Genome-Wide Identification and Expression Analysis of The BHLH Transcription Factor Family and Its Response to Abiotic Stress in Sorghum [Sorghum Bicolor (L.) Moench]

Yu Fan
Guizhou University
Hao Yang
Guizhou University
Dili Lai
Guizhou University
Ailing He
Guizhou University
Guoxing Xue
Guizhou University
Liang Feng
Chengdu Food and Drug Inspection Institute, Chengdu
Long Chen
Department of Nursing, Sichuan Tianyi College, Mianzhu
Xiao-bin Cheng
Department of Environmental and Life Sciences, Sichuan MinZu College
Jingjun Ruan
Guizhou University
Jun Yan
School of Pharmacy and Bioengineering, Chengdu University, Chengdu
Jianping Cheng (chengjianping63@qq.com)
Guizhou University

Research Article

Keywords: Sorghum bicolor, bHLH gene family, genome-wide analysis, abiotic stress

Posted Date: January 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-150639/v1

Declarations

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not Applicable.

Availability of data and materials:

The entire *Sorghum bicolor* genome sequence information was from the Ensembl Genomes website (http://ensemblgenomes.org/). The *Sorghum bicolor* materials (Hongyingzi) used in the experiment were supplied by Prof. Cheng Jianping of Guizhou University. The datasets supporting the conclusions of this article are included in the article and its Additional files.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This research was supported by the National Science Foundation of China (31560578, Cheng JP, http://www.nsfc.gov.cn), Sichuan International Science and Technology Cooperation and Exchange Research and Development Project (2018HH0116, Yan J, http://kjt.sc.gov.cn), Guizhou Science and Technology Support Project (No. 20201Y125).

Authors' contributions:

YF planned and designed the research and analyzed the data. YF and DL wrote the manuscript. HY, LF and LC studied gene expression by qRT-PCR. AH identified the *S. bicolor bHLH* gene family and analyzed gene structure. GX studied chromosome distribution, gene duplication and syntenic analysis of *S. bicolor bHLH* genes. YF and X-bC analyzed the evolutionary relationship of *bHLH* genes in several different species. JC supervised the research. JR and JY revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements:

We thank all of the colleagues in our laboratory for providing useful discussions and technical assistance. We are very grateful to the editor and reviewers for critically evaluating the manuscript and providing constructive comments for its improvement.

Genome-wide identification and expression analysis of the bHLH transcription factor family and its response to abiotic stress in sorghum [*Sorghum bicolor* (L.) Moench]

Yu Fan[1], Hao Yang[1], Dili Lai[1], Ailing He[1], Guoxing Xue[1], Liang Feng[2], Long Chen[3], Xiao-bin Cheng[4], Jingjun Ruan[1], Jun Yan[5*], Jianping Cheng[1*]

1 College of Agriculture, Guizhou University, Guiyang 550025, P.R. China

2 Chengdu Food and Drug Inspection Institute, Chengdu 610000, P.R. China

3 Department of Nursing, Sichuan Tianyi College, Mianzhu 618200, P.R. China

4 Department of Environmental and Life Sciences, Sichuan MinZu College, Kangding 626001, P.R. China

5 School of Pharmacy and Bioengineering, Chengdu University, Chengdu 610106, P.R. China

*** Corresponding author:** Jianping Cheng

Guizhou University, Huaxi District, Guiyang City, Guizhou Province, P.R. China

Tel.: 0851-88305238; Email: chengjianping63@qq.com

Abstract

Background: Basic helix-loop-helix (bHLH) is a superfamily of transcription factors that is widely found in plants and animals, and is the second largest transcription factor family in eukaryotes after MYB. They have been shown to be important regulatory components in tissue development and many different biological processes. However, no systemic analysis of the bHLH transcription factor family has yet been reported in *Sorghum bicolor*.

Results: We conducted the first genome-wide analysis of the bHLH transcription factor family of *Sorghum bicolor* and identified 174 *SbbHLH* genes. Phylogenetic analysis of SbbHLH proteins and 158 *Arabidopsis thaliana* bHLH proteins was performed to determine their homology. In addition, conserved motifs, gene structure, chromosomal spread, and gene duplication of *SbbHLH* genes were studied in depth. To further infer the phylogenetic mechanisms in the *SbbHLH* family, we constructed six comparative syntenic maps of *S. bicolor* associated with six representative species. Finally, we analyzed the gene-expression response and tissue-development characteristics of 12 typical *SbbHLH* genes in plants subjected to six different abiotic stresses. Gene expression during flower and fruit development was also examined.

Conclusions: This study is of great significance for functional identification and confirmation of the *S. bicolor bHLH* superfamily and for our understanding of the *bHLH* superfamily in higher plants.

Keywords: *Sorghum bicolor*, bHLH gene family, genome-wide analysis, abiotic stress

Background

Transcription factors (TFs) play an important role in controlling plant growth and environmental adaptation [1, 2]. They regulate gene expression by combining with specific cis-promoter elements that specifically regulate certain genes or transcription rates, thereby playing a unique regulatory role in plant morphogenesis, cell-cycle processes, and the like [3, 4]. Structurally, the typical TF includes a DNA-binding site, a transcription-activation or repression domain, an oligomerization site, and a nuclear-localization site. TF genes, such as members of the *bHLH*, *WRKY*, *MYB*, *bZIP* and other TF families, constitute a high proportion of all plant genomes, and their target genes are widely involved in physiological processes, such as plant development

and stress responses [5, 6].

Basic helix-loop-helix (bHLH) is a superfamily of TFs that is widely found in plants and animals; it is the second largest TF family among eukaryotic proteins after MYB [7-9]. The first discovered *bHLH* family member was the c-myc proto-oncogene of avian myeloid cell carcinoma virus [10]. The bHLH TFs are so named because of their structural feature of a bHLH domain in all family members. The amino acid sequence of this domain is highly conserved. There are about 50 to 60 amino acid residues that can be divided into two regions based on their functions: a basic region and the HLH [10, 11]. The basic domain is located at the N terminus of the conserved domain of bHLH and contains about 15 amino acids. It can bind to the cis-acting element E-box (5'-canntg-3'). Therefore, the number of basic and key amino acid residues in the basic region determine whether the bHLH TF has DNA-binding activity. The HLH domain is distributed at the C terminus of the gene sequence, where two α-helices are connected by a low-conserved loop, which is essential for the formation of homodimers or heterodimers of bHLH TFs [8, 12, 13]. Based on their ability to bind DNA, bHLH TFs can be divided into two categories: DNA binding and non-DNA binding. These can be further divided into E-box binding and non-E-box binding. The most common method of E-box binding is G-box binding (5'-cacgtg-3') [11, 14, 15]. According to Atchley et al. [11], Glu and Arg at positions 9 and 13 of the basic region, namely E9 and R13, are essential amino acid residues that bind to E-box and H/K5-E9-R13 patterns, and bind to G-box. There are many bHLH families in animals, which are divided into 6 groups (A–F), with a total of 45 families [9, 16]. The study of *bHLH* gene family in different species will help to understand the evolutionary process and biological function. Previous phylogenetic results showed that *bHLH* proteins in plants were divided into 26 subfamilies, 20 of which were found in the common ancestor of vascular and bryophytes plants[17]. Toledo Ortiz et al. [15] divided 147 *AtbHLH* proteins into 21 subfamilies; and Li et al. [18] divided 167 *OsbHLH* proteins into 22 subfamilies.

The bHLH TF family is involved in plants' perception of the external environment, cell-cycle regulation, and tissue differentiation [18, 19]. Different subfamilies regulate different biological processes, such as transduction of light signals [20, 21] and hormone signals [22, 23], and organ development [24-27]. Under stress conditions, certain bHLH TFs are activated; they combine with

146 the promoters of key genes involved in various signaling pathways, and regulate the transcription
147 level of these target genes, thereby regulating the plants' stress tolerance. For example, some
148 researchers have found that the homologous *bHLH* genes *bhlh068* of *Oryza sativa* and *bHLH112*
149 of *Arabidopsis thaliana* play an active role in the response to salt stress, but have opposite effects
150 on regulation of plant flowering [28]. Appropriate TFs, together with *AtbHLH38* and *AtbHLH39*,
151 can regulate iron metabolism in *Arabidopsis* [29]. *Atbhlh112* is a transcriptional activator of
152 drought and other stress signal-transduction pathways, but it has an inhibitory effect on root
153 development [30]. In *Nicotiana tabacum*, plants overexpressing *Ntbhlh123* have enhanced
154 resistance under low-temperature stress [31]. bHLH TFs are involved in regulating the
155 accumulation of secondary metabolites in plants [32]. These examples all show the roles of bHLH
156 TFs in the plant response to stress.
157 The bHLH protein has been conserved throughout the evolutionary history of the plant kingdom.
158 The expansion of this family is closely related to plant evolution and diversity [33, 34], not only in
159 higher plants, but also in lower plants or non-plants, such as algae, mycobacteria, lichens and
160 mosses [34]. As regards abiotic stresses, *bHLH* is mainly involved in the defense responses to
161 drought, high temperature, low temperature, and high salinity, which are unique to the terrestrial
162 environment. Therefore, the evolution of the *bHLH* gene family provides clues to understanding
163 the evolution of green algae to flowering plants through their adaptation to environmental changes.
164 In particular, genome-wide analysis of *bHLH* gene families of different species will help
165 understand the biological function and evolutionary origin of the *bHLH* genes.
166 *Sorghum bicolor* (L.) Moench is an annual herb in the family Gramineae [35]. It is a common
167 grain crop, widely distributed in the tropical, subtropical and temperate regions of the world and
168 cultivated in the northern and southern provinces of China. *S. bicolor* seeds serve as a food source
169 in China, North Korea, the former Soviet Union, India and Africa [36]. *S. bicolor* has rich genetic
170 and phenotypic diversity, especially in plant height, seed color, seed size and branch number.
171 Moreover, *S. bicolor* is a particularly nutritious crop, high in resistant starch, proteins, vitamins
172 and polyphenols [37, 38], and it is widely used in the brewing industry [39]. However, its growth
173 and development are highly susceptible to environmental stress. For example, *S. bicolor* plants
174 show reduced floret fertility and single-grain weight under high temperature, thereby reducing

yield [40, 41]; low temperature leads to weakening of this crop's growth potential, and plants are generally seriously damaged by frost [42]. *S. bicolor* has a well-developed root system that enables it to survive drought to some extent [43, 44]; nevertheless, long-term extreme drought has a huge impact on growth and yield [43]. In the process of *S. bicolor* production, pests, diseases, weeds and other biotic stresses will also cause serious yield losses [44]. Because *S. bicolor* is cultivated throughout the world, it has great economic and research value, and the identification of its functional genes is important.

In 2009, the completion and publication of the whole *S. bicolor* genome sequence enabled us to further explore, clone and verify the *bHLH* genes related to its stress resistance [45]. The *S. bicolor* genome is 750 Mb in length, with about 30,000 genes, ca. 75% more than in rice [46]. The *bHLH* gene family has been widely studied in many plant species, such as *Arabidopsis* [15], rice [18], Chinese cabbage [47], tomato [48], common bean [49], apple [50], peanut [51], *Brachypodium distachyon* [52], potato [26], maize [53], wheat [54], MOSO bamboo [55], *Carthamus tinctorius* [56], Chinese jujube [57], pepper [58], Jilin ginseng [59], pineapple [60], and tartary buckwheat[61], among others. However, at present, our understanding of gene families in *S. bicolor* is very limited. The main gene families identified in this plant are MADS-box [62], Dof [63], CBL [64], ERF [65], SBP-box [66], HSP [67], LEA [68], and NAC [69], among others. Because *bHLH* genes play an important role in various physiological processes, it is of great significance to systematically study the bHLH family in *S. bicolor*. Here, we identified 174 *bHLH* genes in *S. bicolor* and classified them into 24 major groups. Exon–intron structure, motif composition, gene duplication, chromosome distribution, and phylogeny were analyzed. The expression of *bHLH* family members in *S. bicolor* under different biological processes and abiotic stresses was also analyzed. This study provides valuable clues to the functional identification and evolutionary relationships of *S. bicolor*

Results

Identification of *bHLH* genes in *S. bicolor*

To identify all possible *bHLH* members in the *S. bicolor* genome, we used two BLAST methods (Additional file 1: Table S1). To better distinguish these genes, we named them *SbbHLH001* to

SbbHLH174 according to their location on the *S. bicolor* chromosomes (Additional file 1: Table S1) and provide the genes' characteristics, including molecular weight, isoelectric point (pI), protein length, domain information, and subcellular localization (http://cello.life. nctu.edu.tw/) (Additional file 1: Table S1).

Of the 174 SbbHLH proteins, SbbHLH031 and SbbHLH168 were the smallest with 87 amino acids, and the largest protein was SbbHLH040 with 1105 amino acids. The molecular mass of the proteins ranged from 9.67 kDa (SbbHLH168) to124.74 kDa (SbbHLH040), and the pI ranged from 4.53 (SbbHLH081) to 12.05 (SbbHLH004), with a mean of 6.70. Of all of the *SbbHLH* genes, 14 contained the bHLH-MYC-N domain and 172 contained the HLH domain (the exceptions being *SbbHLH097* and *SbbHLH116*). The predicted subcellular localization results showed that 141 SbbHLHs are located in the nucleus, 26 in the cytoplasm, 4 in the mitochondria, 2 (SbbHLH103 and SbbHLH090) in the endoplasmic reticulum, and 1 (SbbHLH095) in the cytoskeleton (Additional file 1: Table S1). The ratio of *SbbHLH* genes to total genes in the *S. bicolor* genome was about 0.58%, which is similar to *Arabidopsis* (0.59%), but more than in rice (0.44%) [18], poplar (0.40%) [27], and tomato (0.46%) [48].

Multiple sequence alignment, phylogenetic analysis, and classification of *SbbHLH* genes

We constructed a phylogenetic tree using the neighbor-joining (NJ) method with a bootstrap value of 1000 based on the amino acid sequences of 174 SbbHLH and 158 AtbHLH proteins (Figure 1; Additional file 1: Table S1). According to the topological structure of the tree and classification method proposed by Pires and Gabriela [15, 17], 332 *bHLH* genes in the phylogenetic tree were divided into 24 clades (groups 1–24) and 1 orphan [1, 6-7]. The unclassified group (UC) contained 8 *SbbHLH* and 6 *AtbHLH* genes, and 149 SbbHLH proteins clustered into 21 subfamilies. This is consistent with the taxonomic group of *bHLH* proteins in *Arabidopsis* [18], indicating no loss of those proteins during the long-term evolution in *S. bicolor* evolution. The bHLH proteins within the reported subfamilies may play a fundamental role in the development, adaptation and evolution in dissimilar plant species, including peanut [51], tomato [48], Chinese cabbage [47], wheat [54], and *Carthamus tinctorius* [56]. Seventeen *S. bicolor* proteins constituted three typical topological structures (groups 22–24), suggesting that these are

new characteristics in the evolution of *S. bicolor* diversity. None of AtbHLHs was assigned into subfamily 23,which contained 7 SbbHLHs (SbbHLH86, SbbHLH87, SbbHLH108, SbbHLH123, SbbHLH124, SbbHLH142, SbbHLH143); this group might indicate a new evolutionary direction for *S. bicolor*. Among the 24 subfamilies, the subfamily 15 had the largest number of members (17 SbbHLHs), and subfamilies 2 (SbbHLH79), 14 (SbbHLH68), and 20 (SbbHLH34) had the fewest (1 SbbHLH). The *SbbHLH* genes, which is not clearly classified into any subfamily, were classified as "orphans" [15, 16] (Figure 1, Additional file 1: Table S1). A phylogenetic tree for *Arabidopsis* showed that some *SbbHLH*s are tightly grouped with the *AtbHLH*s (bootstrap support ≥70). These may be orthologous to the *AtbHLH*s and have similar functions.

The bHLH domain of *Arabidopsis* bHLH proteins and those from subgroups 1–21 were randomly selected as representatives of groups and subgroups for further multiple-sequence comparison (Figure 2, Additional file 1: Table S1). The *SbbHLH* members from groups 22–24 were selected for the comparison. The bHLH domains of *S. bicolor* span approximately 50 amino acids. As shown in Figure 2, although the characteristic bHLH domain is well conserved in *Arabidopsis* and *S. bicolor*, the regions outside of this domain in the rest of the protein are usually differentiate and diversify [13, 14, 18]. We considered the basic region to be 17 amino acids long based on Gabriela's view [15]. In terms of amino acid structure, the loop was the most divergent region of this domain, especially in subfamily 6, 10 and 23, as has been observed for bHLH proteins from other plants, including *Arabidopsis* [18], potato [26], tomato [48] and buckwheat [61].

Conserved motifs and gene structure analysis of *SbbHLH* genes

To understand the structural components of the *SbbHLH* genes, their exon and intron structures were obtained by comparing the corresponding genomic DNA sequences (Figure 3, Additional files 1and 2: Tables S1 and S2). A comparison of the number and position of the exons and introns revealed that the 174 *SbbHLH* genes had different numbers of exons, varying from 1 to 12 (Figure 3A/3B). In addition, 17 (9.77%) genes contained 1 exon, and the remaining genes had 2 or more exons. The 17 intronless genes belonged to four subfamilies (8, 13, 14, 19), but were mainly in subfamilies 8 and 19. Genes with few or no introns are considered to have lower expression levels

in plants [70]. However, the compact gene structure may contribute to the rapid expression of genes in response to endogenous and/or exogenous stimuli [71]. The largest proportion of *SbbHLH* genes (n = 31) had 2 introns. *SbbHLH038* and *SbbHLH054* had the most introns, with 11. Group 1, 2, 4, 10, 20, 21 and 23 members contained 1 or 2 introns. Further analyses indicated that group 18 showed more diversity in the number of introns. In general, members of the same subfamily had similar gene structures.

To further study the characteristic region of the SbbHLH proteins, the motifs of 174 SbbHLH proteins were analyzed using the online tool MEME. A total of 10 distinct conserved motifs (motifs 1–10) were found (Figure 3C, Additional file 2: Table S2). As exhibited in Figure 3C, motifs 1 and 2 were widely distributed in the SbbHLHs, except for SbbHLH001 and SbbHLH017, and the two motifs were very close to each other in the bHLH proteins. SbbHLH members within the same groups were usually found to share a similar motif composition. For example, group 1, 2, 3, 5, 7, 9, 11 and 23 members contained motifs 1, 2, and 4; groups 12 and 17 contained motifs 1, 2, and 5; group 16 contained motifs 3, 1, and 2; and group 22 contained motifs 6, 1, 2, 8, and 4. At the same time, we found that some motifs were only present in specific subfamilies. In addition, motif 5 was specific to groups 12, 17 and 20, whereas motif 8 was specific to groups 5, 10 and 22. Further analysis showed that some of the motifs could only be distributed in specific locations of the pattern. For example, motif 1 was always distributed at the start of the pattern in groups 1, 2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 20, 21, 23 and 24; motif 6 was almost always distributed at the start of groups 7 and 22; motif 3 was almost always distributed at the start of groups 16, 17 and 18. Motif 4 was almost always distributed at the end of the pattern in groups 1, 2, 7, 8, 9, 10, 11, 22 and 23; and motif 10 was distributed at the end of the pattern in the group 6. The functions of most of these conserved motifs remain to be elucidated. Overall, members that belonged to the same subfamily had similar gene structure and motif composition, in accordance with the results of the phylogenetic analysis, and supporting the reliability of the population classification.

Chromosomal spread and gene duplication of *SbbHLH* genes

A map of the physical position of the *SbbHLH* genes was created based on the latest *S. bicolor* genome database (Figure 4, Additional file 3: Table S3). The distribution of the 174 *SbbHLH*

genes on chromosomes (Chr) 1 to 10 was uneven (Figure 4). Each of the *SbbHLH*s' names was given according to its physical position from the top to the bottom on *S. bicolor* Chr1 to Chr10. Chr1 contained the largest number of *SbbHLH* genes (35 genes, ~20.11%), followed by Chr3 (23, ~13.22%), while Chr5 contained the least (5, ~2.87%). Chr2 and Chr4 each contained 21 (~12.07%) *SbbHLH* genes. Chr8 and Chr9 each contained 12 (~6.90%) *SbbHLH* genes. Chr6, Chr7, and Chr10 contained 16 (~9.20%), 19 (~10.92%), and 10 (~5.75%) *SbbHLH* genes, respectively. Interestingly, most *SbbHLH* genes were distributed at the ends of the 10 chromosomes. In addition, we observed a large number of *SbbHLH* gene-duplication events. A chromosomal region within 200 kb exhibiting two or more identical genomic regions is defined as a tandem duplication event [35]. On chromosomes 1, 3, 4, 6, 7 and 8, we discovered 13 tandem duplication events involving 20 *SbbHLH* genes (Figure 4). *SbbHLH132*, *SbbHLH133*, *SbbHLH134*, *SbbHLH147*, *SbbHLH148* and *SbbHLH149* each had two tandem repeat events (*SbbHLH132* and *SbbHLH131* / *SbbHLH133*; *SbbHLH133* and *SbbHLH132* / *SbbHLH134*; *SbbHLH134* and *SbbHLH133* / *SbbHLH135*; *SbbHLH147* and *SbbHLH146* / *SbbHLH148*; *SbbHLH148* and *SbbHLH147* / *SbbHLH149*; *SbbHLH149* and *SbbHLH148* / *SbbHLH150*). All genes that formed tandem repeat events came from the same subfamily. For example, *SbbHLH117* and *SbbHLH118* were tandem repeat genes and they clustered together in subfamily 3 (Figure 4, Additional file 3: Table S3).

In addition, there were 42 pairs of segmental duplications in the *SbbHLH* genes (Figure 5, Additional file 4: Table S4). As shown in Figure 5, 71 (40.8%) paralogs were identified in the *SbbHLH* gene family, indicating an evolutionary relationship among these *bHLH* members. The *SbbHLH* genes were unevenly distributed in 10 *S. bicolor* linkage groups (LGs) (Figure 5). Some LGs had more *SbbHLH* genes than others (LG2, LG7). LG2 had the most *SbbHLH* genes (14), and LG5 had the least (1). Further analysis of the subfamilies of these genes showed that most of them are linked within their subfamily, except for *SbbHLH024* / UC and *SbbHLH056* / 6. For all identified *SbbHLH* genes, group 18 had the largest number of linked genes (9/71). In addition, the group 15 had 8 genes, while groups 13 and 6 had only 1 (Additional file 4: Table S4). These results suggest that some *SbbHLH* genes may have been produced by gene-replication events, and that these replication events played a major role in the occurrence of new functions in *S. bicolor*

evolution and the amplification of the *SbbHLH* gene family.

Synteny analysis of *SbbHLH* genes

To further infer the phylogenetic mechanisms of the *S. bicolor* bHLH family, we constructed six comparative synteny maps of *S. bicolor*'s association with six representative species, including three dicotyledons (*A. thaliana*, *Vitis vinifera* and *Solanum lycopersicum*) and three monocotyledons (*B. distachyon*, *O. sativa* and *Zea mays*) (Figure 6, Additional file 5: Table S5). A total of 150 *SbbHLH* genes showed syntenic relationships with those in *A. thaliana* (16), *V. vinifera* (46), *S. lycopersicum* (37), *B. distachyon* (129), *O. sativa* (135) and *Z. mays* (195) (Additional file 5: Table S5). The numbers of orthologous pairs between the other six species (*A. thaliana*, *V. vinifera*, *S. lycopersicum*, *B. distachyon*, *O. sativa* and *Z. mays*) were 20, 66, 59, 194, 208 and 273, respectively. Some *SbbHLH* genes were associated with at least four syntenic gene pairs (particularly between *S. bicolor* and *Z. mays bHLH*), such as *SbbHLH043*, *SbbHLH049*, *SbbHLH050*, *SbbHLH101*, *SbbHLH137*, *SbbHLH138*, *SbbHLH141* and *SbbHLH166*, hinting at these genes' important role during evolution.

Interestingly, some collinear gene pairs (with 57 *SbbHLH* genes) identified between *S. bicolor* and *B. distachyon*, *O. sativa* or *Z. mays* were not found between *S. bicolor* and *A. thaliana*, *V. vinifera*, or *S. lycopersicum*, such as *SbbHLH001* with KQK12528/BGIOSGA013800-TA/Zm00001d034596_T001, and *SbbHLH004* with KQK12892/BGIOSGA013672-TA/Zm00001d034298_T001. This suggests that these homologous genes may be gradually formed after the independent differentiation of monocotyledons (Additional file 5: Table S5). Similar patterns were also observed between *S. bicolor* and *O. sativa*/ *Z. mays*, which may be related to the phylogenetic relationships between *S. bicolor* and the other six plant species. In addition, some *SbbHLH* genes were found to be associated with at least one syntenic gene pair among the six plants (especially between *S. bicolor* and *Z. mays*), such as *SbbHLH030*, *SbbHLH045*, *SbbHLH050*, *SbbHLH066*, *SbbHLH099*, *SbbHLH136*, *SbbHLH138*, *SbbHLH154*, *SbbHLH166*, suggesting that these orthologous pairs already existed before the ancestral divergence, and thus indicating that these genes may have played an important role in the *bHLH* gene family during evolution. To better understand the evolutionary constraints of the

bHLH gene family in *S. bicolor*, the *SbbHLH* genes were subjected to Tajima's D Neutrality Test. Calculations gave D =7.736378, the large deviation from 0, suggesting that the *SbbHLH* gene family might have experienced strong purifying selective pressure during evolution.

Evolutionary analysis of the *SbbHLH* genes and *bHLH* genes of several different species

To analyze the evolutionary relationship of the trihelix family of *bHLH* proteins among *S. bicolor* and six other plants (*A. thaliana*, *V. vinifera*, *S. lycopersicum*, *B. distachyon*, *O. sativa*, *Z. mays*), an unrooted NJ tree with 10 conserved motifs according to the MEME web server was constructed using the NJ method of Geneious R11 according to the protein sequences of 174 *SbbHLH* genes and the six other plants' trihelix genes (Figure 7, Additional file 2: Table S2). The detailed genetic correspondence is presented in Additional files 1 and 2: Tables S1 and S2. The distribution of *SbbHLH*s in the phylogenetic tree was relatively dispersed. As shown in Figure 7, the SbbHLH proteins tended to gather with the bHLH proteins of *O. sativa* and *Z. mays*. With the exception of a few bHLH proteins, for example ZmbHLH8, ZmbHLH53, SbbHLH001, all other proteins of the six studied plants contained motifs 1 and 2. In addition, several motifs existed only in bHLH proteins of a few specific SbbHLH branches, such as motifs 5, 8 and 10. We also found that the bHLH proteins of *O. sativa*, *Z. mays* and *S. bicolor* on the same branch generally have similar motif compositions, and similar serial motifs tend to cluster in specific bHLH protein families. For example, serial motifs 1, 2, 5 and 10 tended to gather within group 6; and serial motifs 8, 9, 1, 2, 7 and 4 tended to gather within group 8. Thus, SbbHLH proteins may be more closely related to those of *O. sativa* and *Z. mays*.

Expression patterns of *SbbHLH*s in several plant organs

To investigate the potential roles of the *SbbHLH* genes, real-time PCR was used to detect the expression of 12 individual members of the gene family which were homologous to, or had close evolutionary relationships with *AtbHLH* genes with established functions. The accumulation of the transcriptional products of 12 *SbbHLH* genesfrom different subfamilies in six organs (anthers, styles, roots, leaves, fruit and stems) was evaluated (Figure 8A). The results showed that some genes exhibited preferential expression in some tissues of *S. bicolor*. Most of the genes were

expressed in all organs, and 4 genes (*SbbHLH014*, *SbbHLH050*, *SbbHLH079*, *SbbHLH134*) showed their highest expression level in the styles. Two genes (*SbbHLH063* and *SbbHLH110*) showed their highest expression level in the anthers, and the highest expression level of *SbbHLH037* and *SbbHLH125* was in the leaves. Three genes (*SbbHLH045*, *SbbHLH047* and *SbbHLH130*) showed highest expression in the *S. bicolor* stems, and the highest expression of *SbbHLH101* was found in fruit. In addition, correlations of *SbbHLH* expression among the six organs were studied (Figure 8B). We found that the expression of different genes in the plant organs was significantly correlated, indicating that their roles may be synergistic. Most *SbbHLH* genes showed significant positive correlations; for example, we observed four genes—*SbbHLH050*, *SbbHLH079*, *SbbHLH014* and *SbbHLH134*—that had their highest expression in the styles, and were significantly positively correlated; they also showed significant positive correlations with *SbbHLH110*, which is most highly expressed in the anthers. However, four pairs of *SbbHLH* genes (*SbbHLH050* and *SbbHLH125*; *SbbHLH110* and *SbbHLH045*; *SbbHLH079* and *SbbHLH045*; *SbbHLH045* and *SbbHLH134*) were significantly negatively correlated.

Expression patterns of *SbbHLH* genes in response to different treatments

To further determine whether the expression of *SbbHLH* genes was influenced by different abiotic stresses, the expression of 12 *SbbHLH* members was examined under six abiotic stresses: strong ultraviolet (UV) radiation, flooding, polyethylene glycol (PEG), NaCl, heat and cold treatments. qRT-PCR analysis was performed to analyze the 12 *SbbHLH* genes' expression patterns in roots, leaves and stems in response to the different treatments (Figure 9A). Some of the *SbbHLH* genes were significantly induced or repressed by the different treatments. Expression of most of these genes was significantly altered in the early stage of the stress treatment (Figure 9). Some *SbbHLH*s showed changes in expression with time or in different organs, depending on the stress. For example, under cold stress, *SbbHLH037* and *SbbHLH045* were first significantly upregulated, and then downregulated. *SbbHLH063* expression was significantly upregulated in the root, while it was significantly downregulated in the stem and leaf. Under flooding stress, *SbbHLH045* was significantly upregulated in the root, stem and leaf, but *SbbHLH050* was

significantly downregulated. Interestingly, several genes showed opposing expression patterns under different treatments. The transcript levels of many *SbbHLH* genes, such as *SbbHLH063*, were upregulated in stems and leaves by the heat-stress treatment, but downregulated by the cold-stress treatment. Some other genes showed changes in specific organs. For instance, *SbbHLH014* responded significantly to heat treatment in the root. Furthermore, correlations between *SbbHLH* gene-expression patterns were observed (Figure 9B). There were negative correlations among most *SbbHLH* genes. However, a few *SbbHLH* genes were significantly positively correlated, such as *SbbHLH110* and *SbbHLH063*/*SbbHLH134*, with $P < 0.05$ (Figure 9B).

Discussion

Exploration of the *bHLH* gene family at the whole-genome level in any species, and the functional identification of some this family's members, can provide theoretical support for the role of the *bHLH* gene family in the stress signal-transduction process. In this study, 174 *SbbHLH* genes were identified, and all of the encoded proteins showed obvious differences in structure, indicating high complexity. According to Atchley et al. [11] and Toledo-Ortiz et al. [15], we analyzed the DNA-binding ability of the basic region of *SbbHLH*s. The *SbbHLH* gene sequence can be divided into E-box binding genes and non-E-box binding genes (Additional file 1: Table S1). The E-box binding proteins can be subdivided into G-box binding proteins and non-G-box binding proteins [12, 15]. The basic domain of bHLH contains two essential amino acid residues, Glu-13 and Arg-16. If it contains only one of them, it will be classified as a non-E-box binding protein. The G-box binding protein contains three essential amino acid residues (His/Lys-9, Glu-13 and Arg-17) in the basic domain. If only Glu-13 and Arg-17 are present, it is classified as a non-G-box binding protein. In addition, if the number of basic amino acids is less than 4 in the basic domain, and it contains no or only one of Glu-13 and Arg-16, it will be classified as a non-DNA binding protein. These proteins are thought to have no ability to bind directly to DNA. In this study, 119 (68.4%) SbbHLHs were classified as E-box binding proteins: 99 (56.9%) as G-box binding proteins and 20 (11.5%) as non-G-box binding proteins; 30 (17.2%) members were classified as non-E-box binding genes. The remaining 25 (14.4%) members were not considered

436 to have DNA-binding ability due to the lack of Glu-13 or Arg-16 in the alkaline region (Figure 2,
437 Additional file 1: Table S1). Similar to the reports of *O. sativa* (95, 56.9%) and *A. thaliana* (89,
438 60.5%), the highest proportion of *SbbHLH*s were G-box binding proteins [18]. Previous studies
439 have found that some key amino acid residues play important roles in the binding of TFs to DNA
440 and the formation of homodimers or heterodimers between bHLHs or bHLHs and other TFs [15,
441 72]. For example, His-6, Glu-10, and Arg-14 are related to DNA-binding activity, whereas Leu-25
442 and Leu-57 in the helical region determine whether bHLH TFs can form homodimers or
443 heterodimers. In SbbHLHs, the conservation rates of Leu-25 and Leu-57 are 94.3% and 96.0%,
444 respectively, which are lower than in *S. lycopersicum* (99%, 97%) [47] and *Citrus reticulata*
445 (100%, 100%) [73]. Previous studies have found that the formation of such heterodimers can
446 change or expand the diversity of molecular interactions, and generate new functions by
447 identifying new DNA-binding sites [15, 74]. As already noted, a bHLH protein can form a
448 homodimer with itself or a heterodimer with other TFs, such as R2R3-MYBs, BAR1-BES1 and
449 AP2 [75, 76, 77].

450 Based on the constructed phylogenetic tree, we identified at least one bHLH protein from *S.*
451 *bicolor* in each subgroup of AtbHLHs, indicating that the time of differentiation of the bHLH
452 family may have been earlier than that of *S. bicolor* and *A. thaliana*. Compared to *A. thaliana*,
453 *SbbHLH* genes can be divided into 24 subfamilies and 1 orphan subfamily (UC), 4 more than *A.*
454 *thaliana* and 3 more than *O. sativa*. Among them, group 15 (17, 9.8%) and group 18 (15, 8.6%)
455 have more members, which is similar to the results for *A. thaliana* [18], and indicates that those
456 *bHLH* gene groups may have undergone stronger partial differentiation in the long-term
457 evolutionary process. However, there is no research to prove that this kind of differentiation is
458 advantageous in the differentiation process of herbs and woody plants. Seven of the *SbbHLH*s did
459 not have obvious clusters, so they were all classified into the UC group, and those genes all
460 showed non-DNA binding activity, but still a great deal of variability in the base sequence. The
461 gene-structure analysis revealed that *SbbHLH* genes in the same subfamily have similar gene
462 structures, which not only supports our classification of the subgroups to a certain extent, but also
463 indicates that all members of a subfamily are close in evolutionary terms(Figure 3). At the same
464 time, this does not rule out the loss of some independent introns during the long-term evolutionary

development of plants, resulting in the loss of some introns in the domains of some bHLH members. For example, *SbbHLH153* has fewer introns than other members of the same family. Genome-replication events are considered to have occurred in the process of plant evolution, and the expansion of gene families and genome evolution mechanisms mainly depend on gene-replication events [78, 79, 80]. The main replication modes are tandem repeats and fragment replication. These were identified in the *SbbHLH* genes. We discovered 13 tandem duplication events containing 20 *SbbHLH* genes (Figure 4, Additional file 3: Table S3), especially on chromosomes 7 and 8. In addition, there were 42 pairs of segmental duplications of *SbbHLH* genes (Figure 5, Additional file 4: Table S4). Therefore, segment duplication may make a higher contribution to the expansion of the bHLH family in *S. bicolor*. Nevertheless, since there were many duplication events in *S. bicolor*, it is lower than that of the dicotyledonous plants tomato and potato[26, 48]. Similar situations have been reported in studies of other monocotyledonous species [81]. However, the current conclusions cannot explain the significant differences between monocotyledons and dicotyledons.

Analysis of a gene's expression profile can provide important clues to understanding its potential biological function. There are many members of the bHLH TF family with diverse functions, but the current research in plants is not particularly thorough, as it focuses mainly on the two model plants *A. thaliana* and *O. sativa*. The functions of bHLH TFs in other species still need to be explored. In this study, we used 12 *SbbHLH* genes with significant differences in clustering on the phylogenetic tree to study their responses to six abiotic stresses in different developmental organs, and found that almost all of the *bHLH* TF genes have significant differential expression (more than 2-fold difference). For example, under salt stress, 10 *SbbHLH*s were upregulated in leaves, 7 were upregulated in roots, and 8 were upregulated in stems. The expression pattern results indicate that bHLH TFs participate in a complex cross-regulatory network. *SbbHLH079* and *SbbHLH045* were responsive, at the same time, to PEG, NaCl and UV treatments, indicating synergistic or antagonistic regulation under a variety of adverse conditions. Further research is needed to explore the relationship between these genes. Interestingly, most of the *SbbHLH* genes showed significant negative regulation in the expression heat map. If we consider expression patterns and complex protein interactions, then we can suggest that a network

of feedback mechanisms coordinates the expression of multiple genes. In addition, flowers and fruit, as plant reproductive organs, are the main structures in all angiosperms [82]. In this study, we explored the expression of 12 *bHLH* genes in the anthers and styles of *S. bicolor* flowers, as well as in the main organs of plants at the filling stage. Studies have shown that bHLH TFs play an important role in the development of flowers and fruit. The expression levels of *SbbHLH134* and *SbbHLH110* in the anther and style were significantly higher than in roots, stems, leaves and fruit, whereas *SbbHLH101* showed significantly higher expression in fruit at the filling stage (Figure 8A). Therefore, we speculate that *SbbHLH134*, *SbbHLH110* and *SbbHLH101* may also regulate flower and fruit development in *S. bicolor*. However, the specific functions still need to be analyzed through in-depth experiments. In summary, these results reveal the functions and regulation of some bHLH TFs.

Conclusion

In summary, we provided the systematic genome-wide analysis of the *bHLH* gene family in *S. bicolor*. A total of 174 *SbbHLH* genes/proteins were characterized and divided into 24 groups. Furthermore, protein motifs and gene structures of the *SbbHLH*s within the subfamilies were prone to be the similar, which supported the classification predicted. The distribution of the 174 *SbbHLH* genes on 10 *S. bicolor* chromosomes was uneven. We found that gene-replication events may have produced some *SbbHLH* genes, with tandem duplication contributing more to the expansion of the *SbbHLH* gene family than segmental duplication. The qRT-PCR results showed that the 12 studied *SbbHLH*s were all affected by abiotic stresses, and their expression during the development of flowers and fruit was studied. It is speculated that *SbbHLH134*, *SbbHLH110* and *SbbHLH101* also regulate flower and fruit development in *S. bicolor*. Taken together, the results and information described in this work provide a good basis for further investigation of the biological functions and evolution of *bHLH* genes in *S. bicolor*.

Methods

Gene identification

We downloaded the complete *S. bicolor* genome sequence from the Ensembl Genomes website

(http://ensemblgenomes.org/). Based on two BLASTp searches [83, 84], bHLH family members were identified. First, with BLASTp (score value ≥ 100 and e-value ≤ 1e- 10), all possible bHLH proteins were identified from the *S. bicolor* genome referring to trihelix protein sequences of *A. thaliana*. Second, the Hidden Markov Model (HMM) profile consistent with the trihelix domain was obtained from the Pfam protein family database (http://pfam.sanger.ac.uk/). We used both HMMER3.0 (default parameters) with a cutoff of 0.01 (http://plants.ensembl. org/hmmer/index.html) [85] and SMART (http://smart. embl-heidelberg.de/) [86, 87] to ascertain the presence of the bHLH domain, and to further verify the results. In addition, the basic features of the trihelix proteins of the *SbbHLH* gene family were identified: coding sequence length, pI, protein molecular mass, and subcellular localization, from the ExPasy website (http://web.expasy.org/protparam/).

***bHLH* gene structure**

The bHLH domain sequences of the characterized SbbHLH proteins were used to create multiple protein sequence alignments using ClustalW with default parameters [88]. The deduced amino acid sequences in the bHLH domains were then adjusted manually using GeneDoc software. We used Gene Structure Display Server (GSDS: http://GSDS.cbi.pku.edu.cn)[89] to analyze the constituents of the exons/introns of the *SbbHLH* genes. We used MEME to analyze the motifs of SbbHLH proteins, (http://meme-suite.org/tools/meme) [90, 91]. The optimized parameters were as follows: number of repetitions, any; the maximum number of motifs, 10; and the optimum width of each motif, between 6 and 200 residues [84, 91, 92].

Chromosomal distribution and gene duplication

All *SbbHLH* genes were mapped to *S. bicolor* chromosomes based on physical location information from the database of the *S. bicolor* genome using Circos [93]. The Multiple Collinearity Scan toolkit (MCScanX) was adopted to analyze the gene-duplication events, with the default parameters [94]. We analyzed homology of the *bHLH* genes between *S. bicolor* and the other six plants (*A. thaliana*, *V. vinifera*, *S. lycopersicum*, *B. distachyon*, *O. sativa* subsp. *indica*, *Z. mays*) using Dual Synteny Plotter (https://github.com/CJ-Chen/TBtools). Non-synonymous (ka)

and synonymous (ks) substitutions of each duplicated *bHLH* gene were calculated using Ka/Ks-Calculator 2.0 [95].

Phylogenetic analysis and classification of *SbbHLH* gene family

According to the classification of *AtbHLH*s, all of the identified *SbbHLH* genes were divided into groups. The phylogenetic trees were inferred using the NJ method of MEGA X via Geneious R11 with BLOSUM62 cost matrix, the Jukes–Cantor model, global alignment with free end gaps and bootstrap value of 1000. The full-length amino acid sequences of the bHLH proteins (Additional file 1: Table S1) derived from *A. thaliana*, *V. vinifera*, *S. lycopersicum*, *B. distachyon*, *O. sativa* subsp. *indica*, and *Z. mays* (UniProthttps://www.uniprot.org/), combined with newly identified *SbbHLH*s, were used for phylogenetic analysis.

Plant materials, growth conditions, and abiotic stress in *S. bicolor*

Sorghum bicolor cv. Hongyingzi, a typical cultivated variety, was used throughout the study. Since 2019, 'Hongyingzi' has been grown in the greenhouse of Guizhou University. *S. bicolor* was grown in pots filled with soil and vermiculite (1:1) in a growth room with a 16 h/25°C day and 8 h/20°C night regime, and a relative humidity of 75%. We collected the stems, roots, leaves, fruit, anthers and styles separately from five plants with good growth and similar growth conditions, and quickly placed them in liquid nitrogen for storage at -80°C pending further use. To investigate gene-expression patterns in response to various stresses, several *SbbHLH* genes were selected for further analysis. *S. bicolor* plants were subjected to the following abiotic stress treatments at the seedling stage (21 days): salt (5% NaCl), water flooding (whole plant), drought (30% PEG6000), UV exposure (70 μW/cm^2, 220 V, 30 W), high temperature (40°C), and low temperature (4°C); each stress treatment was performed with five replicates, and qRT-PCR analysis was carried out after sampling at 2 h and 24 h, respectively. The collected samples were stored at -80°C for further analysis.

Total RNA extraction, cDNA reverse transcription and qRT-PCR analysis

Total RNA of each sample was extracted with a plant RNA extraction kit (TaKaRa) and used

for cDNA library construction. The sequencing was performed in an Illumina GAII sequencer following the manufacturer's instructions [91, 92]. Gene-expression analysis was performed by qRT-PCR, with primers designed by Primer 5.0 (Additional file 6: Table S6). We used the *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) gene, which was stably expressed at each growth stage in almost all tissues, as an internal control [96]. The qRT-PCR with SYBR Premix Ex Taq II (TaKaRa) was repeated at least three times and the data were analyzed using the $2^{-(\Delta\Delta Ct)}$ method [97].

Statistical analysis

Analysis of variance (ANOVA) was performed with JMP6.0 software (SAS Institute), and compared with least significant difference (LSD) at the 0.05 and 0.01 levels. The histogram was drawn with Origin 8.0 software (OriginLab).

References

[1] Riaño-Pachón DM, Ruzicic S, Dreyer I, Mueller-Roeber B. PlnTFDB: an integrative plant transcription factor database. BMC Bioinformatics. 2007, 8, 42.

[2] Zhang H, Jin J, Tang L, Zhao Y, Gu X, Gao G, Luo J. PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. Nucleic Acids Res. 2011, 39, D1114-D1117.

[3] Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science. 2000,290 (5499), 2105-2110.

[4] Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol. 2003, 20 (9), 1377-1419.

[5] Schwechheimer C, Zourelidou M, Bevan MW. PLANT TRANSCRIPTION FACTOR STUDIES. Annu Rev Plant Physiol Plant Mol Biol. 1998, 49, 127-150.

[6] Jin J, Zhang H, Kong L, Gao G, Luo J. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res. 2014, 42, D1182-D1187.

610 [7] Sun X, Wang Y, Sui N. Transcriptional regulation of bHLH during plant response to stress.
611 Biochem Biophys Res Commun. 2018, 5, 503 (2), 397-401.
612 [8] Buck MJ, Atchley WR. Phylogenetic analysis of plant basic helix-loop-helix proteins. J Mol
613 Evol. 2003, 56 (6), 742-750.
614 [9] Ledent V, Vervoort M. The basic helix-loop-helix protein family: comparative genomics and
615 phylogenetic analysis. Genome Res. 2001, 11 (5), 754-770.
616 [10] Murre C, McCaw PS, Baltimore D. A new DNA binding and dimerization motif in
617 immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell. 1989, 56 (5),
618 777-783.
619 [11] Atchley WR, Terhalle W, Dress A. Positional dependence, cliques, and predictive motifs in
620 the bHLH protein domain. J Mol Evol. 1999, 48 (5), 501-516.
621 [12] Massari ME, Murre C. Helix-loop-helix proteins: regulators of transcription in eucaryotic
622 organisms. Mol Cell Biol. 2000, 20 (2), 429-440.
623 [13] Nair SK, Burley SK. Recognizing DNA in the library. Nature. 2000, 13; 404 (6779), 715,
624 717-718.
625 [14] Shimizu T, Toumoto A, Ihara K, Shimizu M, Kyogoku Y, Ogawa N, Oshima Y, Hakoshima T.
626 Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. EMBO J.
627 1997, 1, 16 (15): 4689-4697.
628 [15] Toledo-Ortiz G, Huq E, Quail PH. The Arabidopsis basic/helix-loop-helix transcription factor
629 family. Plant Cell. 2003, 15 (8), 1749-1770.
630 [16] Atchley WR, Fitch WM. A natural classification of the basic helix-loop-helix class of
631 transcription factors. Proc Natl Acad Sci U S A. 1997, 13; 94 (10), 5172-5176.
632 [17] Pires N, Dolan L. Origin and diversification of basic-helix-loop-helix proteins in plants. Mol
633 Biol Evol. 2010, 27 (4), 862-874.
634 [18] Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, Guo J, Liang W, Chen L, Yin J, Ma H, Wang J,
635 Zhang D. Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and
636 Arabidopsis. Plant Physiol. 2006, 141 (4), 1167-1184.
637 [19] Amoutzias GD, Robertson DL, Oliver SG, Bornberg-Bauer E. Convergent evolution of gene
638 networks by single-gene duplications in higher eukaryotes. EMBO Rep. 2004, 5 (3), 274-279.

639 [20] Stevens JD, Roalson EH, Skinner MK. Phylogenetic and expression analysis of the basic
640 helix-loop-helix transcription factor gene family: genomic approach to cellular differentiation.
641 Differentiation. 2008, 76 (9), 1006-1022.
642 [21] Roig-Villanova I, Bou-Torrent J, Galstyan A, Carretero-Paulet L, Portolés S,
643 Rodríguez-Concepción M, Martínez-García JF. Interaction of shade avoidance and auxin
644 responses: a role for two novel atypical bHLH proteins. EMBO J. 2007, 14; 26 (22), 4756-4767.
645 [22] Leivar P, Monte E, Al-Sady B, Carle C, Storer A, Alonso JM, Ecker JR, Quail PH. The
646 Arabidopsis phytochrome-interacting factor PIF7, together with PIF3 and PIF4, regulates
647 responses to prolonged red light by modulating phyB levels. Plant Cell. 2008, 20 (2), 337-352.
648 [23] Friedrichsen DM, Nemhauser J, Muramitsu T, Maloof JN, Alonso J, Ecker JR, Furuya M,
649 Chory J. Three redundant brassinosteroid early response genes encode putative bHLH
650 transcription factors required for normal growth. Genetics. 2002, 162 (3), 1445-1456.
651 [24] Lu R, Zhang J, Liu D, Wei YL, Wang Y, Li XB. Characterization of bHLH/HLH genes that
652 are involved in brassinosteroid (BR) signaling in fiber development of cotton (Gossypium
653 hirsutum). BMC Plant Biol. 2018, 27, 18 (1), 304.
654 [25] Onohata T, Gomi K. Overexpression of jasmonate-responsive OsbHLH034 in rice results in
655 the induction of bacterial blight resistance via an increase in lignin biosynthesis. Plant Cell Rep.
656 2020, 39 (9), 1175-1184.
657 [26] Wang R, Zhao P, Kong N, Lu R, Pei Y, Huang C, Ma H, Chen Q. Genome-Wide
658 Identification and Characterization of the Potato bHLH Transcription Factor Family. Genes
659 (Basel). 2018 , 22, 9 (1), 54.
660 [27] Fu Y, Win P, Zhang H, Li C, Shen Y, He F, Luo K. PtrARF2.1 Is Involved in Regulation of
661 Leaf Development and Lignin Biosynthesis in Poplar Trees. Int J Mol Sci. 2019, 24; 20 (17),
662 4141.
663 [28] Li Z, Liu C, Zhang Y, Wang B, Ran Q, Zhang J. The bHLH family member ZmPTF1
664 regulates drought tolerance in maize by promoting root development and abscisic acid synthesis. J
665 Exp Bot. 2019, 15, 70 (19), 5471-5486.
666 [29] Chen HC, Hsieh-Feng V, Liao PC, Cheng WH, Liu LY, Yang YW, Lai MH, Chang MC. The
667 function of OsbHLH068 is partially redundant with its homolog, *AtbHLH112*, in the regulation of

668 the salt stress response but has opposite functions to control flowering in Arabidopsis. Plant Mol
669 Biol. 2017, 94 (4-5), 531-548.
670 [30] Yuan Y, Wu H, Wang N, Li J, Zhao W, Du J, Wang D, Ling HQ. FIT interacts with *AtbHLH38*
671 and *AtbHLH39* in regulating iron uptake gene expression for iron homeostasis in Arabidopsis. Cell
672 Res. 2008, 18 (3), 385-397.
673 [31] Liu Y, Ji X, Nie X, Qu M, Zheng L, Tan Z, Zhao H, Huo L, Liu S, Zhang B, Wang Y.
674 Arabidopsis *AtbHLH112* regulates the expression of genes involved in abiotic stress tolerance by
675 binding to their E-box and GCG-box motifs. New Phytol. 2015, 207 (3), 692-709.
676 [32] Zhao Q, Xiang X, Liu D, Yang A, Wang Y. Tobacco Transcription Factor NtbHLH123
677 Confers Tolerance to Cold Stress by Regulating the NtCBF Pathway and Reactive Oxygen Species
678 Homeostasis. Front Plant Sci. 2018, 28, 9, 381.
679 [33] Matus JT, Poupin MJ, Cañón P, Bordeu E, Alcalde JA, Arce-Johnson P. Isolation of WDR and
680 bHLH genes related to flavonoid synthesis in grapevine (*Vitis vinifera* L.). Plant Mol Biol. 2010,
681 72 (6), 607-620.
682 [34] Quattrocchio F, Wing JF, van der Woude K, Mol JN, Koes R. Analysis of bHLH and MYB
683 domain proteins: species-specific regulatory differences are caused by divergent evolution of
684 target anthocyanin genes. Plant J. 1998, 13 (4), 475-488.
685 [35] Chen F , Hu Y , Vannozzi A , Wu KC , Cai HY，Qin Y，Mullis A，Lin ZG，Zhang LS. The
686 WRKY Transcription Factor Family in Model Plants and Crops. Critical Reviews in Plant
687 Sciences, 2018, 36 (5), 1-25.
688 [36] Ali TM , Hasnain A . Morphological, Physicochemical, and Pasting Properties of Modified
689 White Sorghum (Sorghum bicolor) Starch. International Journal of Food Properties, 2014, 17 (3),
690 523-535.
691 [37] Pelpolage S W , Han K , Koaze H , et al. Influence of enzyme-resistant fraction of sorghum
692 (Sorghum bicolor L.) flour on gut microflora composition, short-chain fatty acid production and
693 toxic substance metabolism[J]. Journal of food and nutrition research, 2019, 58 (2), 135-145.
694 [38] Xiong Y , Zhang P , Warner R D , et al. Sorghum Grain: From Genotype, Nutrition, and
695 Phenolic Profile to Its Health Benefits and Food Applications[J]. Comprehensive Reviews in Food
696 ence and Food Safety, 2019 (1).

[39] Zhao ZY, Che P, Glassman K, Albertsen M. Nutritionally Enhanced Sorghum for the Arid and Semiarid Tropical Areas of Africa. Methods Mol Biol. 2019, 1931, 197-207.

[40] Han Y, Song L, Liu S, Zou N, Li Y, Qin Y, Li X, Pan C. Simultaneous determination of 124 pesticide residues in Chinese liquor and liquor-making raw materials (sorghum and rice hull) by rapid Multi-plug Filtration Cleanup and gas chromatography-tandem mass spectrometry. Food Chem. 2018, 15, 241, 258-267.

[41] Prasad PV, Djanaguiraman M, Perumal R, Ciampitti IA. Impact of high temperature stress on floret fertility and individual grain weight of grain sorghum: sensitive stages and thresholds for temperature and duration. Front Plant Sci. 2015, 6, 6, 820.

[42] Prasad PVV , Pisipati SR , Mutava RN , et al. Sensitivity of Grain Sorghum to High Temperature Stress during Reproductive Development[J]. Crop Science, 2008, 48 (5), 1911-1917.

[43] Tsuji W , Ali MEK , Inanaga S , et al. Growth and Gas Exchange of Three Sorghum Cultivars Under Drought Stress[J]. Biologia Plantarum, 2003, 46 (4), 583-587.

[44] Rooney W L . Sorghum Improvement—Integrating Traditional and New Technology to Produce Improved Genotypes[J]. Advances in Agronomy, 2004, 83, 37-109.

[45] Li H , Payne W A , Michels G J , et al. Reducing plant abiotic and biotic stress: Drought and attacks of greenbugs, corn leaf aphids and virus disease in dryland sorghum. Environmental and Experimental Botany, 2008, 63 (1-3), 305-316.

[46] Paterson AH, Bowers JE, Bruggmann R, et al. The Sorghum bicolor genome and the diversification of grasses. Nature. 2009, 29, 457 (7229), 551-556.

[47] 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.

[48] Sun H, Fan HJ, Ling HQ. Genome-wide identification and characterization of the bHLH gene family in tomato. BMC Genomics. 2015, 22, 16 (1),9.

[49] Kavas M, Baloğlu MC, Atabay ES, Ziplar UT, Daşgan HY, Ünver T. Genome-wide characterization and expression analysis of common bean bHLH transcription factors in response to excess salt concentration. Mol Genet Genomics. 2016, 291 (1), 129-143.

[50] Mao K, Dong Q, Li C, Liu C, Ma F. Genome Wide Identification and Characterization of

726 Apple bHLH Transcription Factors and Expression Analysis in Response to Drought and Salt
727 Stress. Front Plant Sci. 2017, 11, 8, 480.
728 [51] Gao C, Sun J, Wang C, Dong Y, Xiao S, Wang X, Jiao Z. Genome-wide analysis of
729 basic/helix-loop-helix gene family in peanut and assessment of its roles in pod development. PLoS
730 One. 2017, 27, 12 (7), e0181843.
731 [52] Niu X, Guan Y, Chen S, Li H. Genome-wide analysis of basic helix-loop-helix (bHLH)
732 transcription factors in Brachypodium distachyon. BMC Genomics. 2017,15,18 (1),619.
733 [53] Zhang T, Lv W, Zhang H, Ma L, Li P, Ge L, Li G. Genome-wide analysis of the basic
734 Helix-Loop-Helix (bHLH) transcription factor family in maize. BMC Plant Biol. 2018, 16, 18 (1),
735 235.
736 [54] Wei K, Chen H. Comparative functional genomics analysis of bHLH gene family in rice,
737 maize and wheat. BMC Plant Biol. 2018, 29, 18 (1), 309.
738 [55] Cheng X, Xiong R, Liu H, Wu M, Chen F, Hanwei Yan, Xiang Y. Basic helix-loop-helix gene
739 family: Genome wide identification, phylogeny, and expression in Moso bamboo. Plant Physiol
740 Biochem. 2018, 132, 104-119.
741 [56] Yingqi H, Ahmad N, Yuanyuan T, Jianyu L, Liyan W, Gang W, Xiuming L, Yuanyuan D,
742 Fawei W, Weican L, Xiaowei L, Xu Z, Na Y, Haiyan L. Genome-Wide Identification, Expression
743 Analysis, and Subcellular Localization of Carthamus tinctorius bHLH Transcription Factors. Int J
744 Mol Sci. 2019, 21, 20 (12), 3044.
745 [57] Li H, Gao W, Xue C, Zhang Y, Liu Z, Zhang Y, Meng X, Liu M, Zhao J. Genome-wide
746 analysis of the bHLH gene family in Chinese jujube (Ziziphus jujuba Mill.) and wild jujube. BMC
747 Genomics. 2019, 10; 20 (1), 568.
748 [58] Zhang Z, Chen J, Liang C, Liu F, Hou X, Zou X. Genome-Wide Identification and
749 Characterization of the bHLH Transcription Factor Family in Pepper (Capsicum annuum L.).
750 Front Genet. 2020 , 25; 11, 570156.
751 [59] Zhu L, Zhao M, Chen M, Li L, Jiang Y, Liu S, Jiang Y, Wang K, Wang Y, Sun C, Chen J,
752 Chen P, Lei J, Su Y, Wang Y, Zhang M. The bHLH gene family and its response to saline stress in
753 Jilin ginseng, Panax ginseng C.A. Meyer. Mol Genet Genomics. 2020, 295 (4), 877-890.
754 [60] Aslam M, Jakada BH, Fakher B, Greaves JG, Niu X, Su Z, Cheng Y, Cao S, Wang X, Qin Y.

755 Genome-wide study of pineapple (Ananas comosus L.) bHLH transcription factors indicates that
756 cryptochrome-interacting bHLH2 (AcCIB2) participates in flowering time regulation and abiotic
757 stress response. BMC Genomics. 2020, 22; 21 (1), 735.
758 [61] Sun W, Jin X, Ma Z, Chen H, Liu M. Basic helix-loop-helix (bHLH) gene family in Tartary
759 buckwheat (Fagopyrum tataricum): Genome-wide identification, phylogeny, evolutionary
760 expansion and expression analyses. Int J Biol Macromol. 2020, 15, 155, 1478-1490.
761 [62] Zhao Y , Li X , Chen W , et al. Whole-genome survey and characterization of MADS-box
762 gene family in maize and sorghum[J]. Plant Cell Tissue and Organ Culture, 2011, 105 (2),
763 159-173.
764 [63] Kushwaha H, Gupta S, Singh VK, Rastogi S, Yadav D. Genome wide identification of Dof
765 transcription factor gene family in sorghum and its comparative phylogenetic analysis with rice
766 and Arabidopsis. Mol Biol Rep. 2011, 38 (8), 5037-5053.
767 [64] Zhang C, Bian M, Yu H, Liu Q, Yang Z. Identification of alkaline stress-responsive genes of
768 CBL family in sweet sorghum (Sorghum bicolor L.). Plant Physiol Biochem. 2011, 49 (11),
769 1306-1312.
770 [65] Yan HW, Hong L, Zhou YQ, Jiang HY, Zhu SW, Fan J, Cheng BJ. A genome-wide analysis of
771 the ERF gene family in sorghum. Genet Mol Res. 2013, 21, 12 (2), 2038-2055.
772 [66] Chang JZ, Yan FX, Qiao LY, Zheng J, Zhang FY, Liu QS. Genome-wide identification and
773 expression analysis of SBP-box gene family in Sorghum bicolor L. Yi Chuan. 2016, 20, 38 (6),
774 569-580.
775 [67] Nagaraju M, Reddy PS, Kumar SA, Kumar A, Rajasheker G, Rao DM, Kavi Kishor PB.
776 Genome-wide identification and transcriptional profiling of small heat shock protein gene family
777 under diverse abiotic stress conditions in Sorghum bicolor (L.). Int J Biol Macromol. 2020, 1, 142,
778 822-834.
779 [68] Nagaraju M, Kumar SA, Reddy PS, Kumar A, Rao DM, Kavi Kishor PB. Genome-scale
780 identification, classification, and tissue specific expression analysis of late embryogenesis
781 abundant (LEA) genes under abiotic stress conditions in Sorghum bicolor L. PLoS One. 2019, 16,
782 14 (1), e0209980.
783 [69] Sanjari S, Shirzadian-Khorramabad R, Shobbar ZS, Shahbazi M. Systematic analysis of NAC

784 transcription factors' gene family and identification of post-flowering drought stress responsive
785 members in sorghum. Plant Cell Rep. 2019, 38 (3), 361-376.
786 [70] Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, Coornaert D, Degnan BM,
787 Vervoort M. Origin and diversification of the basic helix-loop-helix gene family in metazoans:
788 insights from comparative genomics. BMC Evol Biol. 2007, 2, 7:33.
789 [71] Henriksson M, Lüscher B. Proteins of the Myc network: essential regulators of cell growth
790 and differentiation. Adv Cancer Res. 1996, 68, 109-182.
791 [72] Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR,
792 Robertson DL. Genome-wide classification and evolutionary analysis of the bHLH family of
793 transcription factors in Arabidopsis, poplar, rice, moss, and algae. Plant Physiol. 2010, 153 (3),
794 1398-1412.
795 [73] Geng J, Liu JH. The transcription factor CsbHLH18 of sweet orange functions in modulation
796 of cold tolerance and homeostasis of reactive oxygen species by regulating the antioxidant gene. J
797 Exp Bot. 2018, 27, 69 (10), 2677-2692.
798 [74] Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC. The basic
799 helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and
800 functional diversity. Mol Biol Evol. 2003, 20 (5), 735-747.
801 [75] Dubos C, Le Gourrierec J, Baudry A, Huep G, Lanet E, Debeaujon I, Routaboul JM, Alboresi
802 A, Weisshaar B, Lepiniec L. MYBL2 is a new regulator of flavonoid biosynthesis in Arabidopsis
803 thaliana. Plant J. 2008, 55 (6), 940-953.
804 [76] Chandler JW, Cole M, Flier A, Werr W. BIM1, a bHLH protein involved in brassinosteroid
805 signalling, controls Arabidopsis embryonic patterning via interaction with DORNROSCHEN and
806 DORNROSCHEN-LIKE. Plant Mol Biol. 2009, 69 (1-2), 57-68.
807 [77] Yin Y, Vafeados D, Tao Y, Yoshida S, Asami T, Chory J. A new class of transcription factors
808 mediates brassinosteroid-regulated gene expression in Arabidopsis. Cell. 2005, 28, 120 (2),
809 249-259.
810 [78] Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. Evolution's cauldron: duplication,
811 deletion, and rearrangement in the mouse and human genomes. Proc Natl Acad Sci U S A. 2003,

812 30, 100 (20), 11484-11489.
813 [79] Cannon SB, Mitra A, Baumgarten A, Young ND, May G. The roles of segmental and tandem
814 gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol.
815 2004, 1, 4, 10.
816 [80] Mehan MR, Freimer NB, Ophoff RA. A genome-wide survey of segmental duplications that
817 mediate common human genetic variation of chromosomal architecture. Hum Genomics. 2004, 1
818 (5), 335-344.
819 [81] Hudson KA, Hudson ME. A classification of basic helix-loop-helix transcription factors of
820 soybean. Int J Genomics. 2015, 2015, 603182.
821 [82] Gremski K, Ditta G, Yanofsky MF. The HECATE genes regulate female reproductive tract
822 development in Arabidopsis thaliana. Development. 2007, 134 (20), 3593-3601.
823 [83] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped
824 BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids
825 Res. 1997, 1, 25 (17), 3389-3402.
826 [84] Liu M, Ma Z, Wang A, Zheng T, Huang L, Sun W, Zhang Y, Jin W, Zhan J, Cai Y, Tang Y, Wu
827 Q, Tang Z, Bu T, Li C, Chen H. Genome-Wide Investigation of the Auxin Response Factor Gene
828 Family in Tartary Buckwheat (*Fagopyrum tataricum*). Int J Mol Sci. 2018, 9, 19 (11), 3526.
829 [85] Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity
830 searching. Nucleic Acids Res. 2011, 39 (Web Server issue), W29-W37.
831 [86] Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL. The Pfam protein
832 families database. Nucleic Acids Res. 2000, 1, 28 (1), 263-266.
833 [87] Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids
834 Res. 2018, 4, 46 (D1), D493-D496.
835 [88] Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and
836 ClustalX. Curr Protoc Bioinformatics. 2002, Chapter 2: Unit 2. 3.
837 [89] Guo AY, Zhu QH, Chen X, Luo JC. [GSDS: a gene structure display server]. Yi Chuan. 2007,
838 29 (8), 1023-1026.
839 [90] Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS.
840 MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009, 37 (Web Server

841 issue), W202-W208.
842 [91] Xie T, Chen C, Li C, Liu J, Liu C, He Y. Genome-wide investigation of WRKY gene family
843 in pineapple: evolution and expression profiles during development and stress. BMC Genomics.
844 2018, 25, 19 (1), 490.
845 [92] Liu M, Ma Z, Sun W, Huang L, Wu Q, Tang Z, Bu T, Li C, Chen H. Genome-wide analysis of
846 the NAC transcription factor family in Tartary buckwheat (Fagopyrum tataricum). BMC
847 Genomics. 2019, 6, 20 (1), 113.
848 [93] Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA.
849 Circos: an information aesthetic for comparative genomics. Genome Res. 2009, 19 (9),
850 1639-1645.
851 [94] Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H,
852 Kissinger JC, Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene
853 synteny and collinearity. Nucleic Acids Res. 2012, 40 (7), e49.
854 [95] Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating
855 gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics. 2010,
856 8 (1), 77-80.
857 [96] Sudhakar Reddy P, Srinivas Reddy D, Sivasakthi K, Bhatnagar-Mathur P, Vadez V, Sharma
858 KK. Evaluation of Sorghum [*Sorghum bicolor* (L.)] Reference Genes in Various Tissues and under
859 Abiotic Stress Conditions for Quantitative Real-Time PCR Data Normalization. Front Plant Sci.
860 2016 , 25, 7, 529.
861 [97] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time
862 quantitative PCR and the 2 (-Delta Delta C (T)) Method. Methods. 2001, 25 (4), 402-408.
863
864 **Figure legends:**
865 **Figure 1** Unrooted phylogenetic tree showing relationships among bHLH domains of *S. bicolor*
866 and *Arabidopsis*. The phylogenetic tree was derived using the NJ method in MEGA7.0. The tree
867 shows the 24 phylogenetic subfamilies and 1 unclassified group (UC) marked with red font on a
868 white background. bHLH proteins from *Arabidopsis* are marked with the prefix ‘At’.
869 **Figure 2** Multiple sequence alignment of the bHLH domains of the members of 24 phylogenetic

subfamilies and 1 unclassified group (UC) of the SbbHLH protein family. The scheme at the top depicts the locations and boundaries of the basic, helix, and loop regions in the bHLH domain.

Figure 3 Phylogenetic relationships, gene-structure analysis, and motif distributions of *S. bicolor bHLH* genes.

A Phylogenetic tree was constructed by the NJ method with 1000 replicates on each node.

B Exons and introns are indicated by yellow rectangles and gray lines, respectively.

C Amino acid motifs in the SbbHLH proteins (1–10) are represented by colored boxes. The black lines indicate relative protein lengths.

Figure 4 Schematic representation of the chromosomal distribution of the *S. bicolor bHLH* genes. Vertical bars represent the chromosomes of *S. bicolor*. The chromosome number is indicated to the left of each chromosome. The scale on the left represents chromosome length.

Figure 5 Schematic representation of the chromosomal distribution and interchromosomal relationships of *S. bicolor bHLH* genes. Colored lines indicate all synteny blocks in the *S. bicolor* genome and the red lines indicate duplicated *bHLH* gene pairs. Chromosome number is indicated at the bottom of each chromosome.

Figure 6 Synteny analyses of the *bHLH* genes between *S. bicolor* and six representative plant species (*Arabidopsis thaliana*, *Vitis vinifera*, *Solanum lycopersicum*, *Brachypodium distachyon*, *Oryza sativa* subsp. *indica*, *Zea mays*). Gray lines on the background indicate the collinear blocks in *S. bicolor* and other plant genomes; red lines highlight the syntenic *S. bicolor bHLH* gene pairs.

Figure 7 Phylogenetic relationship and motif composition of the bHLH proteins from *S. bicolor* with six different plant species (*Arabidopsis thaliana*, *Vitis vinifera*, *Solanum lycopersicum*, *Brachypodium distachyon*, *Oryza sativa* subsp. *indica*, *Zea mays*).

Outer panel: An unrooted phylogenetic tree constructed using Geneious R11 with the NJ method.

Inner panel: Distribution of the conserved motifs in bHLH proteins. The differently colored boxes represent different motifs and their positions in each bHLH protein sequence. The sequence information for each motif is provided in Additional File 2: Table S2.

Figure 8 Tissue-specific gene expression of 12 *S. bicolor bHLH* genes and the correlation between their expression patterns.

A Expression patterns of 12 *S. bicolor bHLH* genes in the anther, style, leaf, root, stem and fruit

organs were examined by qRT-PCR. Error bars were obtained from three measurements. Lowercase letter above the bar indicates significant difference (α = 0.05, LSD) among the treatments.

B Positive number: positively correlated; negative number: negatively correlated. Red numbers indicate a significant correlation at the 0.05 level.

Figure 9 Gene expression of 12 *S. bicolor bHLH* genes in plants subjected to abiotic stresses (strong UV radiation, flooding, PEG, NaCl, heat and cold treatments) at the seedling stage.

A Expression patterns of 12 *S. bicolor bHLH* genes in leaf, root and stem organs were examined by qRT-PCR. Error bars were obtained from three measurements. Lowercase letter above the bar indicates significant difference (α = 0.05, LSD) among the treatments.

B Positive number: positively correlated; negative number: negatively correlated. Red numbers indicate a significant correlation at the 0.05 level.

Additional files:

Additional file 1: Table S1. List of the 174 *S. bicolor bHLH* genes identified in this study.

Additional file 2: Table S2. Analysis and distribution of the conserved motifs in *S. bicolor* bHLH proteins.

Additional file 3: Table S3. Tandem duplication events of *S. bicolor bHLH* genes.

Additional file 4: Table S4. The 42 pairs of segmental duplications in *S. bicolor bHLH* genes.

Additional file 5: Table S5. One-to-one orthologous genes relationships between *S. bicolor* and other plants.

Additional file 6: Table S6. Primer sequences for qRT-PCR.

Abbreviations:

bHLH: Basic helix-loop-helix

SbbHLH: *Sorghum bicolor* bHLH

GAPDH: Glyceraldehyde-3-phosphate dehydrogenase

qRT-PCR: Quantitative real-time polymerase chain reaction

TF: Transcription factor

928 AtbHLH: *Arabidopsis thaliana* bHLH

929 HMM: Hidden Markov Model

930 pI: Isoelectric point

931 LG: Linkage group

932

933

934

935

936

937

938

939

940

941

942

Figure 1

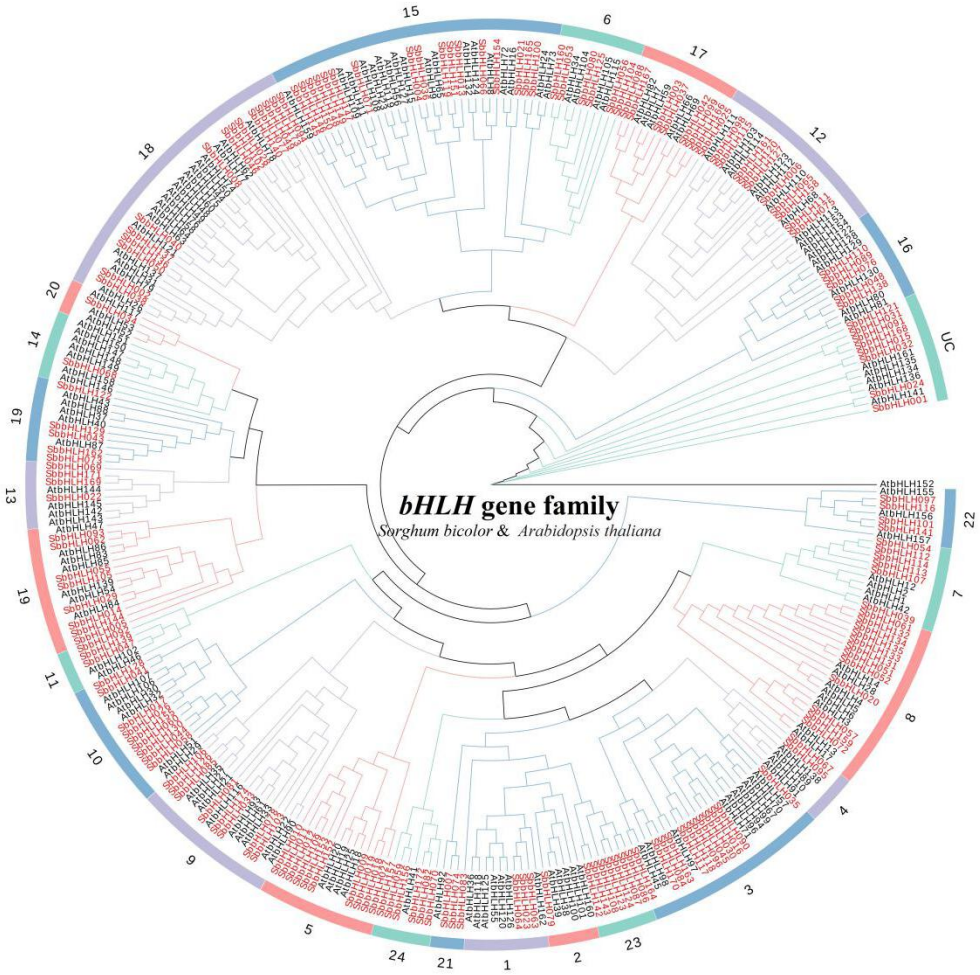
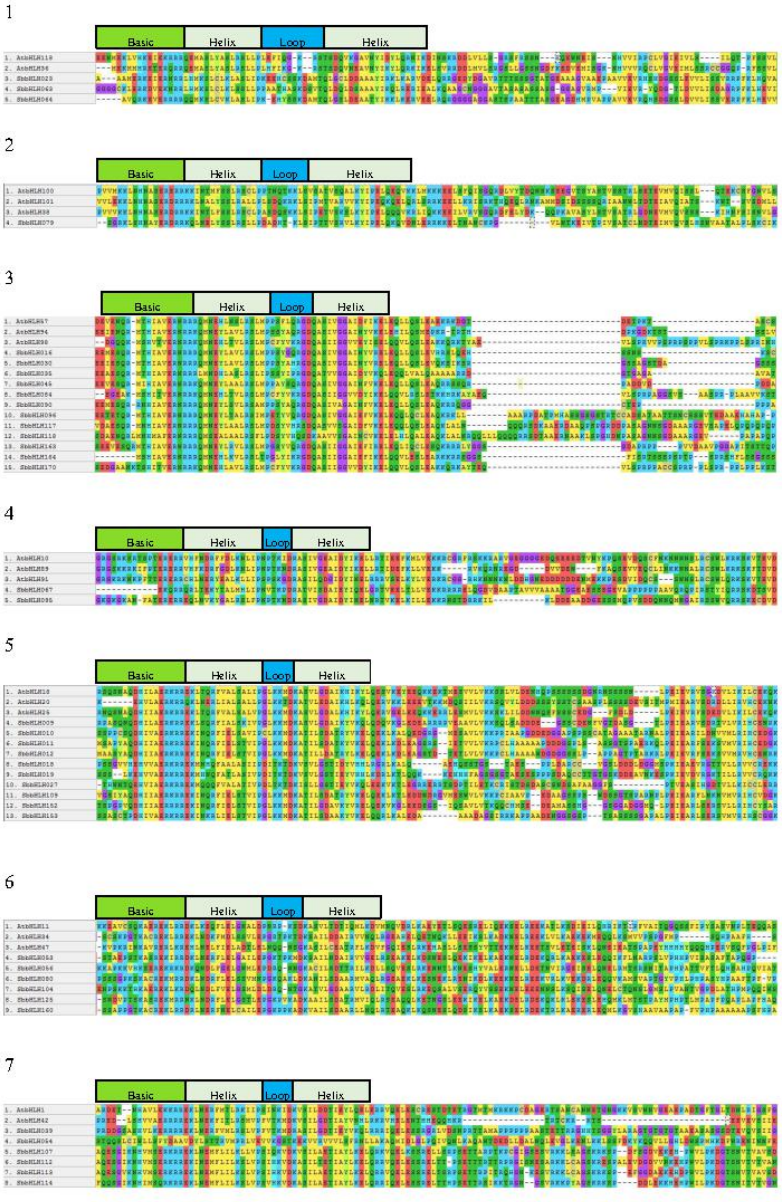
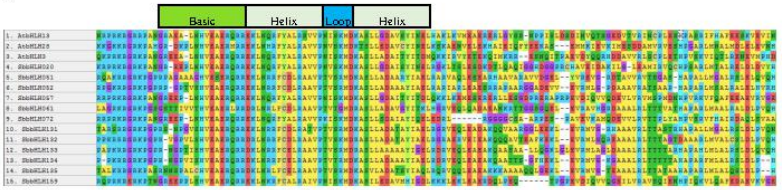


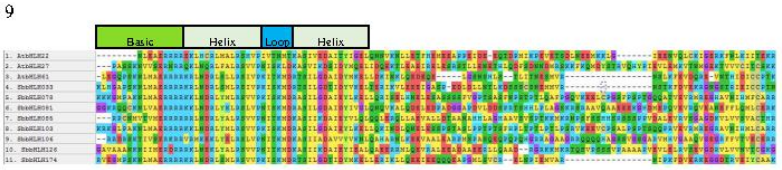
Figure 2


1
Basic
Helix
Loop
Helix
2
Basic
Helix
Loop
Helix
3
Basic
Helix
Loop
Helix
4
Basic
Helix
Loop
Helix
5
Basic
Helix
Loop
Helix
6
Basic
Helix
Loop
Helix
7
Basic
Helix
Loop
Helix

8

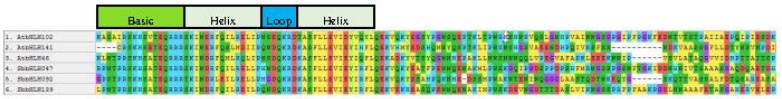


9

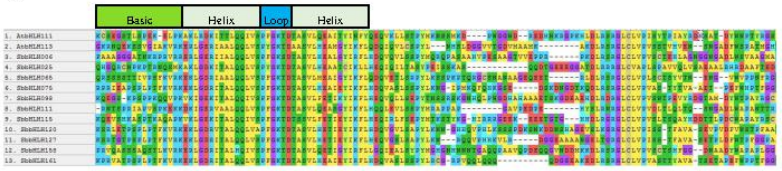


10

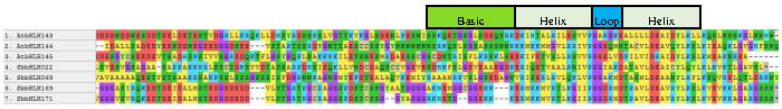
11



12



13

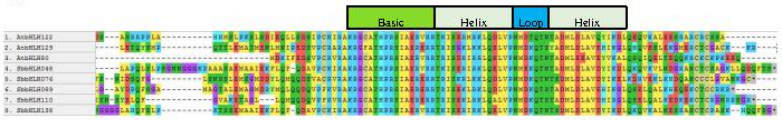


959

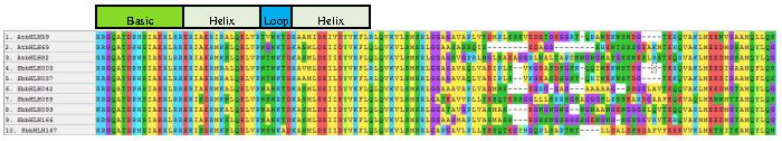
14

15

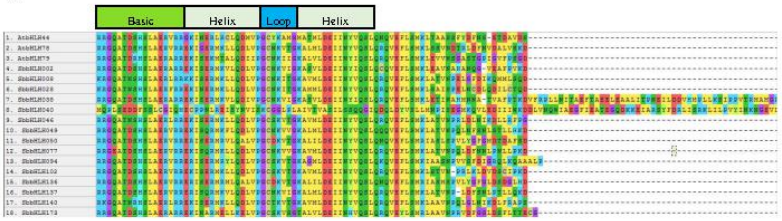
16



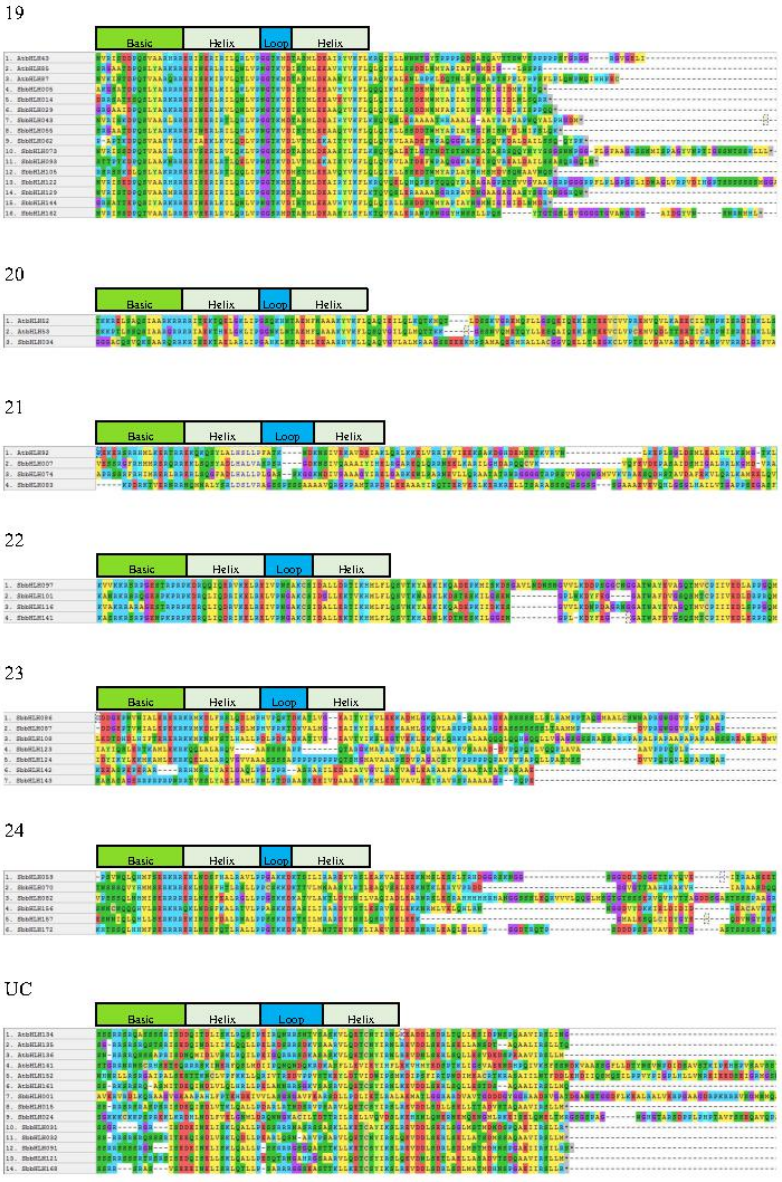
17



18



960



961

962

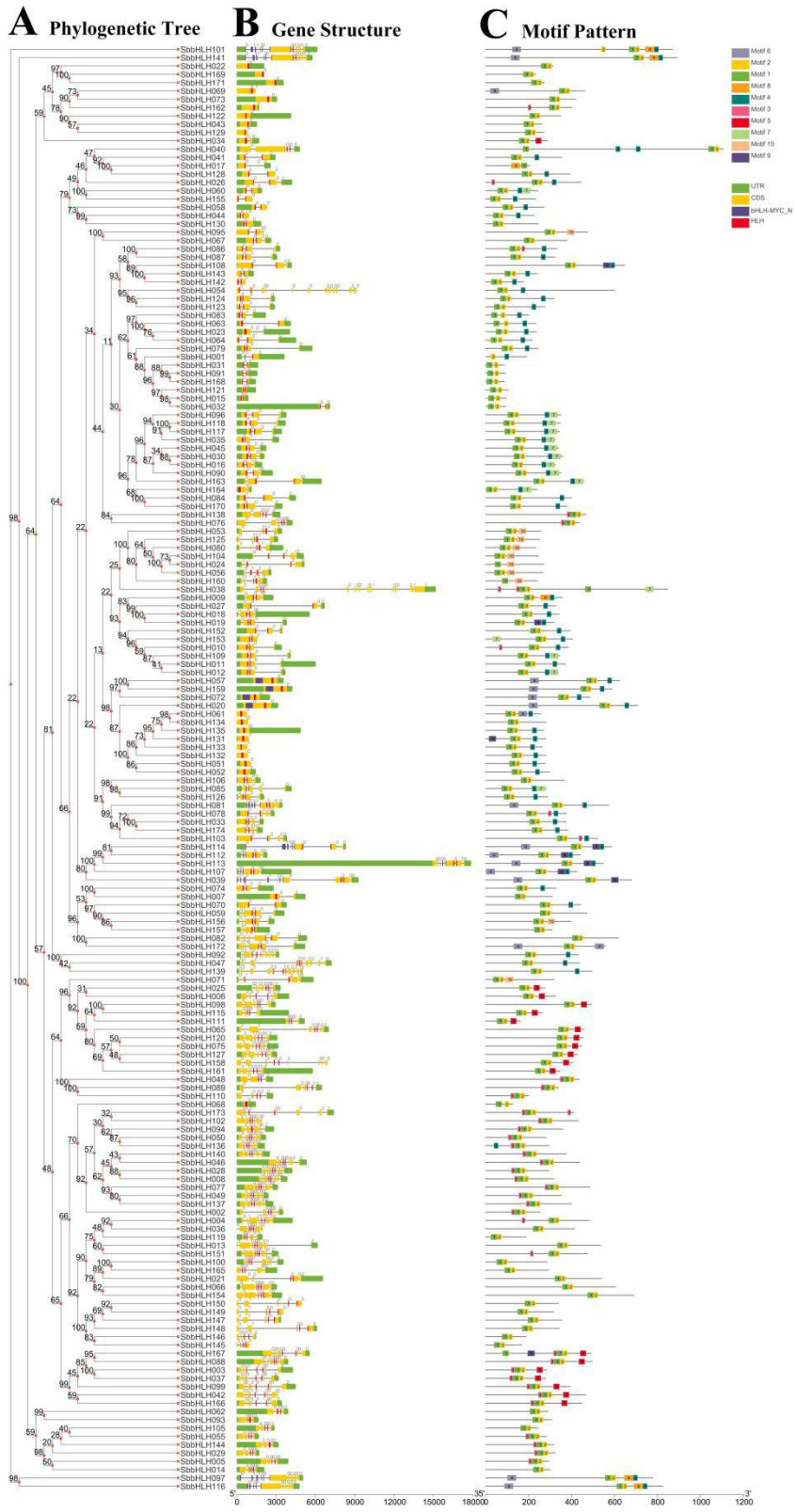
963

964

965

966

967 **Figure 3**



968

Figure 4

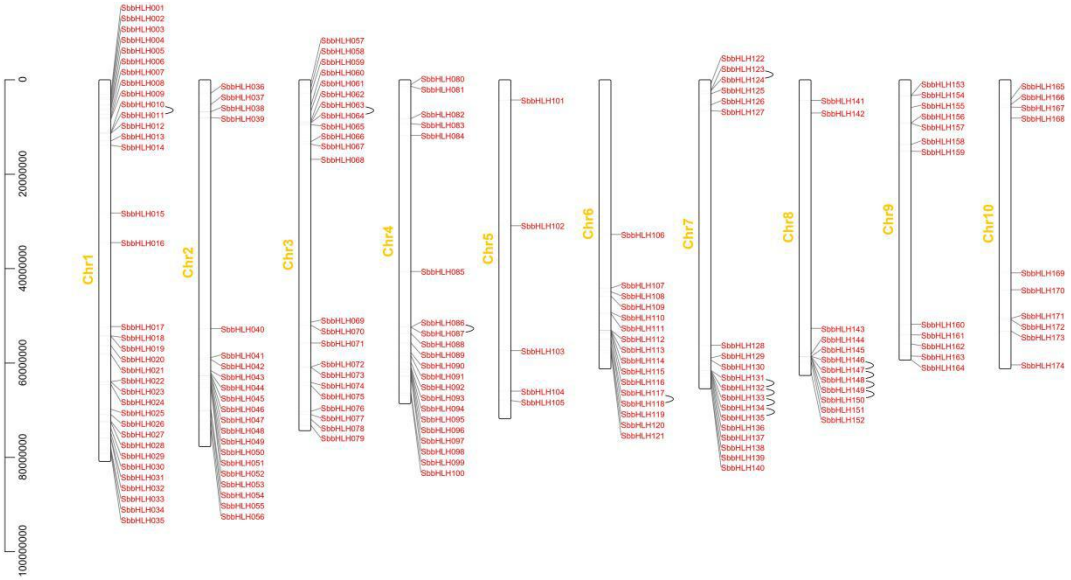

0
20000000
40000000
60000000
80000000
100000000
Chr1
SbbHLH001
SbbHLH002
SbbHLH003
SbbHLH004
SbbHLH005
SbbHLH006
SbbHLH007
SbbHLH008
SbbHLH009
SbbHLH010
SbbHLH011
SbbHLH012
SbbHLH013
SbbHLH014
SbbHLH015
SbbHLH016
SbbHLH017
SbbHLH018
SbbHLH019
SbbHLH020
SbbHLH021
SbbHLH022
SbbHLH023
SbbHLH024
SbbHLH025
SbbHLH026
SbbHLH027
SbbHLH028
SbbHLH029
SbbHLH030
SbbHLH031
SbbHLH032
SbbHLH033
SbbHLH034
SbbHLH035
Chr2
SbbHLH036
SbbHLH037
SbbHLH038
SbbHLH039
SbbHLH040
SbbHLH041
SbbHLH042
SbbHLH043
SbbHLH044
SbbHLH045
SbbHLH046
SbbHLH047
SbbHLH048
SbbHLH049
SbbHLH050
SbbHLH051
SbbHLH052
SbbHLH053
SbbHLH054
SbbHLH055
SbbHLH056
Chr3
SbbHLH057
SbbHLH058
SbbHLH059
SbbHLH060
SbbHLH061
SbbHLH062
SbbHLH063
SbbHLH064
SbbHLH065
SbbHLH066
SbbHLH067
SbbHLH068
SbbHLH069
SbbHLH070
SbbHLH071
SbbHLH072
SbbHLH073
SbbHLH074
SbbHLH075
SbbHLH076
SbbHLH077
SbbHLH078
SbbHLH079
Chr4
SbbHLH080
SbbHLH081
SbbHLH082
SbbHLH083
SbbHLH084
SbbHLH085
SbbHLH086
SbbHLH087
SbbHLH088
SbbHLH089
SbbHLH090
SbbHLH091
SbbHLH092
SbbHLH093
SbbHLH094
SbbHLH095
SbbHLH096
SbbHLH097
SbbHLH098
SbbHLH099
SbbHLH100
Chr5
SbbHLH101
SbbHLH102
SbbHLH103
SbbHLH104
SbbHLH105
Chr6
SbbHLH106
SbbHLH107
SbbHLH108
SbbHLH109
SbbHLH110
SbbHLH111
SbbHLH112
SbbHLH113
SbbHLH114
SbbHLH115
SbbHLH116
SbbHLH117
SbbHLH118
SbbHLH119
SbbHLH120
SbbHLH121
Chr7
SbbHLH122
SbbHLH123
SbbHLH124
SbbHLH125
SbbHLH126
SbbHLH127
SbbHLH128
SbbHLH129
SbbHLH130
SbbHLH131
SbbHLH132
SbbHLH133
SbbHLH134
SbbHLH135
SbbHLH136
SbbHLH137
SbbHLH138
SbbHLH139
SbbHLH140
Chr8
SbbHLH141
SbbHLH142
SbbHLH143
SbbHLH144
SbbHLH145
SbbHLH146
SbbHLH147
SbbHLH148
SbbHLH149
SbbHLH150
SbbHLH151
SbbHLH152
Chr9
SbbHLH153
SbbHLH154
SbbHLH155
SbbHLH156
SbbHLH157
SbbHLH158
SbbHLH159
SbbHLH160
SbbHLH161
SbbHLH162
SbbHLH163
SbbHLH164
Chr10
SbbHLH165
SbbHLH166
SbbHLH167
SbbHLH168
SbbHLH169
SbbHLH170
SbbHLH171
SbbHLH172
SbbHLH173
SbbHLH174

Figure 5

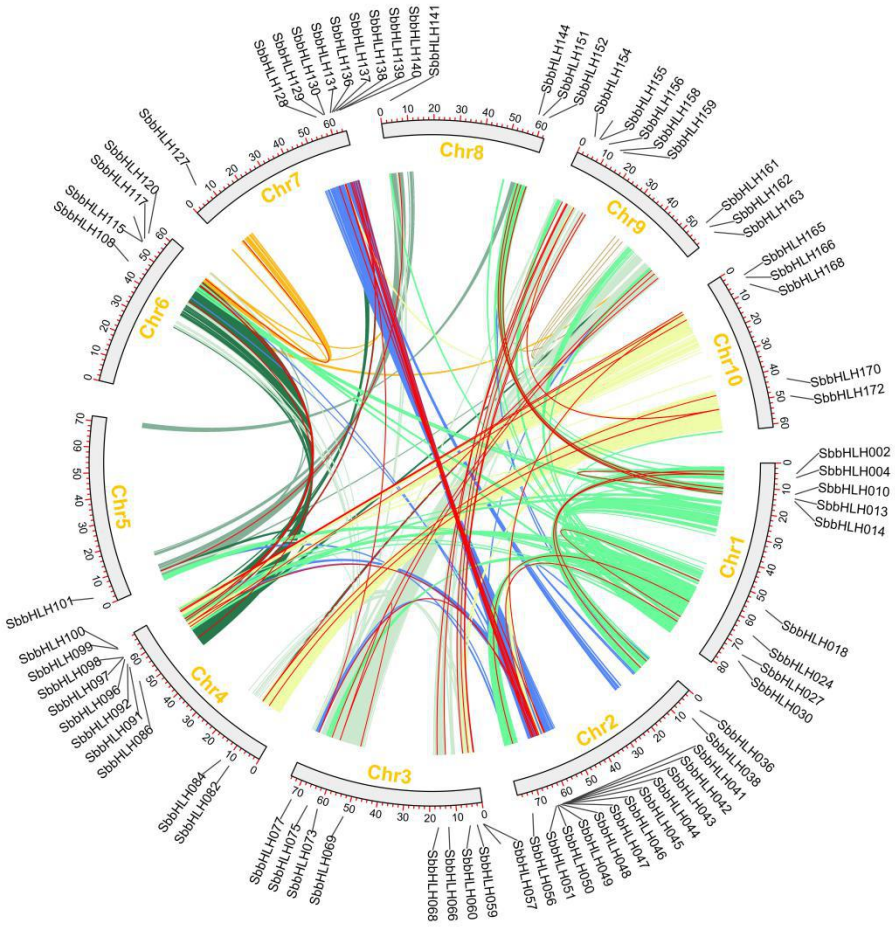

Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
SbbHLH002
SbbHLH004
SbbHLH010
SbbHLH013
SbbHLH014
SbbHLH018
SbbHLH024
SbbHLH027
SbbHLH030
SbbHLH036
SbbHLH038
SbbHLH041
SbbHLH042
SbbHLH043
SbbHLH044
SbbHLH045
SbbHLH046
SbbHLH047
SbbHLH048
SbbHLH049
SbbHLH050
SbbHLH051
SbbHLH056
SbbHLH057
SbbHLH059
SbbHLH060
SbbHLH066
SbbHLH068
SbbHLH069
SbbHLH073
SbbHLH075
SbbHLH077
SbbHLH082
SbbHLH084
SbbHLH086
SbbHLH091
SbbHLH092
SbbHLH096
SbbHLH097
SbbHLH098
SbbHLH099
SbbHLH100
SbbHLH101
SbbHLH109
SbbHLH115
SbbHLH117
SbbHLH120
SbbHLH127
SbbHLH128
SbbHLH129
SbbHLH130
SbbHLH131
SbbHLH136
SbbHLH137
SbbHLH138
SbbHLH139
SbbHLH140
SbbHLH141
SbbHLH144
SbbHLH151
SbbHLH152
SbbHLH154
SbbHLH155
SbbHLH156
SbbHLH158
SbbHLH159
SbbHLH161
SbbHLH162
SbbHLH163
SbbHLH165
SbbHLH166
SbbHLH168
SbbHLH170
SbbHLH172

Figure 6

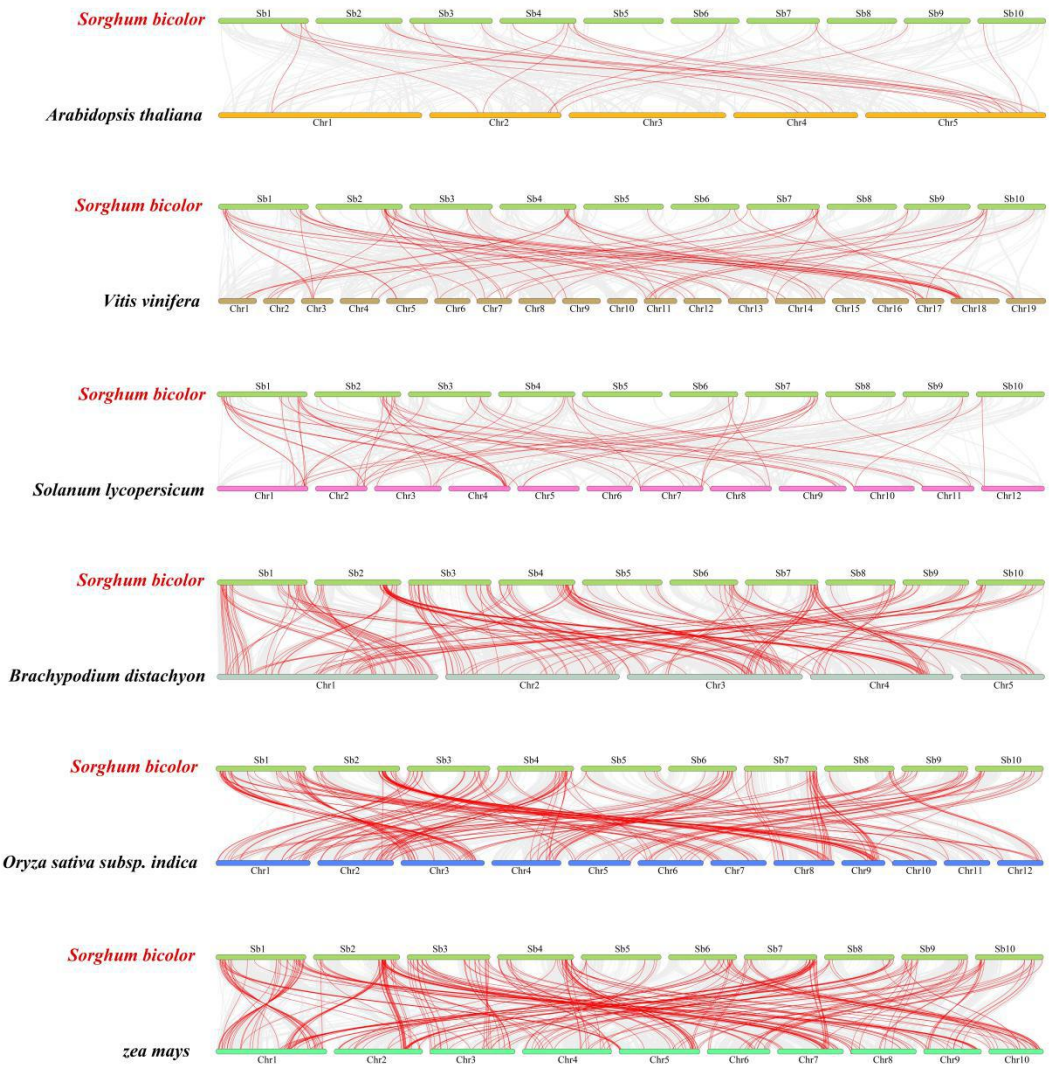

Sorghum bicolor
Sb1 Sb2 Sb3 Sb4 Sb5 Sb6 Sb7 Sb8 Sb9 Sb10
Arabidopsis thaliana
Chr1 Chr2 Chr3 Chr4 Chr5
Sorghum bicolor
Sb1 Sb2 Sb3 Sb4 Sb5 Sb6 Sb7 Sb8 Sb9 Sb10
Vitis vinifera
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12 Chr13 Chr14 Chr15 Chr16 Chr17 Chr18 Chr19
Sorghum bicolor
Sb1 Sb2 Sb3 Sb4 Sb5 Sb6 Sb7 Sb8 Sb9 Sb10
Solanum lycopersicum
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12
Sorghum bicolor
Sb1 Sb2 Sb3 Sb4 Sb5 Sb6 Sb7 Sb8 Sb9 Sb10
Brachypodium distachyon
Chr1 Chr2 Chr3 Chr4 Chr5
Sorghum bicolor
Sb1 Sb2 Sb3 Sb4 Sb5 Sb6 Sb7 Sb8 Sb9 Sb10
Oryza sativa subsp. indica
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12
Sorghum bicolor
Sb1 Sb2 Sb3 Sb4 Sb5 Sb6 Sb7 Sb8 Sb9 Sb10
zea mays
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10

Figure 7

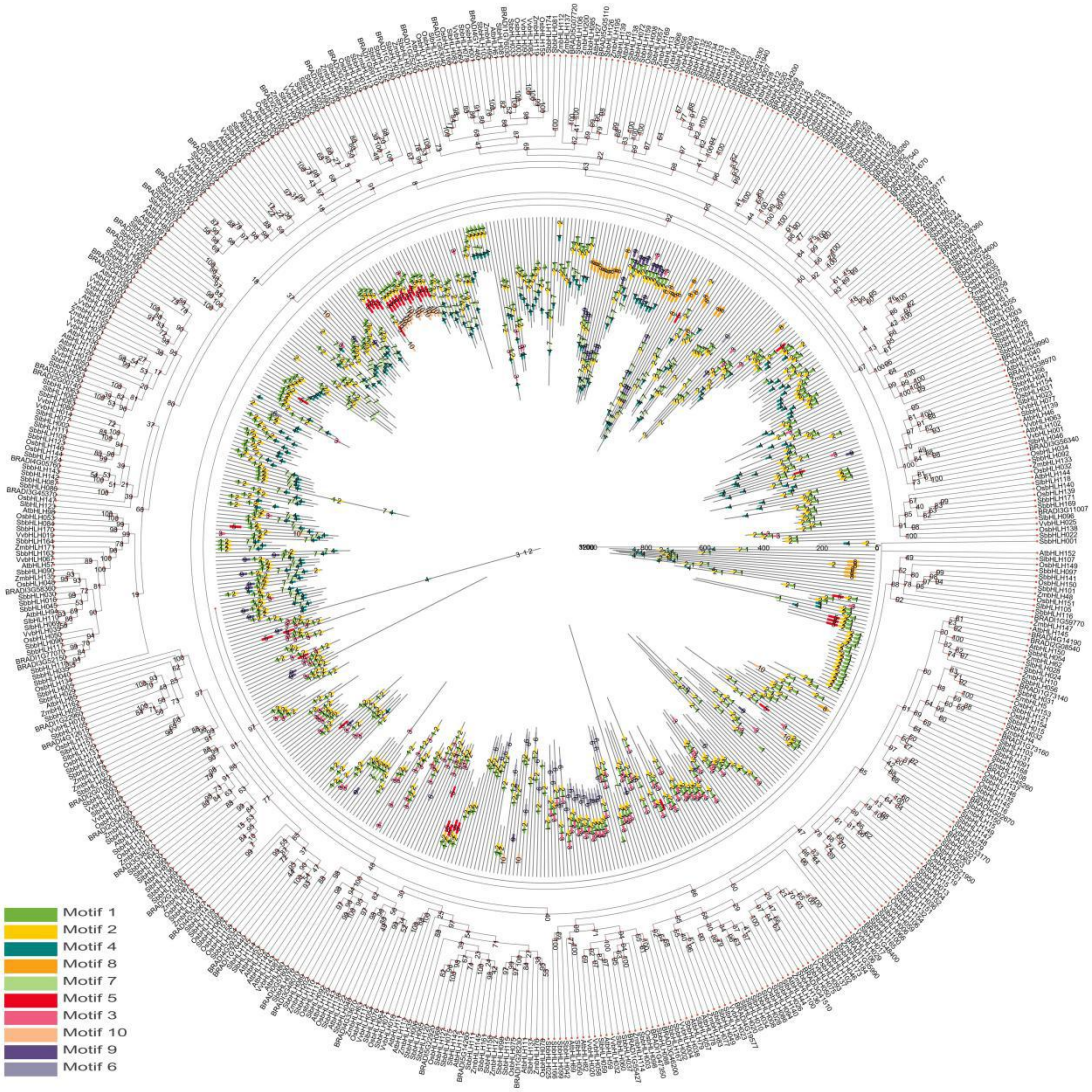

Motif 1
Motif 2
Motif 4
Motif 8
Motif 7
Motif 5
Motif 3
Motif 10
Motif 9
Motif 6

Figure 8A

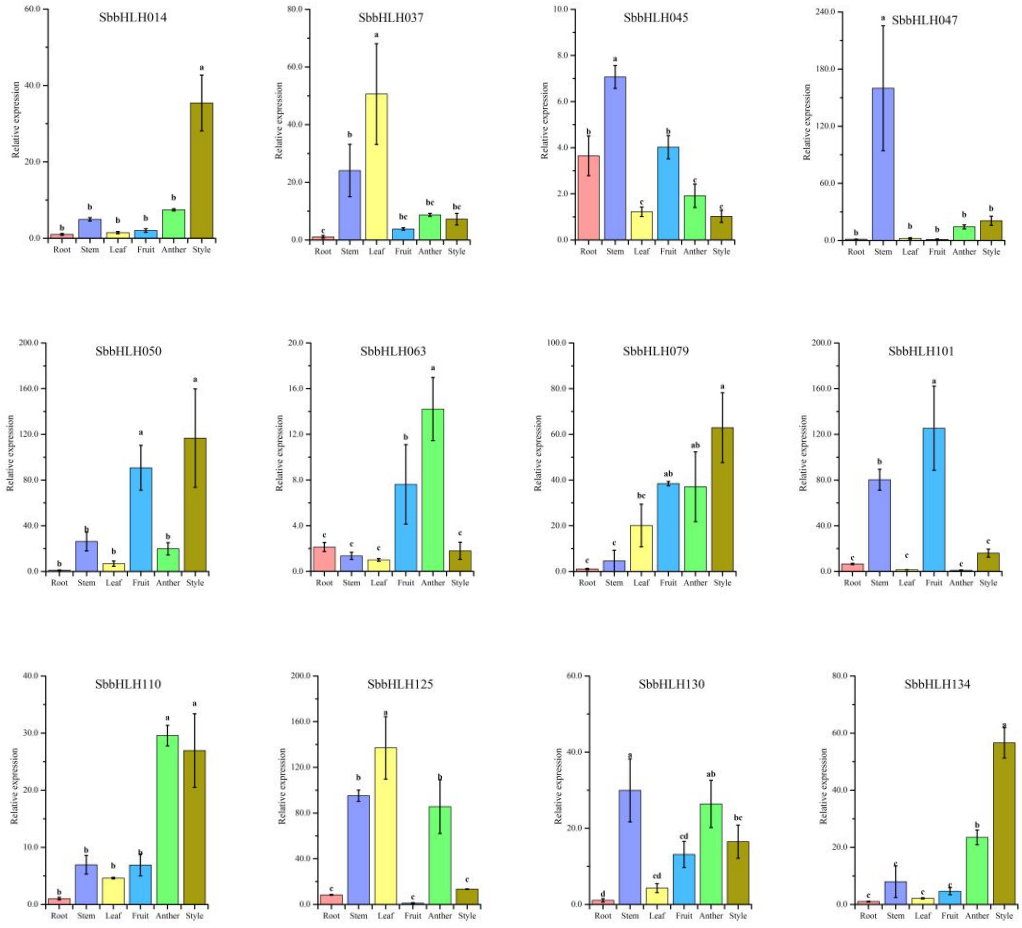

SbbHLH014
SbbHLH037
SbbHLH045
SbbHLH047
SbbHLH050
SbbHLH063
SbbHLH079
SbbHLH101
SbbHLH110
SbbHLH125
SbbHLH130
SbbHLH134
Relative expression
Root
Stem
Leaf
Fruit
Anther
Style

Figure 8B

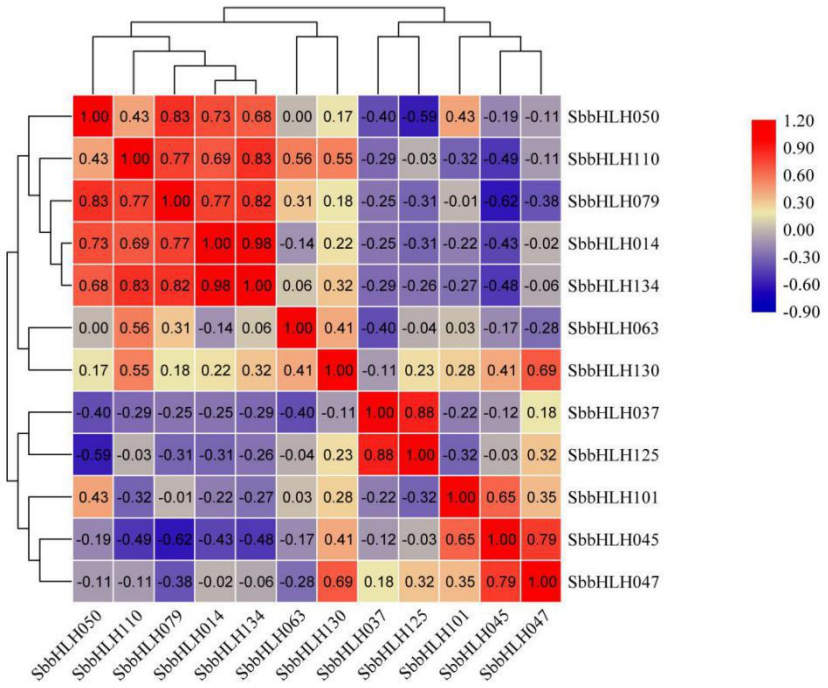
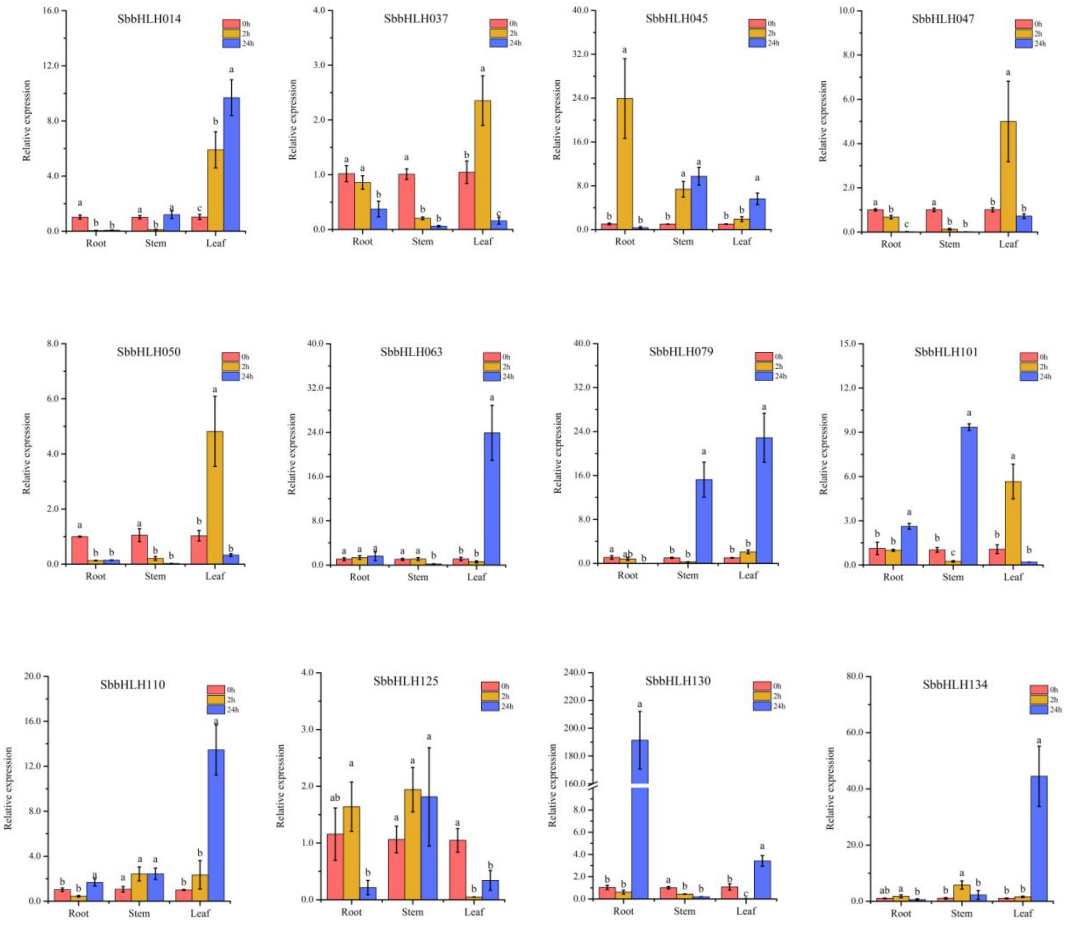
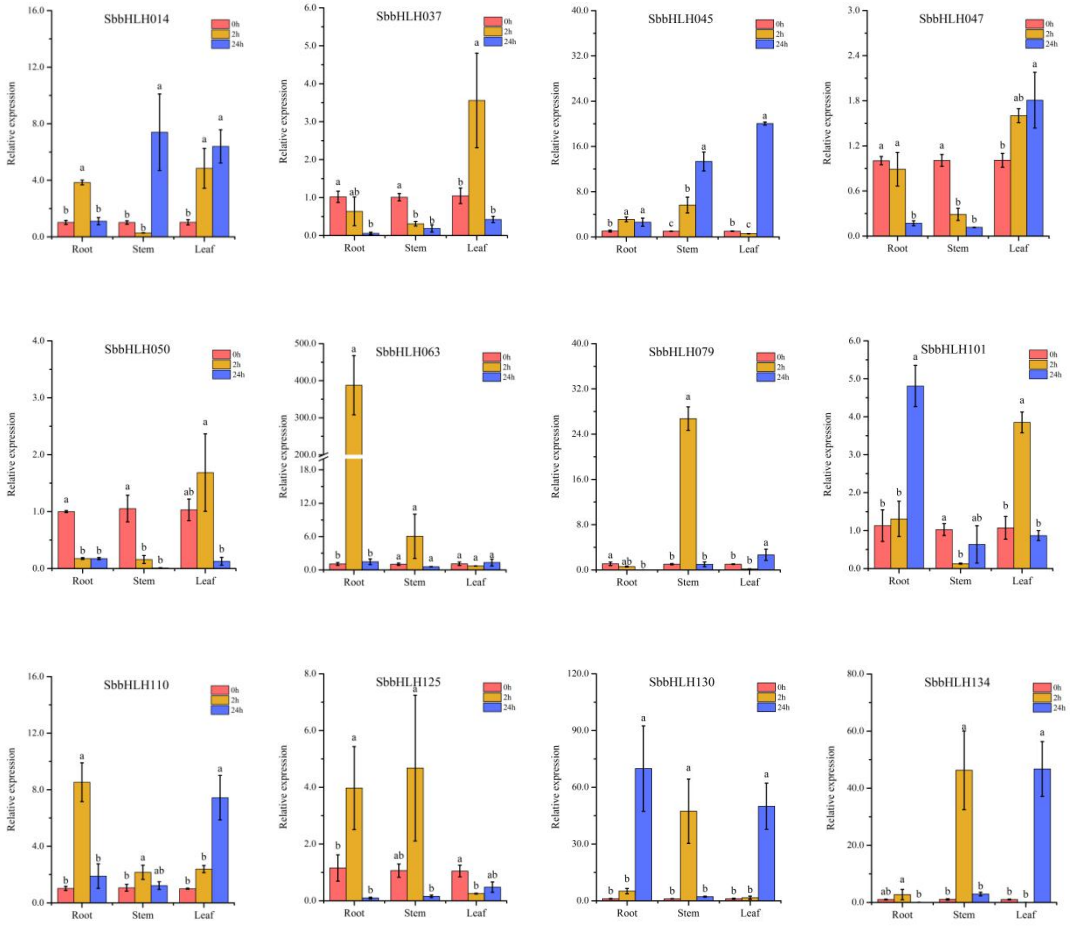

1.00 0.43 0.83 0.73 0.68 0.00 0.17 -0.40 -0.59 0.43 -0.19 -0.11 SbbHLH050
0.43 1.00 0.77 0.69 0.83 0.56 0.55 -0.29 -0.03 -0.32 -0.49 -0.11 SbbHLH110
0.83 0.77 1.00 0.77 0.82 0.31 0.18 -0.25 -0.31 -0.01 -0.62 -0.38 SbbHLH079
0.73 0.69 0.77 1.00 0.98 -0.14 0.22 -0.25 -0.31 -0.22 -0.43 -0.02 SbbHLH014
0.68 0.83 0.82 0.98 1.00 0.06 0.32 -0.29 -0.26 -0.27 -0.48 -0.06 SbbHLH134
0.00 0.56 0.31 -0.14 0.06 1.00 0.41 -0.40 -0.04 0.03 -0.17 -0.28 SbbHLH063
0.17 0.55 0.18 0.22 0.32 0.41 1.00 -0.11 0.23 0.28 0.41 0.69 SbbHLH130
-0.40 -0.29 -0.25 -0.25 -0.29 -0.40 -0.11 1.00 0.88 -0.22 -0.12 0.18 SbbHLH037
-0.59 -0.03 -0.31 -0.31 -0.26 -0.04 0.23 0.88 1.00 -0.32 -0.03 0.32 SbbHLH125
0.43 -0.32 -0.01 -0.22 -0.27 0.03 0.28 -0.22 -0.32 1.00 0.65 0.35 SbbHLH101
-0.19 -0.49 -0.62 -0.43 -0.48 -0.17 0.41 -0.12 -0.03 0.65 1.00 0.79 SbbHLH045
-0.11 -0.11 -0.38 -0.02 -0.06 -0.28 0.69 0.18 0.32 0.35 0.79 1.00 SbbHLH047
SbbHLH050
SbbHLH110
SbbHLH079
SbbHLH014
SbbHLH134
SbbHLH063
SbbHLH130
SbbHLH037
SbbHLH125
SbbHLH101
SbbHLH045
SbbHLH047
1.20
0.90
0.60
0.30
0.00
-0.30
-0.60
-0.90

Figure 9A

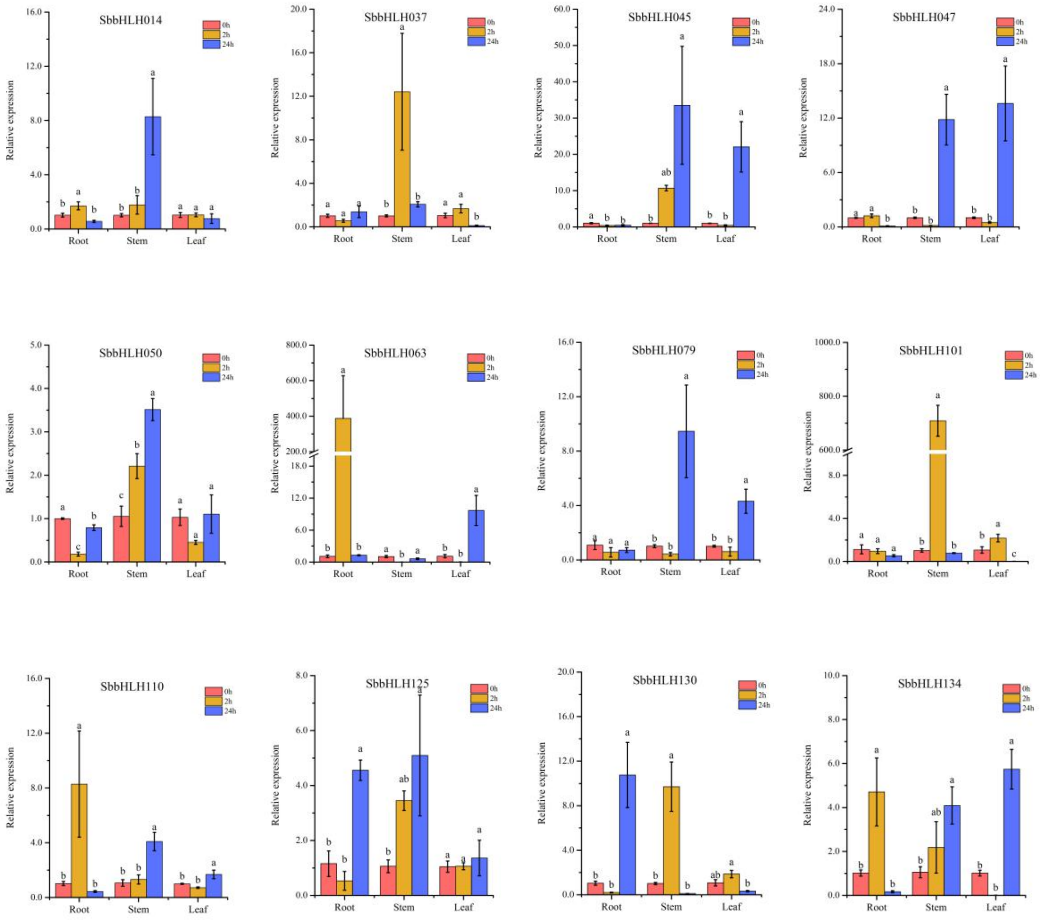
UV


SbbHLH014
SbbHLH037
SbbHLH045
SbbHLH047
SbbHLH050
SbbHLH063
SbbHLH079
SbbHLH101
SbbHLH110
SbbHLH125
SbbHLH130
SbbHLH134
Relative expression
Root
Stem
Leaf
0h
2h
24h

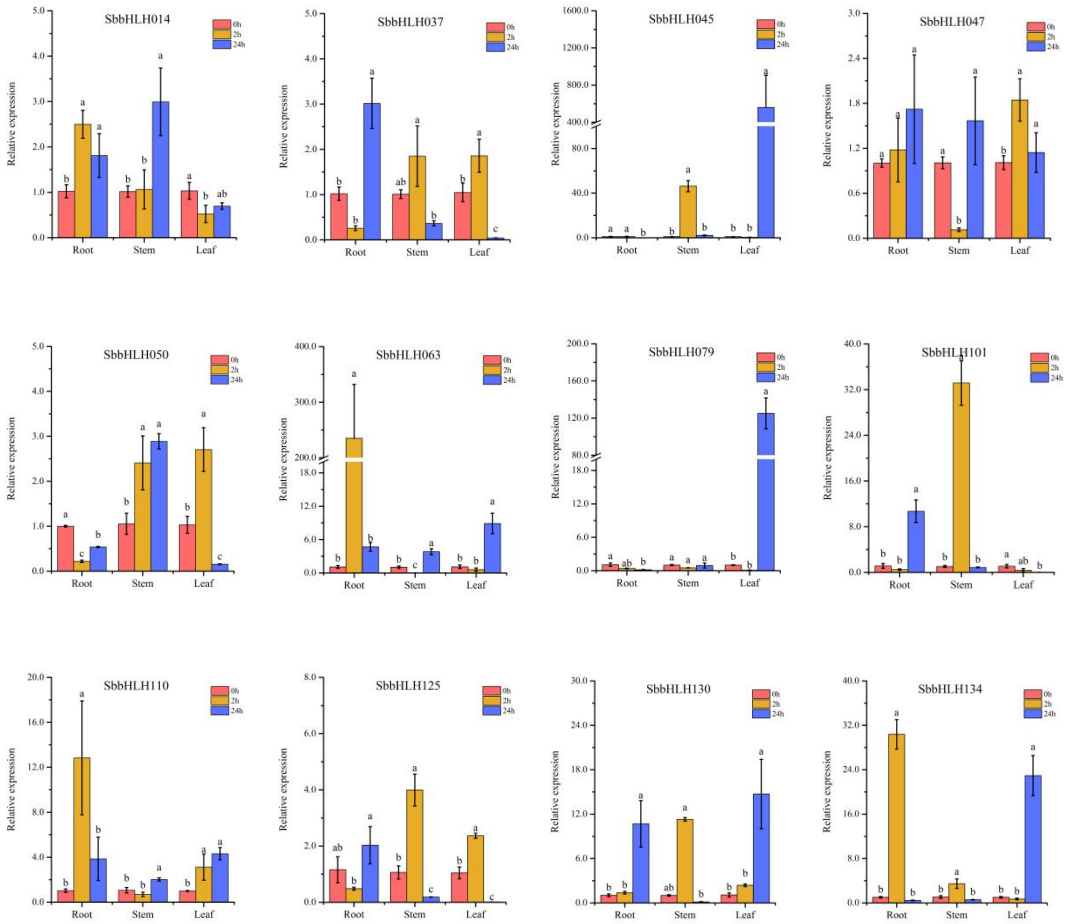
Flooding


SbbHLH014
SbbHLH037
SbbHLH045
SbbHLH047
SbbHLH050
SbbHLH063
SbbHLH079
SbbHLH101
SbbHLH110
SbbHLH125
SbbHLH130
SbbHLH134
Relative expression
Root
Stem
Leaf
0h
2h
24h

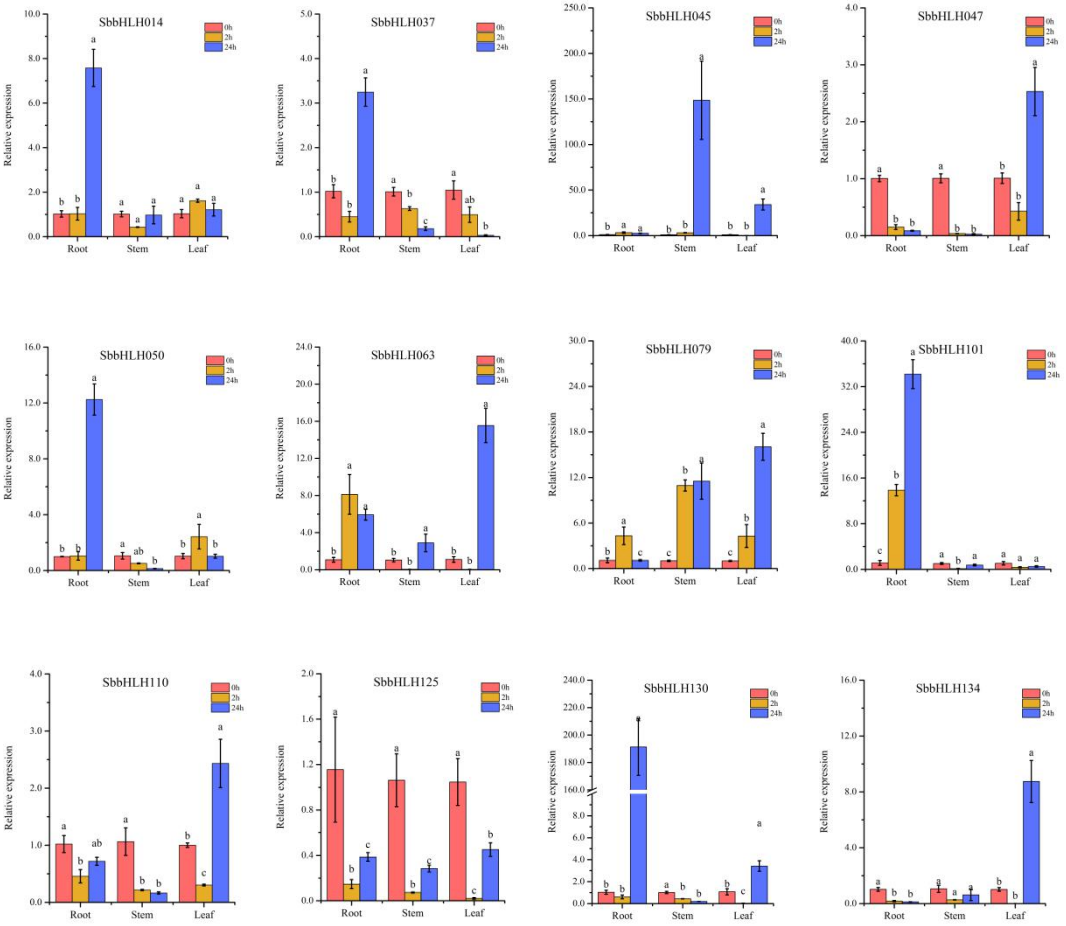
PEG


SbbHLH014
SbbHLH037
SbbHLH045
SbbHLH047
SbbHLH050
SbbHLH063
SbbHLH079
SbbHLH101
SbbHLH110
SbbHLH125
SbbHLH130
SbbHLH134
Relative expression
Root
Stem
Leaf
0h
2h
24h

NaCl


SbbHLH014
SbbHLH037
SbbHLH045
SbbHLH047
SbbHLH050
SbbHLH063
SbbHLH079
SbbHLH101
SbbHLH110
SbbHLH125
SbbHLH130
SbbHLH134
Relative expression
0h
2h
24h
Root
Stem
Leaf

Heat


SbbHLH014
SbbHLH037
SbbHLH045
SbbHLH047
SbbHLH050
SbbHLH063
SbbHLH079
SbbHLH101
SbbHLH110
SbbHLH125
SbbHLH130
SbbHLH134
Relative expression
Root
Stem
Leaf
0h
2h
24h

Cold

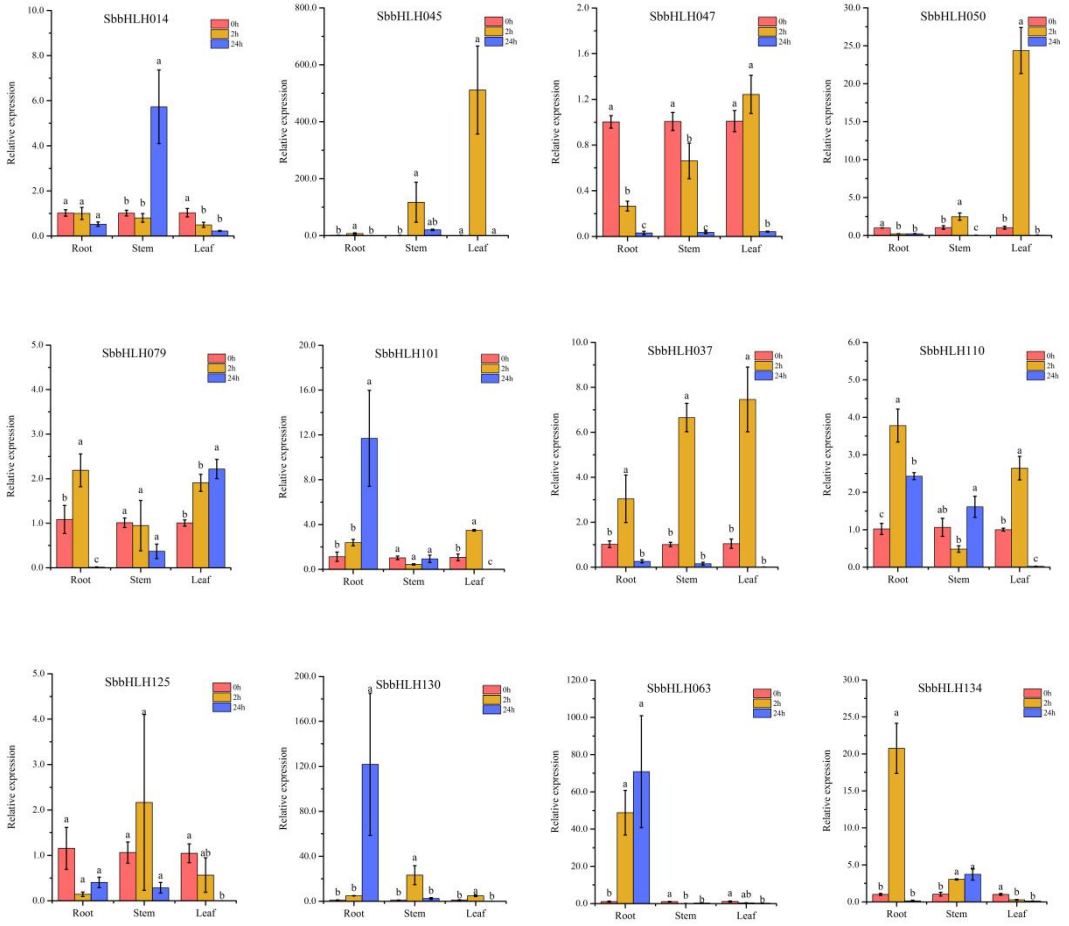
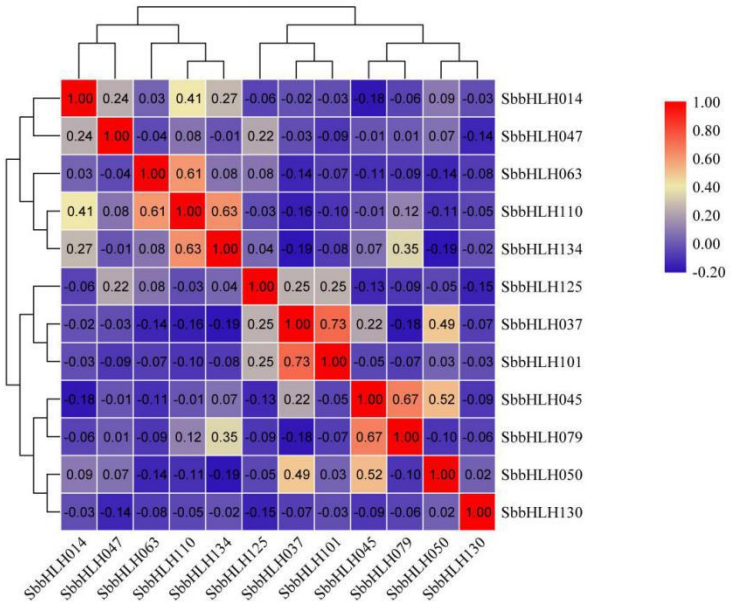

SbbHLH014
SbbHLH045
SbbHLH047
SbbHLH050
SbbHLH079
SbbHLH101
SbbHLH037
SbbHLH110
SbbHLH125
SbbHLH130
SbbHLH063
SbbHLH134
Relative expression
0h
2h
24h
Root
Stem
Leaf

Figure 9B


1.00 0.24 0.03 0.41 0.27 -0.06 -0.02 -0.03 -0.18 -0.06 0.09 -0.03 SbbHLH014
0.24 1.00 -0.04 0.08 -0.01 0.22 -0.03 -0.09 -0.01 0.01 0.07 -0.14 SbbHLH047
0.03 -0.04 1.00 0.61 0.08 0.08 -0.14 -0.07 -0.11 -0.09 -0.14 -0.08 SbbHLH063
0.41 0.08 0.61 1.00 0.63 -0.03 -0.16 -0.10 -0.01 0.12 -0.11 -0.05 SbbHLH110
0.27 -0.01 0.08 0.63 1.00 0.04 -0.19 -0.08 0.07 0.35 -0.19 -0.02 SbbHLH134
-0.06 0.22 0.08 -0.03 0.04 1.00 0.25 0.25 -0.13 -0.09 -0.05 -0.15 SbbHLH125
-0.02 -0.03 -0.14 -0.16 -0.19 0.25 1.00 0.73 0.22 -0.18 0.49 -0.07 SbbHLH037
-0.03 -0.09 -0.07 -0.10 -0.08 0.25 0.73 1.00 -0.05 -0.07 0.03 -0.03 SbbHLH101
-0.18 -0.01 -0.11 -0.01 0.07 -0.13 0.22 -0.05 1.00 0.67 0.52 -0.09 SbbHLH045
-0.06 0.01 -0.09 0.12 0.35 -0.09 -0.18 -0.07 0.67 1.00 -0.10 -0.06 SbbHLH079
0.09 0.07 -0.14 -0.11 -0.19 -0.05 0.49 0.03 0.52 -0.10 1.00 0.02 SbbHLH050
-0.03 -0.14 -0.08 -0.05 -0.02 -0.15 -0.07 -0.03 -0.09 -0.06 0.02 1.00 SbbHLH130
SbbHLH014
SbbHLH047
SbbHLH063
SbbHLH110
SbbHLH134
SbbHLH125
SbbHLH037
SbbHLH101
SbbHLH045
SbbHLH079
SbbHLH050
SbbHLH130
1.00
0.80
0.60
0.40
0.20
0.00
-0.20

Figures

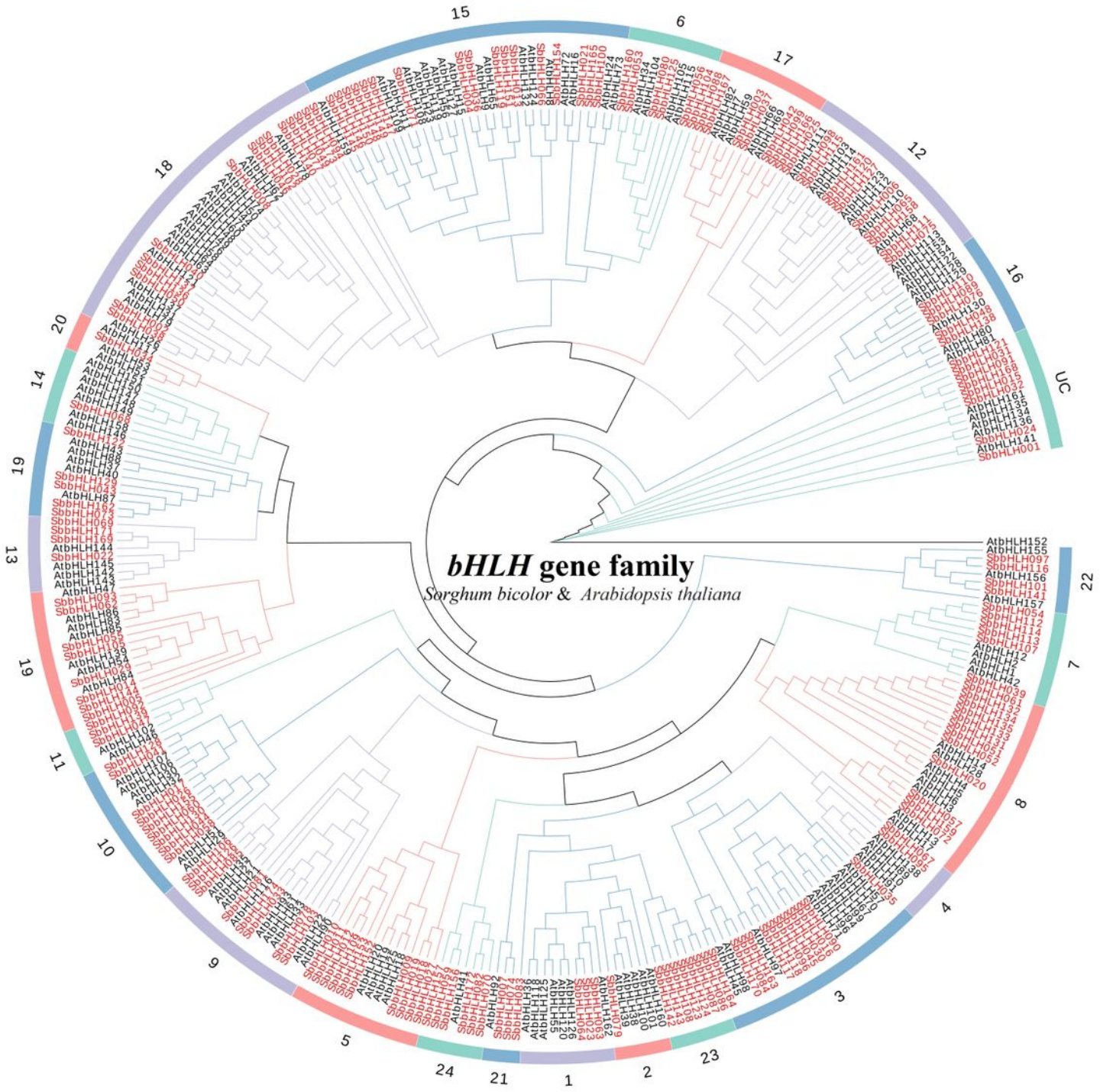


Figure 1

Unrooted phylogenetic tree showing relationships among bHLH domains of S. bicolor and Arabidopsis. The phylogenetic tree was derived using the NJ method in MEGA7.0. The tree shows the 24 phylogenetic

subfamilies and 1 unclassified group (UC) marked with red font on a white background. bHLH proteins from Arabidopsis are marked with the prefix ‘At’.

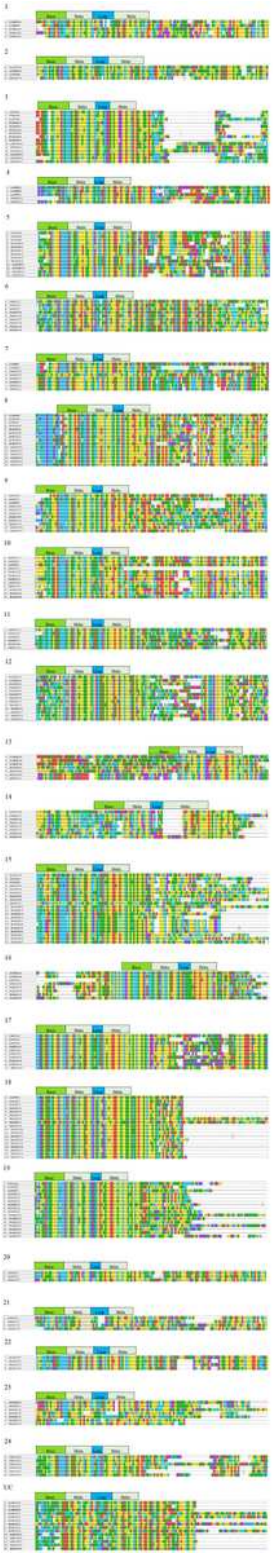


Figure 2

Multiple sequence alignment of the bHLH domains of the members of 24 phylogenetic subfamilies and 1 unclassified group (UC) of the SbbHLH protein family. The scheme at the top depicts the locations and boundaries of the basic, helix, and loop regions in the bHLH domain.

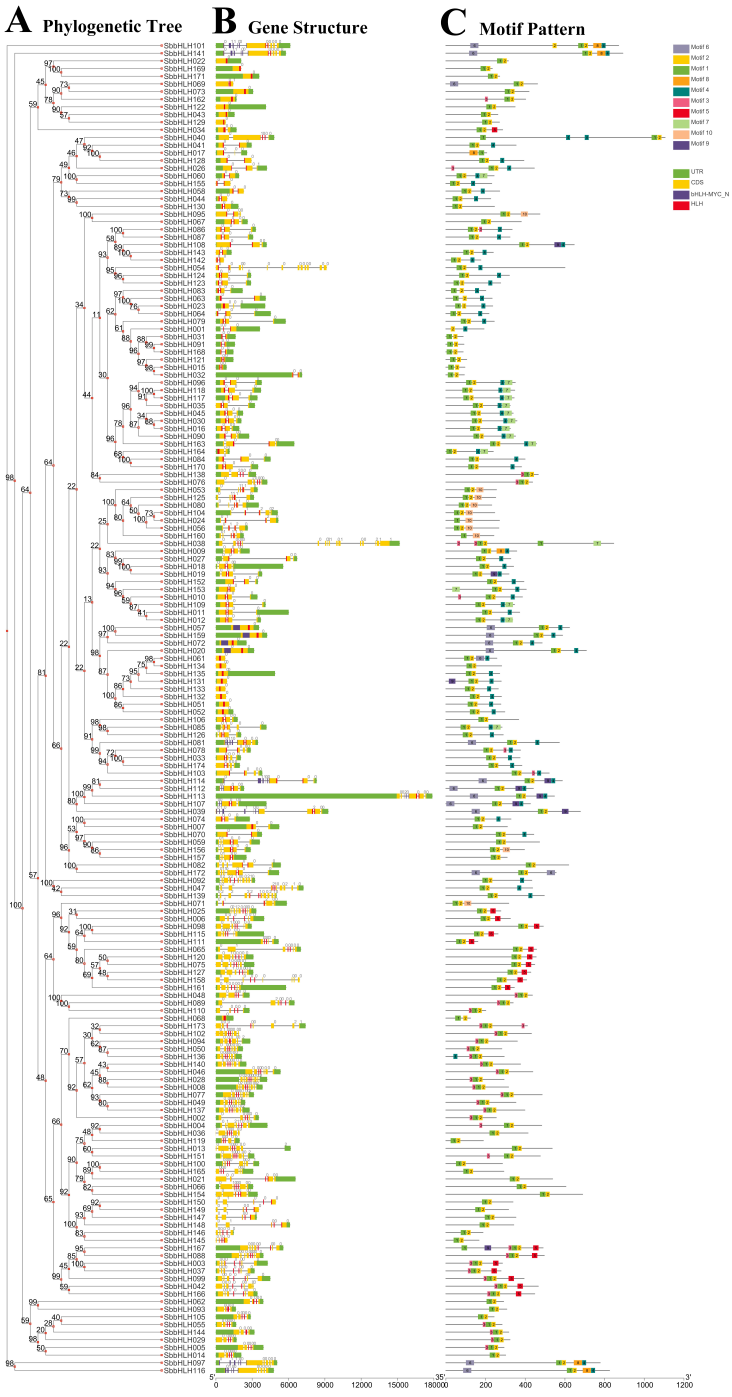


Figure 3

Phylogenetic relationships, gene-structure analysis, and motif distributions of S. bicolor bHLH genes. A Phylogenetic tree was constructed by the NJ method with 1000 replicates on each node. B Exons and introns are indicated by yellow rectangles and gray lines, respectively. C Amino acid motifs in the SbbHLH proteins (1–10) are represented by colored boxes. The black lines indicate relative protein lengths.

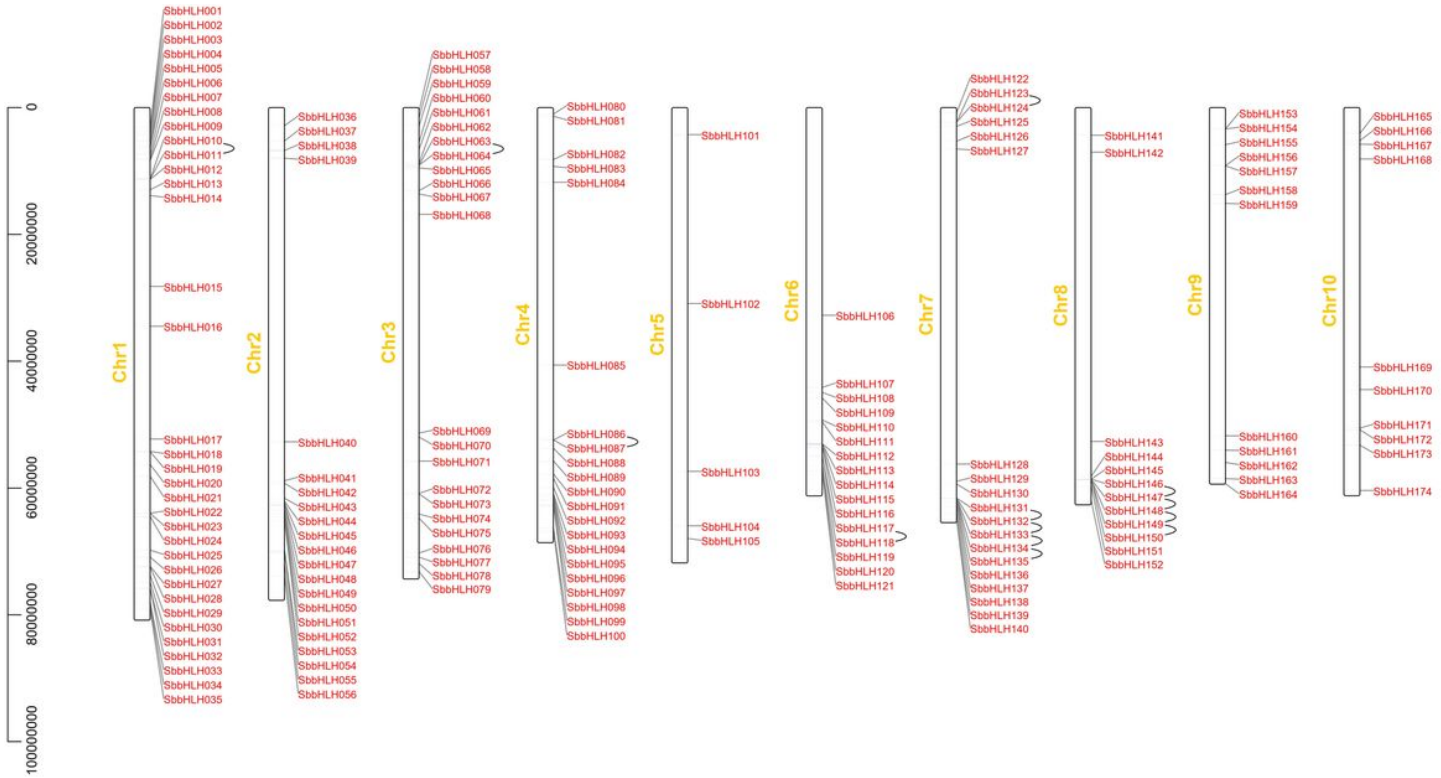


Figure 4

Schematic representation of the chromosomal distribution of the S. bicolor bHLH genes. Vertical bars represent the chromosomes of S. bicolor. The chromosome number is indicated to the left of each chromosome. The scale on the left represents chromosome length.

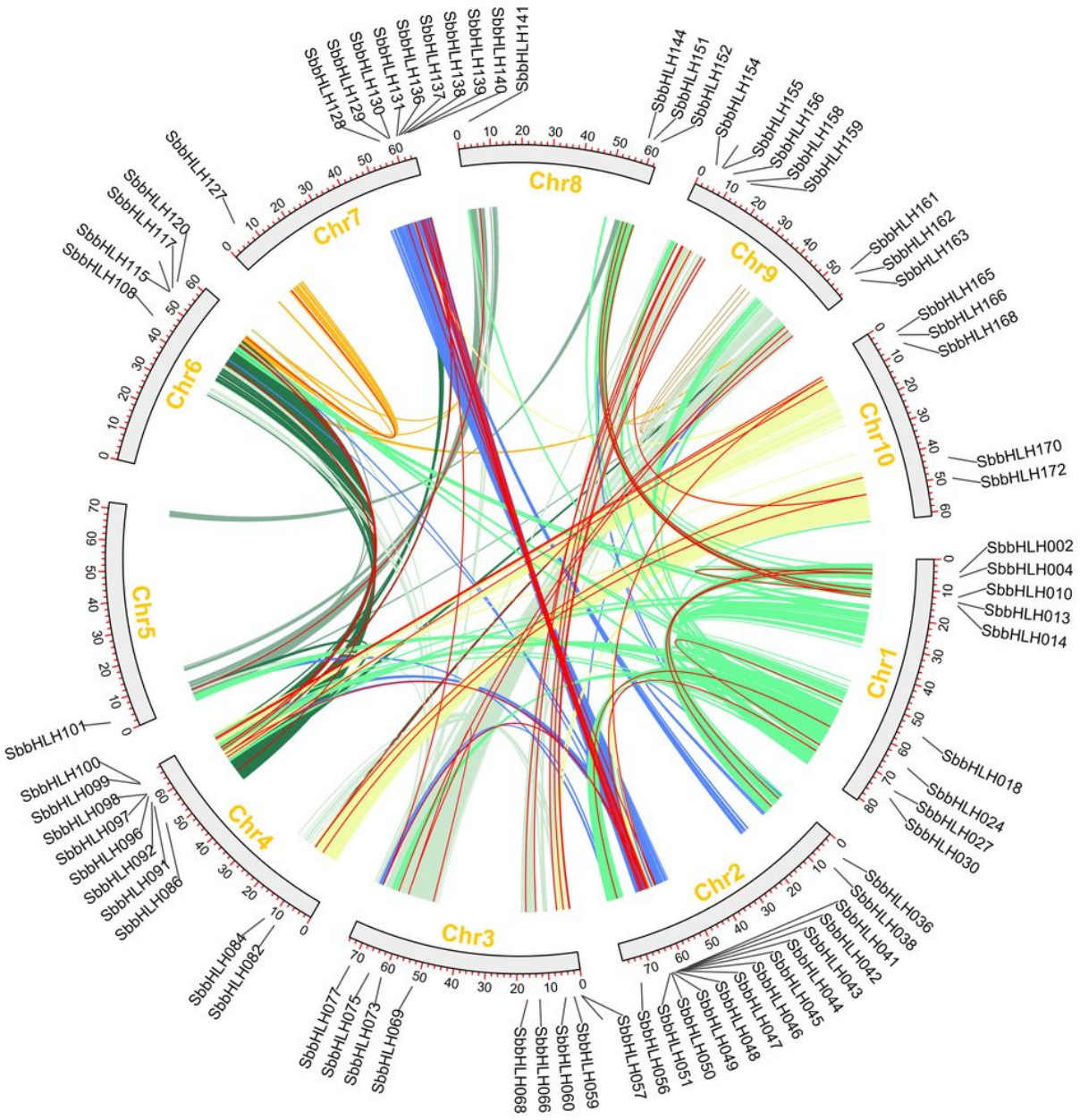


Figure 5

Schematic representation of the chromosomal distribution and interchromosomal relationships of S. bicolor bHLH genes. Colored lines indicate all synteny blocks in the S. bicolor genome and the red lines indicate duplicated bHLH gene pairs. Chromosome number is indicated at the bottom of each chromosome.

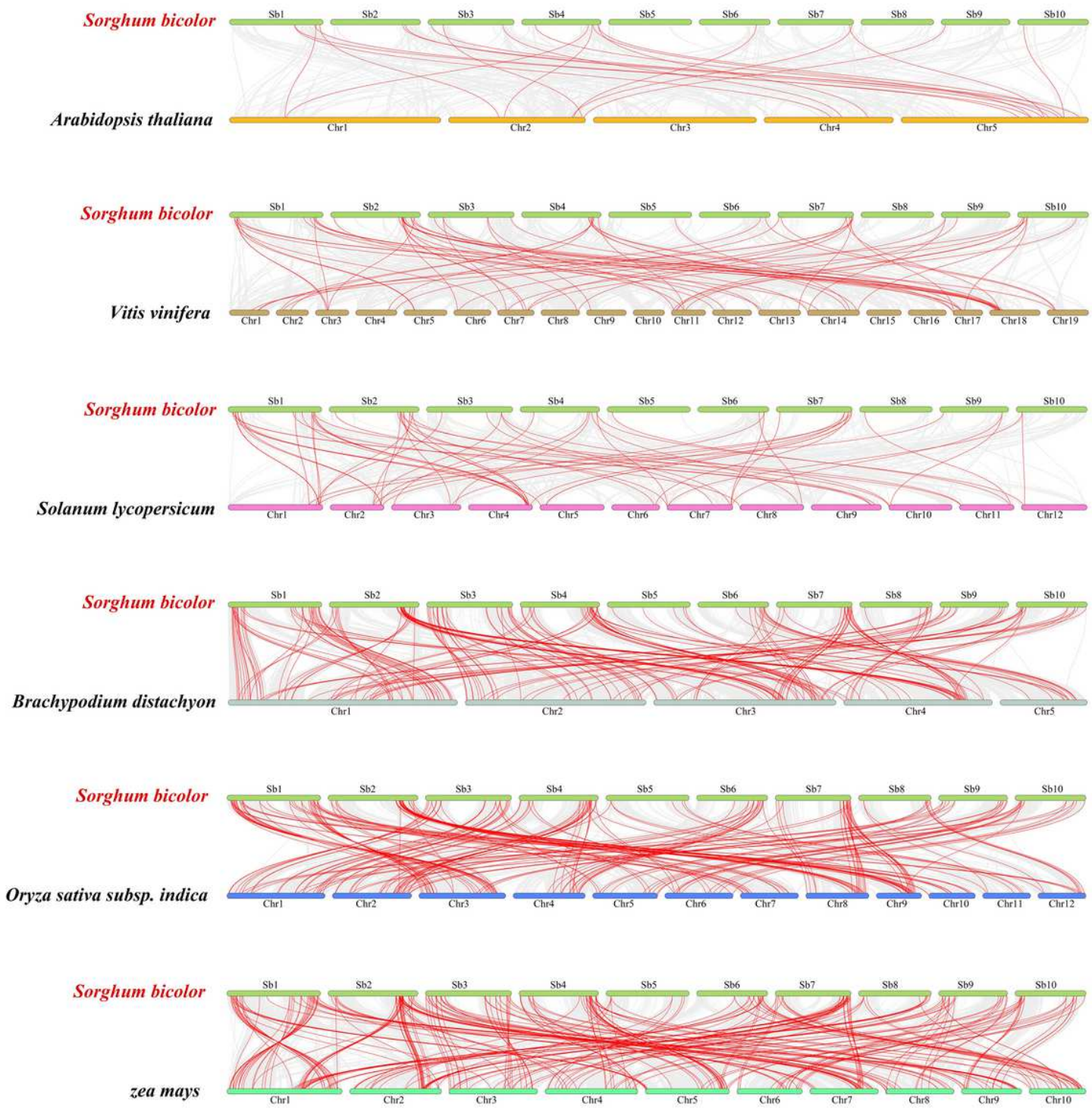


Figure 6

Synteny analyses of the bHLH genes between S. bicolor and six representative plant species (Arabidopsis thaliana, Vitis vinifera, Solanum lycopersicum, Brachypodium distachyon, Oryza sativa subsp. indica, Zea mays). Gray lines on the background indicate the collinear blocks in S. bicolor and other plant genomes; red lines highlight the syntenic S. bicolor bHLH gene pairs.

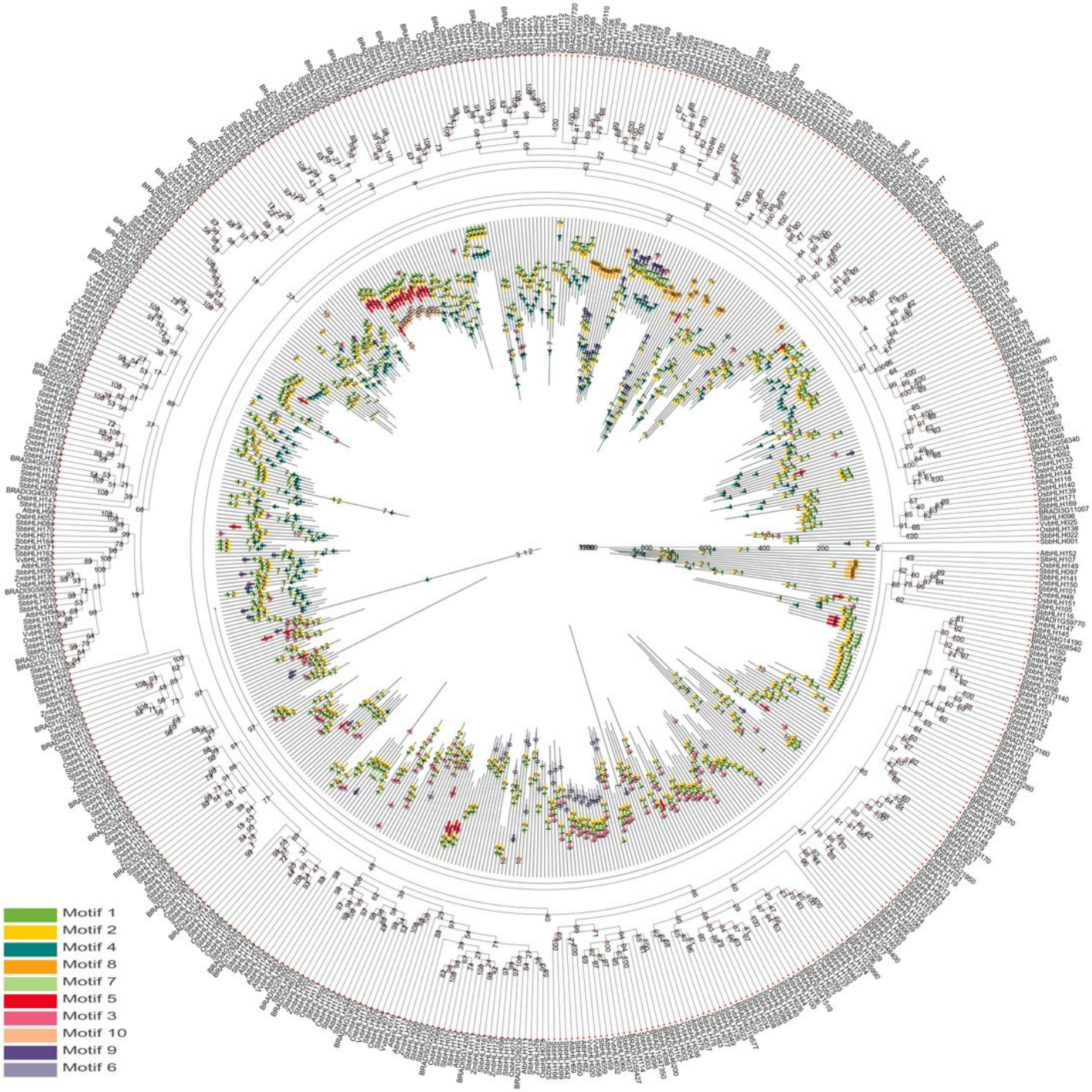


Figure 7

Phylogenetic relationship and motif composition of the bHLH proteins from S. bicolor with six different plant species (Arabidopsis thaliana, Vitis vinifera, Solanum lycopersicum, Brachypodium distachyon, Oryza sativa subsp. indica, Zea mays). Outer panel: An unrooted phylogenetic tree constructed using Geneious R11 with the NJ method. Inner panel: Distribution of the conserved motifs in bHLH proteins.

The differently colored boxes represent different motifs and their positions in each bHLH protein sequence. The sequence information for each motif is provided in Additional File 2: Table S2.

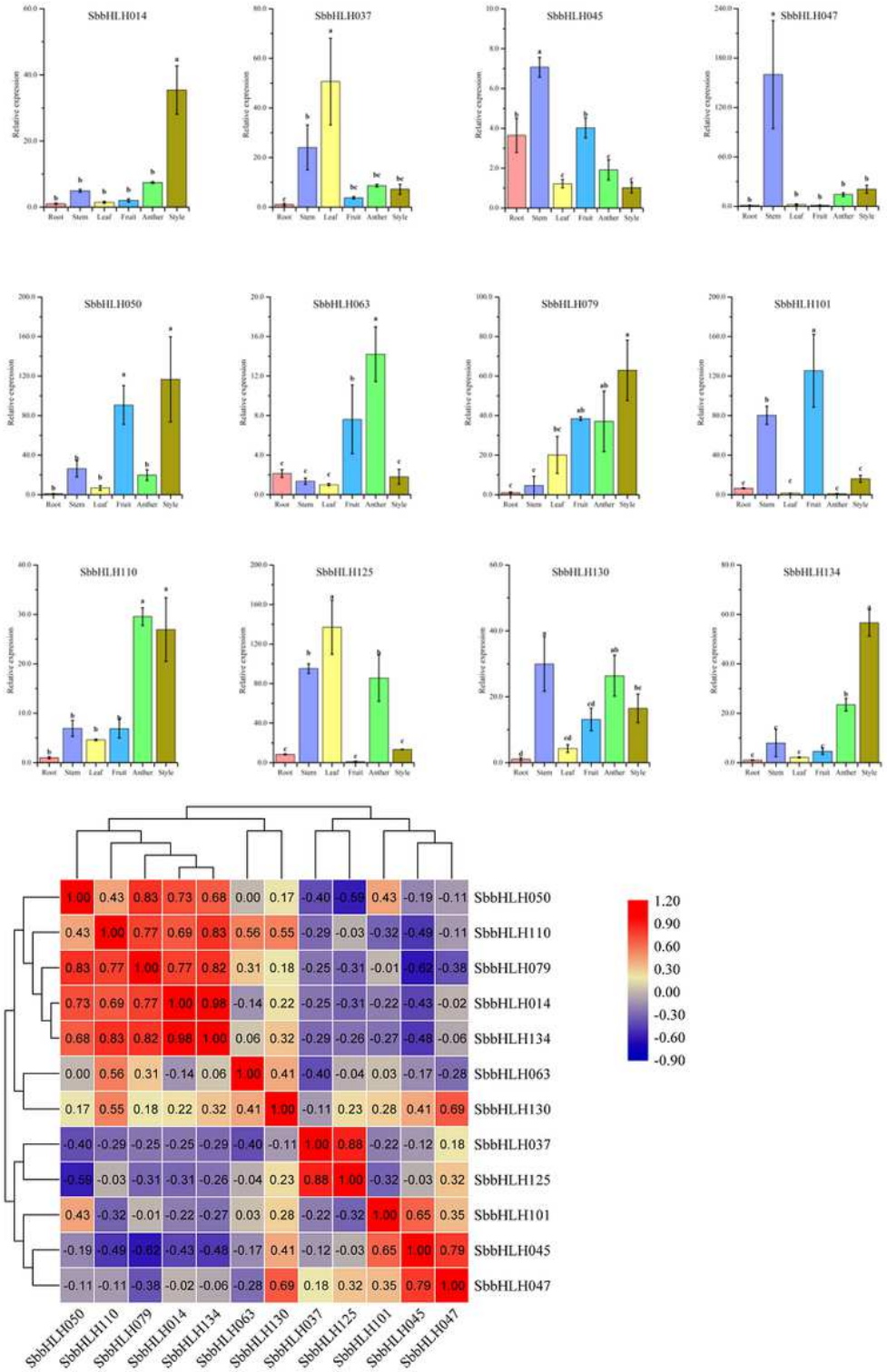


Figure 8

Tissue-specific gene expression of 12 S. bicolor bHLH genes and the correlation between their expression patterns. A Expression patterns of 12 S. bicolor bHLH genes in the anther, style, leaf, root, stem and fruit organs were examined by qRT-PCR. Error bars were obtained from three measurements. Lowercase letter

above the bar indicates significant difference (α = 0.05, LSD) among the treatments. B Positive number: positively correlated; negative number: negatively correlated. Red numbers indicate a significant correlation at the 0.05 level.

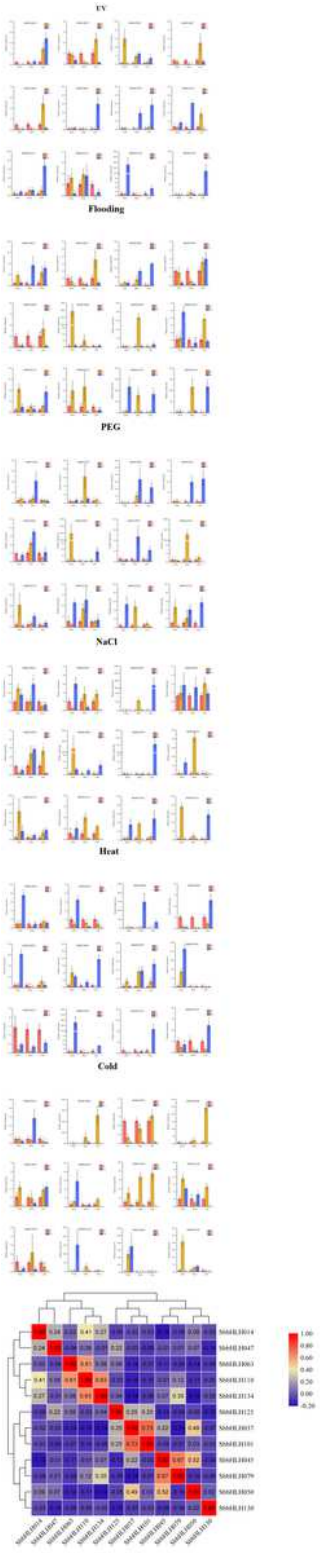


Figure 9

Gene expression of 12 S. bicolor bHLH genes in plants subjected to abiotic stresses (strong UV radiation, flooding, PEG, NaCl, heat and cold treatments) at the seedling stage. A Expression patterns of 12 S.

bicolor bHLH genes in leaf, root and stem organs were examined by qRT-PCR. Error bars were obtained from three measurements. Lowercase letter above the bar indicates significant difference (α = 0.05, LSD) among the treatments. B Positive number: positively correlated; negative number: negatively correlated. Red numbers indicate a significant correlation at the 0.05 level.

Supplementary Files

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

- TableS1.Listofthe174S.bicolorbHLHgenesidentifiedinthisstudy..xlsx
- TableS2.AnalysisanddistributionoftheconservedmotifsinS.bicolorbHLHproteins..xls
- TableS3.TandemduplicationeventsofS.bicolorbHLHgenes..xlsx
- TableS4.The42pairsofsegmentalduplicationsinS.bicolorbHLHgenes..xlsx
- TableS5.OnetooneorthologousgenerelationshipsbetweenS.bicolorandotherplants..xls
- TableS6.PrimersequencesforqRTPCR.xlsx