Genome-wide analysis of the AREB/ABF gene lineage in land plants and functional analysis of TaABF3 in Arabidopsis

Fangfang Li  
Northwest Agriculture and Forestry University

Fangming Mei  
Northwest Agriculture and Forestry University

Yifang Zhang  
Northwest Agriculture and Forestry University

Shumin Li  
Northwest Agriculture and Forestry University

Zhensheng Kang  
Northwest Agriculture and Forestry University

Hude Mao (✉️ mhd163com@163.com)  
Northwest Agriculture and Forestry University  https://orcid.org/0000-0002-2484-9952

Research article

Keywords: ABFs, land plants, phylogenetic relationship, expression analysis, TaABFs, drought stress

DOI: https://doi.org/10.21203/rs.3.rs-19486/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Background: Previous studies have shown that ABFs (abscisic acid-responsive transcription factors) are important ABA-signaling components that participate in abiotic stress response. However, little is known about the function of ABFs in *Triticum aestivum*. In addition, although various ABFs have been identified in other species, the phylogenetic relationship between ABF transcription factors has not been systemically investigated in land plants.

Results: In this study, we systemically collected ABFs from land plants and analyzed the phylogenetic relationship of these *ABF* genes. The *ABF* genes are present in all the land plants we investigated, including moss, lycophyte, monocots, and eudicots. Furthermore, these *ABF* genes are phylogenetically divided into seven subgroups, differentiations that are supported by variation in the gene structure, protein properties, and motif patterns. We further demonstrated that the expression of *ABF* genes varies among different tissues and developmental stages, and are induced by one or more environmental stresses. Furthermore, we found that three wheat *ABFs* (*TaABF1*, *TaABF2*, and *TaABF3*) were significantly induced by drought stress. Compared with wild-type (WT) plants, transgenic *Arabidopsis* plants overexpressing *TaABF3* displayed enhanced drought tolerance.

Conclusions: These results provide important ground work for understanding the phylogenetic relationships between plant *ABF* genes. Our results also indicate that *TaABFs* may participate in regulating plant response to abiotic stresses.

Background

Drought is a major environmental stressor that affects plant growth, survival, distribution, and productivity. Plants have evolved complex mechanisms in molecular, cellular, and physiological processes to respond to environmental stresses in order to survive [1]. Stressful conditions induce the production of stress response genes in plants [2-3]. The phytohormone abscisic acid (ABA) is an important hormone that regulates some critical biological processes in plants, such as stomatal movement, adaptation to drought stress, and seed germination [4-7]. The endogenous ABA is produced when plants encounter adverse environmental stresses such as prolonged periods of osmotic stress. Several stress-responsive genes were expressed due to these increased ABA levels. Additional research indicates that many stress-responsive genes can also be induced by the exogenous application of ABA [2, 7-9].

ABA detects stress in a unique way and acts as an endogenous messenger in plant cells by inducing a double-negative regulatory pathway where ABA is bound to the ABA receptors RCARs/PYR1/PYLs, forming the complex that provides an active site for the PP2Cs. This inhibits the ability of PP2C to act as a negative regulator of the pathway, leading to the induction of SnRK2 as a positive regulator of downstream signalling and subsequent phosphorylation of the target proteins [10-11]. Thus, in the presence of ABA, the PP2Cs are inactivated to repress SnRK2 phosphatase activity. SnRK2 could then
initiate the ABA-responsive regulation pathway and activate the most significant cis-element ABA-responsive element (ABRE) to regulate the expression of many genes under osmotic stress conditions. Subsequently, through the yeast one-hybrid method, a subgroup of bZIP transcription factors was isolated by using ABREs as bait [12-13]. Genes of this subgroup of bZIP transcription factors primarily participated in osmotic stress response by regulating stress-related genes. In Arabidopsis, nine group-A bZIP proteins were found as homologs of AREB/ABFs, and phylogenetically divided into two subfamilies, the AREB/ABF subfamily (ABF1, ABF2/AREB1, ABF3 and ABF4/AREB2) and the ABI5/AtDPBF subfamily (AtDPBF1/ABI5, AtDPBF2, AtDPBF3/AREB3 and AtDPBF4/EEL) [14]. ABF/AREB family members have four conserved domains, including two located in the C-terminus (which includes a highly conserved bZIP domain and a C4 domain) and three located in the N-terminus (which include C1, C2 and C3 domains) [15].

To date, all of these AREB/ABF genes in Arabidopsis have been functionally characterized. These four genes (ABF1, ABF2/AREB1, ABF3 and ABF4/AREB2) are primarily expressed in vegetative tissues [12-13, 16-17]. In addition, the induced ABF1 expression changes in response to abiotic stress are minimal [18], while ABF2/AREB1, ABF3 and ABF4/AREB2 are significantly up-regulated under ABA and osmotic stresses [12-14, 17-20]. Ectopic expression of these four genes in Arabidopsis showed that ABF1 is a functional homolog of ABF2/AREB1, ABF3 and ABF4/AREB2; and ABF2/AREB1, ABF3 and ABF4/AREB2 are the core ABA signaling components responding to abiotic stresses [16-18, 20-21]. Moreover, the areb1areb2abf3 triple mutant and areb1areb2abf3abf1 quadruple mutant showed increased drought sensitivity and decreased ABA sensitivity by impairing the expression of ABA and osmotic stress-responsive genes [18, 22]. Additionally, the overexpression of many AREB/ABFs in various species have been shown to confer increased tolerance to osmotic stress [23-27]. Several studies have reported that AREB/ABF transgenic agricultural plants showed substantial increases in drought tolerance with little or no effect on growth [23, 25-26].

Bread wheat (Triticum aestivum L.) is the most widely cultivated crop on earth, accounting for approximately one-fifth of the total calories consumed by humans [28]. Consequently, wheat yields and production affect the global economy. However, its productivity is frequently hampered by water scarcity, making improved drought tolerance an important goal of many breeding programs. Although several studies have demonstrated the importance of ABFs in response to abiotic stresses, our knowledge of ABFs in wheat is still very limited. In this study, we systemically described the characteristics of plant ABFs, including gene members, phylogenetic relationships, gene structures, protein structural similarities and differences, and gene expression. We performed additional functional analyses of the wheat ABF gene TaABF3 by investigating drought tolerance in transgenic Arabidopsis plants. Our results provide an important framework for understanding the phylogenetic relationship between plant ABF genes and deepens our understanding of the function and mechanism of wheat ABF genes in responses to drought stress.

Results
Identification and analysis of AREB/ABF family in plants

Based on the 34 genomes listed in the Phytozome database, we performed a genome-wide BLAST search using Arabidopsis ABF1, AREB1/ABF2, AREB2/ABF4, and ABF3 amino acid sequences. We found the candidate ABFs in only 29 land plants, including moss, lycophyte, monocots, and eudicots. Among the ABF sequences we identified, some proteins had shorter amino acid residues (fewer than 200 amino acids). These short sequences were eliminated from subsequent analyses. In the end, 190 ABF-like sequences were collected for further analysis. We subjected these 190 protein sequences to SMART and Pfam analyses, and all of them were classified into the protein family containing bZIP domains (Pfam: 00170).

Previous studies have reported that the plant group-A bZIP family proteins can be phylogenetically clustered into two major groups, the AREB/ABF and the ABI5/AtDPBF subfamilies [14]. As such, we constructed a maximum likelihood (ML) tree, using 190 full-length ABF-like gene sequences (Additional file 1: Figure S1). Our results show that these ABF-like sequences are divided into two major clades, designated as group A and B, each having 95 identified sequences. Group A contains all experimentally characterized AREB/ABFs, including Arabidopsis, Thellungiella salsuginea, and rice ABFs [14-15]. According to previously characterized genes, such as ABI5/AtDPBF1, AtDPBF2, AREB3/AtDPBF3 and EEL/AtDPBF4 [14], group B was classified as ABI5/AtDPBF subfamilies. Therefore, group A sequences are designated ABF and were included for further analyses (Additional file 2: Table S1). The number of ABFs in each species is shown in Fig.1. In summary, the moss Physcomitrella patens and the lycophyte Selaginella moellendorffii have two copies of the ABFs. In monocots, all species contain only four copies of ABFs, with the exception of wheat, maize and Panicum virgatum. The ABF copy number differed, from one to seven, in eudicots. This indicates that several duplication incidents took place. The quantity of ABF paralogs in rice, Arabidopsis, and Thellungiella observed by this study are in line with previous research [14-15].

We further analyzed protein length, molecular mass, and the pI values of 95 ABF proteins (Additional file 2: Table S1). According to our results, the length and molecular mass of ABFs ranged from 254 to 485 amino acid residues, and 27.81 to 52.95 kD, with a mean of 389 amino acid residues and 42.16 kD. Aquilegia coerulea_ABFI is the longest and largest ABF (485 amino acid residues and 52.95 kD), while Citrus sinensis_ABFI is the shortest and smallest ABF (254 amino acid residues and 27.81 kD). Zea mays_ABFI has the lowest pI value, with 5.44, while Citrus sinensis_ABFI has the highest value, with 10.42. ABFs in clades III, VI and VII have very close pI values, while the pI values of ABFs in clades I, II, IV, and V varied widely. Interestingly, ABFs from clade V displayed a tendency to maintain acidic pI values, with an average of 6.97, while more alkaline pI values (greater than 7) were observed in 83 out of 95 ABFs belonging to other clades (Table 1; Additional file 2: Table S1).

Phylogenetic and structural analysis of plant ABFs

In order to better understand the evolutionary relationship of AREB/ABF members in land plants, we further constructed an ML tree using full-length protein sequences of 95 ABFs. According to support
values (85% or greater) of the phylogenetic tree, ABFs can be divided into seven clades (clades I to VII) (Fig. 1; Fig. 2). Inside the phylogenetic tree, ABFs from the moss *Physcomitrella patens* and the lycophyte *Selaginella moellendorffii* form two independent clades, assigned as clades I (*P. patens* ABFs) to II (*S. moellendorffii* ABFs). The monocots can be placed into the next two clades, IV and V. The eudicots can be divided into three clades: III, VI, and VII.

It is worth mentioning that the phylogenetic tree (Fig. 2) aligns with the species tree shown in Phytozome (Fig. 1) with the ABFs from moss (*Physcomitrella patens*) and lycophyte (*Selaginella moellendorffii*) forming the two basal lineages of land plants. Monocot and eudicot ABFs are closer on the phylogenetic tree and form two monophyletic clades. To further investigate the accuracy of the ABF phylogenetic tree, we analyzed the exon/intron organization for each individual gene (Additional file 3: Figure S2). Of the 95 ABFs, one has one exon; three have two exons; four have three exons; 71 have four exons, 13 have five exons, two have six exons, and two have seven exons. Within each clade, the gene structure of ABFs is relatively conserved, and the adjacent ABFs have a similar exon/intron structure. We then investigated the intron phases of all ABF gene structures. There are three categories of intron phase: phase 0 intron, phase 1 intron, and phase 2 intron. Our analysis indicated that the intron phase patterns (0, 0, 0) and (0, 0, 0, 0) are the predominant patterns across 95 land plant ABFs (Additional file 3: Figure S2). This analysis indicated that we have constructed a phylogenetic tree of the ABF genes in land plants that is highly accurate.

**Motif composition and arrangement of plant ABFs**

In order to better understand the phylogenetic relationships between plant ABFs, we aligned all of the ABF sequences to better identify the conserved amino acid residues. Based on the alignment, 35 amino acid residues are completely conserved in 88 ABFs (except for eight shorter ABFs). We further identified the conserved motifs in 95 plant ABFs using the SMART program. Finally, we found five conserved protein motifs in all ABFs, which are BRLZ domain and the other four low complexity regions (LCR 1-4; Fig. 3A; Additional file 4: Figure S3). ABFs belong to the basic-leucine zipper (bZIP) domain transcription factor family, and we found that the BRLZ domains are highly conserved in all plant sequences (Fig. 4).

However, the ability of SMART to comprehensively identify the motifs present in ABFs is limited, so we used the MEME program to identify conservation and variation in the motif arrangements among ABFs. We identified 20 distinct motifs in ABFs. The occurrences and arrangements of the motifs in ABFs from seven major clades are shown in Fig. 3B and Additional file 5: Figure S4. Among 20 motifs, 8 motifs are shared by all ABFs, which are components of the BRLZ domain (motif 1 and 2) and the other four conserved low complexity regions (motif 3 and 6 for LCR1, motif 5 for LCR2, motif 4 for LCR3, and motif 7 for LCR4). Next, we examined the non-conserved motif composition in land plant ABFs. We then split the ABFs into four regions, based on the location of the LCR motifs and the BRLZ domain (Fig. 3B): Region 1 is the part before the LCR1, Region 2 is the part between LCR1 and LCR2, Region 3 is the part between LCR3 and the BRLZ domains (there were no motifs between LCR2 and LCR3), and Region 4 is the part between the BRLZ domain and LCR4. Of these four, Regions 2 and 4 are highly conserved in plants on
land (they are mainly comprised of motifs 15 and 16). Less conserved is Region 1, which is primarily comprised of motif 11 in clades III, VI, and VII. Region 3 is the most divergent region: motif 8 was observed in clades I, III, IV, V, VI, and VII; motifs 9 and 10 were found in clades III, IV, VI, and VII; motifs 12 and 17 were found in clades III, V, VI, and VII; motifs 13 and 14 were found in clades III, VI, and VII; motifs 18 were found in clades III and VII; motif 19 was found in clades IV and VI; and motif 20 was found in clade V (Fig. 3B). Taken together, the conserved and non-conserved motif patterns of plant ABFs that we identified match the pattern of clades in the phylogenetic tree.

Expression analysis of plant ABF genes

To obtain the expression profiles of Arabidopsis ABFs, we extracted the expression data from the Arabidopsis eFP Browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi). We found that the expression of Arabidopsis ABF paralogs displayed tissue differentiation. For example, A.thaliana_ABF1 displayed significantly higher expression in roots, and A.thaliana_ABF2 displayed significantly higher expression in seeds, indicating that ABF paralogs have followed the trend of tissue subfunctionalization. We found that ABF paralogs in clade III (A.thaliana_ABF2) have higher expression levels than clade VI paralogs (A.thaliana_ABF1, A.thaliana_ABF3; Additional file 6: Figure S5A). We next investigated the expression profiles of other plant ABF genes. Our results demonstrated that soybean (Glycine max) and common bean (Phaseolus vulgaris) ABF paralogs are expressed more in leaves, roots, and flowers than in other tissues, and that ABF paralogs in clade III (G.max_ABF3, P.vulgaris_ABF2) have higher expression levels than clade VII paralogs (G.max_ABF1, G.max_ABF1, P.vulgaris_ABF1) (Additional file 6: Figure S5B and C). Within monocots, we studied the expression of the four ABF paralogs from rice (Oryza sativa), and found that O.sativa_OsABF2 (clade III) had higher expression levels than O.sativa_TRAB1, O.sativa_OsAREB1, and O.sativa_OsAREB2 (Additional file 6: Figure S5D). The five ABFs in maize (Zea mays) display similar expression patterns among tissues, except for Z.mays_ABF3, which is less expressed among all tissues (Additional file 6: Figure S5E). The expression divergence of plant ABFs indicated the functional differentiation of ABFs.

We also investigated the expression of Arabidopsis ABFs under abiotic stresses using microarray expression data. The results showed that the expression of all Arabidopsis ABFs was induced by ABA, cold temperatures, drought conditions, and high salinity, but the degrees of induction differed. A.thaliana_ABF1 was significantly induced by cold temperatures; A.thaliana_ABF2 was significantly induced by drought conditions; A.thaliana_ABF3 was significantly induced by ABA, drought conditions and salt; and A.thaliana_ABF4 was significantly induced by drought conditions and salt (Additional file 7: Figure S6). We further investigated the expression of other plant ABFs to abiotic stresses (ABA, drought and highly salinity) using quantitative real-time PCR (qRT-PCR). From the heatmap, we found that the expression of most ABFs was induced by ABA, drought conditions, and high salinity. Except for B.rapa_ABF7, G.max_ABF2, and Z.mays_ABF3, all ABFs were significantly induced by drought conditions. ABFs are known for their importance in ABA-mediated abiotic stress responses, meaning a significant induction in the ABF genes might play a crucial role in plant adaptation to environmental stresses (Fig. 5).
Molecular characterization and expression analysis of TaABFs

Phylogenetic analyses suggest that TaABFs might serve a role in regulating abiotic stress response in wheat. We cloned three TaABF genes from the wheat cv. Chinese Spring. Each gene had three homologous components in the A, B, and D genomes of wheat; we named them TaABF1-5A/B/D, TaABF2-7A/B/D, and TaABF3-6A/B/D. Additional phylogenetic analyses indicated that TaABF1 was most closely related to the rice OsAREB2, TaABF2 was most closely related to the rice OsABF2, and TaABF3 was most closely related to the rice OsAREB1 (Fig. 6A). An analysis of the protein sequence revealed that TaABFs displayed 55%-98% sequence similarity (Fig. 6B). We then analyzed the subcellular localization of TaABF3, first constructing the expression cassette and fusing TaABF3 with the GFP protein. The fused proteins were then transiently expressed in Arabidopsis protoplasts. We used fluorescence microscopy to analyze and reveal that the TaABF3-GFP fusion proteins were exclusively localized in the nucleus in the transformed cells, while the control GFP was uniformly distributed throughout the cell (Fig. 7A). These results confirmed that TaABF3 is a nuclear-localized protein.

To examine the expression pattern of TaABFs, we first identified the cis-element in its region of promotion, which was ~ 2 kb upstream of the transcription initiation codon, finding a number of cis-acting elements related to stress response in the promoter of TaABFs. This includes LTR (low temperature-responsive element), MYB (MYB recognition site), MYC (MYC recognition site), MBS (MYB binding site involved in drought-inducibility), ABRE (ABA-responsive element), and DRE (Dehydration-responsive element) (Fig. 7B). In order to better understand the role that TaABFs play in response to drought conditions, we executed quantitative real-time PCR (qRT-PCR) on RNA taken from various tissues and conditions of drought. Considering the highly sequence similarity of wheat homeologous genes, the PCR primers were designed to amplify the conserved locus of three TaABF homeologs; for example, the relative expression level of TaABF1 represents the combined expression of all three TaABF homeologs (TaABF1-5A, TaABF1-5B and TaABF1-5D). The results demonstrated that TaABFs were found in higher levels in the leaves of the seedlings (Fig. 7C) and that under drought stress conditions, all TaABFs in wheat leaves were up-regulated (Fig. 7D).

Overexpression of TaABF3 confers drought tolerance in Arabidopsis

To better understand how TaABFs function in plant abiotic stress tolerance, we generated 35S::TaABF3-GFP transgenic Arabidopsis lines. We then selected three independent transgenic lines for 35S::TaABF3-GFP transgenic Arabidopsis that exhibited higher expression levels of TaABF3 in order to further analyze their response to drought stress (Additional file 8: Figure S7). We then compared the drought tolerance of transgenic and vector-transformed (WT) plants. We grew WT and each 35S::TaABF3-GFP transgenic plants for three weeks in soil before withholding water for ~ 14d. After the drought treatment and six days of re-watering, ~ 65-75% of the transgenic plants survived, while only ~ 8% of the WT plants survived (Fig. 8A and B).

We next assayed the proline contents, malondialdehyde (MDA) contents, and the soluble sugar contents in 35S::TaABF3-GFP transgenic and WT plants (Fig. 8C-E). Our results showed that in transgenic lines the
proline contents and the soluble sugar contents were significantly higher and the MDA contents were significantly lower than in WT under both well-watered and drought conditions. We also detected the expression of several well-known drought-responsive genes in the transgenic lines, including Arabidopsis-homologous LEA14 [29], RD29A [30], DREB2A [31], RAB18 [32], RD20 [33], and GolS2 [34]. These results showed that all of these genes were up-regulated in 35S::TaABF3-GFP transgenic lines (Fig. 8F). Collectively, these findings indicate that the overexpression of TaABF3 in Arabidopsis could enhance the drought tolerance of transgenic plants.

Discussion

Transcription factors (TFs) are a group of regulatory proteins that regulate gene expression by binding to specific cis-acting elements in the promoters of target genes [35]. Despite the fact that many studies have revealed the crucial role of AREB/ABF TFs in response to abiotic stresses [16-18], our knowledge of ABFs is still limited. Previous studies have primarily focused on studying the function of ABF/AREB proteins, whereas phylogenetic studies of ABFs are restricted to some model plants, such as Physcomitrella patens, Selaginella moellendorfii, Arabidopsis, and rice [14]. To advance our understanding of the involvement of ABFs in stress response and other biological processes, it is essential to first understand their evolution and diversity. In this study, we collected most land plant group-A bZIP TFs from available genome databases (Fig. 1; Additional file 2: Table S1) and performed phylogenetic analyses with full coding sequences. This allowed us to identify the ABF clade within the bZIP TFs (Fig. 2), the intron/exon structure of genes (Additional file 3: Figure S2), and the characteristic protein domains (Fig. 3; Fig. 4; Additional file 4: Figure S3; Additional file 5: Figure S4). We next extracted expression profiles of selected plants from a public expression database and explored the functional differences of paralog genes during land plant evolution (Fig. 5; Additional file 6: Figure S5; Additional file 7: Figure S6). In addition, we systemically investigated the function of target genes of TaABF3 (Fig. 6; Fig. 7; Fig. 8). The goal of our study was to provide an overall picture of plant ABFs and deepen our knowledge of the function and mechanism of wheat ABF genes when responding to abiotic stresses.

At the protein level, ABFs in land plants share many of the same structural features, all ABFs have four conserved LCR motifs, and one BRLZ domain (Fig. 3A; Additional file 4: Figure S3). However, the differences of ABF proteins are also existed. For example, the protein structure between the LCR3 motif and the BRLZ domain exhibit the highly variable (Fig. 3B). This region requires additional research to further elucidate the differences in structure and function between the ABFs in land plants from various clades.

A robust phylogenetic tree is essential for tracing the evolutionary history of ABF genes. As sequencing techniques have advanced, increasing amounts of plant genomes have been sequenced and released. In this study, we surveyed 34 different plant genomes and collected 95 ABF genes. With the exception of algae, ABF candidates exist in all land plants, including lowland plants (a moss and a lycophyte) and highland plants (monocots and eudicots). It is increasingly apparent that gene families present in embryophytes (land plants) and absent from sequenced chlorophyte genomes have their origins in the
kind of algae from which the ancestral land plant evolved. This indicates that the ABF gene family originated during the evolution of the algal to land plants. Previous studies favor the single-origin theory of land plants, originating from charophycean green algae [36-37]. Moving from an aqueous to a gaseous environment subjects various plants to different physical conditions, which results in particular changes to their structure and physiology. Significant metabolic pathways, involving flavonoids, lignins, plant hormones, and cutins from vascular plants come from existing structures of the primary metabolism in charophycean algae [36]. During this process, various families of genes evolved and helped land plants adapt to challenging new environmental conditions, which included abiotic stressors. It is possible that ABF genes could be induced by several abiotic stresses, participating in stress response to abiotic factors [16-18]. This evolution of ABFs could have played an important role in allowing plants to adapt to conditions on land.

Our analysis of the phylogenetic relationship demonstrated that the ABF gene family underwent two changes that led to the seven distinct subfamilies (Fig. 2). The first instance happened after Selaginella moellendorfii and Physcomitrella patens diverged from a common ancestor, that of seed plants. This occurrence is consistent with the known patterns of divergence in land plants, where Selaginella moellendorfii and Physcomitrella patens are the precursors of the seed plants. Following this event, the family of ABF genes could have been limited to their historical functions. However, our phylogenetic tree shows that the second instance of duplication that resulted in lineages similar to ABF happened in seed plants. Prior research has found that a whole-genome duplication (WGD) event that occurred in an ancestor of extant angiosperms produced exact copies of each gene [38-39]. Monocots have seen many instances of WGD throughout their history, which are surely responsible for the high instance of ABF genes (Fig. 1) [40]. All of the ABFs in monocots were found in clades IV and V, which is evidence of a duplication event early in the evolutionary history of monocots. In contrast, eudicot ABFs were all found in clades III, VI, and VII. Members of clade III are paralogs of Arabidopsis ABF2, while members of clade VI are paralogs of ABF1, ABF3, and ABF4. Members of clade VII are more similar to members of clade VI but are not found in the paralogs of Arabidopsis. All of this is evidence of the functional similarity of ABFs, while the differences between the clades are indicative of functional differentiation between the clades. Recent research has shown that ABFs are involved in ABA signaling when responding to abiotic stressors [16-27], while the functional differences between ABF genes remain scarce. One study found that ABF2, ABF3, and ABF4 play important roles when regulating the mediation of ABA-triggered Chl degradation as well as leaf senescence in Arabidopsis [41]. This demonstrates evolutionary divergence in the functionality of ABFs, but in order to understand the practical differences within the lineage of ABFs, further research is required.

The responsiveness of ABF genes to abiotic stress strongly suggests that they serve roles in adapting to changing environmental conditions. Our qRT-PCR analyses further revealed that all TaABF genes were induced by drought stress (Fig. 7). To investigate the role of TaABF genes in the abiotic stress response, TaABF3 was transformed into Arabidopsis, and its overexpression was confirmed by RT-PCR (Additional file 8: Figure S7). The transgenic plants showed significantly improved drought and salt tolerance compared to WT plants (Fig. 8). Consistently, several stress-responsive genes, including LEA14, RD29A,
DREB2A, RAB18, RD20, and GolS2 were found to be significantly up-regulated in TaABF3 transgenic Arabidopsis under drought stress (Fig. 8). This research strongly indicates that TaABF3 increases the tolerance of transgenic Arabidopsis plants to drought conditions. Prior research has shown that an overexpression of TF genes can slow the growth of transgenic plants [42-45]. We also closely monitored the growth and morphological features of TaABF3 transgenic Arabidopsis plants, finding that transgenic plants exhibited a slight reduction in the size of rosette leaves (Fig. 8).

Conclusions

In summary, our study provides a comprehensive analysis of the plant ABF genes, including phylogenetic relationships, gene structures, protein structures, and properties, and expression profiles. Phylogenetic analysis combined with gene structure and motif composition clustered the plant ABFs into seven distinct clades. In addition, expression analyses demonstrate that plant ABFs have extensively induced by abiotic stress. Further functional analysis of TaABF3 transgenic Arabidopsis showed that they could confer drought tolerance in plants. Our results will help elucidate the functions of the AREB/ABF lineage in plants, and providing clues for the identification of candidate genes involved in abiotic stress responses in plants.

Methods

Plant materials and stress treatments

ABFs expression was assayed in different stress conditions from four plant species, including Brassica rapa, Glycine max, Oryza sativa, and Zea mays. Brassica rapa cv. ZS11, Glycine max cv. Jidou-7, Oryza sativa cv. Nipponbare, and Zea mays cv. B73 were identified and obtained from the Northwest A&F University and these varieties can also be obtained from the Chinese Crop Germplasm Resources Information System (http://www.cgris.net/zhongzhidinggou/index.php). The growth conditions and stress treatments are simply described as follows: 1‰ (v/v) Topsin-M sterilized seeds were germinated on wet filter paper for 3 days at 25 °C. Then the germinated seeds were hydroponic cultivated using hoagland nutrient solution in an artificial climate chamber with 16 h light/8 h dark photoperiod and temperature of 25 °C. We next applied the abscisic acid (ABA), drought, and high salinity to the cultivated seedlings separately. For high salinity and ABA treatments, the three-week-old seedlings were then transferred into 200 mmol/L NaCl and 100 μmol/L ABA culture solutions, respectively. For drought treatment, the three-week-old seedlings were placed on a clean bench and subjected to dehydration (25 °C, relative humidity of 40-60%). Under these different stress conditions, the whole seedlings were collected at 0, 2 and 10 h after treatments. At least five seedlings of each plant species in each time point were collected and the experiments were conducted for three times. All of the collected samples were froze in liquid nitrogen rapidly and stored at -80 °C refrigerator before RNA extraction.

Identification of ABF genes
We used four *Arabidopsis* ABFs (AREB1/ABF2, AREB2/ABF4, ABF1 and ABF3) to perform a TBLASTN search against the Phytozome databases (http://www.phytozome.net/). Finally, 34 plant species are used, including algae (*Chlamydomonas reinhardtii*, *Coccomyxa subellipsoidea* C-169, *Micromonas pusilla* CCMP1545, *Ostreococcus lucimarinus*, and *Volvox carteri*), the moss *Physcomitrella patens*, the lycophyte *Selaginella moellendorfii*, monocots (*Triticum aestivum*, *Oryza sativa*, *Panicum virgatum*, *Sorghum bicolor*, *Setaria italica*, and *Zea mays*), and eudicots (*Aquilegia coerulea*, *Arabidopsis lyrata*, *Arabidopsis thaliana*, *Brassica rapa*, *Capsella rubella*, *Carica papaya*, *Citrus clementina*, *Citrus sinensis*, *Cucumis sativus*, *Eucalyptus grandis*, *Fragaria vesca*, *Glycine max*, *Gossypium raimondii*, *Manihot esculenta*, *Mimulus guttatus*, *Phaseolus vulgaris*, *Prunus persica*, *Ricinus communis*, *Solanum lycopersicum*, *Thellungiella salsuginea*, *Theobroma cacao*, and *Vitis vinifera*). The cDNA, genomic DNA, and amino acid sequences corresponding to each ABF or putative ABF were downloaded from the Phytozome database. The Simple Modular Architecture Research Tool (SMART; http://smart.embl-heidelberg.de/smart/set_mode.cgi?NORMAL=1) was used to identify ABFs with the presence of bZIP and other typical domains in their protein structure. The theoretical molecular mass and Pi (isoelectric point) values were calculated using the Compute pI/Mw tool provided in ExPASy (http://web.expasy.org/compute_pi/).

**Construction of phylogenetic trees**

The coding sequences (CDS) were aligned using the TranslatorX server (http://translatorx.co.uk/) [46], and jModelTest analysis [47] was further performed to select the best-fit model. ML (maximum likelihood) phylogenetic tree was constructed using the online program RAxML (http://www.trex.uqam.ca/index.php? action=raxml&project=trex) [48] under the best-fit model with 100 bootstrap samples. The phylogenetic tree was visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

**Gene structure analysis and detection of conserved motifs**

Intron/exon distribution and intron phase patterns were analyzed using the online Gene Structure Display Server (GSDS; http://gsds.cbi.pku.edu.cn). Functional motifs of ABF proteins were identified using the Multiple Expectation Maximization for Motif Elicitation program (MEME; http://meme.nbcr.net/meme/cgi-bin/meme.cgi) with the following parameters: distribution of motifs = any number of repetitions, maximum number of motifs = 20, and optimum motif width = 6 to 100 residues.

**Expression profile of ABF genes**

Expression patterns of ABFs in different stress conditions were analyzed by qRT-PCR. Total RNA was isolated using TRIZOL reagent (Biotopped) from no less than five seedlings from three independent experiments. In order to eliminate genomic contamination, total RNA was treated with RNase-free DNase (Takara). The concentration of total RNA was determined using a Nanodrop1000 (Thermo Scientific product, USA). In order to confirm RNA integrity and quantity, 5 µg of total RNA from each sample was run on a 0.8% agarose gel. Recombinant M-MLV reverse transcriptase and 1 µg of total RNA mixed with 1 µg
Oligo (dT) 23 (Promega) were used to synthesize the cDNAs. PCR conditions consisted of an initial denaturation step at 95 °C for 10 min, followed by 40 cycles at 95 °C for 15 s, 60 °C for 30 s. TaActin (TraesCS1A01G274400) was used as the internal control. The quantification method \(2^{-\Delta Ct}\) was used and the variation in expression was estimated using three biological replicates [49]. Primers used in this study are listed in Additional file 9: Table S2.

**Subcellular localization of TaABF3-GFP fusion proteins**

*Triticum aestivum* cv. Chinese spring was identified and obtained from the Prof. Zhensheng Kang's Lab (Northwest A&F University, China) and was used to functional analysis of *TaABF3*. For subcellular localization assay of TaABF3, we amplified the full-length CDS sequence of *TaABF3* from wheat *cv*. Chinese Spring by PCR using specific primers and inserted into the binary vector pCAMV35S::GFP between *Xba*I and *BamH*I to identify the subcellular localization of TaABF3. We then identified positive clones by sequencing, while the construct was transformed into wheat mesophyll protoplasts as previously described [50]. The fluorescence of GFP was observed using a confocal microscope (Olympus, FluoViewTM FV300, Japan).

**TaABF3 isolation and Arabidopsis transformation**

*Arabidopsis* ecotype Columbia was obtained from the Prof. Zhensheng Kang's Lab (Northwest A&F University, China) and was used to transform the *TaABF3*. The full-length opening reading frame of *TaABF3* was amplified from wheat *cv*. Chinese Spring with gene-specific primers and cloned into the pGreen0029-GFP vector driven by cauliflower mosaic virus (CaMV) 35S promoter. The recombinant vector \(35S::TaABF3-GFP\) was introduced into *Agrobacterium tumefaciens* and transformed into *Arabidopsis* (*Arabidopsis thaliana*; ecotype Columbia) using the floral dip method [51]. For selection of transformants, T1 seeds were plated on MS medium containing 2% sucrose and 50 mg/mL kanamycin. Homozygous T3 plants were used for phenotypic analysis.

**Drought tolerance assay**

For the drought tolerance assays, seven-day-old transgenic *Arabidopsis* plants germinated on MS medium were transferred into pots containing 130 g 2:1 mixture of Jiffy mix and vermiculite. Thirty two-day-old plants growing under favorable water conditions (22 °C, relative humidity 60%, and 16/8 h light/dark photoperiod) were exposed to drought stress. Water was withheld from the plants for about 14 days. Watering was then resumed to allow the plants to recover. Six days later, the number of surviving plants was recorded. At least 48 plants of each line were compared with wild-type (WT) plants in each test, and statistical data were based on data obtained from three independent experiments. The student's *t*-test was used to assess the difference between wild-type and transgenic plants.

**Proline, MDA and soluble sugar measurement**
For the WT plants and transgenic plants that underwent drought stress treatment for about 10 days (most leaves begin to wilt), the contents of proline, MDA and soluble sugar were measured according to the manufacturer’s instructions of detection kits (Solarbio).

**Statistical analyses**

We repeated each experiment at least three times, and the data are presented as the mean ± standard deviation (SD) of the independent replicates. The statistical difference analysis was carried out using a Student’s *t*-test. Probability values of *P*<0.05 were designated as statistically significant, and probability values of *P*<0.01 were designated as extremely significant.

**Abbreviations**

ABA: Abscisic acid; ABFs: Abscisic acid-responsive transcription factors; ABRE: ABA-responsive element; BLAST: Basic local alignment search tool; CDS: Coding sequences; DRE: dehydration-responsive element; GEO: Gene Expression Omnibus; GO: Gene Ontology; GSDS: Gene Structure Display Server; LCR: low complexity region; LTR: low temperature responsive element; MBS: MYB binding site involved in drought-inducibility; MEME: Multiple Expectation Maximization for Motif Elicitation; ML: Maximum likelihood; MYB: MYB recognition site; MYC: MYC recognition site; Pi: isoelectric point; SMART: Simple Modular Architecture Research Tool; TFs: Transcription factors

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Availability of data and material**

The relevant data sets supporting the results of this article are included within the article and its additional files.

**Competing interests**

The authors declare that they have no conflict of interest.

**Funding**

This research was supported by Talent Fund of Northwest A&F University (grant no. Z111021602). The funders had no role in the design of the study and collection, analysis, and interpretation of data and in
writing the manuscript.

**Authors' contributions**

HM conceived and initiated the research; HM designed the experiments; FL, FM, YZ, and SL carried out the experiments. HM analyzed the data and wrote the manuscript. All authors have read and approved the final manuscript.

**Acknowledgements**

We thank reviewers for checking our manuscript and the editors for editing the paper. We would like to thank the members of the Bioinformatics Center of Northwest A & F University for their useful input.

**Authors’ information**

State Key Laboratory of Crop Stress Biology for Arid Areas, College of Plant Protection, Northwest A&F University, Yangling 712100, Shaanxi, China.

**References**

1. Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. NAC transcription factors in plant abiotic stress responses. Biochim Biophys Acta. 2012;1819(2):97-103.
2. Yamaguchi-Shinozaki K, Shinozaki K. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol. 2006;57:781-803.
3. Zhu JK. Abiotic stress signaling and responses in plants. Cell. 2016;167(2):313-324.
4. Finkelstein RR, Gampala SS, Rock CD. Abscisic acid signaling in seeds and seedlings. Plant Cell. 2002;14(Suppl):S15-S45.
5. Himmelbach A, Yang Y, Grill E. Relay and control of abscisic acid signaling. Curr Opin Plant Biol. 2003;6(5):470-479.
6. Munemasa S, Hauser F, Park J, Waadt R, Brandt B, Schroeder JI. Mechanisms of abscisic acid-mediated control of stomatal aperture. Curr Opin Plant Biol. 2015;28:154-162.
7. Hauser F, Li Z, Waadt R, Schroeder JI. SnapShot: Abscisic Acid Signaling. Cell. 2017;171(7):1708-1708.
8. Zhu JK. Salt and drought stress signal transduction in plants. Annu Rev Plant Biol. 2002;53:247-273.
9. Fujita Y, Fujita M, Shinozaki K, Yamaguchi-Shinozaki K. ABA-mediated transcriptional regulation in response to osmotic stress in plants. J Plant Res. 2011;124(4):509-525.
10. Ma Y, Szostkiewicz I, Korte A, Moes D, Yang Y, Christmann A, Grill E. Regulators of PP2C phosphatase activity function as abscisic acid sensors. Science. 2009;324 (5930):1064-1068.
11. Park SY, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow TF, Alfred SE, Bonetta D, Finkelstein R, Prokart NJ, Desveaux D, Rodriguez PL, McCourt P, Zhu JK, Schroeder JI, Volkman BF, Cutler SR. Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. Science. 2009;324 (5930):1068-1071.

12. Choi H, Hong J, Ha J, Kang J, Kim SY. ABFs, a family of ABA responsive element binding factors. J Biol Chem. 2000;275(3):1723-1730.

13. Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K. Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc Natl Acad Sci U S A. 2000;97(21):11632-11637.

14. Fujita Y, Yoshida T, Yamaguchi-Shinozaki K. Pivotal role of the AREB/ABF-SnRK2 pathway in ABRE-mediated transcription in response to osmotic stress in plants. Physiol Plant. 2013;147(1):15-27.

15. Yysotskii DA, de Vries-van Leeuwen IJ, Souer E, Babakov AV, de Boer AH. ABF transcription factors of Thellungiella salsuginea: Structure, expression profiles and interaction with 14-3-3 regulatory proteins. Plant Signal Behav. 2013;8(1):e22672.

16. Kang JY, Choi HI, Im MY, Kim SY. Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. Plant Cell. 2002;14(2):343-357.

17. Fujita Y, Fujita M, Satoh R, Maruyama K, Parvez MM, Seki M, Hiratsu K, Ohme-Takagi M, Shinozaki K, Yamaguchi-Shinozaki K. AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. Plant Cell. 2005;17(12):3470-3488.

18. Yoshida T, Fujita Y, Maruyama K, Mogami J, Todaka D, Shinozaki K, Yamaguchi-Shinozaki K. Four Arabidopsis AREB/ABF transcription factors function predominantly in gene expression downstream of SnRK2 kinases in abscisic acid signalling in response to osmotic stress. Plant Cell Environ. 2015;38(1):35-49.

19. Fujii H, Chinnusamy V, Rodrigues A, Rubio S, Antoni R, Park SY, Cutler SR, Sheen J, Rodriguez PL, Zhu JK. In vitro reconstitution of an abscisic acid signalling pathway. Nature. 2009;462(7273):660-664.

20. Furihata T, Maruyama K, Fujita Y, Umezawa T, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K. Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. Proc Natl Acad Sci U S A. 2006;103:1988-1993.

21. Wang Z, Su G, Li M, Ke Q, Kim SY, Li H, Huang J, Xu B, Deng XP, Kwak SS. Overexpressing Arabidopsis ABF3 increases tolerance to multiple abiotic stresses and reduces leaf size in alfalfa. Plant Physiol Biochem. 2016;109:199-208.

22. 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):672-685.

23. Oh SJ, Song SI, Kim YS, Jang HJ, Kim SY, Kim M, Kim YK, Nahm BH, Kim JK. Arabidopsis CBF3/DREB1A and ABF3 in transgenic rice increased tolerance to abiotic stress without stunting...
growth. Plant Physiol. 2005;138(1):341-351.

24. Chinnusamy V, Zhu JK, Sunkar R. Gene regulation during cold stress acclimation in plants. Methods Mol Biol. 2010;639:39-55.

25. Orellana S, Yañez M, Espinoza A, Verdugo I, González E, Ruiz-Lara S, Casaretto JA. The transcription factor SIAREB1 confers drought, salt stress tolerance and regulates biotic and abiotic stress-related genes in tomato. Plant Cell Environ. 2010;33(12):2191-2208.

26. Li C, Yue J, Wu X, Xu C, Yu J. An ABA-responsive DRE-binding protein gene from Setaria italica, SiARDP, the target gene of SiAREB, plays a critical role under drought stress. J Exp Bot. 2014;65(18):5415-5427.

27. Kerr TC, Abdel-Mageed H, Aleman L, Lee J, Payton P, Cryer D, Allen RD. Ectopic expression of two AREB/ABF orthologs increases drought tolerance in cotton (Gossypium hirsutum). Plant Cell Environ. 2018;41(5):898-907.

28. IWGSC. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science. 2018;361(6403).

29. Hundertmark M, Hincha DK. LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC Genomics. 2008;9:118.

30. Nakashima K, Fujita Y, Katsura K, Maruyama K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Transcriptional regulation of ABI3- and ABA-responsive genes including RD29B and RD29A in seeds, germinating embryos, and seedlings of Arabidopsis. Plant Mol Biol. 2006;60(1):51-68.

31. Sakuma Y, Maruyama K, Osakabe Y, Qin F, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Functional analysis of an Arabidopsis transcription factor, DREB2A, involved in drought-responsive gene expression. Plant Cell. 2006;18(5):1292-1309.

32. Nylander M, Svensson J, Palva ET, Welin BV. Stress-induced accumulation and tissue-specific localization of dehydrins in Arabidopsis thaliana. Plant Mol Biol. 2001;45(3):263-279.

33. Aubert Y, Vile D, Pervent M, Aldon D, Ranty B, Simonneau T, Vavasseur A, Galaud JP. RD20, a stress-inducible calleosin, participates in stomatal control, transpiration and drought tolerance in Arabidopsis thaliana. Plant Cell Physiol. 2010;51(12):1975-1987.

34. Selvaraj MG, Ishizaki T, Valencia M, Ogawa S, Dedicova B, Ogata T, Yoshiwara K, Maruyama K, Kusano M, Saito K, Takahashi F, Shinozaki K, Nakashima K, Ishitani M. Overexpression of an Arabidopsis thaliana galactinol synthase gene improves drought tolerance in transgenic rice and increased grain yield in the field. Plant Biotechnol J. 2017;15(11):1465-1477.

35. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, Kuznetsov H, Wang CF, Cobum D, Newburger DE, Morris Q, Hughes TR, Bulyk ML. Diversity and complexity in DNA recognition by transcription factors. Science. 2009;324(5935):1720-1723.

36. Kenrick P, Crane PR. The origin and early evolution of plants on land. Nature. 1997;389:33-39.

37. Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. The timescale of early land plant evolution. Proc Natl Acad Sci USA.
2018;115(10):E2274-E2283.

38. Jiao YN, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang HY, Soltis PS, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW. Ancestral polyploidy in seed plants and angiosperms. Nature. 2011;473:97-100.

39. Jiao YN, Paterson AH. Polyploidy-associated genome modifications during land plant evolution. Philosophical Transactions of the Royal Society of London, Series B. 2014;369:20130355.

40. Jiao YN, Li JP, Tang HB, Paterson AH. Integrated syntenic and phylogenetic analyses reveal and ancient genome duplication in monocots. Plant Cell. 2014;26:2792-2802.

41. Gao S, Gao J, Zhu X, Song Y, Li Z, Ren G, Zhou X, Kuai B. ABF2, ABF3, and ABF4 promote ABA-mediated chlorophyll degradation and leaf senescence by transcriptional activation of chlorophyll catabolic genes and senescence-associated genes in Arabidopsis. Mol Plant. 2016;9(9):1272-1285.

42. Sun Y, Fan XY, Cao DM, Tang W, He K, Zhu JY, He JX, Bai MY, Zhu S, Oh E, Patil S, Kim TW, Ji H, Wong WH, Rhe SY, Wang ZY. Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in Arabidopsis. Dev Cell. 2010;19(5):765-777.

43. Zhang W, Ruan J, Ho TH, You Y, Yu T, Quatrano RS. Cis-regulatory element based targeted gene finding: genome-wide identification of abscisic acid- and abiotic stress-responsive genes in Arabidopsis thaliana. Bioinformatics. 2005;21(14):3074-3081.

44. Lu Z, Yu H, Xiong G, Wang J, Jiao Y, Liu G, Jing Y, Meng X, Hu X, Qian Q, Fu X, Wang Y, Li J. Genome-wide binding analysis of the transcription activator ideal plant architecture1 reveals a complex network regulating rice plant architecture. Plant Cell. 2013;25(10):3743-3759.

45. Birkenbihl RP, Kracher B, Somssich IE. Induced genome-wide binding of three Arabidopsis WRKY transcription factors during early MAMP-triggered immunity. Plant Cell. 2017;29(1):20-38.

46. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations Nucleic Acids Res. 2010;38:W7-13.

47. Posada D. jModelTest: phylogenetic model averaging. Mol Biol Evol. 2008;25(7):1253-1256.

48. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688-2690.

49. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCt Method. Methods. 2001;25(4):402-408.

50. Yoo, S. D., Cho, Y. H., Sheen, J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat Protoc. 2007;2(7):1565-1572.

51. Clough SJ, Bent AF. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 1998;16(6):735-743.

Tables

Table 1 Summary of ABF protein properties.
| Clade | Total No. of ABFs | Protein length (aa) | Molecular Mass (KD) | Theoretical pI | No. of ABFs with pI > 7 |
|-------|------------------|---------------------|--------------------|----------------|-----------------------|
| I     | 2                | 407±60.81           | 42.45±6.76         | 8.60±3.03      | 2                     |
| II    | 2                | 328±101.82          | 35.39±7.72         | 8.05±1.61      | 2                     |
| III   | 23               | 394±55.16           | 42.51±5.82         | 9.63±0.34      | 23                    |
| IV    | 14               | 343±13.91           | 36.99±1.51         | 8.69±1.32      | 12                    |
| V     | 15               | 344±20.08           | 36.67±2.21         | 6.97±1.03      | 5                     |
| VI    | 19               | 417±32.06           | 45.62±3.48         | 9.10±0.66      | 18                    |
| VII   | 21               | 421±24.87           | 46.01±2.68         | 9.40±0.63      | 21                    |
| Minimum | -                | 254                | 27.81              | 5.44           | -                     |
| Maximum | -                | 485                | 52.95              | 10.42          | -                     |

**Additional Files**

**Additional file 1: Figure S1 Phylogenetic relationship of group-A bZIP TFs from 29 plant species.** The land plant group-A bZIP TFs are grouped into two major clades, designated as AREB/ABF and ABI5/AtDPBF subfamilies.

**Additional file 2: Table S1 ABF protein properties and their clade-distributions in land plants.** Gene identifiers are obtained from Phytozome database.

**Additional file 3: Figure S2 Schematic diagram of gene structures of 95 plant ABFs.** The thin lines represent introns and thick bars represent exons. The numbers above the gene structure indicate intron phases. A scale bar with a unit of base pair (bp) is graphed on the bottom.

**Additional file 4: Figure S3 Alignment of 95 plant ABF protein sequences.** The alignment was generated using ClustalW implemented in Geneious software and represented as thick lines (aligned characters) and thin lines (gaps). Overall alignment identity and a scale bar indicating the numbers of amino acid residues are graphed on the top.

**Additional file 5: Figure S4 Combined motif diagram of 95 ABF proteins.** Thick lines represent the ABF proteins. Different colored boxes represent separate and distinct motifs identified using MEME program. A scale bar indicating the numbers of amino acid residues is shown on the top. Motifs are drawn approximately to scale as boxes.

**Additional file 6: Figure S5 Gene expression profile of ABF paralogs in plants.** Gene expression data was extracted from *Arabidopsis thaliana* (http://jsp.weigelworld.org/expviz/expviz.jsp), soybean (*Glycine*...
max, \( http://soybase.org/soyseq/ \), common bean (\textit{Phaseolus vulgaris}, \( http://plantgrn.noble.org/PvGEA/SearchVisual.jsp \)), maize (\textit{Zea mays}, \( http://www.plexdb.org/index.php \)), and rice (\textit{Oryza sativa}, \( http://www.plexdb.org/index.php \)).

Additional file 7: Figure S6 Gene expression profile of \textit{ABF} paralogs in \textit{Arabidopsis thaliana} under different abiotic stresses. The mean-normalized expression values were obtained from the AtGenExpress microarray database via the web \( http://jsp.weigelworld.org/expviz/expviz.jsp \).

Additional file 8: Figure S7 RT-PCR analysis of \textit{TaABF3} transcription levels in the transgenic \textit{Arabidopsis} lines.

Additional file 9: Table S2. Primers used in this research.

\section*{Figures}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
Species & Family & Clade & Copy Number & Total \\
\hline
\textit{Arabidopsis thaliana} & Brassicaceae & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \\
\textit{Arabidopsis lyrata} & Brassicaceae & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 5 \\
\textit{Capsella rubella} & Brassicaceae & 0 & 1 & 0 & 4 & 0 & 0 & 0 & 4 \\
\textit{Brassica rapa} & Brassicaceae & 0 & 1 & 0 & 0 & 6 & 0 & 7 \\
\textit{Thellungiella salsuginea} & Brassicaceae & 0 & 1 & 0 & 0 & 3 & 0 & 4 \\
\textit{Carica papaya} & Caricaceae & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\textit{Gossypium raimondii} & Malvaceae & 0 & 2 & 0 & 0 & 0 & 0 & 3 & 5 \\
\textit{Theobroma cacao} & Malvaceae & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 2 \\
\textit{Citrus sinensis} & Rutaceae & 0 & 2 & 0 & 0 & 0 & 0 & 1 & 3 \\
\textit{Citrus clementina} & Rutaceae & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 2 \\
\textit{Eucalyptus grandis} & Myrtaceae & 0 & 1 & 0 & 0 & 0 & 0 & 2 & 3 \\
\textit{Manihot esculenta} & Euphorbiaceae & 0 & 2 & 0 & 0 & 0 & 0 & 1 & 3 \\
\textit{Ricinus communis} & Euphorbiaceae & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 2 \\
\textit{Glycine max} & Fabaceae & 0 & 2 & 0 & 0 & 0 & 0 & 2 & 4 \\
\textit{Phaseolus vulgaris} & Fabaceae & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 2 \\
\textit{Cucumis sativus} & Cucurbitaceae & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 2 \\
\textit{Fragaria vesca} & Rosaceae & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 2 \\
\textit{Prunus persica} & Rosaceae & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 2 \\
\textit{Vitis vinifera} & Vitaceae & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 4 \\
\textit{Solanum lycopersicum} & Solanaceae & 0 & 1 & 0 & 0 & 0 & 0 & 3 & 4 \\
\textit{Aquilegia coerules} & Ranunculaceae & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\textit{Zea mays} & Poaceae & 0 & 0 & 2 & 3 & 0 & 0 & 5 \\
\textit{Sorghum bicolor} & Poaceae & 0 & 0 & 2 & 3 & 0 & 0 & 5 \\
\textit{Setaria italica} & Poaceae & 0 & 0 & 2 & 2 & 0 & 0 & 4 \\
\textit{Panicum virgatum} & Poaceae & 0 & 0 & 4 & 3 & 0 & 0 & 7 \\
\textit{Triticum aestivum} & Poaceae & 0 & 0 & 2 & 1 & 0 & 0 & 3 \\
\textit{Oryza sativa} & Poaceae & 0 & 0 & 2 & 2 & 0 & 0 & 4 \\
\textit{Selaginella moellendorffii} & Selaginellaceae & 0 & 2 & 0 & 0 & 0 & 0 & 2 \\
\textit{Physcomitrella patens} & Funariaceae & 2 & 0 & 0 & 0 & 0 & 0 & 2 \\
\hline
\end{tabular}
\end{table}

\textbf{Figure 1}
Number of ABF paralogs in each species and their clade distributions. The species tree is based on information in Phytozome (http://www.phytozome.net). The star on the branch point within eudicot species indicates the divergence point between a basal eudicot (A. coerulea) and the core eudicots.

**Figure 2**

Phylogenetic relationship of plant ABF genes. ABFs are grouped into seven distinct clades (I-VII). Numbers above branches represent the support values.
Figure 3

Sequence alignment and motif composition of plant ABFs. A Alignment of protein sequences in representative plant ABFs. The conserved domains are annotated as red lines (LCR1-LCR4, BRLZ domain). The alignment was generated using ClustalW implemented in the Geneious software and represented as thick lines (aligned characters) and thin lines (gaps). B Motif patterns in representative ABFs. Motif occurrences were predicted using the MEME program, and the different colored and numbered boxes represent separate and distinct motifs. Stars indicate the BRLZ domain and four different colored arrows indicate the four conserved motifs (LCR1-LCR4).
Figure 4

The sequences of the BRLZ domain from ABFs are conserved across land plants. A heat map diagram of the protein sequence alignment of BRLZ domain from land plants. B ICE LOGO of the BRLZ domain in the N-terminus of ABF shows that the motif is strictly conserved in land plants.
Figure 5

Expression profiles of representative plant ABF genes under different abiotic stresses. The relative expression levels of ABFs were determined using quantitative real-time RT-PCR in the leaves of three-leaf-stage seedlings treated with drought, NaCl, and ABA, compared with the control.
Figure 6

Phylogeny, subcellular localization, and expression of TaABFs. A Phylogenetic relationship between TaABFs and ABF members from other plant species. The phylogenetic tree was constructed by MEGA6.0 using the neighbor-joining method. The numbers at each node indicate the percentage of bootstrap values from 1000 replicates. B Protein sequence alignment of TaABFs. The locations of the highly conserved BRLZ domain was indicated by black lines.
Figure 7

Molecular characterization and expression analysis of TaABFs. A Subcellular localization of TaABF3 protein. 35S:GFP was used as positive control. B Distribution of several stress-related cis-elements in the promoter region (~2.0 kb) of TaABFs. LTR, low temperature responsive element; ABRE, ABA-responsive element; DRE, dehydration-responsive element; MBS, MYB binding site involved in drought-inducibility; MYB, MYB recognition site; MYC, MYC recognition site. C The expression profiles of TaABFs in different tissues. R, root of wheat seedling at five-leaf stage; S, stem of wheat seedling at five-leaf stage; L, leaf of wheat seedling at five-leaf stage; FL, flag leaf at heading stage; YS5, young spike at early booting stage; GR5, grain of 5 days post-anthesis; GR15, grain of 15 days post-anthesis. D The expression pattern of
TaABFs under drought stress treatment. The numbers on X axis indicate the time point subject to drought stress. The error bars indicate standard deviations derived from three independent biological experiments.

**Figure 8**

Drought tolerance assay of TaABF3 overexpression transgenic Arabidopsis plants. A Performance of TaABF3 overexpression Arabidopsis plants under drought stress. B Statistical analysis of survival rates after the drought stress treatment. The average survival rates and standard errors were calculated based on data obtained from three independent experiments. C-E Proline contents (C), malondialdehyde (MDA) contents (D) and soluble sugar contents (E) of the WT and transgenic lines before and after drought treatment. F qRT-PCR analysis of the selected marker genes involved in water response. TaActin was used as the internal control. WW, well-watered; DT, drought stress.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.
• Additionalfile2TableS1.xlsx
• Additionalfile1FigureS1.tif
• Additionalfile9TableS2.xlsx
• Additionalfile8FigureS7.tif
• Additionalfile4FigureS3.tif
• Additionalfile6FigureS5.tif
• Additionalfile7FigureS6.tif
• Additionalfile5FigureS4.tif
• Additionalfile3FigureS2.tif