Genome wide identification and characterization of cucumber *bHLH* family genes and the functional characterization of *CsbHLH041* in NaCl and ABA tolerance in *Arabidopsis* and cucumber

Jialin Li
  Shandong Agricultural University

Ting Wang
  Shandong Agricultural University

Jing Han
  Shandong Agricultural University

Zhonghai Ren
  Shandong Agricultural University

Research article

Keywords: Abiotic stresses, bHLH family, Cucumber, CsbHLH041, Expression patterns, Regulatory networks

Abstract

Background: The basic/helix-loop-helix (bHLH) transcription factor family exists in all three eukaryotic kingdoms as important regulatory components in biological growth and development. To date, there has been relatively little effort to systematically carry out comprehensive genomic and functional analyses of bHLH genes in cucumber (Cucumis sativus L.).

Results: Here, a total of 142 bHLH genes were identified in the recently released cucumber genome and further classified into 32 subgroups based on the phylogenetic analysis, conserved motifs and gene structures. Multiple sequence alignment analyses showed that the sequences of CsbHLH proteins were highly conserved. The chromosomal distribution, synteny analysis, and gene duplications of these 142 CsbHLHs were further analysed. A cis-element analysis revealed many elements related to stress responsiveness and plant hormones in the promoter regions of CsbHLH genes. Phylogenetic comparison of the bHLH members between cucumber and Arabidopsis , revealed that cucumber bHLH proteins were clustered into the different functional clades of Arabidopsis bHLH members. The transcript abundance analysis of selected CsbHLHs under abiotic stresses (NaCl, ABA and low-temperature treatments) identified five CsbHLH genes that could simultaneously respond to the three abiotic stresses. Tissue-specific expression profiles of these five genes were also analysed. In addition, 35S : CsbHLH041 enhanced the tolerance to salt and ABA in transgenic Arabidopsis and in cucumber seedlings, suggesting CsbHLH041 is an important regulator in response to abiotic stresses. Finally, the functional interoperability network among the CsbHLH proteins was analysed.

Conclusion: This study provided a good foundation for further research into the functions and regulatory mechanisms of CsbHLH proteins and identified candidate genes for stress resistance in cucumber.

Background

Basic helix-loop-helix (bHLH) proteins exist widely in all three eukaryotic kingdoms and form one of the largest families of TFs [1, 2]. The bHLH TFs are named for their own structural characteristics [3], which are mostly composed of conserved 60 amino acid residues. According to the different functions, they can be divided into two parts: the basic region and the HLH region [4]. The basic region is distributed at the N-terminus of the bHLH conserved domain and contains approximately 15 to 20 residues, which are involved in DNA binding [5, 6]. The HLH domain is distributed at the C-terminus of the gene sequence, which is composed of two amphipathic α-helices mainly consisting of hydrophobic residues linked by a loop region of variable sequence and length. The HLH domain is an essential structure for the formation of homologous or heterologous dimers in bHLH transcription factors [6, 7].

According to the evolutionary origin, sequence similarity, DNA binding patterns, and functional types, in animals, bHLH transcription factors are mainly divided into six categories, A-F, containing 45 subgroups [8, 9]. In plants, the bHLH gene family has been divided into 15−26 groups [10], and even up to 32 groups when atypical bHLH proteins are included [2]. In Arabidopsis, 167 bHLH proteins are divided into 21

subfamilies [2,11]; the 165 bHLH family members in rice are classified into 22 subfamilies [12]; and the 159 bHLH proteins are divided into 21 subfamilies in tomato [13]. Currently, increasing numbers of bHLH proteins have been found in plants, and their functional research is gradually increasing.

In plants, the *bHLH* gene family plays important roles in plant growth and development, metabolic regulation, and response to environmental changes. The first member of the *bHLH* family discovered was the maize *R* gene, which was shown to play a key role in anthocyanin synthesis [14]. Subsequently, an increasing number of bHLHs have been shown to be involved in a wider range of physiological pathways. For example, Phytochrome Interacting Factors (PIFs) have been reported to respond to light signals [15]; overexpression of *PRE1* activates gibberellin-dependent responses in *Arabidopsis thaliana* [16]; *AtGL3*, *AtEGL3* and *AtTT8* have been demonstrated to be involved in anthocyanin and PA biosynthesis [17, 18]; while *AtGL3*, *AtEGL3* and *AtMYC1* also regulate trichome formation and root hair patterning [19]. In addition, some bHLH TFs have also been suggested to respond to diverse abiotic stresses and improve plant stress tolerance, including tolerance to cold, salt and drought. In wheat, *bHLH39* increases the expression levels of stress-response genes under salt stress, thus improving the salt tolerance of overexpressing *bHLH39* wheat plants [20]. The bHLH TFs often function by forming homodimers or heterodimers with other proteins. For example, MYC3 and MYC4 transcription factors all can interact with multiple JAZ proteins (such as JAZ1, JAZ4, and JAZ9) to jointly regulate the JA signalling pathway [21]. The MYB-bHLH-WD40 complexes are involved in different processes, such as the biosynthesis of anthocyanins and PAs, leaf trichome formation and root hair patterning [22]. In summary, bHLH in plants can form homologous or heterologous complexes with bHLH proteins or other proteins to extend their biological functions.

Cucumber (*Cucumis sativus* L.) is an economically important crop cultivated worldwide [23]. Although many important functions of the *AtbHLH* family in *Arabidopsis thaliana* have been widely studied [2], genome-wide information about *CsbHLH* family members has not been reported previously. In this study, we identified and characterized 142 *bHLH* family genes in cucumber. They were distributed over seven chromosomes and could be classified into 32 subgroups based on phylogenetic analysis. We also investigated their gene structures, conserved motifs, synteny analysis, gene duplications and cis-elements in promoters. In addition, the expression levels of some *CsbHLH* genes were measured by qRT-PCR to study their responses to low temperature (4°C), salt (NaCl) and ABA stress, for which all tested genes were stress-responsive. The protein interaction network among the CsbHLH proteins was predicted, which could help to understand the possible functional mechanism of CsbHLH proteins. Furthermore, overexpression of *CsbHLH041* showed increased salt resistance and ABA resistance compared with controls in cucumber and *Arabidopsis*. We hope that this work will provide a useful resource for subsequent research on the functions and regulatory mechanisms of potentially important CsbHLH proteins that are crucial in modulating abiotic stress responses in cucumber.

Results

Identification and analysis of cucumber *bHLH* genes

To identify *bHLH* family genes in the cucumber genome, we performed the BlastP programme to search against the cucumber genome database by using 166 *Arabidopsis* bHLH proteins [2, 10] and the consensus protein sequences of the bHLH domain, with Hidden Markov Model (HMM) profile (PF00010) as queries. We obtained 164 putative members of the *CsbHLH* family. To confirm the reliability of the *bHLH* genes in the cucumber genome, we used Pfam (http://pfam.janelia.org/) and SMART (http://smart.embl-heidelberg.de/) [24] to search for the presence of the bHLH domain in the amino acid sequences of all 164 proteins. Only 142 proteins had the corresponding conserved bHLH domain, which were named *CsbHLH1* to *CsbHLH142* based on their phylogenies and sequence similarity corresponding to individual AtbHLH proteins. Finally, the specific information for the 142 typical bHLH genes, including the gene ID, amino acids length, gene length and chromosomal locations, were listed in Table 1. The lengths of the CsbHLH protein sequences ranged from 84 residues (*CsaV3_1G005290*) to 960 residues (*CsaV3_1G043790*), and the isoelectric points (pI) ranged from 4.57 (*CsaV3_2G030090*) to 11.79 (*CsaV3_6G028530*).

Phylogenetic analysis, conserved motif and gene structure analysis of *CsbHLH* gene family

To confirm the structural characteristics of CsbHLHs, multiple sequence alignment (MSA) analysis was performed with the 142 CsbHLH proteins. As shown in Fig. 1, one basic region, one loop region and two helix regions were detected in all 142 CsbHLH proteins. Additionally, conserved amino acids in bHLH domains, with a sequence identity greater than 50%, were showed in purple or light blue colour (Fig. 1a). Sequence logos were produced using the 142 CsbHLH homologous domain amino acid sequences (Fig. 1b). The CsbHLH proteins in cucumber contained 17 conserved amino acids in the bHLH domain, which were present in the *bHLH* gene family of *Arabidopsis* and Moso bamboo [2, 25]. As shown in Fig. 1b, we could clearly observe that key amino acid residues Arg-10, Arg-11, Leu-21 and Leu-53 were highly conserved (92%, 87%, 96%, and 90%, respectively) among the 142 CsbHLH proteins. Subsequently, a phylogenetic tree was constructed based on the alignment of the corresponding 142 bHLH complete protein sequences, and they were subdivided into 32 subgroups, designated C1–C32, according to the clades with at least 50% bootstrap support (Fig. 2a).

To support the phylogenetic analysis, gene structure analysis of *bHLH* family members was performed. The results revealed that *CsbHLHs* in the same subgroups presented similar numbers of exons and introns, and without considering the size of the introns, the *CsbHLH* genes within the same subgroups had similar intron-exon gene structures (Fig. 2c).

To further investigate the specific motifs of CsbHLH proteins within the same subgroup, 10 distinct conserved motifs were identified by the MEME motif search tool (Fig. 2b). All predicted motifs were identified only once in each CsbHLH protein sequences. One hundred forty-two CsbHLH proteins contained different numbers of conserved motifs, ranging from two to seven, and motifs 1 and 2, representing one part of the bHLH domain, were present in almost all CsbHLH proteins (Fig. 2b). Moreover, a similar motif composition existed in CsbHLH proteins of the same subgroup. For example,

members of subgroup 23 all contained motifs 1, 2, 4 and 6, while motif 5 was identified in most CsbHLH proteins. Interestingly, some of the specific motifs were absent in certain subgroups. For example, motif 4 was absent in all members of the 1, 2 and 3 subgroups (Fig. 2b).

In general, the results of conserved motif and gene structure analyses further confirmed the results of the phylogenetic analysis, illustrating that proteins within the same subgroup may have similar functions.

Synteny analysis of *bHLH* genes in cucumber, *Arabidopsis* and tomato

To determine the genomic distribution and duplication of *CsbHLH* genes, first, the DNA sequence of each gene was used to search the cucumber genome database using BLASTN. All 142 *CsbHLH* genes (except *CsaV3_UNG229040*) could be mapped onto chromosomes 1-7 (Fig. 3a; Table 1; Fig. S2). According to the description reported by [26] to determine the duplication of *CsbHLH* genes, the syntenic regions were analysed by MCscanX software. As shown in Table S1, a total of 1468 tandem duplication gene pairs and 231 segmental duplication blocks were identified in the cucumber genome, respectively. Five tandem duplication gene pairs were obtained in *CsbHLH* family (*CsbHLH019* and *CsbHLH020*; *CsbHLH019* and *CsbHLH025*; *CsbHLH120* and *CsbHLH125*; *CsbHLH125* and *CsbHLH126*; *CsbHLH038* and *CsbHLH101*) (Fig. 3a; Table S1). In addition, segmental duplication events with seven *bHLH* gene pairs (*CsbHLH112* and *CsbHLH127*; *CsbHLH040* and *CsbHLH037*; *CsbHLH054* and *CsbHLH085*; *CsbHLH060* and *CsbHLH074*; *CsbHLH001* and *CsbHLH135*; *CsbHLH141* and *CsbHLH046*; *CsbHLH050* and *CsbHLH044*) were also identified in cucumber (Fig. 3a; Table S1).

To further illuminate the phylogenetic mechanisms of cucumber *bHLH* family, we constructed the comparative syntenic maps of cucumber associated with *Arabidopsis* and tomato, respectively (Fig. 3b). Many *CsbHLH* genes showed a syntenic relationship with those in tomato and *Arabidopsis*, respectively. Interestingly, some *CsbHLH* genes were found to be associated with at least two syntenic gene pairs between cucumber and tomato, such as *CsbHLH024*, *CsbHLH040* and *CsbHLH054*, among others. Similarly, some *CsbHLH* genes also corresponded to two syntenic gene pairs between cucumber and *Arabidopsis*, for instance, *CsbHLH020* and *CsbHLH049*, indicating that these genes may have played an important role in the *bHLH* gene family during evolution. In addition, some collinear pairs (with *CsbHLH132*, *CsbHLH135* and *CsbHLH136*, among others) were identified between cucumber and both *Arabidopsis* and tomato (Fig. 3b; Table S2), indicating that these orthologous pairs might have already existed before the ancestral divergence. We also found some *CsbHLH* genes were not associated with syntenic gene pairs in *Arabidopsis* or tomato, indicating that they might have been peculiar to cucumber during the course of evolution.

Cis-element analysis in the promoter regions of *CsbHLH* genes in cucumber

Previous studies have shown that many *bHLH* genes play important roles in response to various abiotic stresses [27]. To analyse the putative functions of the *CsbHLH* genes in response to abiotic stress, the 2-kb promoter regions of the *CsbHLH* genes were isolated for the analysis of the potential *cis*-elements using the PlantCARE software (Table S3), in which most *CsbHLH* genes particularly presented elements related to stress responsiveness (anaerobic induction, low temperature and drought inducibility) and plant hormones (abscisic acid, auxin, gibberellic acid, methyl jasmonate and salicylic acid). Moreover, the promoter regions of some *CsbHLH* genes contained an MYB binding site involved in flavonoid biosynthetic gene regulation, which might be involved in the synthesis of flavonoid in cucumber (Fig. S3; Table S3). In addition, the *CsbHLH* gene promoter regions also contained many light response elements (such as the G-Box and Box-4). In general, the *cis*-regulatory elements present within *CsbHLH* promoters could be divided into three major categories. The first category consisted of the plant light-responsive elements (such as the G-box, AE-box and I-box), the second included plant growth- and development-responsive elements (such as O2-site and ARE), and the third included elements that respond to diverse stresses (such as TC-rich, ABRE, MBS and LTR), which were widely distributed throughout the *CsbHLH* gene family (Table S3).

To further analyse whether there is co-expression of *CsbHLH* genes with the same *cis*-elements, we constructed a co-expression network of *CsbHLH* genes, based on the available RNA-seq data of 10 cucumber tissues regarding correlations between cucumber *bHLH* genes [26]. The co-expression network containing 23 *CsbHLH* genes (nodes) and 191 correlations (edges) showed that each of the *CsbHLH* genes had multiple co-expression genes with same *cis*-elements (Fig. S4; Table S3). The result indicated the co-expression of genes may be related to the same *cis*-elements in their promoter regions.

Function prediction of *CsbHLHs* based on phylogenetic analyses

In *Arabidopsis*, the functions of many bHLH proteins have been characterized and verified [28, 29]. However, little is known about the biological functions of CsbHLHs in cucumber. In this study, we used phylogenetic analyses based on the proteins of 142 CsbHLHs and 166 AtbHLHs to identify putative orthologous and paralogous bHLH proteins in cucumber and *Arabidopsis*, which could preliminarily predict the functions of CsbHLH proteins [2, 10] (Fig. 4).

In all, 23 subfamilies were clustered, and the functions of CsbHLHs were predicted based on their homologs with verified functions in the same cluster (Table S4). As shown in Table S4, most of the members of subfamilies 1, 2, 4, 10, 13, 14 and 18 responded to diverse abiotic and biotic stresses [30, 31], including cold [32], salt [33], and drought [34]. Some proteins in subfamilies 4 and 10 may be related to iron regulation, modulating the homeostasis of iron content [35]. Members of subfamilies 19 and 23 were predicted to regulate flower development [36], while the members of subfamilies 3, 8, 9, 16 and 21 may be involved in the development of various plant organs [37, 38, 39]. There were PIFs in subfamily 17, which are related to light signal transduction and protect the normal growth and development of plants [15]. The

members of subfamily 5 regulate the flavonoid biosynthesis and cell differentiation of root epidermis [22]. The detailed possible functions of CsbHLHs are listed in Table S4.

In general, although the evolutionary relationships could not be clearly deciphered for the functions of all genes, the analysis was meaningful and necessary.

Expression analysis of *CsbHLH* genes under different stress conditions and in different tissues

To identify which *CsbHLH* genes are most important in the response to abiotic stresses, we carefully screened 21, 20 and 25 bHLH genes based on the cis-acting elements containing low temperature, defense and stress responsive and abscisic acid (ABA) elements in the promoters of bHLH genes, respectively, and detected their transcriptional changes with treatments of low temperature (4°C), salt (NaCl) and ABA, respectively. As expected, all the *CsbHLH* genes screened responded to stress treatments under the respective stress conditions (Fig. 5). For example, the expression levels of the 20 *CsbHLHs* were all positive in response to salt stress, and most were upregulated after one hour of treatment and achieved maximum values after 3 h of treatment, decreasing thereafter. The expression levels of *CsbHLH033*, *CsbHLH041* and *CsbHLH082* reached a maximum after just 1 h of NaCl treatment, while the expression levels of *CsbHLH136* were highest after 12 h. The results showed that *CsbHLH041* was the most susceptible to salt stress (increased by approximately 37-fold) (Fig. 5a).

Under ABA treatment, the expression levels of most of the 25 *CsbHLH* genes were up-regulated at different time points. Among them, the expression level of *CsbHLH020*, *CsbHLH041* and *CsbHLH064* increased more than 10-fold compared with the untreated level (*CsbHLH020*: the highest nearly 61-fold; *CsbHLH041*: the highest nearly 55-fold; *CsbHLH064*: the highest nearly 19-fold). In contrast, the expression levels of four of the *CsbHLHs* genes were significantly down-regulated under ABA treatment (*CsbHLH011*, *CsbHLH033*, *CsbHLH034* and *CsbHLH077*), as could be seen in Fig. 5b.

The expression levels of 20 of the 21 *CsbHLH* were up-regulated at some time points after the 4°C treatment, while only *CsbHLH032* was decreased (Fig. 5c). We found that the 20 *CsbHLHs* genes were upregulated at one or two time points,but none of them were upregulated at each time point. As shown in Fig. 5, the *CsbHLH020*, *CsbHLH064*, *CsbHLH086*, *CsbHLH093* and *CsbHLH112* genes could simultaneously respond to the three abiotic stresses.

The expression patterns of genes under different conditions are often related to their functions. Therefore, qRT-PCR were used to detect the expression patterns for *CsbHLH020*, *CsbHLH064*, *CsbHLH086*, *CsbHLH093* and *CsbHLH112* abiotic stress-responsive *CsbHLH* genes in roots, stems, young leaves, male flowers, ovaries and tendrils. The expression patterns of the five *CsbHLH* genes showed different tissue specificities (Fig. 5d). For example, *CsbHLH093* and *CsbHLH112* showed high levels of transcript abundance in roots and ovaries, but low levels in male flowers and tendrils (Fig. 5d). In contrast, both *CsbHLH064* and *CsbHLH086* showed high expression levels in male flowers and tendrils. The expression

levels of *CsbHLH020* were higher in roots and young leaves than in other tissues (Fig. 5d). These results suggested that cucumber *bHLH* genes might be involved in many aspects of physiological and developmental processes.

CsbHLH041 enhanced tolerance to NaCl and ABA in transgenic *Arabidopsis* and cucumber

CsbHLH041 expression was significantly induced by salt and ABA in cucumber (Fig. 5a-b). Therefore, we used *Agrobacterium*-mediated transient transformation of cucumber cotyledons to clarify *CsbHLH041* tolerance to salt and ABA. After 0.5 h of 100 mM NaCl treatment, serious wilting occurred in the seedlings overexpressing *35S* empty vector compared with over-expression *CsbHLH041*, and the wilting difference was more obvious after 3 hour of NaCl treatment (Fig. 6a). After 12 hours, the survival rate of the transgenic seedlings (24%) was markedly higher than that of the *35S* empty vector seedlings (6%), showing that over-expression of *CsbHLH041* resulted in significant salt resistance (Fig. 6c). After 6 hours of ABA treatment, the transgenic seedlings were more vigorous than *35S* empty vector seedlings (Fig. 6b). With the extension of ABA treatment time, the *35S* cucumber seedlings showed visible symptoms of ABA-induced damage, such as drying, wilting, and even death, with survival of only 12%. While some *CsbHLH041* transgenic plants remained green with expanded cotyledons, and the survival rate was up to approximately 40% (Fig. 6b-c).

To clarify the possible factors underlying the enhanced salt and ABA resistance, we examined the enzymatic activities in the ROS clearance system under NaCl and ABA treatments, respectively. Without the NaCl or ABA treatment, the enzymatic activities of POD, SOD and CAT in *35S* and *35S:CsbHLH041* transgenic seedlings were no significant difference (Fig. 6d-f). Nevertheless, both NaCl treatment and ABA treatment were significantly activated more enzymatic scavenging activities in the *CsbHLH041* transgenic plants than in the *35S* empty vector seedlings (Fig. 6d-f).

To further explore the function of *CsbHLH041* resistance to abiotic stress in plants, transgenic *Arabidopsis* plants overexpressing *CsbHLH041* driven by the CaMV35S promoter were generated. Two independent homozygous lines with relatively high expression levels, *CsbHLH041* OX1 and *CsbHLH041* OX2, were selected for the analysis (Fig. 7a). The salt and ABA tolerance of *CsbHLH041* transgenic plants were assessed. There were no differences in seed germination between WT and *CsbHLH041* transgenic *Arabidopsis* on 1/2 MS (Control) (Fig. 7b). However, the germination ratio of transgenic plants seeds was markedly higher than WT seeds in 1/2 MS medium containing 100 mM NaCl or 2 μM ABA (Fig. 7b-d). Subsequently, the 3-week-old seedlings of *CsbHLH041* transgenic lines and wild-type (WT) plants were treated with 200 mM NaCl and 100 μM ABA, respectively. The leaves of WT plants turned severely yellow after 4 days of 200 mM NaCl or 100 μM ABA treatment, while *CsbHLH041* transgenic lines were still growing with green leaves (Fig. 7e-f). After 8 days, the difference in NaCl or ABA resistance between WT plants and *CsbHLH041* transgenic lines was more obvious, which suggested that *CsbHLH041* transgenic plants were more tolerant to salt and ABA stresses than WT plants.

The protein interaction network predictions for CsbHLH orthologs in *Arabidopsis* that were crucial for the abiotic stress response

Network interaction analysis has been demonstrated to be an effective method to analyse the gene function [40]. We used the online software STRING 10 to predict the protein interaction network with the 142 CsbHLH protein sequences as queries, which formed a complex interaction network constructed with orthologs in *Arabidopsis*. Numerous bHLH proteins interacted with more than one bHLH (Fig. 8a), consistent with previous reports demonstrating that the binding activity of specific DNA sequences depends on the formation of homodimers or heterodimers of different bHLH proteins [2]. In all, 21 proteins were showed that could interact with more than four other bHLH proteins, making them important players in regulating plant growth and stress responses, and detailed information about these orthologs was also summarized in Table S6.

We found that *CsbHLH041* showed a significant response to salt and ABA treatments (Fig. 5a-b), and *CsbHLH041* could enhance tolerance to NaCl and ABA in transgenic *Arabidopsis* and cucumber (Fig. 6; Fig. 7). The function of bHLH proteins occurred mainly through the formation of homodimers or heterodimers with other proteins, which are essential for their binding to related downstream genes [2]. The *CsbHLH041* homologous gene, *AT5G56960*, was located at the centre of the predicted gene association network, which showed that it played main roles in regulating various proteins with different functions (Fig. 8b; Table S6). For example, EP3 might play a role in both normal plant growth and disease resistance [41]. VSP1 and VSP2 are anti-insect proteins and respond to methyl jasmonate and wounding, in which their defense function were correlated with its acid phosphatase activity [42, 43]. The predicted gene association network provides useful resources for subsequent research.

Discussion

Characterization of the cucumber *bHLH* family

The basic helix-loop-helix (bHLH) transcription factor family is the second largest family in eukaryotes [10, 44] and extensive studies of *bHLH* families have been identified in various plants [2]. For example, 166 *bHLH* genes have been identified in *Arabidopsis* [2, 10], 115 *bHLH* genes in *Nelumbo nucifera* [45], 188 *bHLH* genes in apple [40], 167 *bHLH* genes in rice [12] and 159 *bHLH* genes in tomato [13]. The bHLH transcription factors have been involved in multiple biological processes in plants, especially in regulating defense against biotic and abiotic stresses [46]. However, very little information is available about *bHLHs* in cucumber. In our study, 142 *bHLH* cucumber genes were identified and characterized based on the published cucumber annotated genes in the cucumber genome through genome-wide analysis, which were further classified into 32 subgroups based on phylogenetic analyses (Fig. 2a). Multiple sequence alignment of the full-length CsbHLH protein sequences showed that the 142 CsbHLHs contained the conserved bHLH domain, and the number and ratio of the conserved amino acids were similar to *Arabidopsis* (Fig. 1). For example, the two amino acid residues Leu-21 and Leu-53 were relatively

conserved in the helical region that determined whether a dimeric complex could be formed. Moreover, the conserved motif analyses showed that almost all the *CsbHLH* family members had the 1 and 2 motif representing the conserved bHLH domain, and the analysis of the motif and gene structure were further performed to gain evidence to support the phylogenetic relationship for the 142 *CsbHLH* gene family (Fig. 2b-c). In summary, these results indicated that the 142 *CsbHLHs* all contained characteristics of the *bHLH* family, confirming the accuracy of the *bHLH* gene family detection in cucumber.

Phylogenetic analysis and evolution of cucumber *bHLH* genes

In the model plant *Arabidopsis*, the bHLH gene family has been systematically analysed [2, 11]. To explore the evolutionary relationships between 142 CsbHLH proteins in cucumber and 166 AtbHLH proteins in *Arabidopsis*, a phylogenetic tree was constructed based on the protein of 308 bHLHs, which clustered into 23 subfamilies (Fig. 4). There are differences in anatomy and physiology between cucumber and *Arabidopsis*, so some clades may have different modes of expansion in the *bHLH* family of cucumber and *Arabidopsis*. As shown in Fig. 4 and Table S4, not all bHLH members in cucumber were included in these 23 subfamilies, which suggested that there were differences between cucumber and *Arabidopsis* during the process of evolution.

Studies had shown that gene duplication events played a crucial role in the rapid expansion and evolution of gene families [26]. In the cucumber genome, a total of 1468 tandem duplication gene pairs and 231 segmental duplication blocks were identified, respectively (Table S1). Seven segmental duplication events were identified in cucumber using MCScanX, and five tandem duplication gene pairs were obtained in the *CsbHLH* family, respectively (Fig. 3a). In general, the gene functions of a clade are highly conserved among different plant species, but it is not absolute. Therefore, it is of great significance to accurately identify the true orthologs between plant species based on synteny analysis. The results showed that the cucumber genome had extensive synteny with the *Arabidopsis* and tomato genomes, and 944 and 983 syntenic blocks between the cucumber and *Arabidopsis* and tomato genome were identified, respectively (Table S5). Many *CsbHLH* genes showed a linear relationship with the tomato and *Arabidopsis* genes, respectively (Fig. 3b; Table S2).

It has been previously suggested that orthologous genes usually have similar functions and are clustered in the same clade and subclades. As shown in Fig. 4, many cucumber bHLH proteins were clustered into some *Arabidopsis* functional clades, which provided valuable information for studying the function of cucumber *bHLH* genes. CsMYC1 and CsbHLH042 were grouped into subfamily5 along with AtGL3, AtEGL3, AtMYC1 and AtTT8, and were highly homologous to these proteins. In *Arabidopsis*, AtGL3, AtEGL3 and AtTT8 have been demonstrated to be key regulators of anthocyanin and PA biosynthesis [22]. Moreover, AtGL3, AtEGL3 and AtMYC1 were shown to regulate trichome formation and root hair patterning [19, 47]. Therefore, it is possible that CsMYC1 and CsbHLH042 may control trichome formation and PA biosynthesis in cucumber.

Cucumber *bHLH* genes may play important roles in abiotic stress tolerance

In the process of plant response to abiotic stress, bHLH transcription factors act as regulatory genes to regulate the expression changes of related stress genes, thus playing an important role in stress responses. Many studies have shown that bHLH TFs can respond to a range of stresses. For example, in addition to being involved in the morphogenesis of stomata, the TFs INDUCER OF CBF EXPESSION1 (ICE1) and ICE2 in *Arabidopsis* and their homologous genes in other species can play key roles in the response to low temperature stress [32, 46]. *RERJ1* is upregulated in the event of physical damage and drought stress to plants [48]. All these examples indicate that bHLH TFs can play a certain role in response to abiotic stress. However, little is known about the functions of the *bHLH* gene family in cucumber. To better analyse the protein functions of the *bHLH* gene family in cucumber, we conducted a preliminary analysis of three aspects to reveal the functions of the *CsbHLH* gene family.

How cis-elements in the promoters of the *bHLH* genes respond to the environment will affect their roles in stimulating and regulating gene expression. Cis-element analyses showed that elements that might respond to diverse stresses (such as LTR, TCA-element and MBS) were widely distributed in the *CsbHLH* gene family (Fig. S3). Most cucumber *bHLH* gene promoters contained MYB binding sites, which are involved in drought response (Table S3), which reflected that they could be regulated by MYB transcription factors under drought stress. Many cucumber *bHLH* gene promoters contained ABRE and TC-rich elements, which are involved in ABA-dependent or independent stress tolerance [49]. Therefore, according to the cis-acting element contained on the promoter, these genes may play important roles in gene regulation in response to different stresses in cucumber. In addition, the functions of 50 CsbHLHs were predicted based on their known and verified homologs in *Arabidopsis* by a phylogenetic tree between cucumber and *Arabidopsis* (Table S4), which were mainly associated with development processes and stress responses (Table S4). For the third aspect, we also used *Arabidopsis* orthologs to predict the regulatory networks for 142 *CsbHLH* genes, and the results suggested that many member genes could respond to stimuli (Table S6). For example, *bHLH093* and *ICE1* were involved in the ABA signalling pathway, and they were crucial for abiotic stress responses in plants [49,50]. All these analyses suggest that the *bHLH* gene family may also be related to plant development, metabolic regulation, and the response to stress in cucumber, consistent with previous research [10, 12]. Subsequently, we analysed and screened *CsbHLH* genes that might respond to stress, as it is very important to improve stress tolerance of cucumber. According to *cis*-element analyses, TC-rich cis-elements that may be involved in defense and stress responses were detected in the promoter regions of 60 *CsbHLH* genes (Fig. S3). Moreover, the promoters of 41 *CsbHLHs* contained the LTR element, which responds to cold stress and 106 *CsbHLHs* contained the ABA-responsive element, which responds to ABA stress. The phylogenetic analyses between cucumber and *Arabidopsis* further showed that 25 *CsbHLHs* might respond to abiotic stresses, including cold, ABA, salt and drought based on their homologs in *Arabidopsis* (Table S4). Through comprehensive analysis, we carefully screened 21, 20 and 25 *bHLH* genes that were likely to respond to low temperature (4°C), salt (NaCl) and ABA, respectively. The screened *CsbHLH* genes all

responded to stress treatments under the respective stress conditions (Fig. 5). *CsbHLH041* was induced by salt and ABA (Fig. 5a-b), and *35S:CsbHLH041* transgenic *Arabidopsis thaliana* and transient transformed cucumber cotyledons were shown to have enhanced tolerance to salt and ABA (Fig. 6; Fig. 7). The results of these analyses indicated that functional prediction of the *CsbHLH* gene family could provide valuable reference data for further functional studies of this gene family in cucumber.

Conclusions

Our study investigated the *bHLH* family genes in detail in cucumber. We also performed expression analyses of the selected genes under different stress treatments, and detailed functions of *CsbHLH041* using the transgenic method. This work provides abundant insights into the functions and regulatory mechanisms of CsbHLH proteins in cucumber abiotic stress tolerance and growth and development.

Methods

Genome-wide identification of the *CsbHLH* genes in *cucumber*

To identify the *CsbHLH* gene family members from the entire cucumber genome database, 166 *Arabidopsis* bHLH proteins were used as query sequences and BlastP searches against the predicted cucumber proteins. In addition, the Hidden Markov Model (HMM) profile of the bHLH domain (PF00010) from the Pfam database (available online: http://pfam.janelia.org) was also applied as a query to search the *bHLH* genes from the cucumber genome database. All candidate genes were further examined the bHLH domains by using the Pfam and Simple Modular Architecture Research Tool (SMART) program (http://smart.embl-heidelberg.de) [24].

Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignments were performed using ClustalX 1.81 with default parameters and DNAMAN software (http://dnaman.software.informer.com/). The sequence logos for bHLHs were obtained by submitting the multiple alignment sequences to the website (http://weblogo.berkeley.edu/logo.cgi) [51]. A phylogenetic tree was constructed with the aligned fully predicted protein sequences of 142 *bHLH* genes using MEGA7 (https://www.megasoftware.net/) [52]. The neighbour-joining (NJ) method was used with the following parameters: Poisson correction, pairwise deletion, and bootstrap (1,000 replicates; random seed). The phylogenetic tree was visualized by plotting it using the EvolView tool (http://www.evolgenius.info). Classification of the *CsbHLH* genes was then performed according to their phylogenetic relationships with their corresponding *Arabidopsis bHLH* genes.

Gene structure and conserved motif analysis

The DNA and cDNA sequences corresponding to each predicted gene from the cucumber genome were downloaded, and then the 142 *CsbHLH* gene structures were analysed using the web-based bioinformatics tool GSDS (http://gsds.cbi.pku.edu.cn/) [53]. Conserved motif structures in CsbHLHs were identified using MEME (http://meme-suite.org/index.html) [26].

Chromosomal distribution and gene duplication

All *CsbHLH* genes were mapped to cucumber chromosomes based on physical location information from the database of the cucumber genome using TBtools [26]. The gene duplication events were assessed using the Multiple Collinearity Scan toolkit (MCScanX), with the default parameters [54]. To identify the synteny relationships of the orthologous *bHLH* genes obtained from cucumber, *Arabidopsis* and tomato, the syntenic analysis maps were constructed using TBtools [26].

Analysis of the *bHLH* gene promoter in cucumber

We downloaded the entire cucumber genome sequence from the cucumber genome database (Chinese Long 9930) and extracted the 2-kb long sequences upstream of the transcription start site of these 142 *CsbHLH* genes. The cis-acting elements on the promoter regions of these genes were analysed using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) software [55].

Plant materials and growth conditions

Cucumber (*Cucumis sativus* L. cv 'Xintaimici') seeds, provided by Professor Chenxing Cao (Shandong Agricultural University), were germinated on moist filter paper in an incubator at 28°C for 1 day. The germinated seeds were sown into soil mixture in an ordinary illuminated incubator at Shandong Agricultural University. After 10 days, batches of 12 seedlings were transferred to a plastic tank filled with an aerated nutrient solution (pH 6.0-6.5) containing the following: Ca $(NO_3)_2$: 3.5 mM, KNO_3: 7 mM, KH_2PO_4: 0.78 mM, $MgSO_4$: 2 mM, H_3BO_3: 29.6 μM, $MnSO_4$: 10 μM, Fe-EDTA: 50 μM, $ZnSO_4$: 1.0 μM, H_2MoO_4: 0.05 μM and $CuSO_4$: 0.95 μM. The experiment was carried out as previously described[56]

RNA extraction and qRT-PCR analysis

Total RNA was isolated from cucumber and *Arabidopsis* plants using an RNAprep pure Plant Kit (TianGen, Beijing, China), following the manufacturer's instructions. Subsequently, reverse transcribed using the PrimeScript®1st Strand cDNA Synthesis Kit (Takara, Japan). The qRT-PCR reactions were performed using the UltraSYBR Mixture (with ROX I; Cwbiotech) with the iCycler iQ5 system (BioRad, CA,

USA). The results were normalized to those of the cucumber *Actin* gene. Three biological replicates were used for each analysis. The primers used in this study are provided in Table S7.

Overexpression vector construction, *Arabidopsis* transformation and transient transformation in cucumber cotyledons

The full-length coding sequence of *CsbHLH041* was recombined into the pCAMBIA1300 vector. The construct was transformed into *Agrobacterium tumefaciens* LBA4404, which was used for transformation of *Arabidopsis* plants and 8-d-old cucumber cotyledons [57]. The *Arabidopsis* seeds were Colombia (Col-0), which were bred in our laboratory. Homozygous T3 transgenic *Arabidopsis* lines were identified by hygromycin (300 mg/L) selection.

Abiotic stress tolerance assays and ABA sensitivity analysis

For *Arabidopsis* salt stress and ABA treatment, the seeds of *CsbHLH041* T3-generation homozygous lines and Col-0 (WT) were sown in vermiculite soil in pots and cultured under normal conditions at 22 °C for 3 weeks. For salt treatment, the 3-week-old seedlings were watered with 200 mM NaCl solution every other day, and the growth of Col-0 (WT) and *CsbHLH041* transgenic lines was observed every 4 days. For ABA treatments, the 3-week-old seedlings were watered with 100 μM ABA solution every other day, and phenotypes were evaluated every 4 days. To check the seed germination rate in response to salt stress and ABA treatment, the seeds of Col-0 (WT) and transgenic lines were surface sterilized and sown in 1/2 MS medium supplemented with 2 μM ABA or 100 mM NaCl, respectively, under normal conditions at 22 °C in a growth chamber. The germination rate was scored on the 7th day after culturing on the plates.

To determine the salt tolerance and ABA sensitivity in cotyledons of 8-d-old cucumber seedlings with transient infiltration of *35S* and *35S:CsbHLH041*, selected seedlings with equivalent growth were transferred to 6 L nutrient solution for hydroponic growth. Hoagland nutrient solution was used for culture, and the seedlings were grown hydroponically for two days before salt and ABA treatment. They were then treated with salt and ABA, and the final concentration in the medium was 100 mM and 100 μM, respectively. To ensure the reliability of the experiment, the cucumber seedlings with transient infiltration of *35S* and *35S:CsbHLH041* were cultured in the same hydroponic box. The changes in transgenic and control seedlings were observed at different time periods.

Determination of physiological parameters

The cucumber cotyledons of *35S* empty vector and *35S:CsbHLH041* seedlings were collected at different time points during salt and ABA stress treatment, then frozen in liquid nitrogen for subsequent

experiments. The activity of superoxide dismutase (SOD), peroxidase (POD), and catalase (CAT) were determined as previously described [58].

Protein association network predictions and functional annotations by STRING

The 142 CsbHLH protein sequences were submitted to the online server STRING (version 10.0; http://string-db.org), with the organism specified as "*Arabidopsis thaliana.*" After the blast step was finished, genes with the highest bitscores were used to construct the network. The bHLHs that did not interact with any others were removed. The functional annotation information was copied manually from the blast results.

Abbreviations

bHLH, basic Helix-Loop-Helix; At, *Arabidopsis thaliana*; Cs, *Cucumis sativus* L; MS, Murashige and Skoog; qRT-PCR, quantitative reverse transcription-PCR; CDS, Coding Sequence; ABA, Abscisic acid; pI, isoelectric point; WT, wild type.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the results are included within the article and its additional files. Other relevant materials are available from the corresponding authors on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by fundings from the National Natural Science Foundation of China (31672170 and 31872950), the Natural Science Foundation of Shandong Province (JQ201309), the Shandong "Double Tops" Program (SYL2017YSTD06) and the 'Taishan Scholar' Foundation of the People's Government of Shandong Province (ts20130932). The funds played no role in study design, data analysis, and manuscript preparation.

Authors' contributions

JL and ZR conceived and designed the experiments. JL, TW and JH performed the experiments. JL analyzed the data and wrote the manuscript. ZR revised the manuscript. All authors have read and approved this manuscript.

Acknowledgments

We thank Dr. Chenxing Cao, College of Horticultural Science and Engineering, Shandong Agricultural University, for providing cucumber (*Cucumis sativus* L. cv 'Xintaimici') seeds.

References

1. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, et al. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science. 2000;290:2105-2110.
2. Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martinez-Garcia JF, Bilbao-Castro JR, Robertson DL. Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in *Arabidopsis*, poplar, rice, moss, and algae. Plant Physiol. 2010;153 3):1398-1412.
3. Ferré-D'Amaré AR, Prendergast GC, Ziff EB, Burley S K. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature. 1993;363:38–45.
4. Murre C, Mccaw PS, Baltimore D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell. 1989;56:777–783.
5. Shimizu T, Toumoto A, Ihara K, Shimizu M, Kyogoku Y, Ogawa N, Oshima Y, Hakoshima T. Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. EMBO J. 1997;16:4689–4697.
6. Massari ME, Murre C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol. 2000;20:429–440.
7. Nair SK, Burley SK. Recognizing DNA in the library. Nature. 2000; 404:715,717–718.

8. Atchley WR, Terhalle W, Dress A. Positional dependence, cliques, and predictive motifs in the bHLH protein domain. Journal of Molecular Evolution. 1999;48:501–516.

9. David C, Pierre K, Morgane TC, Gemma R, Valérie L, Elena S, Degnan BM, Michel V. Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. BMC Evol. Biol. 2007;7:33.

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

11. Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC. The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Mol Biol Evol. 2003;20(5):735-747.

12. Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, et al. Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and *Arabidopsis*. Plant Physiol. 2006; 141(4):1167-1184.

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

14. Ludwig SR, Habera LF, Dellaporta SL, Wessler SR. Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. Proc Natl Acad Sci U S A. 1989;86:7092-7096.

15. Paik I, Kathare PK, Kim JI, Huq E. Expanding Roles of PIFs in Signal Integration from Multiple Processes. Mol Plant. 2017;10(8):1035-1046.

16. Lee S, Lee S, Yang KY, Kim YM, Park SY, Kim SY, et al. Overexpression of PRE1 and its homologous genes activates Gibberellin-dependent responses in *Arabidopsis thaliana*. Plant Cell Physiol. 2006;47(5):591-600.

17. Feyissa DN, Lovdal T, Olsen KM, Slimestad R, Lillo C. The endogenous GL3, but not EGL3, gene is necessary for anthocyanin accumulation as induced by nitrogen depletion in *Arabidopsis* rosette stage leaves. Planta. 2009;230(4):747-754.

18. Appelhagen I, Jahns O, Bartelniewoehner L, Sagasser M, Weisshaar B, Stracke R. Leucoanthocyanidin Dioxygenase in *Arabidopsis thaliana*: characterization of mutant alleles and regulation by MYB-BHLH-TTG1 transcription factor complexes. Gene. 2011;484(1-2):61-68.

19. Ramsay NA, Glover BJ. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. 2005;10(2):63-70.

20. Zhai Y, Zhang L, Xia C, Fu S, Zhao G, Jia J, et al. The wheat transcription factor, TabHLH39, improves tolerance to multiple abiotic stressors in transgenic plants. Biochem Biophys Res Commun. 2016;473(4):1321-1327.

21. Fernández-Calvo P, Chini A, Fernandez-Barbero G, Chico JM, Gimenez-Ibanez S, Geerinck J, et al. The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. Plant Cell. 2011;23(2):701-715.

22. Li S. Transcriptional control of flavonoid biosynthesis. Plant Signaling & Behavior. 2014;9:e27522.

23. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, et al. The genome of the cucumber, *Cucumis sativus* Nat Genet. 2009;41(12):1275-1281.

24. Letunic I, Doerks T, Bork P. SMART 6: recent updates and new developments. Nucleic Acids Res 37 (Database issue). 2009;D229-232.

25. Cheng X, Xiong R, Liu H, Wu M., Chen F, Hanwei Y, Xiang Y. Basic helix-loop-helix gene family: Genome wide identification, phylogeny, and expression in Moso bamboo. Plant Physiol Biochem. 2018;132:104-119.

26. Sun H, Pang B, Yan J, Wang T, Wang L, Chen C, et al. Comprehensive Analysis of Cucumber Gibberellin Oxidase Family Genes and Functional Characterization of CsGA20ox1 in Root Development in *Arabidopsis*. Int J Mol Sci. , 2018;19(10).

27. Babitha KC, Vemanna RS, Nataraja KN, Udayakumar M. Overexpression of EcbHLH57 Transcription Factor from Eleusine coracana L. in Tobacco Confers Tolerance to Salt, Oxidative and Drought Stress. PLoS One. 2015;10(9):e0137098.

28. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell. 2003;15(1):63-78.

29. Petridis A, Doll S, Nichelmann L, Bilger W, Mock HP. Arabidopsis thaliana G2-LIKE FLAVONOID REGULATOR and BRASSINOSTEROID ENHANCED EXPRESSION1 are low-temperature regulators of flavonoid accumulation. New Phytol. 2016;211(3):912-925.

30. Sasaki-Sekimoto Y, Saito H, Masuda S, Shirasu K, Ohta H. Comprehensive analysis of protein interactions between JAZ proteins and bHLH transcription factors that negatively regulate jasmonate signaling. Plant Signal Behav. 2014;9(1):e27639.

31. Song S, Qi T, Fan M, Zhang X, Gao H, Huang H, et al. The bHLH subgroup IIId factors negatively regulate jasmonate-mediated plant defense and development. PLoS Genet. 2013;9(7):e1003653.

32. Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, Zhu JK. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. Genes Dev. 2003;17(8):1043-1054.

33. Ahmad A, Niwa Y, Goto S, Ogawa T, Shimizu M, Suzuki A,et al. bHLH106 Integrates Functions of Multiple Genes through Their G-Box to Confer Salt Tolerance on *Arabidopsis*. PLoS One. 2015;10(5):e0126872.

34. Le Hir R, Castelain M, Chakraborti D, Moritz T, Dinant S, Bellini C. AtbHLH68 transcription factor contributes to the regulation of ABA homeostasis and drought stress tolerance in *Arabidopsis thaliana*. Physiol Plant. 2017;160(3):312-327.

35. Kurt F, Filiz E. Genome-wide and comparative analysis of *bHLH38*, *bHLH39*, *bHLH100* and *bHLH101* genes in *Arabidopsis*, tomato, rice, soybean and maize: insights into iron (Fe) homeostasis. Biometals. 2018;31(4):489-504.

36. Sharma N, Xin R, Kim DH, Sung S, Lange T, Huq E. NO FLOWERING IN SHORT DAY (NFL) is a bHLH transcription factor that promotes flowering specifically under short-day conditions in *Arabidopsis*. Development. 2016;143(4):682-690.

37. Karas B, Amyot L, Johansen C, Sato S, Tabata S, Kawaguchi M, et al. Conservation of lotus and *Arabidopsis* basic helix-loop-helix proteins reveals new players in root hair development. Plant Physiol. 2009;151(3):1175-1185.

38. Yang KZ, Jiang M, Wang M, Xue S, Zhu LL, Wang HZ, et al. Phosphorylation of Serine 186 of bHLH Transcription Factor SPEECHLESS Promotes Stomatal Development in *Arabidopsis*. Mol Plant. 2015;8(5):783-795.

39. Chakraborty M, Gangappa SN, Maurya JP, Sethi V, Srivastava AK, Singh A, et al. Functional interrelation of MYC2 and HY5 plays an important role in *Arabidopsis* seedling development. Plant J. 2019;99(6):1080-1097.

40. Mao K, Dong Q, Li C, Liu C, Ma F. Genome Wide Identification and Characterization of Apple bHLH Transcription Factors and Expression Analysis in Response to Drought and Salt Stress. Front Plant Sci. 2017;8:480.

41. Passarinho PA, Van Hengel AJ, Fransz PF, de Vries SC. Expression pattern of the *Arabidopsis thaliana AtEP3/AtchitIV* endochitinase gene. Planta. 2001;212(4):556-567.

42. Berger S, Bell E, Mullet JE. Two Methyl Jasmonate-Insensitive Mutants Show Altered Expression of AtVsp in Response to Methyl Jasmonate and Wounding. Plant Physiol. 1996;111:525-531.

43. Chi YH, Jing X, Lei J, Ahn JE, Koo YD, Yun DJ, et al. Stability of AtVSP in the insect digestive canal determines its defensive capability. Journal of Insect Physiology. 2011;57(3):391-399.

44. Ledent V, Vervoort M. The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. Genome Res. 2001;11:754-770.

45. Mao TY, Liu YY, Zhu HH, Zhang J, Yang JX, Fu Q, et al. Genome-wide analyses of the bHLH gene family reveals structural and functional characteristics in the aquatic plant *Nelumbo nucifera.* 2019;7:e7153.

46. Feng XM, Zhao Q, Zhao LL, Qiao Y, Xie XB, Li HF, et al. The cold-induced basic helix-loop-helix transcription factor gene *MdCIbHLH1* encodes an ICE-like protein in apple. BMC Plant Biology. 2012;12(1):22.

47. Zhao H, Wang X, Zhu D, Cui S, Li X, Cao Y, et al. A single amino acid substitution in IIIf subfamily of basic helix-loop-helix transcription factor AtMYC1 leads to trichome and root hair patterning defects by abolishing its interaction with partner proteins in *Arabidopsis*. J Biol Chem. 2012;287(17):14109-14121.

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

49. Chen L, Chen Y, Jiang J, Chen S, Chen F, Guan Z, Fang W. The constitutive expression of Chrysanthemum dichrum ICE1 in *Chrysanthemum grandiflorum* improves the level of low temperature, salinity and drought tolerance. Plant Cell Rep. 2012;31(9):1747-1758.

50. An JP, Zhang XW, Bi SQ, You CX, Wang XF, Hao YJ. MdbHLH93, an apple activator regulating leaf senescence, is regulated by ABA and MdBT2 in antagonistic ways. New Phytol. 2019:222(2):735-

751.

51. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14:1188-1190.

52. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870-1874.

53. Guo AY, Zhu QH, Chen X, Luo JC. GSDS: A gene structure display server. Yi Chuan. 2007;29:1023–1026.

54. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40:e49.

55. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, De Peer YV, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002;30:325–327.

56. Li Q, Zhao P, Li J, Zhang C, Wang L, Ren Z. Genome-wide analysis of the WD-repeat protein family in cucumber and Arabidopsis. Mol Genet Genomics. 2014;289:103-124.

57. Liu M, Zhang C, Duan L, Luan Q, Li J, Yang A, et al. CsMYB60 is a key regulator of flavonols and proanthocyanidans that determine the colour of fruit spines in cucumber. J Exp Bot. 2019;70(1):69-84.

58. Zhang X, Chen L, Shi Q, Ren Z. *SlMYB102*, an R2R3-type MYB gene, confers salt tolerance in transgenic tomato. Plant Sci. 2020;291:110356.

Table

Table 1 *bHLH* genes in Cucumber

CsbHLH	Gene ID	Location	Gene length	Amino acid length	pI
001	CsaV3_1G005290	Chr1:3503806-3504411	1909	84	5.04
002	CsaV3_1G011300	Chr1:6972358-6976069	4167	236	5.78
003	CsaV3_3G022420	Chr3:19737551-19739874	958	502	5.7
004	CsaV3_3G049150	Chr3:40071324-40074834	1400	689	5.11
005	CsaV3_3G001710	Chr3:1295970-1297898	5601	643	6.21
006	CsaV3_3G000850	Chr3:656062-658628	6057	448	8.65
007	CsaV3_1G039580	Chr1:24945063-24953384	1684	319	5.78
008	CsaV3_2G007370	Chr2:3725743-3731272	605	707	6.09
009	CsaV3_4G032110	Chr4:22635255-22641052	3486	551	6.25
010	CsaV3_1G043790	Chr1:28804330-28811620	6167	960	6.4
011	CsaV3_2G015700	Chr2:13017831-13023291	1802	336	5.64
012	CsaV3_6G000530	Chr6:351129-356108	2050	645	5.51
013	CsaV3_3G007980	Chr3:6919411-6921670	2306	650	5.83
014	CsaV3_2G010120	Chr2:6877325-6878870	2524	323	6.02
015	CsaV3_7G025510	Chr7:14980031-14984292	3589	533	6.06
016	CsaV3_6G009090	Chr6:7311297-7315687	3711	486	6.35
017	CsaV3_2G028950	Chr2:18953613-18956324	3528	348	8.3
018	CsaV3_7G007460	Chr7:4647975-4650800	3316	335	6.64
019	CsaV3_6G044570	Chr6:26373287-26375739	3723	330	5.3
020	CsaV3_6G044560	Chr6:26366139-26368813	11538	309	5.81
021	CsaV3_5G026500	Chr5:21650877-21655135	2678	624	5.04
022	CsaV3_6G044730	Chr6:26485325-26487050	1415	342	4.86
023	CsaV3_2G030090	Chr2:19685080-19689447	4059	363	4.57
024	CsaV3_6G043370	Chr6:25541506-25544586	8321	416	5.2
025	CsaV3_6G044580	Chr6:26382822-26384588	3478	276	6.05
026	CsaV3_2G035250	Chr2:23586273-23591605	3855	239	6.77
027	CsaV3_2G008770	Chr2:5179234-5186154	7290	246	4.92
028	CsaV3_6G008940	Chr6:7177946-7179613	5695	432	5.42
029	CsaV3_6G014370	Chr6:10430376-10432021	6113	308	4.98
030	CsaV3_5G033960	Chr5:27092009-27094880	4184	372	5.77
031	CsaV3_6G036080	Chr6:20032486-20036831	1493	242	9.03
032	CsaV3_1G033410	Chr1:20481133-20483811	5529	256	9.24
033	CsaV3_1G009900	Chr1:6174783-6178372	6920	551	5.6
034	CsaV3_2G001440	Chr2:370393-376506	1545	543	8.37
035	CsaV3_1G031920	Chr1:18957508-18969046	2275	242	5.03
036	CsaV3_7G004510	Chr7:3234760-3236331	1320	211	6.77
037	CsaV3_4G034440	Chr4:24394485-24395550	3489	239	7.1
038	CsaV3_4G029740	Chr4:19342473-19344831	5460	253	6.17
039	CsaV3_3G000950	Chr3:733695-738681	1011	773	5.11
040	CsaV3_2G026610	Chr2:18201502-18202748	2627	240	7.26
041	CsaV3_1G040580	Chr1:25826012-25829490	1816	492	7.03
042	CsaV3_6G037080	Chr6:20849605-20855801	4137	651	5.91
043	CsaV3_6G041730	Chr6:24304724-24306996	1246	235	6.42
044	CsaV3_1G003910	Chr1:2423148-2424832	2711	266	6.87
045	CsaV3_3G013690	Chr3:10293079-10294740	2245	191	8.64
046	CsaV3_1G006280	Chr1:4002735-4008902	4367	566	9.01
047	CsaV3_1G002260	Chr1:1450554-1451954	1411	261	6.16
048	CsaV3_3G039100	Chr3:32125102-32130846	2115	370	5.74

049	CsaV3_5G033600	Chr5:26846039-26850063	5332	571	6.09
050	CsaV3_1G006650	Chr1:4277323-4279125	2566	279	6.1
051	CsaV3_6G001900	Chr6:1303975-1307218	4986	248	9.26
052	CsaV3_1G037610	Chr1:23567571-23568986	1928	309	4.78
053	CsaV3_6G037070	Chr6:20836169-20845986	5647	263	6.08
054	CsaV3_2G026190	Chr2:17969523-17971339	1844	330	5.15
055	CsaV3_3G034600	Chr3:29292016-29296383	2259	695	5.24
056	CsaV3_3G044120	Chr3:35995420-35997451	4852	272	5.09
057	CsaV3_2G005070	Chr2:2751170-2752663	3663	318	5.74
058	CsaV3_2G014750	Chr2:12321166-12324655	1661	343	6.2
059	CsaV3_7G027630	Chr7:17195587-17199883	1509	317	6
060	CsaV3_2G025890	Chr2:17778099-17780726	3295	342	6.26
061	CsaV3_2G030500	Chr2:20027274-20029389	2181	372	4.71
062	CsaV3_3G015900	Chr3:11810548-11813843	2822	547	6.88
063	CsaV3_7G000080	Chr7:185656-189146	2627	457	5.85
064	CsaV3_1G000190	Chr1:132006-133915	2323	276	7.03
065	CsaV3_3G028610	Chr3:25161463-25169489	3059	540	5.7
066	CsaV3_2G003660	Chr2:1833037-1837221	6474	422	6.1
067	CsaV3_4G002800	Chr4:1745825-1749669	8026	168	7.65
068	CsaV3_1G045830	Chr1:31650489-31656184	4367	395	6.39
069	CsaV3_4G026430	Chr4:15704947-15706451	3065	196	6.51
070	CsaV3_3G007090	Chr3:6381666-6383510	5744	359	5.94
071	CsaV3_3G049050	Chr3:40003473-40010806	1802	322	7.73
072	CsaV3_1G005810	Chr1:3727778-3731264	2031	443	9.02
073	CsaV3_2G026540	Chr2:18154564-18158701	3855	380	5.01
074	CsaV3_4G035310	Chr4:24884098-24888333	2900	403	5.7
075	CsaV3_3G020750	Chr3:16961726-16964548	7333	248	7.09
076	CsaV3_5G018750	Chr5:14293739-14297926	3510	534	5.2
077	CsaV3_4G034660	Chr4:24537693-24539435	1700	409	6.38
078	CsaV3_7G026520	Chr7:16047841-16051016	3844	490	6.04
079	CsaV3_1G001960	Chr1:1286661-1290828	5186	196	7.64
080	CsaV3_5G040480	Chr5:31836009-31840665	2228	245	5.67
081	CsaV3_6G002130	Chr6:1472037-1473566	1504	161	5.12
082	CsaV3_1G028780	Chr1:15708212-15711935	2358	423	6.34
083	CsaV3_5G024030	Chr5:18711066-18713891	3741	285	6.21
084	CsaV3_4G000380	Chr4:228607-230307	5797	342	4.89
085	CsaV3_4G034980	Chr4:24696914-24698535	1065	245	6.18
086	CsaV3_3G014190	Chr3:10642450-10643959	1742	203	7.81
087	CsaV3_7G035000	Chr7:22142359-22144409	1621	394	6.54
088	CsaV3_2G016810	Chr2:14091802-14092813	4235	216	9.77
089	CsaV3_6G012850	Chr6:8973920-8976038	2826	262	9.23
090	CsaV3_5G031540	Chr5:25725825-25732273	1729	679	7.3
091	CsaV3_2G011050	Chr2:8296805-8299080	1004	496	5.5
092	CsaV3_2G013060	Chr2:10626443-10627763	5030	280	9.14
093	CsaV3_3G048260	Chr3:39394424-39397324	4187	326	4.94
094	CsaV3_1G039160	Chr1:24662874-24666933	2825	359	5.94
095	CsaV3_7G008580	Chr7:5322491-5324405	2588	254	7.22
096	CsaV3_2G029940	Chr2:19575197-19577442	4258	308	6.01
097	CsaV3_3G011010	Chr3:8707121-8711973	6448	382	5.08
098	CsaV3_3G021970	Chr3:19073471-19076098	2460	377	5.08
099	CsaV3_6G028530	Chr6:16812784-16815830	2030	207	11.79

100	CsaV3_6G037460	Chr6:21181685-21184516	4024	299	5.66
101	CsaV3_4G029750	Chr4:19361035-19364776	2871	211	9.21
102	CsaV3_6G033930	Chr6:18737868-18742274	1923	333	5.71
103	CsaV3_6G036240	Chr6:20141242-20142683	4656	97	9.18
104	CsaV3_3G012210	Chr3:9421374-9425037	4979	227	5.7
105	CsaV3_3G022870	Chr3:20405212-20408271	3243	236	6.13
106	CsaV3_3G042970	Chr3:34884792-34886594	1529	244	8.4
107	CsaV3_6G018830	Chr6:13512926-13515085	1667	253	7.06
108	CsaV3_2G030310	Chr2:19895721-19897132	4390	240	9.2
109	CsaV3_1G002670	Chr1:1668618-1674219	2118	359	8.65
110	CsaV3_3G045440	Chr3:37108680-37112535	1645	430	6.69
111	CsaV3_5G012430	Chr5:7900420-7905450	2159	451	6.33
112	CsaV3_1G011460	Chr1:7106592-7110120	3046	360	4.75
113	CsaV3_7G008090	Chr7:5057906-5059973	4406	255	6.11
114	CsaV3_5G026380	Chr5:21538651-21541239	4345	173	9.15
115	CsaV3_UNG229040	scaffold115:93241-95547	1441	274	9.42
116	CsaV3_3G027730	Chr3:24067204-24073678	9817	529	5.88
117	CsaV3_7G003870	Chr7:2853486-2854424	6196	313	5.32
118	CsaV3_3G016560	Chr3:12365460-12367641	2831	253	8.6
119	CsaV3_1G012350	Chr1:7668571-7671887	2272	205	6.16
120	CsaV3_5G003430	Chr5:2204438-2205442	3080	256	7.75
121	CsaV3_6G047120	Chr6:27815147-27818761	2674	337	6.13
122	CsaV3_1G042640	Chr1:27559193-27563048	2452	438	7.72
123	CsaV3_3G005540	Chr3:4716554-4722201	1766	439	6.41
124	CsaV3_6G046660	Chr6:27537459-27541954	1725	357	8.44
125	CsaV3_5G003420	Chr5:2191536-2193265	1497	261	5.59
126	CsaV3_5G003410	Chr5:2180337-2183163	4495	252	7.01
127	CsaV3_6G049510	Chr6:28902722-28905982	3614	405	5.35
128	CsaV3_4G003860	Chr4:2368080-2373266	3260	357	8.51
129	CsaV3_7G031270	Chr7:19779388-19783470	749	420	8.28
130	CsaV3_3G039080	Chr3:32107556-32110621	3490	367	9.16
131	CsaV3_6G045070	Chr6:26672336-26673833	938	229	10.26
132	CsaV3_6G051560	Chr6:29996501-29997250	1571	250	5.66
133	CsaV3_7G027460	Chr7:17033739-17042484	2825	692	5.66
134	CsaV3_5G037950	Chr5:30085680-30087603	1391	92	9.09
135	CsaV3_1G002240	Chr1:1440343-1441301	2067	93	9.09
136	CsaV3_5G032530	Chr5:26313766-26315796	1914	96	9.17
137	CsaV3_1G009880	Chr1:6153304-6155828	4261	373	6.61
138	CsaV3_7G033460	Chr7:21082838-21087117	3175	298	6.78
139	CsaV3_7G007860	Chr7:4913547-4914938	8745	211	6.35
140	CsaV3_4G010010	Chr4:7769516-7771744	4296	333	5.11
141	CsaV3_1G003270	Chr1:2026829-2032886	4082	619	9.19
142	CsaV3_5G031750	Chr5:25868912-25871372	4279	365	5.84

Figures

a

b

Sequence logo
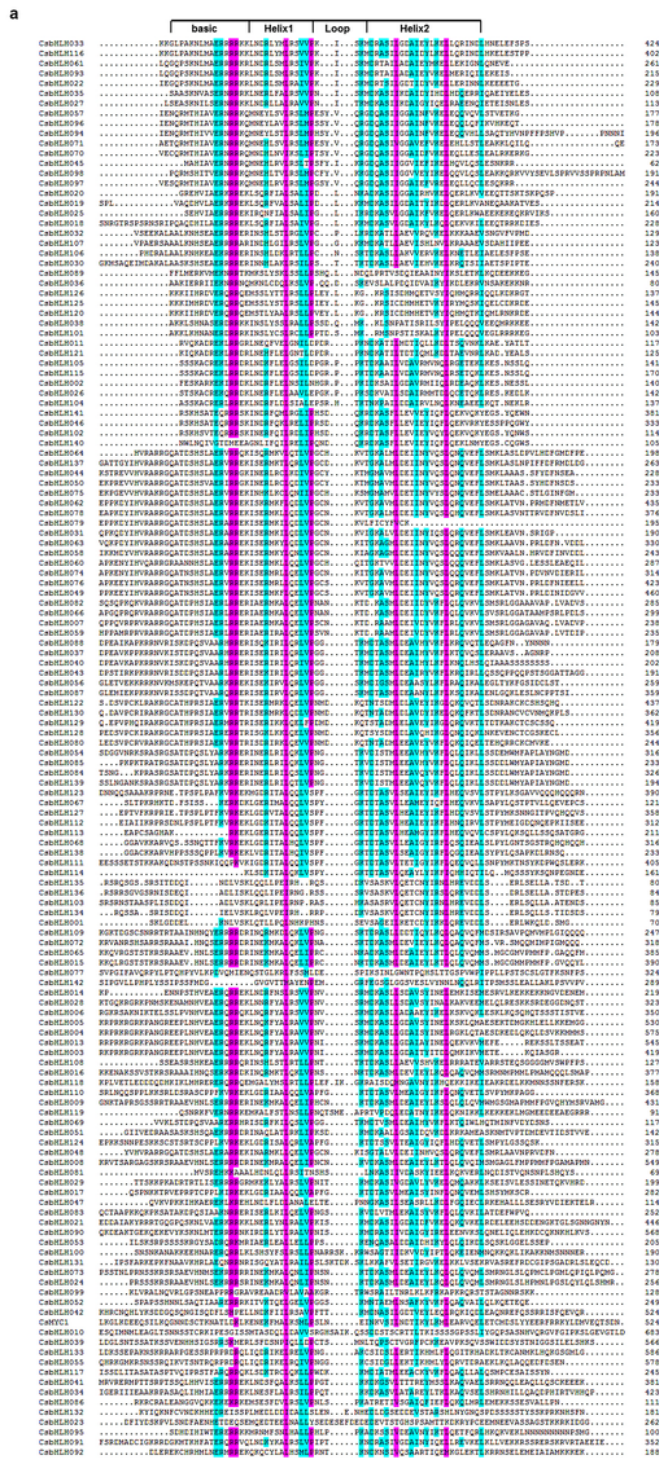
Figure 1

Conserved amino acids and multiple sequence alignment schematic diagrams of the CsbHLHs bHLH domains. (a) Multiple sequence alignments of CsbHLH proteins. The CsbHLH conserved sequences were marked with a purple background for an amino acid identity greater than 75% and a light blue background for an amino acid identity greater than 50%. The bHLH domains were labelled. (b) Sequence

logo of CsbHLH domains. The overall height of each stack represented the conservation of the sequence at that position.
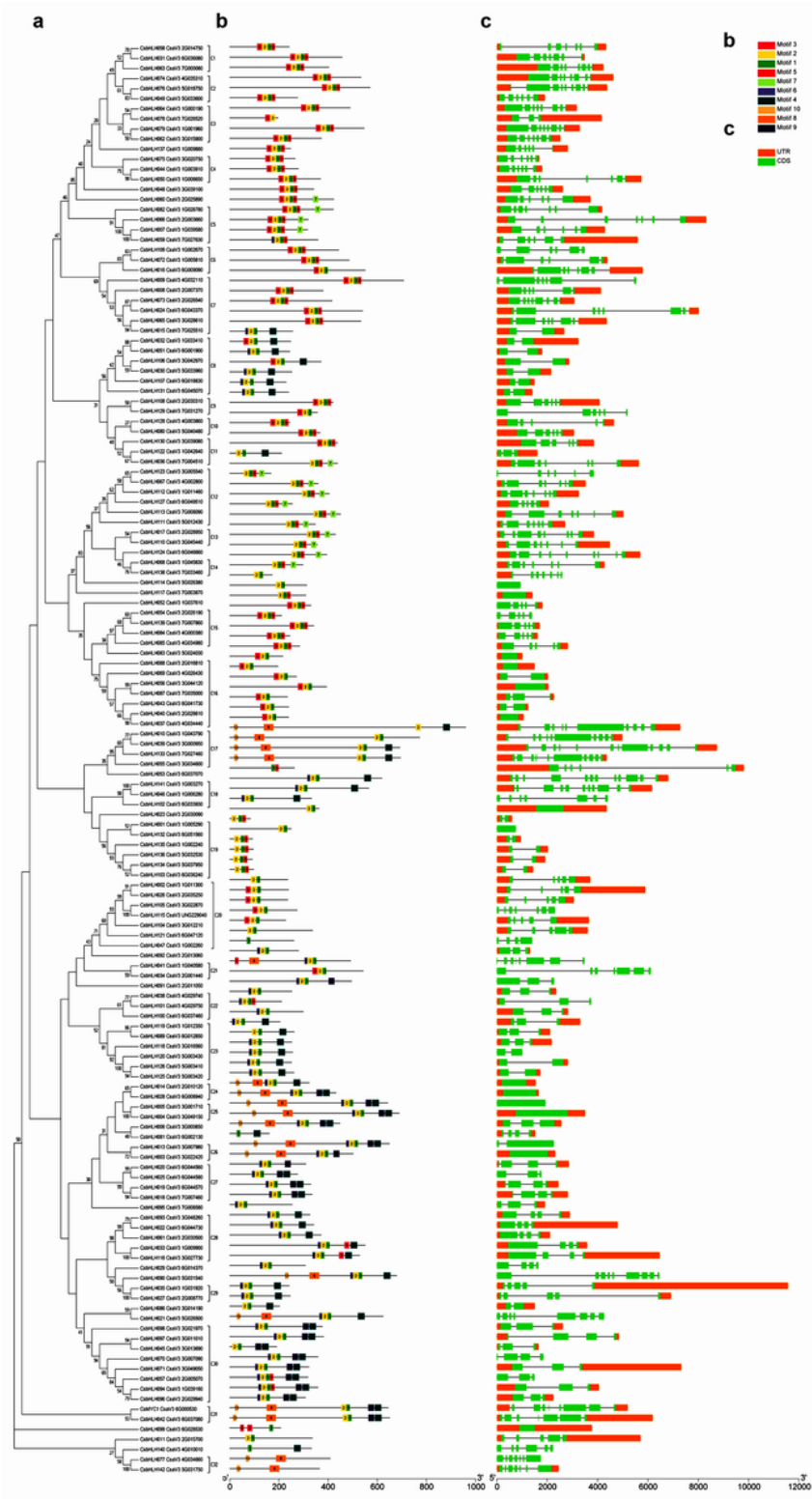
Figure 2

Phylogenetic relationships, gene structure and conserved protein motifs in bHLH genes from cucumber. (a) The phylogenetic tree was constructed based on the full-length protein sequences of 142 CsbHLH proteins using MEGA 7.0 software. The tree showed the 32 phylogenetic subgroups (C1-C32) with high

bootstrap value. (b) Conserved motifs in CsbHLH proteins. The motifs, numbers 1-10, were displayed in different coloured boxes. The sequence logos and E values for each motif were shown in Fig. S1. (c) Exon-intron structure of CsbHLH genes. Exons and introns were indicated by green boxes and single lines, respectively. Blue boxes represented upstream or downstream. The length of each gene was listed in Table 1.
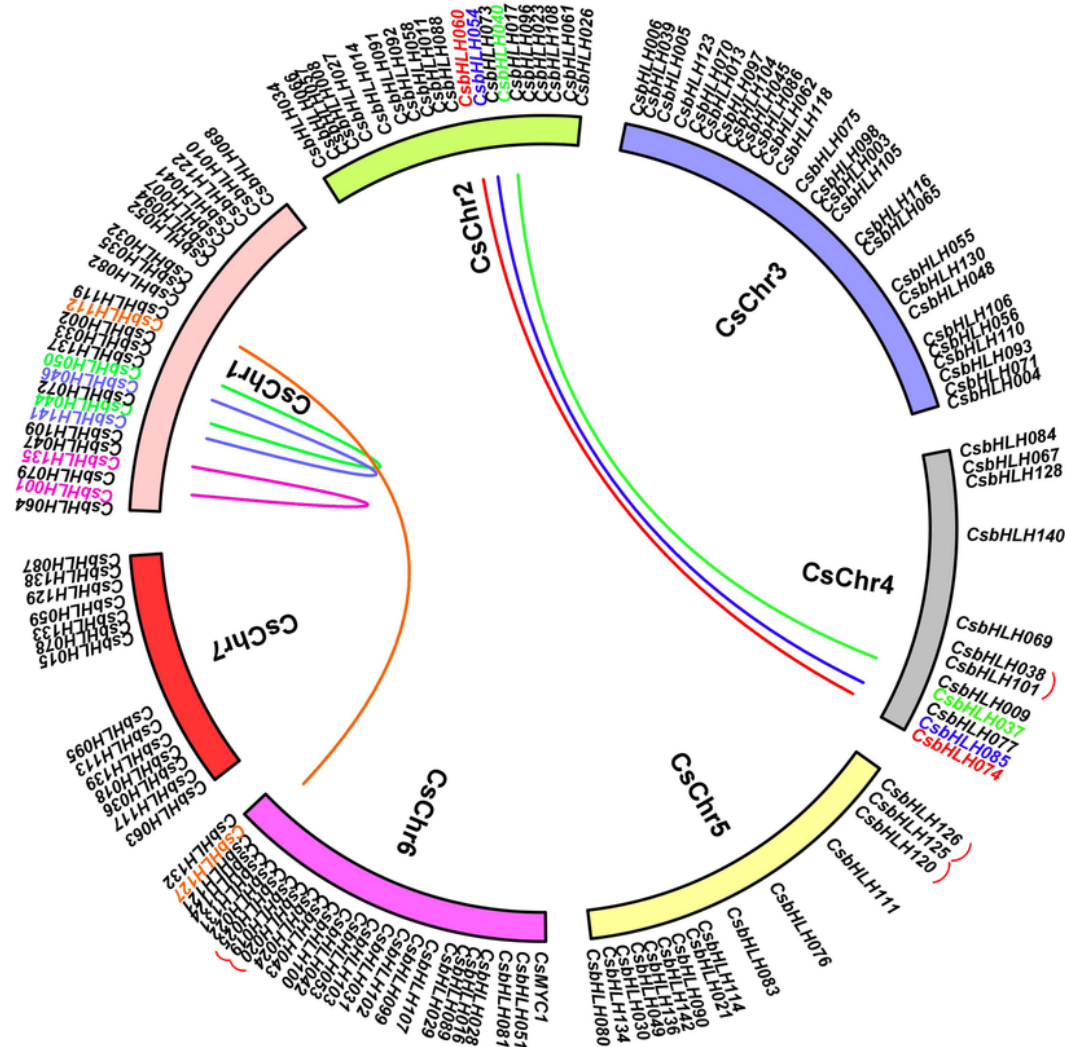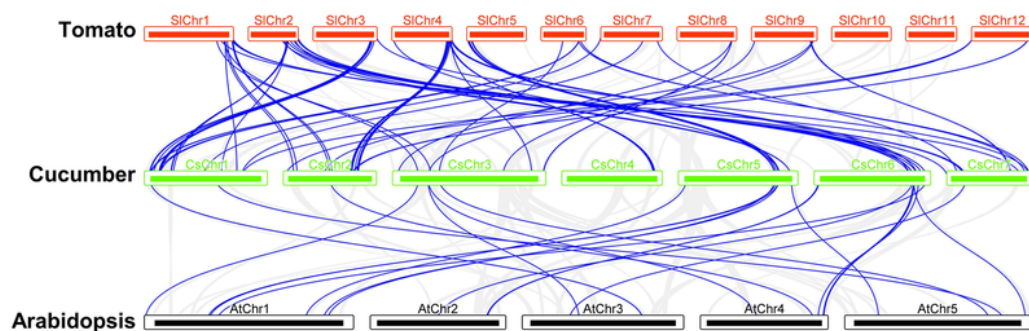
Figure 3

Gene duplication and synteny analysis of CsbHLH genes. (a) Schematic representations of the chromosomal distribution and interchromosomal relationships of CsbHLH genes. Different line colours represented different segmental duplicated CsbHLH gene pairs, among which the two genes of the same segmental duplicated gene pair were labelled in the same colour. The red lines in the outer ring indicated tandem duplication gene pairs. (b) Synteny analysis of bHLH genes between cucumber and Arabidopsis and tomato. Blue lines indicated collinear blocks of the bHLH gene within the cucumber and Arabidopsis and tomato genomes.
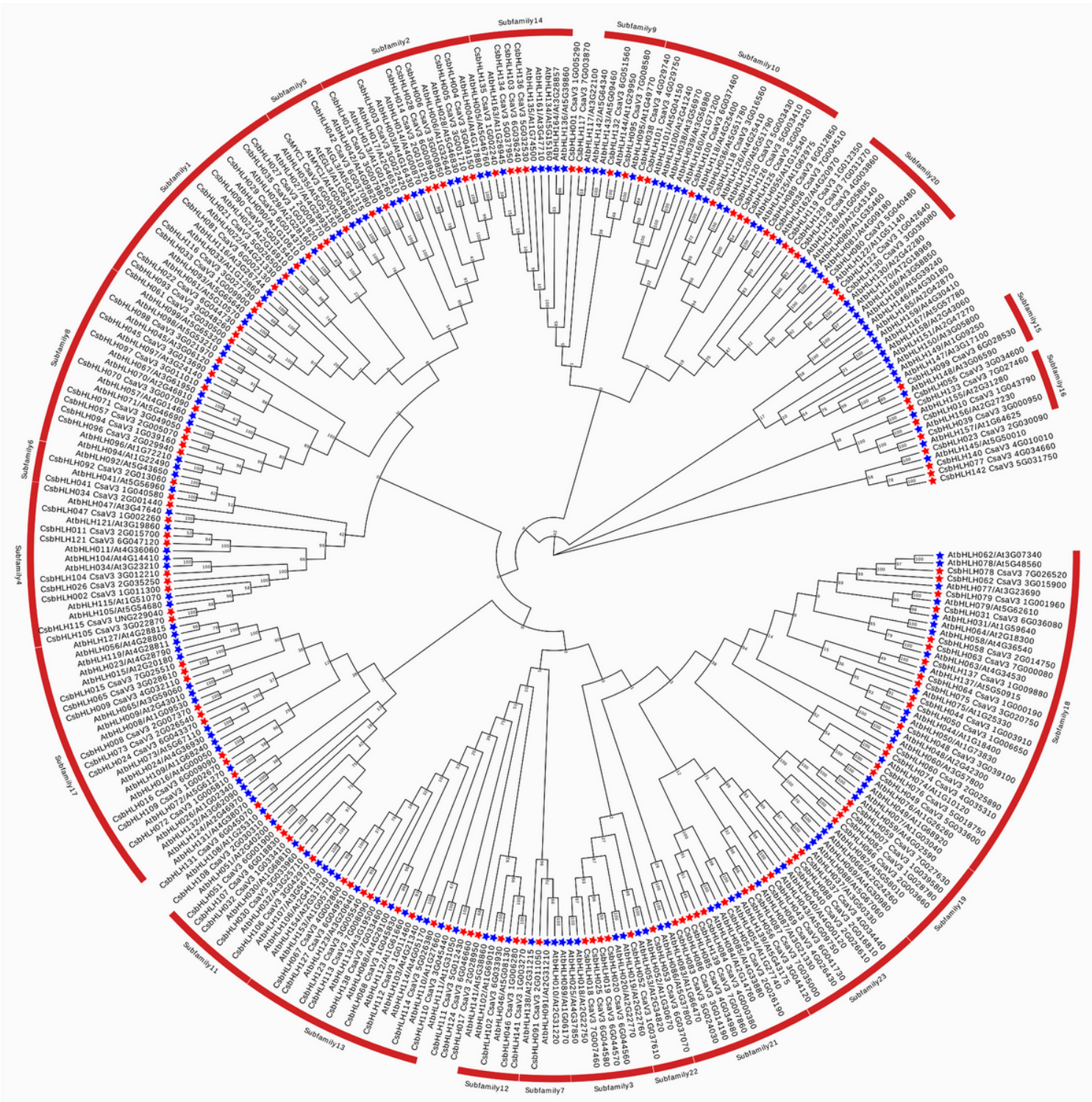
Figure 4

Evolutionary tree analysis (circle tree) and subfamily classifications of bHLHs proteins in cucumber and Arabidopsis thaliana. The evolutionary tree was constructed using the Neighbour-Joining method with 1000 bootstrap replication. The evolutionary distances were computed using poisson correction. The analysis involved 142 cucumber bHLH protein sequences and 166 Arabidopsis thaliana bHLH protein. Red stars represented the CsbHLH proteins and blue represented the AtbHLH proteins.
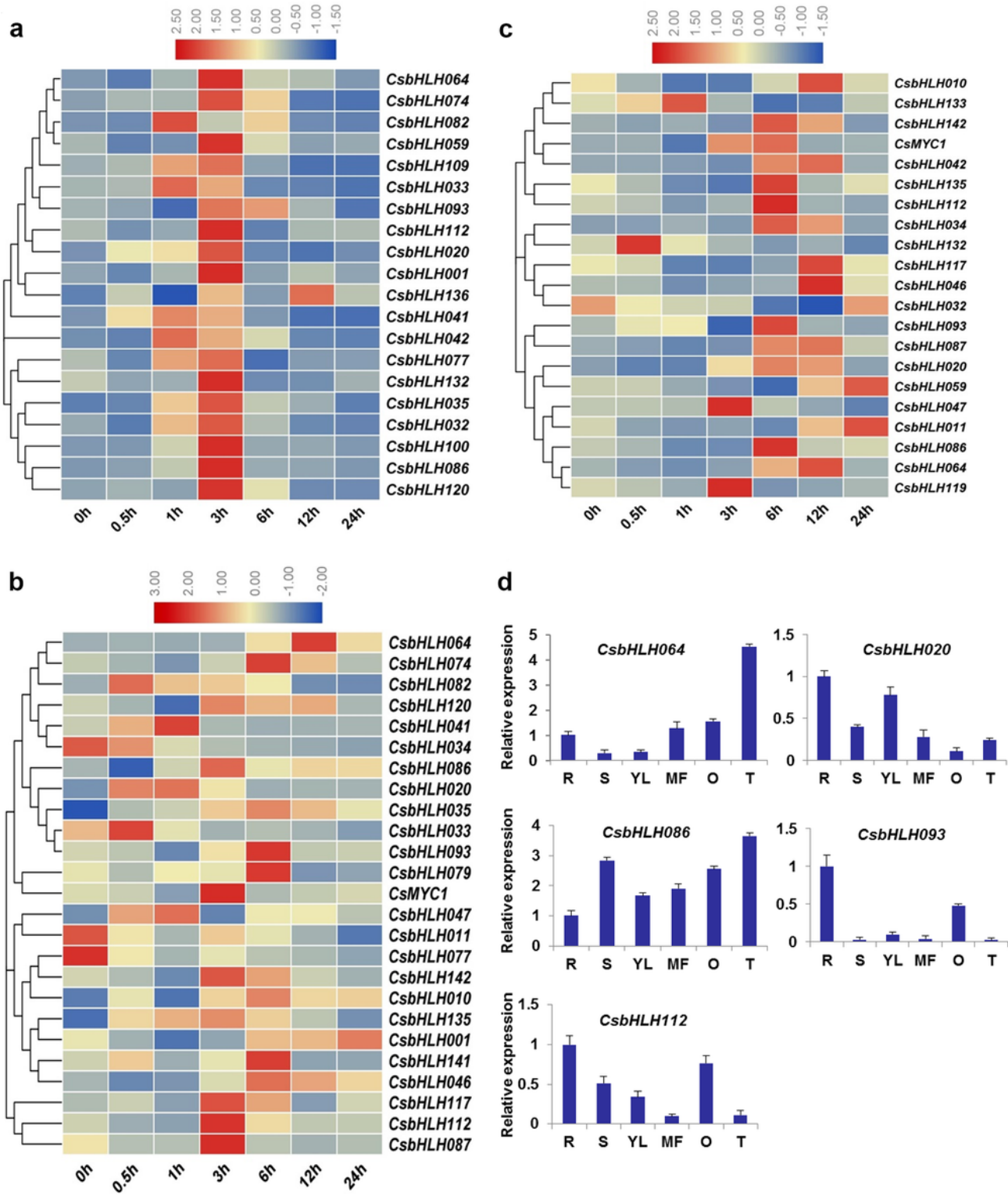
Figure 5

Relative expression analysis of the CsbHLH genes under different stress conditions and different tissues. Expression patterns of CsbHLH genes under NaCl (100 mM) treatment (a), ABA (100 μM) treatment (b) and low temperature (4°C) treatment (c). (d) Tissue-specific expression profiles of five cucumber bHLH genes. Total RNA was isolated from roots (R), stems (S), young leaves (YL), male flowers (MF), ovary (O) and tendrils (T), respectively. The cucumber β-actin gene was performed as an internal control, and three independent samples were used for these experiments. Error bars indicated standard errors (SE).
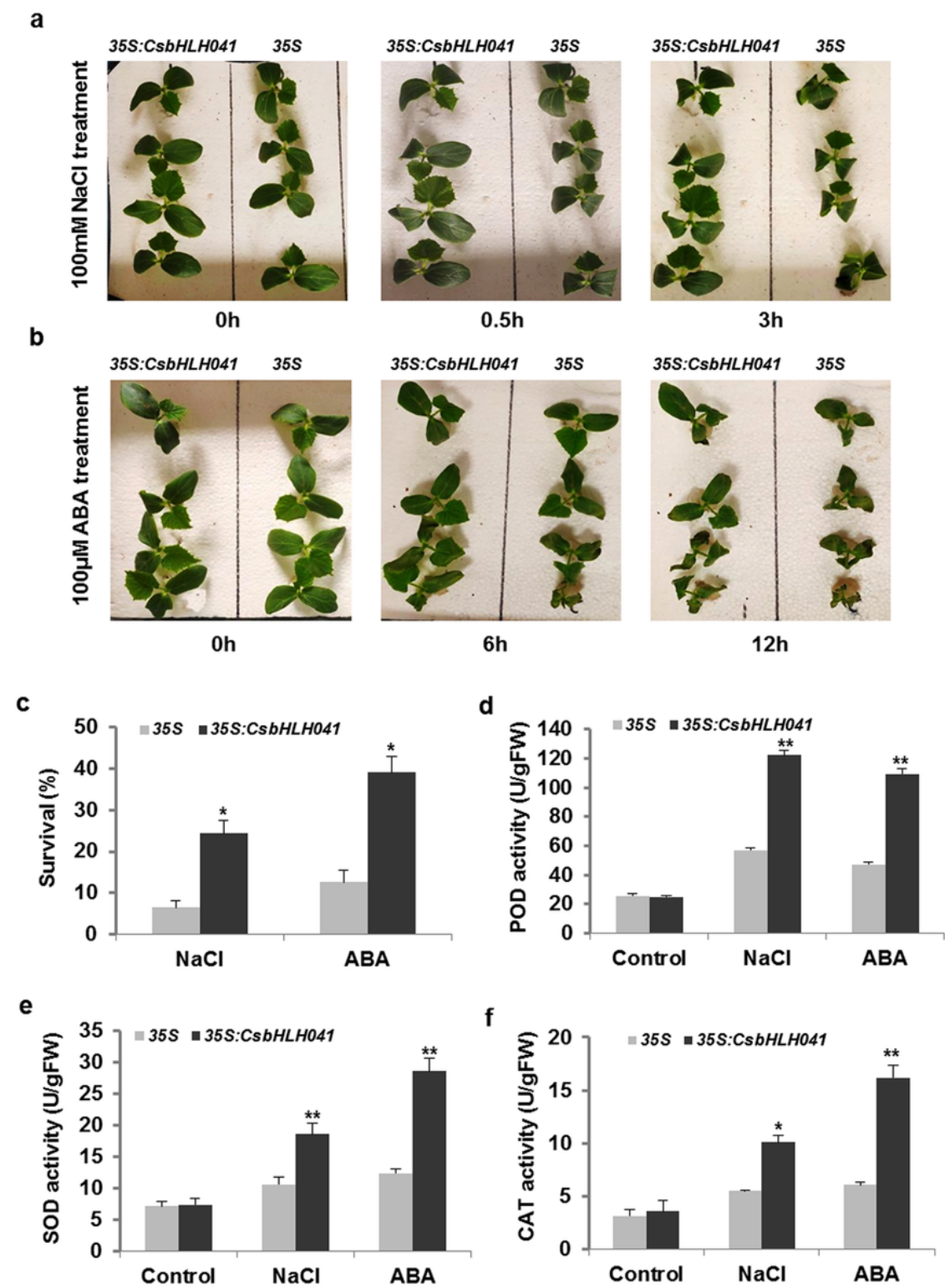
Figure 6

Overexpression of CsbHLH041 increased salt and ABA tolerance in cucumber seedlings. Phenotypes of 35S empty vector and 35S:CsbHLH041 cucumber seedlings treated with 100 mM NaCl (a) and 100 μM ABA (b) at different time periods during hydroponic growth. (c) Survival of 35S empty vector and 35S:CsbHLH041 cucumber seedlings after 12 hours of salt and ABA treatments. Comparison of the antioxidant enzyme activity between 35S empty vector and 35S:CsbHLH041 cucumber seedlings under salt and ABA treatment: (d) peroxidase (POD) activity, (e) superoxide dismutase (SOD) activity, (f) catalase (CAT) activity. The bars showed the SE. * and ** indicate significant differences at $P < 0.05$ and $P < 0.01$, respectively.
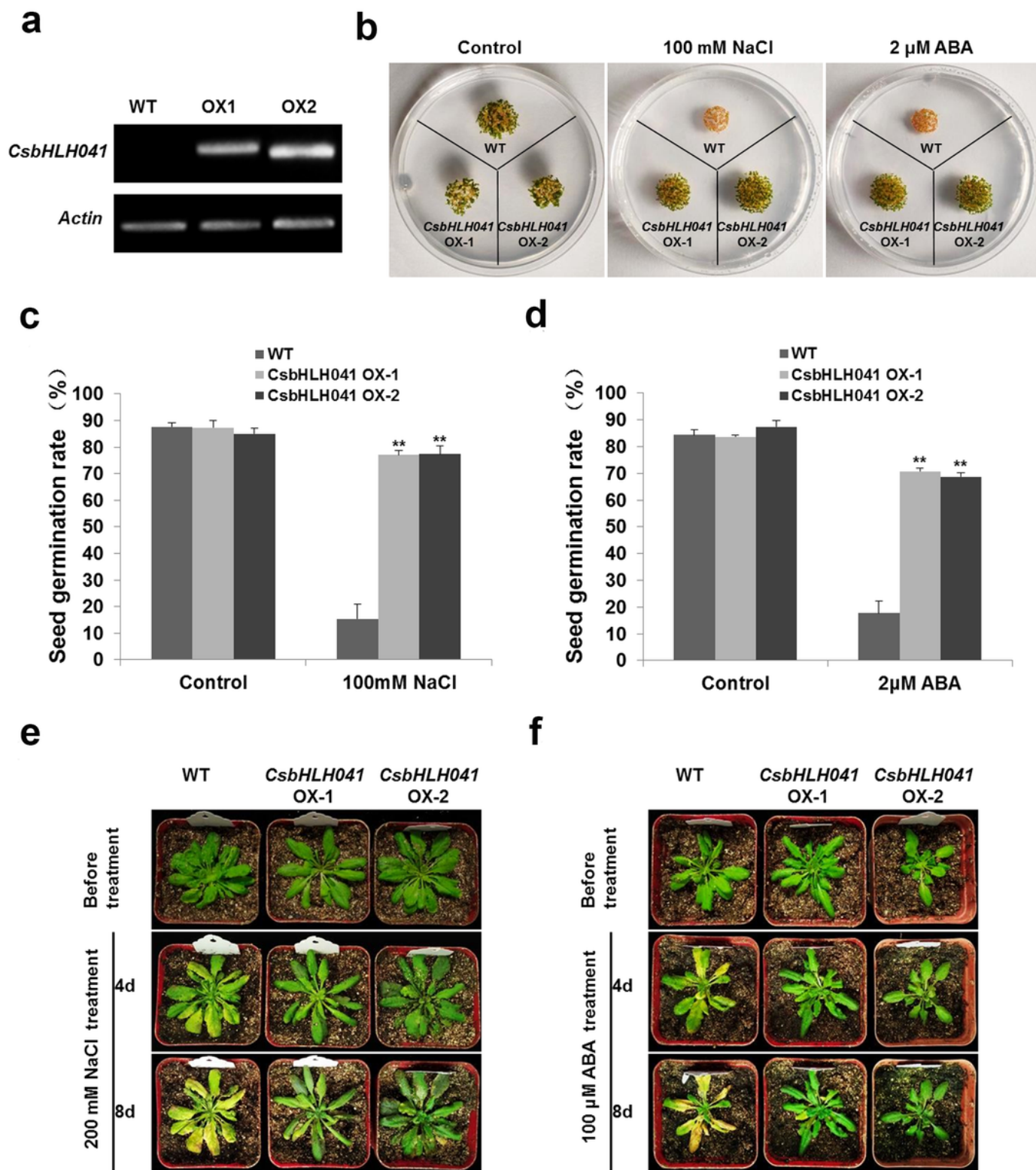
Figure 7

CsbHLH041 transgenic Arabidopsis showed enhanced salt and ABA tolerance. (a) Relative expression of CsbHLH041 in Col-0 (WT) and two T3 generation transgenic lines by semi-quantitative PCR. The actin8 gene was used as an internal control. The original, uncropped gel image was provided as Additional file 9. (b) Germination of WT seeds of Col-0 and CsbHLH041 transgenic lines OX-1, OX-2 on 1/2 MS supplemented with 100 mM NaCl and 2 µM ABA after 7 days of cultivation at 22°C. (c) and (d) Seed

germination rate for the corresponding (b), respectively. Three biological replications were performed. Asterisks indicated a significant difference **p < 0.01 compared with the corresponding controls. The growth of Col-0 (WT) and CsbHLH041 transgenic lines after 200 mM NaCl (e) and 100 μM ABA (f) treatments.

a

b
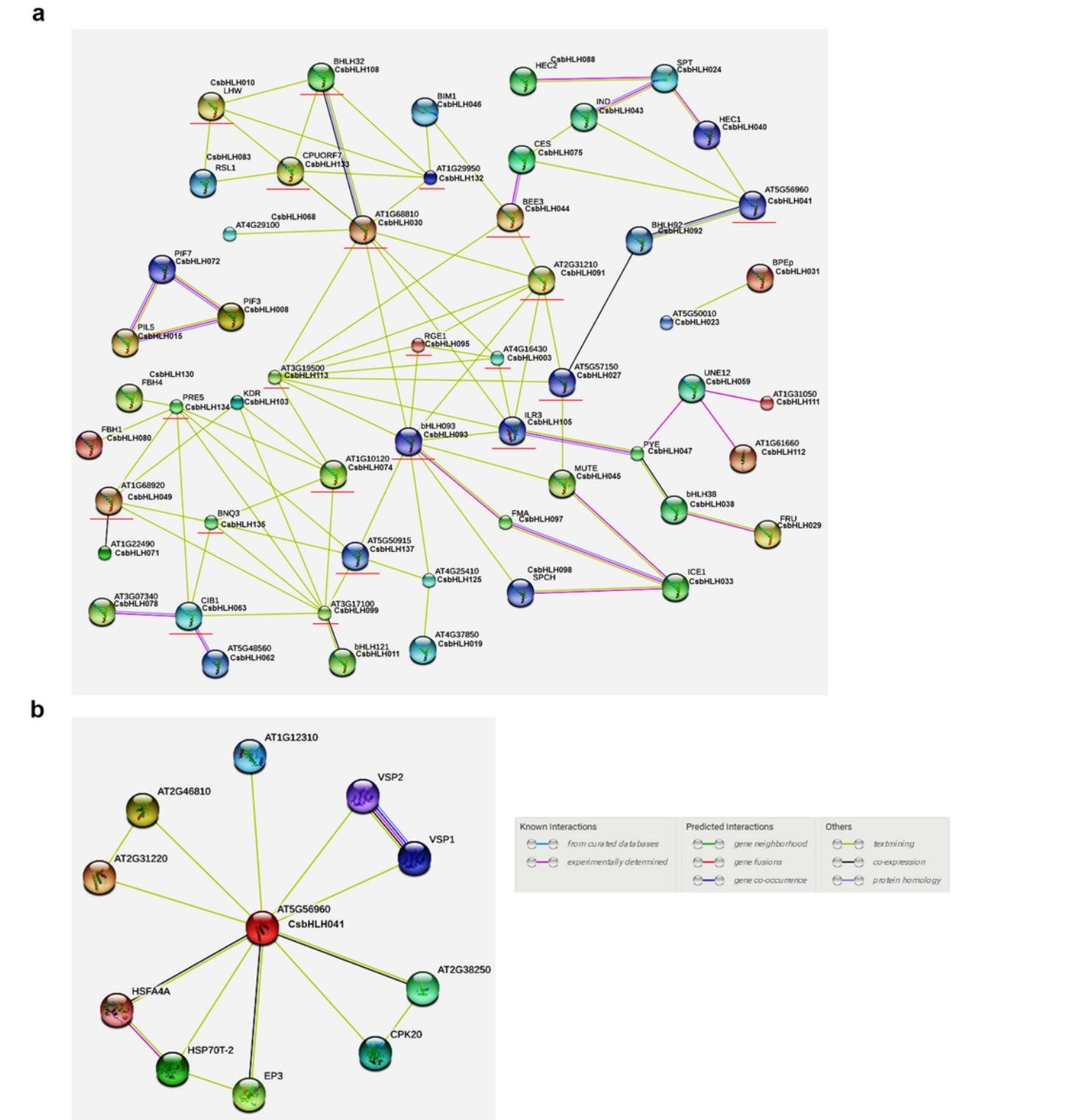
Figure 8

Protein interaction network for CsbHLHs based on CsbHLH orthologs in Arabidopsis. Protein interaction network predictions of CsbHLHs (a) and CsbHLH041 (b), based on CsbHLH orthologs in Arabidopsis. Red lines indicated proteins that were predicted to interact with more than four other bHLH proteins. CsbHLH proteins were shown next to Arabidopsis orthologs.