

RESEARCH

Open Access



Genome-wide identification and expression analysis of the bZIP transcription factor family genes in response to abiotic stress in *Nicotiana tabacum* L.

Lili Duan^{1,2,3}, Zejun Mo^{1,2}, Yue Fan⁴, Kuiyin Li¹, Mingfang Yang¹, Dongcheng Li^{2,3}, Yuzhou Ke^{2,3}, Qian Zhang^{2,3}, Feiyan Wang^{2,3}, Yu Fan^{5*} and Renxiang Liu^{2,3*}

Abstract

Background: The basic leucine zipper (bZIP) transcription factor (TF) is one of the largest families of transcription factors (TFs). It is widely distributed and highly conserved in animals, plants, and microorganisms. Previous studies have shown that the bZIP TF family is involved in plant growth, development, and stress responses. The bZIP family has been studied in many plants; however, there is little research on the *bZIP* gene family in tobacco.

Results: In this study, 77 *bZIPs* were identified in tobacco and named *NtbZIP01* through to *NtbZIP77*. These 77 genes were then divided into eleven subfamilies according to their homology with *Arabidopsis thaliana*. *NtbZIPs* were unevenly distributed across twenty-two tobacco chromosomes, and we found sixteen pairs of segmental duplication. We further studied the collinearity between these genes and related genes of six other species. Quantitative real-time polymerase chain reaction analysis identified that expression patterns of *bZIPs* differed, including in different organs and under various abiotic stresses. *NtbZIP49* might be important in the development of flowers and fruits; *NtbZIP18* might be an important regulator in abiotic stress.

Conclusions: In this study, the structures and functions of the *bZIP* family in tobacco were systematically explored. Many *bZIPs* may play vital roles in the regulation of organ development, growth, and responses to abiotic stresses. This research has great significance for the functional characterisation of the tobacco bZIP family and our understanding of the bZIP family in higher plants.

Keywords: bZIP family, *Nicotiana tabacum* L., Gene expression patterns, Abiotic stress

Background

Transcription factors (TFs) are proteins that combine with specific short sequences of DNA to regulate the expression of genes [1, 2]. Important regulatory factors,

TFs bind to the promoter regions of target genes and activate or inhibit their transcription. Among the TF families, the basic leucine zipper (bZIP) TF family is the most widely distributed and conserved in the domain of eukaryotes, found in humans, plants, animals, and microorganisms [3, 4].

The bZIP TF family is named after the bZIP domain [5]. The bZIP protein sequence contains 2 conserved bZIP domains with a total length of ~60–80 amino acids. The first is a highly conserved domain containing an

*Correspondence: fandavi@163.com; rxliu@gzu.edu.cn

² Guizhou Key Laboratory for Tobacco Quality Research, Guizhou University, Guiyang 550025, People's Republic of China

⁵ School of Food and Biological Engineering, Chengdu University, Chengdu 610106, People's Republic of China

Full list of author information is available at the end of the article



N- \times 7-R/K sequence, which consists of approximately 16 amino acid residues and functions in localization and specific binding to DNA [6, 7]. The second is a relatively poorly conserved leucine zipper region characterised by leucine in the last position of every seven amino acids, and hydrophobic residues at positions 3 and 4. The α -helix is a necessary structure of dimer required for bZIP proteins to bind to DNA [8–10]. The core sequence of bZIP TFs is the cis acting element of ACGT, e.g., as found in the G-box (CACGTG), C-box (GACGTC), the abscisic acid (ABA)-responsive element, and the A-box (ACGTA) [11–14].

Research into the bZIP family in different species will bring new insights into its evolution and function. Previous studies have shown that the bZIP TF family is involved in the regulation of genes involved in organ and tissue differentiation [15–18], seed storage [19, 20], pathogen defence [21], the unfolded protein response [22–24], photomorphogenesis and light signal regulation [25, 26], hormone and sugar signaling processes [27, 28], and responses to abiotic stress [3, 29, 30]. Different subfamilies regulate different biological processes. Studies have shown that *AtbZIP30* is involved in regulating cell development, organ development, hormone homeostasis, and meristematic processes [16]. Some bZIP TFs are activated under abiotic stress; they can combine with the promoter regions of key genes in signaling pathways to regulate their transcription levels and, in turn, plant resistance to stress. For example, research has identified that the expression of *EcbZIP60*, a gene homologous to *AtbZIP60*, is upregulated in response to drought and salt stresses [25]. In *Arabidopsis thaliana*, HY5, a member of the bZIP TF family, has been shown to inhibit hypocotyl growth and lateral root development, and promote pigment accumulation in a light-dependent manner [15]. *OsbZIP12* is a member of the bZIP family in rice that has been shown to be involved in ABA signal transduction and regulation of drought resistance [26]. *RsbZIP031* and *RsbZIP059* in radish were significantly downregulated under high temperature and salt stress, respectively [27]. In pineapple root tissues, *AcobZIP24* has a significant response to salt, drought, cold, and heat stresses at 24 and 48 h [28]. These examples illustrate the response of bZIP TFs in plants to various abiotic stresses.

Nicotiana tabacum L. is a tobacco plant of the Solanaceae. Tobacco is an important cash crop in agriculture, and is widely cultivated in more than 120 countries around the world [31–34]. Besides *A. thaliana*, tobacco is also often used as a model plant for the study of biological mechanisms because of the advantages conferred by its short growth time, disease susceptibility, and easy genetic transformation [35]. In addition, there is new interest in tobacco for its potential in biopharmaceuticals

and biofuel production [36, 37]. Thriving in the natural environment, tobacco is stressed by salt, drought, cold, heat, and flooding, which seriously impact its growth, yield, and quality [17, 18, 38–40]. The bZIP family has been comprehensively identified and studied, with 75 bZIPs identified in *A. thaliana* [7, 8], 89 in rice [4], 131 in soybean [9], 55 in grape [41], 90 in potato [42], and 52 in *Carthamus tinctorius* [43], with further bZIP genes identified in other plant species [44–46]. However, research into the bZIP family of tobacco is lacking. The development of the Tobacco Genome Initiative has provided many opportunities for the study of genome-wide gene families in tobacco [19, 20]. Furthermore, the emergence of next-generation sequencing technology enabled the problems of high cost and low yield to be overcome, making the development of tobacco genome sequencing more feasible [8, 9, 47–49]. In this study, we identified bZIPs from *Nicotiana tabacum* L., and studied their gene structure, chromosomal location, duplication, gene expression patterns, and responses to 6 abiotic stresses. We also studied the changes in enzyme activity occurring in plants under stress. In addition, an analysis of the homology between bZIPs in *A. thaliana*, *Oryza sativa*, *Fagopyrum tataricum*, *Solanum lycopersicum*, *Hordeum vulgare*, *Vitis vinifera*, *Solanum tuberosum* and tobacco was conducted. This study provides a theoretical basis for the mechanisms of stress resistance in tobacco and other species, and could open new research avenues in the breeding of high-quality tobacco varieties.

Results

Identification of *NtbZIPs* genes in tobacco

In this study, we identified 77 *NtbZIPs* in the tobacco database using two BLAST methods, and named them *NtbZIP01* through to *NtbZIP77* according to their chromosomal locations (from top to bottom) (Additional file 1: Table S1). The physical properties and basic characteristics of the predicted 77 *NtbZIPs* were analysed, including the coding sequence length (CDS), protein molecular weight (Mw), isoelectric point (pI), and predicted subcellular location (W) (Additional file 1: Table S1). The protein length of the predicted 77 *NtbZIPs* ranged from 85 (*NtbZIP43*) to 973 amino acids (*NtbZIP76*), with an average of 369 amino acids. The lowest Mw was 10.11 kDa (*NtbZIP43*), whilst the highest was 104.64 kDa (*NtbZIP76*); the average Mw was 40.77 kDa. The pI ranged from 4.17 (*NtbZIP32*) to 10.19 (*NtbZIP52*), with a mean of 7.11. Furthermore, analysis of predicted subcellular locations indicated that 64 *NtbZIPs* were located to the nuclear, 8 to chloroplasts, 3 (*NtbZIP21*, *NtbZIP43*, *NtbZIP68*) to the cytoplasm, 1 (*NtbZIP02*) to vacuole, and 1 (*NtbZIP51*) to the endoplasmic reticulum. The ratio of *NtbZIPs* to total genes in the tobacco

genome was approximately 0.10%, similar to Chinese jujube (0.13%) [45], but lower than *Chenopodium quinoa* (0.16%) [41], castor bean (0.19%) [50], *A. thaliana* (0.20%) [8], and banana (0.35%) [51].

Phylogenetic analysis and classification of bZIPs in tobacco

To study the evolutionary relationship between the bZIP families of *A. thaliana* and tobacco, a phylogenetic tree

was constructed based on the amino acid sequences of bZIPs in *A. thaliana* and tobacco (Fig. 1; Additional file 1: Table S1). We used the method proposed by Jakoby et al. [8] to construct an unrooted neighbour-joining tree for classification. The 77 identified *NtbZIPs* were clustered into 11 subfamilies (A-G, I, J, K, S) and four ‘orphan genes.’ Among these 11 subfamilies, 9 (A-G, I, and S) were similar to those found in *A. thaliana*, and the

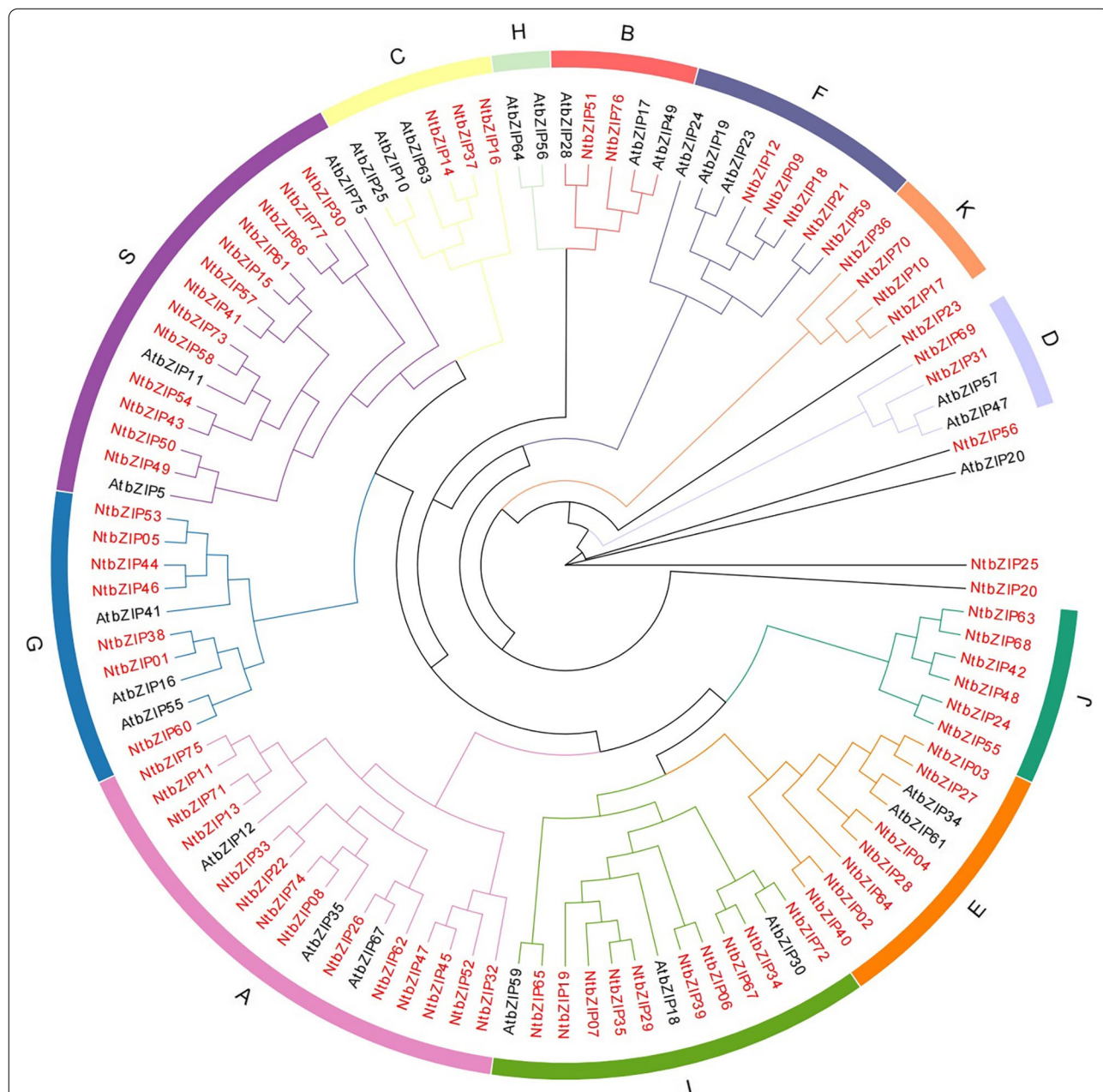


Fig. 1 Unrooted phylogenetic tree showing relationships between the bZIP domains of *Nicotiana tabacum* Land *Arabidopsis thaliana*. The phylogenetic tree was derived using the neighbour-joining method in MEGA X. The tree shows the 11 phylogenetic subfamilies (red). bZIP proteins from *Arabidopsis* are indicated by the prefix ‘At’. bZIP: basic leucine zipper; MEGA: Molecular Evolutionary Genetics Analysis

subfamilies were named to be consistent with *A. thaliana* [8]. We also identified 2 new subfamilies, J and K. These results indicated that the *bZIP* family is conserved in the evolutionary history of tobacco; these conserved subfamilies may play an important role in the process of evolution. The A subfamily was the largest subfamily identified, with 14 *NtbZIP* members; the S subfamily was second largest, containing 13 *NtbZIP*s. The smallest subfamilies were the B and D subfamilies, each containing 2 genes. The C, E, F, G, and I subfamilies had 3, 7, 5, 7, and 10 members, respectively. Subfamily H from *A. thaliana* (consisting of only *AtbZIP56* and *AtbZIP64*), contained no *NtbZIP* members, indicating that this group was lost during tobacco evolution (Fig. 1). Conversely, subfamilies J and K contained only *NtbZIP*s (6 and 4, respectively), suggesting that this may be a unique evolutionary direction in tobacco. Aside from the 11 subfamilies, 4 genes (*NtbZIP20*, *NtbZIP23*, *NtbZIP25*, and *NtbZIP56*) were not divided into any subfamily, which suggested that these 4 genes may have unique functions. The phylogenetic trees of *A. thaliana* and tobacco identified some *NtbZIP*s that were closely linked to *AtbZIP*s; these genes could potentially be homologous with the *AtbZIP*s and have similar physiological functions.

Gene structure and motif analysis of the *NtbZIP* genes

By comparing the corresponding genomic DNA sequences, all *NtbZIP*s were further analysed to better understand the structural characteristics, including the number and distribution of exons and introns (Fig. 2a,b). The numbers of introns present in the *NtbZIP*s ranged from 0 to 17 (Additional files 1: Tables S1). In total, 13 *NtbZIP* genes (16.88%) had no intron, 8 (10.39%) had 1 intron, 6 (7.79%) had 2 introns, 22 (28.57%) had 3 introns, 4 (5.19%) had 4 introns, 2 (2.60%) had 5 introns, 1 (1.30%) had 6 introns, 3 (3.90%) had 7 introns, 1 (1.30%) had 8 introns, 4 (5.19%) had 11 introns, 1 (1.30%) had 16 introns, 5 (6.49%) had 17 introns, respectively. The genes with no intron belonged to the A, F, and S subfamilies, and predominantly in the S subfamily. The 22 *NtbZIP*s with 3 introns accounted for the largest proportion of genes (28.57%). *NtbZIP24*, 42, 48, 63, and 68 had the most number of introns, 17, and all of these genes belonged to the J subfamily. In general, genes in the same subgroup had a similar structure. For example, all genes in the E subfamily (except *NtbZIP40*) had 4 exons and 3 introns; all genes in the J subfamily (except *NtbZIP55*) had 18 exons and 17 introns.

In order to study the characteristic regions of the *NtbZIP* proteins, we used the MEME online tool to predict conserved motifs. A total of 10 conserved motifs (motifs 1–10) were identified from the 77 *NtbZIP* proteins (Fig. 2c, Additional file 2: S2). The length of the

motifs ranged from 15 (motif 3) to 50 (motif 2, 4, 7, 8, 9, 10) (Additional file 2: S2). Motif 1 and 3 were widely distributed in the majority of *NtbZIP*s and these 2 motifs were very closely positioned in proteins, suggesting that the two motifs may play an important role in particular functions. As expected, members from the same subfamily shared similar motif compositions, particularly between a few close members. For example, members of the J subfamily had the same motifs: 3, 4, 6, 7, and 8. In subfamily S, apart from *NtbZIP43* which only contained motif 3, the rest of the members included both motifs 1 and 3; in addition, motif 9 was found in all the members of subfamily F. Furthermore, we also found that some motifs only appeared in certain subfamilies; e.g., motif 8 only appeared in subfamily J, indicating that this motif is specific to this subfamily. Analysis also identified that some of the 10 motifs only occurred in particular orders. In the J subfamily, except for where motifs were repeated, the motifs only appeared in a specific order: 6–3–4–7–8. Similarly, this also occurred in the F subfamily, in which the order of motifs was 9–1–3–10. In summary, the genes that in the same subfamily shared similar gene structure and motif composition, providing further support for the phylogenetic analysis and subfamily classifications.

Chromosome location and duplication of *bZIP*s in tobacco

On the basis of the tobacco genome sequence, we constructed a map of the physical positions of the *NtbZIP*s (Fig. 3). As shown in Fig. 3, all the identified 77 *NtbZIP*s were located on tobacco chromosomes, and the nomenclature of the genes was guided by their physical positions on tobacco Chr1 to Chr24. Aside from Chr8 and Chr14, on which no *NtbZIP* was distributed, the number of genes distributed on the remaining chromosomes ranged from 1 to 15. Chr17 contained the largest number of *NtbZIP*s (15 genes, 19.48%), followed by Chr24 (8 genes, 10.39%), Chr19 (7 genes, 9.09%) and Chr5 (6 genes, 7.79%). Chr1, Chr3, Chr7, Chr11, Chr13, Chr16, Chr18, Chr20, and Chr21 included just 1 gene (~1.30%), Chr22 contained 2 *NtbZIP*s (~2.60%), and Chr2, Chr9, Chr15, and Chr23 each contained 3 *NtbZIP*s (~3.90%). Chr6 and Chr10 had 4 (~5.19%) *NtbZIP*s, whilst Chr4 and Chr12 all included 5 *NtbZIP*s (6.49%). There was a correlation between chromosome size and the number of genes on chromosome. Such as, the largest chromosome (Chr17) contains 15 genes, and the smallest chromosome (Chr21) contains only one gene.

Furthermore, 16 pairs of segmental duplication were observed in the *NtbZIP*s (Fig. 4, Additional file 3: Table S3) and no tandem duplication was observed. Twenty-nine (37.66%) paralogs were identified in the *NtbZIP* family, indicating an evolutionary relationship between

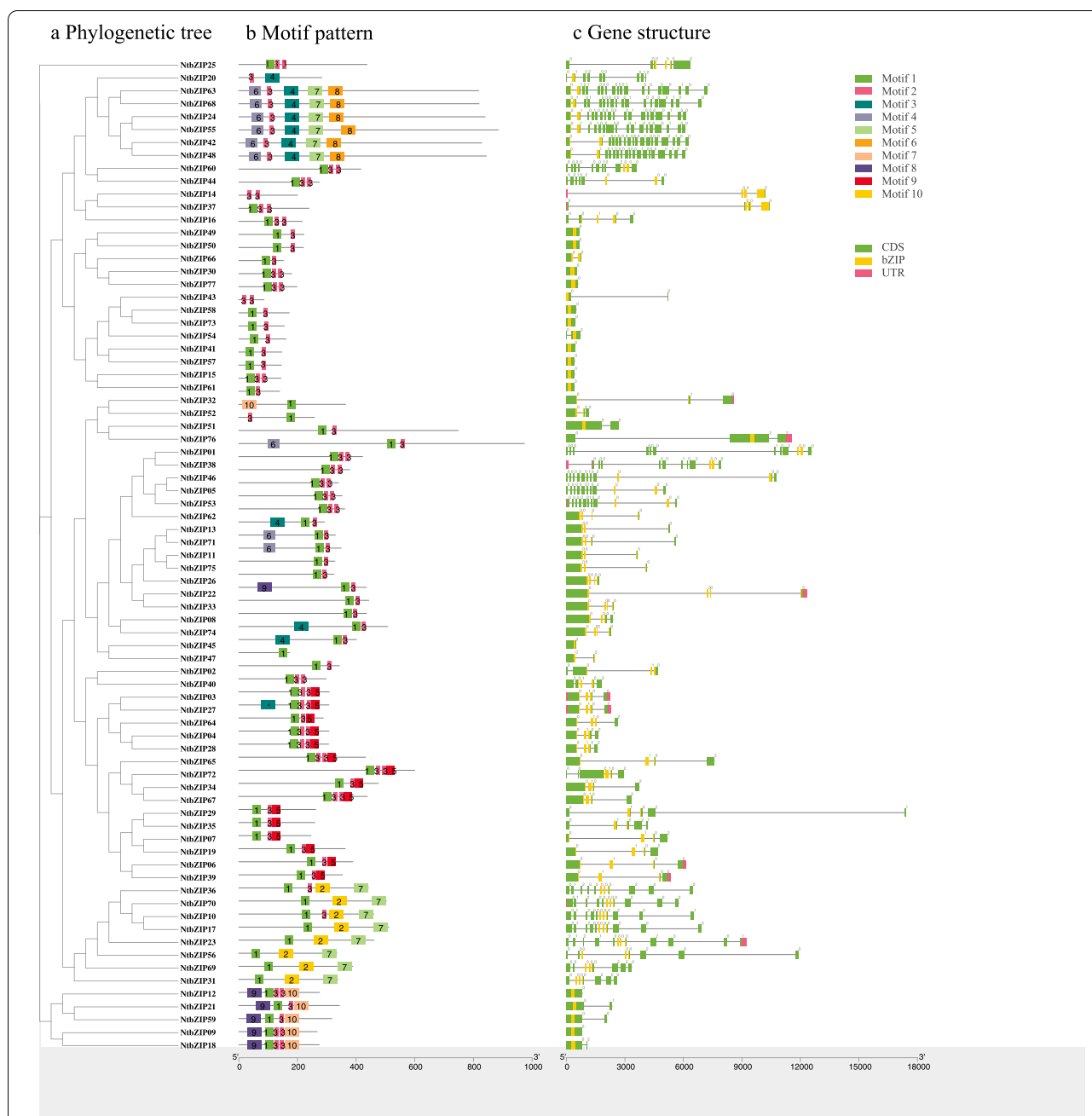


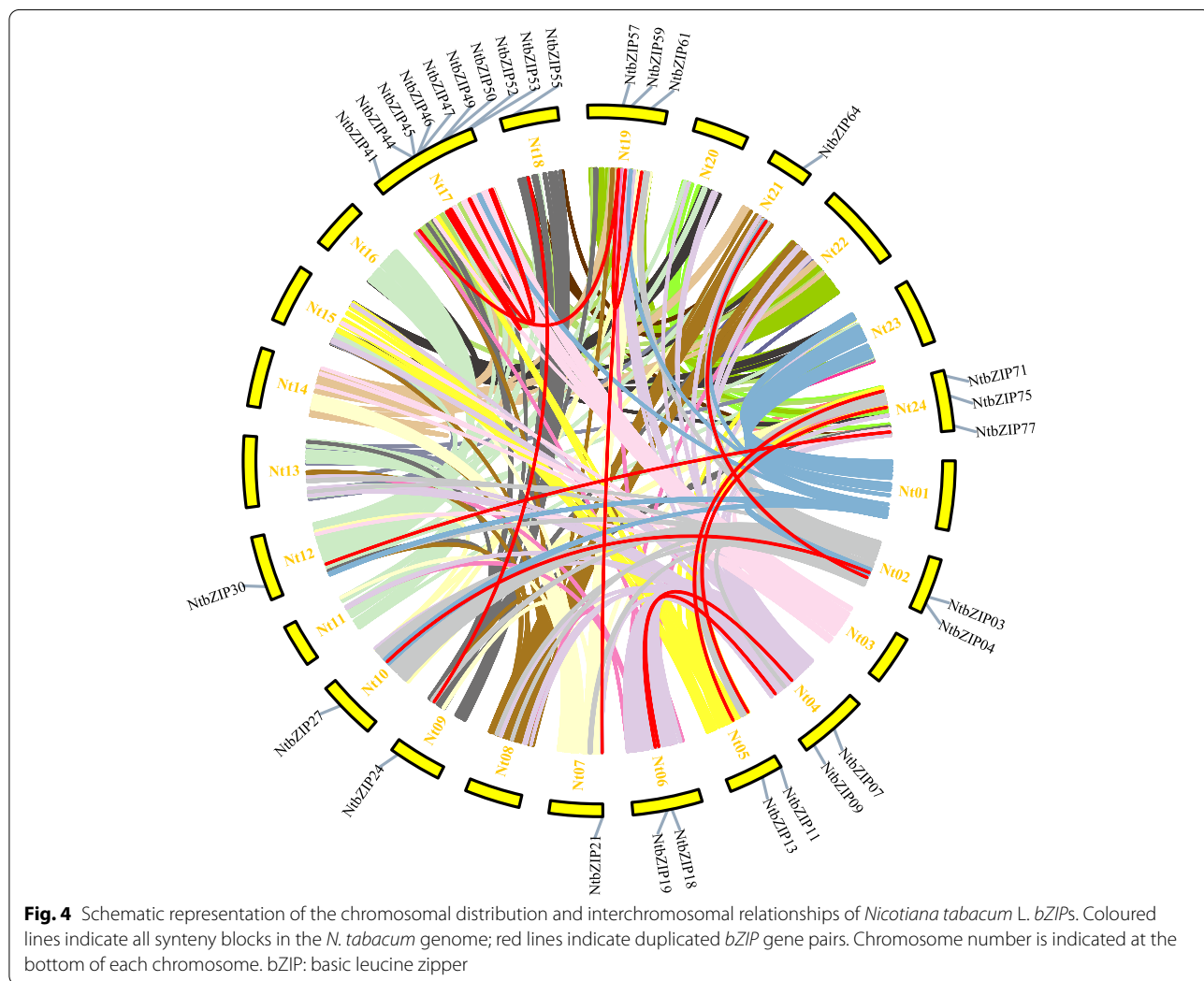
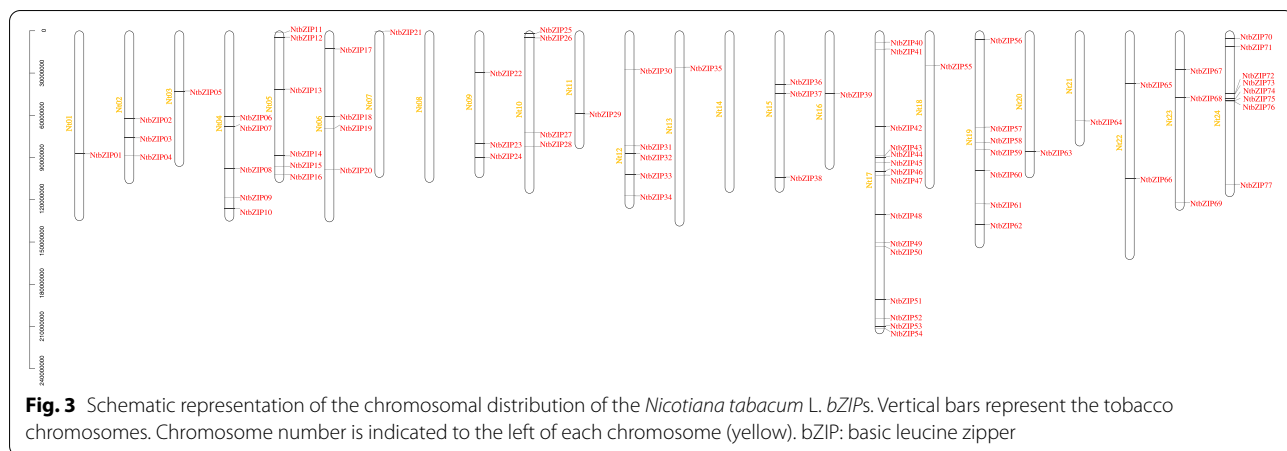
Fig. 2 Phylogenetic relationships, gene structure analysis, and motif distributions of *Nicotiana tabacum* L. bZIPs. bZIP: basic leucine zipper. **a** Phylogenetic tree constructed by the neighbour-joining method, with 1000 replicates on each node. **b** Amino acid motifs in the bZIP proteins (1–10) are represented by coloured boxes. Black lines indicate relative protein lengths. **c** Exons and introns are indicated by green rectangles, and grey lines, respectively

these bZIP members (Fig. 4). The paralogous NtbZIPs were unevenly distributed between 13 tobacco linkage groups (LGs) (Fig. 4). LG17 contained the most NtbZIPs (9), whilst LG7, LG9, LG10, LG12, LG18, and LG21 had the least (1). LG2, LG4, LG5, and LG6 each contained 2 genes, and there were 3 genes in LG19 and LG24. We also

found that the 16 pairs of paired genes were linked within their subfamilies of S, A, E, G, and I.

Synten analysis of NtbZIPs

To analyse the evolutionary relationship of the tobacco bZIP family, we constructed 7 synten maps, comparing



tobacco with six representative species, including *A. thaliana*, *S. lycopersicum*, *F. tataricum*, *O. sativa*, *H. vulgare*, *V. vinifera* and *S. tuberosum* (Fig. 5, Additional file 4:

Table S4). In total, 68 *NtbZIPs* displayed a syntenic relationship across the 7 species, indicating that the tobacco bZIPs were highly conserved on the corresponding

chromosomes in the process of evolution. The largest number of *NtbZIPs* displaying a syntenic relationship occurred when comparing the tobacco genes to *S.tuberosum* (60), followed by *S. lycopersicum* (59), *V. vinifera* (45), *A. thaliana* (29), *F. tataricum* (27), *O. sativa* (9), and lastly *H. vulgare* (3) (Additional file 4: Table S4). The number of orthologous pairs between the other 7 species analysed (*A. thaliana*, *O. sativa*, *F. tataricum*, *S. lycopersicum*, *H. vulgare*, *V. vinifera* and *S.tuberosum*) was 44, 14, 35, 97, 4, 59 and 92, respectively. These result illustrated that the genetic relationship of *NtbZIP* genes, *SlbZIP* genes and *StbZIP* genes were close, which was reasonable because they are all Solanaceae plants. Several *NtbZIPs* were associated with at least 6 syntenic gene pairs: *NtbZIP07*, *NtbZIP21*, *NtbZIP24*, *NtbZIP30*, *NtbZIP33*, *NtbZIP38*, *NtbZIP41*, *NtbZIP43*, *NtbZIP49*, *NtbZIP50*, *NtbZIP52*, *NtbZIP57*, *NtbZIP59*, *NtbZIP60*, and *NtbZIP73*, indicating that these genes may have played an important role in the evolution of the *bZIP* family in tobacco. Furthermore, *NtbZIP33* was found in tobacco and all of the other 7 species, indicating that this orthologous pair appeared earlier than ancestral differentiation, and that *NtbZIP33* may have played a vital role in the evolution of the tobacco *bZIP* family.

Expression patterns of *NtbZIPs* in different plant tissues and organs

To investigate the physiological roles of the *NtbZIPs*, we selected 11 *NtbZIPs* from different subfamilies to study gene expression patterns. The expression levels of these *NtbZIPs* in eight tissues (calyx, petal, anther, stigma, leaf, stem, fruit, and root) were evaluated by quantitative real-time polymerase chain reaction (qRT-PCR) (Fig. 6a). The different expression patterns found for each gene in different organs indicate that *NtbZIPs* have different functions in tobacco. Most of the genes were expressed in all organs, except *NtbZIP04*, which was not expressed in the leaf or stem. *NtbZIP69* genes was more highly expressed in the calyx than in other organs. *NtbZIP29* and *NtbZIP69* were highly expressed in petal tissue, and *NtbZIP70* was highly expressed in anther tissue. *NtbZIP01* was more highly expressed in both stigma and leaf tissue than in other tissues, and its expression in these tissues was higher than other *NtbZIPs*; similarly *NtbZIP49* was more highly expressed in stem, fruit, and root tissue than other genes. In addition, we analysed whether the expression

patterns of the 11 *NtbZIPs* in the 8 organs were correlated (Fig. 6b). This analysis found that the expression of different genes in the plant organs were significantly correlated; some genes were positively correlated (*NtbZIP49* with *NtbZIP37*, *NtbZIP69*, and *NtbZIP70*; *NtbZIP01* with *NtbZIP04*, *NtbZIP76*, and *NtbZIP18*), whilst some were negatively correlated (*NtbZIP37* and *NtbZIP29*; *NtbZIP70* with *NtbZIP01*, *NtbZIP18*, *NtbZIP29*, *NtbZIP76*). All of the 11 *NtbZIPs* showed positive correlation with different tissues (Additional file 5: Figure S1). For example, the *NtbZIP70*, *NtbZIP74*, *NtbZIP29* showed highly positive correlation with the anthers, the stem and the petals, respectively. However, *NtbZIP63* showed negative correlation with anthers.

Expression patterns of *NtbZIPs* in response to different stresses

To study the expression of *NtbZIPs* under abiotic stress, we used qRT-PCR to analyse the expression of 11 *NtbZIPs* in roots, stems, and leaves under 6 abiotic stresses: polyethylene glycol (PEG), flooding, heat, cold, NaCl, and strong ultraviolet (UV) radiation treatments (Fig. 7a). The expression patterns of *NtbZIPs* differed in response to the 6 abiotic stresses, with some *NtbZIPs* either significantly induced or inhibited. In the early stages of stress treatments, the expression patterns of most genes changed significantly (Fig. 7a); in response to different stress treatments, the expression of some *NtbZIPs* also changed with time or in different organs. For example, in response to PEG stress, *NtbZIP04*, *NtbZIP18*, and *NtbZIP49* were first significantly upregulated in root tissue, and then downregulated; *NtbZIP49* also demonstrated this pattern of expression in stem and leaf tissue. In response to heat stress, *NtbZIP18* and *NtbZIP29* were significantly downregulated in the root, stem, and leaf, whereas *NtbZIP74* was first significantly upregulated, and then downregulated in these tissues. Some other genes showed changes in specific organs: for example, the response of *NtbZIP76* to cold stress was significant only in the stem. In addition, we analysed whether there were correlations between the expression patterns of these *NtbZIPs* (Fig. 7b), detecting positive correlations among most *NtbZIPs*. *NtbZIP49* and *NtbZIP63* were significantly positively correlated, and 4 genes (*NtbZIP04*, *NtbZIP18*, *NtbZIP37* and *NtbZIP49*) were significantly positively correlated with *NtbZIP29*. Under different abiotic stresses, plant tissues showed

(See figure on next page.)

Fig. 5 Synteny analyses of the *bZIPs* between *Nicotiana tabacum* and 7 representative plant species (*Arabidopsis thaliana*, *Oryza sativa*, *Fagopyrum tataricum*, *Solanum lycopersicum*, *Hordeum vulgare*, *Vitis vinifera* and *Solanum tuberosum*). Grey lines on the background indicate the collinear blocks in *N. tabacum* and other plant genomes; red lines highlight the syntenic *N. tabacum* *bZIP* gene pairs. *bZIP*: basic leucine zipper. Outer panel: An unrooted phylogenetic tree constructed using Geneious R11 with the neighbour-joining method. Inner panel: Distribution of the conserved motifs in *bZIP* proteins. The differently coloured boxes represent different motifs and their positions in each *bZIP* protein sequence. The sequence information for each motif is provided in Additional file 2: Table S2

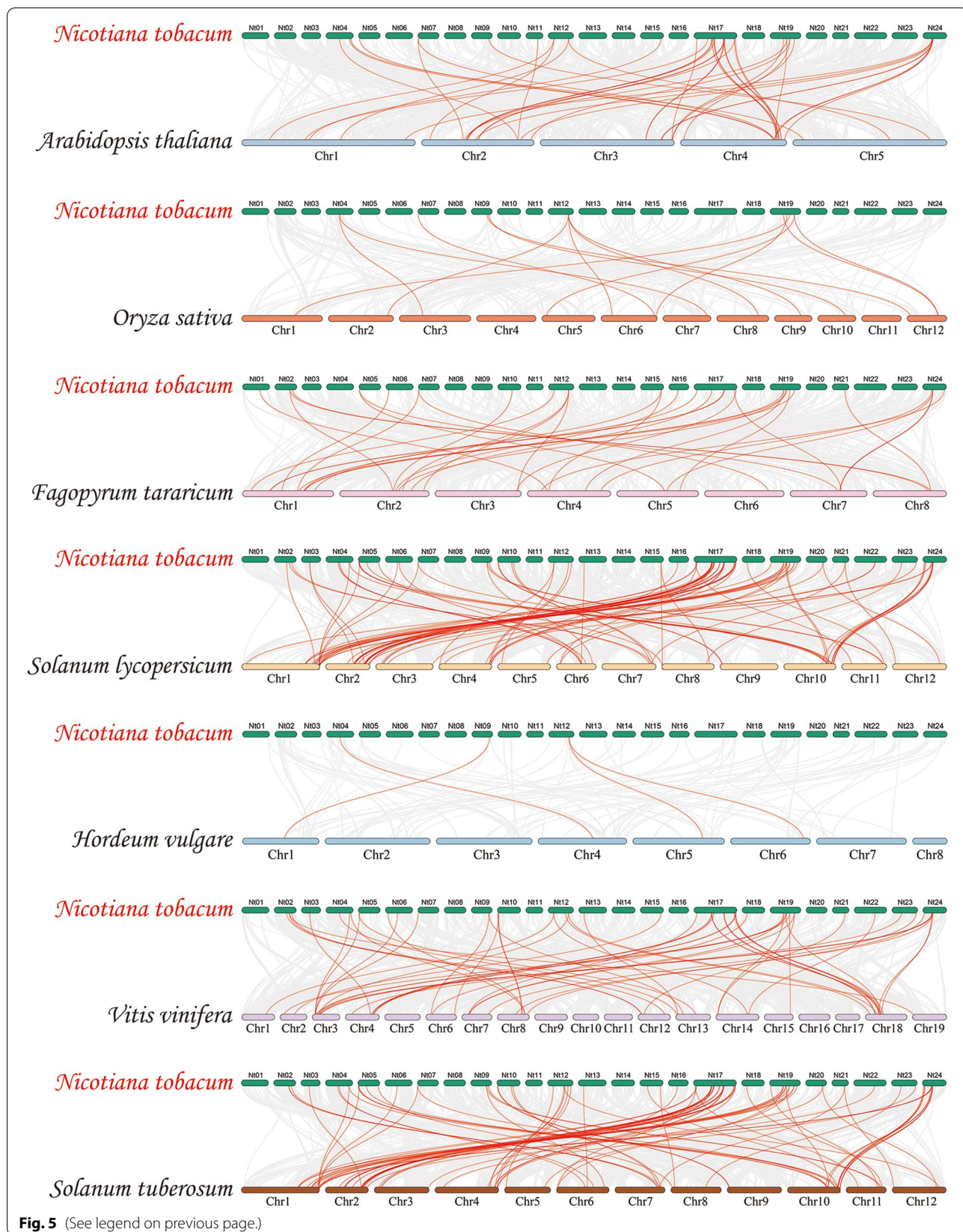


Fig. 5 (See legend on previous page.)

different correlations with gene expression patterns (Additional file 6: Figure S2). Under flooding stress, leaf tissue showed significant positive correlations with *NtbZIP69*, but a negative correlation with *NtbZIP29*. Similarly, under heat stress, root tissue was significantly positively correlated with *NtbZIP29*, *NtbZIP18* and *NtbZIP01*, but significantly negatively correlated with *NtbZIP04* and *NtbZIP76*. Furthermore, The number of genes that positively related to stem tissue were more than that of leaf and root tissue under 6 abiotic stresses.

Discussion

Tobacco is an important commercial crop and model plant. The bZIP family is one of the most important TF families found in the plant kingdom. Increasing evidence suggests that bZIP TFs play important roles in regulating growth and development, and responses to biotic and abiotic stresses [3, 15, 52]. Previous research on the functions of bZIP TFs has mainly focused on *A. thaliana*, and there were also relevant studies in other crops, vegetables, and fruits [4, 8, 9, 28, 42, 53]. Compared with other important plants, there has been little research on the bZIP family in tobacco; therefore, we systematically analysed the bZIP family in tobacco.

In this study, 77 bZIPs were identified from tobacco (3734 Mb) on the basis of a constructed phylogenetic tree. Previous studies have identified 75 bZIPs in *A. thaliana* (117 Mb) [8], 89 in rice (466 Mb) [4], 125 in maize (2182 Mb) [54, 55], 160 in soybean [56] (915 Mb), 55 in grape [53] (490 Mb), and 96 in tartary buckwheat [44] (489 Mb). Results from this study suggest that the number of bZIPs in tobacco is similar to that found in *A. thaliana*, more than in grape and in *Carthamus tinctorius*, and fewer than in rice and soybean. There is no correlation between the number of bZIPs in a species and genome size. According to cluster analysis and the similarity of protein sequences with *A. thaliana*, the 77 *NtbZIPs* divided into 11 subfamilies, and 2 subfamilies (J and K) were newly identified. The number of subfamilies was consistent with the buckwheat [44], castor bean [57], more than *A. thaliana* [8] and grape [53], lower than Chinese jujube [45] and poplar [46]. The J subfamily is consistent with grapevine [53] and Chinese jujube [45], whereas the K subfamily is the same as that in soybean [56] and radish [27]. Subfamilies A and S had the largest number of bZIPs, similar to as seen in *A. thaliana* [8]. Four of the *NtbZIPs* did not have obvious clusters;

this also occurred in grape [53], and banana [51]. The H subfamily did not contain any tobacco genes, which suggested that the H subfamily was lost in tobacco.

Gene replication, including tandem, segment, and whole genome replication, plays an important role in the evolution of organisms and the expansion of gene family [58]. In this study, 16 pairs of segmental duplications were observed in the *NtbZIP* genes, without any tandem replications, which also occurred in poplar [46]. These results suggest that segmental replication may contribute more than tandem duplications to the expansion of the bZIP family in tobacco, consistent with previous studies in other species, including rice [4], maize [59], grape [60], and cucumber [61]. The number of bZIPs in tobacco is similar to that found in *A. thaliana* [8], more than in grape [53] and in *Carthamus tinctorius* [43] and fewer than in rice [8] and soybean [56]. This difference may be caused by the whole-genome duplication event that occurred after the earliest ancestors of the other species diverged [62]. These segment duplication events may be promote the occurrence and evolution of some *NtbZIP* genes, which also occurred in *Populus trichocarpa* [63] and *Sorghum bicolor* [62]. From the results of chromosome distribution, there were no *NtbZIP* genes on Chr8 and Chr14, illustrating that the *NtbZIP* gene family may be affected by gene deletion during the process of evolution [53]. Therefore, some bZIP genes may be caused by some replication events, which further illustrates that replication events play an important role in the rapid expansion of bZIP family members in plants.

In addition, the phylogenetic analysis was confirmed by gene structure and conserved motif analysis of the 77 bZIPs. Analysis of the gene structure of the *NtbZIP* family identified several differences among the *NtbZIPs*. We determined that the number of introns present varied from 0 to 17, with 16.88% of the genes containing no introns, similar to as seen in rice [8], soybean [56], and sorghum [64]. The two duplicated genes tended to cluster in a subfamily, and previous studies have also shown such results [64]. The results of phylogenetic analysis confirmed this, with *NtbZIP30* and *NtbZIP77* (1 exon and 0 introns), *NtbZIP11* and *NtbZIP75* (3 exons and 2 introns), and *NtbZIP21* and *NtbZIP59* (2 exons and 1 intron) clustered to subfamilies S, A, and F, respectively. The subfamilies G, J, and K contained genes with more than 9 introns, while other subfamilies had genes with fewer introns. Nuruz-zaman et al. [65] studied the bZIP TF family in rice, and

(See figure on next page.)

Fig. 6 Tissue-specific gene expression of 11 *Nicotiana tabacum* L. bZIPs and the correlation between their expression patterns. bZIP: basic leucine zipper; qRT-PCR: Real-time quantitative PCR. **a** Expression patterns of 11 *Nicotiana tabacum* L. bZIPs in the calyx, petal, anther, stigma, leaf, stem, fruit, and root organs, as examined by qRT-PCR. Error bars represent standard error. Lowercase letters above bars indicate a significant difference ($P < 0.05$, least significant difference test) among the treatments. **b** Positive number: positively correlated; negative number: negatively correlated. Red numbers indicate a significant correlation at the 0.05 level

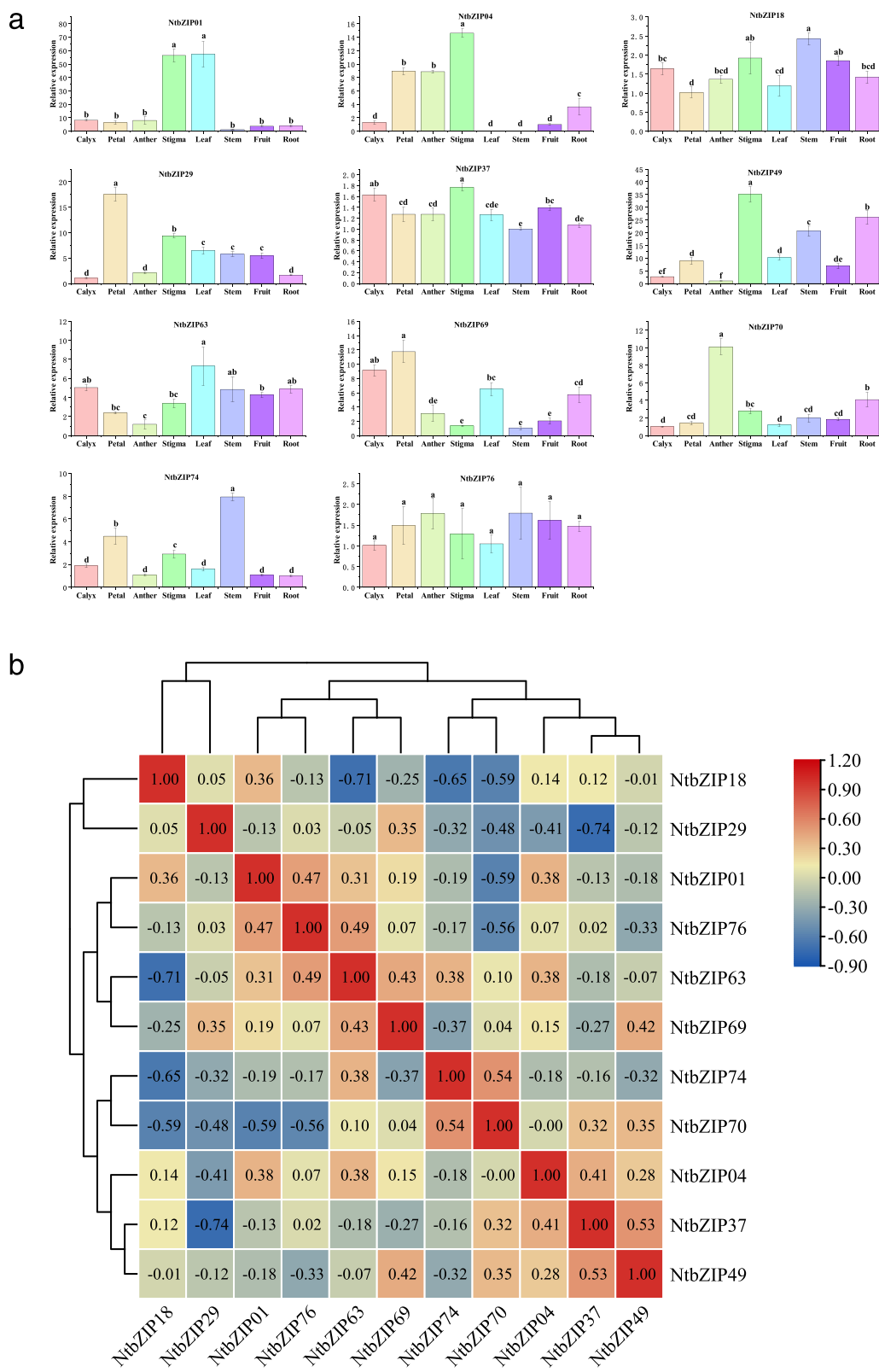


Fig. 6 (See legend on previous page.)

found that the rate of intron acquisition was slower than the rate of intron deletion after segmental duplication. Therefore, we can infer that the G, J, and K subfamilies may contain primitive genes. In most gene families, the increase and loss of introns and exons is one of the mechanisms for diversification [66], and this was also observed here. For example, *NtbZIP47* had 1 intron and 2 exons, while its paralogous gene *NtbZIP45* contained 0 introns and 1 exon, indicating the loss of an exon during evolution. A similar phenomenon also occurred in the *bZIP* gene families of apple [67] and soybean [56]. The increase of introns is conducive to the adaptation of plants to the environment, which is beneficial to evolution. However, genes with fewer introns may be conducive to rapid response to the environment [68]. In this study, a total of 10 motifs were identified in the 77 *NtbZIP* proteins. Although different subfamilies have different motifs and the motif encoding *bZIP* is conservative. Most members contained the motif 1 conserved domain, which contains an N- \times 7-R/K conserved sequence, indicating conservation of the *bZIP* TF family in tobacco. Furthermore, analysis of the conserved motifs indicated that each subfamily member shared similar motifs, which were the same as those in buckwheat [27, 44]. For example, all *NtbZIP* proteins shares motifs 1, 3 and 5 in group I. Both Group S and I contain motifs 1 and 3. These results not only support the classification of the subfamilies, but also indicate that members in the same subfamily have similar evolutionary histories and physiological functions [50].

The *bZIP* family plays an important regulatory role in growth and development, including the development of flowers and fruits [51, 69, 70]. Analysis of gene expression profiles can provide an important basis for understanding the potential biological functions of genes. The functions of *bZIP* TFs in tobacco still need to be further studied. Here, we selected 11 *NtbZIPs*, clustered in different subfamilies and with significant differences, to study the expression patterns in different organs. All of the *NtbZIP* genes investigated were expressed in at least 1 of the organs tested, indicating that they may play a broad role in the growth and development of tobacco, similar to as the former reported research in soybean [62] and buckwheat [71]. Most genes were expressed in all organs, consistent with previous studies [8, 43, 45, 72], however there were some genes that were not expressed in some of the organs, as previously seen in studies of grape, soybean, and apple [53,

56, 67]. For instance, the *NtbZIP04* gene was not expressed in leaf and stem. In this study, according to the phylogenetic tree constructed by *NtbZIP* genes homologous to *AtbZIP* genes (Fig. 1), the function of *NtbZIP* gene homologous to *AtbZIP* can be predicted. Clarifying the function of these *NtbZIP* genes can lay a foundation for the growth and development of tobacco and other species in future. As the reproductive organs of plants, flowers and fruits are the main structures of all angiosperms [73]. In comparison to non-reproductive tissues (root, stem, and leaf), the expression levels of *NtbZIP29* was significantly higher in the petal, stigma and anther tissues. *AtbZIP18*, a gene in *A. thaliana* homologous to the tobacco gene *NtbZIP29*. The study of *AtbZIP18* gene in *A. thaliana* showed that *AtbZIP18* had high expression in pollen and played an important role in male gametophyte [72]. Therefore, we speculated that *NtbZIP29* is involved in the development of flowers and fruits. *NtbZIP01* showed high expression in leaves. *AtbZIP41*, a gene homologous to *NtbZIP01*, has been proven to recognise and specifically bind with the G-box to form the G-box-binding factor 1/*bZIP41* binding factor, which is constitutively expressed in *A. thaliana* leaves [74–76]. According to the subfamily classification result (Fig. 1), *NtbZIP01* belongs to G subfamily that contains the G-box-binding factor. The specific expression of *NtbZIP01* gene in leaf may be due to its special structure. The expression level of *NtbZIP04* was highest in stigma; the *A. thaliana* genes homologous to *NtbZIP04*, *AtbZIP34* and *AtbZIP61*, have been confirmed in several studies to interact with other *bZIP* TFs to participate in plant development [48, 74, 77]. The *NtbZIP04* gene may be has similar function with the homologous genes of *AtbZIP34* and *AtbZIP61*.

bZIP genes play an vital role in plant stress response, which have illustrated in previous studies [4, 8, 9]. We also studied the gene expression patterns of these 11 *NtbZIPs* under 6 abiotic stresses. The results showed that the expression of most of the *NtbZIPs* were significantly different under abiotic stresses, which was consistent with previous studies [28, 53, 78]. For example, under drought stress, *NtbZIPs* were upregulated in the root (9), stem (9), and leaf (7). There were 8, 5, 11 *NtbZIPs* upregulated in the root, stem, and leaf, respectively. Additionally, the same gene showed opposite expression patterns under different stress. *NtbZIP04* gene was up-regulated in leaves under heat stress, but down-regulated under cold stress. These results

(See figure on next page.)

Fig. 7 Expression of 11 *Nicotiana tabacum* L. *bZIPs* in plants subjected to abiotic stresses (strong UV radiation, flooding, PEG, NaCl, and hot and cold treatments) at the seedling stage. *bZIP*: basic leucine zipper; qRT-PCR: Real-time quantitative PCR. **a** Expression patterns of 11 *Nicotiana tabacum* L. *bZIPs* in leaf, root, and stem organs were examined by qRT-PCR. Error bars represent standard error. Lowercase letters above bars indicate a significant difference ($P < 0.05$, LSD) among the treatments. *bZIP*: basic leucine zipper; qRT-PCR: Real-time quantitative PCR. **b** Positive number: positively correlated; negative number: negatively correlated. Red numbers indicate a significant correlation at the 0.05 level

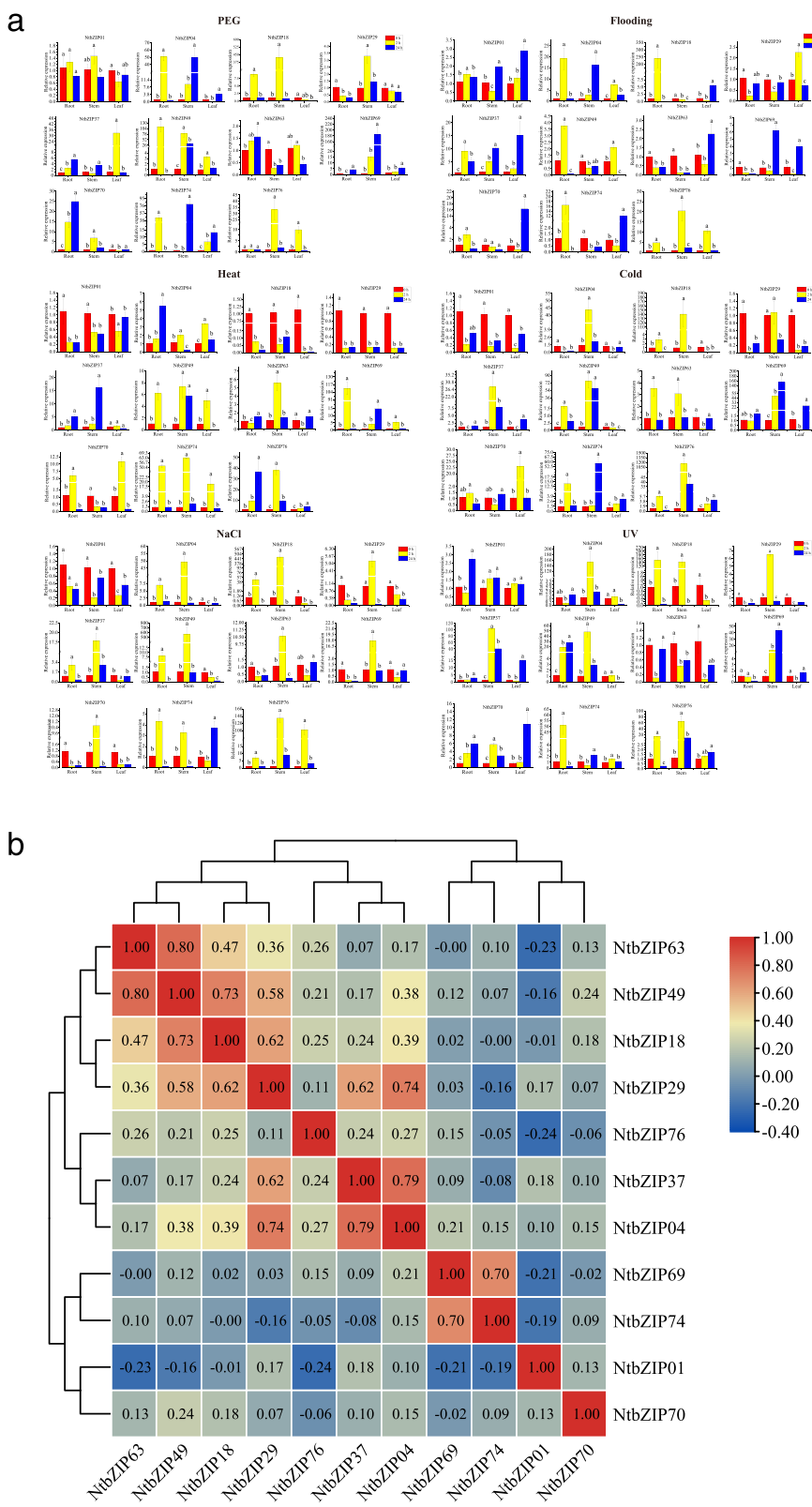


Fig. 7 (See legend on previous page.)

illustrate that *bZIP* genes are widely involved in the response to abiotic stress and *bZIP* TFs participate in a complex cross regulatory network [64]. *NtbZIP49* and *NtbZIP76* responded to PEG, flooding, heat, cold, and NaCl abiotic stresses at the same time, indicating that these two genes are co-expressed under a variety of adverse conditions, which was consistent with the study of *Sorghum bicolor* [64]. We found high expression of *NtbZIP18* under all 6 abiotic stresses. *NtbZIP18* is homologous with subfamily F members *AtbZIP19*, *AtbZIP23*, and *AtbZIP24* of *A. thaliana*. Previous studies showed that *AtbZIP24* is an important regulator of the salt stress response in plants, and regulates complex transcription networks involved in abiotic stress resistance [79, 80]. The *NtbZIP18* gene may have similar functions to *AtbZIP24* gene. Subsequent studies can verify the function of *NtbZIP18* gene to prove whether it can be used to improve the resistance of varieties to various abiotic stress. *NtbZIP49* showed the highest expression levels in stem tissue under NaCl stress; this gene is homologous with *AtbZIP75* of *A. thaliana*, reported to respond to salt stress [81]. Among the six abiotic stress, *NtbZIP74* showed high expression in root. *NtbZIP74* gene and *AtbZIP67* are homologous genes and both come from subfamily A that is mainly related to abscisic acid or some stress signal pathways [82]. Previous confirmed that *AtbZIP67* played a positive role in ABA-induced AHT1 expression. The up-regulated expression of *NtbZIP74* gene in roots may be due to its participation in ABA signaling pathway. In this analysis, *NtbZIP76* showed significant responses to heat stress. This gene is homologous to *AtbZIP17* (*A. thaliana*), which plays a vital role in the endoplasmic reticulum heat stress pathway, supporting our findings in tobacco [83,84]. From the expression heat map, we found that most of the *NtbZIPs* were positively correlated. By considering the interactions between the expression patterns and complex proteins, we could find a feedback mechanism network to coordinate the expression of multiple genes.

Conclusion

In summary, we analysed the whole genome of the *bZIP* family in tobacco, including gene structure, chromosomal location, sequence homology, synteny, and expression patterns. In this study, a total of 77 *NtbZIPs* were identified and divided into 11 subfamilies, providing preliminary information for the characterisation of *NtbZIP* function. Although there were many types of proteins, the exon–intron structure of genes, and the motif arrangement, of *bZIP* proteins from the same subfamily, supported the classification predicted by the

phylogenetic tree. The 77 *NtbZIPs* were unevenly distributed across 22 tobacco chromosomes. A total of 16 pairs of segmental duplication were identified, illustrating the important role of gene duplication in the expansion of the *NtbZIP* family. In addition, expression analysis of *NtbZIPs* in different tissues in response to 6 abiotic stresses implied that *NtbZIPs* participate in the development of tobacco. *NtbZIP49* may be important in the development of flowers and fruits; *NtbZIP18* is an important regulator in abiotic stress. These results will be helpful for the further study of the physiological processes performed by the *bZIP* family in tobacco.

Methods

Identification of *bZIP* family in tobacco

The entire genome sequence of *Nicotiana tabacum* L. was downloaded from the solgenomics website (https://solgenomics.net/organism/Nicotiana_tabacum/genome). Two BLASTp methods were used to obtain *NtbZIPs* [85, 86]. Firstly, all *AtbZIPs* were used to search for possible *NtbZIPs* from the tobacco genome by using BLASTp (score value ≥ 100 and E value $\leq 1e^{-10}$). Secondly, the Hidden Markov Model file of the *bZIP* domain was downloaded from the Pfam protein family database (<http://pfam.xfam.org>) [87]. Both HMMER 3.0 (<http://hmmer.org/>), with a cutoff of 0.01, and SMART (Simple Modular Architecture Research Tool) (<https://academic.oup.com/nar/article/49/D1/D458/5940513>), were used to identify and verify *bZIPs*. Furthermore, we analysed the basic characteristics of the identified *bZIP* proteins using the ExpASY website (<https://web.expasy.org/>), including the CDS, pI, and Mw.

Exon/intron structures and conserved motif analysis

ClustalW was used to align the default parameters of the multiple amino acid sequences of the *bZIPs* identified [88]. GeneDoc software (<https://genedoc.software.informer.com/>) was used to manually adjust the deduced amino acid sequences in the *bZIP* domain. The exon/intron composition of the *bZIPs* was analysed by the Gene Structure Display Server (GSDS: <http://GSDS.cbi.pku.edu.cn>) [89]. MEME (<http://meme-suite.org/tools/meme>) was used to analyse the motifs of the *bZIP* proteins [90, 91]. Optimisation parameters were set as follows: repetition times, arbitrary; maximum number of motifs, 10; optimal width of each motif, between 6 and 200 residues [86, 91, 92].

Chromosomal distribution, gene duplication and collinear analysis with other species

All *NtbZIPs* were mapped to tobacco chromosomes using Circos [93], according to the physical location information in the tobacco genome database. We used the default parameters of the Multiple Collinearity Scan toolkit (MCScanX) to detect gene duplication [94]. Homology between *bZIPs* in tobacco and the other 7 species (*A. thaliana*, *O. sativa*, *F. tataricum*, *S. lycopersicum*, *H. vulgare*, *V. vinifera* and *S. tuberosum*) was analysed through the Dual Synteny Plotter software (<https://github.com/CJ-Chen/TBtools>).

Phylogenetic analysis and classification of the *NtbZIP* family

All identified *NtbZIPs* were divided into several subfamilies, according to the classification of the *AtbZIP* family. Using the neighbour-joining method of Molecular Evolutionary Genetics Analysis (MEGA) X, the phylogenetic tree was built using Geneious R11 with a BLOSUM62 matrix, using the Jukes Cantor substitution model, a global alignment with free end gaps, and a bootstrap value of 1000. The full-length amino acid sequences of *bZIP* proteins were compared with the newly identified *NtbZIPs* for phylogenetic analysis (Additional file 1: Table S1), derived from *A. thaliana*, *O. sativa*, *F. tataricum*, *S. lycopersicum*, *H. vulgare*, *V. vinifera* and *S. tuberosum* (Uniprot: <https://www.uniprot.org/>).

Plant materials, growth conditions, and abiotic stress in tobacco

'K326', which is a typical, cultivated, common tobacco variety with good yield and quality, was used in this study. 'K326' has been grown in the greenhouse of the Guizhou University tobacco research base since 2020. The seeds of 'K326' were sown in a special substrate for flue-cured tobacco and raised by floating seedling in a growth room with a 16 h/25 °C day and 8 h/20 °C night regime, and a relative humidity of 75%. We collected the calyx, petal, anther, stigma, leaf, stem, fruit, and root separately from 3 plants showing good and similar growth, 71 days after transplanting, and repeated sampling for three times. To investigate the expression patterns in response to different abiotic stresses, 11 *bZIPs* were selected for further study. Tobacco plants were subjected to the following abiotic stresses 63 days post-sowing: salt (5% NaCl), water flooding (whole plant), drought (20% PEG6000), UV exposure (70 μW/cm², 220 V, 30 W), high temperature (38 °C), and low temperature (4 °C). Every abiotic stress treatment was performed with 3 replicates; materials were sampled for qRT-PCR analysis and enzyme

activity determination after treatment for 2 h and 24 h. Collected samples were kept at -80 °C until analysis.

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

The RNAPrep Pure Plant Kit (TIANGEN, DP432) was used to extract total RNA from samples. The cDNA was generated using the FastKing gDNA Dispelling RT Super-Mix kit (TIANGEN, KR118), following manufacturer's instructions. We selected 11 *NtbZIPs* that all with close evolutionary relationships to *AtbZIPs* from different subfamilies to study gene expression patterns. The qRT-PCR primers used to analyse gene expression were designed by Primer Premier 6 (<http://www.premierbiosoft.com/>) (Additional file 7: Table S5). We used the *Actin* gene as an internal reference, as it is stably expressed at each growth stage in almost all tissues. The qRT-PCR analysis, performed with Talent qPCR PreMix (SYBR Green) (TIANGEN, FP209), was repeated at least 3 times and the 2^{-ΔΔCt} method was used to calculate the final test results [95].

Statistical analysis

DPS (Data Processing System) software was used to analyse the variance of the data, and data were compared with the least significant difference test at the 0.05 and 0.01 levels. OriginPro 2016 software (OriginLab, <https://www.originlab.com/>) was used for the generation of graphs.

Abbreviations

bZIP: Basic leucine zipper; CDS: Coding sequence length; DPS: Data Processing System; LG: Linkage group; MEGA: Molecular Evolutionary Genetics Analysis; MEME: Multiple Expectation maximizations for Motif Elicitation; Mw: Protein molecular weight; PEG: Polyethylene glycol; pI: Isoelectric point; qRT-PCR: Real-time quantitative PCR; SMART: Simple Modular Architecture Research Tool; TF: Transcription factor; UV: Ultraviolet; W: Predicted Subcellular location.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08547-z>.

Additional file 1: Table S1. List of the 77 *Nicotianatabacum* L. *bZIPs* identified in this study. *bZIP*: basic leucine zipper.

Additional file 2: Table S2. Analysis and distribution of the conserved motifs in *Nicotiana tabacum* L. *bZIP* proteins. *bZIP*: basic leucine zipper.

Additional file 3: Table S3. The 16 pairs of segmental duplications in *Nicotiana tabacum* L. *bZIPs*. *bZIP*: basic leucine zipper.

Additional file 4: Table S4. One-to-one orthologous gene relationships between *Nicotiana tabacum* L. and other plants.

Additional file 5: Figure S1. The correlation analysis between 8 different tissues and expression patterns of the 11 *Nicotiana tabacum* L. *bZIP* genes.

Additional file 6: Figure S2. The correlation analysis between 3 different tissues and expression patterns of the 11 *Nicotiana tabacum* L. *bZIP* genes under abiotic stresses.

Additional file 7: Table S5. Primer sequences used for real-time quantitative PCR.

Acknowledgements

Not applicable.

Authors' contributions

LD planned and designed the research, analysed the data, and wrote the manuscript. ZM and YF studied the gene expression by qRT-PCR. KL, MY, DL and YK identified the gene family and analysed gene structure, chromosome distribution, the gene duplication and syntenic analysis. QZ and FW analysed the evolutionary relationship of bZIPs in several different species. RL supervised the research. YF and RL revised the manuscript. All authors read and approved the final manuscript. All authors reviewed the manuscript.

Funding

This study was financially supported by the National Science Foundation of China (32060510), the Guizhou Province High-level Innovative Talents Training Project Hundred Talents Program (Qiankehepingtairencai[2016]5663), the Guizhou Provincial Key Foundation (Qiankehejichu[2019]1405), and the Science and Technology Project of Guizhou Tobacco Company (201904).

Availability of data and materials

The whole *Nicotiana tabacum* genome sequence information was obtained from the solgenomics website (https://solgenomics.net/organism/Nicotiana_tabacum/genome) and the website is open to all researchers. The tobacco materials ('K326') used in this study were supplied by Prof. Liu Renxiang of Guizhou University. The datasets supporting the conclusions of this article are included in the article and its Additional files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Agriculture, Guizhou University, Guiyang 550025, People's Republic of China. ²Guizhou Key Laboratory for Tobacco Quality Research, Guizhou University, Guiyang 550025, People's Republic of China. ³College of Tobacco, Guizhou University, Guiyang 550025, People's Republic of China. ⁴College of Food Science and Engineering, Xinjiang Institute of Technology, Aksu 843100, People's Republic of China. ⁵School of Food and Biological Engineering, Chengdu University, Chengdu 610106, People's Republic of China.

Received: 10 December 2021 Accepted: 13 April 2022

Published online: 22 April 2022

References

- Zhang B, Liu J, Yang ZE, Chen EY, Zhang CJ, Zhang XY, et al. Genome-wide analysis of GRAS transcription factor gene family in *Gossypium hirsutum* L. *BMC Genomics*. 2018;19:1–12. <https://doi.org/10.1186/s12864-018-4722-x>.
- Martinez-García JF, Moyano E, Alcocer MJC, Martin C. Two bZIP proteins from *Antirrhinum* flowers preferentially bind a hybrid C-box/G-box motif and help to define a new sub-family of bZIP transcription factors. *Plant J*. 1998;13:489–505. <https://doi.org/10.1046/j.1365-3113x.1998.00050.x>.
- Rodriguez-Urbe L, O'Connell MA. A root-specific bZIP transcription factor is responsive to water deficit stress in tepary bean (*Phaseolus acutifolius*) and common bean (*P. vulgaris*). *J Exp Bot*. 2006;57:1391–8. <https://doi.org/10.1093/jxb/erj118>.
- 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;146:333–50. <https://doi.org/10.1104/pp.107.112821>.
- Zhang Lina. Functional study of wheat transcription factors bZIP and NAC Genes involved in Abiotic stress tolerance. Chinese Academy of Agricultural Sciences. 2014.
- Suckow M, Schwamborn K, Kisters-woike B, Von Wilcken-bergmann B, Müller-hill B. Replacement of invariant bzip residues within the basic region of the yeast transcriptional activator GCN4 can change its DNA binding specificity. *Nucleic Acids Res*. 1994;22:4395–404. <https://doi.org/10.1093/nar/22.21.4395>.
- Guedes Corrêa LG, Riaño-Pachón DM, Guerra Schrago C, Vicentini dos Santos R, Mueller-Roeber B, Vincenz M. The role of bZIP transcription factors in green plant evolution: Adaptive features emerging from four founder genes. *PLoS One*. 2008;3:1–12. <https://doi.org/10.1371/journal.pone.0002944>.
- Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, et al. bZIP transcription factors in Arabidopsis. *Trends Plant Sci*. 2002;7:106–11. [https://doi.org/10.1016/S1360-1385\(01\)02223-3](https://doi.org/10.1016/S1360-1385(01)02223-3).
- Liao Y, Zou HF, Wang HW, Zhang WK, Ma B, Zhang JS, et al. Soybean GmMYB76, GmMYB92, and GmMYB177 genes confer stress tolerance in transgenic Arabidopsis plants. *Cell Res*. 2008;18:1047–60. <https://doi.org/10.1038/cr.2008.280>.
- Zhang Hongjing. Bioinformatic and expression analysis of the basic leucine zipper gene family in grape. Northwest Agriculture and Food University. 2014.
- Izawa T, Foster R, Chua NH. Plant bZIP protein DNA binding specificity. *J Mol Biol*. 1993;230:1131–44. <https://doi.org/10.1006/jmbi.1993.1230>.
- Foster R, Izawa T, Chua N. Plant bZIP proteins gather at ACGT elements. *FASEB J*. 1994;8:192–200. <https://doi.org/10.1096/fasebj.8.2.8119490>.
- E ZG, Zhang YP, Zhou JH, Wang L. Roles of the bZIP gene family in rice. *Genet Mol Res*. 2014;13:3025–36. <https://doi.org/10.4238/2014.April.16.11>
- Ali Z, Sarwat SS, Karim I, Jaskani MJ, Khan AA. Functions of plant's bZIP transcription factors. *Pakistan J Agric Sci*. 2016;53:303–14. <https://doi.org/10.21162/PAKJAS/16.2043>.
- Gangappa SN, Crocco CD, Johansson H, Datta S, Hettiarachchi C, Holm M, et al. The Arabidopsis B-BOX protein BBX25 interacts with HY5, negatively regulating BBX22 expression to suppress seedling photomorphogenesis. *Plant Cell*. 2013;25:1243–57. <https://doi.org/10.1105/tpc.113.109751>.
- Lozano-Sotomayor P, Chávez Montes RA, Silvestre-Vañó M, Herrera-Ubaldo H, Greco R, Pablo-Villa J, et al. Altered expression of the bZIP transcription factor DRINK ME affects growth and reproductive development in Arabidopsis thaliana. *Plant J*. 2016;88:437–51. <https://doi.org/10.1111/tpl.13264>.
- Zhang X, Lin HM, Hu H, Hu X, Hu L. Gamma-aminobutyric acid mediates nicotine biosynthesis in tobacco under flooding stress. *Plant Divers*. 2016;38:53–8. <https://doi.org/10.1016/j.pld.2016.05.004>.
- Tan SL, Yang YJ, Liu T, Zhang SB, Huang W. Responses of photosystem I compared with photosystem II to combination of heat stress and fluctuating light in tobacco leaves. *Plant Sci*. 2019;2020(292):110371. <https://doi.org/10.1016/j.plantsci.2019.110371>.
- Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB. A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Mol Plant-Microbe Interact*. 2012;25:1523–30. <https://doi.org/10.1094/MPMI-06-12-0148-TA>.
- Edwards KD, Fernandez-Pozo N, Drake-Stowe K, Humphry M, Evans AD, Bombarely A, et al. A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics*. 2017;18:1–14. <https://doi.org/10.1186/s12864-017-3791-6>.
- Alves MS, Dadalto SP, Gonçalves AB, De Souza GB, Barros VA, Fietto LG. Plant bZIP transcription factors responsive to pathogens: A review. *Int J Mol Sci*. 2013;14:7815–28. <https://doi.org/10.3390/ijms14047815>.
- Iwata Y, Koizumi N. An Arabidopsis transcription factor, AtbZIP60, regulates the endoplasmic reticulum stress response in a manner unique to plants. *Proc Natl Acad Sci U S A*. 2005;102:5280–5. <https://doi.org/10.1073/pnas.0408941102>.
- Kondo S, Saito A, Asada R, Kanemoto S, Imaizumi K. Physiological unfolded protein response regulated by OASIS family members,

- transmembrane bZIP transcription factors. *IUBMB Life*. 2011;63:233–9. <https://doi.org/10.1002/iub.433>.
24. Takasaki H, Maruyama K, Kidokoro S, Ito Y, Fujita Y, Shinozaki K, et al. The abiotic stress-responsive NAC-type transcription factor OsNAC5 regulates stress-inducible genes and stress tolerance in rice. *Mol Genet Genomics*. 2010;284:173–83. <https://doi.org/10.1007/s00438-010-0557-0>.
 25. Babitha KC, Ramu SV, Nataraja KN, Sheshshayee MS, Udayakumar M. EcZIP60, a basic leucine zipper transcription factor from *Eleusine coracana* L. improves abiotic stress tolerance in tobacco by activating unfolded protein response pathway. *Mol Breed*. 2015;35:1–7. <https://doi.org/10.1007/s11032-015-0374-6>.
 26. Joo J, Lee YH, Song SI. Overexpression of the rice basic leucine zipper transcription factor OsbZIP12 confers drought tolerance to rice and makes seedlings hypersensitive to ABA. *Plant Biotechnol Rep*. 2014;8:431–41. <https://doi.org/10.1007/s11816-014-0335-2>.
 27. Fan L, Xu L, Wang Y, Tang M, Liu L. Genome-and transcriptome-wide characterization of bZIP gene family identifies potential members involved in abiotic stress response and anthocyanin biosynthesis in radish (*Raphanus sativus* L.). *Int J Mol Sci*. 2019;20:1–17. <https://doi.org/10.3390/ijms20246334>.
 28. Liu Y, Chai M, Zhang M, He Q, Su Z, Priyadarshani SVGN, et al. Genome-Wide Analysis, Characterization, and Expression Profile of the Basic Leucine Zipper Transcription Factor Family in Pineapple. *Int J Genomics*. 2020;5:1–14. <https://doi.org/10.1155/2020/3165958>.
 29. Rolly NK, Mun BG, Yun BW. Insights into the transcriptional regulation of branching hormonal signaling pathways genes under drought stress in *Arabidopsis*. *Genes (Basel)*. 2021;12:1–17. <https://doi.org/10.3390/genes12020298>.
 30. Liu JX, Srivastava R, Howell SH. Stress-induced expression of an activated form of AtbZIP17 provides protection from salt stress in *Arabidopsis*. *Plant, Cell Environ*. 2008;31:1735–43. <https://doi.org/10.1111/j.1365-3040.2008.01873.x>.
 31. Siervo N, Battley JND, Ouadi S, Bakaher N, Bovet L, Willig A, et al. The tobacco genome sequence and its comparison with those of tomato and potato. *Nat Commun*. 2014;5:1–9. <https://doi.org/10.1038/ncomms4833>.
 32. Song Z, Wang D, Gao Y, Li C, Jiang H, Zhu X, et al. Changes of lignin biosynthesis in tobacco leaves during maturation. *Funct Plant Biol*. 2020. <https://doi.org/10.1071/FP20244>.
 33. Jing Z, Xiquan Li, Fei D, Jianpeng Z, Ying H, Guangqin W, Po W. Comprehensive Evaluation of Tobacco- growing Soil Fertility in Karst Mountain Area — Take Xixiu District of Anshun As an Example. *J Mountain Agric Biol*. 2021;40:29–35. <https://doi.org/10.15958/j.cnki.sdnyxwb.2021.02.005>.
 34. Sifola MI, Carrino L, Cozzolino E, Del Piano L, Graziani G, Ritienni A. Potential of pre-harvest wastes of tobacco (*Nicotiana tabacum* L.) crops, grown for smoke products, as source of bioactive compounds (phenols and flavonoids). *Sustain*. 2021;13:1–13. <https://doi.org/10.3390/su13042087>.
 35. Xiang X, Wu X, Chao J, Yang M, Yang F, Chen G, Liu G, Wang Y. Genome-wide identification and expression analysis of the WRKY gene family in common tobacco (*Nicotiana tabacum* L.). *Hereditas*. 2016;38:840–56. <https://doi.org/10.16288/jyczz.16-016>.
 36. Ma JKC, Drossard J, Lewis D, Altmann F, Boyle J, Christou P, et al. Regulatory approval and a first-in-human phase I clinical trial of a monoclonal antibody produced in transgenic tobacco plants. *Plant Biotechnol J*. 2015;13:1106–20. <https://doi.org/10.1111/pbi.12416>.
 37. Tusé D, Tu T, McDonald KA. Manufacturing economics of plant-made biologics: Case studies in therapeutic and industrial enzymes. *Biomed Res Int*. 2014;2014:1–17. <https://doi.org/10.1155/2014/256135>.
 38. Li Y, Ren K, Hu M, He X, Gu K, Hu B, et al. Cold stress in the harvest period: effects on tobacco leaf quality and curing characteristics. *BMC Plant Biol*. 2021;21:131. <https://doi.org/10.1186/s12870-021-02895-w>.
 39. Xu J, Zhou Y, Xu Z, Chen Z, Duan L. Combining physiological and metabolomic analysis to unravel the regulations of coronatine alleviating water stress in tobacco (*Nicotiana tabacum* L.). *Biomolecules*. 2020;10:1–16. <https://doi.org/10.3390/biom10010099>.
 40. Sun H, Sun X, Wang H, Ma X. Advances in salt tolerance molecular mechanism in tobacco plants. *Hereditas*. 2020;157:1–6. <https://doi.org/10.1186/s41065-020-00118-0>.
 41. Li F, Liu J, Guo X, Yin L, Zhang H, Wen R. Genome-wide survey, characterization, and expression analysis of bZIP transcription factors in *Chenopodium quinoa*. *BMC Plant Biol*. 2020;20:1–11. <https://doi.org/10.1186/s12870-020-02620-z>.
 42. Kumar P, Kumar P, Sharma D, Verma SK, Halterman D, Kumar A. Genome-wide identification and expression profiling of basic leucine zipper transcription factors following abiotic stresses in potato (*Solanum tuberosum* L.). *PLoS One*. 2021;16(3):e0247864. <https://doi.org/10.1371/journal.pone.0247864>.
 43. Li H, Li L, ShangGuan G, Jia C, Deng S, Noman M, et al. Genome-wide identification and expression analysis of bZIP gene family in *Carthamus tinctorius* L. *Sci Rep*. 2020;10:1–15. <https://doi.org/10.1038/s41598-020-72390-z>.
 44. Liu M, Wen Y, Sun W, Ma Z, Huang L, Wu Q, et al. Genome-wide identification, phylogeny, evolutionary expansion and expression analyses of bZIP transcription factor family in tartary buckwheat. *BMC Genomics*. 2019;20:1–18. <https://doi.org/10.1186/s12864-019-5882-z>.
 45. Zhang Y, Gao W, Li H, Wang Y, Li D, Xue C, et al. Genome-wide analysis of the bZIP gene family in Chinese jujube (*Ziziphus jujuba* Mill.). *BMC Genomics*. 2020;21:1–14. <https://doi.org/10.1186/s12864-020-06890-7>.
 46. Zhao K, Chen S, Yao W, Cheng Z, Zhou B, Jiang T. Genome-wide analysis and expression profile of the bZIP gene family in poplar. *BMC Plant Biol*. 2021;21:1–16. <https://doi.org/10.1186/s12870-021-02879-w>.
 47. Chern MS, Bobb AJ, Bustos MM. The regulator of MAT2 (ROM2) protein binds to early maturation promoters and represses PvALF-activated transcription. *Plant Cell*. 1996;8:305–21. <https://doi.org/10.1105/tpc.8.2.305>.
 48. Shen H, Cao K, Wang X. A conserved proline residue in the leucine zipper region of AtbZIP34 and AtbZIP61 in *Arabidopsis thaliana* interferes with the formation of homodimer. *Biochem Biophys Res Commun*. 2007;362:425–30. <https://doi.org/10.1016/j.bbrc.2007.08.026>.
 49. Bennetzen JL. Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol*. 2000;42:251–69. <https://doi.org/10.1023/A:1006344508454>.
 50. Jin Z, Xu W, Liu A. Genomic surveys and expression analysis of bZIP gene family in castor bean (*Ricinus communis* L.). *Planta*. 2014;239:299–312. <https://doi.org/10.1007/s00425-013-1979-9>.
 51. Hu W, Wang L, Tie W, Yan Y, Ding Z, Liu J, et al. Genome-wide analyses of the bZIP family reveal their involvement in the development, ripening and abiotic stress response in banana. *Sci Rep*. 2016;6:1–15. <https://doi.org/10.1038/srep30203>.
 52. BabuRajendra Prasad V, Gupta N, Nandi A, Chattopadhyay S. HY1 genetically interacts with GBF1 and regulates the activity of the Z-box containing promoters in light signaling pathways in *Arabidopsis thaliana*. *Mech Dev*. 2012;129:298–307. <https://doi.org/10.1016/j.mod.2012.06.004>.
 53. Liu J, Chen N, Chen F, Cai B, Dal Santo S, Torrielli GB, et al. Genome-wide analysis and expression profile of the bZIP transcription factor gene family in grapevine (*Vitis vinifera*). *BMC Genomics*. 2014;15:1–18. <https://doi.org/10.1186/1471-2164-15-281>.
 54. Bolle C. The role of GRAS proteins in plant signal transduction and development. *Planta*. 2004;218:683–92. <https://doi.org/10.1007/s00425-004-1203-z>.
 55. Rogozin IB, Carmel L, Csuros M, Koonin EV. Origin and evolution of spliceosomal introns. *Biol Direct*. 2012;7:1–28. <https://doi.org/10.1186/1745-6150-7-11>.
 56. Zhang M, Liu Y, Shi H, Guo M, Chai M, He Q, et al. Evolutionary and expression analyses of soybean basic leucine zipper transcription factor family. *BMC Genomics*. 2018;19:1–14. <https://doi.org/10.1186/s12864-018-4511-6>.
 57. Wang J, Zhou J, Zhang B, Vanitha J, Ramachandran S, Jiang SY. Genome-wide Expansion and Expression Divergence of the Basic Leucine Zipper Transcription Factors in Higher Plants with an Emphasis on Sorghum. *J Integr Plant Biol*. 2011;53:212–31. <https://doi.org/10.1111/j.1744-7909.2010.01017.x>.
 58. Xu G, Guo C, Shan H, Kong H. Divergence of duplicate genes in exon-intron structure. *Proc Natl Acad Sci U S A*. 2012;109:1187–92. <https://doi.org/10.1073/pnas.1109047109>.
 59. Wei K, Chen J, Wang Y, Chen Y, Chen S, Lin Y, et al. Genome-wide analysis of bZIP-encoding genes in maize. *DNA Res*. 2012;19:463–76. <https://doi.org/10.1093/dnares/dss026>.
 60. Gao M, Zhang H, Guo C, Cheng C, Guo R, Mao L, et al. Evolutionary and Expression Analyses of Basic Zipper Transcription Factors in the Highly Homozygous Model Grape PN40024 (*Vitis vinifera* L.). *Plant Mol Biol Report*. 2014;32:1085–102. <https://doi.org/10.1007/s1105-014-0723-3>.

61. Baloglu MC, Eldem V, Hajyzadeh M, Unver T. Genome-wide analysis of the bZIP transcription factors in cucumber. *PLoS One*. 2014;9:1–14. <https://doi.org/10.1371/journal.pone.0096014>.
62. Li K, Duan L, Zhang Y, Shi M, Chen S, Yang M, et al. Genome-wide identification and expression profile analysis of trihelix transcription factor family genes in response to abiotic stress in sorghum [*Sorghum bicolor* (L.) Moench]. *BMC Genomics*. 2021;22:1–17. <https://doi.org/10.1186/s12864-021-08000-7>.
63. Wang Z, Liu Q, Wang H, Zhang H, Xu X, Li C, et al. Comprehensive analysis of trihelix genes and their expression under biotic and abiotic stresses in *Populus trichocarpa*. *Sci Rep*. 2016;6:1–15. <https://doi.org/10.1038/srep36274>.
64. Fan Y, Yan J, Lai D, Yang H, Xue G, He A, et al. Genome-wide identification, expression analysis, and functional study of the GRAS transcription factor family and its response to abiotic stress in sorghum [*Sorghum bicolor* (L.) Moench]. *BMC Genomics*. 2021;22:1–18. <https://doi.org/10.1186/s12864-021-07848-z>.
65. Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, et al. Genome-wide analysis of NAC transcription factor family in rice. *Gene*. 2010;465:30–44. <https://doi.org/10.1016/j.gene.2010.06.008>.
66. Wang L, Yin X, Cheng C, Wang H, Guo R, Xu X, et al. Evolutionary and expression analysis of a MADS-box gene superfamily involved in ovule development of seeded and seedless grapevines. *Mol Genet Genomics*. 2015;290:825–46. <https://doi.org/10.1007/s00438-014-0961-y>.
67. Zhao J, Guo R, Guo C, Hou H, Wang X, Gao H. Evolutionary and expression analyses of the apple basic leucine zipper transcription factor family. *Front Plant Sci*. 2016;7:1–13. <https://doi.org/10.3389/fpls.2016.00376>.
68. Shabalina SA, Ogurtsov AY, Spiridonov AN, Novichkov PS, Spiridonov NA, Koonin EV. Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes. *Mol Biol Evol*. 2010;27:1745–9. <https://doi.org/10.1093/molbev/msq086>.
69. Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M, Lohmann JU, et al. Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science*. 2005;309:1056–9. <https://doi.org/10.1126/science.1114358>.
70. Xu Z, Ali Z, Xu L, He X, Huang Y, Yi J, et al. The nuclear protein GmbZIP110 has transcription activation activity and plays important roles in the response to salinity stress in soybean. *Sci Rep*. 2015;2016(6):1–12. <https://doi.org/10.1038/srep20366>.
71. Ma Z, Liu M, Sun W, Huang L, Wu Q, Bu T, et al. Genome-wide identification and expression analysis of the trihelix transcription factor family in tartary buckwheat (*Fagopyrum tataricum*). *BMC Genomics*. 2019;19:1–16. <https://doi.org/10.1186/s12870-019-1957-x>.
72. Zou X, Li L, Liao F, Chen W. iTRAQ-based quantitative proteomic analysis reveals NtGNL1-dependent regulatory network underlying endosome trafficking for pollen tube polar growth. *Plant Physiol Biochem*. 2020;2021(161):200–9. <https://doi.org/10.1016/j.plaphy.2021.02.010>.
73. Gibalová A, Reňák D, Matczuk K, Dupl'Áková N, Cháb D, Twell D, et al. AtbZIP34 is required for *Arabidopsis* pollen wall patterning and the control of several metabolic pathways in developing pollen. *Plant Mol Biol*. 2009;70:581–601.
74. Shaikhali J. GIP1 protein is a novel cofactor that regulates DNA-binding affinity of redox-regulated members of bZIP transcription factors involved in the early stages of *Arabidopsis* development. *Protoplasma*. 2015;252:867–83. <https://doi.org/10.1007/s00709-014-0726-9>.
75. Shaikhali J, Norén L, De Dios B-López J, Srivastava V, König J, Sauer UH, et al. Redox-mediated mechanisms regulate DNA binding activity of the G-group of basic region leucine zipper (bZIP) transcription factors in *Arabidopsis*. *J Biol Chem*. 2012;287:27510–25.
76. Smykowski A, Fischer SM, Zentgraf U. Phosphorylation affects DNA-binding of the senescence-regulating bZIP transcription factor GBF1. *Plants*. 2015;4:691–709.
77. Gremski K, Ditta G, Yanofsky MF. The HECATE genes regulate female reproductive tract development in *Arabidopsis thaliana*. *Development*. 2007;134:3593–601. <https://doi.org/10.1242/dev.011510>.
78. Fujita Y, Fujita M, Satoh R, Maruyama K, Parvez MM, Seki M, et al. AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in *Arabidopsis*. *Plant Cell*. 2005;17:3470–88.
79. Popova OV, Yang O, Dietz KJ, Gollmack D. Differential transcript regulation in *Arabidopsis thaliana* and the halotolerant *Lobularia maritima* indicates genes with potential function in plant salt adaptation. *Gene*. 2008;423:142–8. <https://doi.org/10.1016/j.gene.2008.07.017>.
80. Yang O, Popova OV, Süthoff U, Lüking I, Dietz KJ, Gollmack D. The *Arabidopsis* basic leucine zipper transcription factor AtbZIP24 regulates complex transcriptional networks involved in abiotic stress resistance. *Gene*. 2009;436:45–55. <https://doi.org/10.1016/j.gene.2009.02.010>.
81. Liu Z, Zheng L, Pu L, Ma X, Wang X, Wu Y, et al. ENO2 Affects the Seed Size and Weight by Adjusting Cytokinin Content and Forming ENO2-bZIP75 Complex in *Arabidopsis thaliana*. *Front Plant Sci*. 2020;11:1–14. <https://doi.org/10.3389/fpls.2020.574316>.
82. Choi HI, Hong JH, Ha JO, Kang JY, Kim SY. ABFs, a family of ABA-responsive element binding factors. *J Biol Chem*. 2000;275:1723–30. <https://doi.org/10.1074/jbc.275.3.1723>.
83. Chopperla R, Singh S, Mohanty S, Reddy N, Padaria JC, Solanke AU. Isolation and expression analysis of EcbZIP17 from different finger millet genotypes shows conserved nature of the gene. *3 Biotech*. 2017;7:1–10. <https://doi.org/10.1007/s13205-017-0984-2>.
84. Ramakrishna C, Singh S, Raghavendrarao S, Padaria JC, Mohanty S, Sharma TR, et al. The membrane tethered transcription factor EcbZIP17 from finger millet promotes plant growth and enhances tolerance to abiotic stresses. *Sci Rep*. 2018;8:1–14.
85. Zhuang Y, Zhou X, Wang S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Plant Syst Evol*. 2012;298:3389–402. <https://doi.org/10.2503/jshs.58.977>.
86. Liu M, Ma Z, Wang A, Zheng T, Huang L, Sun W, et al. Genome-wide investigation of the auxin response factor gene family in tartary buckwheat (*Fagopyrum tataricum*). *Int J Mol Sci*. 2018;19:1–18. <https://doi.org/10.3390/ijms19113526>.
87. Shamloo S. TportHMM: Predicting the substrate class of transmembrane transport proteins using profile Hidden Markov Models. *Concordia University*. 2020. <https://doi.org/10.1109/BIBM49941.2020.9313229>.
88. Thompson JD, Gibson TJ, Higgins DG. Multiple Sequence Alignment Using ClustalW and ClustalX. *Curr Protoc Bioinforma*. 2003;00:1–22. <https://doi.org/10.1002/0471250953.bi0203s00>.
89. Guo AY, Zhu QH, Chen X, Luo JC. GSDS: a gene structure display server. *Hereditas*. 2007;29:1023–6.
90. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37(SUPPL. 2):1–7. <https://doi.org/10.1093/nar/gkp335>.
91. Zhang M, Liu Y, He Q, Chai M, Huang Y, Chen F, et al. Genome-wide investigation of calcium-dependent protein kinase gene family in pineapple: Evolution and expression profiles during development and stress. *BMC Genomics*. 2020;21:1–18. <https://doi.org/10.1186/s12864-020-6501-8>.
92. Liu M, Ma Z, Sun W, Huang L, Wu Q, Tang Z, et al. Genome-wide analysis of the NAC transcription factor family in Tartary buckwheat (*Fagopyrum tataricum*). *BMC Genomics*. 2019;20:1–16. <https://doi.org/10.1186/s12864-019-5500-0>.
93. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: An information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639–45. <https://doi.org/10.1101/gr.092759.109>.
94. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40:1–14. <https://doi.org/10.1093/nar/gkr1293>.
95. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta C_T$ method. *Methods*. 2001;25:402–8. <https://doi.org/10.1006/meth.2001.1262>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.