes are highly expressed in roots, indicating a previously-ignored important roles of BnbHLHs in roots.

It is common that members of a given subfamily generally exhibit the same/similar expression profile. For example, members in S2 subfamily are mainly expressed in root and leaf at seedling, budding, and flowering stages (Additional file 6: Table S6). Moreover,
the bHLHs in the same subfamily probably process the same or similar expression profiles across different species as well, thus may share conserved functions during evolution. For example, AtbHLH155/CPU and AtbHLH156/LHW in subfamily S23 play an essential role in establishing vascular cells and the size of vascular initial population in root meristem [38]. The corresponding homologs in B. napus are also expressed in roots (Additional file 6: Table S6). The homologs of AtbHLH155/CPU and AtbHLH156/LHW in B. napus have relatively low expression levels in root after germinating 24 to 72h but high expression levels in mature roots, which is corresponding to their possibly functions in root vascular cells. Moreover, expression profile of 28 pairs of tandem duplication genes, and 79 pairs of segmental duplication genes show similar expression patterns, indicating the functionally redundancy of duplicates (Additional file 5,6: Table S5, S6).

Overall, the BnbHLHs have a widespread expression patterns at different levels, especially in root, which provide an important clue for the possible roles of this gene family. Moreover, BnbHLHs in a certain subfamily tend to process conserved expression patterns and high structural similarity across different species, indicating a possible function conservation during evolution.

**Plenty of BnbHLHs are induced by hormones treatments in root**

As discussed above, plenty of BnbHLHs are highly expressed in roots, implying possible roles in root related biological processes. To further explore their functional characteristics in roots, a comprehensive expression analysis of candidate BnbHLHs in roots under five hormone treatments (auxin indole-3-acetic acid, IAA; abscisic acid, ABA; cytokinin 6-benzyladenine, 6-BA; ethylene precursor 1-aminocyclopropane-1-carboxylic
acid, ACC; and gibberellic acid, GA) is performed, based on the RNA-seq data (BioProject ID PRJNA358784).

Our results show that many *BnbHLHs* (221, 36.7%) response to more than one hormone treatments in roots, and most of which are mainly clustered in ten subfamilies (S2, S16, S18, S19, S20, S21, S25, S27, S35 and S39) taking up over 50% 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) are partly induced by the five hormones in roots. Interestingly, *BnbHLHs* in subfamilies S16, S18 and S21 are up-regulated in roots under all of the five hormones inductions, while *BnbHLHs* in subfamilies S7, S23 and S34 are all down-regulated instead. Notably, *BnbHLHs* in S12 with low or no expression levels in roots (Fig. 5) are highly expressed after the five hormone treatments (Additional file 7: Table S7), in contrary, some gene that are normally highly expressed in roots are barely responded to hormone treatments. Meanwhile, *BnbHLHs* in S3, S11, S17, S22, S23, S28, S30, S31, S33, S36 and S37 hardly response to any hormone treatments in roots (Additional file 7: Table S7). Overall, most members of the *B. napus* bHLH gene family response to hormone treatments in roots, implying their important roles in hormone response in *B. napus* roots.

To further verify the results by RNA-Seq analysis, five *BnbHLHs* that have high expression levels in roots (Additional file 6: Table S6) and obviously respond to hormone inductions (Additional file 7: Table S7) are selected to analyse their expression profiles under hormone inductions (IAA, ABA, 6-BA, ACC and GA₃) by qRT-PCR method. Among them, three genes (*BnbHLH033, BnbHLH041* and *BnbHLH269*) are orthologs of *Arabidopsis ILR3* gene which was demonstrated to participate in auxin-conjugate metabolism [39]; two
genes (BnbHLH453 and BnbHLH126) are orthologs of Arabidopsis MYC2 gene which was involved in ABA, JA and light signalling pathways. As shown in Fig. 6, our results by qRT-PCR are similar to that of the RNA-seq analyses. These five BnbHLHs are positively induced by all of the five hormone treatments, with BnbHLH126 generally has relative higher expression levels than the others. Moreover, the genes from the same clade show similar expression patterns under a certain hormone induction. For example, BnbHLH033, BnbHLH041 and BnbHLH269 from ILR3 clade in S4 subfamily have similar expression patterns under IAA, ACC and ABA treatments (Fig. 6a-c). Meanwhile, their expression profiles are different under GA3 and 6-BA inductions, where BnbHLH033 is significantly down-regulated while BnbHLH041 and BnbHLH269 are significantly up-regulated (Fig. 6d-e). Similarly, the expressions of BnbHLH453 and BnbHLH126 from MYC2 clade in S2 subfamily show similar expression pattern under the five hormone treatments (Fig. 6). In addition, cis-acting elements analysis revealed that the promoter regions of these five BnbHLHs contain more than one cis-acting elements that are related to hormone response (Additional file 8: Table S8). This further supports our above results.

In a word, expression profile analyses reveals that a large proportion of BnbHLHs are induced by more than one hormone treatment in roots, indicating possibly roles of this gene family in root development in B. napus. qRT-PCR experiment confirms the hormone-induced expression characteristics of BnbHLH033, BnbHLH041, BnbHLH269, BnbHLH453 and BnbHLH126 which provides 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 widely performed in a mass of species [7-10]. However, we find that there is still lacking a uniform subfamily classification of this gene family in plants, resulting in many sorts of 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 2010 [6] and Carretero-Paulet et al. 2010 [7], respectively. Although both of these two studies divided this gene family into 28 subfamilies, in fact, their results are not exactly the same where 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 phylogenetic analyses, as well as the criterion applied in the classification on the basis of the phylogenetic trees. In Pires and Dolan’ study, they adopted the Arabidopsis bHLHs proposed by Heim et al. 2003 [3] which included 118 AtbHLHs. However, there are a large number of bHLHs were missed in their study due to the restriction of lower Arabidopsis genome version used. By contrast, there was a total of 162 AtbHLHs in the study of Pires and Dolan resulting in 12 newly identified orphan genes, such as AtbHLH022/DYT1, AtbHLH159/P1r2 and AtbHLH102/BIM2 etc. Generally, the division of gene family is performed on the basis of the topology and bootstrap value of the phylogenetic tree [40, 41]. However, we find that subfamilies VII (a+b), IX and IIIf in the result of Pires and Dolan did not consist of a consensus node, but across different branches/clades instead [6]. By contrast, the situation was well defined in the results of Carretero-Paul, indicating it is relatively more credible. Furthermore, most of the orphan genes (62 genes) identified by Pires and Dolan were classified into different subfamilies by Carretero-Paulet et al. Consequently, an obvious decrease of orphan gene proportions (total 15 genes, 2%) was observed, suggesting a relative better solution for the classification of orpan genes. To date, this gene family have been genome-wide characterized in various plants, such as Z. mays [8], tomato [9], and B. rapa [10], etc.,
and the classification in most of those studies referred to that of Pires and Dolan. As a result, there did exist several inadequacies for the subfamily classification in those researches. For instance, in Z. mays, subfamilies VII, VIII, IX, XVI and XVIII did not clustered in a consensus node respectively [8]; the same situation were observed in B. rapa, e.g. Illf and Ib (2) [10]. Together, the criterion of Carretero-Pault is relative more credible.

In this study, the 602 candidate BnbHLHs are classified into 43 subfamilies, as refered to the method of Carretero-Pault et al [7] (Fig. 2). Of which, onesubfamily (S39) is newly identified (AtbHLH151 homolgs), while 11 ones are separated from four former subfamilies that had a relatively low bootstrap value support in previous study [7], including three subfamilies are separated into two new subfamilies (S5 and S33; S17 and S34; S21 and S35; S30 and S38, respectively) and one subfamily is divided into three new subfamilies (S24, S36 and S37) in our study. The difference may be attributed to more sequences from Brassicaceae species applied that have close evolutionary relationship. It is reported that many other characteristics of candidates generally supply important clues to support the subfamily classification, including highly conserved intron pattern and motif distribution within each subfamily [40, 41]. Similarly, the sequence characteristics within each subfamily are also highly conserved in our results, which provides an independent support to our phylogeny analysis and classification as well. For example, subfamily S39 had a relatively high bootstrap value (Fig. 2a), share conserved non-E-box and k intron pattern (Fig. 2b, c); S24 subfamily defined by Carretero-Pault et al. [7] is separated into subfamilies S24, S36 and S37 in this study (Fig. 2a), and these three new subfamilies all contained subfamily-specific DNA binding type, intron insertion patterns and motifs (Fig. 2b, c); the former S17 subfamily is separated into S17 and S34 subfamilies in present
study (Fig. 2a), and the BnbHLHs in the new S17 subfamily had a E-box in DNA binding domain while these in S33 are all non-DNA binding types; Similarly, the former S5 subfamily in Carretero-Paulet’s study [7] is divided into two new subfamilies in our study as well.

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

The potential roles of bHLHs in B. napus roots

Given their crucial roles in diverse plant biological processes, functional researches of plant bHLHs have been a world-wide research hotspot in the past three decades. The majority of known functions of bHLHs are regulatory components in transcriptional networks controlling a number of biological processes, including metabolic and developmental processes. However, an increasing number of studies have demonstrated that they play important roles in plant root development as well. Till now, the functions of bHLHs in plant roots mainly focused on the efficiency of iron-uptake, salt and drought stress response, hormone mediated regulation (ABA, JA, BR and IAA), size control of vascular in root and meristem (Table 1).

As mentioned above, the bHLHs are widely expressed in root tissues in B. napus. Accordingly, to date, 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 [38], and

AtbHLH024/SPT in S24 regulates root growth by controlling the size of root meristem [42];

homologs of AtbHLH002/GL3 (S5), AtLRLs (S26) and AtRSLs (S28) are involved in root hair
development [35, 43-52]; AtbHLH92 in S7 and AtbHLH129 in S27 regulate root elongation
[53, 54]; AtbHLH74 of S25 regulates seedling root growth [55]. Accordingly, BnbHLH
orthologs of those functionally-characterized bHLHs also showed high expression levels in
roots (Fig. 5), indicating their potential roles in root development.

bHLHs are also proved to involve in hormone signaling pathways and environmental stress
in plant roots. Until now, many members in 11 subfamilies (S1-S4, S7, S12, S15, S16, S19,
S26 and S27) have been identified to response to Fe, salt, and drought stresses, and
involve in ABA and JA pathways (Table 1). For example, members in subfamilies S1, S4 and
S12 are affected Fe-uptake [56-58]; AtbHLH17/AIB (S2), AtbHLH92 (S7), AtbHLH112 (S15)
and bHLH122 (S27) response to salt stress [53, 59-61]; AtbHLH68 (S15) and ZmPTF1 (S26)
response to drought stress [45, 62]. Meanwhile, many MYC2 homologous genes in S2
subfamily e.g. Catharanthus roseus CrMYC2 [63] and Salvia miltiorrhiza SmMYC2 [64], are
known to involve in ABA and JA [17-19] members in subfamilies S15 and S26 have been
demonstrated to response to ABA [45, 62]; members in S19 negative circuit of BR
signaling pathway [65]; and the genes in S4 participate in auxin-conjugate metabolism
[39, 66] (Table 1). Accordingly, our RNA-seq data analyses shows that many BnbHLHs
(221, 36.7%) response to more than one hormone treatments in B. napus roots (Additional
file 7: Table S7); and our qRT-PCR analysis further confirmed that two BnbHLHs
(BnbHLH126 and BnbHLH453) in S2 and three BnbHLHs (BnbHLH033, BnbHLH041 and
BnbHLH269) in S4 are obviously induced by hormone treatments in B. napus roots (Fig. 6).
Together, these results suggested possible wide roles of bHLHs hormone signaling
pathways in plant roots.

To date, the mechanisms of bHLHs function in plant development processes have been extensively studied in numerous plants at biochemical, genetic, and molecular levels [7]. And it is revealed that bHLH proteins usually functioned in homo- or heterodimeric forms by interacting with bHLHs or other types of transcription factors. The MBW complex is just one of the classic such examples, which consisted of a MYB (CPC/WER), a bHLH (GL3/EGL3), and a WD40 (TTG1) protein. In this complex, both the MYB and the WD40 proteins bind to the bHLH protein [35, 47, 48]. Moreover, this complex is proved to be conserved in plants during evolution [73, 76, 77], which regulates five processes, namely the production of anthocyanidin, proanthocyanidin, seed-coat mucilage, and trichomes and root hairs development [70]. For example, the roles of this complex in controlling the development of trichomes and root hairs are verified in many species, ranging from monocots (Z. mays) to dicots (Arabidopsis, Arabisalpina, Gossypium hirsutum and Petunia hybrida) [23, 25, 78]. And four residues of GL3 and EGL3 proteins are identified to be vital for the protein interaction, including Phe-177 of AtGL3, Pro-377 of AtGL3, Asp-477 of AtGL3 and Ser-589 of AtGL3 [70]. Similarly, these key residues are observed in BnbHLH orthologs as well (Additional file 1: Table S1), suggesting they may have the same/or similar roles in B. napus. To further verify the interaction relationship among these three types of proteins as well as the existence of the MBW complex in B. napus, the yeast-two-hybrid (Y2H) experiment is applied in present study. Our result shows that BnbHLH544 (ortholog of EGL3) can interact with BnCPC (BnaA05g01400D, ortholog of CPC)/BnWER (BnaA02g02300D, ortholog of WER), and BnTTG (XP_013643414, ortholog of TTG) (Fig. 7), indicating this complex is existed in B. napus as well, which probably functions through the same mechanism.
Taken together, expression profile analyses along with previous gene function researches implied that BnbHLHs may widely involve in root biological process in B. napus, including root growth, hormone signal responses, etc. Y2H experiment further confirmed the interaction relationships of candidate BnbHLHs in the MBW complex. Our study provided valuable foundation for further gene function researches.

Declarations

Acknowledgments

Not applicable.

Funding

This work was supported by National Natural Science Foundation of China (31471528 and 31671727), National Program on Key Basic Research Project (973 Program, 2015CB150201), Fundamental Research Funds for the Central Universities of China (XDJK2014B035).

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 contribution in the 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., P.C.G and M.M.L are contributed to data analysis. Y.Z.K., W.Y.W., H.D., C.P., M.M.W., J.W. and P.F.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 interest.

References

1. Amoutzias GD, Veron AS, Weiner J, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG, Robertson DL: 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, Adam L, Pineda O, Ratcliffe OJ, Samaha RR: 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, Li G: 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, Hou XL: 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. Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, Guo J, Liang W, Chen L, Yin J et al: Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. *Plant Physiol* 2006, 141(4):1167-1184.

12. Xiang L, Jian D, Zhang F, Yang C, Bai G, Lan X, Chen M, Tang K, Liao Z: The cold-induced bHLH transcription factor AabHLH112 promotes artemisinin biosynthesis in Artemisia annua. *J Exp Bot* 2019:pii: erz220.

13. Xu YH, Liao YC, Lv FF, Zhang Z, Sun PW, Gao ZH, Hu KP, Sui C, Jin Y, Wei JH: 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.

14. de Lucas M, Daviere JM, Rodriguez-Falcon M, Pontin M, Iglesias-Pedraz JM, Lorrain S, Fankhauser C, Blazquez MA, Titarenko E, Prat S: A molecular framework for light and gibberellin control of cell elongation. *Nature* 2008, 451(7177):480-484.

15. Liu Y, Lin-Wang K, Espley RV, Wang L, Li Y, Liu Z, Zhou P, Zeng L, Zhang X, Zhang J et
al: StMYB44 negatively regulates anthocyanin biosynthesis at high temperatures in tuber flesh of potato. *J Exp Bot* 2019, 70(15):3809-3824.

16. 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.

17. 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.

18. 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.

19. 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.

20. Zhou M, Memelink J: Jasmonate-responsive transcription factors regulating plant secondary metabolism. *Biotechnol Adv* 2016, 34(4):441-449.

21. Ogo Y, Itai RN, Nakanishi H, Kobayashi T, Takahashi M, Mori S, Nishizawa NK: 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD et al: The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res 2001, 29(1):37-40.

27. 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.

28. Katoh K, Standley DMB: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology Evolution 2013, 30(4):772-780.

29. 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.

30. Zaharia M, Chowdhury M, Franklin MJ: Spark: Cluster computing with working sets. HotCloud 2010, 10(10-10):95.

31. 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.

32. Guo P, Wen J, Yang J, Ke Y, Wang M, Liu M, Ran F, Wu Y, Li P, Li J 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.

33. Lorenzo CP, Anahit G, Irma RV, Martínez-García JF, Bilbao-Castro JR, Robertson DL:
Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. *Plant Physiology* 2010, 153(3):1398-1412.

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, Zhou XE, Chen J, Brunzelle J, Griffin PR 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, Lim YP: Genome-wide identification, characterization, and comparative phylogeny analysis of MADS-box transcription factors in Brassica rapa. *Genes Genom* 2014, 36(4):509-525.

38. Ohashi-Ito K, Bergmann DC: Regulation of the Arabidopsis root vascular initial population by LONESOME HIGHWAY. *Development* 2007, 134(16):2959-2968.

39. Rampey RA, Woodward AW, Hobbs BN, Tierney MP, Lahner B, Salt DE, Bartel B: An Arabidopsis basic helix-loop-helix leucine zipper protein modulates metal homeostasis and auxin conjugate responsiveness. *Genetics* 2006, 174(4):1841-1857.

40. Du H, Liang Z, Zhao S, Nan MG, Tran LS, Lu K, Huang YB, Li JN: The Evolutionary History of R2R3-MYB Proteins Across 50 Eukaryotes: New Insights Into Subfamily Classification and Expansion. *Sci Rep* 2015, 5:11037.

41. 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.

42. 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.

43. 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.

44. 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.

45. 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.

46. Menand B, Yi K, Jouannic S, Hoffmann L, Ryan E, Linstead P, Schaefer DG, Dolan L: An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* 2007, 316(5830):1477-1480.

47. 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.

48. 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.

49. 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.

50. Zhang C, Simpson RJ, Kim CM, Warthmann N, Delhaize E, Dolan L, Byrne ME, Wu Y, Ryan PR: 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.

51. 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.

52. 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.

53. 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.

54. Tian H, Guo H, Dai X, Cheng Y, Zheng K, Wang X, Wang S: An ABA down-regulated bHLH transcription repressor gene, bHLH129 regulates root elongation and ABA response when overexpressed in Arabidopsis. *Sci Rep* 2015:5:17587.

55. Bao M, Bian H, Zha Y, Li F, Sun Y, Bai B, Chen Z, Wang J, Zhu M, Han N: 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.

56. Gratz R, Manishankar P, Ivanov R, Koster P, Mohr I, Trofimov K, Steinhorst L, Meiser J, Mai HJ, Drerup M et al: CIPK11-Dependent Phosphorylation Modulates FIT Activity to Promote Arabidopsis Iron Acquisition in Response to Calcium Signaling. *Dev Cell* 2019, 48(5).

57. 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.

58. 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.

59. 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.

60. Chen HC, Hsieh-Feng V, Liao PC, Cheng WH, Liu LY, Yang YW, Lai MH, Chang MC: 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.

61. 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.

62. 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.

63. Zhang H, Hedhili S, Montiel G, Zhang Y, Chatel G, Pre M, Gantet P, Memelink J: 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.

64. Du T, Niu J, Su J, Li S, Guo X, Li L, Cao X, Kang J: SmbHLH37 Functions Antagonistically With SmMYC2 in Regulating Jasmonate-Mediated Biosynthesis of Phenolic Acids in Salvia miltiorrhiza. *Front Plant Sci* 2018, 9:1720.

65. 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.
66. Samira R, Li B, Kliebenstein D, Li C, Davis E, Gillikin JW, Long TA: The bHLH transcription factor ILR3 modulates multiple stress responses in Arabidopsis. *Plant Mol Biol* 2018, 97:297-309.

67. 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.

68. Penuelas M, Monte I, Schweizer F, Vallat A, Reymond P, Garcia-Casado G, Franco-Zorrilla JM, Solano R: Jasmonate-related MYC Transcription Factors are Functionally Conserved in Marchantia polymorpha. LID - tpc.00974.2018 [pii] LID - 10.1105/tpc.18.00974 [doi]. (1532-298X (Electronic)).

69. Chiasson DM, Loughlin PC, Mazurkiewicz D, Mohammadidehcheshmeh M, Fedorova EE, Okamoto M, McLean E, Glass AD, Smith SE, Bisseling T 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.

70. 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.

71. 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.

72. Li L, Gao W, Peng Q, Zhou B, Kong Q, Ying Y, Shou H: Two soybean bHLH factors regulate response to iron deficiency. *J Integr Plant Biol* 2018, 60:608-622.

73. Xu W, Dubos C, Lepiniec L: Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci* 2015, 20:176-185.

74. Zheng K, Wang Y, Zhang N, Jia Q, Wang X, Hou C, Chen JG, Wang S: Involvement of
PACLOBUTRAZOL RESISTANCE6/KIDARI, an Atypical bHLH Transcription Factor, in Auxin Responses in Arabidopsis. *Front Plant Sci* 2017:8:1813.

75. Gajewska P, Janiak A, Kwasniewski M, Kedziorski 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.

76. Ramsay NA, Glover BJ: MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci* 2015, 10:63-70.

77. Zhang B, Schrader A: TRANSPARENT TESTA GLABRA 1-Dependent Regulation of Flavonoid Biosynthesis. *Plants (Basel)* 2017, 6(4):pii: E65.

78. Zhang B, Hulskamp M: Evolutionary Analysis of MBW Function by Phenotypic Rescue in Arabidopsis thaliana. *Front Plant Sci* 2019:10:375.

### Tables

Table 1. Functionally Characterized plant bHLH Proteins Related to Root.

| Sub-family | Gene name       | Species                                  | Function                                                                 |
|------------|-----------------|------------------------------------------|--------------------------------------------------------------------------|
| S1         | AtbHLH029/FIT   | Arabidopsis                              | Affects the plant's Fe-uptake ability[56].                               |
| S2         | bHLH006/MYC2    | Arabidopsis, Oryza sativa, Solanum lycopersicum, Marchantia polymorpha, Catharanthus roseus, Salvia miltiorrhiza, Triticum aestivum, Artemisia annua, Aquilaria sinensis | Regulate salt tolerance AtNHX6 and AtNHX6 in Promotes ABA and JA responsiveness[61, 63, 68]. |
| S3         | GmbHLHm1        | Glycine max                              | Linked to the activity of a class of ammonium channels and to signaling cascades influencing a nodule circadian clock[69]. |
| S4         | AtbHLH17/AtAIB  | Arabidopsis                              | Response to NaCl, Man oxidative stress[59].                              |

35
| Page | Protein Name | Organism | Function |
|------|--------------|----------|----------|
| S4   | AtbHLH11     | Arabidopsis | High expression in root negative regulator of Fe homeostasis[57]. |
|      | AtbHLH105/ILR3 | Arabidopsis | Regulatory network that controls wounding pathogen response by modulating Fe accumulation under iron deficiency. Participate in conjugate metabolism[57]. |
| S5   | bHLH001/EGL3 | Arabidopsis, Arabis alpina, Gossypium hirsutum, Petunia hybrida, Zea mays | ENHANCER OF GLABRA |
|      | bHLH002/GL3 | Arabidopsis | Partly redundantly regulate anthocyanin biosynthesis: trichome and root hair development[25, 35, 47]. |
| S7   | AtbHLH92     | Arabidopsis | Response to NaCl, dehydration, and cold stress and regulate root elongation[52]. |
| S12  | GmbHLH57 GmbHLH300 AtbHLH38 AtbHLH39 AtbHLH100 sativa, Glycine max, Z. mays | Arabidopsis, S. lycopersicum, O. sativa, S. lycopersicum, S. lycopersicum | Induced by Fe deficiency, roots and shoots[56, 57]. |
| S15  | AtbHLH112    | Arabidopsis | Regulation of the salt stress response and regulate root elongation[60]. |
|      | AtbHLH68     | Arabidopsis | Regulation of ABA homeostasis and drought stress tolerance. Regulate lateral root development[62]. |
| S16  | AtbHLH163/PRE6 | Arabidopsis | Negatively regulates auxin responses[74]. |
| S19  | AtbHLH148/AIF2 | Arabidopsis, O. sativa, S. lycopersicum | BIN2/AIF2-mediated negative circuit of BR signaling pathways[65]. |
| S23  | AtbHLH156/LHW | Arabidopsis | Regulates the size of the vascular initial population in the root meristem[38]. |
| S24  | AtbHLH024/SPT | Arabidopsis | Regulates root growth controlling the size of the root meristem[42]. |
| S25  | AtbHLH74     | Arabidopsis | Root growth in seedlings[42]. |
| S26  | bHLH066/LRL1 | Arabidopsis, Lotus japonicas, Physcomitrella patens, Z. mays, O. sativa, Hordeum vulgare. | Act redundantly to positively regulate the development of root hairs[43, 44, 75]. |
bHLH069/LRL2  
bHLH082/LRL3  
bHLH007/LRL4  
Negatively control root growth by repressing the expression of LRL3

bHLH059/LRL5  
ZmPTF1  
Z. mays  
Regulates drought tolerance in maize by promoting root development and ABA synthesis[45].

S27  
AtbHLH129  
Arabidopsis  
Regulates root elongation and ABA response[54].

bHLH122  
Arabidopsis, O. sativa, S. lycopersicum  
Regulates salt tolerance in root. Promotes ABA and JA responsiveness[61].

S28  
bHLH086/RSL1  
Arabidopsis, M. polymorpha, Brachypodium, P. patens  
Partially redundant and involved in root hair development[46, 49-52].

bHLH085/RSL2  
bHLH084/RSL3  
bHLH054/RSL4  
bHLH139/RSL5

Figures
Figure 1

Sequence characteristics of the bHLH domains in different DNA-binding types.

Multiply sequence alignments are conducted with the bHLH domains of all candidate proteins or of different DNA-binding types, respectively. Number in bracket indicates the amount of BnbHLHs in a certain category. Protein secondary structures are illustrated under the sequences. Red asterisks stand for the residues with over 90% similarity; black asterisk indicates over 70% similarity. Black triangle at the top of the sequence indicates the E-box recognition sites; red triangle indicates 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 would be 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 36 bHLH subfamilies. a Phylogenetic relationships of 1243 bHLHs from B. napus (602), B. rapa (230), B. oleracea (244) and Arabidopsis (170). The phylogenetic tree is generated based on the alignment of 1243 bHLH domains of the corresponding bHLH proteins with 1000
bootstrap replicates. The subfamilies marked in gray are the newly identified ones in this study as compared to the results in Arabidopsis [7].

b DNA-binding types and intron insertion types of BnbHLHs in each subfamily. The illustration of intron insertion patterns are shown in Fig. 3.

c Architecture of conserved protein motifs of BnbHLHs in each subfamily by MEME analysis. Block with black background indicates the bHLH domain. Block with solid line represents the motif distributed in all proteins in a certain subfamily, while these with dotted line represents 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 BnbHLHs. The intron patterns are classified into 8 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. Black triangle at the top of the sequence indicates the conserved E-box recognition sites. The intron pattern of each subfamily is
The distributions of candidate BnbHLHs on each chromosome and the expansion mechanism of BnbHLHs. 

**Figure 4**

The distributions of candidate BnbHLHs on each chromosome and the expansion mechanism of BnbHLHs. a Distribution of BnbHLHs on the 19 B. napus chromosomes. Blue box indicates chromosomes in An-subgenome while orange implies those in Cn-subgenome. Chromosome number is showed on the left and the number of BnbHLHs on a certain chromosome is listed on the right. b Percentage of BnbHLHs derives from An- or Cn-subgenome.
Expression profiles of candidate BnbHLHs in 50 B. napus tissues or organs across different developmental stages. Six major blocks of different expression patterns are illustrated on the right. The tissues that are used for expression analysis are indicated at the bottom 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; SP, silique; Se, seed; SC, seed coat; Em, embryo; Co, cotyledon. s, seedling stage; b, bud stage; i, initial flowering stage; and f, full-bloom stage. The colour bar represents log10 (FPKM+1).
Expressions of five BnbHLHs under hormone treatments. The transcript levels are determined in seedling roots by qRT-PCR under five hormone inductions. 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. Bars represent means ± SEM (n = 3 independent biological replicates). Comparison between treatments and CK (0 h) according to Welch’s t-test (*P < 0.05; **P < 0.01).
Yeast two-hybrid analysis of protein–protein interactions in MBW complex.

Interactions, including BnGL3, BnWER, BnCPC and BnTTG are verified by culturing the transformed AH109 cells on (a) DDO (SD/-Leu/-Trp) and (b) QDO (SD/-Ade/-His/-Leu/-Trp).

Supplementary Files

This is a list of supplementary files associated with the primary manuscript. Click to download.

Table S6.xlsx
Table S1.xlsx
Table S8.xlsx
Table S9.xlsx
Table S7.xlsx
Table S3.xlsx
Table S5.xlsx
Table S2.xlsx
Table S4.xlsx