Comparative analysis of basic helix–loop–helix gene family among Brassica oleracea, Brassica rapa, and Brassica napus

Liming Miao
Zhejiang University
Yingying Gao
Zhejiang University
Kun Zhao
Zhejiang University
Lijun Kong
Zhejiang University
Shubo Yu
Zhejiang University
Rongrong Li
Zhejiang University
Kaiwen Liu
Zhejiang University
Xiaolin Yu (xlyu@zju.edu.cn)
Zhejiang University https://orcid.org/0000-0003-3256-0638

Research article

Keywords: Brassica oleracea; Brassica rapa; Brassica napus; bHLH gene family; Whole genome identification; Phylogeny; Expression analysis; Molecular characteristic analysis

Posted Date: September 4th, 2019

DOI: https://doi.org/10.21203/rs.2.10332/v1

Version of Record: A version of this preprint was published on February 24th, 2020. See the published version at https://doi.org/10.1186/s12864-020-6572-6.

Comparative analysis of basic helix–loop–helix gene family among *Brassica oleracea*, *Brassica rapa*, and *Brassica napus*

Liming Miao[1,2], Yingying Gao[1,2], Kun Zhao[1,2], Lijun Kong[1,2], Shubo Yu[1,2], Rongrong Li[1,2], Kaiwen Liu[1,2], Xiaolin Yu[1,2*]

[1]*Laboratory of Cell and Molecular Biology, Institute of Vegetable Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou, Zhejiang 310058, China*

[2]*Key Laboratory of Horticultural Plant Growth, Development, and Quality Improvement, Ministry of Agriculture, Zhejiang Provincial Key Laboratory of Horticultural Plant Integrative Biology, Hangzhou, Zhejiang 310058, China*

*To whom reprint requests should be addressed. Email address: xlyu@zju.edu.cn

E-mail addresses : Liming Miao, 11616010@zju.edu.cn; Yingying Gao, 21716057@zju.edu.cn; Kun Zhao, 21616070@zju.edu.cn; Lijun Kong, 11516009@zju.edu.cn; Shubo Yu, 21716153@zju.edu.cn; Rongrong Li, 21716403@zju.edu.cn; Kaiwen Liu, 3160100649@zju.edu.cn; Xiaolin Yu (Corresponding author), xlyu@zju.edu.cn.

Abstract

Background: The basic helix–loop–helix (bHLH) transcription factors exist widely in eukaryotes and play important roles in development and stress response regulation in plants. The bHLH gene family has been identified in many species, except for *Brassica oleracea* and *Brassica napus* thus far. This study aims to identify the bHLH family members in *B. oleracea*, *Brassica rapa*, and *B. napus* and elucidate the expression, duplication, phylogeny and evolution characters of these genes.

Result: A total of 268 bHLH genes in *B. oleracea*, 440 genes in *B. napus*, and 251 genes in *B. rapa*, including 21 new bHLH members, have been identified. Subsequently, the analysis of the phylogenetic tree, conserved motifs and gene structures showed that the members in the same subfamily were highly conserved. Most *Ka/Ks* values of the homologous gene were <1, which indicated that the homologous genes suffered from strong purifying selection for retention. The *Ks* values of the three *Brassica* crops were concentrated in the range of 0.3–0.5. Hence, the divergence time of the bHLH gene family between *Brassica* crops and *Arabidopsis thaliana* is approximately 10–18 MYA. The retention rates of BrabHLH and BolbHLH genes were 51.6% and

55.1%, respectively. A total of 182 genes were lost in *B. napus* after tetraploid. GO annotations of *BolbHLH* genes showed that most genes focused on DNA-binding transcription factor, DNA-binding, and protein dimerization activities. The temporal and spatial expression patterns of 50 *BolbHLH* genes were diverse, some of which showing high expression in the reproduction tissue, while some had high expression in the root. The comparison of expression patterns between *B. rapa* and *B. napus* showed that they had similar expression patterns in the root and contrasting patterns in the stems, leaves, and reproduction tissues. However, the expression patterns of *B. oleracea* and *B. napus* were different.

Conclusion: This study is the first to report about the gene family analysis of the bHLH gene in *B. oleracea* and *B. napus.* Our results not only offer useful information on the functional analysis of the bHLH gene but also provide new insights into the evolution of *Brassica* spp.

Keywords: *Brassica oleracea*; *Brassica rapa*; *Brassica napus*; bHLH gene family; Whole genome identification; Phylogeny; Expression analysis; Molecular characteristic analysis

Background

Brassica oleracea L. is an extremely important cruciferous vegetable worldwide that contains rich nutrients and have a morphologically abundant variation. *B. oleracea* L. generally, includes many common foods as cultivars, including cabbage, broccoli, cauliflower, kale, Brussels sprouts, collard greens, savoy, kohlrabi, and gai lan. Some varieties also have excellent ornamental properties. Many gene families have been identified in *Brassica* crops belonging to the U triangle since the completed sequencing of many species [1]. However, the basic helix–loop–helix (bHLH) gene family has not been identified in *B. oleracea and Brassica napus.*

The bHLH gene family was named from its bHLH domain. This domain is composed of 50–60 amino acids that can be divided into basic amino acids with 10–15 amino acids in the N-terminal and HLH region of approximately 40 amino acids in the C-terminal [2, 3]. The bHLH transcription factors generally function as homodimers or heterodimer and interact with the E-Box (5′–CANNTG-3′), most commonly G-Box (5′-CACGTG-3′), which is the element of the gene promoter region [4]. The bHLH transcription factors exist widely in eukaryotes and play important roles in the regulation of development and stress response. The number of bHLH family members is extremely large, which is only next to MYB transcription factors. In the previous study, 133 bHLH genes were classified into 12 subfamilies in *Arabidopsis*; with the determination of new bHLH genes, the family was divided into 21 subfamilies [5, 6]. A total of 162 *Arabidopsis* bHLH and 167

rice bHLH genes were clustered into 25 subfamilies (A–Y) [3].

The bHLH transcription factors are also involved in many developmental processes in the plant. *LAX* (*OsbHLH164*) is expressed in the boundary between the shoot apical meristem and the region of new meristem formation and involved in the formation of all types of axillary meristems throughout the ontogeny of a rice plant [7]. Meanwhile, many bHLH genes respond to many types of stress, such as drought, salt, and cold stresses. *OsbHLH148* and *OsbHLH006* (*RERT1*) respond to drought stress through the jasmonic acid signaling pathway [8-10]. *OrbHLH2* (a bHLH gene cloned from *Oryzaru fipogon* Griff.) overexpression can enhance salt tolerance and osmotic stress resistance in *A. thaliana* transgenic plants [11]. *OsbHLH1*, which is independent of ABA, plays a transcriptional role in cold signal transduction [12]. Some bHLH genes are also involved in plant reproductive development. Silique and septum development are restricted to the basal half, and seed set is limited to the apex in the *spt-2* mutant [13]. *SPT* (*SPATULA*) and *IND* (*INDEHISCENT*) interact to mediate gynoecium and fruit development by controlling auxin distribution through cooperative binding to regulatory sequences in downstream target genes [14]. *HEC2-RNAi hec1 hec3* gynoecia lack any stigmatic development and have longer styles than the wild-type one, while pin-shaped inflorescences were observed in *HEC* overexpression lines [15]. *ETTIN* is also a negative regulator of *HEC* gene expression. *HEC* genes are possibly involved in the auxin-mediated control of gynoecium patterning [15]. Some bHLH genes regulate seed development, especially the color of seed coat. A loss-of-function mutant, that is, *rc* mutant, causes rice seed coat color to change from red to white [16]. *BrTT8* regulates seed coat pigment accumulation in *Brassica* crops by modulating the expression of the late biosynthetic genes of flavonoid [17].

Many bHLH families have been identified in various plant species to date. Ninety five members of the bHLH superfamily genome were classified into 19 subfamilies in peach [18]. A total of 113 bHLH transcription factors were found in strawberry [19]. bHLH genes were also identified in potato, grape, and peanut [20-22]. In *Brassica* spp., bHLH family was identified in *Brassica rapa*, which contains 230 bHLH transcription factors and is classified into 24 subfamilies [23]. However, the bHLH members in *B. oleracea* and *B. napus* have not been reported. In the present study, 268, 251, and 440 genes belonging to the bHLH family were identified in *B. oleracea*, *B. rapa*, and *B. napus*, respectively. Gene and protein sequences were evaluated and compared by analyzing the protein sequences, evolution of the gene family, chromosomal localization, and structure. The expression characteristics of bHLH family genes have been analyzed in different tissues and floral developmental stages to determine their function. These results will provide some useful clues for further studies on bHLH family in *Brassica* crops.

Results

Identification and phylogenetic tree analysis of bHLH genes

According to the BLAST results and domain verification, we obtained 268 bHLH genes in *B. oleracea* (Table S1), 440 genes in *B. napus* (Table S2 and S3), and 251 bHLH members in *B. rapa*, while a previous study reported 230 identified bHLH genes in *B. rapa* (Table S4) [24]. The newly identified genes in *B. rapa* and their names are listed in Table S4. All bHLH genes were named based on their gene ID (Table S1, S2, S3 and S4). To divide all bHLH genes from *B. oleracea*, *B. rapa*, and *B. napus* into different subfamilies, we constructed NJ phylogenetic trees by using the domain sequences of bHLH genes from *B. oleracea*, *B. rapa*, and *B. napus*, as shown in Fig. S1–S4. The analysis of *B. napus* was divided into AA and CC genomes.

Analysis of conserved short amino acid sequence and structure

We analyzed the conserved motifs of all bHLH members from *B. oleracea*, *B. rapa*, and *B. napus* by using MEME; searched 15 conserved motifs of these amino acids sequences; and integrated the results according to the NJ phylogenetic trees, as shown in Fig. S5–S8. The analysis of *B. napus* was divided into the AA and CC genomes. Then, we divided all bHLH genes into different subfamilies according the phylogenetic trees and conserved motifs. Finally, the bHLH genes from *B. oleracea* and *B. rapa*, *B. napus* AA genome, and *B. napus* CC genome were classified into 25, 26, 23, and 22 subfamilies, respectively. *B. oleracea* had the same subfamilies, and *B. rapa* was divided into Ib (1) and Ib (2) in *B. rapa* except subfamily Ib. Subfamilies X and XIV had no members in *B. napus* AA genome, while subfamilies X, XIV, and IIIb had no members in *B. napus* CC genome. In *B. oleracea*, the largest subfamily was Ib, which had 32 members, and the smallest was orphan, which had one member. Subfamily XII was the largest one in *B. rapa* and *B. napus*, while subfamilies IVd and XIV were the smallest subfamilies in *B. rapa.* Subfamilies IVd, IIIf, and orphans were the smallest subfamilies in the *B. napus*, which had one and six members, respectively.

For the conserved short amino acid sequence analysis, two motifs (motif 1 and motif 2) were highly conserved throughout all bHLH genes. Genes on near evolutionary branches had highly conserved motifs in *B. oleracea*, *B. rapa*, and *B. napus*. Highly conserved motifs were found in the same subfamily among the three *Brassica* crops. For example, subfamily III (d+e) had the same conserved motifs (motif 1, 2, 4, 5, 8, and 12) in the three *Brassica* crops. Results indicated that motifs 1 and 2 were the most conserved motifs in all bHLH genes, and the motifs in the same subfamily from different *Brassica* crops were highly conserved.

To determine whether the number of exons and introns in different subfamilies is conserved, we analyze

the intron–exon location of all bHLH members from three *Brassica* crops and integrated the results according to the NJ phylogenetic trees, as shown in Fig. S9–S12. The analysis of *B. napus* was divided into the AA and CC genomes. In *B. oleracea*, most members in some subfamilies had the similar gene structure. For instance, most genes in subfamilies VIIIb and III(d+e) had one exon and no intron. However, the numbers of some other subfamilies were diverse. For instance, subfamily Ib was the largest subfamily. In this subfamily, 16 members had 3 exons and 2 introns, 13 of which had 2 exons and 1 introns; one gene had 1 exon and 0 intron, *BolbHLH089* had four exons and three introns, and *BolbHLH255* have six exons and five introns. *B. rapa* and *B. napus* had similar situations in which some subfamilies were conserved, while some were diverse. For the conserved subfamily (III(d+e)) in different *Brassica* crops, the gene structures of most members were conserved. In contrast to the conservatism in short amino acid sequence, the gene structures in the same subfamilies were not highly conserved.

Molecular characteristics analysis and chromosomal localization of bHLH genes in *B. oleracea, B. rapa*, and *B. napus*

Subsequently, we analyzed the molecular characteristics of bHLH genes in *B. oleracea*, *B. rapa*, and *B. napus* (Table S1, S2, S3 and S4). In *B. oleracea*, the ORF length of all bHLH genes ranged from 273 bp (*BolbHLH143*) to 2985 bp (*BolbHLH063* and *BolbHLH114*), and the encoded polypeptides had 91 (BolbHLH143) to 994 amino acids (BolbHLH063 and BolbHLH114). The gene with the largest molecular weight was BolbHLH063 (112.881 kDa), while the minimum molecular weight of all genes was 10.253 kDa (BolbHLH143). The theoretical isoelectric point (pI) ranged from 4.49 (BolbHLH072) to 10.26 (BolbHLH250). The PIs of 174 genes were <7 (64.9%), thereby indicating that these encoded proteins were acidic proteins (Table S1). In *B. rapa*, the ORF length of all bHLH genes ranged from 255b bp (*BrabHLH067*) to 3606 bp (*BrabHLH198*), and the maximum to minimum molecular weight was 9.799 (BrabHLH067) to 132.383 (BrabHLH198). Among all BrabHLH genes, the pI of 156 genes (62.2%) was <7, thereby indicating that these genes were acidic proteins. In *B. napus*, the ORF length of all bHLH genes ranged from 243 bp (*BnabHLH020*) to 4320 bp (*BnabHLH017*). The pI of 295 genes (67.0%) was <7, including 141 genes in the AA genome (65.3%) and 154 genes in CC genome (68.8%).

In *B. rapa*, 247 bHLH genes were distributed on 10 chromosomes (i.e., A01–A10) (Fig. 1A). For the 21 newly identified genes, *BrabHLH237*, *BrabHLH242*, and *BrabHLH251* were distributed on A01 chromosomes; *BrabHLH234* and *BrabHLH248* were distributed on A02; and *BrabHLH*231 and *BrabHLH*243 were distributed on A03 and A06. Two genes were located on A04 and A10, three genes were distributed on

A05 and A07, and four members were mapped on A09. A total of 222 BolbHLH genes were distributed in 9 chromosomes (i.e., C01–C09) (Fig. 1B). The location of all BolbHLH genes on the chromosome can be seen in Table S1. The number of bHLH genes distributed on C06 was the least, with only 16 genes, while 35 genes were distributed on C04. A total of 33 genes were distributed on the C03 chromosome, 28 on C08, 25 on C01 and C07, 21 on C02 and C09, and 18 on C05. Finally, 46 BolbHLH genes were distributed on different scaffolds. We also mapped the bHLH genes from *the* AA and CC genomes of *B. napus* (Fig. 2). For *B. napus* AA genome, 190 bHLH genes were mapped on 10 chromosomes (A01–A10). A07 had the most members (32 genes), while A08 and A10 had the minimum number of bHLH genes (11 genes). For the CC genome, 188 bHLH genes were distributed on 9 chromosomes (i.e., C01–C09). C03 and C04 had the maximum number of bHLH genes (i.e., 35 genes), and the number of bHLH genes distributed on C08 was the least, with only 13 members.

***Ka* and K_S calculation of orthologous bHLH genes between *A. thaliana* and *Brassica* crops**

First, we searched orthologous bHLH genes between *A. thaliana* and three *Brassica* crops. Then, we obtained 261 pairs of orthologous genes between *A. thaliana* and *B. oleracea* (Table S5), 252 pairs between *A. thaliana* and *B. rapa* (Table S6), and 278 pairs between *A. thaliana* and *B. napus* (Table S7). We also obtained the paralogous gene pairs of bHLH genes in *B. oleracea*. A total of 158 pairs of paralogous gene pairs were listed with different duplication modes (Table S8). Then, we calculated the *Ka*, *Ks*, and their ratio *Ka/Ks* for these bHLH gene pairs (Table S5, S6, S7, and S8). The results showed that most *Ka/Kss* values of the orthologous bHLH genes between *A. thaliana* and *Brassica* crops were <1, thereby indicating that the orthologous genes suffered from strong purifying selection for retention. A total of 12, 7, and 9 pairs of genes had the *Ka/Ks* value of >1 in *B. oleracea*, *B. rapa*, and *B. napus*, which suffered from positive Darwinian selection. For paralogous gene pairs, most *Ka/Ks* values were <1 except for eight gene pairs. For *B. oleracea,* we linearized 211 pairs of ortholog genes distributed on the chromosomes by using the Circos program (Fig. 3). For the bHLH gene family, chromosome Chr1 had additional orthologous genes with chromosomes C05 and C06 in *B. oleracea*. Chr2 and C04, Chr4 and C01, Chr5 and C02, and C03 had additional orthologous genes. The orthologous bHLH genes on Chr3 were dispersed on *B. oleracea* chromosomes.

The *Ks* values orthologs genes can be used to calculate the divergence time between *B. oleracea* and *A. thaliana*. We first counted the distribution frequency of the *Ks* value and then calculated the divergence time on the basis of the neutral substitution rate of 1.5×10^{-8} substitutions per site per year for *Chs* [24]. In the present study, the *Ks* values had a concentrated location between 0.3 and 0.5 in *B. oleracea*, *B. rapa*, and *B. napus* (Fig.

4). This result indicated that the divergence time of the bHLH gene family between the three *Brassica* crops and *A. thaliana* was approximately in the range of 10–18 MYA, which coincided well with the result of a previous study (13–17 MYA) [25].

The retention rates of the identified BrabHLH and BolbHLH genes were 51.6% and 55.1%, respectively (251/486 and 268/486). We calculated the retention rate of the AtbHLH orthologous bHLH, core, and random genes on different subgenomes of varying species (Fig. 5). All genes had the highest retention rate in the LF subgenome. In the LF and MF1 subgenomes of *B. rapa*, AtbHLH orthologous bHLH genes had much higher retention rate than the core and random genes but had lower retention rate than the core genes in *B. napus* AA genome (Fig. 5 A, C). The same trend was observed between *B. oleracea* and *B. napus* CC genome (Fig. 5 B and D). In the MF2 subgenome, core genes had the highest retention rate in the three species. We also calculated the retention rates of Arabidopsis bHLH genes and determined the core and random genes in *B. rapa*, *B. oleracea*, and *B. napus* (AA and CC genome). The retention rates of the AtbHLH orthologous genes in *B. rapa* and *B. oleracea* were higher than those in *B. napus* (AA and CC genome) at 56% and 57%, and higher than those of the core and random genes (Fig. 5E), respectively.

We studied the bHLH gene loss, gain, and retention in *B. napus* after tetraploid (Table 1). In theory, the number of bHLH genes in *B. napus* is the sum of bHLH genes in *B. rapa* and *B. oleracea.* In fact, the number was lower than the theoretical value due to the loss of genes. A total of 182 genes were lost, of which more genes were lost on the CC genome than in the AA genome. Some new genes were obtained in the sites without genes in *B. rapa* and *B. oleracea* genome. However, the gain of the genes was extremely rare. Only 13 genes, including 7 genes, were obtained in the MF2 sub genome. Some genes did not belong to the bHLH gene family, although they were the paralogous genes of a bHLH gene in *B. rapa* and *B. oleracea*. A total of 40 domain loss genes in *B. napus* were found in this study. Some identified genes did not have information about their distribution in the subgenomes, which we count as no hits. In this case, two no hits genes were found in *B. rapa*, 16 in *B. oleracea,* and 148 in *B. napus*. Hence, the most frequent occurrence of species in polyploidy is gene loss.

Signal peptide and subcellular localization prediction of bHLH family of proteins

We predicted the signal peptides of 268, 251, and 440 bHLH proteins in *B. oleracea*, *B. rapa*, and *B. napu*, respectively. Only one and two members had signal peptides in *B. oleracea* and *B. rapa*, which were BolbHLH128 (Bol021805), BrabHLH084 (Bra014653), and BrabHLH168 (Bra029354, Fig. S13). According to the C, S, and Y values, the site near serine may be a potential signal peptide shear site.

To determine the location of all bHLH proteins in the cell, we performed subcellular localization prediction by using WoLF PSORT. In *B. rapa*, 223 genes had a high possibility of being located in the nucleus. Some genes were also likely to be located in the cytoplasm, some in peroxisome, and some in mitochondria, vacuole, chloroplast, Golgi apparatus, and endoplasmic reticulum. For *B. oleracea*, the results showed that the subcellular localizations of most members of the bHLH family proteins were likely to be in the nucleus except for 21 proteins (Table 2). BolbHLH019, BolbHLH024, BolbHLH044, BolbHLH068, BolbHLH077, BolbHLH112, BolbHLH128, BolbHLH203, and BolbHLH230 were located in the chloroplast; BolbHLH188, BolbHLH199, BolbHLH210 and BolbHLH239 were located in cytoplasm; BolbHLH071 and BolbHLH141 were located in the peroxisome; and BolbHLH255 was located in the plasma membrane. In *B. napus*, 409 genes were located in the nucleus with the highest locational coefficient. Among the 31 remaining genes, 6 were located in the chloroplast, 6 in the cytoplasm, 11 in the mitochondria, 4 in the peroxisome, and 2 in the Golgi apparatus. BnaHLH230 and BnaHLH370 were likely to be located in the plasma membrane and cytoskeleton, respectively.

Gene function annotation of bHLH proteins in *B. oleracea*

We performed BLAST and GO annotation of 268 bHLH proteins with Blast2go (Fig. 6). GO annotation have three parts, namely, molecular function, cellular component, and biological process. GO annotation (BLAST and InterProScan) involves two methods. The first method was utilized in this study, and one member cannot blast to the eclectic database (BolbHLH139). Finally, 267 members obtained their own function annotation. In the molecular function, most of the gene annotations were focused on DNA-binding transcription factor, DNA-binding, and protein dimerization activities. In the biological process, most annotations were concentrated in the cellular N compound metabolic, biosynthetic, and cellular processes. In the cellular component, most gene annotations centered on nucleus, and this result was consistent with the subcellular localization prediction (Fig. 6 and Table 1).

Gene expression analysis of bHLH gene family in *B. oleracea, B. rapa, and B. napus*

We analyzed the temporal and spatial expression patterns of 50 *BrabHLH*, 50 *BolbHLH*, and 65 *BnabHLH* genes in different tissues or organs. The primers used for qRT-PCR in B. *oleracea*, *B. rapa*, and *B. napus* are shown in Table S9, S10, and S11, respectively. All the relative expression results underwent logarithmic transformation. In *B. oleracea*, the expression patterns of the bHLH gene family were diverse. Some genes were highly expressed in vegetative tissues, while others were highly expressed in reproductive tissues (Fig. 7). For example, BolbHLH053 had extremely high expression in the root and almost no expression in the pods and

buds. By contrast, BolbHLH239 had a low expression in the root, stem, and leaf and high expression in the flower, middle buds, and small buds. The expression level decreased again with reproductive organ development.

In studying whether the expression patterns of genes changed after the formation of tetraploid AACC genomes from the AA and CC genomes, these selected genes were orthologous genes to each other. The heat map showed that the expression patterns of bHLH genes in different species were diverse. However, some similarity or complementarity was observed. In *B. rapa* and *B. napus,* bHLH genes almost had no expression in the root. However, their expression patterns in the stems, leaves, and reproductive organs were complementary. Most genes had high expression in the stem, flower, silique, and floral buds and low expression in *B. napus* leaves, while BrabHLH genes had high expression in the leaves and low expression in the other tissues. Most genes were highly expressed in *B. oleracea* roots, which was the opposite of that in *B. rapa* and *B. napus*. Analyzing the temporal and spatial expression patterns of orthologous genes showed insignificant similarity or complementarity between them (Fig. 8). For the newly identified BrabHLH genes, we analyzed the expression patterns of the two newly identified BrabHLH genes (*BrabHLH238* and *BrabHLH242* belonged to subfamily X) and found that they had the opposite expression pattern. *BrabHLH238* had no expression in the root and high expression in the stem, leaf, flower, silique, and flower buds. Meanwhile, *BrabHLH242* had a slightly high expression in the root and low expression in other tissues.

Discussion

Brassica plant is an important economic crop in the world, and *B. oleracea* is a widely known species. With the rapid development of bioinformatics analysis, many genome sequences of *Brassica* plants have been completed at present [1, 26, 27]. Several gene families have been identified in Brassica to date [28-30]. After the discovery of bHLH motif with DNA binding and dimerization, increasing bHLH protein super families have been identified in plants and animals [31]. In animals, bHLH proteins are divided into six main group (groups A to F) according to their phylogenetic relationship, motifs, and functions [32]. bHLH gene family is a large family that have been identified in many species [24, 33-34]. A total of 230 bHLH genes have been identified in *B. rapa*, and the expression patterns of some genes with different treatments have also been analyzed [24]. In the present study, we identified 268 bHLH genes in *B. oleracea*, 440 genes in *B. napus*, and 21 new bHLH genes in *B. rapa*. These bHLH genes were classified into 25, 23, and 22 subfamilies in *B. oleracea*, *B. rapa*, and *B. napu*, respectively [24, 36]. The molecular characteristics were analyzed, and the

results showed that most bHLH proteins were acidic proteins in *Brassica* crops. The total lengths of the gene coding regions were in the ranges of 273–2985 bp in *B. oleracea*, 255–3606 bp in *B. rapa*, and 243–4320 bp in *B. napus*. The molecular weights were in the ranges of 10–113 kDa in *B. oleracea*, 9.799–132.383 kDa in *B. rapa*, and 9.178–165.835 kDa in *B. napus*, which indicated that all the members of the bHLH family differed considerably in terms of sequence length. In the three *Brassica* crops, more than 60% proteins were acidic.

A. thaliana has experienced three whole genome duplication (WGD) events, as follows: a γ event shared with most dicots and two subsequent genome duplications (α and β) shared with other members of the order Brassicales [35]. *B. oleracea* has a common ancestor with *A. thaliana* and also underwent WGD events. At approximately 13–17 MYA, *B. oleracea* experienced a whole genome triplication (WGT) event, thereby resulting in the divergence of the genome between *B. oleracea* and *A. thaliana* [36, 37]. In this study, the divergence time of the bHLH gene family between *B. oleracea* and *A. thaliana* was approximately 10–18 MYA, which was consistent with the divergence times of *B. oleracea* and *A. thaliana*. To understand the selection pressure on bHLH genes in *B. oleracea*, *B. rapa*, and *B. napu* during evolution, we calculated the *ka*/*ks* value, and most homologous gene pairs experienced purifying selection, thereby indicating that these genes were strongly controlled in evolution. Only 12, 7, and 9 orthologous gene pairs and 8 paralogous gene pairs had a *ka*/*ks* ratio of >1 (Table S5, S6, S7, and S8), thereby suggesting that novel functions were likely to generate among these genes. Similar findings were observed in other gene families in plants, such as the TCS of tomato and PMEI of Brassica, of which most homolog pairs evolve through purifying selection, and a few or even no gene experiences positive selection [38, 39]. Theoretically, after WGD and WGT events, the genome size of *B. oleracea* should be threefold larger than that of *A. thaliana*. However, the size was much smaller than the theoretical value. The retention rate of AtbHLH orthologous genes was 57% in *B. oleracea*, which was close to its retention rate (56%) in *B. rapa*, while the actual retention rates in *B. rapa* and *B. oleracea* were 51.6% and 55.1%. The retention rate of BolbHLH was close to that of the core and random genes and slightly higher than that of the other gene families, such as the PMEI (52%) [39] and SDG gene families (43%) in *B. rapa* [40]. These retentions indicated that almost half of the genes were lost after WGT event. In *B. napus*, the retention rates of AtbHLH orthologous, core, and random genes were low. This finding suggested that some genes have been lost during the formation of allotetraploid. Genome replication and tandem repetition were the main driving forces of genome expansion. Studies in *A. thaliana*, rice, and poplar showed that genome replication is the only driving force for the expansion of the SDG family [41]. However, a study of the SDG family of *B. rapa* showed that the expansion of the SDG family depends not only on genomic replication but also on

tandem repeat events [40]. In our study, tandem repeat events occurred in 14 loci. This finding implied that the expansion of the bHLH family depends upon genomic replication and tandem repeat events in *B. oleracea.*

bHLH gene family is an extremely large family that is involved in many regulation processes, such as stress response [42] and seed development [17, 43]. To understand the possible gene function of bHLH genes in *B. oleracea* further, we selected 50 genes distributed in different subfamilies for spatiotemporal expression pattern analysis. The heat map showed that the expression patterns of 50 BolbHLH genes were diverse. Although the studies on the function of bHLH genes in *B. oleracea* were lacking, its gene function can be deduced through its expression pattern and the gene function of *A. thaliana* bHLH genes in the same subfamilies. *INDEHISCENT*, *HECATE*, and *SPATULA* are involved in the pistil development, and they belong to the VIIIb and I (a+b) subfamily, respectively [14, 44]. Our analysis of the expression patterns of BolbHLH genes belonging to the same subfamilies showed that these genes had relatively high expression in the reproductive organs. Thus, our research on the expression patterns of BolbHLH genes can guide the study of their functions. Two newly identified genes (i.e., *BrabHLH238* and *BrabHLH242* belonging to subfamily X) had the opposite expression pattern. After WGT, a series of events, such as chromosome rearrangement, gene loss, and epigenetic modification, often emerges [45, 46]. The remained genes after the loss events are often associated with dosage effects or with new or subfunctionalized genes in *A. thaliana* [47, 48]. Homologous genes with similar expression patterns are preserved because of the dosage effect, while homologous genes with different expression patterns are retained due to new functionalization and subfunctionalization [49]. In the present study, we selected 50 BolbHLH genes, among which *BolbHLH098*/*BolbHLH155*, *BolbHLH154*/*BolbHLH226*, *BolbHLH141*/*BolbHLH090*, *BolbHLH133*/*BolbHLH159*, and *BolbHLH198*/*BolbHLH167* were homologous genes, and 4 gene pairs had similar expression pattern, except *BolbHLH141*/*BolbHLH090*. According to the heat map, most BolbHLH genes had similar expression patterns (Fig. 8). According to the results above, we speculated that the retention of bHLH family genes was mainly due to the dose effect in *B. oleracea*. In our study, bHLH genes had high expression in *B. oleracea* leaves, while they had high expression in reproduction tissues in *B. napus*. The main economic organ of *B. oleracea* was the leaf, while that of *B. napus* was the seed. The different expression patterns may be correlated with the agronomic traits they selected during the course of evolution. In some studies, researchers have found that bHLH genes are related to plant biomass. For example, OsbHLH107 overexpression can enhance the grain size in rice, and a *Vitis vinifera* bHLH transcription factor (*VvCEB1*(opt)) enhances plant cell size, vegetative biomass, and reproductive yield [50, 51].

An important regulation pathway is that transcription factors interact with *cis*-acting elements to express the genes involved in stress response and developmental processes specifically. In the present study, the functional annotation of the BolbHLH genes showed that they were mainly concentrated in the DNA-binding transcription factor, DNA binding, transcription regulator, and protein dimerization activities in the molecular function section, which was consistent with the way in which the bHLH gene functions by forming homodimers or heterodimers [31]. bHLH transcription factors can regulate many genes involved in different regulatory pathways [52]. bHLH transcription factors also participate in the regulation of metabolic processes, such as the biosynthesis of alkaloids and nicotine [53]. bHLH genes are also involved in plant response to biotic and abiotic stresses. *TcMYC* is highly expressed in the xylem and leaves and upregulated by drought and high-salinity stresses in yew trees [54]. The *CsbHLH18* of sweet orange functions in the modulation of cold tolerance by regulating the antioxidant gene [55]. Some bHLH transcription factors can also respond to salinity and Fe-deficient abiotic stress [56, 57]. In the present study, the biological process of GO annotation showed that most genes had the GO terms of cellular N compound, metabolic process, biosynthetic process, anatomical structure development, response to stress, regulation of transcription, DNA template, and reproduction. These enriched GO annotations were also consistent with the known functional bHLH transcription factors in some studies.

Conclusion

A total of 268, 251, and 440 bHLH genes are present in *B. oleracea, B. rapa*, and *B. napus*, respectively. The short motifs of the bHLH genes in the same subfamilies of different *Brassica* crops are highly conserved, while the gene structures are diverse in some subfamilies. Most *Ka/Ks* values are <1, which indicates that these genes suffered from strong purifying selection. The divergence time of the bHLH gene family between the three *Brassica* crops and *A. thaliana* is approximately in the range of 10–18 MYA. The expression patterns between *B. rapa* and *B. napus* shows that they have the similar expression pattern in the root and opposite patterns in the stems, leaves, and reproduction tissues. The expression patterns of *B. oleracea* and *B. napus* are the opposite.

Methods

Identification of bHLH family genes in *B. oleracea*, *B. rapa*, and *B. napus*

To identify bHLH family genes in *B. oleracea* and *B. napus*, we searched the protein sequences of the 162 reported *Arabidopsis* bHLH from TAIR (http://www.arabidopsis.org/) and downloaded the protein sequences of 230 bHLH genes that were identified in *B. rapa* from the *Brassica* Database (http://brassicadb.org/brad/index.php) according to previous studies [5, 23, 33, 58]. We obtained the candidate genes of bHLH in *B. oleracea* and *B. napus* by analyzing the amino acid sequences of *Arabidopsis thaliana* and *B. rapa* bHLH family members with BLASTP in the *Brassica* Database. We also searched all the orthologous genes of 162 *A. thaliana* bHLH genes. Then, we downloaded the amino acid sequences of all candidate genes, and the Pfam database (E value = 1.0, http://pfam.xfam.org/) was used to determine whether each sequence harbored the conserved domains (HLH, PF00010.25). Genes that did not contain the known conserved domains of the gene families were excluded from further analysis. All amino acid sequences were subjected to gene ontology (GO) annotation by using Blast2GO with default parameters [59].

Chromosomal localization of bHLH family genes in *B. oleracea*, *B. rapa*, and *B. napus*

We searched the start and stop locations on the *B. oleracea*, *B. rapa*, and *B. napus* chromosomes of all bHLH family members in the *Brassica* database. Considering the numerous members in *B. napus*, the analysis of *B. napus* was divided into AA and CC genomes. Subsequently, the chromosomal localization of these members was performed using the MapChart software on the basis of the relative location of each gene on each chromosome [60].

Gene structure and phylogenetic analyses of bHLH family genes in *B. oleracea, B. rapa*, and *B. napus*

We searched the full-length DNA and cDNA sequences of *B. oleracea*, *B. rapa*, and *B. napus* bHLH genes from the National Center for Biotechnology Information, (https://www.ncbi.nlm.nih.gov/). Then, the Gene Structure Display Server database (http://gsds.cbi.pku.edu.cn/index.php) was utilized to analyze the gene structure. The homologous sequence alignment of all bHLH genes from *B. oleracea*, *B. rapa*, and *B. napus* was performed using ClustalW [61]. Then, the sequence alignment results were considered the basis in generating the unrooted phylogenetic tree of the *B. oleracea*, *B. rapa*, and *B. napus* bHLH genes with MEGA (version 5.0) [62]. Considering the numerous members in *B. napus*, the analysis of *B. napus* was divided into the AA and CC genomes. All parameters used were default parameters. Phylogenetic trees were generated with the value of the

1000 bootstrap samples by the neighbor-joining (NJ) method [62]. Then, the results of gene structure analysis were integrated with phylogenetic trees by using Photoshop CS3.

Nonsynonymous substitution rate, synonymous substitution rate, and gene retention analysis

The nonsynonymous substitution rates (*Ka*) and synonymous substitution rates (*Ks*) of the paralogous bHLH gene pairs in *B. oleracea* and orthologous bHLH gene pairs among the three *Brassica* crops and *A. thaliana* were calculated. First, the multiple sequence alignments of the CDS sequence pairs were performed using MEGA5. Then, the alignment results were analyzed, and the bottom named Distance (Compute Pairwise Distances) was chosen. The next step was choosing a substitution model; in this step, we chose synonymous–nonsynonymous method. The options of substitution include "nonsynonymous only," which refers to *Ka*, and "synonymous only," which is *Ks*. Finally, the *Ka* and *Ks* values were obtained using MEGA5 [62]. The relationships of orthologous genes between *A. thaliana* and *B. oleracea* were also visualized using the Circos program [63].

The retention rates of *Arabidopsis* bHLH, core, and random genes and the retention rate in the different subgenomes (i.e., LF, MF1, and MF2) of *B. rapa*, *B. oleracea*, and *B. napus* (AA and CC genomes) were counted. The gene loci with tandem repeats were calculated using one gene. The 458 core genes and 459 random genes of *A. thaliana* were downloaded from the CEGMA database (http://korflab.ucdavis.edu/Datasets/cegma); the homologous genes of 917 genes in *B. rapa*, *B. oleracea*, and *B. napus* were searched in BRAD, and their retention rates were counted [64, 65]. The retention and loss of genes on the AA and CC genome in *B. napus* during tetraploid were analyzed.

Motif analysis, signal peptide, and subcellular localization prediction in *B. oleracea, B. rapa*, and *B. napus*

We obtained 268, 251, and 440 amino acid sequences of *B. oleracea, B. rapa*, and *B. napus* by searching the amino acid sequences from the Brassica database, respectively. Then, MEME was used to analyze the common conserved the short amino acid sequence of bHLH family members [66]. In this study, we searched 15 bHLH gene motifs. Phylogenetic trees were generated with the value of 1000 bootstrap resampling by the NJ method [62]. Considering the numerous bHLH members in *B. napus*, the analysis of *B. napus* was divided into the AA and CC genomes. Then, the conserved motifs were integrated with phylogenetic trees by using Photoshop CS3. SignalP-4.1 (http://www.cbs.dtu.dk/services/SignalP/) was utilized to predict the signal peptide of the bHLH genes in *B. oleracea, B. rapa*, and *B. napus*, all of which are set at default [67]. The subcellular localizations of

these bHLH genes were predicted using WolfPsort (https://wolfpsort.hgc.jp) [68].

Plant material and expression pattern analyses of bHLH genes in *B. oleracea, B. rapa*, and *B. napus*

The plant materials *(B. rapa* and *B. napus)* were planted in the experimental farm of Zhejiang University. The original source of *B. rapa* seeds came from Dr. Xiaowu Wang of the Institute of vegetables and Flowers Chinese Academy of Agricultural Sciences. The seeds of *B. napus* were provided by Dr. Jianbin Li of Jiangsu Academy of Agricultural Sciences. *B. oleracea* was grown in the experimental farm of Zhejiang Academy of Agricultural Science, and It was provided by Dr. Xinmin Zhong. In this study, Dr. Xiaolin Yu undertook the formal identification of all plant material. However, all plant materials were not deposited in a publicly available herbarium. If anyone require these seeds for study only, the corresponding author would like to send the seeds to them according to the relative agreement. Roots, stems, leaves, flowers, silique, big bud (>2.0 mm), middle buds, and small bud (<2.0 mm) in the flowering period were sampled in liquid nitrogen and stored in a refrigerator at −80 °C.

Total RNA was extracted from previous materials by using the TRIzol reagent (Invitrogen, USA) following the manufacturer's instructions. The first cDNA strand was generated following the manufacturer's protocol by using the Takara Reverse Transcription System (Japan). A total of 50 *B. oleracea*, 50 *B. rapa*, and 65 *B. napus* bHLH genes were chosen to analyze the expression patterns, and these genes were orthologous genes. The primers were designed using Integrated DNA Technologies (https://sg.idtdna.com/pages) and the Primer (version 5.0) software. All cDNA samples were adjusted to a uniform concentration. A total of 1 μL of cDNA were subjected to real-time PCR at a final volume of 15 μL, containing 7.5 μL of SYBR Green Master Mix Reagent (TOYOBO, Japan), 0.3 μL of specific primers, and 5.9 μL ddH_2O. Three technical replicates for each sample were performed in a real-time PCR machine (Bio-Rad CFX Manager). The heating program is as follows: 95 °C for 30 s, followed by 95 °C at 10 s for 40 cycles, and 57 °C at 30 s (varying with specific primers). To normalize the total amount of cDNA present in each reaction, we amplified the gene *UBC10* of *B. rapa*, *GAPDH* of *B. oleracea*, and *25S* of *B. napu* as endogenous controls in calibrating the relative expression. The $2^{-\Delta\Delta CT}$ method of the relative gene quantification recommended by Applied Biosystems (PE Applied Biosystems, USA) was used to calculate the expression level of different tissues [69].

Abbreviations

bHLH	Basic helix–loop–helix;
MYA	Millions of years ago;
SPT	*SPATULA*;
IND	*INDEHISCENT*;
HEC	*HECATE*;
CDS	Coding sequence;
ORF	Open reading frame;
WGD	Whole genome duplication;
WGT	Whole genome triplication;
TCS	Two component signaling system;
PMEI	Pectin methylesterase inhibitors;
SDG	SET domain group.

Declarations

Acknowledgements

The authors gratefully acknowledge Dr. Longjiang Fan and Dr. Li Huang for their technological guidance and Dr. Zhenning Liu for stimulating discussions and critical reading of the manuscript. We also thank Dr. Xiaowu Wang and Dr. Jianbin Li for providing the seeds of *B. rapa* and *B. napus*. We are all grateful to Dr. Xinmin Zhong for offering the plant materials of *B. oleracea.*

Funding

This work was partially supported by the National Key Research and Development Program of China (grant no. 2016YFD0100204-31), the 948 Project of Agricultural Ministry of China (grant no. 2014-Z28), National Natural Science Foundation of China (grant nos. 31460521, 31700272 and 31872110), the Breeding Project of the Sci-tech Foundation of Zhejiang Province (grant no. 2016C02051-6-1), the Project of SRTP in Zhejiang University (grant no. 2018R401080), the Development Project of Wuxi City (grant no. CLE02N1603), and the Project of Application on Public Welfare Technology in Zhejiang Province (grant no. LGN18C150003). The funding agencies had no role in the design, analysis, and interpretation of the data or writing of the manuscript. All funders we mentioned provide financial support for our study.

Availability of data and materials

Plant materials are available under request to the respective owner institutions. The datasets supporting the results of this article are included within the article and its additional files.

Authors' contributions

XLY and LMM proposed and supervised the research and gave final approval of the version to be published. LMM and YYG performed sequence analysis. YYG, KZ, and LJK performed qRT-PCR analysis. LMM, SBY, RRL, and KWL performed statistical analyses and interpreted the experimental results. LMM and XLY wrote the manuscript.

Ethics approval and consent to participate

Not applicable. This study has not directly involved humans, animals. We comply with the Convention on the Trade in Endangered Species of Wild Fauna and Flora.

Consent for publication

Not applicable.

Competing interests

All authors declare that they have no competing interests.

References

1. Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IA *et al*. **The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes**. *Nat. commun.* 2014, **5**:3930.
2. William R. Atchley WT, Andreas Dress. **Positional dependence, cliques, and predictive motifs in the bHLH protein domain**. *J Mol Evol* 1999, **48**:501–516.
3. 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.
4. Cornelis Murre PSM, David Baltimore **A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins**. *Cell* 1989, **56**(5):777-783.

5. 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.

6. William R. Atchley WMF. **A natural classification of the basic helix-loop-helix class of transcription factors**. *P Natl Acad Sci U S A* 1997, **94**:5172-5176.

7. Komatsu K, Maekawa M, Ujiie S, Satake Y, Furutani I, Okamoto H *et al*. ***LAX* and *SPA*: major regulators of shoot branching in rice**. *Proc Natl Acad Sci U S A* 2003, **100**(20):11765-11770.

8. Seo JS, Joo J, Kim MJ, Kim YK, Nahm BH, Song SI *et al*. **OsbHLH148, a basic helix-loop-helix protein, interacts with OsJAZ proteins in a jasmonate signaling pathway leading to drought tolerance in rice**. *Plant J* 2011, **65**(6):907-921.

9. Kiribuchi K, Sugimori M, Takeda M, Otani T, Okada K, Onodera H *et al*. ***RERJ1*, a jasmonic acid-responsive gene from rice, encodes a basic helix-loop-helix protein**. *Biochem Biophys Res Commun* 2004, **325**(3):857-863.

10. Kiribuchi K, Jikumaru Y, Kaku H, Minami E, Hasegawa M, Kodama O *et al*. **Involvement of the basic helix-loop-helix transcription factor RERJ1 in wounding and drought stress responses in rice plants**. *Bios Biotechnol Biochem* 2005, **69**(5):1042-1044.

11. Zhou J, Li F, Wang JL, Ma Y, Chong K.Xu YY. **Basic helix-loop-helix transcription factor from wild rice (OrbHLH2) improves tolerance to salt- and osmotic stress in Arabidopsis**. *J Plant Physiol* 2009, **166**(12):1296-1306.

12. Wang YJ, Zhang ZG, He XJ, Zhou HL, Wen YX, Dai JX *et al*. **A rice transcription factor OsbHLH1 is involved in cold stress response**. *Theor Appl Genet* 2003, **107**(8):1402-1409.

13. Marcus G. B. Heisler AA, Yasmin H. Bylstra, Rebecca Walsh and David R. Smyth*. ***SPATULA*, a gene that controls development of carpel margin tissues in *Arabidopsis*, encodes a bHLH protein**. *Development* 2001, **128**:1089-1098.

14. 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.

15. Gremski K, Ditta G.Yanofsky MF. **The *HECATE* genes regulate female reproductive tract development in *Arabidopsis thaliana***. *Development* 2007, **134**(20):3593-3601.

16. Sweeney MT, Thomson MJ, Pfeil BE.McCouch S. **Caught red-handed: *Rc* encodes a basic**

helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* 2006, **18**(2):283-294.

17. Li X, Chen L, Hong M, Zhang Y, Zu F, Wen J *et al*. **A large insertion in bHLH transcription factor *BrTT8* resulting in yellow seed coat in *Brassica rapa***. *PLoS One* 2012, **7**(9):e44145.

18. Zhang C, Feng R, Ma R, Shen Z, Cai Z, Song Z *et al*. **Genome-wide analysis of basic helix-loop-helix superfamily members in peach**. *PLoS One* 2018, **13**(4):e0195974.

19. Zhao F, Li G, Hu P, Zhao X, Li L, Wei W *et al*. **Identification of basic helix-loop-helix transcription factors reveals candidate genes involved in anthocyanin biosynthesis from the strawberry white-flesh mutant**. *Sci Rep* 2018, **8**(1):2721.

20. Wang R, Zhao P, Kong N, Lu R, Pei Y, Huang C *et al*. **Genome-wide identification and characterization of the potato bHLH transcription factor family**. *Genes* 2018, **9**(1).

21. Wang P, Su L, Gao H, Jiang X, Wu X, Li Y *et al*. **Genome-wide characterization of bHLH genes in grape and analysis of their potential relevance to abiotic stress tolerance and secondary metabolite biosynthesis**. *Front Plant Sci* 2018, **9**:64.

22. Gao C, Sun J, Wang C, Dong Y, Xiao S, Wang X *et al*. **Genome-wide analysis of basic helix-loop-helix gene family in peanut and assessment of its roles in pod development**. *PLoS One* 2017, **12**(7):e0181843.

23. 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.

24. Koch MA, Haubold B.Mitchell-Olds T. **Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae)**. *Mol Biol Evol* 2000, **17**(10):1483-1498.

25. Mun JH, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA *et al*. **Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication**. *Genome Biol* 2009, **10**(10).

26. Wang XW, Wang HZ, Wang J, Sun RF, Wu J, Liu SY *et al*. **The genome of the mesopolyploid crop species *Brassica rapa***. *Nat Genet* 2011, **43**(10):1035-U1157.

27. Yang JH, Liu DY, Wang XW, Ji CM, Cheng F, Liu BN *et al*. **The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection**. *Nat Genet* 2016, **48**(10):1225-1232.

28. Liu Z, Lv Y, Zhang M, Liu Y, Kong L, Zou M *et al*. **Identification, expression, and comparative genomic analysis of the *IPT* and *CKX* gene families in Chinese cabbage (*Brassica rapa* ssp *pekinensis*)**. *BMC Genomics* 2013, **14:594**.

29. Miao LM, Lv YX, Kong LJ, Chen QZ, Chen CQ, Li J *et al*. **Genome-wide identification, phylogeny, evolution, and expression patterns of MtN3/saliva/SWEET genes and functional analysis of *BcNS* in *Brassica rapa***. *BMC Genomics* 2018, **19:174**.

30. Kong LJ, Zhao K, Gao YY, Miao LM, Chen CQ, Deng H *et al*. **Comparative analysis of *cytokinin response factors* in *Brassica* diploids and amphidiploids and insights into the evolution of *Brassica* species**. *BMC Genomics* 2018, **19:728**.

31. 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.

32. Atchley WR.Fitch WM. **A natural classification of the basic helix-loop-helix class of transcription factors**. *Proc Natl Acad Sci U S A* 1997, **94**(10):5172-5176.

33. Toledo-Ortiz G, Huq E.Quail PH. **The Arabidopsis basic helix-loop-helix transcription factor family**. *Plant Cell* 2003, **15**(8):1749-1770.

34. Sun H, Fan HJ.Ling HQ. **Genome-wide identification and characterization of the bHLH gene family in tomato**. *BMC Genomics* 2015, **16:9**.

35. Bowers JE, Chapman BA, Rong JK.Paterson AH. **Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events**. *Nature* 2003, **422**(6930):433-438.

36. Town CD, Cheung F, Maiti R, Crabtree J, Haas BJ, Wortman JR *et al*. **Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy**. *Plant Cell* 2006, **18**(6):1348-1359.

37. Yang YW, Lai KN, Tai PY.Li WH. **Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages**. *J Mol Evol* 1999, **48**(5):597-604.

38. He YJ, Liu X, Ye L, Pan CT, Chen LF, Zou T *et al*. **Genome-wide identification and expression analysis of two-component system genes in tomato**. *Int J Mol Sci* 2016, **17**(8):1204.

39. Liu TT, Yu H, Xiong XP, Yue XY, Yu YJ, Huang L *et al*. **Genome-wide identification, molecular evolution, and expression profiling analysis of pectin methylesterase inhibitor genes in *Brassica campestris* ssp *chinensis***. *Int J Mol Sci* 2018, **19**(5):1338.

40. Dong H, Liu DD, Han TY, Zhao YX, Sun J, Lin S *et al*. **Diversification and evolution of the SDG gene family in *Brassica rapa* after the whole genome triplication**. *Sci Rep* 2015, **5:16851**.
41. Yi-Fei ZL-SMC-RJQW. **Genome-wide identification, classification and expression analyses of SET domain gene family in Arabidopsis and rice**. *Hereditas (Beijing)* 2009, **31**(2):186-198.
42. Mao K, Dong QL, Li C, Liu CH.Ma FW. **Genome wide identification and characterization of apple bHLH transcription factors and expression analysis in response to drought and salt stress**. *Frontiers in Plant Science* 2017, **8:480**.
43. Chen HC, Cheng WH, Hong CY, Chang YS.Chang MC. **The transcription factor OsbHLH035 mediates seed germination and enables seedling recovery from salt stress through ABA-dependent and ABA-independent pathways, respectively**. *Rice* 2018, **11:50**.
44. Gremski K, Ditta G.Yanofsky MF. **The *HECATE* genes regulate female reproductive tract development in *Arabidopsis thaliana***. *Development* 2007, **134**(20):3593-3601.
45. Jackson S.Chen ZJ. **Genomic and expression plasticity of polyploidy**. *Curr Opin Plant Biol* 2010, **13**(2):153-159.
46. Hufton AL.Panopoulou G. **Polyploidy and genome restructuring: a variety of outcomes**. *Curr Opin Genet Dev* 2009, **19**(6):600-606.
47. Blanc G.Wolfe KH. **Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution**. *Plant Cell* 2004, **16**(7):1679-1691.
48. Van de Peer Y, Maere S.Meyer A. **The evolutionary significance of ancient genome duplications**. *Nat Rev Genet* 2009, **10**(10):725-732.
49. Kim J, Lee J, Choi JP, Park I, Yang K, Kim MK *et al*. **Functional innovations of three chronological mesohexaploid Brassica rapa genomes**. *Bmc Genomics* 2014, **15**.
50. Yang XM, Ren YL, Cai Y, Niu M, Feng ZM, Jing RN *et al*. **Overexpression of *OsbHLH107*, a member of the basic helix-loop-helix transcription factor family, enhances grain size in rice (*Oryza sativa* L.)**. *Rice* 2018, **11:41**.
51. Lim SD, Yim WC, Liu DG, Hu RB, Yang XH.Cushman JC. **A *Vitis vinifera* basic helix-loop-helix transcription factor enhances plant cell size, vegetative biomass and reproductive yield**. *Plant Biotechnol J* 2018, **16**(9):1595-1615.
52. Castelain M, Le Hir R.Bellini C. **The non-DNA-binding bHLH transcription factor *PRE3/bHLH135/ATBS1/TMO7* is involved in the regulation of light signaling pathway in**

Arabidopsis. *Physiol Plantarum* 2012, **145**(3):450-460.

53. Zhang HB, Bokowiec MT, Rushton PJ, Han SC.Timko MP. **Tobacco transcription factors NtMYC2a and NtMYC2b form nuclear complexes with the NtJAZ1 repressor and regulate multiple jasmonate-inducible steps in nicotine biosynthesis**. *Mol Plant* 2012, **5**(1):73-84.

54. Yang YF, Zhang KK, Yang LY, Lv X, Wu Y, Liu HW *et al*. **Identification and characterization of MYC transcription factors in *Taxus sp.*** *Gene* 2018, **675**:1-8.

55. Geng JJ.Liu JH. **The transcription factor *CsbHLH18* of sweet orange functions in modulation of cold tolerance and homeostasis of reactive oxygen species by regulating the antioxidant gene**. *J. Exp. Bot.* 2018, **69**(10):2677-2692.

56. Huang DQ.Dai WH. **Molecular characterization of the basic helix-loop-helix (bHLH) genes that are differentially expressed and induced by iron deficiency in Populus**. *Plant Cell Rep* 2015, **34**(7):1211-1224.

57. Babitha KC, Vemanna RS, Nataraja KN.Udayakumar M. **Overexpression of *EcbHLH57* transcription factor from *Eleusine coracana* L. in tobacco confers tolerance to salt, oxidative and drought stress**. *PLoS One* 2015, **10**(9):e0137098.

58. Pires N.Dolan L. **Origin and diversification of basic helix-loop-helix proteins in plants**. *Mol Biol Evol* 2010, **27**(4):862-874.

59. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ *et al*. **High-throughput functional annotation and data mining with the Blast2GO suite**. *Nucleic Acids Res* 2008, **36**(10):3420-3435.

60. Voorrips RE. **MapChart: Software for the graphical presentation of linkage maps and QTLs**. *J Hered* 2002, **93**(1):77-78.

61. Thompson JD, Higgins DG.Gibson TJ. **Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic Acids Res* 1994, **22**(22):4673-4680.

62. Saitou N.Nei M. **The Neighbor-joining method - a new method for reconstructing phylogenetic trees**. *Mol Biol Evol* 1987, **4**(4):406-425.

63. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D *et al*. **Circos: An information aesthetic for comparative genomics**. *Genome Res* 2009, **19**(9):1639-1645.

64. Parra G, Bradnam K.Korf I. **CEGMA: a pipeline to accurately annotate core genes in eukaryotic**

genornes. *Bioinformatics* 2007, **23**(9):1061-1067.

65. Cheng F, Wu J, Fang L.Wang XW. **Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species**. *Front Plant Sci* 2012, **3**.

66. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L *et al*. **MEME SUITE: tools for motif discovery and searching**. *Nucleic Acids Res* 2009, **37**:W202-W208.

67. Petersen TN, Brunak S, von Heijne G.Nielsen H. **SignalP 4.0: discriminating signal peptides from transmembrane regions**. *Nat Methods* 2011, **8**(10):785-786.

68. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ *et al*. **WoLF PSORT: protein localization predictor**. *Nucleic Acids Res* 2007, **35**:W585-W587.

69. Livak KJ.Schmittgen TD. **Analysis of relative gene expression data using real-time quantitative PCR and the 2 $^{-\Delta\Delta CT}$ method**. *Methods* 2001, **25**(4):402-408.

Legends

Fig. 1 Chromosome location of bHLH genes in *B. rapa* (A), *B. oleracea* (B). The symbols + and − indicate the gene located in sense or antisense strands, respectively.

Fig. 2 Chromosome location ofAA genome of *B. napus* (A), and CC genomes of *B. napus* (B). The symbols + and − indicate the gene located in sense or antisense strands, respectively.

Fig. 3 Nine *B. oleracea* (C01–C10) and five *Arabidopsis* chromosome (Chr1–Chr5) maps were based on orthologous pair positions and demonstrated highly conserved synteny. The curves of rose red, purple, green, orange, and reddish red link the bHLH gene on the Chr1, Chr2, Chr3, Chr4, and Chr5 chromosomes of *A. thaliana* and its orthologous genes in *B. olerace*, respectively.

Fig. 4 Ks value distribution of orthologous bHLH genes between *A. thaliana* and *B. rapa* (A), *A. thaliana* and *B. oleracea* (B), and *A. thaliana* and *B. napus* (C). The vertical axis indicates the frequency of paired sequences. The peaks of Ks value appeared between 0.4 and 0.5.

Fig. 5 Retention rates of Arabidopsis bHLH, core, and random genes in *B. oleracea*, *B. rapa*, and *B. napus*. The retention rates in the different subgenomes in (A) *B. rapa*, (B) *B. oleracea*, (C) *B. napus*, and (D) *B. napus*. (E) The retention rates of Arabidopsis bHLH, core, and random genes in the whole genome of *B. oleracea*, *B. rapa*, and *B. napus*.

Fig. 6 GO annotation of bHLH genes in *B. oleracea* with Blast2GO.

Fig. 7 Expression pattern of 50 bHLH genes from different subfamilies in varying tissues in *B. oleracea.* All values underwent logarithmic transformation. R: root, Ste: stem, L: leaves, F: flower, Si: silique, LB: large bud, MB: middle buds, and SB: small bud.

Fig. 8 Expression patterns of 50 BolbHLH, 50 BrabHLH, and 65 BnabHLH genes in different tissues. All values underwent logarithmic transformation. R: root, Ste: stem, L: leaves, F: flower, Si: silique, LB: large bud, MB: middle buds, and SB: small bud.

Fig. S1 Phylogenetic tree of *B. oleracea* bHLH genes with domain sequences. The numbers on the branches indicate the bootstrap percentage values calculated from 1000 replicates.

Fig. S2 Phylogenetic tree of *B. rapa* bHLH genes with domain sequences. The numbers on the branches indicate the bootstrap percentage values calculated from 1000 replicates.

Fig. S3 Phylogenetic tree of bHLH genes in *B. napus*. The numbers on the branches indicate the bootstrap percentage values calculated from 1000 replicates.

Fig. S4 Phylogenetic tree of bHLH genes in the CC genome of *B. napus*. The numbers on the branches indicate the bootstrap percentage values calculated from 1000 replicates.

Fig. S5 MEME analysis of conserved motifs of bHLH genes in *B. oleracea*, which was integrated with BolbHLH gene phylogenetic tree.

Fig. S6 MEME analysis of conserved motifs of bHLH genes in *B. rapa*, which was integrated with BrabHLH gene phylogenetic tree.

Fig. S7 MEME analysis of conserved motifs of bHLH genes in AA genome of *B. napus*, which was integrated with BnabHLH gene phylogenetic tree.

Fig. S8 MEME analysis of conserved motifs of bHLH genes in CC genome of *B. napus* which was integrated with BnabHLH gene phylogenetic tree.

Fig. S9 Evolutionary and exon–intron analyses of bHLH genes in *B. oleracea,* which were integrated with BolbHLH gene phylogenetic tree. Exons and introns are represented by boxes and lines, respectively.

Fig. S10 Evolutionary and exon–intron analyses of bHLH genes in *B. rapa,* which was integrated with BrabHLH gene phylogenetic tree. Exons and introns are represented by boxes and lines, respectively.

Fig. S11 Evolutionary and exon–intron analyses of bHLH genes in the AA genome of *B. napus*, which was integrated with BnabHLH gene phylogenetic tree. Exons and introns are represented by boxes and lines, respectively.

Fig. S12 Evolutionary and exon–intron analyses of bHLH genes in the CC genome of *B. napus*, which was integrated with BnabHLH gene phylogenetic tree. Exons and introns are represented by boxes and lines, respectively.

Fig. S13 Signal peptide prediction of BolbHLH128 (A), BrabHLH084 (B), and BrabHLH168 (C).

Table 1 Statistics of bHLH gene loss, gain, and retention in *B. napus* after tetraploid.

Table 2 Statistical analysis of subcellular localization with maximum localization coefficient of genes.

Table S1 Identified bHLH genes in *B. oleracea* and their molecular characteristics.

Table S2 Identified bHLH genes in the AA genome of *B. napus* and their molecular characteristic.

Table S3 Identified bHLH genes in the CC genome of *B. napus* and their molecular characteristics.

Table S4 Identified bHLH genes in *B. rapa* with 21 newly identified BrabHLH genes and their molecular characteristics. The genes with yellow background were newly identified BrabHLH genes in this study, while the others were identified in a previous study (Xiao-Ming Song et al., 2014).

Table S5 Nonsynonymous and synonymous substitution rates of orthologous bHLH genes between *B. oleracea* and *A. thaliana*.

Table S6 Nonsynonymous and synonymous substitution rates of orthologous bHLH genes between *A. thaliana* and *B. rapa*.

Table S7 Nonsynonymous and synonymous substitution rates of orthologous bHLH genes in *B. napus*.

Table S8 Nonsynonymous and synonymous substitution rates of paralogous bHLH genes in *B. oleracea*.

Note: “W” indicates whole genome duplication, and “T” indicates tandem duplication.

Table S9 Primers of selected bHLH genes for qRT-PCR in *B. oleracea*.

Table S10 Primers of selected bHLH genes for qRT-PCR in *B. rapa.*

Table S11 Primers of selected bHLH genes for qRT-PCR in *B. napus.*

Table 1 Statistics of bHLH gene loss, gain, and retention in *B. napus* after tetraploid

	LF		MF1		MF2		Sum
	BnaA	**BnaC**	**BnaA**	**BnaC**	**BnaA**	**BnaC**	
Loss	39	42	30	34	16	21	182
Gain	2	3	1	0	3	4	13
Retention	61	54	48	44	39	33	279
Domain loss	9	12	3	7	4	5	40
No hits	/	/	/	/	/	/	148

739

740 **Table 2 Statistical analysis of subcellular localization with maximum localization coefficient of genes.**

Subcellular location	***Brassica rapa***	***Brassica oleracea***	***Brassica napus***
Nucleus	223	247	409
Cytoplasm	6	4	6
Peroxisome	3	2	4
Mitochondria	7	4	11
Vacuole	2	0	0
Chloroplast	8	9	6
Golgi apparatus	1	0	2
Endoplasmic Reticulum	1	1	0
Plasma membrane	0	1	1
Cytoskeleton	0	0	1

741

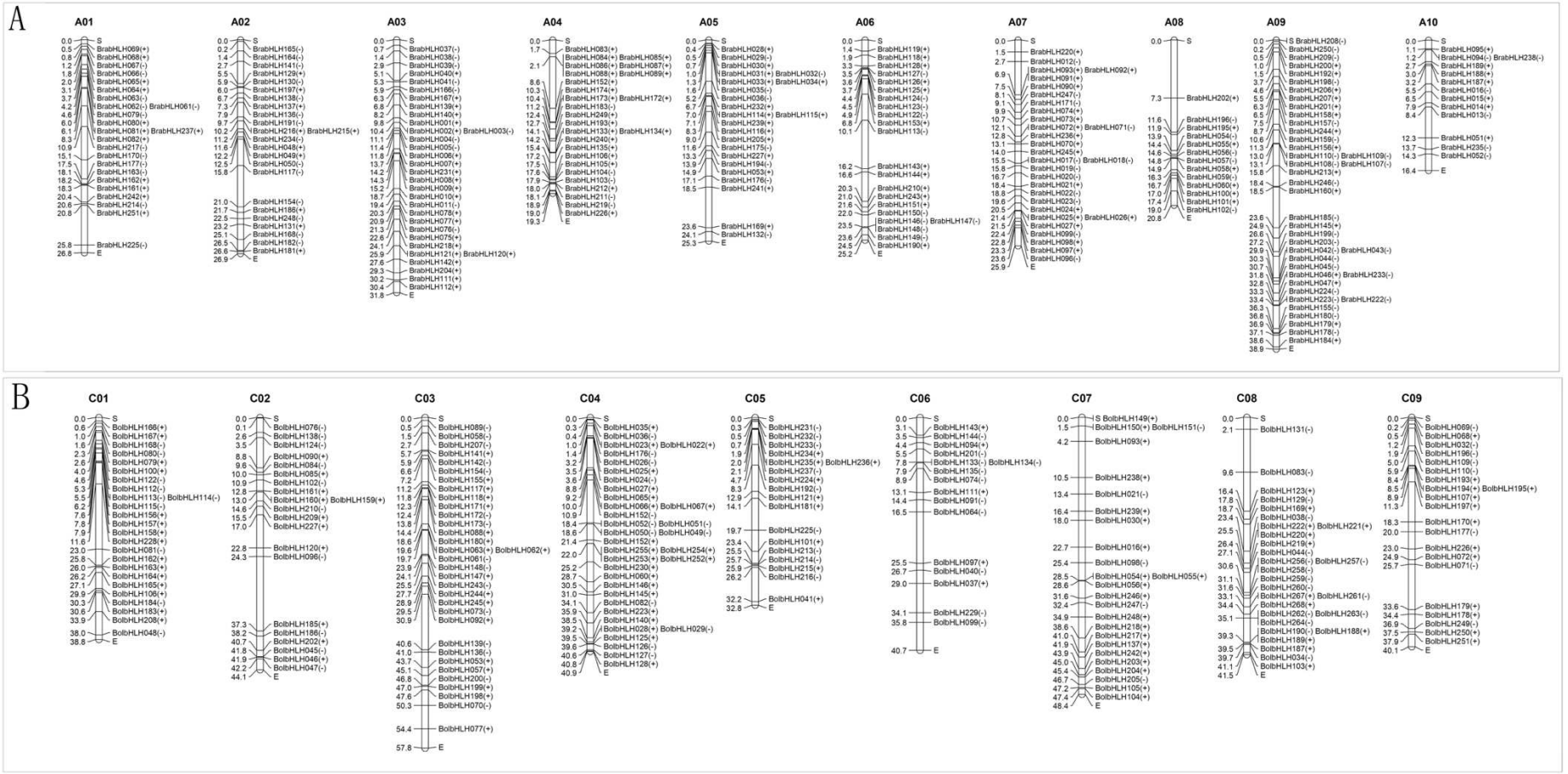


Fig. 1 Chromosome location of bHLH genes in *B. rapa* (A), *B. oleracea* (B). The symbols + and − indicate the gene located in sense or antisense strands, respectively.

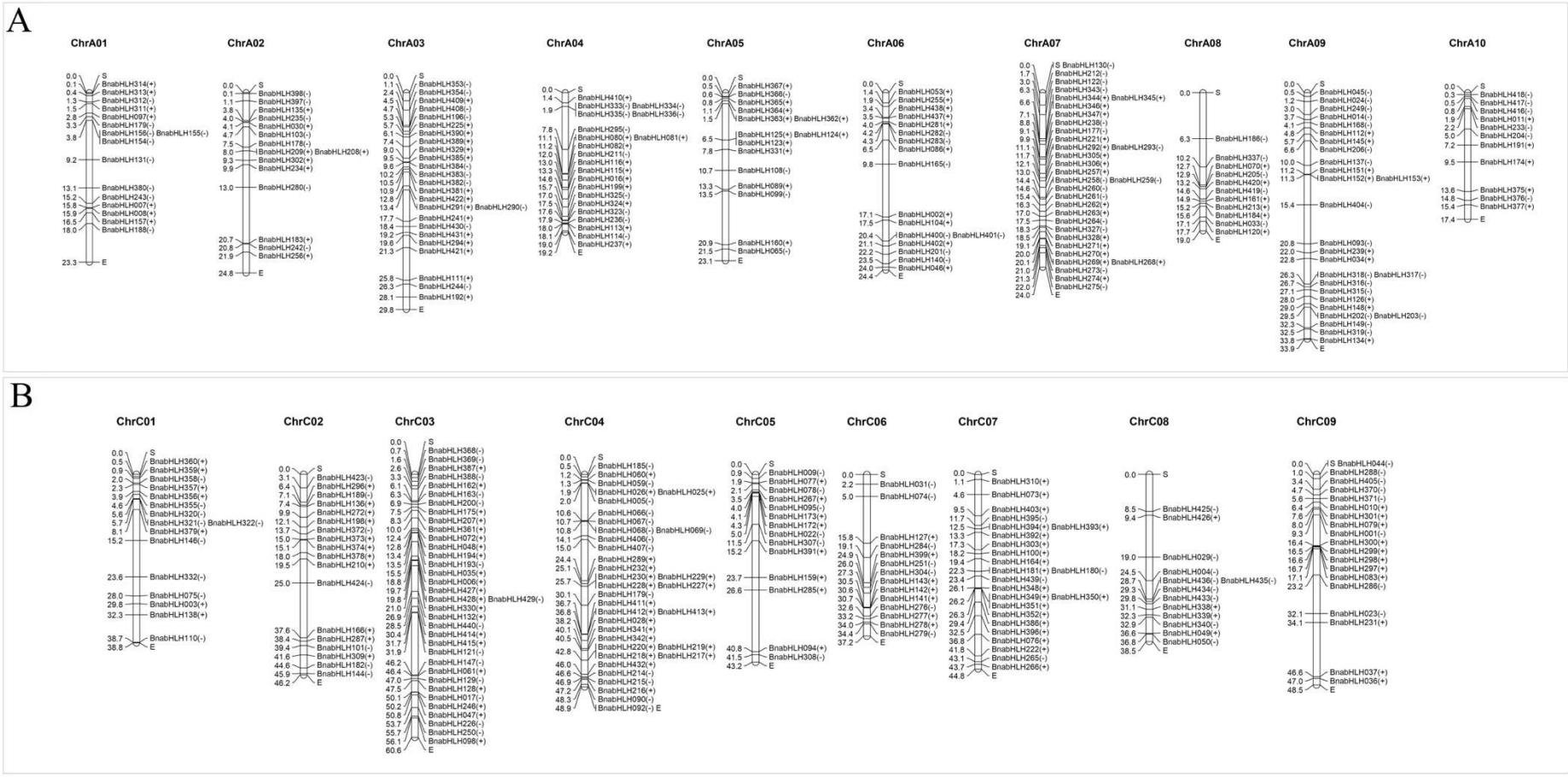


Fig. 2 Chromosome location ofAA genome of *B. napus* (A), and CC genomes of *B. napus* (B). The symbols + and − indicate the gene located in sense or antisense strands, respectively.

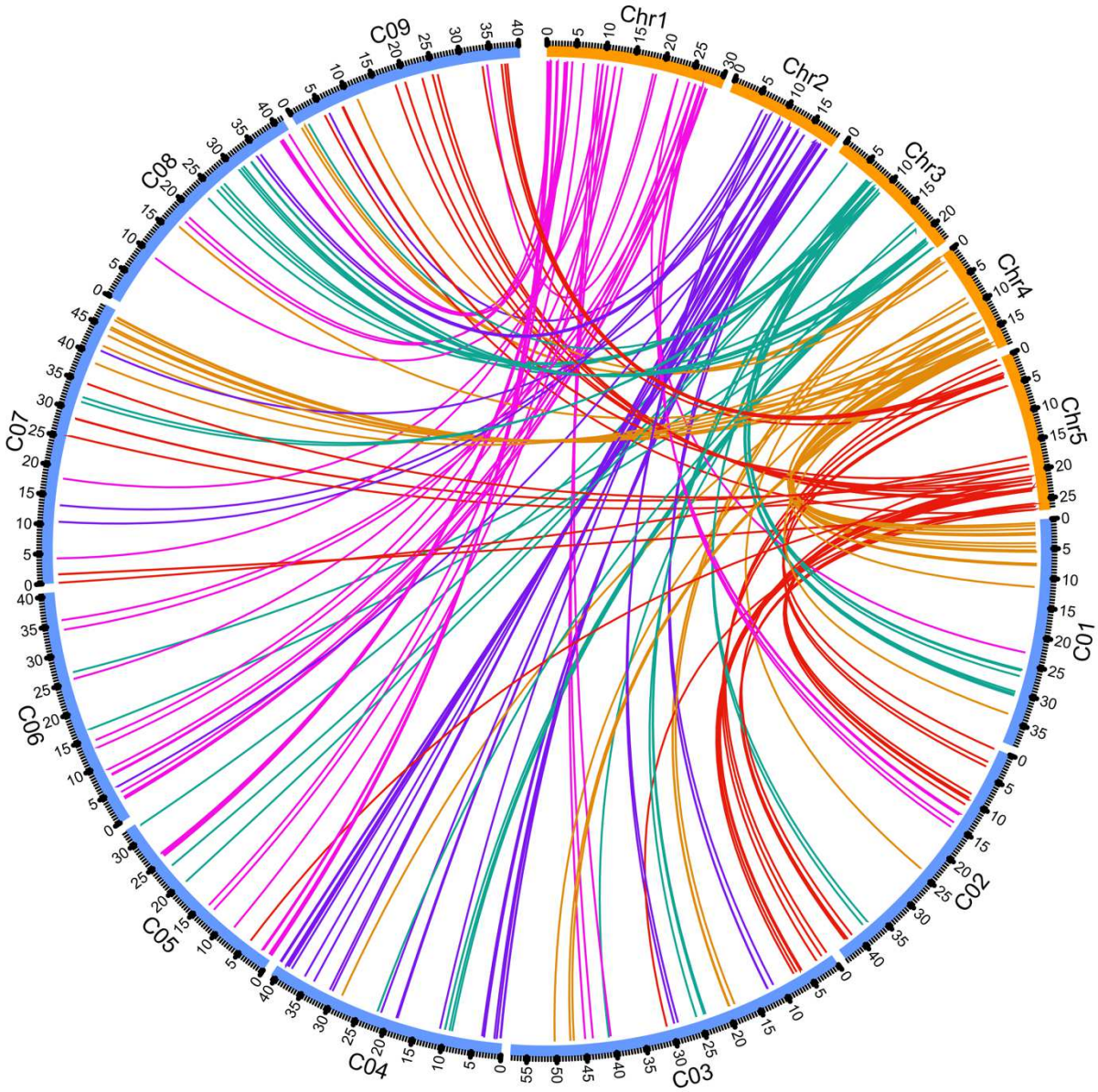


Fig. 3 Nine *B. oleracea* (C01–C10) and five *Arabidopsis* chromosome (Chr1–Chr5) maps were based on orthologous pair positions and demonstrated highly conserved synteny. The curves of rose red, purple, green, orange, and reddish red link the bHLH gene on the Chr1, Chr2, Chr3, Chr4, and Chr5 chromosomes of *A. thaliana* and its orthologous genes in *B. olerace*, respectively.

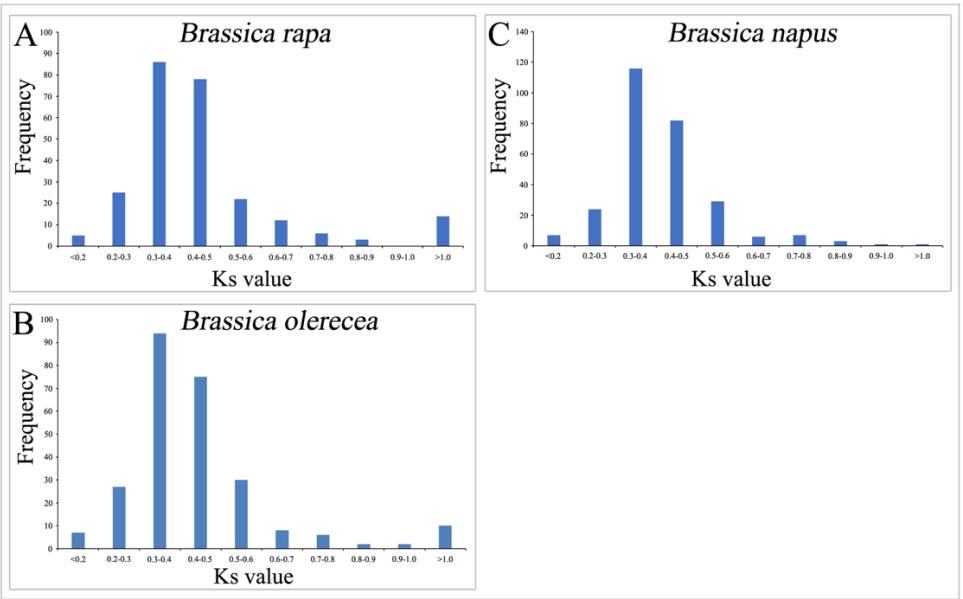


Fig. 4 Ks value distribution of orthologous bHLH genes between *A. thaliana* and *B. rapa* (A), *A. thaliana* and *B. oleracea* (B), and *A. thaliana* and *B. napus* (C). The vertical axis indicates the frequency of paired sequences. The peaks of Ks value appeared between 0.4 and 0.5.

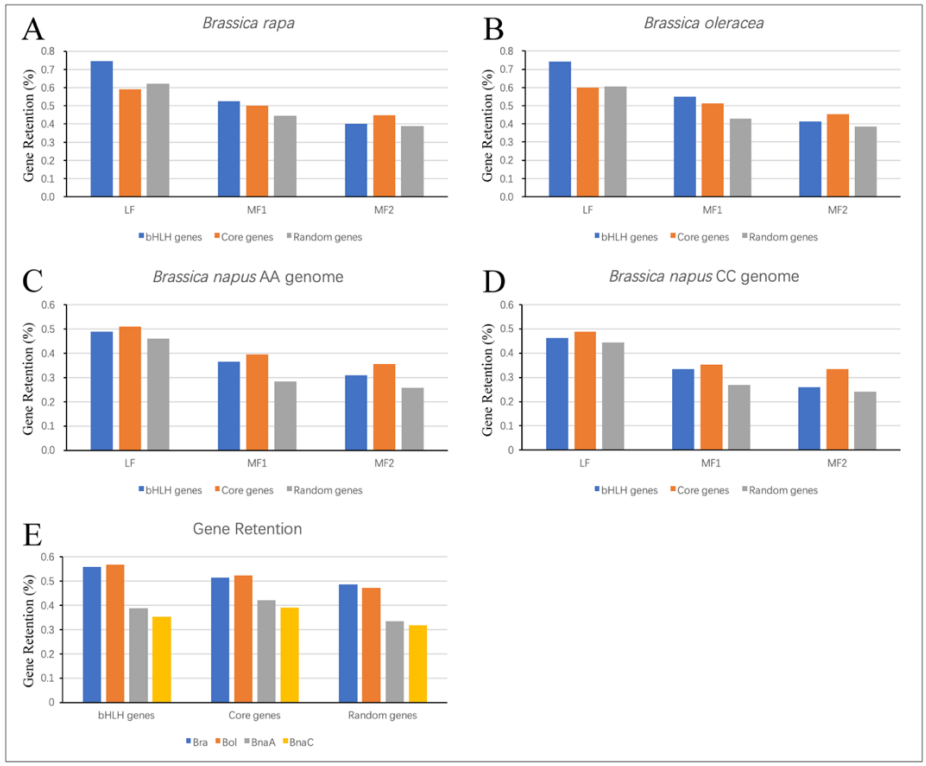
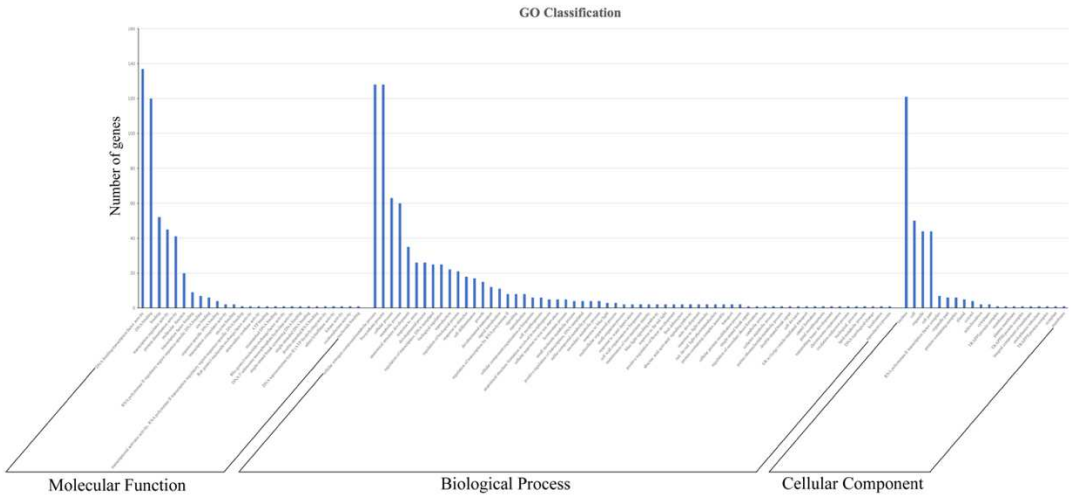


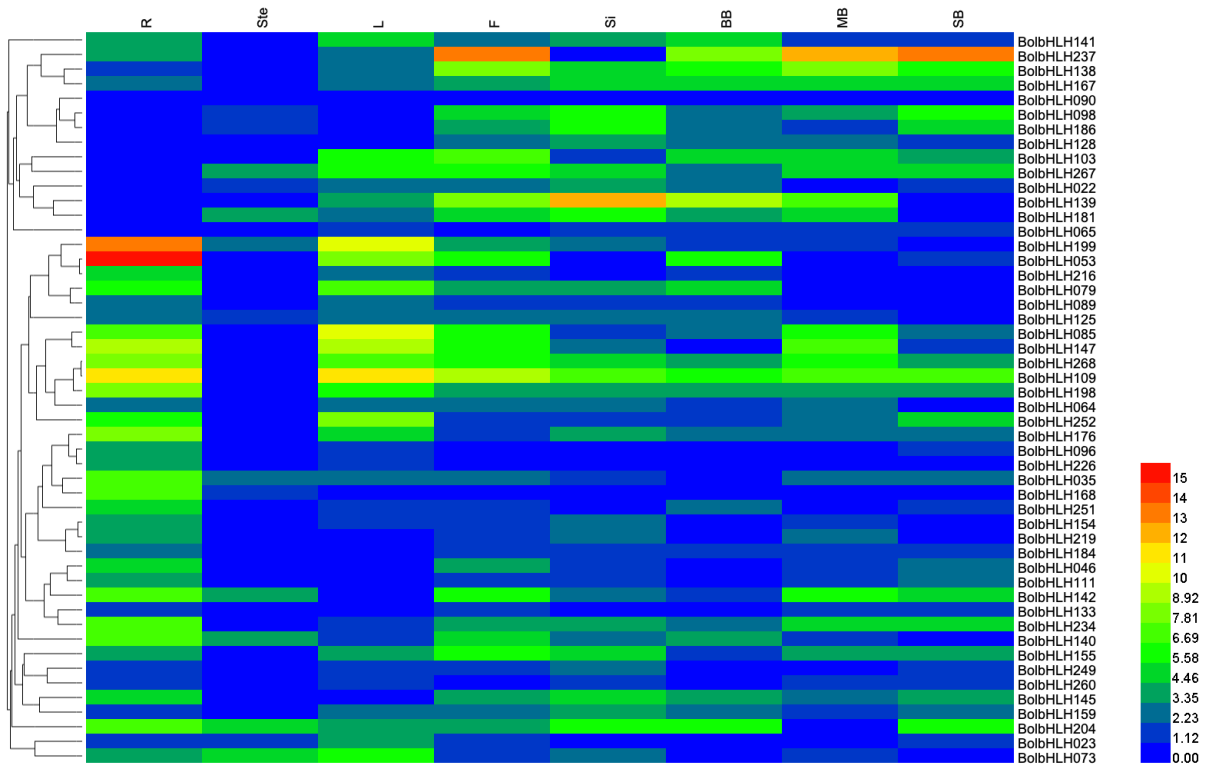
Fig. 5 Retention rates of Arabidopsis bHLH, core, and random genes in *B. oleracea*, *B. rapa*, and *B. napus*. The retention rates in the different subgenomes in (A) *B. rapa*, (B) *B. oleracea*, (C) *B. napus*, and (D) *B. napus*. (E) The retention rates of Arabidopsis bHLH, core, and random genes in the whole genome of *B. oleracea*, *B. rapa*, and *B. napus*.



765

766 **Fig. 6 GO annotation of bHLH genes in *B. oleracea* with Blast2GO.**

767



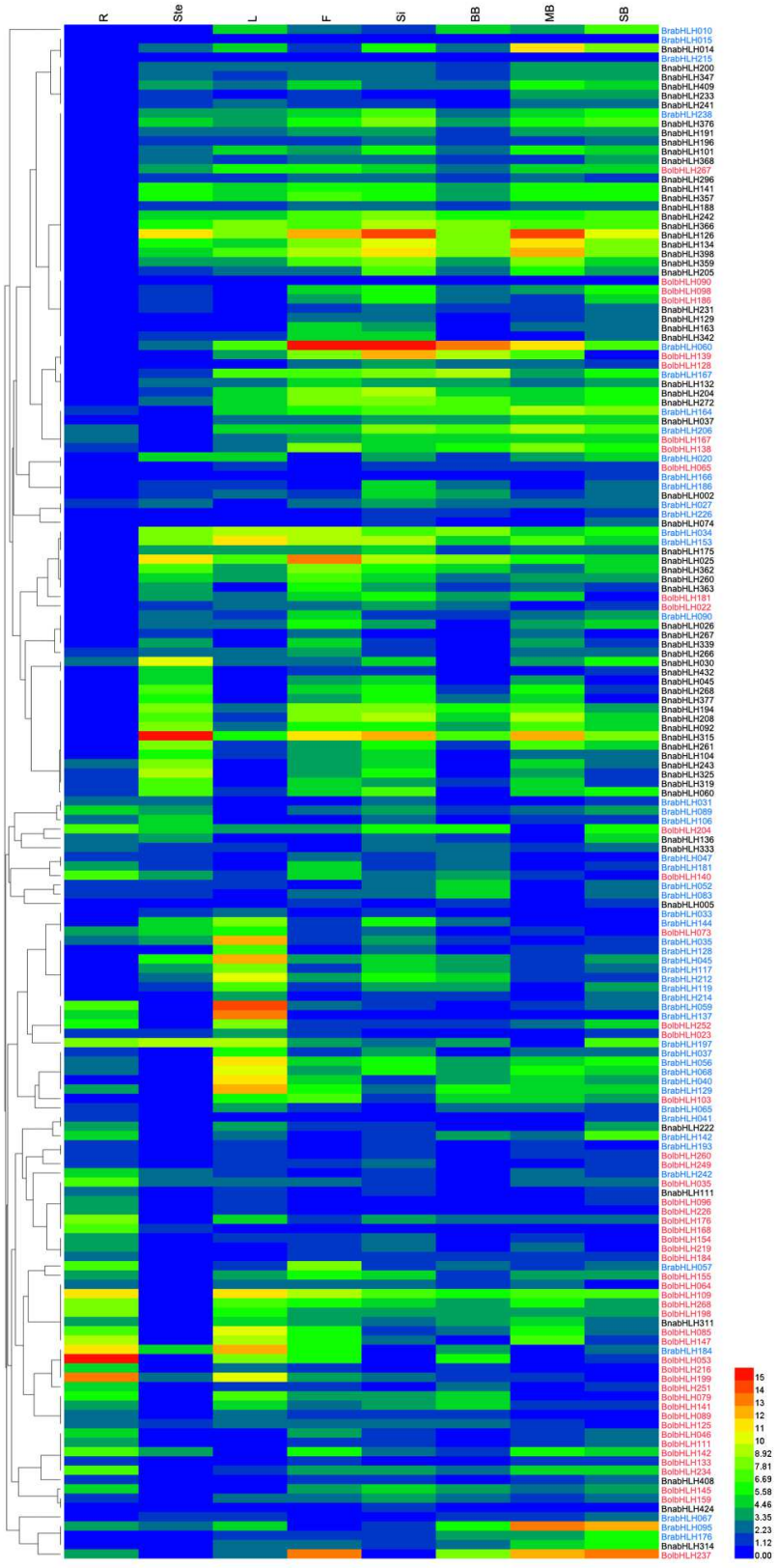
768

769 **Fig. 7 Expression pattern of 50 bHLH genes from different subfamilies in varying tissues in *B. oleracea.***

770 All values underwent logarithmic transformation. R: root, Ste: stem, L: leaves, F: flower, Si: silique, LB: large bud, MB: middle

771 buds, and SB: small bud.

772



773

774 **Fig. 8 Expression patterns of 50 *BolbHLH*, 50 *BrabHLH*, and 65 *BnabHLH* genes in different tissues.** All

775 values underwent logarithmic transformation. R: root, Ste: stem, L: leaves, F: flower, Si: silique, LB: large bud, MB: middle buds,

776 and SB: small bud.

Figures

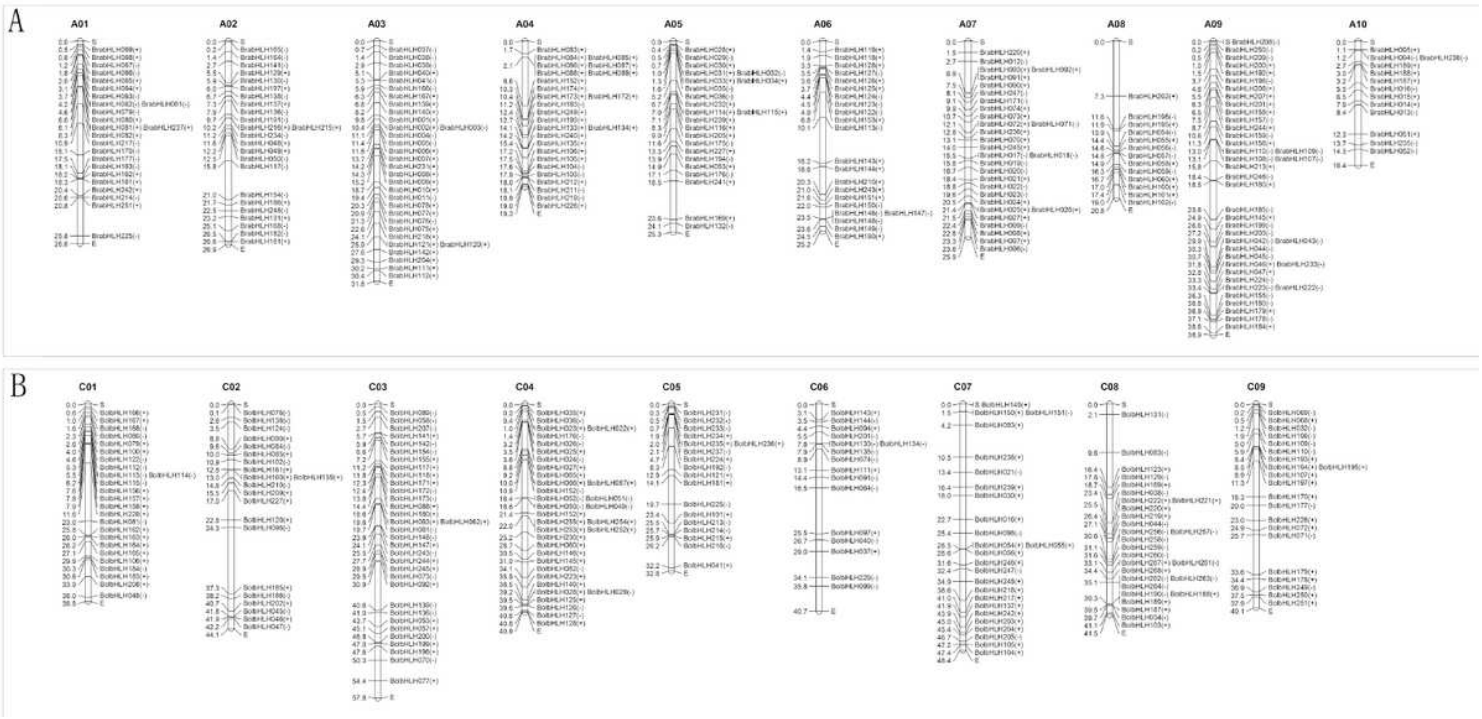


Figure 1

Chromosome location of bHLH genes in B. rapa (A), B. oleracea (B). The symbols + and − indicate the gene located in sense or antisense strands, respectively.

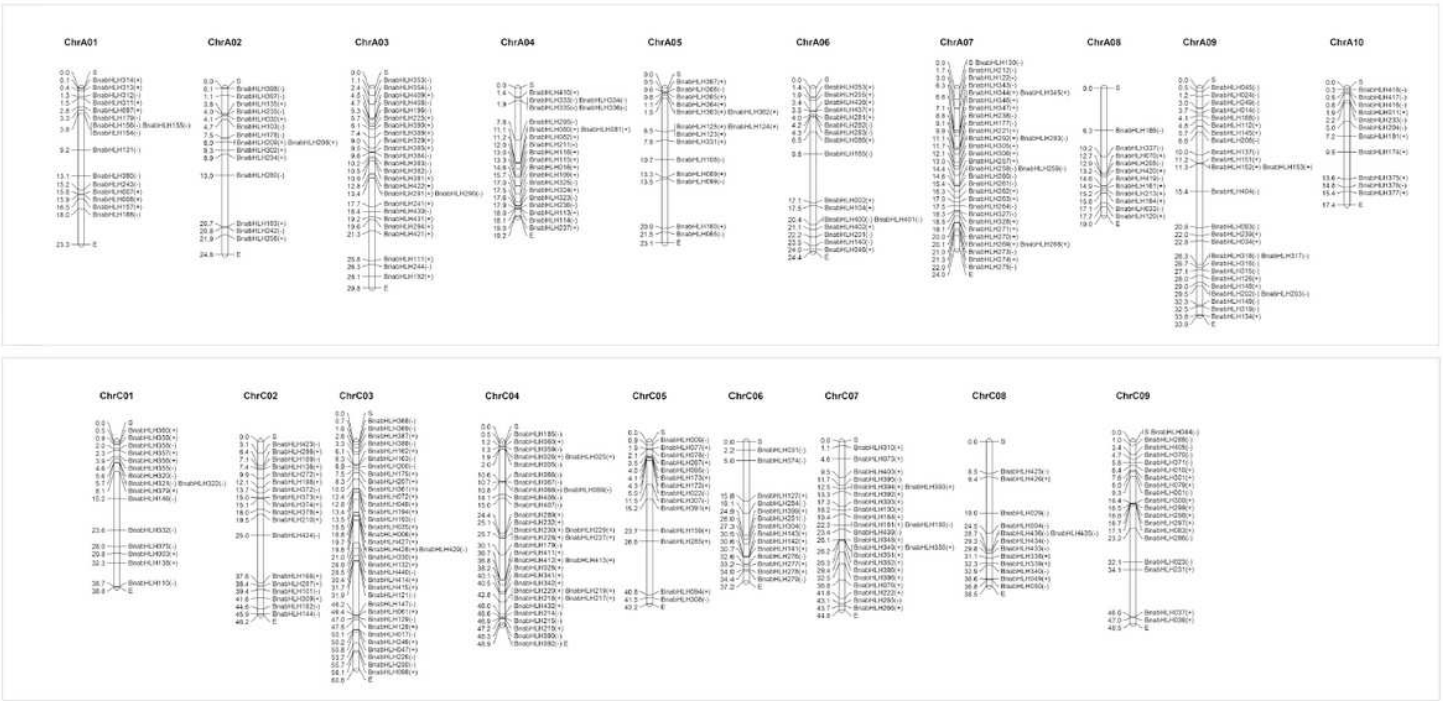


Figure 2

Chromosome location ofAA genome of B. napus (A), and CC genomes of B. napus (B). The symbols + and − indicate the gene located in sense or antisense trands, respectively.

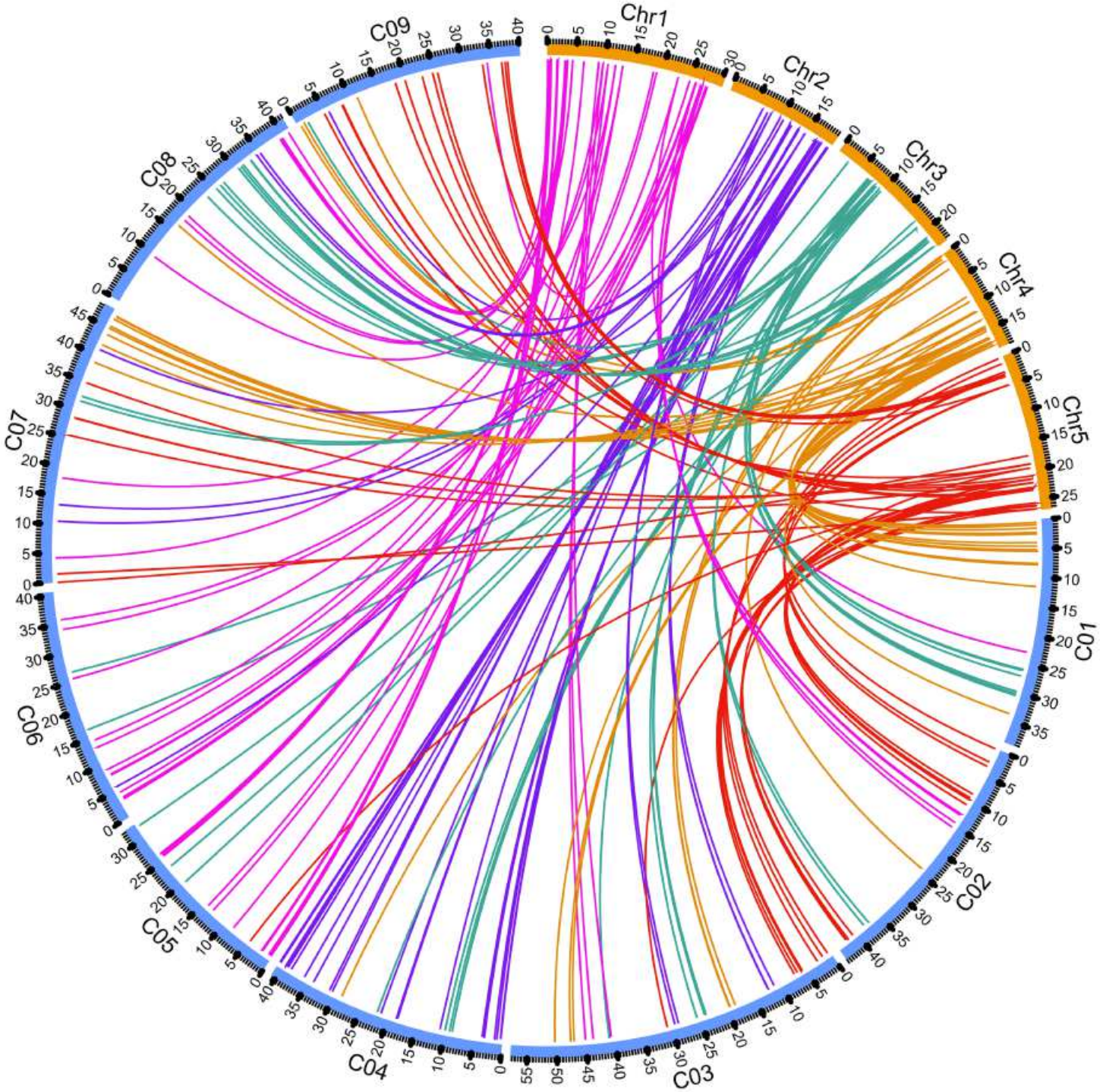


Figure 3

Nine B. oleracea (C01–C10) and five Arabidopsis chromosome (Chr1–Chr5) maps were based on orthologous pair positions and demonstrated highly conserved synteny. The curves of rose red, purple, green, orange, and reddish red link the bHLH gene on the Chr1, Chr2, Chr3, Chr4, and Chr5 chromosomes of A. thaliana and its orthologous genes in B. olerace, respectively.

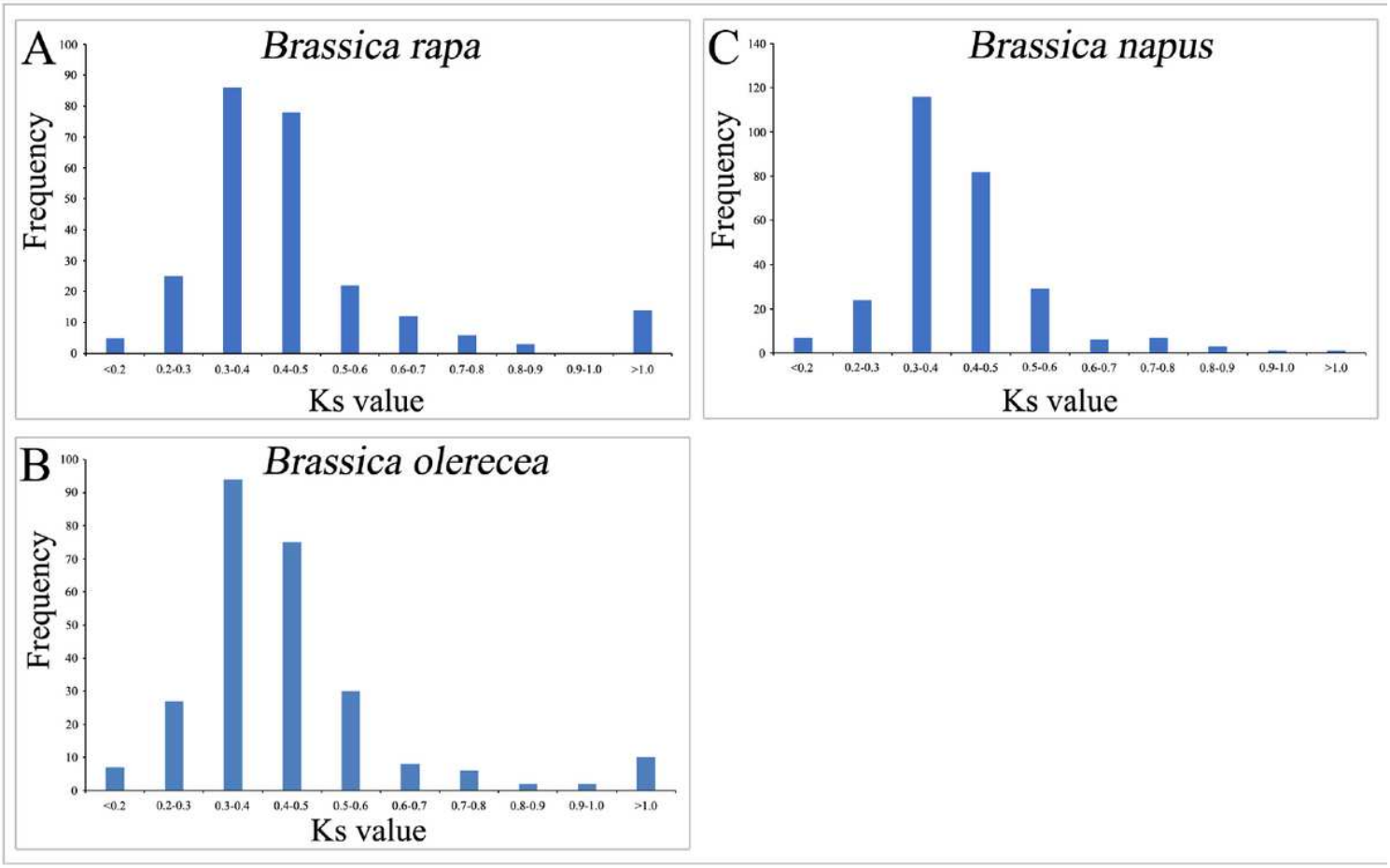


Figure 4

4 Ks value distribution of orthologous bHLH genes between A. thaliana and B. rapa (A), A. thaliana and B. oleracea (B), and A. thaliana and B. napus (C). The vertical axis indicates the frequency of paired sequences. The peaks of Ks value appeared between 0.4 and 0.5.

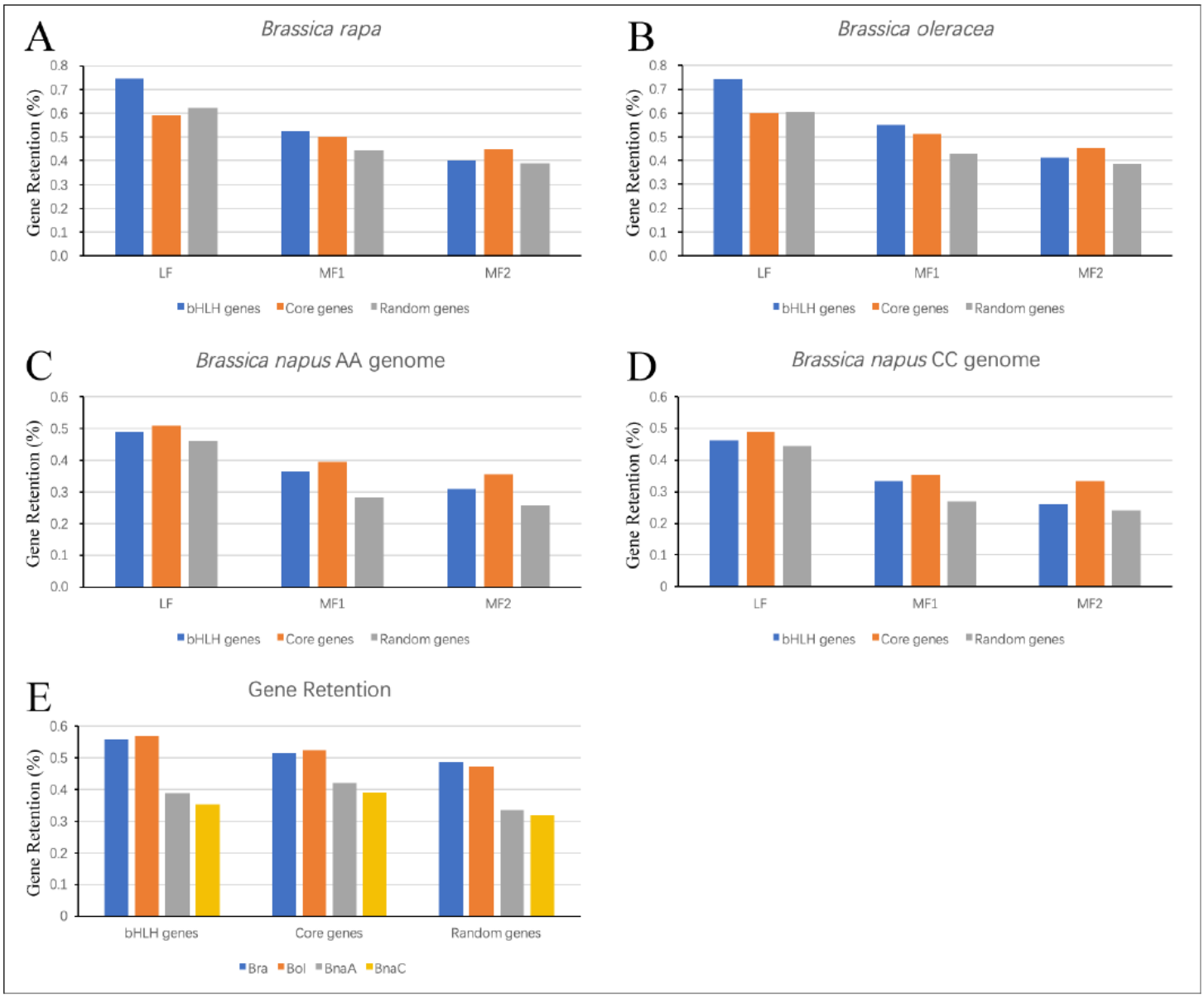


Figure 5

Retention rates of Arabidopsis bHLH, core, and random genes in B. oleracea, B. rapa, and B. napus. The retention rates in the different subgenomes in (A) B. rapa, (B) B. oleracea, (C) B. napus, and (D) B. napus. (E) The retention rates of Arabidopsis bHLH, core, and random genes in the whole genome of B. oleracea, B. rapa, and B. napus.

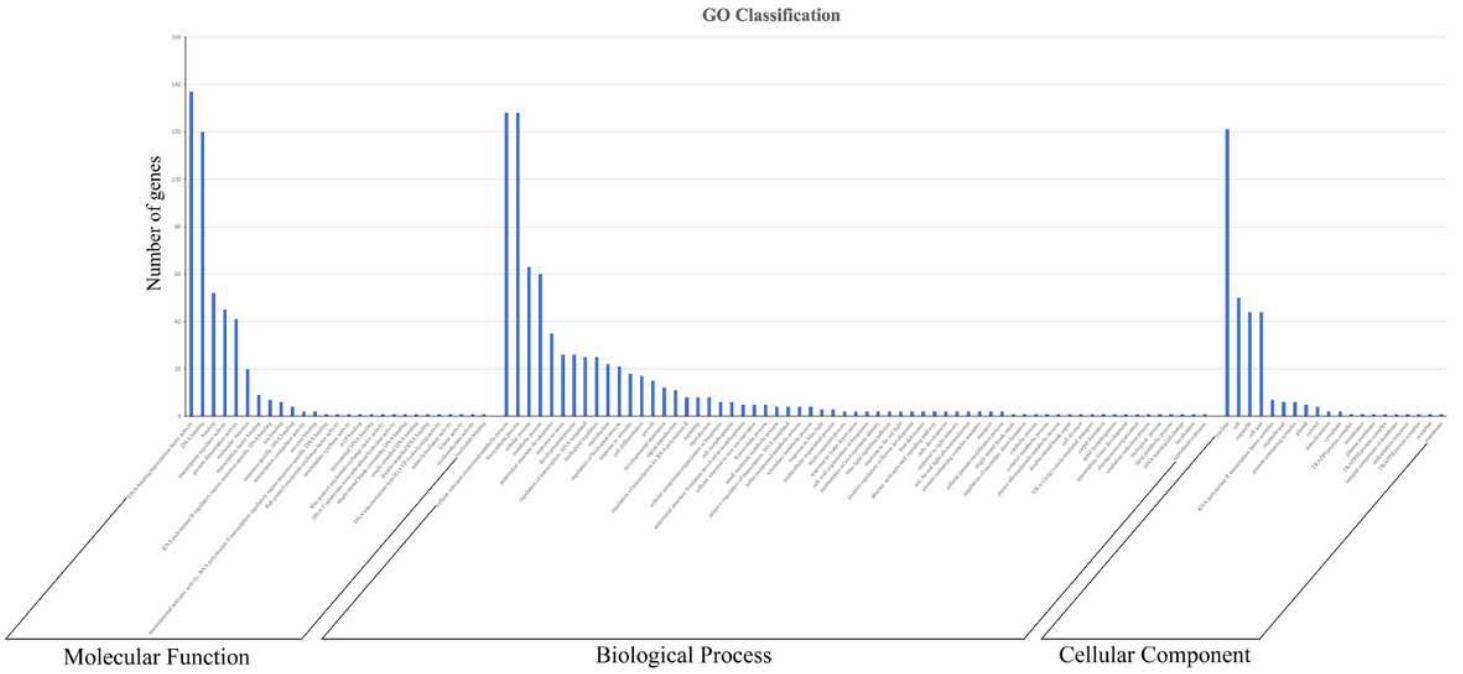


Figure 6

GO annotation of bHLH genes in B. oleracea with Blast2GO.

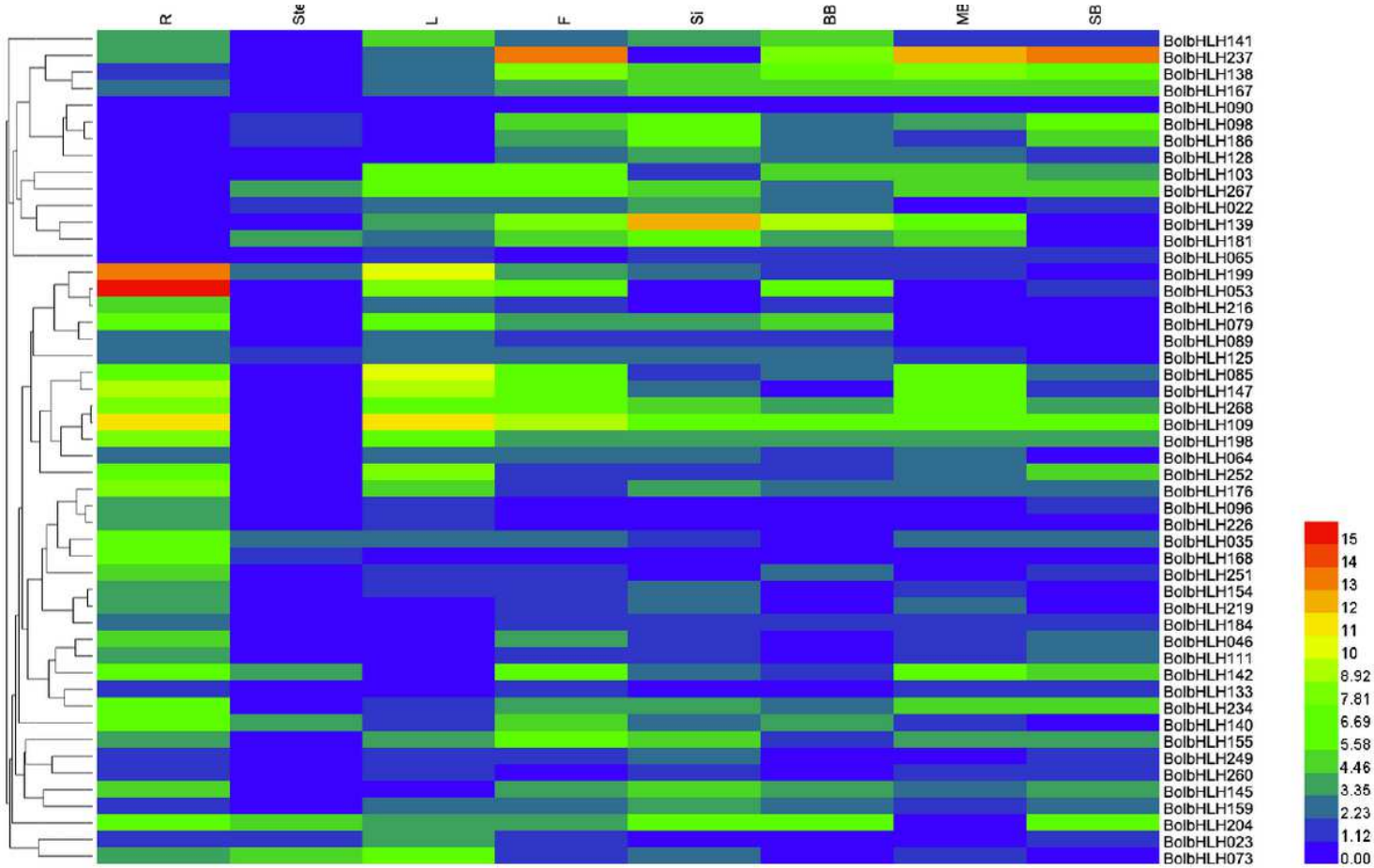


Figure 7

Expression pattern of 50 bHLH genes from different subfamilies in varying tissues in B. oleracea. All values underwent logarithmic transformation. R: root, Ste: stem, L: leaves, F: flower, Si: silique, LB: large bud, MB: middle buds, and SB: small bud.

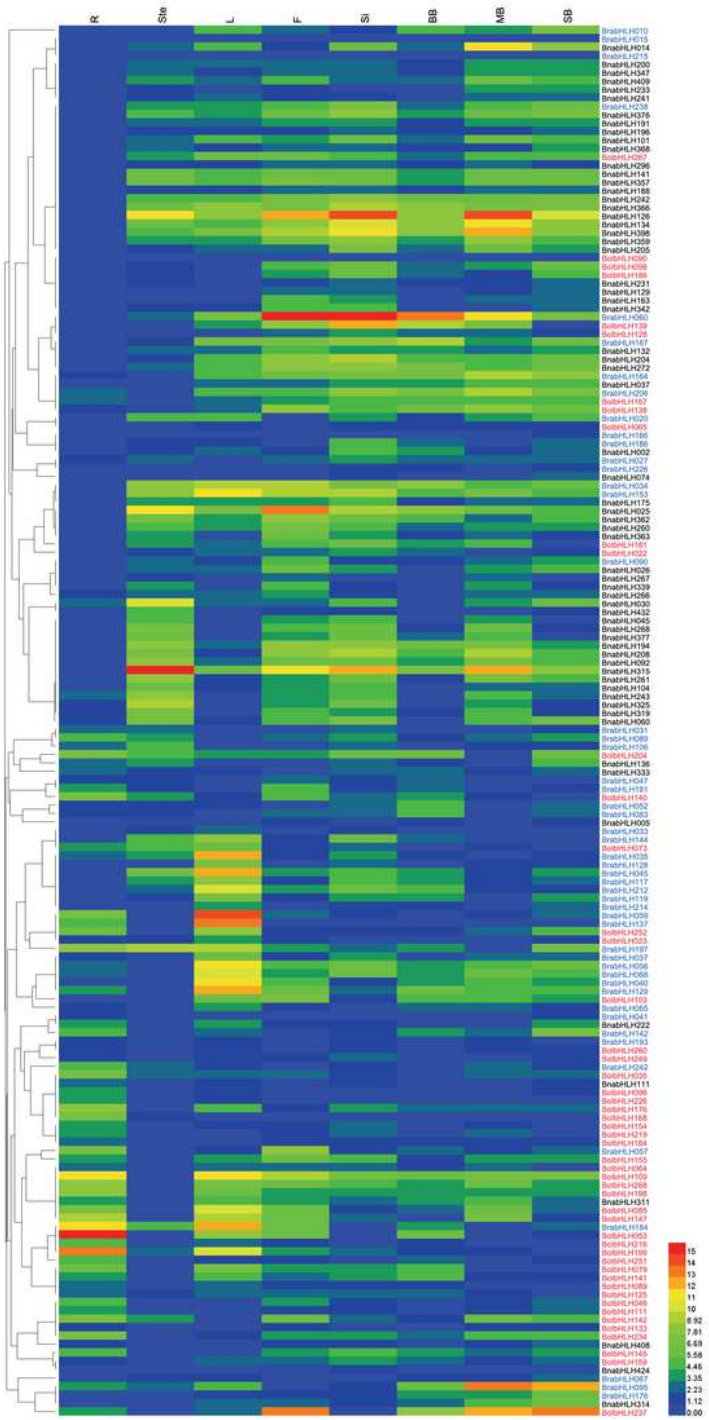


Figure 8

Expression patterns of 50 BolbHLH, 50 BrabHLH, and 65 BnabHLH genes in different tissues. All values underwent logarithmic transformation. R: root, Ste: stem, L: leaves, F: flower, Si: silique, LB: large bud, MB:

middle buds, and SB: small bud.

Supplementary Files

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

- supplement1.pdf
- supplement2.pdf
- supplement3.pdf
- supplement4.pdf
- supplement5.pdf
- supplement6.pdf
- supplement7.pdf
- supplement8.pdf
- supplement9.pdf
- supplement10.pdf
- supplement11.pdf
- supplement12.pdf
- supplement13.pdf
- supplement14.pdf
- supplement15.pdf
- supplement16.pdf
- supplement17.pdf
- supplement18.pdf
- supplement19.pdf
- supplement19.pdf
- supplement21.pdf
- supplement22.pdf
- supplement23.pdf
- supplement24.pdf