Characterization of the basic helix-loop-helix gene family and its tissue-differential expression in response to salt stress in poplar

Kai Zhao ¹, Shuxuan Li ¹, Wenjing Yao ¹,², Boru Zhou ¹, Renhua Li ¹, Tingbo Jiang Corresp. ¹

¹ State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, Harbin, China
² Northeast Institute of Geography and Agroecology, Chinese Academy of Science, Harbin, China

Corresponding Author: Tingbo Jiang
Email address: tbjiang@yahoo.com

The basic helix-loop-helix transcription factor gene family is one of the largest gene families and extensively involved in plant growth, development, and stress responses. However, limited studies are available on the gene family in poplar. In this study, we focused on 202 bHLH genes, exploring their DNA and protein sequences and physicochemical properties. According to their protein sequence similarities, we classified the genes into 25 groups with specific motif structures. In order to explore their expressions, we performed gene expression profiling using RNA-Seq and identified 19 genes that display tissue-differential expression patterns without treatment. Furthermore, we also performed gene expression profiling under salt stress. We found 74 differentially expressed genes, which are responsive to the treatment. Eighteen of the 19 genes correspond well to the differentially expressed genes. We validated the results using quantitative real-time PCR. This study lays the foundation for future studies on gene cloning, transgenes, and biological mechanisms.
Characterization of the basic helix-loop-helix gene family and its tissue-differential expression in response to salt stress in poplar

Kai Zhao¹, Shuxuan Li¹, Wenjing Yao¹,², Boru Zhou¹, Renhua Li¹, Tingbo Jiang¹

¹State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, 51 Hexing Road, Harbin 150040, China
²Northeast Institute of Geography and Agroecology, Chinese Academy of Science, 138 Haping Road, Harbin 150081, China

Corresponding Author:
Tingbo Jiang¹
Email address: tbjiang@yahoo.com

Abstract

The basic helix-loop-helix transcription factor gene family is one of the largest gene families and extensively involved in plant growth, development, and stress responses. However, limited studies are available on the gene family in poplar. In this study, we focused on 202 bHLH genes, exploring their DNA and protein sequences and physicochemical properties. According to their protein sequence similarities, we classified the genes into 25 groups with specific motif structures. In order to explore their expressions, we performed gene expression profiling using RNA-Seq and identified 19 genes that display tissue-differential expression patterns without treatment. Furthermore, we also performed gene expression profiling under salt stress. We found 74 differentially expressed genes, which are responsive to the treatment. Eighteen of the 19 genes correspond well to the differentially expressed genes. We validated the results using quantitative real-time PCR. This study lays the foundation for future studies on gene cloning, transgenes, and biological mechanisms.

Introduction

The bHLH transcription factor gene family is widely existed in eukaryotes and plays an important role in plant growth and development. However, limited studies are available in plants, especially in poplar. Since the first bHLH protein structure was analyzed in 1989 (Murre et al. 1989), more bHLH proteins have been identified. The bHLH gene family is named for its highly conserved bHLH domain, which consists of about 60 amino acids, including the basic region distributed at the N-terminus and the HLH region distributed at the C-terminus of the polypeptide chain (Feller et al. 2011). The basic region contains about 15 amino acids and plays an important role in target DNA recognizing and binding (Li et al. 2006). The HLH region, ranging from 40 to 50 amino acids in length, consists of two amphiphilic alpha helices and a loop
structure with uncertain length and sequence. The two amphiphilic alpha helices can form homodimer or heterodimeric to interact with other bHLH proteins (Ellenberger et al. 1994). The bHLH transcription factors can identify the element called E-box (5'-CANNTG-3'), with the most common one is G-Box (5'-CACGTG-3') (Li et al. 2006). Studies have shown that the nucleotides in the flank of the external of core element also impact specific binding (Mart et al. 2000). Previous studies indicated that the known bHLH proteins in animals can be divided into six groups of A-F (Atchley & Fitch 1997). In plants, however, many of the identified bHLH proteins belong to group B (Buck & Atchley 2003), the most members of which are characterized by binding to G-box (Li et al. 2006).

The bHLH transcription factors have various functions in plant growth and development. The bHLH genes SPT and ALC impact the development of pistils (Groszmann et al. 2010; Heisler et al. 2001), and PIL5 affects the germination of seeds in Arabidopsis thaliana (Oh et al. 2004). The bHLH gene RGE1 in Arabidopsis thaliana plays an important role in controlling the growth of embryos (Kondou et al. 2008), and the bHLH gene SPT is associated with final leaf size (Ichihashi et al. 2010). Rice bHLH gene LAX is the main regulator of leaf meristem (Komatsu et al. 2003), and rice bHLH gene bHLH142 is involved in the development process of pollens (Ko et al. 2014). The bHLH genes also regulate plant response to various abiotic stresses, such as drought (Abe et al. 1997; Seo et al. 2011), salinity (Jiang et al. 2009; Zhou et al. 2009), cold (Chinnusamy et al. 2003; Wang et al. 2003), high temperature (Koini et al. 2009), aluminum and iron deficiency (Kumari et al. 2008; Ling et al. 2002; Zhang et al. 2015), and low phosphorus stress (Yi et al. 2005). In addition, the bHLH genes play an important role in regulating multiple signal transduction pathways and impacting biosynthesis (Carretero-Paulet et al. 2010).

Poplar is an economically important tree in northern China, where salinity is one of the major abiotic stresses that limit poplar survival and growth. Since poplar is susceptible to salt, drought, cold, heavy metals, and other abiotic stresses, it is critical to identify important transcription factor gene families. In this study, we focus on 202 bHLH genes in poplar, in order to investigate their structures and functions, especially tissue-differential gene expression patterns in response to salt stress.

Materials and methods

Plant culture and stress treatment

We cut the twigs from the di-haploid Populus simonii × Populus nigra growing in the greenhouse. The twigs were cultivated in the same beakers with water in order to obtain new branches and roots. The conditions of culture include 60-70% relative humidity, 14-h light / 10-h dark cycles, and average temperature of 25 °C. After two months, we selected 20 plants with similar growth status, which were at random divided into two equal groups with two biological replicates. One of the groups was then treated with 150 mM NaCl for 24 hours. The other was
used as a control. For gene expression profiling, we collected root, stem, and leaf samples from each group. The 12 samples from each group were frozen in liquid nitrogen immediately, and stored in a refrigerator at -80 °C.

**Gene expression analysis based on RNA – Seq**

In order to explore tissue-differential expression patterns within the bHLH gene family in response to salinity, we sent the 12 samples to the GENEWIZ Company (https://www.genewiz.com/) for RNA-Seq using the Illumina HiSeq2500 platform. The sequencing generated pair-end reads with 150 bases. Construction of the RNA libraries was described in our previous study (Yao et al. 2016). Data processing methods for sequence reads are as follows: first, we used the Trimmomatic software (Bolger et al. 2014) to remove the adaptor sequences and low quality sequences from the original sequencing data; second, the high quality reads were mapped to the reference genome using the STAR software (Dobin et al. 2013) with default parameters. Finally, gene expression levels were calculated as fragments per kilo-bases per million mapped reads (FPKM) applying RSEM software (Li & Dewey 2011).

We identified differentially expressed genes (DEGs) by contrasting the treatment group with the control, using DESeq package (Anders & Huber 2010) from Bioconductor (http://www.bioconductor.org/). We used two thresholds for DEGs selection, that is, fold change >= 2 and padj <= 0.05 (p value adjusted for multiple testing at the false discovery rate of 0.05).

**Phylogenetic relationship and Physicochemical properties of the bHLH gene family**

In order to investigate the bHLH gene family in poplar, we obtained their amino acid sequences from the PlantTFDB database (Jin et al. 2017). Then the multiple sequence alignment was carried out by use of ClustalX 1.83, and the phylogenetic trees were constructed with MEGA 5.05 (Tamura et al. 2011; Thompson et al. 1997). The parameters are as follows: the neighbor-joining (NJ) method, 1000 iterations of bootstrap resample and poisson model. Since we are interested in general sequences similarities, we then classified the proteins into grand groups. We used bootstrap resample to support the classification. A threshold of the bootstrap score 50 is applied, which means 50 percent of the chances proteins can be found in the same group.

In addition, we constructed an evolutionary tree, using the best model selected by the SMS software and maximum likelihood (ML) method(Lefort et al. 2017). We obtained the DNA binding domain sequences of poplar bHLH family from the PlantTFDB database, Pfam database and SMART (Finn et al. 2016; Jin et al. 2017; Letunic & Bork 2018; Letunic et al. 2015), followed by performing multiple sequence alignment. The DNA binding domain obtained can be visualized by WebLogo (Crooks et al. 2004; Schneider & Stephens 1990). The physicochemical properties of proteins, including length, molecular weight, theoretical isoelectric point, aliphatic index, and grand average of hydropathicity, were predicted by ProtParam (Gasteiger et al. 2005).
Protein sequence motif analysis

Protein sequence motifs were identified by the MEME method (Bailey et al. 2009), and then the motif logos were obtained using the TBtools (https://github.com/CJ-Chen/TBtools). Annotations of motifs were obtained from Pfam database and InterProScan (Finn et al. 2017; Finn et al. 2016; Jones et al. 2014).

Gene structure analysis

We also downloaded the gene sequences from Phytozome (Goodstein et al. 2012; Tuskan et al. 2006). Structures of the bHLH gene family members were derived using Gene structure Display Server Program (Hu et al. 2015).

Subcellular location and Gene coordinate analysis

Subcellular localization of bHLH proteins was predicted by WoLF PSORT (Horton et al. 2007). Gene coordinate was identified by PopGenIE v3.0 database (Sjodin et al. 2009).

Verification by quantitative real-time PCR

In order to validate our results from RNA-Seq, we performed quantitative real-time PCR with 30 DEGs selected at random. Actin is used as a reference gene (Regier & Frey 2010). The primer sequences are listed in Table S1. The experimental procedures were described in our previous study (Yao et al. 2016). The expression level of each gene was calculated as relative to corresponding gene expression in leaf with no treatment.

Results

Identification of bHLH gene family and analysis of its physicochemical properties

We obtained 202 bHLH gene members in poplar from the PlantTFDB 4.0 (Jin et al. 2017). The gene family shares the DNA-binding domain. We named them *PtrbHLH1* to *PtrbHLH202* (Table S2). The aligned DNA binding domain sequences of poplar bHLH family are showed in Table S3, and the DNA binding domain alignment logo with the length of 60 amino acids is displayed in Fig. 1. Evidence from physicochemical properties indicted that lengths and molecular weights of the 202 corresponding proteins vary substantially (Table S2). The average length is 339.4 amino acids, ranging from 60 to 742. The mean of molecular weights is 37,634.23 Da (6,917.34-79,784.41 Da). The theoretical isoelectric points of these proteins are between 4.63 and 9.92. The aliphatic indexes are in the range of 32.67-108.33, which indicate that thermal stability of the proteins varies substantially. The grand average of hydropathicity runs the gamut of -1.885 and -0.06, indicating that these proteins are hydrophilic.
Phylogenetic trees and protein sequence motif analysis

In order to explore relationships of the bHLH gene family members in poplar, we constructed a dendrogram and a NJ-phylogenetic tree with their protein sequences (Fig. 2, S1). The proteins can be divided into 25 groups (A-Y). The reliability of the results is supported by the bootstrap resample test (Fig. S1). The average size of the groups has approximately eight members, ranging from two to twenty-six. To validate our phylogenetic tree, we also constructed a ML-phylogenetic tree with the best model of JTT+G+F (Fig. S2). All of groups have a good consistence on the ML-phylogenetic tree except group J and V with only six members.

As expected, members in each group, especially closely related proteins, have the same or similar conserved motifs. We obtained 15 conserved protein sequence motifs based on the MEME (Bailey et al. 2009). It is shown that motifs 1, 2, 3, 4, 6, 7 were annotated to the bHLH domain, by use of Pfam and InterProScan (Finn et al. 2017; Finn et al. 2016; Jones et al. 2014). Motifs 2 and 4 share the E-box/N-box specific site. Motif 5 is the ACT domain. Motif 8 is achaete-scute transcription factor-related. Motif 9 is the bHLH-MYC and R2R3-MYB transcription factors N-terminal. The others have no annotation (Table S4).

Based on MEME (Bailey et al. 2009), all the bHLH family genes contain the motif which was annotated as the bHLH domain, except bHLH202 without any motif (Fig. S3). Group A, B, C, E mainly contain motif 1 and motif 3; thirteen groups largely share motif 2; group O harbors motif 4; seven groups mainly have motif 6 and motif 7; and several proteins in multiple groups contain motif 1, 2 or 3. Regarding other motifs, motif 5 is mainly distributed in group K-R; motif 8 in group T; motif 9, 11, 12, 13 in group Q; motif 10 in group B and Y; motif 14 in group I; motif 15 in group K, N, O, P, Q. In addition, group F only contains motif 2, group U only harbors motif 6 and 7, and group Y only has motif 2 and 10. Motif 4 only occurs in group O.

Gene structure at the DNA level

In order to compare similarities of the gene structure in each group, we characterized introns and exons of each gene (Fig. S3). In general, the 202 genes differ substantially by length and the number of introns and exons. The majority of the bHLH gene family members have multiple introns, except for a few genes without a single intron (Fig. S3).

Within each of the 25 protein groups shown in the Fig. 2, multiple subgroups occur based on gene structure. For example, the first four members in group A share the similar number of introns and exons, as well as the length of the exons. Similar situation is observed for the last four members in group C (Fig. S3).

Subcellular localization and Gene coordinate
In order to characterize cellular distribution of the bHLH proteins, we predicted the localization of these proteins using the WoLF PSORT procedure (Horton et al. 2007). The results show that the 181 bHLH proteins were predicted to locate in the nucleus, six proteins in the cytoplasm, ten proteins in the chloroplast, three proteins in the mitochondria, and one protein in the Golgi apparatus. In addition, one protein was predicted to be double localization with shuttling between the cytoplasm and the nucleus (Table S5).

In order to examine chromosomal distribution of the bHLH gene family, we applied the PopGenIE v3.0 database (Sjodin et al. 2009). The results indicate that the bHLH gene family members are unevenly distributed on all 19 chromosomes. Chromosome 2 harbors a maximum number of 22 genes; in contrast, each of chromosomes 16, 17, and 18 contains a minimum of only six genes (Table S5). And the distribution is independent on the size of the chromosome. For example, chromosomes 1, the largest one, harbors 18 genes. Chromosomes 9, the minimal one, contains nine genes.

### Tissue-differential gene expression without salt treatment

The sequencing results are given in Table S6. The sequencing depths are approximately tenfold. Total reads for each sample range from 31.8 to 38.3 million. The total mapped reads run the gamut of 27.3 (77.3054%) and 36.4 (95.044%) million. The pearson correlation coefficients of uniquely mapped reads between biological replicates are from 0.961 to 0.994, which are highly significant (p-values < 2.2e-16).

To explore tissue-differential expression of the gene family without salt treatment, we first compared three sets of the tissue pairs, that is, leaf-root, leaf-stem, and root-stem (Data S1). Then we contrasted between the tissue pairs and retrieved genes that are shared. Results are shown in Fig. 3A-D.

In the comparison of the leaf-root pair, we identified 73 genes whose expression levels differ significantly. Similarly, we found 45 and 42 genes in the root-stem and leave-stem pairs, respectively (Data S1). We then focused on inter-pair comparisons. Regarding the root-leaf and root-stem pairs, forty genes are shared (Fig. 3B), indicating the gene expressions are different in roots relative to both leaves and stems. Similarly, thirty-four genes are shared in the leaf-root and leaf-stem pairs, followed by 22 genes in the stem-leaf and stem-root pairs (Fig. 3A, 3C). Among the three sets of shared parts, nineteen genes are shared (Fig. 3D and Table S7).

A heatmap regarding the 19 gene expressions across the three tissues is given in Fig. 4. As expected, there are distinct gene expression patterns for the 19 genes. In general, the 19 genes can be classified into three clusters that are tissue differential. Cluster 1 genes are highly expressed in leaves, lowly expressed in roots, and moderately expressed in stems. Cluster 2
genes are greatly expressed in stems, and lowly expressed in roots and leaves. Cluster 3 genes are highly, moderately, and lowly expressed in roots, stems, and leaves. Cluster 1 genes display an opposite pattern to cluster 3 genes.

In order to explore the direction of differential expression in different comparisons, we classified the tissue-differential genes into up- and down-regulated ones (Fig. S4). We found that most of the tissue-differential genes in leaves have similar trends, for instance, there are respective 15 genes consistently down-regulated and 17 genes agreeably up-regulated in comparison between the leaf-root and leaf-stem tissue pairs. In contrast, only two genes display inconsistency. Similar results are observed in the comparison between the root-leaf and root-stem tissue pairs. However, there is no such trend in the stems.

**Differential expression genes in response to salinity**

Screen for salinity responsive members in the bHLH gene family is critical for mechanistic understanding of gene regulation of stress tolerance. In the leaves treated with salt stress, we identified 26 DEGs in the family. Eight of the genes are up-regulated and 18 are down-regulated in treatment samples compared to controls. In roots, forty-seven DEGs were found, with 24 up-regulated and 23 down-regulated. In stems, nineteen DEGs were obtained, with nine up-regulated and 10 down-regulated (Data S2). The number of up-regulated DEGs in roots is larger than the down-regulated DEGs, which is opposite to the DEGs in leaves or stems. Among the URGs, variation of gene expression is remarkably greater in roots compared to that in the other two tissues. In contrast, among the DRGs, leaf tissue displays the widest gene expression variation (Fig. S5).

The expression degrees of the DEGs in each tissue are shown in Fig. S6. The majority of the DEGs are concentrated at 2-8 fold changes. It is only in roots that six genes with over-expressing of >16 fold changes. In contrast, 11 genes that are down-regulated with >16 fold changes are mainly in leaf tissue.

**Tissue-differential gene expression in response to salinity**

In order to observe the distribution of DEGs in different tissues, we drew the Venn Diagram (Fig. 3E). There are 74 DEGs that are responsive to salt stress in at least one tissue. There exist 16 (21.6%) of the DEGs in any two of the three tissue combinations, with nine (12.2%) in the leaf-stem pair, seven (9.5%) in the root-stem pair, and four (5.4%) in the leaf-root pair. Only 2 (2.7%) DEGs are shared across the three tissues. There are 38 (51.4%) DEGs that are specifically occurred in roots, followed by 15 (20.3%) in leaves, and 5 (6.8%) in stems (Fig. 3E). The gene list is given in Table S8. Number of respective DRGs and URGs are shown in Fig. 3F and 3G. Across the three tissues, the majority of DRGs and URGs are in the root tissue.
Two heatmap plots regarding the expression of 74 DEGs across the three tissues are displayed in Fig. 5. In general, the URGs, which are classified into four clusters, display distinct patterns that contrast between the tissues (Fig. 5A). For example, the top four genes in cluster 1 are highly expressed in roots and stems, but lowly expressed in leaves. Conversely, all the genes in cluster 3 display over expression in roots, but low expression in other two tissues. The DRGs, which are classified into four clusters, exhibit similar patterns (Fig. 5B). For example, all the genes in cluster 1 are highly expressed in leaves and stems, but lowly expressed in roots. In contrast, cluster 4 genes display an opposite pattern compared to the cluster 1 genes.

To connect the tissue-differential genes without salt treatment with the DEGs in response to salinity, we mapped the 19 genes onto the heatmaps of the DEGs (Fig. 5). In general, there is a good correspondence between the matched genes, regarding tissue difference. For example, the cluster 1 genes from Fig. 4 and highlighted in blue in Fig. 5 are highly or moderately expressed in leaves and stems.

**Validation of the DEGs by RT-qPCR**

In order to validate the DEGs in response to salt stress, which are identified by RNA-Seq, we selected at random 30 genes for RT-qPCR. Results turn out to be that the trend of the relative gene expression is comparable to that from RNA-Seq (Fig. 6). For example, the salt-responsive DEGs in leaf, such as *Potri.002G054100.1* and *Potri.002G248500.1*, are up-regulated with over two fold changes in both RNA-Seq and RT-qPCR. In contrast, DEGs in leaf, such as *Potri.012G055700.1* and *Potri.009G117300.1*, are down-regulated with over eight fold changes.

Congruent results are observed for DEGs in stems. For example, the salt-responsive DEGs in stems, such as *Potri.005G158100.1*, *Potri.002G172100.1* and *Potri.008G165700.1*, are over-expressed with over two fold changes in both experiments. *Potri.012G055700.1* is down-regulated with over seven fold changes.

Similar results are obtained for DEGs in roots. For example, the salt-responsive DEGs in roots, such as *Potri.012G104900.1* and *Potri.004G156000.1*, are up-regulated over two fold changes in both experiments. In contrast, *Potri.002G054100.1* is down-regulated over 3.5 fold changes.

**Discussion**

The bHLH gene family is one of the largest families in plants. In previous studies, the bHLH genes in Arabidopsis, poplar, rice, moss and algae had been updated and classified in 2010, and found 183 bHLH genes in *Populus trichocarpa* using Joint Genome Institute (JGI) Ptri version 1.1 (Carretero-Paulet et al. 2010). However, the gene ID used in the previous study is inconvenient to be found in the updated database. And the gene annotations in the updated database are also more complete. In this study, we obtained 202 poplar bHLH genes from the
PlantTFDB database which identifies transcription factors applying Joint Genome Institute (JGI) Ptri version 3.0 (Jin et al. 2017). And we have characterized the gene and protein structures of the 202 bHLH gene family members. According to the bootstrap values in the NJ-phylogenetic tree, the gene family can be classified into 25 distinct groups. And the groups correspond well to the ML-phylogenetic tree. Each group harbors specific sequence motifs that are corresponding to different protein domains. In addition, each group shares similar intron and exon structures. These results of motif and gene structure analysis are similar to the study of the poplar trihelix family (Wang et al. 2016). The previous study found that the domains outside the bHLH domain may have diverse molecular functions in Arabidopsis (Toledo-Ortiz et al. 2003). For example, the PAS domain of PIF3 plays a major role in the interaction with phytochrome B (Zhu et al. 2000). The poplar bHLH transcription factors also contain many different motifs outside the bHLH domain, which suggests that poplar bHLH gene family members have different functions. Since the functional studies of bHLH transcription factors are limited in poplar and there are many amino acid motifs whose functions are unknown in this family, the functional diversity of poplar bHLH transcription factors need to be explored in the further study.

For the subcellular localization prediction, most of the bHLH proteins were predicted to locate in the nucleus except for 21 members locating in other organelles. We speculate that they may play a role in these organelles and enter the nucleus by interacting with other transcription factors or under stress conditions. And we found that nine of the 21 members inducible by salt stress. Due to lack of studies on poplar regarding the gene family, we focused on investigating tissue-differential gene expression. We have identified 19 genes that demonstrate tissue-differential gene expression patterns. The 19 genes can be classified into three distinct clusters.

To better understand functions of the 19 genes, we then annotated the genes onto the genome of Arabidopsis thaliana using Phytozome (Table S9) (Goodstein et al. 2012; Lamesch et al. 2012). The best-hit homologous gene of Potri.002G108400.1 is AT5G65640, which promotes flowering under short-day condition (Sharma et al. 2016). AT2G24260, the homologous gene of Potri.006G186600.1, plays a role in root development (Lin et al. 2015). ATIG73830, the homologous gene of Potri.012G055700.1, is required for normal growth of Arabidopsis thaliana (Friedrichsen et al. 2002). AT2G18300, the homologous gene of Potri.007G023600.1, is associated with Arabidopsis thaliana growth and immune antagonism (Malinovsky et al. 2014a). These studies suggest that the genes are important in the growth and development of poplar.

Based on the analysis of the expression levels of bHLH genes under salt stress, we have identified 74 DEGs that are responsive to the stress in at least one tissue. There exist 16 (21.6%) DEGs in the three tissue pairs, with four (5.4%) DEGs in the leaf-root pair, nine (12.2%) in the leaf-stem pair, and seven (9.5%) in the root-stem pair. Only 2 (2.7%) DEGs are shared across the three tissues, suggesting that different tissues may have diversified mechanisms in the regulation of response of salinity in poplar. There are 38 (51.4%) DEGs that are specifically occurred in
roots, followed by 15 (20.3%) in leaves and 5 (6.8%) in stems (Fig.4e). It is clear that the root and leaf tissues play a more significant role in salt stress responses, compared to stem tissue. The enrichment of DEGs in root is associated with the fact that root is a primary stress perception and response organ.

By annotating the sequences of the DEGs onto the genome of Arabidopsis thaliana, we found that homologous genes in Arabidopsis are associated with plant growth, development and stress response (Table S9) (Goodstein et al. 2012; Lamesch et al. 2012). For example, AT2G18300, the best-hit homologous gene of Potri.005G121900.1, is related to the growth and immunity of Arabidopsis (Malinovsky et al. 2014b). AT1G59640, the best-hit homologous gene of Potri.008G190800.1, plays a part in Arabidopsis petal growth (Varaud et al. 2011). AT3G47640, the best-hit homologous gene of Potri.003G207200.1, impacts stomatal opening (Takahashi et al. 2013). In this study, the Potri.003G207200.1 gene was down-regulated with > 32 fold changes in poplar roots challenged with salinity, compared to the control (Data S2). Since the stomata plays an important role in regulating water balance under stress, we speculate that Potri.003G207200.1 gene expresses in roots may regulate signal transduction pathways that are related to stomatal opening and closing in response to the stress.

Then, we mapped these genes into the gene family groups, and found that the genes, which show tissue-differential expression patterns, distribute in nine groups unequally, and salt stress response genes widely distribute in 20 groups. It should be noted that many gene pairs among these genes are in the closest position of the phylogenetic tree. It also demonstrates that the genes, which have closer genetic relationship, may have similar functions.

It is interesting to compare the genes that have tissue-differential expression before and after salt treatment. Before the treatment, we identified 19 significant genes, with 11 over expressed in leaf, followed by five in roots and three in stems (Fig. 4). After the treatment, we found 74 significant DEGs in response to salinity (Fig. 3E). The DEGs specific to root, leaf, and stem are 38, 15, and 5, respectively (Fig. 3E).

It is noteworthy that the 18 of the 19 bHLH genes showing tissue-differential expression without treatment correspond well to those DEGs that are responsive to salinity (Fig. 5). We then selected at random 30 DEGs for RT-qPCR for validation. The results from RNA-Seq and RT-qPCR are congruent (Fig. 6).

**Conclusion**

In this study, we focus on the 202 bHLH gene family members in poplar, starting from analysis of their physicochemical properties, evolutionary relationship, and gene structures, followed by
tissue differential gene expression and differential expression in response to salinity. Around one-third of the genes are found to play a significant role in regulating salinity response. The majority of the DEGs display gene expression patterns in a tissue-differential fashion. This study lays the foundation for future work in gene cloning, transgenes, and biological mechanisms.

Reference

Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, and Shinozaki K. 1997. Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. The Plant Cell 9:1859-1868.

Anders S, and Huber W. 2010. Differential expression analysis for sequence count data. Genome Biol 11:R106. 10.1186/gb-2010-11-10-r106

Atchley WR, and Fitch WM. 1997. A natural classification of the basic helix–loop–helix class of transcription factors. Proceedings of the National Academy of Sciences 94:5172-5176.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, and Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202-208. 10.1093/nar/gkp335

Bolger AM, Lohse M, and Usadel B. 2014. Trimmomatic: a flexible trimer for Illumina sequence data. Bioinformatics 30:2114.

Buck MJ, and Atchley WR. 2003. Phylogenetic analysis of plant basic helix-loop-helix proteins. J Mol Evol 56:742-750. 10.1007/s00239-002-2449-3

Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martinez-Garcia JF, Bilbao-Castro JR, and Robertson DL. 2010. Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. Plant Physiol 153:1398-1412. 10.1104/pp.110.153593

Chinnusamy V, Ohta M, Kanrar S, Lee BH, Hong X, Agarwal M, and Zhu JK. 2003. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Genes Dev 17:1043-1054. 10.1101/gad.1077503

Crooks GE, Hon G, Chandonia JM, and Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14:1188-1190. 10.1101/gr.849004

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15.

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

Feller A, Machemer K, Braun EL, and Grotewold E. 2011. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant J 66:94-116. 10.1111/j.1365-313X.2010.04459.x

Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztanyi Z, El-Gebali S, Fraser M, Gough J, Haft D, Hollday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Oregan CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto SC, Wu CH, Xenarios I, Yeh LS, Young SY, and Mitchell AL. 2017. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res 45:D190-D199. 10.1093/nar/gkw1107

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, and Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279-285. 10.1093/nar/gkv1344

Friedrichsen DM, Nemhauser J, Muramitsu T, Maloof JN, Alonso J, Ecker JR, Furuya M, and Chory J. 2002. Three
redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth. *Genetics* 162:1445-1456.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, and Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, ed. *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press, 571-607.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, and Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40:D1178-1186. 10.1093/nar/gkr944

Groszmann M, Bylstra Y, Lampugnani ER, and Smyth DR. 2010. Regulation of tissue-specific expression of SPATULA, a bHLH gene involved in carpel development, seedling germination, and lateral organ growth in Arabidopsis. *J Exp Bot* 61:1495-1508. 10.1093/jxb/erq015

Heisler MG, Atkinson A, Bylstra YH, Walsh R, and Smyth DR. 2001. SPATULA, a gene that controls development of carpel margin tissues in Arabidopsis, encodes a bHLH protein. *Development* 128:1089-1098.

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, and Nakai K. 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Res* 35:W585-587. 10.1093/nar/gkm259

Hu B, Jin J, Guo AY, Zhang H, Luo J, and Gao G. 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31:1296-1297. 10.1093/bioinformatics/btu817

Ichihashi Y, Horiguchi G, Gleissberg S, and Tsukaya H. 2010. The bHLH transcription factor SPATULA controls final leaf size in Arabidopsis thaliana. *Plant Cell Physiol* 51:252-261. 10.1093/pcp/pcp184

Jiang Y, Yang B, and Deyholos MK. 2009. Functional characterization of the Arabidopsis bHLH92 transcription factor in abiotic stress. *Mol Genet Genomics* 282:503-516. 10.1007/s00438-009-0481-3

Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, and Gao G. 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res* 45:D1040-D1045. 10.1093/nar/gkw982

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, and Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236-1240. 10.1093/bioinformatics/btu031

Ko SS, Li MJ, Sun-Ben Ku M, Ho YC, Lin YJ, Chuang MH, Hsing HX, Lien YC, Yang HT, Chang HC, and Chan MT. 2014. The bHLH142 Transcription Factor Coordinates with TDR1 to Modulate the Expression of EAT1 and Regulate Pollen Development in Rice. *Plant Cell* 26:2486-2504. 10.1105/tpc.114.126292

Koini MA, Alvey L, Allen T, Tilley CA, Harberd NP, Whitelam GC, and Franklin KA. 2009. High temperature-mediated adaptations in plant architecture require the bHLH transcription factor PIF4. *Curr Biol* 19:408-413. 10.1016/j.cub.2009.01.046

Komatsu K, Maekawa M, Ujiie S, Satake Y, Furutani I, Okamoto H, Shimamoto K, and Kyozuka J. 2003. LAX and SPA: major regulators of shoot branching in rice. *Proc Natl Acad Sci U S A* 100:11765-11770. 10.1073/pnas.1932414100

Kondou Y, Nakazawa M, Kawashima M, Ichikawa T, Yoshizumi T, Suzuki K, Ishikawa A, Koshi T, Matsui R, Muto S, and Matsui M. 2008. RETARDED GROWTH OF EMBRYO1, a new basic helix-loop-helix protein, expresses in endosperm to control embryo growth. *Plant Physiol* 147:1924-1935. 10.1104/pp.108.118364

Kumari M, Taylor GJ, and Deyholos MK. 2008. Transcriptomic responses to aluminum stress in roots of Arabidopsis

PeerJ reviewing PDF | (2017:11:22030:1:1:NEW 19 Feb 2018)
thaliana. *Mol Genet Genomics* 279:339-357. 10.1007/s00438-007-0316-z

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, and Huala E. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res* 40:D1202-1210. 10.1093/nar/gkr1090

Lefort V, Longueville J-E, and Gascuel O. 2017. SMS: Smart Model Selection in PhyML. *Mol Biol Evol* 34:2422-2424. 10.1093/molbev/msx149

Letunic I, and Bork P. 2012. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res* 40:D1202-1210. 10.1093/nar/gkr1090

Letunic I, Doerks T, and Bork P. 2015. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res* 43:D257-260. 10.1093/nar/gku949

Li B, and Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *Bmc Bioinformatics* 12:323.

Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, Guo J, Liang W, Chen L, Yin J, Ma H, Wang J, and Zhang D. 2006. Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. *Plant Physiol* 141:1167-1184. 10.1104/pp.106.080580

Lin Q, Ohashi Y, Kato M, Tsuge T, Gu H, Qu L-J, and Aoyama T. 2015. GLABRA2 directly suppresses basic helix-loop-helix transcription factor genes with diverse functions in root hair development. *The Plant Cell* 27:2894-2906.

Ling HQ, Bauer P, Bereczky Z, Keller B, and Ganal M. 2002. The tomato fer gene encoding a bHLH protein controls iron-uptake responses in roots. *Proc Natl Acad Sci U S A* 99:13938-13943. 10.1073/pnas.212448699

Long TA, Tsukagoshi H, Busch W, Lahner B, Salt DE, and Benfey PN. 2010. The bHLH transcription factor POPEYE regulates response to iron deficiency in Arabidopsis roots. *Plant Cell* 22:2219-2236. 10.1105/tpc.110.074096

Malinovsky FG, Batoux M, Schwessinger B, Youn JH, Stransfeld L, Win J, Kim S-K, and Zipfel C. 2014a. Antagonistic regulation of growth and immunity by the Arabidopsis basic helix-loop-helix transcription factor homolog of brassinosteroid enhanced expression2 interacting with increased leaf inclination1 binding bHLH1. *Plant Physiol* 164:1443-1455.

Malinovsky FG, Batoux M, Schwessinger B, Youn JH, Stransfeld L, Win J, Kim SK, and Zipfel C. 2014b. Antagonistic regulation of growth and immunity by the Arabidopsis basic helix-loop-helix transcription factor homolog of brassinosteroid enhanced expression2 interacting with increased leaf inclination1 binding bHLH1. *Plant Physiol* 164:1443-1455. 10.1104/pp.113.234625

Mart, iacute, nez G, iacute, and a JF. 2000. Direct Targeting of Light Signals to a Promoter Element-Bound Transcription Factor. *Science* 288:859-863. 10.1126/science.288.5467.859

Murre C, McCaw PS, and Baltimore D. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56:777-783. 10.1016/0092-8674(89)90682-x

Oh E, Kim J, Park E, Kim JJ, Kang C, and Choi G. 2004. PIL5, a phytochrome-interacting basic helix-loop-helix protein, is a key negative regulator of seed germination in Arabidopsis thaliana. *Plant Cell* 16:3045-3058. 10.1105/tpc.104.025163

Regier N, and Frey B. 2010. Experimental comparison of relative RT-qPCR quantification approaches for gene
expression studies in poplar. *BMC Mol Biol* 11:57. 10.1186/1471-2199-11-57

Schneider TD, and Stephens RM. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 18:6097-6100. 10.1093/nar/18.20.6097

Seo JS, Joo J, Kim MJ, Kim YK, Nam BH, Song SI, Cheong JJ, Lee JS, Kim JK, and Choi YD. 2011. OsbHLH148, a basic helix-loop-helix protein, interacts with OsJAZ proteins in a jasmonate signaling pathway leading to drought tolerance in rice. *Plant J* 65:907-921. 10.1111/j.1365-313X.2010.04477.x

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

Sjodin A, Street NR, Sandberg G, Gustafsson P, and Jansson S. 2009. The Populus Genome Integrative Explorer (PopGenIE): a new resource for exploring the Populus genome. *New Phytol* 182:1013-1025. 10.1111/j.1469-8137.2009.02807.x

Takahashi Y, Ebisu Y, Kinoshita T, Doi M, Okuma E, Murata Y, and Shimazaki K. 2013. bHLH transcription factors that facilitate K(+) uptake during stomatal opening are repressed by abscisic acid through phosphorylation. *Sci Signal* 6:ra48. 10.1126/scisignal.2003760

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731-2739. 10.1093/molbev/msr121

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, and Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876-4882.

Toledo-Ortiz G, Huq E, and Quail PH. 2003. The Arabidopsis Basic/Helix-Loop-Helix Transcription Factor Family. *The Plant Cell* 15:1749-1770. 10.1105/tpc.013839

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, and Salamov A. 2006. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science* 313:1596-1604.

Varaud E, Brioudes F, Szecsi J, Leroux J, Brown S, Perrot-Rechenmann C, and Bendahmane M. 2011. AUXIN RESPONSE FACTOR8 regulates Arabidopsis petal growth by interacting with the bHLH transcription factor BIGPETALp. *Plant Cell* 23:973-983. 10.1105/tpc.110.081653

Wang YJ, Zhang ZG, He XJ, Zhou HL, Wen YX, Dai JX, Zhang JS, and Chen SY. 2003. A rice transcription factor OsbHLH1 is involved in cold stress response. *Theor Appl Genet* 107:1402-1409. 10.1007/s00122-003-1378-x

Wang Z, Liu Q, Wang H, Zhang H, Xu X, Li C, and Yang C. 2016. Comprehensive analysis of trihelix genes and their expression under biotic and abiotic stresses in Populus trichocarpa. *Scientific Reports* 6:36274.

Yao W, Wang S, Zhou B, and Jiang T. 2016. Transgenic poplar overexpressing the endogenous transcription factor ERF76 gene improves salinity tolerance. *Tree Physiol* 36:896-908. 10.1093/treephys/tpw004

Yi K, Wu Z, Zhou J, Du L, Guo L, Wu Y, and Wu P. 2005. OsPTF1, a novel transcription factor involved in tolerance to phosphate starvation in rice. *Plant Physiol* 138:2087-2096. 10.1104/pp.105.063115

Zhang J, Liu B, Li M, Feng D, Jin H, Wang P, Liu J, Xiong F, Wang J, and Wang HB. 2015. The bHLH transcription factor bHLH104 interacts with IAA-LEUCINE RESISTANT3 and modulates iron homeostasis in Arabidopsis. *Plant Cell* 27:787-805. 10.1105/tpc.114.132704

Zhou J, Li F, Wang JL, Ma Y, Chong K, and Xu YY. 2009. Basic helix-loop-helix transcription factor from wild rice
(OrbHLH2) improves tolerance to salt- and osmotic stress in Arabidopsis. *J Plant Physiol* 166:1296-1306. 10.1016/j.jplph.2009.02.007

Zhu Y, Tepperman JM, Fairchild CD, and Quail PH. 2000. Phytochrome B binds with greater apparent affinity than phytochrome A to the basic helix-loop-helix factor PIF3 in a reaction requiring the PAS domain of PIF3. *Proc Natl Acad Sci U S A* 97:13419-13424. 10.1073/pnas.230433797
**Figure 1** (on next page)

The DNA binding domain alignment logo of poplar bHLH family.
Figure 2 (on next page)

Dendrogram of bHLH gene family proteins.

Each color represents a special group.
Figure 3 (on next page)

Venn diagrams of tissue-differential bHLH genes and DEGs challenged with salinity.

(A-C) Number of genes displaying distinct and shared expression in tissue-pairs without salt treatment. The shared genes express differently in two tissue-pairs. (A) Comparison between leaf-root and leaf-stem pairs. (B) Comparison between root-leaf and root-stem pairs. (C) Comparison between stem-leaf and stem-root pairs. (D) Comparison between the three shared parts. We extracted the shared genes from each of the comparisons mentioned above, and compared them. (E-G) Number of DEGs, DRGs, or URGs in response to salt stress in each tissue.
A heatmap of genes displaying tissue-differential expression without treatment.

We focused on the 19 genes that have tissue-differential expression. The gene expression values are square-root transformed FPKM (fragments per kilo-bases per million mapped reads). We used the Z-score as standardization method for each gene. We used dist (Distance Matrix Computation) function with euclidean method to compute and return the distance matrix and hclust (Hierarchical Clustering) function with complete method to perform a hierarchical cluster analysis. The colorful vertical bars on the left side denote gene cluster 1-3 orderly.
Two heatmaps of URGs (A) and DRGs (B).

The gene expression values are square-root transformed FPKM (fragments per kilo-bases per million mapped reads). We used the Z-score as standardization method for each gene. We used dist (Distance Matrix Computation) function with euclidean method to compute and return the distance matrix and hclust (Hierarchical Clustering) function with complete method to perform a hierarchical cluster analysis. The colorful vertical bars on the left side denote gene cluster 1-4 orderly. The genes, which are highlighted in blue, red, and green, belong to the cluster1-3 in Fig. 4 orderly.
**Figure 6** (on next page)

Barplots of DEGs relative expression levels based on RT-qPCR and RNA-Seq.

The expression levels of each gene were calculated as relative to corresponding gene expression in leaf with no treatment. Error bars represent SD of biologic replicates.
