Genome-wide analyses of the bHLH gene family reveals structural and functional characteristics in the aquatic plant *Nelumbo nucifera*

Tian-Yu Mao [1, 2], **Yao-Yao Liu** [1, 2], **Huan-Huan Zhu** [1, 2], **Jie Zhang** [Corresp., 1, 2], **Ju-Xiang Yang** [1, 2], **Qiang Fu** [1, 2], **Nian Wang** [1, 2], **Ze Wang** [1, 2]

[1] Key Laboratory of Horticultural Plant Biology, College of Horticulture and Forestry Sciences, Huazhong Agriculture University, Wuhan, China

[2] Key Laboratory of Urban Agriculture in Central China, Ministry of Agriculture and Rural Affairs, Wuhan, China

Corresponding Author: Jie Zhang
Email address: flybebrave@mail.hzau.edu.cn

Lotus (*Nelumbo nucifera* Gaertn.) is an economically important aquatic plant with multiple applications, but water salinity and cold stress seriously affect lotus yield and distribution. The basic helix-loop-helix (bHLH) transcription factors play a vital role in plant growth and development, metabolic regulation processes and responses to environmental changes. However, systematic analyses of the bHLH transcription factor family in lotus has not yet been reported. Here, we report the identification and description of bHLH genes in lotus (*NnbHLHs*) with a focus on functional prediction, particularly for those involved in stress resistance. In all, 115 *NnbHLH*s were identified in the lotus genome and classified into 19 subfamilies. The chromosomal distribution, physicochemical properties, bHLH domain, conserved motif compositions and evolution of these 115 *NnbHLHs* were further analyzed. To better understand the functions of the lotus bHLH family, gene ontology, cis-element, and phylogenetic analyses were conducted. *NnbHLHs* were predicted to be involved in plant development, metabolic regulation and responses to stress, in accordance with previous findings. Overall, 15 *NnbHLHs* were further investigated with functional prediction via quantitative real-time PCR (qRT-PCR) analyses. Meanwhile, expression profiles of *NnbHLHs* in four tissues indicated that many *NnbHLHs* showed tissue preference in their expression. This study is supposed to provide a good foundation for further research into the functions and evolution of *NnbHLHs*, and identifies candidate genes for stress resistance in lotus.

Genome-wide analyses of the bHLH gene family reveals structural and functional characteristics in the aquatic plant *Nelumbo nucifera*

Tian-Yu Mao[12], Yao-Yao Liu[12], Huan-Huan Zhu[12], Jie Zhang[12*], Ju-Xiang Yang[12], Qiang Fu[12], Nian Wang[12], Ze Wang[12]

[1] Key Laboratory of Horticultural Plant Biology, College of Horticulture and Forestry Sciences, Huazhong Agriculture University, Wuhan 430070, China;
[2] Key Laboratory of Urban Agriculture in Central China, Ministry of Agriculture and Rural Affairs, Wuhan 430070, China

Corresponding Author:
Jie Zhang[12]
No.1, Shizishan Street, Wuhan, Hubei, 430070, China
Email address: flybebrave@mail.hzau.edu.cn

Abstract

Lotus (*Nelumbo nucifera* Gaertn.) is an economically important aquatic plant with multiple applications, but water salinity and cold stress seriously affect lotus yield and distribution. The basic helix-loop-helix (bHLH) transcription factors play a vital role in plant growth and development, metabolic regulation processes and responses to environmental changes. However, systematic analyses of the bHLH transcription factor family in lotus has not yet been reported. Here, we report the identification and description of bHLH genes in lotus (*NnbHLHs*) with a focus on functional prediction, particularly for those involved in stress resistance. In all, 115 *NnbHLH*s were identified in the lotus genome and classified into 19 subfamilies. The chromosomal distribution, physicochemical properties, bHLH domain, conserved motif compositions and evolution of these 115 *NnbHLHs* were further analyzed. To better understand the functions of the lotus bHLH family, gene ontology, cis-element, and phylogenetic analyses were conducted. *NnbHLHs* were predicted to be involved in plant development, metabolic regulation and responses to stress, in accordance with previous findings. Overall, 15 *NnbHLHs* were further investigated with functional prediction via quantitative real-time PCR (qRT-PCR) analyses. Meanwhile, expression profiles of *NnbHLHs* in four tissues indicated that many *NnbHLHs* showed tissue preference in their expression. This study is supposed to provide a good foundation for further research into the functions and evolution of *NnbHLHs*, and identifies candidate genes for stress resistance in lotus.

Introduction

The basic helix-loop-helix (bHLH) family is the second largest gene family in plants, after the MYB family (*Feller et al., 2010*). With the greater availability of genome sequence data, genes in the bHLH family have been identified and characterized in various plant species (*Li et al., 2006; Niu et al., 2017; Sun et al., 2015; Toledo-Ortiz et al., 2003*). The bHLH gene family is named after its distinctive structure of a bHLH domain, consisting of two conserved regions, a basic region and a helix–loop–helix region (HLH region), of approximately 60 amino acids. The basic region is located at the N-terminus of the bHLH domain and consists of approximately 17 amino acids, 6 of which are basic amino acid residues. This domain is a DNA-binding region that enables bHLH transcription factors (TFs) to bind to E-box (CANNTG) (*Carretero-Paulet et al., 2010; Niu et al., 2017*). The HLH region, containing two amphipathic α-helices linked by a loop region with a variable sequences (*Murre et al., 1989*), is located at the C-terminus and participates in protein dimerization (*Brownlie et al., 1997*). The regions outside the bHLH domain are distinctly divergent. In animals, the bHLH genes are classified into six groups (from A to F), which contain 45 subgroups based on their various functions in the regulation of gene expression, target DNA elements, and phylogenetic analyses (*Atchley et al., 1999*). However, research on bHLH proteins in plants has been limited compared to that in animals, and therefore the exact organization of bHLH genes is unclear. Generally, the bHLH gene family in plants has been divided into 15–26 groups (*Toledo-Ortiz et al., 2003*), and sometimes up to 32 when atypical bHLH proteins are included (*Carretero-Paulet et al., 2010*). Based on sequence homology and phylogenetic relationships, 147 bHLH genes were identified in Arabidopsis and grouped into 21 subfamilies (*Toledo-Ortiz et al., 2003*), while 167 bHLH genes were detected in rice and formed 25 subfamilies (*Li et al., 2006*).

In plants, the bHLH gene family plays important roles in plant growth and development, metabolic regulation, and response to environmental changes. In Arabidopsis, the BR-Enhanced Expression and Phytochrome Interacting Factors (*PIFs*) had been reported to response to cold (*Kim et al., 2010*) and light (*Paik et al., 2017*). Paclobutrazol Resistance and Cryptochrome 2 Interacting BHLH are involved in flowering initiation and root initiation (*Ikeda et al., 2012; Liu et al., 2013*), with the MYCs acting as positive regulators of jasmonate biosynthesis (*Fernández-Calvo et al., 2011; Song et al., 2014a)*. Furthermore, Glabra3, Enhancer of Glabra3 and Transparent Testa8 are involved in anthocyanin biosynthesis in Arabidopsis (Petridis et al., 2016; Wen et al., 2018). Recently, numerous bHLH TFs in plants have been suggested to respond to diverse abiotic stresses and improve plant stress tolerance, including to cold, salt and drought (Sun et al., 2018). Under salt stress, *bHLH39* increased the expression levels of stress-response genes in wheat , and thus improved the salt tolerance of *bHLH39*-overexpressing wheat plants (*Zhai et al., 2016*). A total of 18 bHLH genes in poplar have been found to respond to salt stress (*Zhao et al., 2018a*). In addition, *CgbHLH001* in *Chenopodium glaucum* can interact with *CgCDPK* in a signal transduction pathway under salt stress (*Wang et al., 2017b*). A study on grape validated the activity of *VabHLH1* and *VvbHLH1* as positive regulators under cold stress (*Xu et al., 2014*).

NtbHLH123 in *Nicotiana tabacum* is a transcriptional activator that can bind to the G-box/E-box motif in the *NtCBF* gene promoter, thus regulating the expression of stress-responsive genes and increasing the cold tolerance of *N. tabacum* (*Zhao et al., 2018c*). *FtbHLH2* in *Fagopyrum tataricum* is significantly induced under cold treatment and overexpression of *FtbHLH2* results in better cold/oxidative tolerance in transgenic Arabidopsis (*Yao et al., 2018*). Furthermore, *StbHLH45* in *Solanum tuberosum* (*Wang et al., 2018*) acts as positive regulatory factor under cold stress. The functions of bHLH gene family in plant under biotic and abiotic stresses are gaining increasing attention, and will be an important direction of future research (Sun et al., 2018).

Lotus is an economically important aquatic plant, that has been widely used for food, medicinal, and ornamental purposes. Being a basal eudicot plant with numerous monocot characteristics, it has been an important subject of evolutionary and taxonomic studies (*Group 2009; Ming et al., 2013; Zhao et al., 2018b*). Meanwhile, lotus plays a vital role in cultural and religious activities and is extensively distributed throughout Asia and Northern Australia. With soil and water salinity being one of the most serious environmental stresses affecting crop yields worldwide, it is essential to improve the tolerance of plants to salinization of soil and water (*Munns & Tester 2008*). In addition, low temperatures can severely impact the growth, yield and distribution area of plants (*Sanghera et al., 2011*). The dry matter content of lotus is minimal during colder seasons at high latitudes (*Bullard & Crawford 2010; Halling et al., 2010*), causing substantial economic losses. Hence, identification of salt and cold tolerance and response genes is expected to improve salt and cold resistance of lotus, which may increase the yield and expand the distribution of lotus to some extent. As noted above, the bHLH gene family responds to environmental stimuli, and thus structural and functional analyses of the *NnbHLH* gene family in lotus are necessary to improve stress tolerance in lotus. However, the bHLH gene family has not yet been comprehensively studied in lotus. Although the identification and structural properties of some bHLH family members in lotus were reported in 2014, systematic structural and functional analyses of the bHLH gene family in lotus have remained scarce (*Hudson & Hudson 2014*). In this study, the bHLH gene family of lotus was systematically identified using bioinformatics methods. Then, thorough analyses of the gene sequences, gene structures and conserved motifs, phylogenetic relationships, gene ontology (GO) annotations, function prediction, and expression patterns were conducted. The results offer an effective framework for further functional characterization of the lotus bHLH gene family, and in particular their role under stress.

Materials & Methods

1.1. Plant materials and treatments of the stress experiments

Seedlings (25 days old) of *N. nucifera* were grown in artificial climate chambers (Model:RXZ-500C, NingboJiangnan) with 16 h light and 8 h dark at 24°C, following the conditions for lotus cultivation described in *Diao et al.* (*2014*). For cold treatment, seedlings were placed in an Intellus Ultra Controller (Model:LT-36VLC8, Percival Scientific, Inc.) with the temperature at 4°C, a common condition for cold treatment, and the same photoperiod as described above. Leaves were collected before cold exposure (control), after 2 h, 4 h, 8 h, and 12 h of cold exposure for RNA

extraction before expression analyses. For salt stress treatment, roots of the lotus seedlings were exposed to 50 mM NaCl (*Diao et al., 2014*), and the leaves were collected before treatment (control) and after 2 h, 4 h, 8 h, and 12 h of salt exposure. All of the samples were collected with three biological replications and were immediately frozen in liquid nitrogen. The samples were stored at -80°C for RNA isolation.

1.2. Identification of the *NnbHLH* gene family

The lotus genome sequence and the gene annotation data were obtained from the lotus genome database (Lotus-DB: http://lotus-db.wbgcas.cn)(*Wang et al., 2015*). First, a Hidden Markov Model search (HMMsearch) was performed using HMMER software with the seed profile of the bHLH domain (PF00010) (*Cheng et al., 2018; Sun et al., 2015*), which was downloaded from the PFAM database (http://pfam.xfam.org/) (*Finn et al., 2014*). HMMER software was used to search for bHLH protein in the entire protein dataset with the E-value cut-off set to 10^{-5}. Then, a Basic Local Alignment Search Tool protein (BLASTp) alignment against all lotus protein sequences was subjected to perform an extensive search for candidate bHLH genes using bHLH protein sequences from Arabidopsis and rice as queries. The protein sequences of Arabidopsis and rice were downloaded from PlantTFDB (http://planttfdb.cbi.pku.edu.cn) (*Jin et al., 2017*). The Simple Modular Architecture Research Tool (*Letunic et al., 2015*) and Conserved Domains Search (*Marchler-Bauer et al., 2017*) were employed to detect the bHLH domain in candidate protein sequences.

1.3. Chromosomal distribution, gene duplication analyses and calculation of synonymous (Ks) and non-synonymous (Ka) substitution rates

The distribution of each *NnbHLH* on megascaffolds was obtained based on the GFF3 file (*Wang et al., 2015*) and analyzed using Map Gene 2 Chromosome V2 (http://mg2c.iask.in/mg2c_v2.0/). Gene duplication analyses for lotus was conducted using the Multiple Collinearity Scan Toolkit (MCScanX) (*Wang et al., 2012*). To identify candidate homologous gene pairs ($E<1e^{-5}$), BLASTp alignment was carried out across the whole lotus genome. The potential homologous gene pairs were identified, and then loaded into the program MCScanX with the default parameters to identify syntenic chains. MCScanX was used to further distinguish among whole-genome duplication (WGD)/segmental, dispersed proximal, and tandem duplication events in the *NnbHLH* gene family. Candidate homologous gene pairs identified in the same synteny block were applied to calculation of Ka and Ks values using DNAsp5 (*Librado et al., 2009*).

1.4. Sequence alignment and phylogenetic, gene structure and conserved motif analyses

Multiple domain alignments were performed using MEGA (vision 6.0) (*Tamure et al., 2013*) and loaded into Geneious to visualize. The phylogenetic tree was constructed using MEGA 6.0 with the neighbor joining (NJ) method and the following parameters: pairwise deletion and 1000 bootstrap replications. The phylogenetic tree was visualized by plotting it using the EvolView tool (http://www.evolgenius.info).

The intron-exon organization of *NnbHLHs* was visualized using the Gene Structure Display Server 2.0 (GSDS: http://gsds.cbi.pku.edu.cn/) (*Hu et al., 2014*). Conserved motifs in *NnbHLHs*

were identified using the MEME (http://meme-suite.org/index.html) (*Bailey et al., 2009*) server with the maximum number of motifs set to 20.

1.5. GO annotation and analyses of cis-regulatory elements

GO analyses of *NnbHLHs* was performed using the Blast2GO program (*Conesa et al., 2005*), and the NCBI database was selected as the reference database. The sequences 1,500 bp upstream of the translation initiation codon ATG for each *NnbHLH* were selected for analyses of the promoters using a C script. Cis-regulatory elements for each promoter sequence were predicted through searching the PlantCARE database (*Lescot et al., 2002*) with the false discovery rate (FDR) < 0.1%.

1.6. Expression profiles and qRT-PCR analyses of *NnbHLHs*

RNA-seq data of four lotus tissues (leaf, petiole, rhizome, root) were downloaded from Lotus-DB. Total RNA was isolated using an RNA extraction kit (Aidlab, Beijing, China) according to the manufacturer's instructions. 5X All-In-One RT MasterMix (abm) was used for reverse transcription. Specific primers for qRT-PCR were designed using Primer Premier 5 software (*Lalitha et al., 2000*). qRT-PCR was conducted using the LightCycler®/LightCycler®96 System Real Time PCR (Roche, Basel, Switzerland) with SYBR Premix Ex Taq II (TaKaRa, Japan). The *actin* gene (Genebank: XM_010252745) was used as an internal control for normalization of the expression levels of candidate *NnbHLHs* among different samples, following previous studies (*Cheng et al., 2013; Diao et al., 2014; Jin et al., 2016*), and expression levels were calculated using the delta-delta CT method (*Livak & Schmittgen 2001*).

Results

2.1. Identification, chromosomal distribution and physicochemical properties of *NnbHLHs*

Members of the bHLH family in lotus were identified in the lotus genome using two strategies, i.e., Hidden Markov Model search and BLASTp search, as described in the Materials & Methods section. Then, to verify the sequences, all candidates were checked for the presence of a complete bHLH domain via CDD and SMART. In total, 115 sequences were confirmed as lotus bHLHs and named *NnbHLH1* to *NnbHLH115*. The encoded putative NnbHLH proteins were predicted to be 89 (*NnbHLH110*) to 779 (*NnbHLH61*) amino acids in length, with molecular weights (Mw) ranging from 10.07 kDa (*NnbHLH110*) to 82.64 kDa (*NnbHLH61*) (Table S1). The grand average of hydropathicity values (GRAVY) of all candidate NnbHLH proteins were negative, ranging from -0.887 to -0.059, representing a hydrophilic characteristic. The predicted isoelectric points (PIs) of NnbHLH proteins were 4.72 (*NnbHLH72*) to 10.62 (*NnbHLH110*). The predicted numbers of negatively charged residues (Asp and Glu) were 7 (*NnbHLH110*) to 104 (*NnbHLH1*), while the predicted number of positively charged residues (Arg and Lys) ranged from 18 (*NnbHLH110*) to 83 (*NnbHLH13*) (Table S1). Detailed information about these characteristics is listed in Table S1. The ratio of *NnbHLHs* in the *Nelumbo nucifera* genome was about 0.425%, similar to that in rice (0.44%) (Li et al., 2006) and poplar (0.40%) (*Carretero-Paulet et al., 2010*) but lower than the ratios in Arabidopsis (0.59%) (*Toledo-Ortiz et al., 2003*) and *Brachypodium distachyon* (0.55%) (*Niu et al., 2017*). Based on the results of lotus genome annotation (*Wang et al., 2015*), the 115

predicted *NnbHLHs* were distributed unevenly, localized on 20 megascaffolds of lotus, with 168 megascaffolds in the lotus genome (Table S1). Among them, megascaffolds 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, and 13 contained most *NnbHLHs*, while megascaffolds 9, 11, 15, 17, 24, 33, 43, 59, and 114 and scaffolds 554 and 634 only possessed 1–2 *NnbHLHs* each (Table S1).

2.2. Multiple sequence alignment, prediction of DNA-binding and protein dimerization activity of *NnbHLHs*

To further clarify the structural characteristics of *NnbHLHs*, multiple sequence alignment analyses of the bHLH domain were carried out. As shown in Fig. 1, one basic region, one loop region and two helix regions were detected in all 115 NnbHLH proteins. And conserved amino acids in bHLH domains, with sequence identity more than 50%, were shaded in grey or black color (Fig. 1A). A sequence logo of the conserved amino acids in the NnbHLH domain was presented in Fig. 1B. The bHLH gene family in Arabidopsis, rice and *Brachypodium distachyon* all contain 17 conserved amino acids in the bHLH domain, which were also detected in the NnbHLH proteins (Fig. 1C, Table 1). The consensus ratios of the 17 conserved amino acids among the three species were also calculated (Table 1). Arg-16, Arg-17, Leu-27, and Leu-61 showed extremely high sequence identity (98%, 93%, 100% and 94%, respectively) among the 115 NnbHLH proteins (Table1). The HLH region may be essential for dimerization, particularly Leu-27 in helix 1 and Leu-61 in helix 2 (*Carretero-Paulet et al., 2010*). Thus, we speculated that NnbHLH proteins may also have the capacity to form protein complexes.

Based on the criteria developed by *Toledo-Ortiz et al.(2003)*, bHLH proteins with more than five basic amino acid residues in the basic region were identified as DNA-binding proteins. NnbHLH proteins were divided into two major groups accordingly, including 101 DNA-binding proteins and 14 non-DNA-binding proteins. Then we subdivided the 101 DNA-binding NnbHLH proteins into two groups: 89 E-box-binding proteins (based on the presence of Glu-6 and Arg-9) and 12 non-E-box-binding proteins (without the simultaneous presence of Glu-6 and Arg-9), according to previous research (*Ellenberger et al., 1994*). The E-box-binding NnbHLH proteins were further subdivided into two groups, including 75 G-box-binding proteins and 14 non-G-box-binding proteins, based on the presence or absence of His/Lys-9, Glu-13, and Arg-17, which may be critical for binding to the G-box (CACGTG) (*Niu et al., 2017; Sun et al., 2015*).

2.3. Gene structure and conserved motif analyses of *NnbHLHs*

A neighbor-joining phylogenetic tree was constructed using the alignment results of the NnbHLH proteins (Fig. 2A). Based on the bootstrap values, the NnbHLH proteins were divided into 19 subfamilies (Fig. 2A). Among those subfamilies, subfamily 16 was the largest, containing 16 proteins, while subfamilies 8 and 18 had only had 1 protein each.

Next, the exon/intron structures of *NnbHLHs* were analyzed using the GSDS 2.0 online tool. *NnbHLHs* in the same subfamily appeared to share similar numbers of exons and introns (Fig. 2B). Members of subfamilies 6, 7, 10 and 12 all contained the same number of exons and introns, while the majority of subfamilies 2, 9, 11 and 13 shared the same number of exons and introns. Meanwhile, the numbers of exons and introns were variable among subfamilies 1, 3, 15 and 19 (Fig. 2B).

To further reveal the specific regions of NnbHLH proteins, 20 conserved motifs that varied from 16 to 100 residues in length were detected using the MEME tool. All predicted motifs were identified only once in each NnbHLH protein sequence. NnbHLH proteins contained different numbers of conserved motifs, ranging from 2 to 8, and all NnbHLH proteins possessed motifs 1 and 2, representing the location of the bHLH domain (Fig. 2C). Meanwhile, NnbHLH proteins in the same subfamily always shared similar motif composition. For example, members of subfamily 9 all contained motifs 1, 2, 4, 10 and 11, while motifs 5, 17, and 19 were found exclusively in subfamilies 1, 12, and 14, respectively (Fig. 2C).

2.4. Gene duplication analyses of *NnbHLHs*

Duplication events usually contribute to the expansion of gene families and genome evolution, and in particular, WGD/segmental and tandem duplication events have been important in the expansion of multigene families (*Kong et al., 2010*). In this study, almost all 115 detected *NnbHLHs* had experienced duplication events, except *NnbHLH21*. Because duplication usually contributes to the expansion of gene families, we further investigated the duplication patterns of each *NnbHLH*. In total, 80% (92/115) of the *NnbHLHs* were retained from WGD/ segmental duplication events (Table S2), with only 3.47% (4) from tandem, 13.9% (16) from dispersed and 1.73% (2) from proximal duplication events.

The substitution rate ratio, Ka/Ks, is an effective criterion for the selective pressure during gene duplication. Ka/Ks values less than one indicate negative selection, those equal to one indicate neutral selection, and values greater than one indicate positive selection (*Yang 2007*). Among the 118 pairs of WGD, the Ka/Ks ratios of 27 pairs were greater than 1, with those of *NnbHLH1* and *NnbHLH59*, *NnbHLH6* and *NnbHLH97*, *NnbHLH7* and *NnbHLH76*, *NnbHLH26* and *NnbHLH93* being 1.808, 1.640, 1.510, and 1.679, respectively, indicating that these duplication pairs have been strongly positively selected during their evolutionary history (*Cheng et al., 2018*). The remaining *NnbHLH* genes, with Ka/Ks<1, may have undergone negative selection, suggesting that most bHLH genes have evolved slowly (Table S2). For the two tandem duplication pairs, the Ka/Ks of the gene pair *NnbHLH112* and *NnbHLH115* was 1.334, while that of *NnbHLH117* and *NnbHLH120* was 0.453.

As Ks values can be further be applied to trace the divergence time of duplication events (*Lynch & Conery 2000*), the divergence times of 118 WGD gene pairs were calculated. As shown in Table S2, the divergence times of these 118 WGD pairs ranged from 8.39 Mya to 167.35 Mya. The two peaks observed in Fig. 3 show that the lotus bHLH gene family has undergone two larger duplication events, an ancient one that occurred ~120 Mya, and a recent one that occurred ~50 Mya. The earliest duplication event in the *NnbHLH* family appeared at 176 Mya. Since 147 Mya, bHLH family genes began to duplicate frequently, and the expansion of the gene family accelerated.

2.5. GO annotation and cis-element analyses of NnbHLH proteins

To understand the specific functions of bHLH proteins, GO annotation of NnbHLH proteins was performed. As shown in Table. S7, most NnbHLH proteins were annotated as being associated with protein dimerization activity, the development process, and response to stimulus. Within the

cellular component, most genes were assigned to the nucleus (95/115). Very few genes had predicted distributions in organelles, such as the chloroplast envelope (8), mitochondrion (8), cytosol (4), and endosome (1). In addition, 19 NnbHLH proteins were predicted to be located on the plasmodesma (Table. S7). In terms of molecular functions, almost all NnbHLH proteins were predicted to be involved in protein dimerization activity (113/115), with 83 proteins related to DNA binding transcription factors and 79 proteins associated with DNA-binding activity. These results are consistent with the DNA-binding analyses described in subsection 2.2. Within the biological process category, 102 NnbHLHs were predicted to be involved in development processes, including the development of the roots (52/115), carpels (43/115), floral organs (34/115), and petals (25/115). Moreover, 104 NnbHLH proteins were predicted to respond to stimuli, with 65 supposedly reacting to abiotic stresses. A total of 28 NnbHLH proteins could respond to low temperature, 35 were associated with responses to radiation (blue light, red light, or far-red light) and 13 were predicted to respond to salt stress. The number of NnbHLH proteins predicted to respond to abiotic stresses was high, suggesting that the *NnbHLH* gene family may play a vital role in lotus tolerance to abiotic stresses (Table. S7).

Cis-element analyses were also conducted applying the 1,500 bp upstream sequence of *NnbHLHs* in the PlantCARE database. In the *NnbHLH* promoter region, G-box and Sp1 (response to light), MBS (response to drought), Skn-1_motif (required for endosperm expression), and ARE (essential for anaerobic induction) were the most common elements (Table S4). All *NnbHLH* family members except *NnbHLH 58*, *86*, *90*, *100,* and *101* possessed at least one cis-element involved in stress responses (Table S4). The cis-regulatory elements present within *NnbHLH* promoters could be divided into three main categories. The first category was a ubiquitous class of plant light-responsive elements (such as G-box and Sp1), and the second category included plant growth- and development-responsive elements (such as Skn-1_motif and ARE) (Table S4). The last category included elements that respond to diverse stresses (such as TC-rich), including elements responding to biotic stresses (CGTCA-motif and ABRE) and abiotic stresses (HSE, MBS and LTR). Elements in this category were widely distributed throughout the *NnbHLH* gene family (Table S4).

2.6. Function prediction of *NnbHLHs* based on phylogenetic analyses

To date, the biological functions of most *NnbHLHs* have remained unclear. Meanwhile, in Arabidopsis and rice, the functions of many bHLH proteins have been characterized and verified. In this study, phylogenetic analyses allowed us to identify putative orthologous and paralogous bHLH genes in lotus, Arabidopsis and rice. In general, homologous genes share similar structures and are clustered in the same clades, and these genes possess similar functions. To predict the gene functions of each *NnbHLH*, another NJ phylogenetic tree was constructed based on the protein of 115 *NnbHLHs*,132 *OsbHLHs* of rice and 160 *AtbHLHs* of Arabidopsis(*Li et al., 2006; Toledo-Ortiz et al., 2003*) (Fig. S1). In total, 24 subgroups were clustered, and the functions of *NnbHLHs* were predicted based on their homologs with verified functions in the same cluster (Table S5).

In summary, the majority of the members of subgroups 1a, 3, 4, 6, 7, 15a and 19 may be able to enhance stress tolerance (*Kim et al., 2005*) and response to diverse abiotic and biotic stresses

(*Sasaki-Sekimoto et al., 2014; Song et al., 2013*), including cold (*Kim et al., 2010; Deng et al., 2017*), salt (*Ahmad et al., 2015; Sakai et al., 2013; Jiang et al., 2009*), and drought (*Le Hir et al., 2017*). In subgroups 8 and 18, proteins may be related to Fe regulation, modulating the homeostasis of Fe content (*Kurt & Filiz 2018; Wang et al., 2017a; Kurt et al., 2018*). Members of subgroups 1a, 2 and 15a were predicted to regulate flower development of flower (*Sharma et al., 2016; Choi et al., 2018; Ito et al., 2012*), while those in subgroups 1b, 3, 4 and 17 may be involved in the development of various plant organs (*Karas et al., 2009; Le Hir et al., 2017; An et al., 2014; Lee et al., 2017; Ohashi-Ito & Bergmann 2006; Cui et al., 2016; Ohashi-Ito et al., 2006; Feng et al., 2017*). Subgroup 7 members may regulate the flavonoid (*Rai et al., 2016*) and anthocyanin biosynthesis (*Wang et al., 2017a*). In subgroup 10, many proteins were predicted to be PIFs, which are related with photo-induced signal transduction and may optimize plant growth and development (*Paik et al., 2017*). Other members of subgroup 10 and subgroup 13 have also showed regulatory abilities (*Leivar et al., 2008; Zumajo-Cardona et al., 2017; Cifuentes-Esquivel et al., 2014*). Members of subgroups 15b and 16 have been predicted to be involved in diverse processes (*Mai et al., 2015; Ferguson et al., 2017; Gaillochet et al., 2018; Zhu et al., 2016*). *NnbHLHs* in subgroup 21 should negatively control thermospermine biosynthesis and xylem differentiation (*Cai et al., 2016; Yamamoto & Takahashi 2017*), while members of subgroup 24 may play an essential role in establishing vascular cells (*Ohashi-Ito et al., 2013*). The detailed function prediction of *NnbHLHs* can be found in Table S5.

2.7. Expression profiles of *NnbHLHs*

Transcriptional data of four different lotus tissues were employed to analyze the expression patterns of *NnbHLHs* (Table S6). In all tissues, *NnbHLHs* with FPKM>1 were used for further analyses (*Zhuo et al., 2018*). Among 115 *NnbHLH* genes, 91 *NnbHLHs* were expressed in at least one of the tissues, while the expression levels of *NnbHLH 2, 12, 19, 21, 23, 30, 36, 40, 44, 45, 47, 48, 49, 61, 64, 65, 84, 88, 89, 90, 92, 96, 104* and *110* were low, with FPKM<1 in all four tissues (Table S6). Some *NnbHLH* genes exhibited tissue preferences, with expression levels in the preferred tissue that were more than two times higher than those in other tissues (*Sun et al., 2015*). Overall, *NnbHLH 20, 52, 84, 106, 108*, and *115* in the leaves; *NnbHLH 9, 24, 55, 97, 98,* and *109* in petioles, 14 genes (*NnbHLH 1, 4, 16, 29, 38, 50, 57, 59, 72, 74, 81, 93, 99* and *102*) in rhizomes; and 22 genes (*NnbHLH 5, 11, 14, 26, 34, 39, 41, 46, 53, 54, 56, 60, 63, 67, 68, 69, 80, 82, 83, 91, 113*, and *114*) in roots were suggested to have tissue expression preferences (Table S6). In addition, most members of subfamilies 4, 5, 6, 7, 8, 14 and 18 had no expression or low expression in the four tissues. The majority of the members of subfamilies 10, 11, and 19 exhibited high expression levels in all four tissues (Table S6).

Because stress-resistance is very important in lotus, many resistance-related *NnbHLHs* were predicted based on the structural and predictive analysis. In all, 15 *NnbHLHs* were selected for analyses of their expression under low temperatures and salt stress, and the specific primers used are listed in Table S3. The expression levels of the 12 *NnbHLHs* increased initially under 4°C treatment (Fig. 4). Most were upregulated and reached their maximum values after 4 h of treatment, decreasing thereafter. The expression levels of *NnbHLH13* and *NnbHLH66* reached a maximum

after just 2 h of treatment, while the expression levels of *NnbHLH16*, *NnbHLH38* and *NnbHLH98* were highest after 8 h. We also found that *NnbHLH 15*, *17*, *37*, *38*, *55*, *66*, *70*, *81*, *93*, and *98* were significantly (P<0.01) upregulated at all time points. Among these genes, the expression levels of *NnbHLH 15*, *17*, *55*, *66*, and *81* were significantly increased by 10-fold or more (Fig. 4); in particular, *NnbHLH17* and *NnbHLH81* were strongly induced under 4°C treatment, with expression levels that increased by nearly 25-fold and 30-fold, respectively. Under NaCl treatment, eight *NnbHLHs* (*NnbHLH1*, *NnbHLH13*, *NnbHLH17*, *NnbHLH29*, *NnbHLH37*, *NnbHLH66*, *NnbHLH70* and *NnbHLH81*) were upregulated initially and then decreased. Among these genes, only the expression level of *NnbHLH81* increased more than 10-fold compared to its untreated level. The other six *NnbHLHs* were downregulated. Although some *NnbHLHs* were significantly upregulated (P<0.01) at one or two time points, none of them were upregulated at all time points (Fig. 4).

Discussion

3.1. Comprehensive genome-wide detection of *NnbHLHs* in lotus

Numerous studies of the bHLH family in various species, such as Arabidopsis (*Toledo-Ortiz et al., 2003*), rice (*Li et al., 2006*), peanut (*Gao et al., 2017*), chinses cabbage (*Song et al., 2014b*), tomato (*Sun et al., 2015*), apple (*Yang et al., 2017*), and bamboo (*Cheng et al., 2018*), have been reported in recent years. In our study, 115 bHLH genes of lotus were identified, which is two fewer than in a previous study (*Hudson & Hudson 2014*). This difference may be the result of a more restrictive selection in *NnbHLH* in our study, in which genes without complete bHLH domain were removed. These 115 genes were further classified into 19 subfamilies based on phylogenetic analyses (Fig. 2A). Multiple sequence alignment of the full-length NnbHLH protein sequences showed that all putative *NnbHLHs* contained the classic bHLH domain, and the number and ratio of the conserved amino acid were consistent with the Arabidopsis, rice, and *Brachypodium distachyon* (Table 1). Further structural analyses indicated that most of the *NnbHLHs* were associated with DNA-binding and homodimer formation activities. In addition, the conserved motif analyses showed that motifs representing the conserved domain were apparent in almost all the *NnbHLH* family members (Fig. 2C). Together, these results indicate that the 115 *NnbHLHs* all show characteristics of the bHLH family, which confirms the accuracy of the detection of bHLH gene family in lotus. Meanwhile, the expression profiles of *NnbHLHs* in four different tissues were analyzed, revealing that expression of many *NnbHLH* showed tissue preferences.

3.2. *NnbHLHs* may play a special role in lotus evolution

The duplication pattern of a gene usually reveals how the gene was generated, how its function has evolved, and what roles it may plat in plant growth and development (*Xia et al., 2017*). In our analyses, 80% (92/115) of *NnbHLHs* were duplicated through WGD (Table S2), similar to previous studies that have reported 48% in rice and 58.2% in *Brachypodium distachyon* (*Li et al., 2006*; *Niu et al., 2017*) .Because the major duplication type in lotus was WGD, the bHLH gene family in lotus is presumed to have existed in ancient times (*Xia et al., 2017*). New genes produced through WGD usually increase the ability of plants to adapt to various growth conditions,

suggesting that *NnbHLHs* have been indispensable to the growth and development of lotus (*Flagel & Wendel 2009; Van De Peer et al., 2009*). Meanwhile, most Ka/Ks values of WGD *NnbHLH*s gene pairs were less than 1 (Table S2), showing that the lotus bHLH gene family has evolved slowly (*Wang et al., 2013*).

Based on Fig. 3, the *NnbHLH* gene family underwent a large WGD duplication event in modern times (about 50 Mya). Interestingly, two extant species of Nelumbonaceae, the *Nelumbo nucifera* and *Nelumbo lutea*, reportedly diverged in modern times (*Xue et al., 2012*). Thus, the WGD duplication event observed in our study may be related to the divergence event of *Nelumbo nucifera* and *Nelumbo lutea*. Meanwhile, another peak appears around 120 Mya in Fig. 3, which suggests that the *NnbHLH* family experienced relatively frequent duplication events. Coincidentally, the Nelumbonaceae are closely related to Platanaceae phylogenetically, and their divergence time has been estimated to be around 75.2–122.8 Mya (*Xue et al., 2012*).A duplication event appeared around 120 Mya in our study, which may be related to the divergence of the Nelumbonaceae and Platanaceae. WGD events are related to the divergence of plant taxa and often appear to be accompanied by marked and sudden increases in species richness (*Van De Peer et al., 2009*), and therefore the high frequency of WGD events in the *NnbHLH* gene family are likely to have been important to lotus evolution.

3.3. Functional prediction of *NnbHLH*s in lotus

Research on the functional and structural genomics of Arabidopsis and rice has shown that bHLH transcription factors are involved in stress responses and plant development, including cold stress, heat stress, abscisic acid, jasmonic acid and light response signaling pathways (*Ikeda et al., 2012; Kim et al., 2010; Liu et al., 2013; Paik et al., 2017*). Meanwhile, bHLH genes are involved in plant metabolite biosynthesis and trait development, including the formation of root hairs, anther development, and axillary meristem generation (*Fernández-Calvo et al., 2011; Song et al., 2014a; Wen et al., 2018*). However, little is known about the functions of the bHLH gene family in lotus. To better understand the second-largest gene family in plants, preliminary analyses of three aspects were conducted to reveal the functions of *NnbHLH* family genes in lotus for the first time.

Cis-element analyses showed that elements that can respond to diverse stresses (such as LTR, ABRE, TC-rich and HSE) were widely distributed in the *NnbHLH* gene family (Table S4). In addition, GO annotation revealed that most NnbHLH proteins were probably involved in the plant development process and responses to stimuli (Table. S7). Furthermore, the functions of 69 *NnbHLHs* were predicted based on their known and verified homologs in Arabidopsis and rice (Table S5), which were mainly associated with development processes (root hair development, seed dormancy, fruit dehiscence and flowering initiation) and stress responses (responses to low-temperature and salt stress and enhancement of stress tolerance) (Table S5). These analyses suggest that the bHLH gene family is also related to plant development, metabolic regulation, and the response to stress in lotus, in accordance with previous studies (*Li et al., 2006; Niu et al., 2017; Toledo-Ortiz et al., 2003; Zhang et al., 2015*). Next, we analyzed the candidate stress-response *NnbHLHs*, as improving stress tolerance in lotus is vital. Based on cis-element analyses, TC-rich cis-elements that may be involved in defense and stress responses were detected in the promoter

regions of 58 *NnbHLH* genes (Table S4). Meanwhile, 38 *NnbHLHs* contained LTR box, which responds to cold stress and 44 members contained HSE box, which responds to heat stress (Table S4). The results of GO annotation suggested that 109 member genes can respond to stimuli (Table. S7), with 13 and 28 *NnbHLHs* predicted to play roles in salt and cold stress, respectively (Table. S7). Phylogenetic analyses further suggested that 25 *NnbHLHs* may respond to stresses, including cold, salt, and drought, based on their homologs in Arabidopsis and rice (Table S5). Comprehensive analyses of these three datasets suggested that *NnbHLH 1*, *15*, *37*, *38*, *55*, *70*, *81*, *93*, and *98* and *NnbHLH13*, *15*, *16*, *17*, *27*, *29*, *66*, *70*, and *93* were likely to respond to cold and salt stress, respectively.

To validate the functional prediction of *NnbHLH*s, qRT-PCR analyses were conducted for 15 *NnbHLH* genes under cold (4°C) and salt (50 mM NaCl) treatments. *NnbHLH 1*, *15*, *37*, *38*, *55*, *70*, *81*, *93*, and *98* were all significantly ($P<0.01$) upregulated at all treatment time points, and were thus candidates of *NnbHLHs* that respond to cold stress. *NnbHLH 13*, *17*, *29*, *66*, and *70* were also upregulated under salt stress, as expected. We found that *NnbHLH 1*, *13*, *17*, *37*, *66*, *70*, and *81* were upregulated under both cold and salt stresses, in accordance with the results of GO annotations and cis-element and homolog analyses. Thus, *NnbHLH 1*, *13*, *17*, *37*, *66*, *70*, and *81* are the strongest candidates for lotus resistance genes due to their responsiveness to various stressors. The results of qRT-PCR analyses suggested that functional prediction of the *NnbHLH* gene family could provide valuable reference data for further functional research of this gene family.

Conclusions

In summary, we conducted a genome-wide evaluation of the bHLH gene family in lotus. The structural characteristics of this gene family were thoroughly investigated. Functional prediction of the *NnbHLH* family was systematically conducted for the first time using three methods. We also analyzed the expression patterns of 15 candidate genes under cold and salt treatments at several time points based on functional prediction. Taken together, the results and findings described in this study provide a strong basis for further investigation of the function and evolution of *NnbHLH*s. In addition, candidate genes for stress resistance in lotus were identified.

Acknowledgements

We appreciate Dr. Tao Shi for his kindly help during data analyses.

References

Ahmad A, Niwa Y, Goto S, Ogawa T, Shimizu M, Suzuki A, Kobayashi K, Kobayashi H. 2015. bHLH106 Integrates Functions of Multiple Genes through Their G-Box to Confer Salt Tolerance on Arabidopsis. *Plos One* 10:e0126872.

An R, Liu X, Wang R, Wu H, Liang S, Shao J, Qi Y, An L, Yu F. 2014. The over-expression of two transcription factors, ABS5/bHLH30 and ABS7/MYB101, leads to upwardly curly leaves. *Plos One* 9:e107637.

Atchley WR, Terhalle W, Dress A. 1999. Positional Dependence, Cliques, and Predictive Motifs in the bHLH Protein Domain. *Journal of Molecular Evolution* 48:501-516.

Bailey TL, Bodén M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009 MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* 37:W202-W208

Brownlie P, Ceska TA, Lamers M, Romier C, Stier G, Teo H, Suck D. 1997. The crystal structure of an intact human Max–DNA complex: new insights into mechanisms of transcriptional control. *Structure* 5:509.

Bullard MJ, Crawford TJ. 2010. Productivity of Lotus corniculatus L. (bird's-foot trefoil) in the UK when grown under low-input conditions as spaced plants, monoculture swards or mixed swards. *Grass & Forage Science* 50:439-446.

Cai Q, Fukushima H, Yamamoto M, Ishii N, Sakamoto T, Kurata T, Motose H, Takahashi T. 2016. The SAC51 Family Plays a Central Role in Thermospermine Responses in Arabidopsis. *Plant & Cell Physiology* 57:1583.

Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martínez-García JF, Bilbao-Castro JR, Robertson DL. 2010. Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. *Plant Physiology* 153:1398-1412.

Cheng L, Li S, Hussain J, Xu X, Yin J, Zhang Y, Chen X, and Li L. 2013. Isolation and functional characterization of a salt responsive transcriptional factor, LrbZIP from lotus root (Nelumbo nucifera Gaertn). *Molecular Biology Reports* 40:4033-4045.

Cheng X, Xiong R, Liu H, Wu M, Chen F, Yan H, Xiang Y. 2018. Basic helix-loop-helix gene family Genome wide identification, phylogeny, and expression in Moso bamboo. *Plant Physiology and Biochemistry* 132:104-119.

Choi B, Yoo J-M, Lee S-B, Kim G-J, Kim K-W, Yoo J-H, Oh K-S, Moon B-C, Kim W-H, Cho Y-H, Park S-W, Park Y-J. 2018. Genome-Wide Association Study of Rice Core Set Related to Grain Weight. *Plant Breeding and Biotechnology* 6:293-304. 10.9787/pbb.2018.6.3.293

Cifuentes-Esquivel N, Bou-Torrent J, Galstyan A, Gallemí MA, Sessa G, Salla Martret M, Roig-Villanova I, Ruberti I, Martínez-García JF. 2014. The bHLH proteins BEE and BIM positively modulate the shade avoidance syndrome in Arabidopsis seedlings. *Plant Journal for Cell & Molecular Biology* 75:989-1002.

Conesa A, Götz S, García-Gómez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674-3676.

Cui J, You C, Zhu E, Huang Q, Ma H, Chang F. 2016. Feedback Regulation of DYT1 by Interactions with Downstream bHLH Factors Promotes DYT1 Nuclear Localization and Anther Development. *Plant Cell* 28:1078.

Deng C, Ye H, Fan M, Pu T, Yan J. 2017. The rice transcription factors OsICE confer enhanced cold tolerance in transgenic Arabidopsis. *Plant Signaling & Behavior* 12:1.

Diao Y, Xu H, Li G, Yu A, Yu X, Hu W, Zheng X, Li S, Wang Y, Hu Z. 2014. Cloning a glutathione peroxidase gene from Nelumbo nucifera and enhanced salt tolerance by overexpressing in rice. *Molecular Biology Reports* 41:4919-4927.

Ellenberger T, Fass D, Arnaud M, Harrison SC. 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev* 8:970-980.

Feller A, Machemer K, Braun EL, Grotewold E. 2011. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant Journal for Cell & Molecular Biology* 66:94-116.

Feng Y, Xu P, Li B, Li P, Wen X, An F, Gong Y, Xin Y, Zhu Z, Wang Y. 2017. Ethylene promotes root hair growth through coordinated EIN3/EIL1 and RHD6/RSL1 activity in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* 114:13834.

Ferguson AC, Pearce S, Band LR, Yang C, Ferjentsikova I, King J, Yuan Z, Zhang D, Wilson ZA. 2017. Biphasic regulation of the transcription factor ABORTED MICROSPORES (AMS) is essential for tapetum and pollen development in Arabidopsis. *New Phytologist* 213:778-790.

Fernández-Calvo P, Chini A, Fernández Barbero G, Chico J-M, Gimenez-Ibanez S, Geerinck J, Eeckhout D, Schweizer F, Godoy M, Franco-Zorrilla JM. 2011. 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* 23:701.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry JP. 2014. The protein families database. *Nucleic Acids Res* 42:D222–D230.

Flagel LE, Wendel JF. 2009. Gene duplication and evolutionary novelty in plants. *New Phytologist* 183:557-564.

Gaillochet C, Jamge S, Van dWF, Angenent G, Immink R, Lohmann JU. 2018. A molecular network for functional versatility of HECATE transcription factors. *Plant Journal* 95.

Gao C, Sun J, Wang C, Dong Y, Xiao S, Wang X, Jiao Z. 2017. Genome-wide analysis of basic/helix-loop-helix gene family in peanut and assessment of its roles in pod development. *Plos One* 12:e0181843.

Group TAP. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161:105-121.

Halling MA, Topp CFE, Doyle CJ. 2010. Aspects of the productivity of forage legumes in Northern Europe. *Grass & Forage Science* 59:331-344.

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. 2014. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31:1296.

Hudson KA, Hudson ME. 2014. The Basic Helix-Loop-Helix Transcription Factor Family in the Sacred Lotus, Nelumbo Nucifera. *Tropical Plant Biology* 7:65-70.

Ikeda M, Fujiwara S, Mitsuda N, Ohme-Takagi M. 2012. A triantagonistic basic helix-loop-helix system regulates cell elongation in Arabidopsis. *Plant Cell* 24:4483-4497.

Ito S, Song YH, Josephson-Day AR, Miller RJ, Breton G, Olmstead RG, Imaizumi T. 2012. FLOWERING BHLH transcriptional activators control expression of the photoperiodic flowering regulator CONSTANS in Arabidopsis. *Proc Natl Acad Sci U S A* 109:3582-3587.

562 Jiang Y, Yang B, Deyholos MK. 2009. Functional characterization of the Arabidopsis bHLH92
563 transcription factor in abiotic stress. *Molecular Genetics and Genomics* 282:503-516.
564 Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G. 2017. PlantTFDB 4.0: toward a central
565 hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*
566 45:D1040-D1045.
567 Jin Q, Hu X, Li X, Wang B, Wang Y, Jiang H, Mattson N, and Xu Y. 2016. Genome-Wide
568 Identification and Evolution Analysis of Trehalose-6-Phosphate Synthase Gene Family in
569 Nelumbo nucifera. Frontiers in Plant Science 7:1445.
570 Karas B, Amyot L, Johansen C, Sato S, Tabata S, Kawaguchi M, Szczyglowski K. 2009.
571 Conservation of lotus and Arabidopsis basic helix-loop-helix proteins reveals new players
572 in root hair development. *Plant Physiology* 151:1175-1185.
573 Kim JA, Yun J, Kim YS, Woo JC, Park CM. 2005. A basic helix-loop-helix transcription factor
574 regulates cell elongation and seed germination. *Molecules & Cells* 19:334.
575 Kim SY, Kim BH, Lim CJ, Lim CO, Nam KH. 2010. Constitutive activation of stress-inducible
576 genes in a brassinosteroid-insensitive 1 (bri1) mutant results in higher tolerance to cold.
577 *Physiologia Plantarum* 138:191-204.
578 Kong H, Landherr LL, Frohlich MW, Leebens Mack J, Ma H, Depamphilis CW. 2010. Patterns
579 of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple
580 mechanisms of rapid gene birth. *Plant Journal* 50:873-885.
581 Kurt F, Filiz E. 2018. Genome-wide and comparative analysis of bHLH38, bHLH39, bHLH100
582 and bHLH101 genes in Arabidopsis , tomato, rice, soybean and maize: insights into iron
583 (Fe) homeostasis. *Biometals* 31:1-16.
584 Lalitha S. 2000. Primer Premier 5. *Biotech Software & Internet Report* 1: 270–272
585 Le Hir R, Castelain M, Chakraborti D, Moritz T, Dinant S, Bellini C. 2017. AtbHLH68
586 transcription factor contributes to the regulation of ABA homeostasis and drought stress
587 tolerance in Arabidopsis thaliana. *Physiologia Plantarum* 160:312-327.
588 Ledent V, Paquet O, Vervoort M. 2002. Phylogenetic analysis of the human basic helix-loop-helix
589 proteins. *Genome Biology* 3:1-18.
590 Ledent V, Vervoort M. 2001. The basic helix-loop-helix protein family: comparative genomics
591 and phylogenetic analysis. *Genome Research* 11:754-770.
592 Lee JH, Jung JH, Park CM. 2017. Light Inhibits COP1-Mediated Degradation of ICE Transcription
593 Factors to Induce Stomatal Development in Arabidopsis. *Plant Cell* 29:tpc.00371.02017.
594 Leivar P, Monte E, Al-Sady B, Carle C, Storer A, Alonso JM, Ecker JR, Quail PH. 2008. The
595 Arabidopsis phytochrome-interacting factor PIF7, together with PIF3 and PIF4, regulates
596 responses to prolonged red light by modulating phyB levels. *Plant Cell* 20:337-352.
597 Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S. 2002.
598 PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in
599 silico analysis of promoter sequences. Nucleic Acids Research 30:325-327.
600 Letunic I, Doerks T, Bork P. 2015. SMART: recent updates, new developments and status in 2015.
601 *Nucleic Acids Research* 43:257-260.
602 Li X, Duan X, Jiang H. 2006. Genome-wide analysis of basic/helix-loop-helix transcription factor
603 family in rice and Arabidopsis. *Plant Physiology* 141:1167.
604 Lian T, Xu Y, Li L, Su XD. 2017. Crystal Structure of Tetrameric Arabidopsis MYC2 Reveals the
605 Mechanism of Enhanced Interaction with DNA. *Cell Reports* 19:1334-1342.
606 Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA
607 polymorphism data. *Bioinformatics* 25: 1451-1452.

608 Liu Y, Li X, Li K, Liu H, Lin C. 2013. Multiple bHLH proteins form heterodimers to mediate
609 CRY2-dependent regulation of flowering-time in Arabidopsis. *Plos Genetics* 9:e1003861.
610 Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time
611 quantitative PCR and the 2(-delta delta CT) method. *Methods* 25:402–408.
612 Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science*
613 290:1151-1155.
614 Mai HJ, Lindermayr C, Von TC, Fink-Straube C, Durner J, Bauer P. 2015. Iron and FER-LIKE
615 IRON DEFICIENCY-INDUCED TRANSCRIPTION FACTOR-dependent regulation of
616 proteins and genes in Arabidopsis thaliana roots. *Proteomics* 15:3030-3047.
617 Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC,
618 Gonzales NR. 2017. CDD/SPARCLE: functional classification of proteins via subfamily
619 domain architectures. *Nucleic Acids Research* 45:D200-D203.
620 Munns R, Tester M. 2008. Mechanisms of salinity tolerance. *Annual Review of Plant Biology*
621 59:651-681.
622 Murre C, Mccaw PS, Baltimore D. 1989. A new DNA binding and dimerization motif in
623 immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56:777-
624 783.
625 Niu X, Guan Y, Chen S, Li H. 2017. Genome-wide analysis of basic helix-loop-helix (bHLH)
626 transcription factors in Brachypodium distachyon. *Bmc Genomics* 18:619.
627 Ohashi-Ito K, Bergmann DC. 2006. Arabidopsis FAMA controls the final
628 proliferation/differentiation switch during stomatal development. *Plant Cell* 18:2493-
629 2505.
630 Ohashi-Ito K, Matsukawa M, Fukuda H. 2013. An atypical bHLH transcription factor regulates
631 early xylem development downstream of auxin. *Plant & Cell Physiology* 54:398-405.
632 Paik I, Kathare PK, Kim JI, Huq E. 2017. Expanding Roles of PIFs in Signal Integration from
633 Multiple Processes. *Molecular Plant* 10:1035-1046.
634 Petridis A, Döll S, Nichelmann L, Bilger W, Mock HP. 2016. Arabidopsis thaliana G2-LIKE
635 FLAVONOID REGULATOR and BRASSINOSTEROID ENHANCED EXPRESSION1
636 are low-temperature regulators of flavonoid accumulation. *New Phytologist* 211:912-925.
637 Rai A, Umashankar S, Rai M. 2016. Coordinate Regulation of Metabolites Glycosylation and
638 Stress Hormones Biosynthesis by TT8 in Arabidopsis. *Plant Physiology and Biochemistry*
639 pp:00421.
640 Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, Wakimoto H, Yang C, Iwamoto M,
641 Abe T. 2013. Rice Annotation Project Database (RAP-DB): An Integrative and Interactive
642 Database for Rice Genomics. *Plant & Cell Physiology* 54:E6.
643 Sanghera GS, Wani SH, Hussain W, Singh NB. 2011. Engineering cold stress tolerance in crop
644 plants. *Current Genomics* 12:-.
645 Sasaki-Sekimoto Y, Saito H, Masuda S, Shirasu K, Ohta H. 2014. Comprehensive analysis of
646 protein interactions between JAZ proteins and bHLH transcription factors that negatively
647 regulate jasmonate signaling. *Plant Signaling & Behavior* 9:-.
648 Sharma N, Xin R, Kim DH, Sung S, Lange T, Huq E. 2016. NO FLOWERING IN SHORT DAY
649 (NFL) is a bHLH transcription factor that promotes flowering specifically under short-day
650 in Arabidopsis. *Development* 143:682.
651 Song S, Huang H, Gao H, Wang J, Wu D, Liu X, Yang S, Zhai Q, Li C, Qi T. 2014a. Interaction
652 between MYC2 and ETHYLENE INSENSITIVE3 modulates antagonism between
653 jasmonate and ethylene signaling in Arabidopsis. *Plant Cell* 26:263-279.

654 Song S, Qi T, Fan M, Zhang X, Gao H, Huang H, Wu D, Guo H, Xie D. 2013. The bHLH subgroup
655 IIId factors negatively regulate jasmonate-mediated plant defense and development. *Plos*
656 *Genetics* 9:e1003653.
657 Song XM, Huang ZN, Duan WK, Ren J, Liu TK, Li Y, Hou XL. 2014b. Genome-wide analysis of
658 the bHLH transcription factor family in Chinese cabbage (Brassica rapa ssp. pekinensis).
659 *Molecular Genetics and Genomics*.
660 Sun H, Fan HJ, Ling HQ. 2015. Genome-wide identification and characterization of the bHLH
661 gene family in tomato. *Bmc Genomics* 16:9.
662 Sun X, Wang Y, Sui N. 2018. Transcriptional regulation of bHLH during plant response to stress.
663 *Biochemical & Biophysical Research Communications* 503:397-401.
664 Tamure K, Stecher G, Pterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary
665 Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30:2725-2729
666 Toledo-Ortiz G, Huq E, Quail PH. 2003. The Arabidopsis basic/helix-loop-helix transcription
667 factor family. *The Plant Cell* 15:1749-1770.
668 Van De Peer Y, Maere S, Meyer A. 2009. The evolutionary significance of ancient genome
669 duplications. *Nature Reviews Genetics* 10:725-732.
670 Wang C, Yao X, Yu D, Liang G. 2017a. Fe-deficiency-induced expression of bHLH104 enhances
671 Fe-deficiency tolerance of Arabidopsis thaliana. *Planta* 246:1-11.
672 Wang J, Cheng G, Wang C, He Z, Lan X, Zhang S, Lan H. 2017b. The bHLH transcription factor
673 CgbHLH001 is a potential interaction partner of CDPK in halophyte Chenopodium
674 glaucum. *Scientific Reports* 7:8441.
675 Wang K, Deng J, Damaris RN, Yang M, Xu L, Yang P. 2015. LOTUS-DB: an integrative and
676 interactive database for Nelumbo nucifera study. *Database* 2015:bav023.
677 Wang N, Xiang Y, Fang L, Wang Y, Xin H, Li S. 2013. Patterns of Gene Duplication and Their
678 Contribution to Expansion of Gene Families in Grapevine. *Plant Molecular Biology*
679 *Reporter* 31:852-861.
680 Wang R, Zhao P, Kong N, Lu R, Pei Y, Huang C, Ma H, Chen Q. 2018. Genome-Wide
681 Identification and Characterization of the Potato bHLH Transcription Factor Family.
682 *Genes* 9:54.
683 Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee T-h, Jin H, Marler B, Guo H. 2012.
684 MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and
685 collinearity. 40:e49-e49.
686 Wen J, Li Y, Qi T, Gao H, Liu B, Zhang M, Huang H, Song S. 2018. The C-terminal domains of
687 Arabidopsis GL3/EGL3/TT8 interact with JAZ proteins and mediate dimeric interactions.
688 *Plant Signal Behav* 13:00-00.
689 Xia W, Yu H, Cao P, Luo J, Wang N. 2017. Identification of TIFY Family Genes and Analysis of
690 Their Expression Profiles in Response to Phytohormone Treatments andMelampsora
691 larici-populinaInfection in Poplar. *Frontiers in Plant Science* 8.
692 Xu W, Zhang N, Jiao Y, Li R, Xiao D, Wang Z. 2014. The grapevine basic helix-loop-helix
693 (bHLH) transcription factor positively modulates CBF-pathway and confers tolerance to
694 cold-stress in Arabidopsis. *Molecular Biology Reports* 41:5329-5342.
695 Xue JH, Dong WP, Botany IO. 2012. Nelumbonaceae: Systematic position and species
696 diversification revealed by the complete chloroplast genome. *Journal of Systematics &*
697 *Evolution* 50:477-487.
698 Yamamoto M, Takahashi T. 2017. Thermospermine enhances translation of SAC51 and SACL1
699 in Arabidopsis. *Plant Signaling & Behavior* 12:1583-1592.

Yang J, Gao M, Huang L, Wang Y, Van NS, Wan R, Guo C, Wang X, Gao H. 2017. Identification and expression analysis of the apple (Malus Ã— domestica) basic helix-loop-helix transcription factor family. *Sci Rep* 7:28.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology & Evolution* 24:1586-1591.

Yao P, Sun Z, Li C, Zhao X, Li M, Deng R, Huang Y, Zhao H, Chen H, Wu Q. 2018. Overexpression of Fagopyrum tataricum FtbHLH2 enhances tolerance to cold stress in transgenic Arabidopsis. *Plant Physiol Biochem* 125:85-94.

Zhai Y, Zhang L, Xia C, Fu S, Zhao G, Jia J, Kong X. 2016. The wheat transcription factor, TabHLH39, improves tolerance to multiple abiotic stressors in transgenic plants. *Biochemical & Biophysical Research Communications* 473:1321-1327.

Zhao K, Li S, Yao W, Zhou B, Li R, Jiang T. 2018a. Characterization of the basic helix–loop–helix gene family and its tissue-differential expression in response to salt stress in poplar. *Peerj* 6:e4502.

Zhao M, Yang JX, Mao TY, Zhu HH, Xiang L, Zhang J, Chen LQ. 2018b. Detection of Highly Differentiated Genomic Regions Between Lotus (Nelumbo nucifera Gaertn.) With Contrasting Plant Architecture and Their Functional Relevance to Plant Architecture. *Frontiers in Plant Science* 9.

Zhao Q, Xiang X, Liu D, Yang A, Wang Y. 2018c. Tobacco Transcription Factor NtbHLH123 Confers Tolerance to Cold Stress by Regulating the NtCBF Pathway and Reactive Oxygen Species Homeostasis. *Frontiers in Plant Science* 9:381.

Zhu L, Xin R, Bu Q, Shen H, Dang J, Huq E. 2016. A Negative Feedback Loop between PHYTOCHROME INTERACTING FACTORs and HECATE Proteins Fine-Tunes Photomorphogenesis in Arabidopsis. *Plant Cell* 28:855-874.

Zhuo X, Zheng T, Zhang Z, Zhang Y, Jiang L, Ahmad S, Sun L, Wang J, Cheng T, Zhang Q. 2018. Genome-Wide Analysis of the NAC Transcription Factor Gene Family Reveals Differential Expression Patterns and Cold-Stress Responses in the Woody Plant Prunus mume. *Genes* 9.

Zumajo-Cardona C, Ambrose BA, Pabón-Mora N. 2017. Evolution of the SPATULA/ ALCATRAZ gene lineage and expression analyses in the basal eudicot, Bocconia frutescens L. (Papaveraceae). *Evodevo* 8:5.

Supplementary materials legends

Supplemental Raw data. **The gene sequences of the *NnbHLHs* family of lotus.**

Supplemental Figure S1. **Phylogeny analysis of the bHLHs from lotus, Arabidopsis and rice.**

Supplemental Table S1. **Detailed Characteristics of *NnbHLHs* in lotus.**

Supplemental Table S2. **The Ka/Ks ratios and estimated divergence time for WGD/segmentally duplicated *NnbHLHs*.**

Supplemental Table S3. **Primers used in the qRT-PCR experiment.**

745
746 Supplemental Table S4. **Cis-elements analyses of the bHLH proteins in lotus.**
747
748 Supplemental Table S5. **Predicted functions of the 69 *NnbHLHs* and their functional verified**
749 **homologs in Arabidopsis and rice by phylogenetic analysis.**
750
751 Supplemental Table S6 **FPKM values of the *NnbHLHs* in four tissues.** FPKM: fragments per
752 kilobase of exon per million fragments mapped.
753
754 Supplemental Table S7 **Gene ontology (GO) annotation of lotus bHLH protein.**

Figure 1(on next page)

Conserved amino acids and multiple sequence alignment schematic diagrams of the *NnbHLH*s bHLH domains.

(A) **Conserved amino acids across *NnbHLH* domain**. The amino acid with sequence identity more than 50% was labeled with grey or black shading (B) **Sequence logo of *NnbHLH* domains.** The overall height of each stack represents the conservation of the sequence at that position. (C) **Multiple sequence alignment of the *NnbHLH* doamins.** Shading represents the degree of amino acids identity at each position, with the black shading indicating 100 % sequence identity.

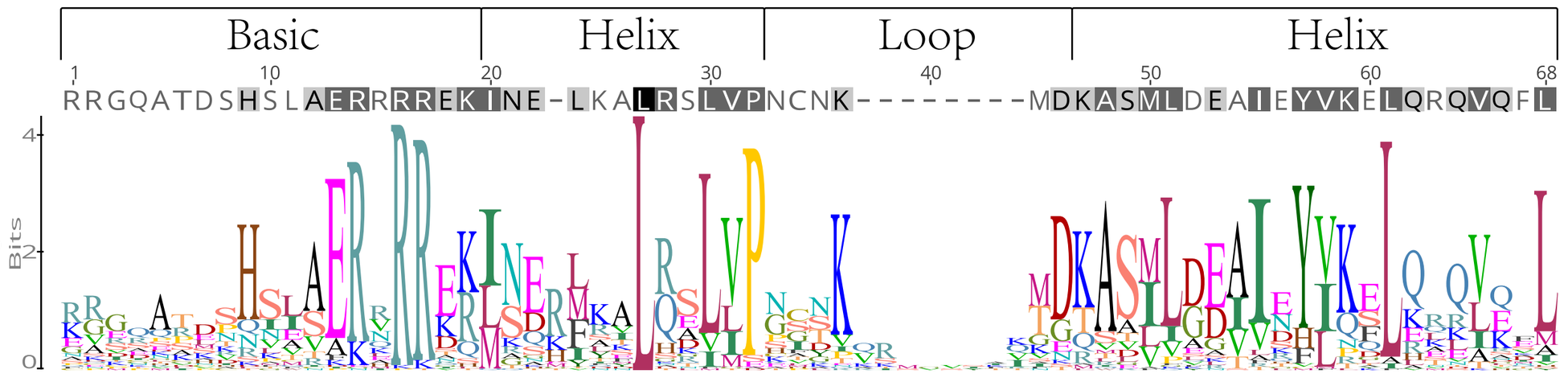
PeerJ

Manuscript to be reviewed

A

B

C



1. NnbHLH1
2. NnbHLH2
3. NnbHLH3
4. NnbHLH4
5. NnbHLH5
6. NnbHLH6
7. NnbHLH7
8. NnbHLH8
9. NnbHLH9
10. NnbHLH10
11. NnbHLH11
12. NnbHLH12
13. NnbHLH13
14. NnbHLH14
15. NnbHLH15
16. NnbHLH16
17. NnbHLH17
18. NnbHLH18
19. NnbHLH19
20. NnbHLH20
21. NnbHLH21
22. NnbHLH22
23. NnbHLH23
24. NnbHLH24
25. NnbHLH25
26. NnbHLH26
27. NnbHLH27
28. NnbHLH28
29. NnbHLH29
30. NnbHLH30
31. NnbHLH31
32. NnbHLH32
33. NnbHLH33
34. NnbHLH34
35. NnbHLH35
36. NnbHLH36
37. NnbHLH37
38. NnbHLH38
39. NnbHLH39
40. NnbHLH40
41. NnbHLH41
42. NnbHLH42
43. NnbHLH43
44. NnbHLH44
45. NnbHLH45
46. NnbHLH46
47. NnbHLH47
48. NnbHLH48
49. NnbHLH49
50. NnbHLH50
51. NnbHLH51
52. NnbHLH52
53. NnbHLH53
54. NnbHLH54
55. NnbHLH55
56. NnbHLH56
57. NnbHLH57
58. NnbHLH58
59. NnbHLH59
60. NnbHLH60
61. NnbHLH61
62. NnbHLH62
63. NnbHLH63
64. NnbHLH64
65. NnbHLH65
66. NnbHLH66
67. NnbHLH67
68. NnbHLH68
69. NnbHLH69
70. NnbHLH70
71. NnbHLH71
72. NnbHLH72
73. NnbHLH73
74. NnbHLH74
75. NnbHLH75
76. NnbHLH76
77. NnbHLH77
78. NnbHLH78
79. NnbHLH79
80. NnbHLH80
81. NnbHLH81
82. NnbHLH82
83. NnbHLH83
84. NnbHLH84
85. NnbHLH85
86. NnbHLH86
87. NnbHLH87
88. NnbHLH88
89. NnbHLH89
90. NnbHLH90
91. NnbHLH91
92. NnbHLH92
93. NnbHLH93
94. NnbHLH94
95. NnbHLH95
96. NnbHLH96
97. NnbHLH97
98. NnbHLH98
99. NnbHLH99
100. NnbHLH100
101. NnbHLH101
102. NnbHLH102
103. NnbHLH103
104. NnbHLH104
105. NnbHLH105
106. NnbHLH106
107. NnbHLH107
108. NnbHLH108
109. NnbHLH109
110. NnbHLH110
111. NnbHLH111
112. NnbHLH112
113. NnbHLH113
114. NnbHLH114
115. NnbHLH115

100% Identity
80 to 100% Identity
50 to 80% Identity
Less than 50% Identity

Figure 2(on next page)

Gene structures, phylogenetic relationships and conserved motifs analyses of the *NnbHLHs*.

(A) **Neighbor-joining phylogenetic tree of NnbHLHs**. (B) **Gene structure of *NnbHLH*s**. Orange box represent exon, blue box represent UTR and the black line represent intron. The sizes of exons can be estimated by the scale at bottom. (C) **Conserved motifs in the NnbHLH proteins.** 20 predicted motifs are represented by different colored boxes and motif sizes can be estimated by the scale at the bottom.

PeerJ

Manuscript to be reviewed

A

B

C

NnbHLH 15
NnbHLH 16
NnbHLH 17
NnbHLH 93
NnbHLH 66
NnbHLH 74
NnbHLH 13
NnbHLH 27
NnbHLH 25
NnbHLH 26
NnbHLH 110
NnbHLH 64
NnbHLH 65
NnbHLH 82
NnbHLH 83
NnbHLH 106
NnbHLH 108
NnbHLH 81
NnbHLH 44
NnbHLH 45
NnbHLH 37
NnbHLH 38
NnbHLH 39
NnbHLH 63
NnbHLH 75
NnbHLH 11
NnbHLH 12
NnbHLH 3
NnbHLH 4
NnbHLH 2
NnbHLH 21
NnbHLH 40
NnbHLH 41
NnbHLH 113
NnbHLH 114
NnbHLH 112
NnbHLH 115
NnbHLH 18
NnbHLH 19
NnbHLH 49
NnbHLH 46
NnbHLH 47
NnbHLH 48
NnbHLH 20
NnbHLH 89
NnbHLH 62
NnbHLH 104
NnbHLH 52
NnbHLH 84
NnbHLH 53
NnbHLH 54
NnbHLH 72
NnbHLH 99
NnbHLH 42
NnbHLH 43
NnbHLH 111
NnbHLH 68
NnbHLH 69
NnbHLH 70
NnbHLH 92
NnbHLH 105
NnbHLH 8
NnbHLH 9
NnbHLH 10
NnbHLH 14
NnbHLH 107
NnbHLH 50
NnbHLH 6
NnbHLH 7
NnbHLH 51
NnbHLH 77
NnbHLH 22
NnbHLH 23
NnbHLH 90
NnbHLH 24
NnbHLH 28
NnbHLH 30
NnbHLH 5
NnbHLH 88
NnbHLH 96
NnbHLH 73
NnbHLH 76
NnbHLH 58
NnbHLH 61
NnbHLH 87
NnbHLH 59
NnbHLH 60
NnbHLH 94
NnbHLH 95
NnbHLH 103
NnbHLH 85
NnbHLH 86
NnbHLH 34
NnbHLH 35
NnbHLH 97
NnbHLH 98
NnbHLH 55
NnbHLH 109
NnbHLH 78
NnbHLH 79
NnbHLH 80
NnbHLH 36
NnbHLH 56
NnbHLH 67
NnbHLH 91
NnbHLH 1
NnbHLH 29
NnbHLH 71
NnbHLH 102
NnbHLH 57
NnbHLH 101
NnbHLH 31
NnbHLH 100
NnbHLH 32
NnbHLH 33

Motif 1
Motif 2
Motif 3
Motif 4
Motif 5
Motif 6
Motif 7
Motif 8
Motif 9
Motif 10
Motif 11
Motif 12
Motif 13
Motif 14
Motif 15
Motif 16
Motif 17
Motif 18
Motif 19
Motif 20

5'
0kb 1kb 2kb 3kb 4kb 5kb 6kb 7kb 8kb 9kb 10kb 11kb 12kb 13kb 14kb 15kb 16kb 17kb 18kb 19kb 20kb 21kb 22kb 23kb 24kb 25kb 26kb 27kb
3'

100 200 300 400 500 600 700 800

Legend:
CDS
upstream/ downstream
Intron

PeerJ reviewing PDF | (2019:01:34453:2:0:NEW 14 May 2019)

Figure 3(on next page)

Distribution of divergence time of WGD pairs of the *NnbHLHs*.

The x-axis represents divergence time; y-axis represents the density of the distribution.

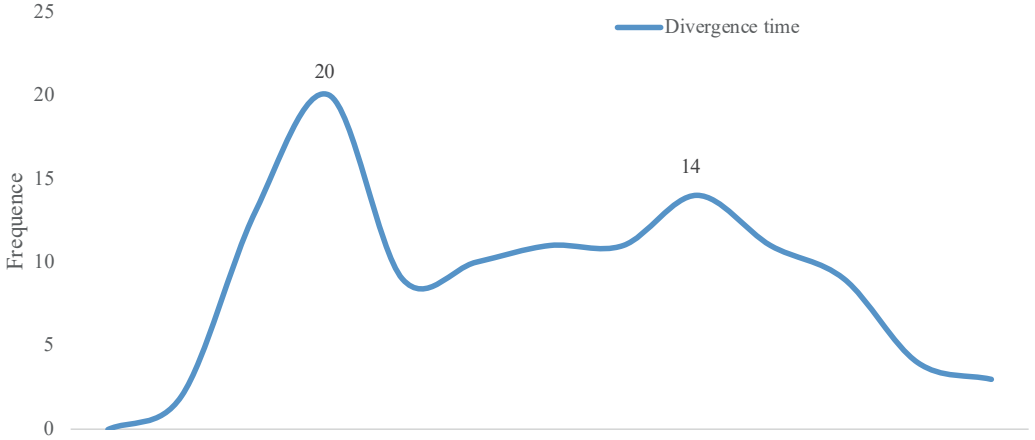

Divergence time
20
14
Frequence
Million years ago
0
5
10
15
20
25
8
22
36
50
64
78
92
106
120
134
147
161
175

Figure 4(on next page)

Expression patterns of the *NnbHLH* candidate genes in lotus under cold and salt treatment by qRT-PCR.

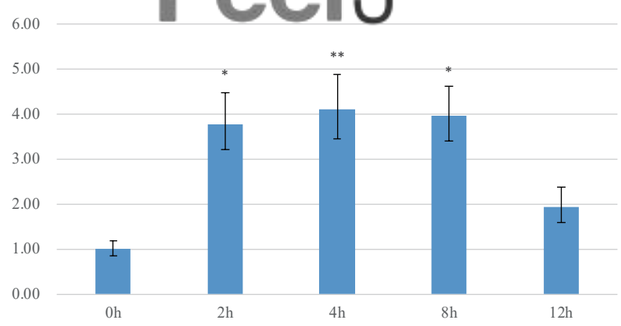
(A) **Expression patterns of *NnbHLH* candidate genes under cold stress treatment.**

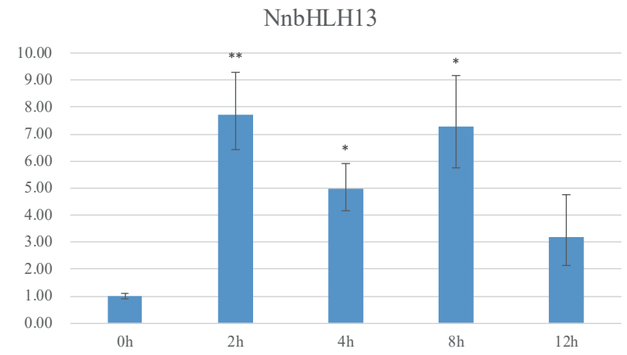
(B) **Expression patterns of *NnbHLH* candidate genes under salt stress treatment.** Y-axis: relative expression levels; X-axis: the time course of stress treatments; Error bar represent standard errors from three biological replicates. The asterisks indicate significant differences ($P<0.05$) between treatment group and control group (0h) with two asterisks indicate extremely significant differences ($P<0.01$).

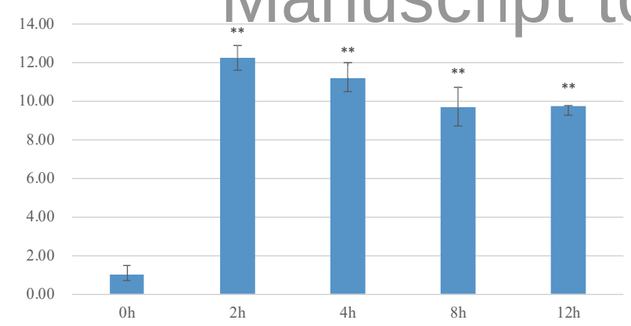
PeerJ

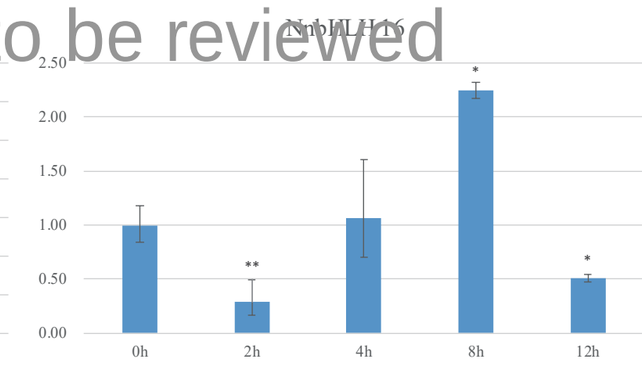
Manuscript to be reviewed

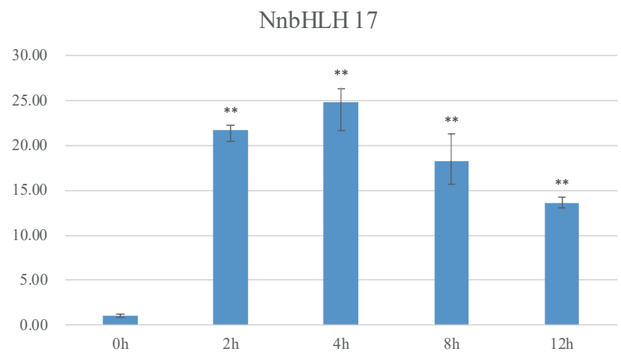
A

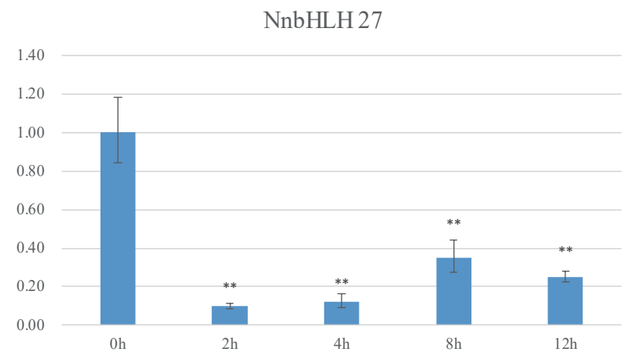


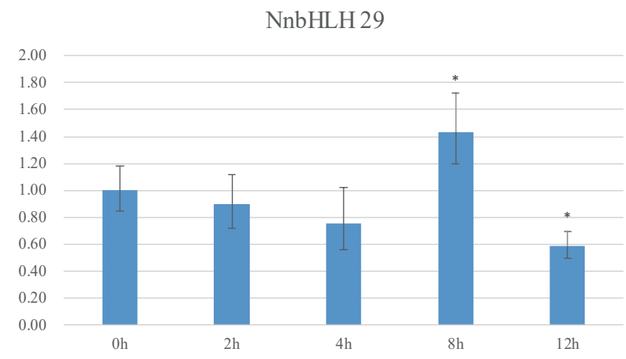


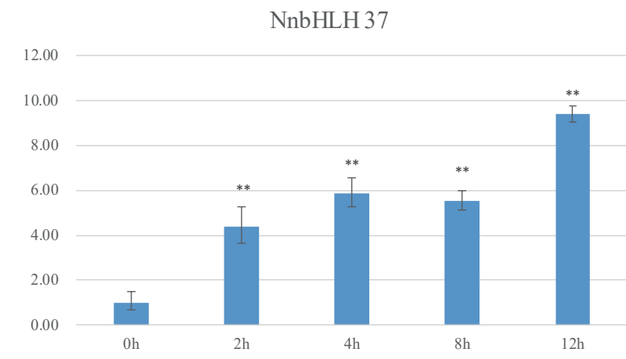


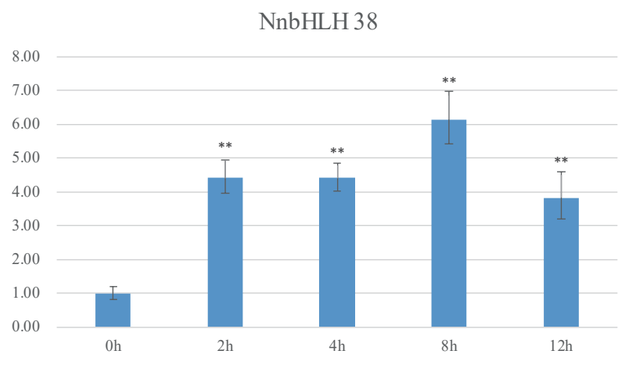


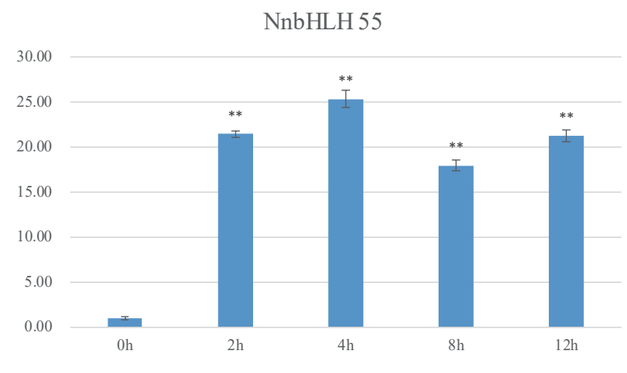


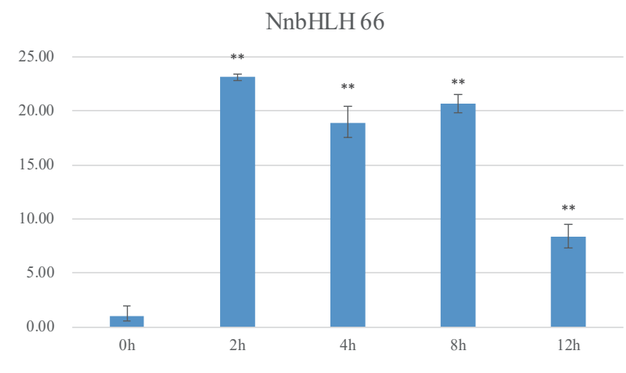


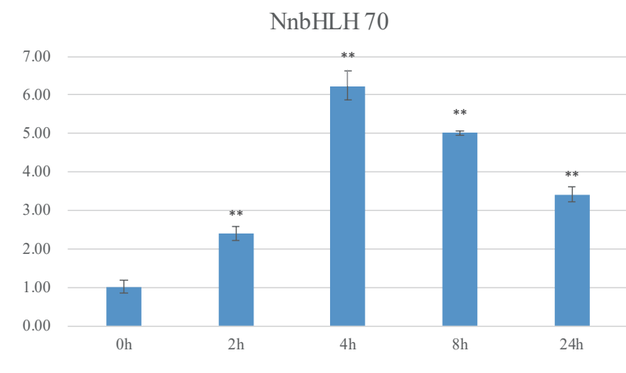


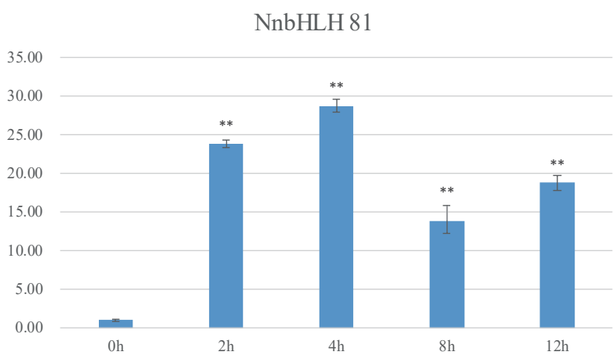


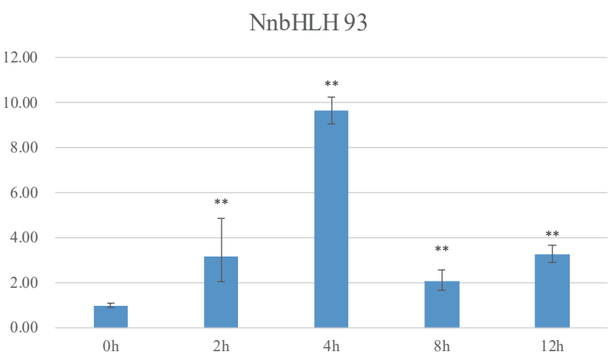


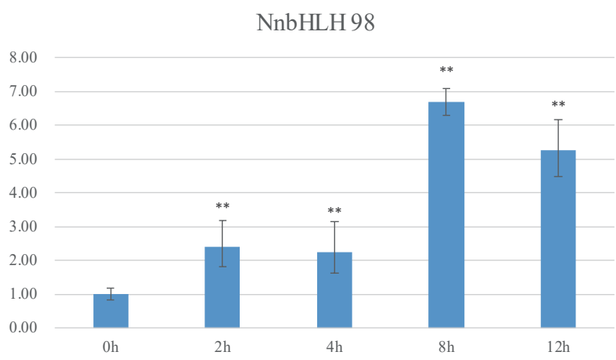




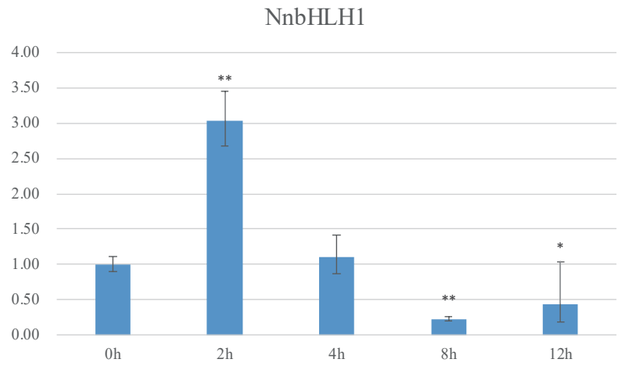


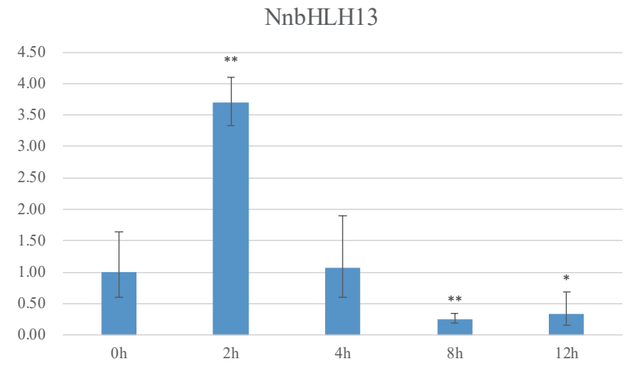


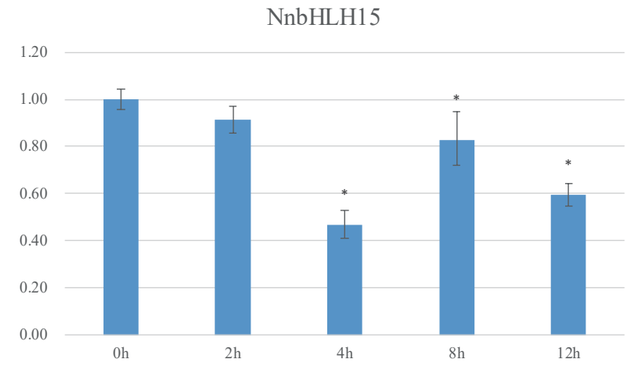


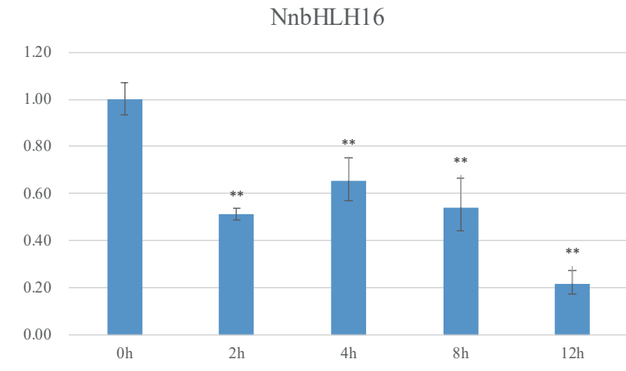


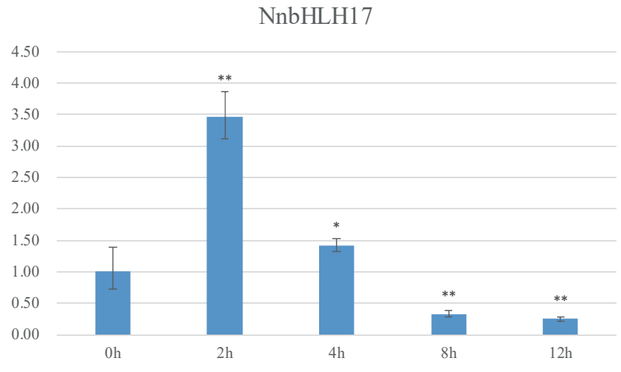
B

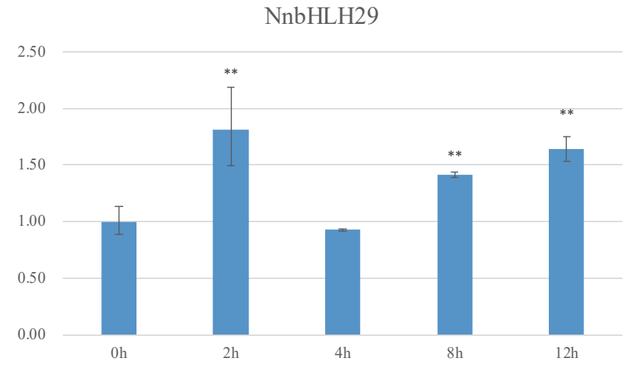


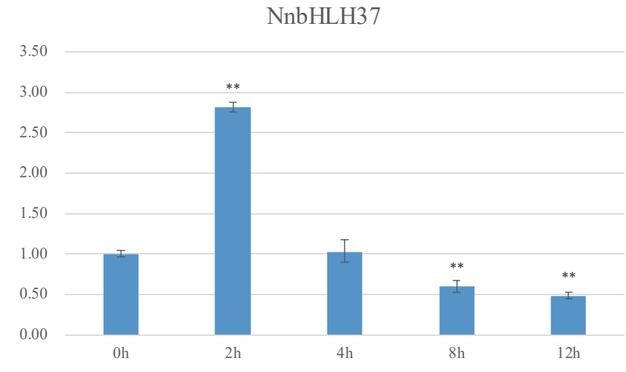


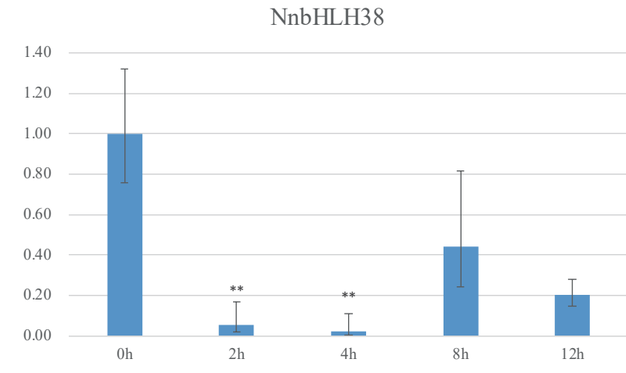


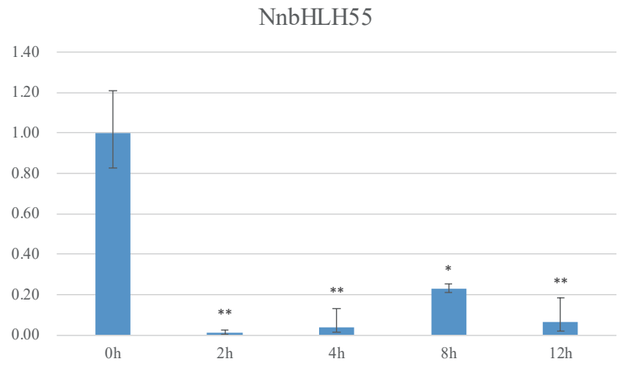


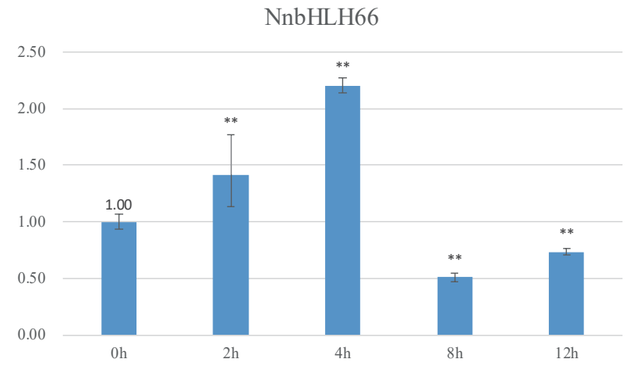


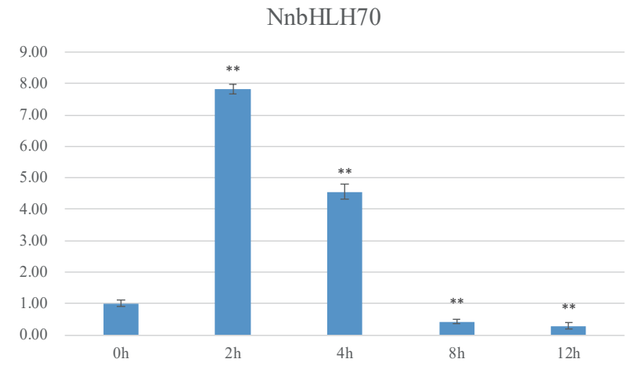


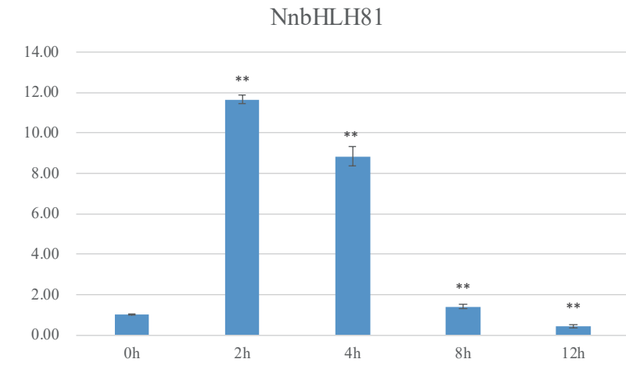


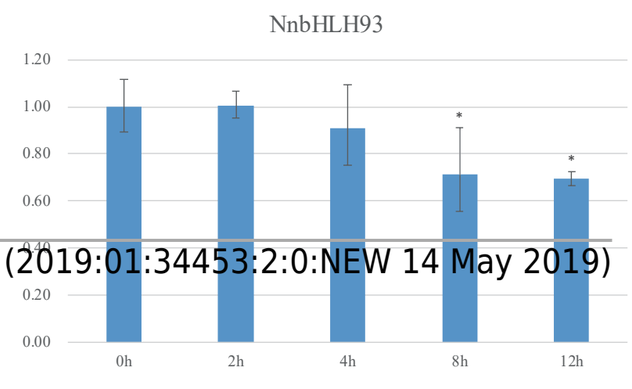












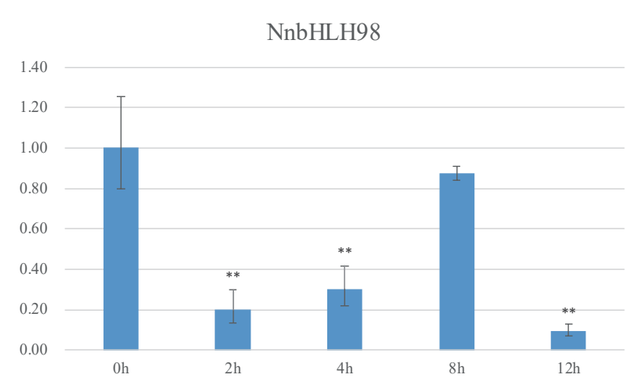


Table 1(on next page)

Consensus conserved amino acids in the bHLH domains between lotus, Arabidopsis, rice and *Brachypodium distachyon*.

Table 1. Consensus conserved amino acids in the bHLH domains between lotus, Arabidopsis, rice and *Brachypodium distachyon*.

Position in the alignment	Consensus amino acids and their ratios within the lotus bHLH domain	Position in the alignment	Consensus amino acids and their ratios within the Arabidopsis bHLH domain (*Toledo-Ortiz et al., 2003*)	Position in the alignment	Consensus amino acids and their ratios within the rice bHLH domain (*Li et al., 2006*)	Position in the alignment	Consensus amino acids and their ratios within the *Brachypodium distachyon* bHLH domain *(Niu et al., 2017*)
13	E (82%)	13	E (76%)	13	E (68%)	14	E (74%)
14	R (85%)	14	R (74%)	14	R (67%)	15	R (73%)
16	R (98%)	16	R (91%)	16	R (90%)	18	R (88%)
17	R (93%)	17	R (86%)	17	R (84%)	19	R (88%)
27	L (100%)	27	L (100%)	27	L (99%)	29	L (99%)
30	L (81%)	30	L (65%)	30	L (68%)	32	L (69%)
32	P (92%)	32	P (88%)	32	P (82%)	34	P (56%)
36	K (70%)			39	K (76%)	38	K (64%)
46	D (66%)	41	D (64%)	50	D (71%)	59	D (68%)
48	A (77%)	43	A (73%)	52	A (74%)	61	A (69%)
49	S (63%)	44	S (53%)	53	S (57%)	62	S (51%)
51	L (75%)	46	L (76%)	55	L (86%)	64	L (84%)
54	A (51%)	49	A (60%)	58	A (57%)	67	A (58%)
55	I (72%)	50	I (63%)	59	I (64%)	68	I (55%)
57	Y (77%)	52	Y (78%)	61	Y (83%)	70	Y (77%)
59	K (61%)	54	K (55%)	63	K (62%)	72	K (53%)
61	L (94%)	56	L (93%)	65	L (96%)	77	L (99%)

1