RESEARCH ARTICLE   **Open Access**

Genome-wide analysis and expression profile of the bZIP gene family in poplar

Kai Zhao[1], Song Chen[1], Wenjing Yao[1,2], Zihan Cheng[1], Boru Zhou[1*] and Tingbo Jiang[1*]

Abstract

Background: The bZIP gene family, which is widely present in plants, participates in varied biological processes including growth and development and stress responses. How do the genes regulate such biological processes? Systems biology is powerful for mechanistic understanding of gene functions. However, such studies have not yet been reported in poplar.

Results: In this study, we identified 86 poplar bZIP transcription factors and described their conserved domains. According to the results of phylogenetic tree, we divided these members into 12 groups with specific gene structures and motif compositions. The corresponding genes that harbor a large number of segmental duplication events are unevenly distributed on the 17 poplar chromosomes. In addition, we further examined collinearity between these genes and the related genes from six other species. Evidence from transcriptomic data indicated that the bZIP genes in poplar displayed different expression patterns in roots, stems, and leaves. Furthermore, we identified 45 bZIP genes that respond to salt stress in the three tissues. We performed co-expression analysis on the representative genes, followed by gene set enrichment analysis. The results demonstrated that tissue differentially expressed genes, especially the co-expressing genes, are mainly involved in secondary metabolic and secondary metabolite biosynthetic processes. However, salt stress responsive genes and their co-expressing genes mainly participate in the regulation of metal ion transport, and methionine biosynthetic.

Conclusions: Using comparative genomics and systems biology approaches, we, for the first time, systematically explore the structures and functions of the bZIP gene family in poplar. It appears that the bZIP gene family plays significant roles in regulation of poplar development and growth and salt stress responses through differential gene networks or biological processes. These findings provide the foundation for genetic breeding by engineering target regulators and corresponding gene networks into poplar lines.

Keywords: Poplar bZIP gene family, Tissue-differential expression, Salt stress, Co-expression analysis

Background

The basic leucine zipper (bZIP) represents a super gene family that encodes transcription factors. This gene family is widely distributed in eukaryotes. The bZIP proteins, which are defined by the conserved bZIP domain [1, 2], play significant roles in the regulation of various biological processes, such as plant growth and development and salt stress responses.

Transcription factor proteins coded by the bZIP gene family contain a highly conserved bZIP domain. The structure is composed of 60–80 amino acids, including a basic DNA binding region and adjacent leucine zipper [2]. The binding region contains nuclear localization signals and an $N-X_7-R/K$ motif with constant precise intervals to contact target DNA [3]. The leucine zipper region is composed of heptad repeats of leucine or other large hydrophobic amino acids, and the number of repeats in

---

* Correspondence: boruzhou@yahoo.com; tbjiang@yahoo.com
[1]State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, 51 Hexing Road, Harbin 150040, China
Full list of author information is available at the end of the article

different genes may vary greatly [3, 4]. Leucine is located at the seventh amino acid position of the heptapeptide sequence and may be replaced by isoleucine, valine, phenylalanine or methionine [3]. The bZIP proteins usually function by forming dimers through the leucine zipper [4]. Plant bZIP transcription factors have a binding preference for ACGT core sequences, such as A-box (TACGTA), C-box (GACGTC), and G-box (CACGTG). In addition, they also bind to other DNA sequence motifs [3, 4]. The outer flanks of the core element regulate specificity of protein-DNA interactions [3]. Previous studies have indicated that segmental genome duplications and whole-genome duplication events may explain expansion of the bZIP gene family [2, 5]. Regarding classification, researchers initially divided the bZIP gene family members into 10 groups in Arabidopsis, based on common domains [3]. Then the Arabidopsis bZIP genes were updated and further divided into 13 groups (A-M) [4].

The bZIP gene family plays important roles in biological processes, such as growth and development, maturation of flowers, and stress responses in plants. The Arabidopsis *bZIP11* gene impacts root development by linking low-energy signals to auxin-mediated control of primary root growth [6]. *HY5* encodes a bZIP protein, which is involved in the regulation of Arabidopsis root and hypocotyl development [7]. Overexpression of *ZmbZIP4* in maize leads to an increase in the number of lateral roots, longer primary roots, and an improved root system [8]. The bZIP proteins also regulate plant responses to abiotic stresses, such as salt stress. Over-expression of *SlAREB* gene in tomato can improve plant tolerance to water deficiency and salt stress [9]. Similar results were observed for the *GhABF2* gene in Arabidopsis and cotton [10], as well as for *GmbZIP2* in transgenic soybean [11].
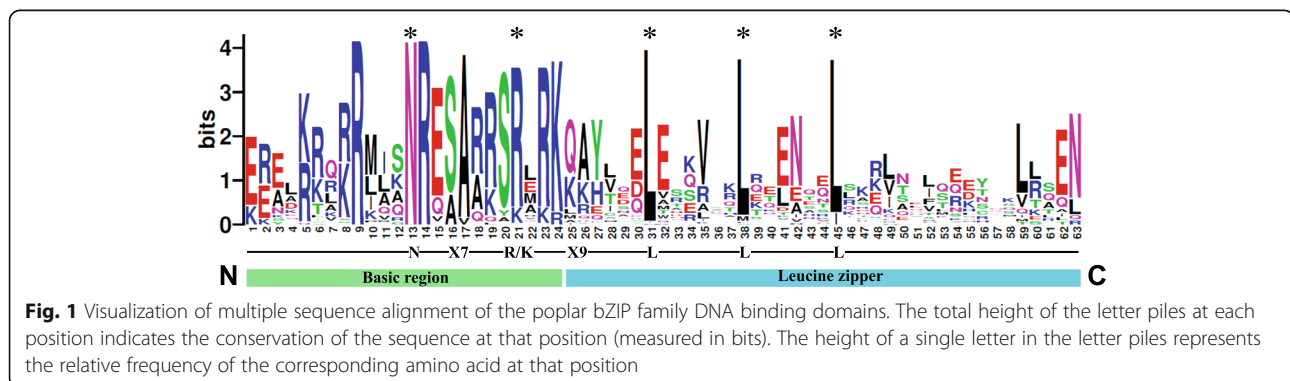
Studies on the bZIP gene family in poplar have focused only on its functions in root development and drought resistance. For example, poplar *PtabZIP1L* is mainly expressed in roots and can mediate the development of lateral roots and drought resistance by regulating various metabolic pathways [12]. Poplar *bZIP53* is inducible in gene expression by salt stress and negatively regulates the

development of adventitious roots [13]. The poplar *AREB1* can regulate drought responses and tolerance of *Populus trichocarpa* by affecting histone acetylation [14]. The comprehensive analysis and characterization of the poplar bZIP gene family and the screening of tissue differentially expressed genes (DEGs) and salt stress response genes using transcriptome sequencing can provide an important reference for gene function researches and genetic engineering breeding. In addition, using bioinformatics methods to reveal the biological processes that genes may participate in will help to analyze the regulatory mechanism of genes. In the present study, we performed systematic investigation on the bZIP genes in poplar, including identification of gene family members; protein sequence analyses and phylogenetic relationships; chromosomal distribution of the genes; genomic tandem duplications and segmental duplications; and collinearity analysis across species. Furthermore, based on transcriptome profiling data, we also explored differential expression patterns of the bZIP genes across different tissues, and their responses to salt stress. Finally, we performed gene co-expression and network analyses on the key genes, followed by gene set enrichment analyses. Our systems biology approach has shed light on differential gene networks or biological pathways associated with varied biological processes.

Results

Identification and characterization of the bZIP transcription factor family in poplar

In this study, we identified 86 proteins from the bZIP family, by use of HMMER analysis (E-value $< 1 \times 10^{-5}$). We then used Pfam and SMART databases to verify the results [15–17]. Evidence from verification by Pfam and SMART indicated that the 86 poplar proteins shared the bZIP domain, which is congruent to our predictions. Thus we extracted the amino acid sequences of the bZIP conserved domain from each member and performed multiple sequence alignment [18]. The results are visualized in Fig. 1. The bZIP domain is composed of a basic DNA-binding region and an adjacent leucine zipper structure. The basic

Fig. 1 Visualization of multiple sequence alignment of the poplar bZIP family DNA binding domains. The total height of the letter piles at each position indicates the conservation of the sequence at that position (measured in bits). The height of a single letter in the letter piles represents the relative frequency of the corresponding amino acid at that position

Zhao *et al. BMC Plant Biology*   (2021) 21:122

Page 3 of 16

region contains an invariable N-X$_7$-R/K motif, while the ZIP domain is composed of heptapeptide repeat of Leucine (L) or related hydrophobic amino acid. The highly conserved leucine residues are occasionally replaced by isoleucine, methionine, etc. (Fig. 1). Compared to the previous studies in Arabidopsis [3, 4], our results are congruent to those.

Phylogenetic tree and sequence structure analysis

To explore the evolutionary relationships and classification of the bZIP family, we constructed a phylogenetic tree, using the entire amino acid sequences of each member from both poplar and Arabidopsis. The Arabidopsis bZIP family members are divided into 13 groups using letters representing some of their important members (A for ABF/AREB/ABI5, C for CPRF2-like, G for GBF, H for HY5), protein size (B for big, S for small), or alphabetically [3, 4]. As shown in Fig. 2, we divided the poplar bZIP proteins into 12 groups, based on previous studies in *Arabidopsis thaliana*. The size of the 12 groups varies. The three largest groups have 19 (S group), 17 (A group), and 14 (D group) members, while
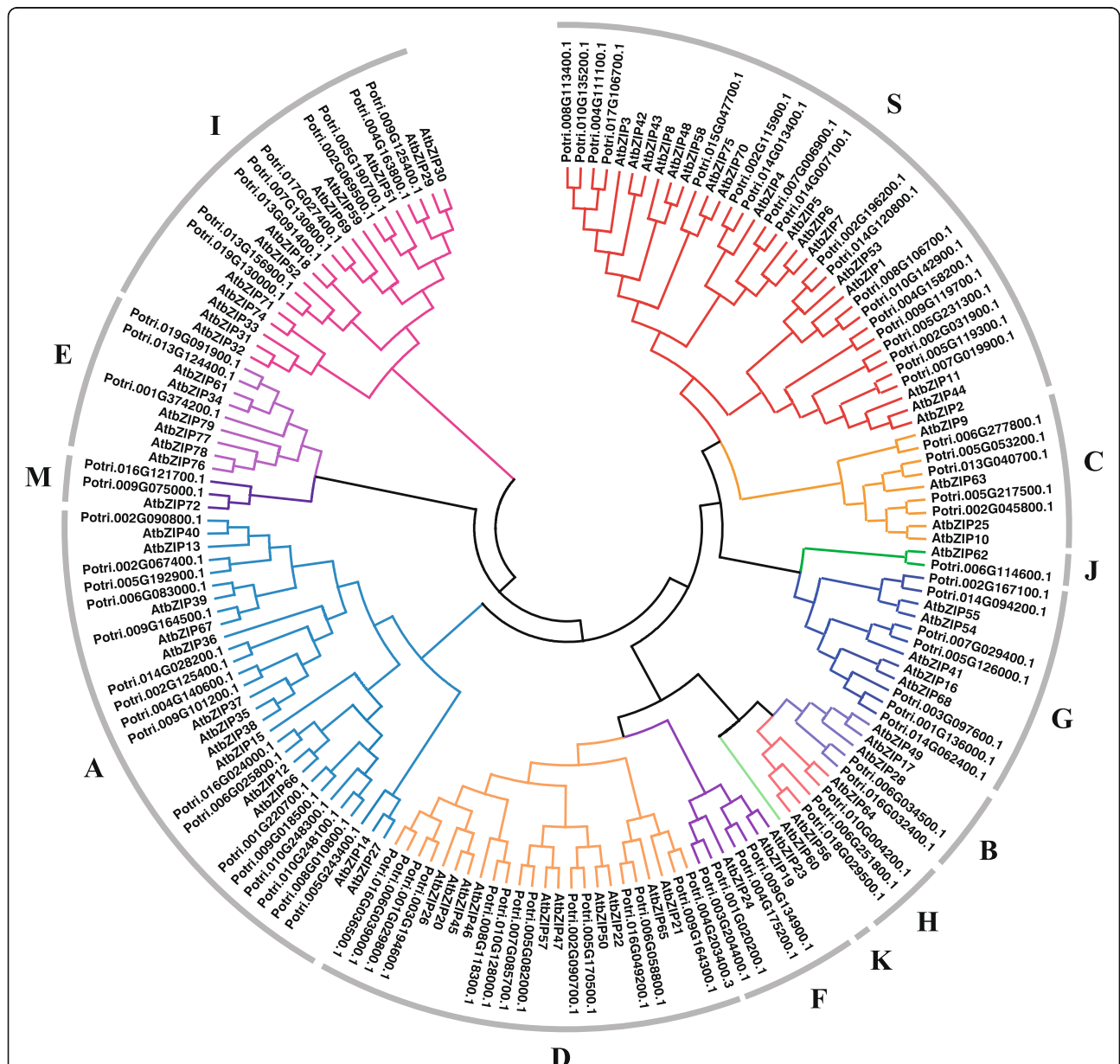

Fig. 2 Dendrogram of poplar and Arabidopsis bZIP members. The dendrogram was drew by MEGA7 with the Maximum Likelihood method and JTT + G + F model. Different groups are marked with different colors. The groups were named with letters representing some of their important members (A for ABF/AREB/ABI5, C for CPRF2-like, G for GBF, H for HY5), protein size (B for big, S for small), or alphabetically

groups J, B, M, and K have only 1, 2, 2, and 0 members, respectively.

In addition, to explore the sequence structure of the poplar bZIP family, we analyzed the intron/exon structure and motif composition of each member (Fig. 3). As

expected, members of the same group, especially some close members, share similar gene structures. For example, out of the 19 members in the S group, 18 contain only one exon. All members in group C harbor 6 exons and 5 introns. Out of the 7 members in group G, 6 have
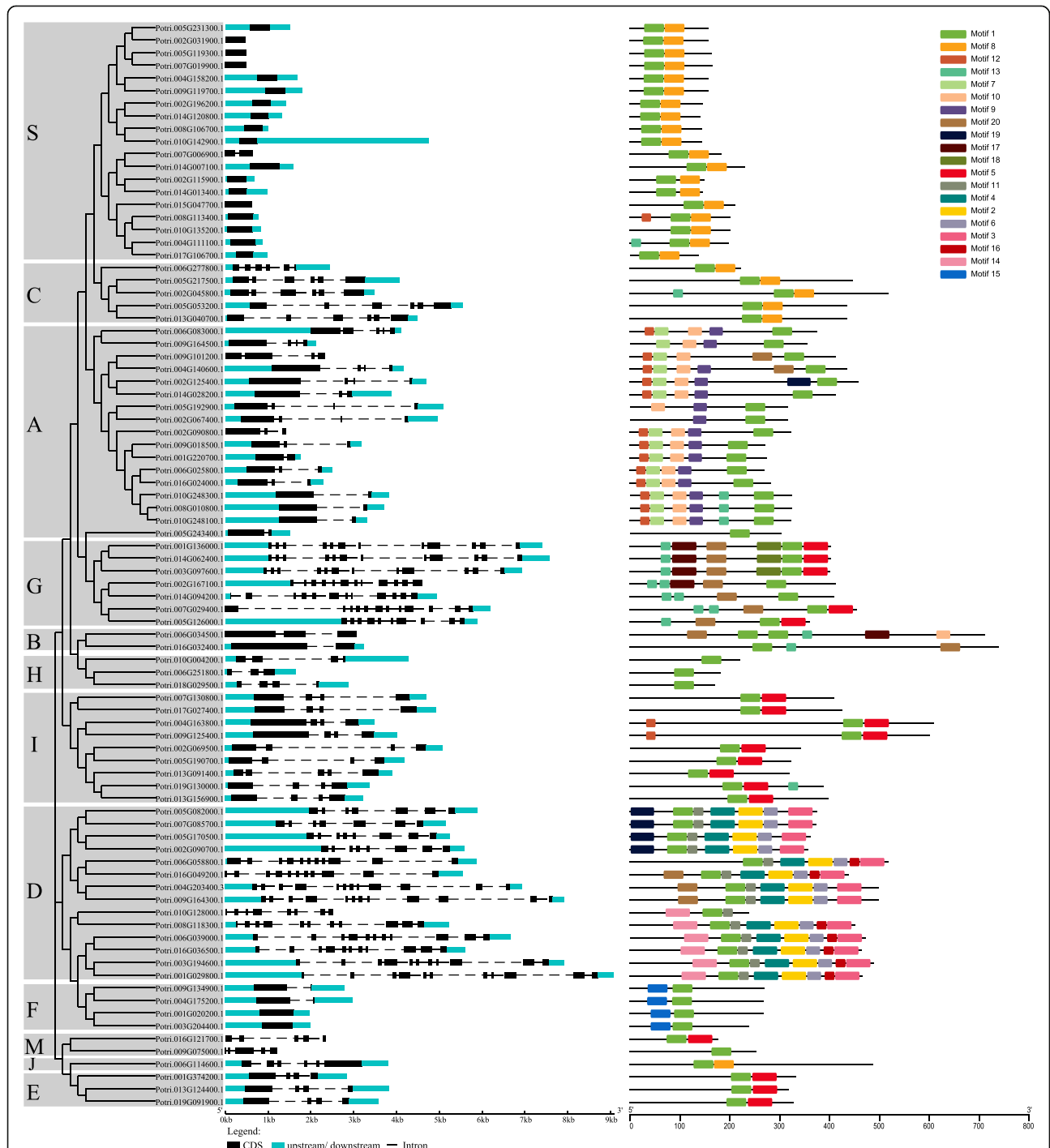
Fig. 3 DNA structures and protein motifs of the bZIP gene family in poplar. **a** Gene structures. Exons and 5′ UTR/3′ UTR are displayed using black and cyan bars. Black dotted lines denote introns. **b** Protein motifs in the bZIP members. The colorful boxes delineate different motifs. The clustering is performed according to the results of phylogenetic analysis

12 exons and 11 introns. Interestingly, many members with a closer relationship also share similar exon lengths (Fig. 3).

Using MEME [19], we found a total of 20 conserved motifs (Fig. 3). Through the annotations of SMART and Pfam databases, we found that motif 1 is a bZIP domain, motif 2 and 4 are DOG1, and other motifs have no specific annotation information (Supplemental Table 1). As expected, all of the poplar bZIP members share the motif 1. In contrast, motifs 2, 3, 4, and 6 exist only in the majority of members in group D. Similarly, motifs 7 and 9 occur only in Group A. Motif 8 presents only in groups S, C, and J. Motif 10 rests only in group A and B. Motif 11 is shared only by all members in group D. Motif 14 exists only in the six members of group D. Motif 15 exists only in all members of group F. Motif 16 occurs only in the 7 members of group D. Motif 17 exists only in the 5 members of group G and B. Motif 18 presents only in the three members of group G. Many motifs exist in specific groups, which might be related to specific biological functions.

Chromosomal location and collinearity analysis of the bZIP gene family in poplar

Using poplar genome annotation information and TBtools [20, 21], we visualized the chromosomal distribution of the bZIP gene family. Results from Fig. 4 indicated that the 86 bZIP genes are unevenly distributed on the 17 chromosomes, and the number of genes on each chromosome is irrelevant to chromosome size. For example, the largest chromosome (Chr 1) contains only 5 genes, while the smallest chromosome (Chr 9) contains 8 bZIP genes. No genes were located on chromosomes 11 or 12.

Subsequently, using TBtools with the Multiple Collinearity Scan toolkit (MCScanX) method [21, 22], we analyzed the tandem duplication events among the genes. Interestingly, no tandem duplication events were identified. Then, we used TBtools with the BLASTP and
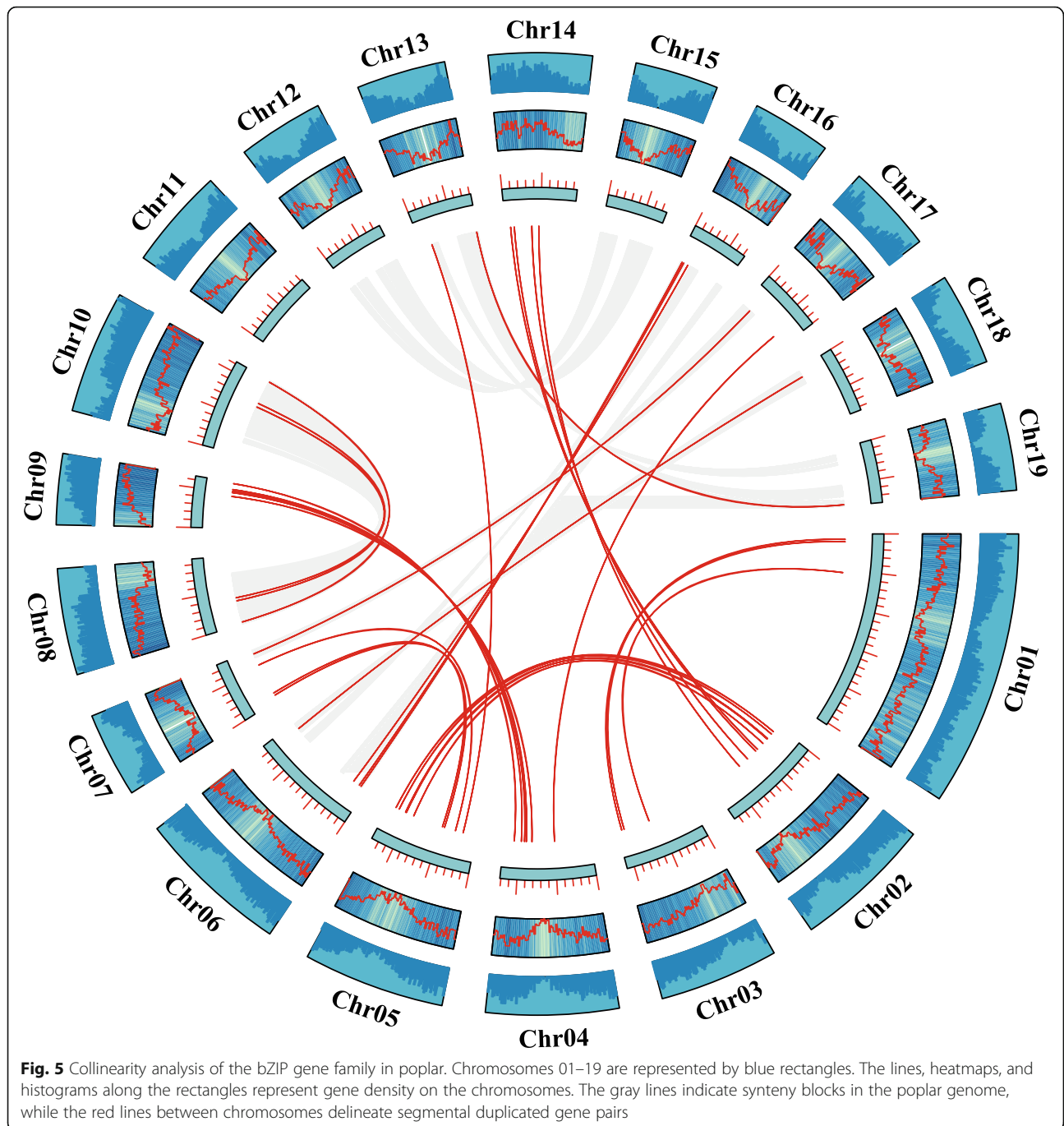
MCScanX methods to analyze the segmental duplication events [21, 22]. Our results are shown in Fig. 5 and Supplemental Table 2. We identified a total of 31 gene pairs with segmental duplication events, which occurred on 16 of the 19 chromosomes. These lines of evidence suggest that segmental duplication events are the main driving force for the diversity of the bZIP genes in poplar.

Furthermore, we also explored the collinearity relationships between the poplar bZIP genes and related genes from six representative species, including three eudicots (*Arabidopsis thaliana*, *Glycine max*, and *Solanum lycopersicum*) and three monocots (*Oryza sativa*, *Zea mays*, and *Ananas comosus*), to explore orthologs (Fig. 6, Supplemental Table 3). A total of 55 poplar genes have collinearity relationships with 11 Arabidopsis genes, 74 soybean genes, 20 tomato genes, and 5 pineapple genes. However, there is no such relationship between poplar genes and rice or maize genes (Supplemental Table 3). The number of orthologous gene pairs is 16 between poplar and Arabidopsis, 127 between poplar and soybean, 34 between poplar and tomato, and 9 between poplar and pineapple. However, no such gene pairs were identified between poplar and rice, and between poplar and maize. These may be explained by the closer phylogenetic relationships among the dicots relative to monocots.

It is worth noting that there was great collinearity between the poplar genes and the soybean genes, than those found with the other five species. This may be related to the fact that both poplar and soybean belong to fabids. In addition, we found that a large number of the poplar bZIP genes have collinearity relationships with three to four soybean genes, suggesting that these genes may play important roles in the evolution of the gene family. In addition, we also found that two poplar bZIP genes (*Potri.006G251800.1* and *Potri.018G029500.1*) are collinear with pineapple genes. However, there is no such relationship with the genes from the other five plants, suggesting that these genes have been retained in
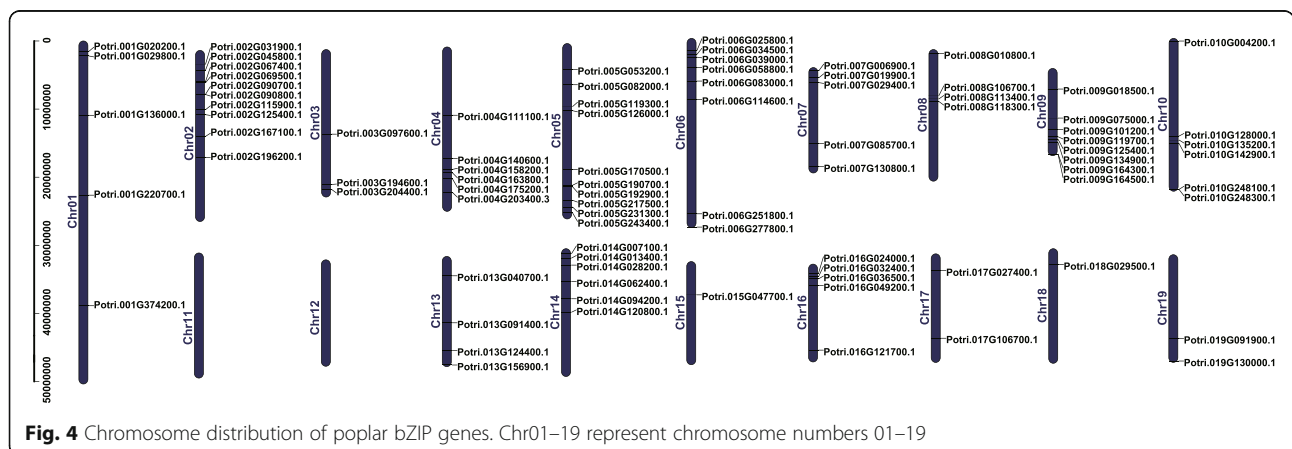
Fig. 4 Chromosome distribution of poplar bZIP genes. Chr01–19 represent chromosome numbers 01–19
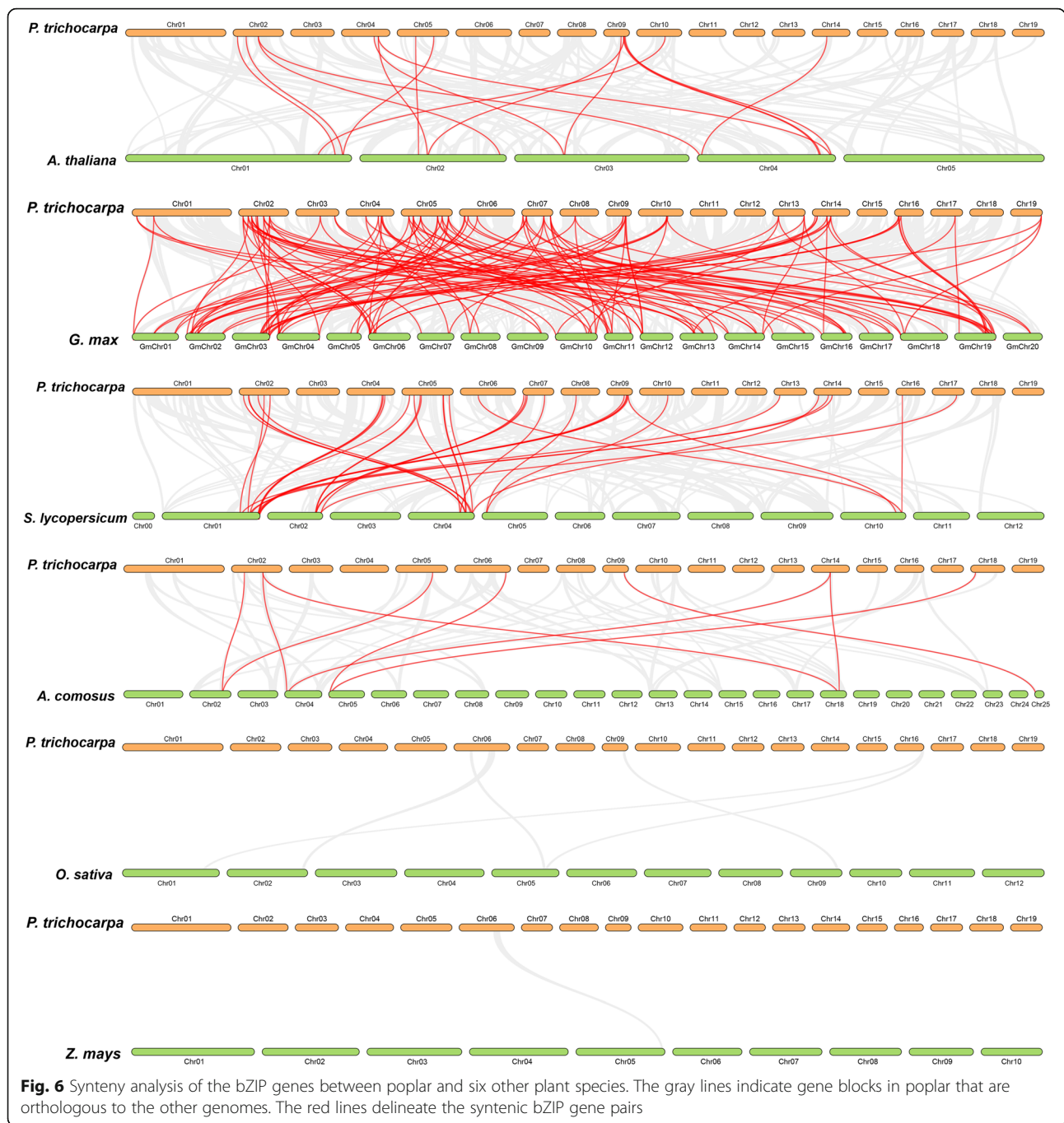
Zhao *et al. BMC Plant Biology*   (2021) 21:122

Page 6 of 16

Fig. 5 Collinearity analysis of the bZIP gene family in poplar. Chromosomes 01–19 are represented by blue rectangles. The lines, heatmaps, and histograms along the rectangles represent gene density on the chromosomes. The gray lines indicate synteny blocks in the poplar genome, while the red lines between chromosomes delineate segmental duplicated gene pairs

pineapple and poplar and have been lost in the rest of the plants analyzed.

We also used TBtools to calculate the Non-synonymous (Ka) / synonymous (Ks) ratios for each gene pair, to explore the evolutionary constraints of the poplar bZIP genes [21]. The Ka/Ks ratios of both segmental duplication and collinearity gene pairs are less than 1. This means that the poplar bZIP gene family might have experienced strong purifying selection pressure in the process of evolution.
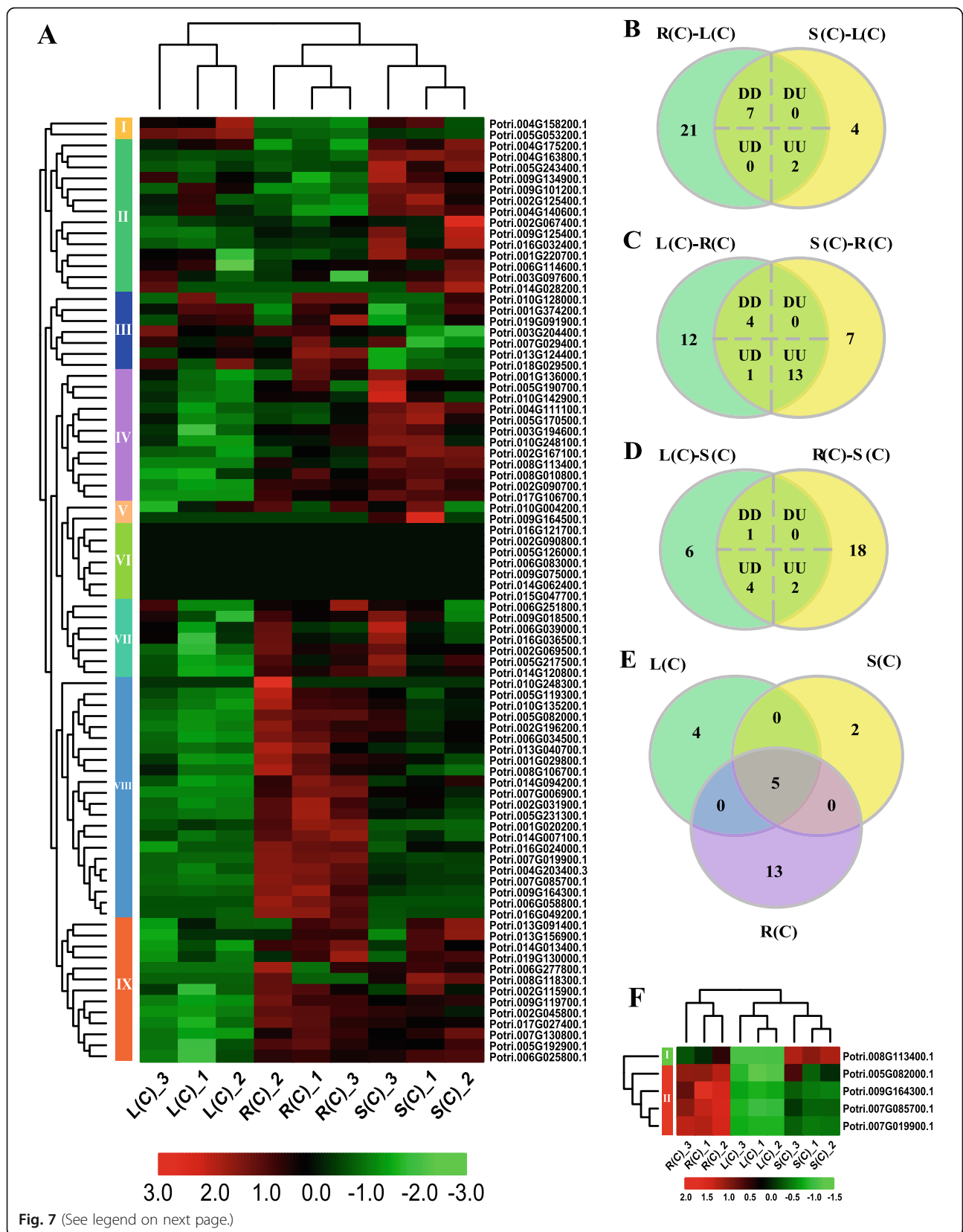
Tissue-differential gene expression of poplar bZIP genes

To explore expression patterns of the poplar bZIP gene family in different tissues, we used transcriptome profiling data from RNA-Seq to analyze the bZIP gene expression in poplar roots, stems, and leaves. As shown in Fig. 7a, we divided these genes into 9 groups with specific patterns. For example, genes in groups 2 and 4 are highly expressed in stems. In contrast, those of group 8 have higher expression levels on roots.

Fig. 6 Synteny analysis of the bZIP genes between poplar and six other plant species. The gray lines indicate gene blocks in poplar that are orthologous to the other genomes. The red lines delineate the syntenic bZIP gene pairs

We then used the DESeq method to identify DEGs across the tissues [23]. Our results indicated that there were 30 DEGs between leaves and roots (8 and 22 down- and up-regulated in roots, respectively), 13 between leaves and stems (5 and 8 down- and up-regulated in stems, respectively), and 25 between roots and stems (15 and 10 down- and up-regulated in stems, respectively) (Supplemental Data 1). Subsequently, we identified genes that are differentially expressed in one tissue relative to the other two tissues. As shown in Fig.

7b-d and Supplemental Data 1, the majority of DEGs (18) were identified in roots. Among them, 4 and 13 genes were down- and up-regulated in roots relative to the other two tissues, respectively. And there is also a gene that was up-regulated in roots relative to leaves, but was down-regulated in roots relative to stems. Similarly, we identified 9 DEGs in the leaves. Among them, respective 7 and 2 genes were down- and up-regulated in leaves compared to stems and roots. Seven DEGs were identified in the stems. Among them, one gene was

Fig. 7 (See legend on next page.)

down-regulated and 2 genes were up-regulated in stems compared to leaves and roots. Furthermore, four genes were up-regulated in stems relative to leaves, but were down-regulated in stems relative to roots. Finally, we compared the three sets of genes identified above, and obtained 5 shared genes, which were differentially expressed in any tissue relative to the other two tissues (Fig. 7e). We then drew a heatmap, based on the expression data of these 5 genes (Fig. 7f). It appears clear tissue-specific expression patterns; that is, one gene showed high, moderate, or low expressed in stems, roots, and leaves, respectively. However, the rest of the genes displayed a different pattern (Fig. 7f).

Gene expression in response to salt stress in poplar

To explore expression patterns of the bZIP gene family in response to salt stress, we analyzed expression changes of the responsive genes before and after salt stress, based on the RNA-Seq data. As shown in Fig. 8a-c and Supplemental Data 2, we identified the majority of DEGs in roots (16 and 15 down- and up-regulated, respectively), followed by 19 genes in leaves (8 and 11), and by 10 genes in stems (4 and 6). We then compared the three groups of genes and drew Venn diagrams (Fig. 8d-f). As shown in Fig. 8d, the majority of genes (22) responsive to salt stress were found to be specific to roots, followed by 8 genes to leaves, and by 2 genes to stems. In addition, 5 genes were responsive to the salt stress in both leaves and roots (leaf-root), 4 genes in leaf-stem, and 2 genes in root-stem. Among them, 2 genes overlap between the three groups of DEGs (Fig. 8d). Comparisons of down- and up-regulated DEGs are shown in Fig. 8e, f.

Validation of DEGs by qRT-PCR

To verify the results of RNA-Seq, we used qRT-PCR to quantify expression levels of the 23 DEGs in roots, stems and leaves, before and after the salt stress. As shown in Fig. 9, results from RNA-Seq and qRT-PCR are congruent.
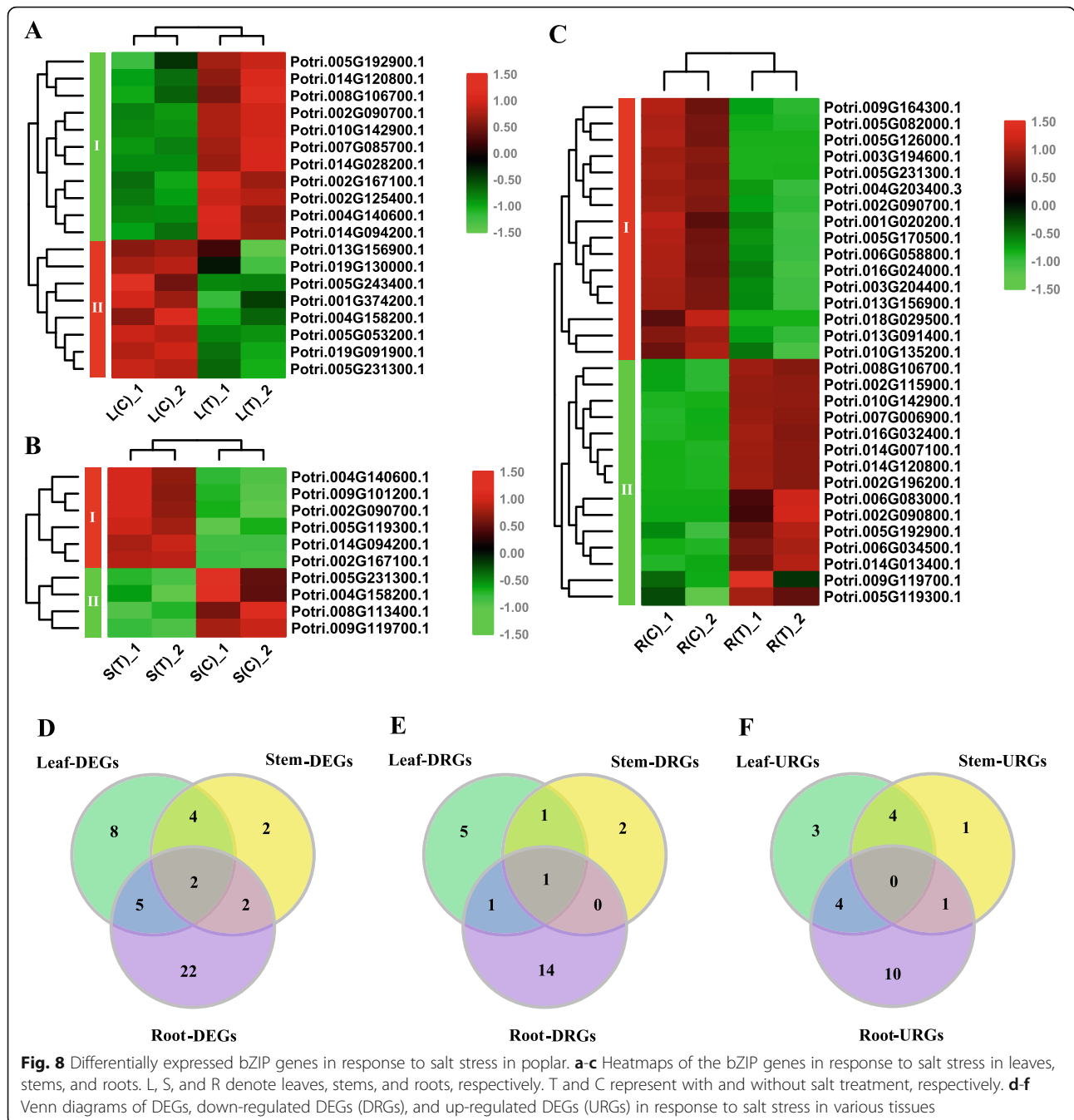
Gene co-expression analysis

Co-expression analysis can help find genes with similar expression patterns. These genes may be closely co-regulated, tightly related in function, or members involved in the same signaling pathway or physiological process. In this study, we used the Weighted Correlation

Network Analysis (WGCNA) method and the RNA-Seq data of 21 samples to construct a co-expression network centered on the 5 tissue-differential expressed genes and the 2 salt responsive genes described above [24]. As shown in Fig. 10 and Supplemental Data 3, we obtained a total of 7 co-expression networks. Among them, the network centered on *Potri.005G231300.1* is the largest (855 genes). In contrast, the network centered on *Potri.002G090700.1* is the smallest (27 genes).

To explore the biological processes that these genes may participate in, we then conducted gene set enrichment analysis on the 7 sets of co-expressed genes identified above (Fig. 11). Four of the five tissue-differential expressed genes (*Potri.005G082000.1, Potri.007G085700.1, Potri.007G019900.1, Potri.009G164300.1*), share 17 significantly enriched GO Terms. The shared GO Terms include secondary metabolic process, secondary metabolite biosynthetic process, phenylpropanoid metabolic process, and phenylpropanoid biosynthetic process. This suggests that the four genes may play important roles in regulation of poplar growth and development and stress responses. Interestingly, the four genes have the same expression pattern across the tissues (Fig. 7f).

Other genes involved in tissue differential expression and the co-expressed gene network are significantly enriched in cell wall biogenesis, cell wall macromolecule biosynthetic process, xylan metabolic process, xylan biosynthetic process, hemicellulose metabolic process, and other biological processes, suggesting that they may be related to formation of cell walls.

Regarding the genes that respond to salt stress in the root, stem and leaf tissues, *Potri.002G090700.1* and its co-expressed gene network are significantly enriched in GO terms, such as regulation of potassium ion transport, hyperosmotic salinity response, regulation of metal ion transport, and hyperosmotic response. It is suspected that these genes may respond to salt or osmotic stresses by regulation of ion balance. *Potri.005G231300.1* and the genes in its network are significantly enriched in GO terms, such as methionine biosynthetic process and methionine metabolic process. The S-Adenosyl-L-methionine synthetase (SAMS) in plants can catalyze the reaction of methionine and ATP to produce S-Adenosyl-L-methionine, which is a key enzyme that regulates the methionine cycle. Previous studies have shown that *SAMS* is a key gene for organisms to resist adversity

Fig. 8 Differentially expressed bZIP genes in response to salt stress in poplar. **a**-**c** Heatmaps of the bZIP genes in response to salt stress in leaves, stems, and roots. L, S, and R denote leaves, stems, and roots, respectively. T and C represent with and without salt treatment, respectively. **d**-**f** Venn diagrams of DEGs, down-regulated DEGs (DRGs), and up-regulated DEGs (URGs) in response to salt stress in various tissues

and stress [25]. Therefore, we speculate that these genes might be involved in the processes of methionine synthesis and metabolism, which are related to stress resistance.

Discussion

The bZIP transcription factors are found in the plant kingdom, which play important roles in regulating growth and development and responses to biotic and abiotic stresses [4, 26]. Previous studies on poplar were
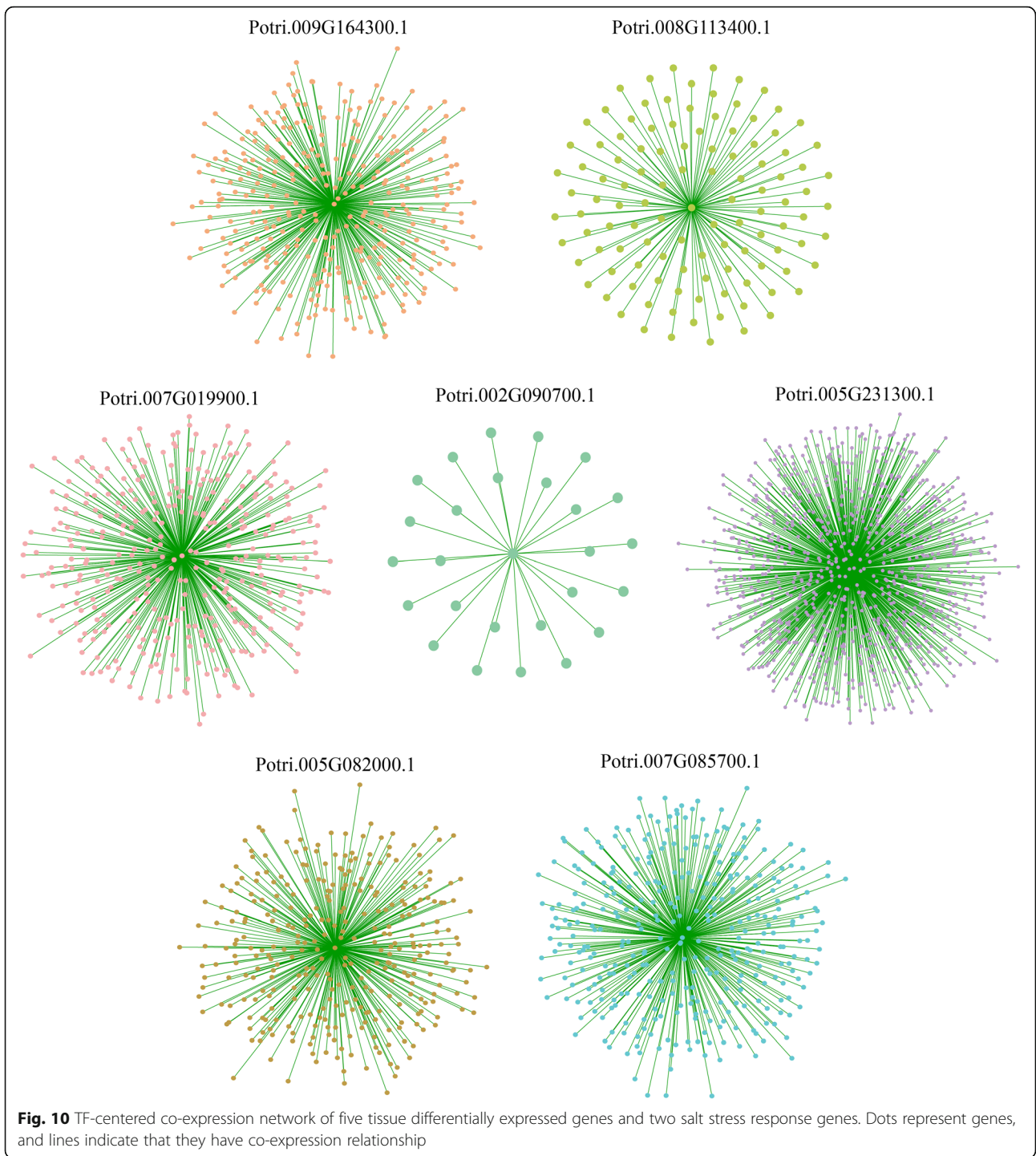
constrained on only a few bZIP genes, which regulate root development and drought stress. Thus systematic studies on the poplar bZIP gene family have not been reported. In this study, we used strict standards to identify 86 bZIP genes from poplar. Then we extracted the bZIP protein domain sequences of the members. Evidence from multiple sequence alignment indicated that both poplar and Arabidopsis share the same bZIP domain. Compared to Arabidopsis [3, 4], the poplar bZIP domains also consist of a DNA binding region containing

Fig. 9 Gene expression levels based on RNA-Seq and qRT-PCR. L, S, and R denote leaves, stems, and roots, respectively. T and C represent with and without salt treatment, respectively. Error bars are standard deviations from the biologic replicates

the N-X$_7$-R/K motif and a leucine zipper structure. We then divided these 86 members into 12 groups, based on similarity of their protein sequences. It is worth noting that many motifs exist in specific groups. Given that the bZIP transcription factors have varied functions, these

motifs might perform specific functions. This phenomenon deserves further studies.

Previous studies indicated that in the evolutionary process gene families usually undergo tandem duplication or large-scale segmental duplication, to maintain a

Zhao *et al. BMC Plant Biology*   (2021) 21:122

Page 12 of 16

Fig. 10 TF-centered co-expression network of five tissue differentially expressed genes and two salt stress response genes. Dots represent genes, and lines indicate that they have co-expression relationship

large size of each family [27]. Unlike birch, which has not undergo recent whole genome duplication [28], poplar has undergone at least three rounds of whole-genome duplication, followed by multiple segmental duplication, tandem duplication and transposition events [20, 29]. Since the bZIP gene family is a relatively large one, thus we analyzed both the tandem duplication events and the segmental duplication events. Interesting,

we found that there were no tandem duplication events in the poplar bZIP genes. However, a large number of segmental duplication events occurred, which is consistent with previous studies in rice [2]. These results indicate that segmental duplication events play an important role in the expansion of the bZIP gene family. Furthermore, we analyzed the collinearity between the poplar bZIP genes and the counterparts from three eudicots
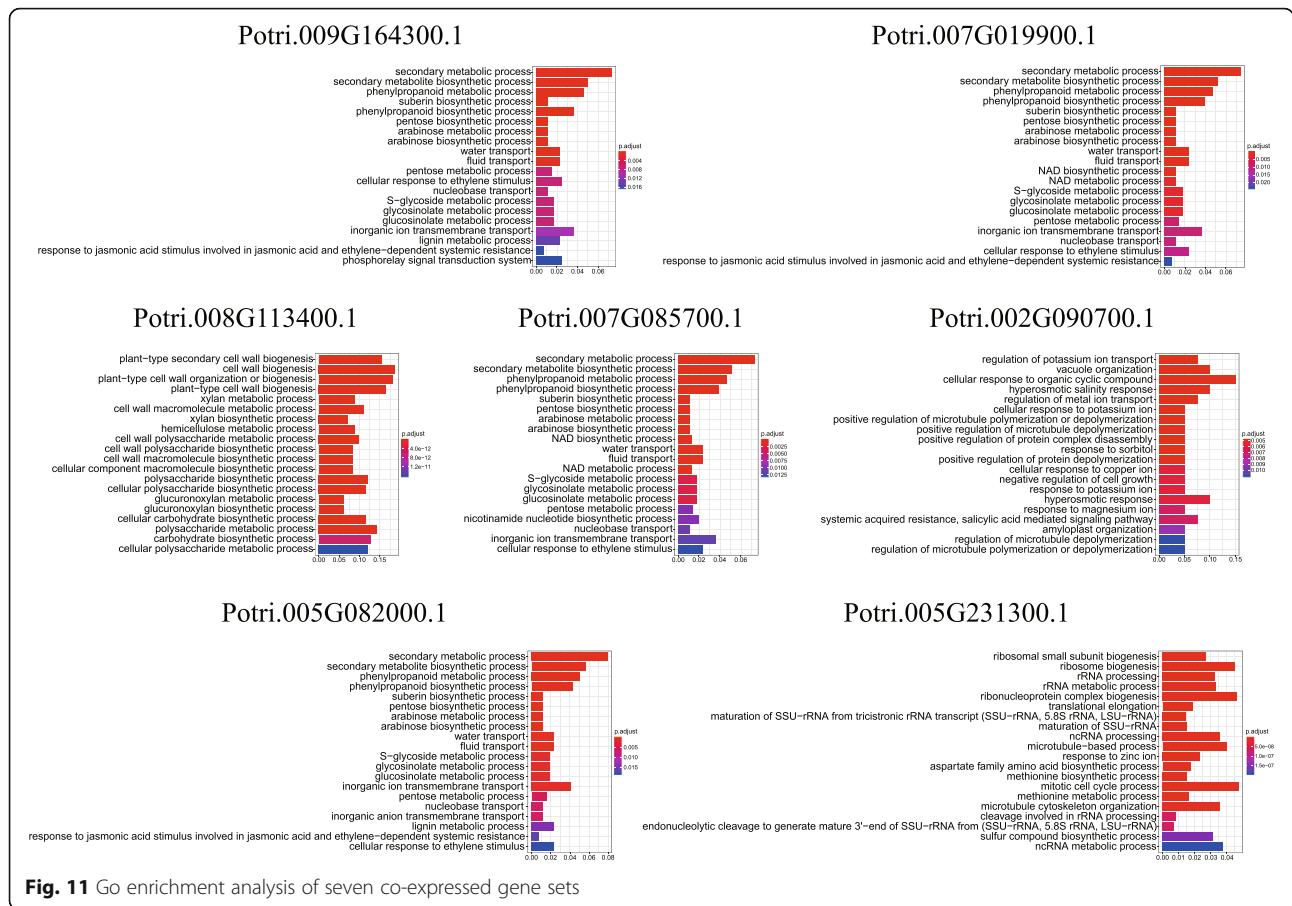
Fig. 11 Go enrichment analysis of seven co-expressed gene sets

and three monocots. The results demonstrated that there were significantly more collinearity gene pairs between poplar and eudicots than between poplar and monocots. Species with relatively close evolutionary relationships, such as poplar and soybean, appeared to have more collinear gene pairs. We also identified many poplar bZIP genes that existed either in multiple such gene pairs or only in the collinearity with monocots. The calculated values of the Ka/Ks ratios for all gene pairs were less than 1, suggesting that these genes might have experienced strong purifying selection pressure in evolution.

We explored expression patterns of the bZIP genes in a tissue-specific fashion. We have identified many DEGs between tissues, of which five genes are differentially expressed in any two of the three tissues. These genes showed low expression in leaves and high expression in roots or stems. We then mapped these genes to the Arabidopsis genome to understand their possible functions (Supplemental Table 4) [30, 31]. The best match for *Potri.005G082000.1* and *Potri.007G085700.1* is *AT5G65210.1*, which is an essential cofactor in BOP-dependent regulation of development in Arabidopsis [29]. *AT1G08320.1* is the counterpart for *Potri.009G164300.1*, which is involved

in the Arabidopsis anther development [32]. *AT1G75390.1* is the best match for *Potri.007G019900.1*, which affects the germination process of seeds [33]. Taken together, these tissue-specific DEGs are likely to play important roles in poplar growth and development.

Since little is known in poplar about the functions of the bZIP genes in regulation of salt responses, we explored their expression patterns in response to salt stress. Comparing the expression data from samples both before and after salt stress, we identified a total of 45 genes that respond to salt stress in roots, stems, and leaves. We also mapped them to the Arabidopsis genome (Supplemental Table 4) [30, 31]. The homologous genes are involved in growth and development of Arabidopsis, as well as in responses to abiotic stresses. For example, *AT2G40950.1*, a homologous gene of *Potri.006G034500.1* and *Potri.016G032400.1*, is a transcription factor regulating cellular responses to salinity through osmotic stress adjustment [34]. The best matches for *Potri.001G020200.1* (*AT4G35040.1*) and for *Potri.003G204400.1* (*AT2G16770.1*) were reported to regulate adaptation of Arabidopsis to zinc deficiency [35]. *AT1G45249.1*, the counterpart for *Potri.009G101200.1*, *Potri.014G028200.1*, and *Potri.002G125400.1*, were able to

regulate ABRE-dependent ABA signaling pathway related to drought stress tolerance [36].

To understand their function, we performed co-expression based gene network analysis, focusing on the seven key genes we identified, followed by gene set enrichment analysis. Our results from gene ontology analyses indicated that the DEGs across tissues and their corresponding network genes are enriched in secondary metabolic process, and secondary metabolite biosynthetic process. However, the salt-inducible genes and their corresponding network genes are enriched in regulation of metal ion transport and methionine biosynthetic processes. These suggest that different regulators and regulated gene networks play important roles in specific biological functions.

Conclusions
In this study, we identified 86 bZIP gene family members in poplar and characterized their conserved bZIP domains. We then conducted systematic analyses of the gene family. On the basis of the phylogenetic analyses, these members can be divided into 12 groups and each group has specific gene structure and motif composition. The bZIP genes are unevenly distributed on the 17 chromosomes of poplar. It is worth noting that no tandem duplication events were identified between poplar bZIP genes, however, we detected a large number of segmental duplication events, which suggest that segmental duplication events are the main driving force for the evolution of the bZIP gene family in poplar. In addition, we investigated the collinearity relationship between the poplar bZIP genes and the homologous genes from six representative species. These will benefit future comparative gene functions studies. Furthermore, we identified tissue-specific gene expression patterns, as well as salt-inducible gene expression patterns. On the basis of the seven key regulators and their corresponding gene networks, differential regulators and corresponding gene networks have been found to associate with varied biological processes.

Methods
Identification of poplar bZIP transcription factors and their conserved domains
The amino acid sequences of all poplar proteins were downloaded from the Phytozome database [20, 30], and that of bZIP_1 (PF00170) from the Pfam database [17]. First, we used hmmsearch (http://www.hmmer.org/) with bZIP_1 to search the poplar amino acid sequences, with a threshold of E-value $< 1 \times 10^{-5}$. After obtaining the candidate genes, we applied Pfam and SMART databases to further verify our results [15–17]. Then, we extracted sequences of the conserved domains from the poplar bZIP proteins identified. Finally, we used

ClustalX 1.83 and WebLogo for multiple sequence alignment and visualization, respectively [18, 37].

Classification and sequence analysis on the bZIP members
The amino acid sequences of both poplar and Arabidopsis family members were downloaded from the Phytozome database [20, 30, 31]. Members of the Arabidopsis bZIP gene family were identified from previous studies [4].

We used the MEGA 7 with Maximum Likelihood method and the best model selected by SMS to construct a phylogenetic tree [38, 39]. Classification of the poplar bZIP protein family refers to previous studies in Arabidopsis [4].

To analyze structure of the bZIP gene family, we downloaded the genome sequences and coding sequences of the genes from the Phytozome database [20, 30]. We then used the Gene Structure Display Server to draw gene structure diagrams [40]. The order and grouping of genes in the gene structure diagrams refer to the results of the phylogenetic tree. Using MEME, we identified conserved motifs contained in the bZIP gene family [19]. Finally, we imported the generated file into TBtools for visualization [21]. Motif annotation information comes from the SMART and Pfam databases [15–17].

Chromosome distribution and collinearity analysis
Using the TBtools and the poplar genome data downloaded from the Phytozome database [20, 21, 30], we visualized the chromosomal distribution of the bZIP genes in poplar. In addition, using the TBtools with MCScanX, we analyzed the tandem duplication events of the bZIP gene family [21, 22]. Similarly, using the TBtools with MCScanX and BLASTP methods, we investigated segmental duplication events and the collinearity relationship for gene pairs from different species [21, 22]. Ka and Ks substitutions between gene pairs were also calculated, by use of the TBtools [21].

Plant materials and gene expression analysis
The plant materials used in this study were di-haploid *Populus simonii* × *Populus nigra* seedlings from a wild-type clone growing in the experimental forest of Northeast Forestry University. Using the RNA-Seq data described in our previous study [41], we explored tissue differential expression patterns of the bZIP genes. To investigate expression patterns of the genes in response to salt stress, we used the RNA-Seq data collected from root, stem, and leaf tissues under treatment with 0 or 150 mM NaCl for 24 h [42]. All of the samples have 10X sequencing depth. Using the DESeq package in R, we identify DEGs (fold change $>= 2$ and padj $<= 0.05$) [23].

Zhao *et al. BMC Plant Biology*   (2021) 21:122

Page 15 of 16

Using qRT-PCR to validate differentially expressed genes

To verify the gene expression data from RNA-Seq, we used qRT-PCR to quantify expression levels of the DEGs in response to salt stress. The qRT-PCR were performed according to the published studies [43, 44]. We use actin as a reference gene [45]. The relative expression level was calculated, based on the expression level of each gene without salt treatment. The primers are listed in Supplemental Table 5.

Gene co-expression networks and gene ontology analyses

Using the WGCNA package in R, we identified gene co-expression based gene networks [24]. Cytoscape was then used to visualize the results [46]. The 21 RNA-Seq data mentioned above were used for co-expression analysis. Only genes with sum of expression levels are greater than 10 across the samples are used for the analysis. Regarding the parameters used, the best β (soft threshold power) value was set to 18 (after 20 iterations), the "deepSplit" value to 2, the "minModuleSize" value to 30, and the "mergeCutHeight" value to 0.15. The correlation coefficient is calculated using Pearson algorithm. The top 20% genes with the strongest weights were used to draw the co-expression networks.

Gene set enrichment analysis was performed using the clusterprofiler package in R [47]. We focused on biological processes. The *P*-value of each pathway was calculated and adjusted using the Benjamin-Hochberg method [48].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-021-02879-w.

Additional file 1: Supplemental Table 1. Annotations of bZIP protein sequence motifs.

Additional file 2: Supplemental Table 2. Segmentally duplicated bZIP gene pairs in poplar.

Additional file 3: Supplemental Table 3. List of syntenic gene pairs.

Additional file 4: Supplemental Table 4. Annotations of DEGs.

Additional file 5: Supplemental Table 5. Primer sequences.

Additional file 6: Supplemental Data 1. Pairwise gene expressions. C denotes without salt treatment. The DEGs were identified by use of the DESeq software and the thresholds were set at both fold change > = 2 and padj <= 0.05.

Additional file 7: Supplemental Data 2. DEGs in response to salt stress. L, S, and R denote leaves, stems, and roots, respectively. T and C represent with and without salt treatment, respectively. DEGs were identified by use of the DESeq software and the thresholds were set at both fold change > = 2 and padj <= 0.05.

Additional file 8: Supplemental Data 3. Co-expression gene sets.

Abbreviations
DEGs: Differentially expressed genes; SAMS: S-Adenosyl-L-methionine synthetase; MCScanX: Multiple Collinearity Scan toolkit; Ka: Non-synonymous; Ks: Synonymous; WGCNA: Weighted Correlation Network Analysis

Author details
[1]State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, 51 Hexing Road, Harbin 150040, China. [2]Co-Innovation Center for Sustainable Forestry in Southern China/Bamboo Research Institute, Nanjing Forestry University, 159 Longpan Road, Nanjing 210037, China.

References
1. Landschulz WH, Johnson PF, McKnight SL. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science. 1988;240(4860):1759–64.
2. Nijhawan A, Jain M, Tyagi AK, Khurana JP. Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. Plant Physiol. 2008;146(2):333–50.
3. Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, Parcy F. bZIP transcription factors in Arabidopsis. Trends Plant Sci. 2002;7(3):106–11.
4. Dröge-Laser W, Snoek BL, Snel B, Weiste C. The Arabidopsis bZIP transcription factor family-an update. Curr Opin Plant Biol. 2018;45(Pt A):36–49.
5. Zhou Y, Xu D, Jia L, Huang X, Ma G, Wang S, Zhu M, Zhang A, Guan M, Lu K. Genome-wide identification and structural analysis of bZIP transcription factor genes in Brassica napus. Genes. 2017;8(10):288.
6. Weiste C, Pedrotti L, Selvanayagam J, Muralidhara P, Fröschel C, Novák O, Ljung K, Hanson J, Dröge-Laser W. The Arabidopsis bZIP11 transcription factor links low-energy signalling to auxin-mediated control of primary root growth. PLoS Genet. 2017;13(2):e1006607.
7. Oyama T, Shimura Y, Okada K. The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus-induced development of root and hypocotyl. Genes Dev. 1997;11(22):2983–95.
8. Ma H, Liu C, Li Z, Ran Q, Xie G, Wang B, Fang S, Chu J, Zhang J. ZmbZIP4 contributes to stress resistance in maize by regulating ABA synthesis and root development. Plant Physiol. 2018;178(2):753–70.

9. Hsieh T-H, Li C-W, Su R-C, Cheng C-P, Tsai Y-C, Chan M-T. A tomato bZIP transcription factor, SlAREB, is involved in water deficit and salt stress response. Planta. 2010;231(6):1459–73.

10. Liang C, Meng Z, Meng Z, Malik W, Yan R, Lwin KM, Lin F, Wang Y, Sun G, Zhou T. GhABF2, a bZIP transcription factor, confers drought and salinity tolerance in cotton (Gossypium hirsutum L.). Sci Rep. 2016;6(1):1–14.

11. Yang Y, Yu T-F, Ma J, Chen J, Zhou Y-B, Chen M, Ma Y-Z, Wei W-L, Xu Z-S. The soybean bZIP transcription factor gene GmbZIP2 confers drought and salt resistances in transgenic plants. Int J Mol Sci. 2020;21(2):670.

12. Dash M, Yordanov YS, Georgieva T, Tschaplinski TJ, Yordanova E, Busov V. Poplar PtabZIP 1-like enhances lateral root formation and biomass growth under drought stress. Plant J. 2017;89(4):692–705.

13. Zhang Y, Yang X, Cao P, ZA X, Zhan C, Liu M, Nvsvrot T, Wang N. The bZIP53–IAA4 module inhibits adventitious root development in Populus. J Exp Bot. 2020;71(12):3485–98.

14. Li S, Lin Y-CJ, Wang P, Zhang B, Li M, Chen S, Shi R, Tunlaya-Anukit S, Liu X, Wang Z. The AREB1 transcription factor influences histone acetylation to regulate drought responses and tolerance in Populus trichocarpa. Plant Cell. 2019;31(3):663–86.

15. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. Nucleic Acids Res. 2015;43(D1):D257–60.

16. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46(D1):D493–6.

17. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47(D1):D427–d432.

18. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997;25(24):4876–82.

19. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(suppl_2):W202–8.

20. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. 2006;313(5793):1596–604.

21. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13(8):1194–202.

22. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.

23. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.

24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

25. Gong B, Li X, VandenLangenberg KM, Wen D, Sun S, Wei M, Li Y, Yang F, Shi Q, Wang X. Overexpression of S-adenosyl-l-methionine synthetase increased tomato tolerance to alkali stress through polyamine metabolism. Plant Biotechnol J. 2014;12(6):694–708.

26. Ritonga FN, Chen S. Physiological and Molecular Mechanism Involved in Cold Stress Tolerance in Plants. Plants (Basel). 2020;9(5):560.

27. Cannon SB, Mitra A, Baumgarten A, Young ND, May G. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. BMC Plant Biol. 2004;4(1):10.

28. Gang H, Li R, Zhao Y, Liu G, Chen S, Jiang J. Loss of GLK1 transcription factor function reveals new insights in chlorophyll biosynthesis and chloroplast development. J Exp Bot. 2019;70(12):3125–38.

29. Wang Y, Salasini BC, Khan M, Devi B, Bush M, Subramaniam R, Hepworth SR. Clade I TGACG-motif binding basic leucine zipper transcription factors mediate BLADE-ON-PETIOLE-dependent regulation of development. Plant Physiol. 2019;180(2):937–51.

30. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012;40(D1):D1178–86.

31. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M. The Arabidopsis information resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 2012;40(D1):D1202–10.

32. Murmu J, Bush MJ, DeLong C, Li S, Xu M, Khan M, Malcolmson C, Fobert PR, Zachgo S, Hepworth SR. Arabidopsis basic leucine-zipper transcription factors TGA9 and TGA10 interact with floral glutaredoxins ROXY1 and ROXY2 and are redundantly required for anther development. Plant Physiol. 2010;154(3):1492–504.

33. Iglesias-Fernández R, Barrero-Sicilia C, Carrillo-Barral N, Oñate-Sánchez L, Carbonero P. Arabidopsis thaliana bZIP 44: a transcription factor affecting seed germination and expression of the mannanase-encoding gene AtMAN7. Plant J. 2013;74(5):767–80.

34. Cifuentes-Esquivel N, Celiz-Balboa J, Henriquez-Valencia C, Mitina I, Arraño-Salinas P, Moreno AA, Meneses C, Blanco-Herrera F. bZIP17 regulates the expression of genes related to seed storage and germination, reducing seed susceptibility to osmotic stress. J Cell Biochem. 2018;119(8):6857–68.

35. Assunção AG, Herrero E, Lin YF, Huettel B, Talukdar S, Smaczniak C, Immink RG, van Eldik M, Fiers M, Schat H, et al. Arabidopsis thaliana transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. Proc Natl Acad Sci U S A. 2010;107(22):10296–301.

36. Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. Plant J. 2010;61(4):672–85.

37. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14(6):1188–90.

38. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.

39. Lefort V, Longueville J-E, Gascuel O. SMS: Smart model selection in PhyML. Mol Biol Evol. 2017;34(9):2422–4.

40. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31(8):1296–7.

41. Yao W, Li C, Lin S, Wang J, Zhou B, Jiang T. Transcriptome analysis of salt-responsive and wood-associated NACs in Populus simonii×Populus nigra. BMC Plant Biol. 2020;20(1):1–12.

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

43. Wang S, Huang H, Han R, Chen J, Jiang J, Li H, Liu G, Chen S. BpAP1 directly regulates BpDEF to promote male inflorescence formation in Betula platyphylla × B. pendula. Tree Physiol. 2019;39(6):1046–60.

44. Wang S, Huang H, Han R, Liu C, Qiu Z, Liu G, Chen S, Jiang J. Negative feedback loop between BpAP1 and BpPI/BpDEF heterodimer in *Betula platyphylla × B. pendula*. Plant Sci. 2019;289:110280.

45. Regier N, Frey B. Experimental comparison of relative RT-qPCR quantification approaches for gene expression studies in poplar. BMC Mol Biol. 2010;11:57.

46. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.

47. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.

48. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300.

Publisher's Note