

Genome-wide identification and expression analysis of bZIP transcription factors under abiotic stress in alfalfa (*Medicago sativa* L.)

Atit Parajuli

Washington State University

Karen A. Sanguinet

Washington State University

Zhiwu Zhang (✉ zhiwu.zhang@wsu.edu)

Washington State University

Research Article

Keywords: Alfalfa, bZIP transcription factor, phylogenetic analysis, expression pattern, abiotic stress

Posted Date: March 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1405748/v1>

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

[Read Full License](#)

Abstract

Background

Alfalfa (*Medicago sativa* L.) is the most cultivated forage legume around the world. Under a variety of growing conditions, forage yield in alfalfa is stymied by biotic and abiotic stresses including heat, salt, drought, and disease. Given the sessile nature of plants, they use strategies such as differential gene expression to respond to environmental cues. Transcription factors control the expression of genes that contribute to or enable tolerance and survival during periods of stress. Basic-leucine zipper (bZIP) transcription factors have been demonstrated to play a critical role in regulating plant growth and development as well as mediate the responses to abiotic stress in several species, including *Arabidopsis thaliana*, *Oryza sativa*, *Lotus japonicus* and *Medicago truncatula*. However, there is little information about bZIP transcription factors in cultivated alfalfa.

Result

In the present study, 237 bZIP genes were identified in alfalfa from publicly available sequencing data. Multiple sequence alignments showed the presence of intact bZIP motifs in the identified sequences. Based on previous phylogenetic analyses in *Arabidopsis thaliana*, alfalfa bZIPs were similarly divided and fell into 10 groups. The physico-chemical properties, motif analysis and phylogenetic study of the alfalfa bZIPs revealed high specificity within groups. The differential expression of alfalfa bZIPs in a suite of tissues indicates that bZIP genes are specifically expressed at different developmental stages in alfalfa. Similarly, expression analysis in response to ABA, cold, drought and salt stresses, indicates that a subset of bZIP genes are also differentially expressed and likely play a role in abiotic stress signaling and/or tolerance.

Conclusions

Taken together, this work provides a framework for the future study of bZIPs in alfalfa and presents candidate bZIPs involved in stress-response signaling.

Introduction

Alfalfa (*Medicago sativa* L.) is a perennial and highly outcrossing forage legume crop grown predominantly for hay, silage, and pasture. It is the most widely cultivated forage legume in the United States with approximately 16 million hectares planted (1). The high nutritional value of alfalfa with about 15–22% crude protein and an abundance of vitamins and minerals makes it well suited for animal and livestock feed. Alfalfa also brings long-term ecological benefits to society by improving soil fertility through its symbiotic association with the soil bacterium *Sinorhizobium meliloti* for atmospheric nitrogen fixation, which augments the nitrogen content in the soil for future crops (2–4). The perennial nature of alfalfa along with its deep root system (up to 15 m) helps to prevent soil erosion. However, genetic improvement in terms of forage yield has been relatively stagnant in alfalfa (1).

In fact, the high out-crossing nature, genomic complexity, severe inbreeding depression upon selfing and self-incompatibility complicates alfalfa breeding. Although the multi-purpose use of alfalfa is increasing in demand, production is hindered by changing environmental conditions leading to abiotic stresses like heat, drought, and salinity. In the context of stagnant genetic improvement, cultivation of stress-resilient alfalfa germplasm will help to improve production in response to climate change. However, identification of stress-resilient germplasm requires identification of stress-responsive genes, which, in alfalfa, are very few due to incomplete genomic information and limited expression profile data.

The sessile nature of plants inevitably exposes them to adverse environmental conditions such as abiotic stress. Plants have developed diverse mechanisms to cope with these abiotic stresses. One mechanism is the synthesis of proteins, metabolites, and other compounds to aid in survival through abiotic stress, which are often controlled by transcription factors (TFs). Transcription factors play a critical role in responses to environmental stresses via binding to cis-regulatory elements in promoters to regulate downstream gene expression. In plants, approximately 7% of the genome codes for transcriptional regulators, which bind promoter elements of downstream genes through their conserved sequence specific DNA-binding domain (5). Among the 64 families (6) of transcription factors identified in the plant kingdom, the bZIP (basic leucine zipper) family is one of the largest and most diverse (6–8).

The basic leucine zipper (bZIP) family is distinguished by its highly conserved bZIP domain composed of 60–80 amino acids (9). Structurally, the bZIP domain is divided into two functionally distinct regions: a basic region and a leucine zipper motif (9). The basic region is composed of an invariant motif (N-x7-R/K-x9) of 18 amino acids residues that facilitates sequence-specific DNA binding, while the leucine zipper contains several heptad repeats of leucine or other bulky hydrophobic amino acids such as isoleucine, valine, phenylalanine, or methionine, for dimerization specificity (7, 10). Molecular studies of bZIP genes in *Arabidopsis thaliana* show that they are involved in the regulation of diverse biological processes including pathogen defense, light and stress signaling, seed maturation and flower development (10). Additional information on the bZIP transcription factor family has provided evidence of their role in response to biotic and abiotic stresses in a diversity of plant species (10, 11).

The availability of whole genome sequences for plants allows the identification or prediction of bZIP TF family members at the genome-wide level. The number of bZIP TFs identified in different plant and crop species varies from 78 (AtbZIPs) in *Arabidopsis thaliana* (8), 89 (OsbZIPs) in *Oryza sativa subs. japonica* (7), 125 (ZmbZIPs) in *Zea mays* (11), 131 (GmbZIPs) in *Glycine max* (12), 92 (SbbZIPs) in *Sorghum bicolor* (13), 55 (VvbZIPs) in *Vitis vinifera* (14), 64 (CsbZIPs) in *Cucumis sativus* (15) and 247 (BnbZIPs) in *Brassica napus* (16). The bZIP transcription factors play crucial roles in developmental processes and environmental tolerance in response to multiple stresses. They are involved in the regulation of seed development (17, 18), cell elongation (19, 20), vascular development (20), flower development (21–24), somatic embryogenesis (25), as well as in nitrogen/carbon and energy metabolism (26–28).

In addition to functions in plant growth and development, bZIPs also play an important role in responses to abiotic and biotic stresses. Several bZIPs from *A. thaliana* (AtbZIP17, AtbZIP24, AtbZIP12), rice

(OsZIP12, OsZIP72, OsABF1) and soybean (GmZIP44, GmZIP62, GmZIP78) were found to positively regulate salt stress adaptation in plants either directly or indirectly (12, 29–34). Several bZIPs from rice (OsZIP52/RISBZ5, OsZIP16, OsZIP23, OsZIP45, AREB1, AREB2, ABF3) were also found to be involved in drought tolerance (35–38). OsZIP52/RISBZ5 negatively regulates cold stress responses (36) while OsZIP72 was a positive regulator of ABA responses (33). Similarly, overexpression of *GmZIP44*, *GmZIP62* and *GmZIP78* reduced ABA sensitivity (12). Interestingly, group D or so-called TGA bZIPs play a role in systemic acquired resistance (SAR) and pathogen resistance (39, 40). However, there is little published information about the bZIP transcription factor family in cultivated alfalfa and its role in stress resistance.

With the availability of a chromosome-level genome assembly in alfalfa (41) we conducted a genome-wide search to identify and characterize the alfalfa bZIP transcription factors. Since bZIP transcription factors were identified to play significant roles in the regulation of abiotic stress tolerance (10, 11), we speculated various bZIP transcription factors would be differentially expressed throughout distinct developmental stages and abiotic stresses in alfalfa as well. The present study identifies several bZIPs from a proteomic database in tetraploid alfalfa (*Medicago sativa*). We also analyzed differential gene expression from transcriptome sequences during ABA, drought, salt, and cold stress conditions. This study will facilitate functional analysis of the bZIP transcription factor family in alfalfa. The identification of functions of alfalfa bZIP transcription factors during abiotic stress conditions will further help breeding efforts for improved stress tolerance.

Materials And Methods

Identification of bZIP transcription factor gene family in alfalfa

For comprehensive identification and analysis of the bZIP transcription factor (TF) gene family in alfalfa, the sequences of bZIP transcription factors from model and known species were downloaded from the Plant Transcription factor database (<http://planttfdb.cbi.pku.edu.cn/>), which included 127 sequences from *Arabidopsis thaliana*, 93 from *Lotus japonicus*, 124 from *Medicago truncatula* and 140 from *Oryza sativa*. The number of bZIPs used were more than that is mentioned in Table 1 as it included spliced variants as well. A local protein database was created using Basic Local Alignment Search Tool (54) with protein sequences from chromosome level assembly of alfalfa (41). A BLASTp search was conducted in the local database created using the protein sequences from alfalfa, taking the bZIP sequences from model organisms as a query with an E-value cut-off of 1E-05 (0.00001). The bZIP sequences obtained from the search were further confirmed based on the presence of the bZIP domain (N-x(7)-R/K-x(9)-L-x(6)-L-x(6)-L) using the Pfam web program (<https://pfam.xfam.org/>) with an E-value of 1.0. Further, the bZIP domain was used to search against the database of the identified bZIP sequences using the Prosite program of the ExPASy bioinformatics resource (<http://prosite.expasy.org>). The identified sequences with intact bZIP domains were predicted to be bonafide bZIP sequences.

Evolutionary analysis, protein properties and detection of conserved motifs in the bZIPs

To analyze the sequence features of bZIP transcription factors, multiple sequence alignment of 237 bZIP proteins were performed using multiple sequence comparison by log-expectation (MUSCLE) (55) command using default parameters. The output of the multiple sequence alignment was visualized using Unipro UGENE v.33. (56) For evolutionary analysis, 549 sequences were used which included sequences from *Medicago sativa* (237), *Arabidopsis thaliana* (78), *Lotus japonicus* (70), *Medicago truncatula* (75) and *Oryza sativa* (89). Multiple sequence alignment was carried out by CLUSTALW with default parameters. Subsequently, the phylogenetic tree was constructed by the Neighbor Joining method using 1000 bootstraps replicates. Phylogenetic analyses were conducted using MEGA version X (57).

Other physical properties of the identified sequences like the molecular weight and theoretical isoelectric point (pI) were determined using Compute pI/Mw tools (http://web.expasy.org/compute_pi/) of ExPASy bioinformatics resource. The MEME (44) program was used to identify the conserved motifs within the full-length Alfalfa. The parameters used were maximum number of motifs to be 25, distribution of motifs = any number of repetitions, optimum motif width = 6 to 50 residues.

In silico functional analysis of bZIP genes

For predicting the MsbZIP protein function (gene ontology) GO annotation was performed using the web-accessible Blast2GO v4.1 annotation system (<https://www.blast2go.com/>) (58). Briefly, the MsbZIP protein sequences were used to search for similar sequences against the NCBI non-redundant (Nr) database using the Blast tool in the Blast2GO software, with an E-value of 10^{-3} (1e-03). Next, mapping and annotation were performed on Blast2GO using default parameters. Finally, functional classification was also performed by Blast2GO.

Expression analysis during plant development

The raw RNA sequence data was downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA), SRP055547 (45). The data was generated from six tissues at different growth stages of Alfalfa namely, root, nodule, elonged stem, pre-elonged stem, leaf, and flower. The tissue sample for RNA-Seq was collected at the respective stage of alfalfa plants. Fastqc version 0.11.7 was used for quality check of the raw sequences. The reads passing the minimum Phred quality score of 30 were selected. The RNA-Seq analysis was carried out following the method described by (59), in which the filtered reads were aligned with the reference genome using HISAT2 version 2.1.0 (60) and sorted by Samtools ver 1.9. Transcript assembly and quantification was carried out using Stringtie version 2.1.1 (61). A python script was used to extract read count information directly from the files generated from Stringtie and edgeR package (62) in R was used for differential gene expression analysis. TBtools version 1.0692 (63) was used to generate heatmaps for the differentially expressed genes.

Transcriptome analysis of bZIP genes in response to abiotic stresses

The raw RNA sequence data from previous studies were downloaded from the National Center for Biotechnological Information (NCBI) Sequence Read Archive (SRA). The transcriptome data consist of cold treatment (SRR7091780-SRR7091794, (64, 65), and ABA, drought, and salt treatments (SRR7160313-SRR7160357, (64, 66)). All these samples were collected from 12 days old alfalfa seedlings for RNA-Seq. Fastqc version 0.11.7 was used for quality check of the raw sequences. The reads passing the minimum Phred quality score of 40 were selected. The RNA-Seq analysis was carried out following the method described by (59), in which the filtered reads were aligned with the reference genome using HISAT2 ver2.1.0 and sorted by Samtools ver1.9. Transcript assembly and quantification was carried out using Stringtie version 2.1.1. A python script was used to extract read count information directly from the files generated from Stringtie and edgeR package in R was used for differential gene expression analysis. TBtools version 1.0692 was used to generate heatmaps for the differentially expressed genes.

Analysis of cis-regulatory elements

For this analysis, the bZIP genes with changed expression during abiotic stress were visualized using Integrated Genome Browser 9.1.4 (67) to locate the promoter sequences. Samtools (ver. 1.9) was used to extract the 2000 base pair sequence from the promoter of these changed bZIP genes to investigate the potential cis-regulatory elements by querying them through the PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>). In total six cis-regulatory elements responsive to stress were analyzed. These elements included abscisic acid responsive (ABRE), methyl jasmonate responsive (CGTCA-motif), light inducible G-box motif, low-temperature responsive (LTR), drought responsive (MBS binding site) and defense and stress responsive (TC-rich repeats).

Results

Identification of the alfalfa bZIP gene family

We identified 237 bZIP sequences with the intact bZIP domain in alfalfa (*Medicago sativa*). These sequences were named MsbZIP1 to MsbZIP237 based on the order identified in the protein sequence database (41). We compared the genome size and number of bZIPs in different models and crop species (Table 1). The comparison shows that alfalfa has the highest number of bZIP sequences. Since the diploid model legume *Medicago truncatula* with genome size of 390 Mega Base (Mb) has 75 bZIP sequences, tetraploid alfalfa is expected to have double the number of bZIP sequences. Not surprisingly, the number of bZIP TFs identified in alfalfa was 237, which likely represents the complete number of bZIP for tetraploid alfalfa.

Table 1
Comparative genome size and number of bZIP proteins in different model crops used in the study

Species	Chromosome	Genome Size	Number of bZIPs
<i>Arabidopsis thaliana</i> (8,10)	$2n = 2x = 10$	135 Mb	78
<i>Brassica napus</i> (16)	$2n = 2x_1 + 2x_2 = 38$	1.16 Gb	247
<i>Lotus japonicus</i> (6)	$2n = 2x = 12$	470 Mb	70
<i>Medicago truncatula</i> (42)	$2n = 2x = 16$	390 Mb	75
<i>Oryza sativa</i> (7)	$2n = 2x = 20$	430 Mb	89
<i>Medicago sativa</i> (Current Study)	$2n = 4x = 32$	3,150 Mb	237

Multiple Sequence Alignment and Phylogenetic Analysis

To identify common conserved domains amongst the sequences, we carried out multiple sequence alignment. The alignment of 237 bZIP protein sequences showed the presence of intact and highly conserved bZIP domains (N-x(7)-R/K-x(9)-L-x(6)-L-x(6)-L) (Fig. 1, **Supplementary Fig. 1**). The domain is divided into the basic region with ~ 18 amino acids residues containing nuclear localization signal followed by an intact N-x(7)-R/K motif while the leucine zipper region contains heptad repeats of leucines or other bulky hydrophobic amino acids with nine amino acids towards the C-terminus (10). The presence of the intact bZIP domain further validates the identified sequences as bZIP proteins. The identified 237 bZIP proteins were divided into 10 groups (A, C, D, E, F, G, H, I, M, and S) based on the topology of the tree developed in *Arabidopsis thaliana* (8, 10) and were used to generate a phylogenetic tree along with protein sequences from *Arabidopsis thaliana*, *Lotus japonicus*, *Medicago truncatula* and *Oryza sativa* (Fig. 2). Alfalfa bZIP proteins fell into 10 different groups and numbers ranged from four (H) to forty-three (A); however, no members were identified for groups B, J and K.

Conserved protein domain analysis

Identification of conserved protein motifs helps to elucidate protein functions and bZIPs usually possess additional conserved motifs that could provide sites for activation (43). Using the “MEME” (Multiple Em for Motif Elicitation) program (44), 25 conserved motifs were identified in the 237 bZIPs (**Supplementary Table 2**). Among the identified motifs, the basic region of the bZIP, containing an invariant motif (N-x7-R/K-x9) with 18 amino acid residues was found (Fig. 3A), while the leucine zipper region that contains the

heptad repeat of leucine or other bulky hydrophobic amino acids was also identified (Fig. 3B). The basic region facilitates sequence specific DNA binding whereas the leucine zipper region is important for dimerization specificity. However, the function of the 23 motifs that were also identified in the bZIP sequences are unknown and require further study.

In silico functional classification of MsbZIP transcription factors

Among the 237 MsbZIP, 21 GO (Gene Ontology) categories were assigned to 203 of the MsbZIPs identified (Fig. 4). The major molecular functions of these bZIPs were DNA-binding transcription factor activity, which is consistent with their demonstrated role as transcription factors in other species. In the biological process category, most of bZIPs were assigned to the regulation of transcription category and almost all these proteins were predicted to localize to the nucleus in the cellular component category. Transcription factors provide binding sites through which they can regulate gene expression. They may act as either positive or negative regulators of downstream genes depending upon the environmental condition. The current functional classification (GO terms) of these bZIP proteins further supports their regulatory nature.

Tissue-specific expression profile analysis of alfalfa bZIPs

After analysis of publicly available RNA-Seq data (45), we found differential expression of 177 bZIP genes. These genes were selected for having expression values in at least one of the tissues: stem, flowers, leaves, root nodules, roots, and pre-elongated stems (PES). They were then displayed in a heatmap to visualize the expression profile in different tissues and organs (Fig. 5). Differential gene expression was observed for different developmental stages. Most of the genes were highly expressed in nodules and roots. Apart from nodules and roots, genes that were upregulated in one developmental stage were downregulated in other developmental stages which can be observed in the heatmap. Even within a group, the genes were differentially expressed across all developmental stages suggesting different *bZIP* genes are required for growth and development at different stages.

Alfalfa bZIP genes are differentially expressed in response to abiotic stresses

Analysis from the publicly available RNA-seq datasets showed differential expression of 146 genes during ABA, drought, and salt stress as well as 152 bZIP proteins during cold stress at 0, 2, 6, 24, and 48h, respectively. The expression pattern of *MsbZIP* genes during different abiotic stress conditions of cold, ABA, drought and salt showed differential expression. Across different time points of abiotic stress, the expression was different for different genes and even within a group the genes were expressed differently for different abiotic stress. Among 4 different time points of cold treatment (2h, 6h, 24h and 48 h), different genes were upregulated at different time points (**Supplementary Fig. 2**). Even within a group, at

different time points, different genes were upregulated and downregulated at different intervals of cold treatment. Like the cold treatment, abiotic stress of ABA, drought and salt treatment also showed multiple genes upregulated at different time points of stress treatment (Fig. 6). However, no genes were actively expressed during different time points of the same treatment condition among ABA, drought, and salt, which indicates different transcription factors are active during different abiotic stress as well as different time points of stress.

Cis-regulatory elements in bZIP gene promoter

The expression pattern of stress responsive-genes are often controlled by cis-regulatory elements. These elements are typically located 5' upstream of the gene coding sequences. These elements provide a binding site for the transcription factors to switch on or off the gene based on the environmental condition. In this study, we analyzed 135 stress-responsive bZIP promoters, we identified 875 cis-regulatory elements distributed along these 135 bZIP promoter. The detailed distribution of these cis-elements along the bZIP promoters was performed (**Supplementary Fig. 3**). We focused on cis-elements implicated in abiotic stress responses and found an abundance of the following cis-regulatory elements: abscisic acid responsive element (ABRE), methyl jasmonate responsive motif (CGTCA-motif), light inducible G-box motif, low-temperature responsive (LTR), drought responsive (MBS binding site) and defense and stress responsive (TC-rich repeats). Among the 875 cis-elements, light inducing G-box motif was the highest with 274 followed by abscisic stress responsive element (ABRE) with 234 while low temperature responsive (LTR) with 50 was the lowest.

Discussion

In the present study, we identified 237 bZIP sequences from tetraploid alfalfa that contained both a highly conserved basic region and the heptad repeat leucine zipper region, suggesting they are functional bZIPs. As predicted, the number of bZIP in tetraploid alfalfa (237) is more than double to that of diploid model legume *Medicago truncatula* (75). Not surprisingly, the number of bZIP genes varied amongst plant species with *Arabidopsis thaliana* (78), *Lotus japonicus* (70), *Medicago truncatula* (75) and *Oryza sativa* (89) (6, 10, 42, 46, 47). Similarly, the allotetraploid *B. napus* genome contained 247 bZIP genes, which is roughly double that of the number found in the related diploid *A. thaliana*.

Based on phylogenetic analysis and previous analyses from *A. thaliana*, *Medicago truncatula*, *Lotus japonicus* and *Oryza sativa*, we classified the alfalfa bZIP genes into 10 groups (A, C, D, E, F, G, H, I, M, and S). The most recent classification of bZIPs from *A. thaliana* (8) sorted AtbZIPs into 13 groups. Notably, groups B, J and K are missing in our analysis of alfalfa. In *A. thaliana* there are three members of group B (bZIP17, bZIP28, and bZIP49) and one group K member (bZIP60), which are implicated in endoplasmic reticulum stress responses (48), but both these groups are missing in alfalfa which begs the question of which groups perform this function in alfalfa. Group J in *A. thaliana* is made up of a single copy gene, bZIP62, which is related to Group G bZIP GBF1 – a negative regulator of blue-light responsive hypocotyl growth that acts antagonistically to HY5 and HYH, two group H bZIPs important in photomorphogenic

growth (49). Another remarkable difference between groups is the group M bZIP72, which is single copy in *A. thaliana* but contains 13 members in alfalfa. It will be interesting to determine the role M group bZIPs play in alfalfa and it is intriguing to postulate why this group has increased in number.

It is well established that bZIP transcription factors have a myriad of roles in plant development such as seed maturation and germination (18), floral induction and development (21, 24). Not surprisingly, tissue-specific expression of 177 *bZIP* genes in nodules, flowers, roots, leaves, and stems was found in alfalfa as well (Fig. 5). Interestingly, group E members were most specifically expressed in stems, roots, and flowers, whereas several group F members were expressed in pre-elongated stems. In *A. thaliana* the group E member bZIP34 has been linked to pollen germination and pollen tube growth (23). In contrast, group F members regulate zinc (Zn) transporters and salt stress responses (34, 50). Group C and S bZIPs are known to heterodimerize in the so-called C/S1 bZIP network involved in nutrient and energy metabolism (28, 51). Likewise, group C and S bZIPs are co-expressed in some tissues such as roots and nodules in alfalfa.

In addition to regulating development, bZIPs play a wide array of roles in biotic and abiotic stress responses (10, 11) in different crop species. Zou et al., (2008) (52) identified the OsABI5 bZIP TF that was involved in rice fertility and stress tolerance. Nijhawan et al., (2008) (7) related *bZIP* genes in rice to drought tolerance through genomic survey and gene expression analysis. Similarly, a root-specific bZIP transcription factor was isolated in tepary beans and found to be responsive to water-stress conditions (53). Liao et al., (2008) (12) isolated three bZIP genes (*GmbZIP44*, *GmbZIP62*, *GmbZIP78*) and found a negative regulator of ABA and tolerance to salt and freezing stress by over-expression in *A. thaliana*. As several studies have shown the role of bZIP transcription factors in the response to plant stress, Liu et al., (2012) (36) further added to it by cloning a bZIP gene and measuring physiological changes mediated by it in alfalfa under different stress conditions. Additionally, they over-expressed cloned alfalfa *bZIP* genes in tobacco plants which resulted in transgenic tobacco plants conveying salt and drought tolerance. These results indicate that over-expression of certain *bZIP* genes increases tolerance of plants to different abiotic stresses. In the present study, *MsbZIP* genes were found to be differentially expressed in response to cold, ABA, drought and salt indicating the involvement of distinct bZIPs in response to abiotic stresses. Notably, group A *MsbZIP88* was strongly upregulated in response to 24h salt stress and in *Arabidopsis* group A members are involved in ABA signaling and abiotic stress responses. Similarly, another group A member *MsbZIP80* showed upregulation 3h of ABA treatment as well as 3h drought exposure. Our results are in line with previously published work on bZIPs in abiotic stress responses and provide candidates for functional analyses in alfalfa.

Conclusion

Here we report the first comprehensive analysis of the bZIP transcription factor family in alfalfa (*Medicago sativa*). We identified 237 *bZIP* genes and named them *MsbZIP1* to *MsbZIP237*. Phylogenetic analysis of these bZIP genes using *Arabidopsis thaliana* as reference divided the sequences into 10 groups, with B, J, and K missing in alfalfa. The physico-chemical analysis and motif analysis showed

high specificity within each group. The expression profile of bZIPs from suggest bZIPs are expressed in a tissue-specific manner. Finally, the expression profiles of bZIP genes during different abiotic stress conditions (cold, ABA, drought, and salt) showed specific response of a few bZIP at specific timepoints during the stress response making them good candidates for stress-responsive transcription factors. Taken together, this work provides a framework for the future study of bZIPs in alfalfa and presents candidate bZIPs involved in stress-response signaling.

Abbreviations

ABA:	Abscisic acid
BLAST:	Basic local alignment search tool
bZIP:	Basic-leucine zipper
DNA:	Deoxyribonucleic acid
GO:	Gene ontology
Mb:	Mega base
MEGA:	Molecular evolutionary genetics analysis
MEME:	Multiple Em for Motif Elicitation
MUSCLE:	Multiple Sequence Comparison by Log-Expectation
NASS:	National Agricultural Statistics Service
NCBI:	National center for biotechnology information
PES:	Pre-elongated stem
RNA-Seq:	Ribonucleic acid sequencing
SAR:	Systemic acquired resistance
SRA:	Sequence read archive
TFs:	Transcription Factors
TGA:	TGACG-Binding
USDA:	United States Department of Agriculture
Zn:	Zinc

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during the current study are available in the, Plant Transcription Factor database (<http://planttfdb.gao-lab.org/family.php?fam=bZIP>),

Zeng, Yan (2020): genome fasta sequence and annotation files. figshare. Dataset., <https://doi.org/10.6084/m9.figshare.12327602.v3>

Raw reads for RNA-seq are available with Sequence Read Archive SRP055547, SRR7091780 to SRR7091794, and SRR7160313 to SRR7160357.

Competing interest

Not applicable.

Funding

This project was funded by USDA NIFA (Hatch project 1014919, 2018-70005-28792, 2019-67013-29171, and 2020-67021-32460), and Washington Grain Commission (Endowment and Awards #s 126593 and 134574)

Contributions

All Authors contributed equally to this work.

Acknowledgement

Not applicable

References

1. Statistics USD of ANA. United States Department of Agriculture National Agricultural Statistics Services. 2018.
2. Carelli M, Gnocchi S, Fancelli S, Mengoni A, Paffetti D, Scotti C, et al. Genetic Diversity and Dynamics of Sinorhizobium meliloti Populations Nodulating Different Alfalfa Cultivars in Italian Soils. Appl

- Environ Microbiol [Internet]. 2000 Nov;66(11):4785–9. Available from: <https://journals.asm.org/doi/10.1128/AEM.66.11.4785-4789.2000>
3. Pellock BJ, Cheng H-P, Walker GC. Alfalfa Root Nodule Invasion Efficiency Is Dependent on *Sinorhizobium meliloti* Polysaccharides. J Bacteriol [Internet]. 2000 Aug;182(15):4310–8. Available from: <https://journals.asm.org/doi/10.1128/JB.182.15.4310-4318.2000>
 4. Donnarumma F, Bazzicalupo M, Blažinkov M, Mengoni A, Sikora S, Babić KH. Biogeography of *Sinorhizobium meliloti* nodulating alfalfa in different Croatian regions. Res Microbiol [Internet]. 2014 Sep;165(7):508–16. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0923250814000898>
 5. Udvardi MK, Kakar K, Wandrey M, Montanari O, Murray J, Andriankaja A, et al. Legume Transcription Factors: Global Regulators of Plant Development and Response to the Environment. Plant Physiol [Internet]. 2007 Jun 7;144(2):538–49. Available from: <https://academic.oup.com/plphys/article/144/2/538/6106874>
 6. Pérez-Rodríguez P, Riaño-Pachón DM, Corrêa LGG, Rensing SA, Kersten B, Mueller-Roeber B. PlnTFDB: updated content and new features of the plant transcription factor database. Nucleic Acids Res [Internet]. 2010 Jan;38(suppl_1):D822–7. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkp805>
 7. 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 [Internet]. 2008 Feb 4;146(2):323–4. Available from: <https://academic.oup.com/plphys/article/146/2/323/6107127>
 8. Dröge-Laser W, Snoek BL, Snel B, Weiste C. The Arabidopsis bZIP transcription factor family – an update. Curr Opin Plant Biol [Internet]. 2018 Oct;45:36–49. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1369526617302157>
 9. Hurst HC. Transcription factors. 1: bZIP proteins. Protein Profile [Internet]. 1994;1(2):123–68. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8528902>
 10. Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, et al. bZIP transcription factors in Arabidopsis. Trends Plant Sci [Internet]. 2002 Mar;7(3):106–11. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1360138501022233>
 11. 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 [Internet]. 2012 Dec 1;19(6):463–76. Available from: <https://academic.oup.com/dnaresearch/article-lookup/doi/10.1093/dnares/dss026>
 12. Liao Y, Zou H-F, Wei W, Hao Y-J, Tian A-G, Huang J, et al. Soybean GmbZIP44, GmbZIP62 and GmbZIP78 genes function as negative regulator of ABA signaling and confer salt and freezing tolerance in transgenic Arabidopsis. Planta [Internet]. 2008 Jul 26;228(2):225–40. Available from: <http://link.springer.com/10.1007/s00425-008-0731-3>
 13. Wang J, Zhou J, Zhang B, Vanitha J, Ramachandran S, Jiang S-Y. Genome-wide Expansion and Expression Divergence of the Basic Leucine Zipper Transcription Factors in Higher Plants with an

- Emphasis on SorghumF. *J Integr Plant Biol* [Internet]. 2011 Mar;53(3):212–31. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1744-7909.2010.01017.x>
14. Liu J, Chen N, Chen F, Cai B, Dal Santo S, Tornielli GB, et al. Genome-wide analysis and expression profile of the bZIP transcription factor gene family in grapevine (*Vitis vinifera*). *BMC Genomics* [Internet]. 2014 Dec 13;15(1):281. Available from: <https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-281>
 15. Baloglu MC, Eldem V, Hajyzadeh M, Unver T. Genome-Wide Analysis of the bZIP Transcription Factors in Cucumber. Zhang B, editor. *PLoS One* [Internet]. 2014 Apr 23;9(4):e96014. Available from: <https://dx.plos.org/10.1371/journal.pone.0096014>
 16. Zhou Y, Xu D, Jia L, Huang X, Ma G, Wang S, et al. Genome-Wide Identification and Structural Analysis of bZIP Transcription Factor Genes in *Brassica napus*. *Genes (Basel)* [Internet]. 2017 Oct 24;8(10):288. Available from: <http://www.mdpi.com/2073-4425/8/10/288>
 17. Izawa T, Foster R, Nakajima M, Shimamoto K, Chua NH. The rice bZIP transcriptional activator RITA-1 is highly expressed during seed development. *Plant Cell* [Internet]. 1994 Sep;6(9):1277–87. Available from: <https://academic.oup.com/plcell/article/6/9/1277-1287/5984775>
 18. Toh S, McCourt P, Tsuchiya Y. HY5 is involved in strigolactone-dependent seed germination in *Arabidopsis*. *Plant Signal Behav* [Internet]. 2012 May 28;7(5):556–8. Available from: <http://www.tandfonline.com/doi/abs/10.4161/psb.19839>
 19. Fukazawa J, Sakai T, Ishida S, Yamaguchi I, Kamiya Y, Takahashi Y. REPRESSION OF SHOOT GROWTH, a bZIP Transcriptional Activator, Regulates Cell Elongation by Controlling the Level of Gibberellins. *Plant Cell* [Internet]. 2000 Jun;12(6):901–15. Available from: <https://academic.oup.com/plcell/article/12/6/901-915/6008839>
 20. Yin Y. RF2a, a bZIP transcriptional activator of the phloem-specific rice tungro bacilliform virus promoter, functions in vascular development. *EMBO J* [Internet]. 1997 Sep 1;16(17):5247–59. Available from: <http://emboj.embopress.org/cgi/doi/10.1093/emboj/16.17.5247>
 21. Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, et al. FD, a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex. *Science (80-)* [Internet]. 2005 Aug 12;309(5737):1052–6. Available from: <https://www.science.org/doi/10.1126/science.1115983>
 22. Chuang C-F, Running MP, Williams RW, Meyerowitz EM. The PERIANTHIA gene encodes a bZIP protein involved in the determination of floral organ number in *Arabidopsis thaliana*. *Genes Dev* [Internet]. 1999 Feb 1;13(3):334–44. Available from: <http://www.genesdev.org/cgi/doi/10.1101/gad.13.3.334>
 23. 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* [Internet]. 2009 Jul 18;70(5):581–601. Available from: <http://link.springer.com/10.1007/s11103-009-9493-y>

24. Iven T, Strathmann A, Böttner S, Zwafink T, Heinekamp T, Guivarc'h A, et al. Homo- and heterodimers of tobacco bZIP proteins counteract as positive or negative regulators of transcription during pollen development. *Plant J* [Internet]. 2010 Apr 16;no-no. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2010.04230.x>
25. Guan Y, Ren H, Xie H, Ma Z, Chen F. Identification and characterization of bZIP-type transcription factors involved in carrot (*Daucus carota* L.) somatic embryogenesis. *Plant J* [Internet]. 2009 Oct;60(2):207–17. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2009.03948.x>
26. Baena-González E, Rolland F, Thevelein JM, Sheen J. A central integrator of transcription networks in plant stress and energy signalling. *Nature* [Internet]. 2007 Aug 1;448(7156):938–42. Available from: <http://www.nature.com/articles/nature06069>
27. Ciceri P, Locatelli F, Genga A, Viotti A, Schmidt RJ. The Activity of the Maize Opaque2 Transcriptional Activator Is Regulated Diurnally. *Plant Physiol* [Internet]. 1999 Dec 1;121(4):1321–7. Available from: <https://academic.oup.com/plphys/article/121/4/1321/6085486>
28. Ehlert A, Weltmeier F, Wang X, Mayer CS, Smeekens S, Vicente-Carbajosa J, et al. Two-hybrid protein-protein interaction analysis in *Arabidopsis* protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors. *Plant J* [Internet]. 2006 Jun;46(5):890–900. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2006.02731.x>
29. Hossain MA, Cho J-I, Han M, Ahn C-H, Jeon J-S, An G, et al. The ABRE-binding bZIP transcription factor OsABF2 is a positive regulator of abiotic stress and ABA signaling in rice. *J Plant Physiol* [Internet]. 2010 Nov;167(17):1512–20. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0176161710002348>
30. Amir Hossain M, Lee Y, Cho J-I, Ahn C-H, Lee S-K, Jeon J-S, et al. The bZIP transcription factor OsABF1 is an ABA responsive element binding factor that enhances abiotic stress signaling in rice. *Plant Mol Biol* [Internet]. 2010 Mar 29;72(4–5):557–66. Available from: <http://link.springer.com/10.1007/s11103-009-9592-9>
31. Ji X, Liu G, Liu Y, Zheng L, Nie X, Wang Y. The bZIP protein from *Tamarix hispida*, ThbZIP1, is ACGT elements binding factor that enhances abiotic stress signaling in transgenic *Arabidopsis*. *BMC Plant Biol* [Internet]. 2013 Dec 4;13(1):151. Available from: <https://bmcpantbiol.biomedcentral.com/articles/10.1186/1471-2229-13-151>
32. LIU J-X, SRIVASTAVA R, HOWELL SH. Stress-induced expression of an activated form of AtbZIP17 provides protection from salt stress in *Arabidopsis*. *Plant Cell Environ* [Internet]. 2008 Dec;31(12):1735–43. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-3040.2008.01873.x>
33. Lu G, Gao C, Zheng X, Han B. Identification of OsbZIP72 as a positive regulator of ABA response and drought tolerance in rice. *Planta* [Internet]. 2009 Feb 2;229(3):605–15. Available from: <http://link.springer.com/10.1007/s00425-008-0857-3>

34. Yang O, Popova O V, Süthoff U, Lüking I, Dietz K-J, Golldack D. The Arabidopsis basic leucine zipper transcription factor AtbZIP24 regulates complex transcriptional networks involved in abiotic stress resistance. *Gene* [Internet]. 2009 May;436(1–2):45–55. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0378111909000857>
35. Chen H, Chen W, Zhou J, He H, Chen L, Chen H, et al. Basic leucine zipper transcription factor OsbZIP16 positively regulates drought resistance in rice. *Plant Sci* [Internet]. 2012 Sep;193–194:8–17. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0168945212001008>
36. Liu C, Wu Y, Wang X. bZIP transcription factor OsbZIP52/RISBZ5: a potential negative regulator of cold and drought stress response in rice. *Planta* [Internet]. 2012 Jun 22;235(6):1157–69. Available from: <http://link.springer.com/10.1007/s00425-011-1564-z>
37. Park S-H, Jeong JS, Lee KH, Kim YS, Do Choi Y, Kim J-K. OsbZIP23 and OsbZIP45, members of the rice basic leucine zipper transcription factor family, are involved in drought tolerance. *Plant Biotechnol Rep* [Internet]. 2015 Mar 10;9(2):89–96. Available from: <http://link.springer.com/10.1007/s11816-015-0346-7>
38. Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, et al. 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* [Internet]. 2010 Feb;61(4):672–85. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2009.04092.x>
39. Fu ZQ, Dong X. Systemic Acquired Resistance: Turning Local Infection into Global Defense. *Annu Rev Plant Biol* [Internet]. 2013 Apr 29;64(1):839–63. Available from: <https://www.annualreviews.org/doi/10.1146/annurev-arplant-042811-105606>
40. Gatz C. From Pioneers to Team Players: TGA Transcription Factors Provide a Molecular Link Between Different Stress Pathways. *Mol Plant-Microbe Interact* [Internet]. 2013 Feb;26(2):151–9. Available from: <https://apsjournals.apsnet.org/doi/10.1094/MPMI-04-12-0078-IA>
41. Chen H, Zeng Y, Yang Y, Huang L, Tang B, Zhang H, et al. Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nat Commun* [Internet]. 2020;11(1):2494. Available from: <https://doi.org/10.1038/s41467-020-16338-x>
42. Zhang Z, Liu W, Qi X, Liu Z, Xie W, Wang Y. Genome-wide identification, expression profiling, and SSR marker development of the bZIP transcription factor family in *Medicago truncatula*. *Biochem Syst Ecol* [Internet]. 2015 Aug;61:218–28. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0305197815001520>
43. Yang Y, Li J, Li H, Yang Y, Guang Y, Zhou Y. The bZIP gene family in watermelon: genome-wide identification and expression analysis under cold stress and root-knot nematode infection. *PeerJ* [Internet]. 2019 Oct 16;7:e7878. Available from: <https://peerj.com/articles/7878>
44. 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* [Internet]. 2009 Jul 1;37(Web Server):W202–8. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkp335>

45. O'Rourke JA, Fu F, Bucciarelli B, Yang SS, Samac DA, Lamb JFS, et al. The *Medicago sativa* gene index 1.2: a web-accessible gene expression atlas for investigating expression differences between *Medicago sativa* subspecies. *BMC Genomics* [Internet]. 2015 Dec 7;16(1):502. Available from: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-015-1718-7>
46. E ZG, Zhang YP, Zhou JH, Wang L. Mini Review Roles of the bZIP gene family in rice. *Genet Mol Res* [Internet]. 2014;13(2):3025–36. Available from: <http://www.funpecrp.com.br/gmr/year2014/vol13-2/pdf/gmr4227.pdf>
47. Wang Z, Cheng K, Wan L, Yan L, Jiang H, Liu S, et al. Genome-wide analysis of the basic leucine zipper (bZIP) transcription factor gene family in six legume genomes. *BMC Genomics* [Internet]. 2015 Dec 10;16(1):1053. Available from: <http://www.biomedcentral.com/1471-2164/16/1053>
48. Howell SH. Endoplasmic Reticulum Stress Responses in Plants. *Annu Rev Plant Biol* [Internet]. 2013 Apr 29;64(1):477–99. Available from: <https://www.annualreviews.org/doi/10.1146/annurev-arplant-050312-120053>
49. Gangappa SN, Srivastava AK, Maurya JP, Ram H, Chattopadhyay S. Z-Box Binding Transcription Factors (ZBFs): A New Class of Transcription Factors in *Arabidopsis* Seedling Development. *Mol Plant* [Internet]. 2013 Nov;6(6):1758–68. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1674205214602628>
50. Assuncao AGL, Herrero E, Lin Y-F, Huettel B, Talukdar S, Smaczniak C, et al. *Arabidopsis thaliana* transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. *Proc Natl Acad Sci* [Internet]. 2010 Jun 1;107(22):10296–301. Available from: <http://www.pnas.org/cgi/doi/10.1073/pnas.1004788107>
51. Weltmeier F, Ehlert A, Mayer CS, Dietrich K, Wang X, Schütze K, et al. Combinatorial control of *Arabidopsis* proline dehydrogenase transcription by specific heterodimerisation of bZIP transcription factors. *EMBO J* [Internet]. 2006 Jul 12;25(13):3133–43. Available from: <http://emboj.embopress.org/cgi/doi/10.1038/sj.emboj.7601206>
52. Zou M, Guan Y, Ren H, Zhang F, Chen F. A bZIP transcription factor, OsABI5, is involved in rice fertility and stress tolerance. *Plant Mol Biol* [Internet]. 2008 Apr;66(6):675–83. Available from: <http://link.springer.com/10.1007/s11103-008-9298-4>
53. 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* [Internet]. 2006 Mar 1;57(6):1391–8. Available from: <http://academic.oup.com/jxb/article/57/6/1391/516687/A-rootspecific-bZIP-transcription-factor-is>
54. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* [Internet]. 1990 Oct;215(3):403–10. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0022283605803602>
55. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* [Internet]. 2004 Mar 8;32(5):1792–7. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkh340>

56. Okonechnikov K, Golosova O, Fursov M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* [Internet]. 2012 Apr;28(8):1166–7. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts091>
57. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Battistuzzi FU, editor. *Mol Biol Evol* [Internet]. 2018 Jun 1;35(6):1547–9. Available from: <https://academic.oup.com/mbe/article/35/6/1547/4990887>
58. Conesa A, Götz S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *Int J Plant Genomics* [Internet]. 2008 Apr 30;2008:1–12. Available from: <https://www.hindawi.com/journals/ijpg/2008/619832/>
59. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* [Internet]. 2016 Sep 11;11(9):1650–67. Available from: <http://www.nature.com/articles/nprot.2016.095>
60. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* [Internet]. 2015 Apr 9;12(4):357–60. Available from: <http://www.nature.com/articles/nmeth.3317>
61. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* [Internet]. 2015 Mar 18;33(3):290–5. Available from: <http://www.nature.com/articles/nbt.3122>
62. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* [Internet]. 2010 Jan 1;26(1):139–40. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp616>
63. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant* [Internet]. 2020 Aug;13(8):1194–202. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1674205220301878>
64. Luo D, Wu Y, Liu J, Zhou Q, Liu W, Wang Y, et al. Comparative Transcriptomic and Physiological Analyses of *Medicago sativa* L. Indicates that Multiple Regulatory Networks Are Activated during Continuous ABA Treatment. *Int J Mol Sci* [Internet]. 2018 Dec 22;20(1):47. Available from: <http://www.mdpi.com/1422-0067/20/1/47>
65. Zhou Q, Luo D, Chai X, Wu Y, Wang Y, Nan Z, et al. Multiple Regulatory Networks Are Activated during Cold Stress in *Medicago sativa* L. *Int J Mol Sci* [Internet]. 2018 Oct 15;19(10):3169. Available from: <http://www.mdpi.com/1422-0067/19/10/3169>
66. Luo D, Zhou Q, Wu Y, Chai X, Liu W, Wang Y, et al. Full-length transcript sequencing and comparative transcriptomic analysis to evaluate the contribution of osmotic and ionic stress components towards salinity tolerance in the roots of cultivated alfalfa (*Medicago sativa* L.). *BMC Plant Biol* [Internet]. 2019 Dec 21;19(1):32. Available from: <https://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-019-1630-4>

67. Freese NH, Norris DC, Loraine AE. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* [Internet]. 2016 Jul 15;32(14):2089–95. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw069>

Figures

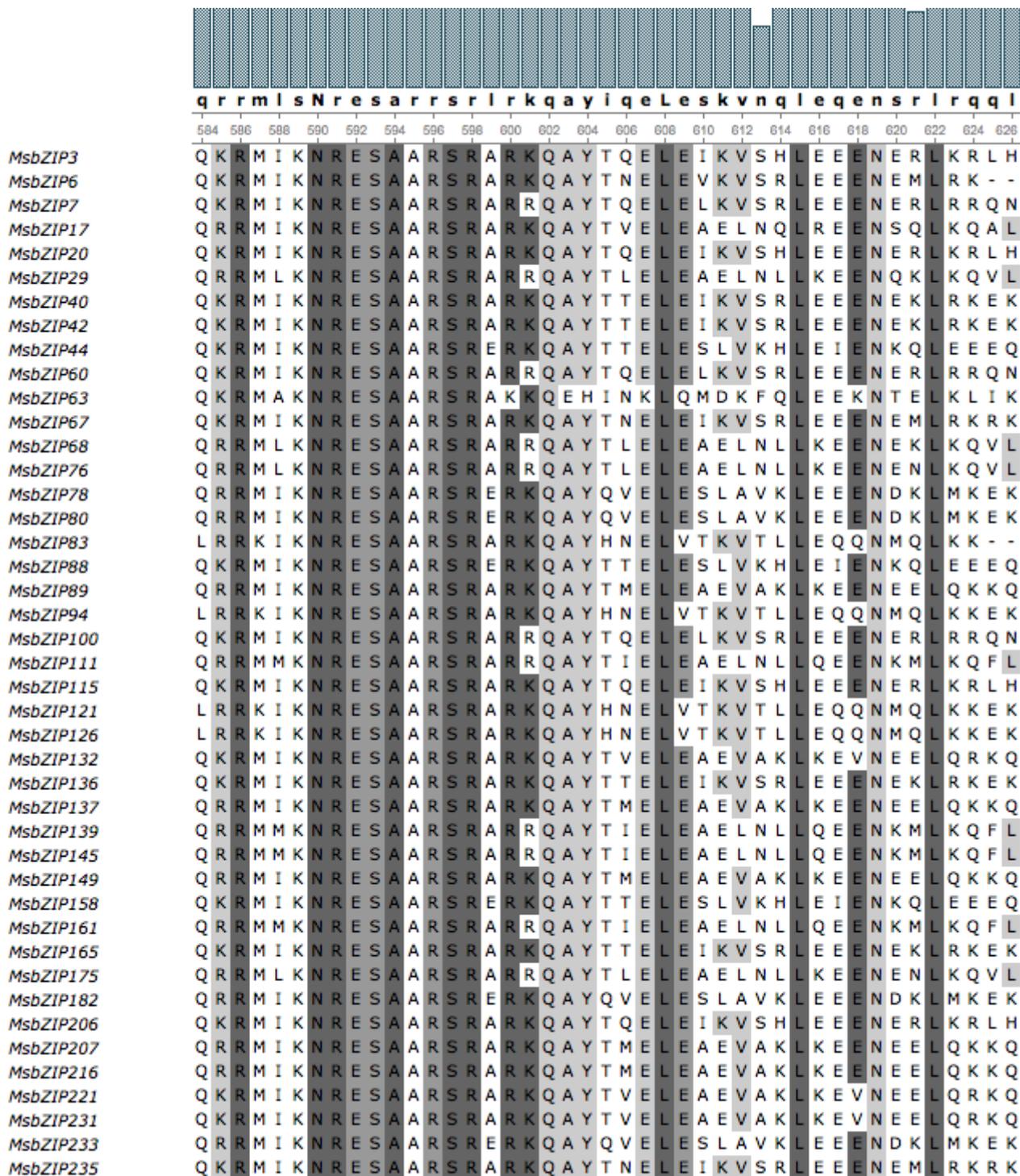


Figure 1

Multiple sequence alignment of alfalfa bZIP proteins of group S. The alignment was performed using MUSCLE 3.8.31 and visualized using Unipro UGENE v.33. The S group contains 32 bZIPs in alfalfa, which are highly expressed during abiotic stress. A consensus sequence is provided at the top of the figure (in bold). The bars above the consensus sequence present the percentage of consensus amino acid base between the aligned sequences. The ruler below the consensus sequence provides the position of the amino acid base on the aligned sequences. The color changes from light to dark with dark color indicating highly conserved amino acid bases. The consensus sequence at the top indicates the level of conservation with a capital letter indicating high conservation and a small letter with low conservation. The highly conserved bZIP domain (N-x(7)-[RK]-x(9)-L-x(6)-L-x(6)-L), between 590 to 622, is presented in dark color.

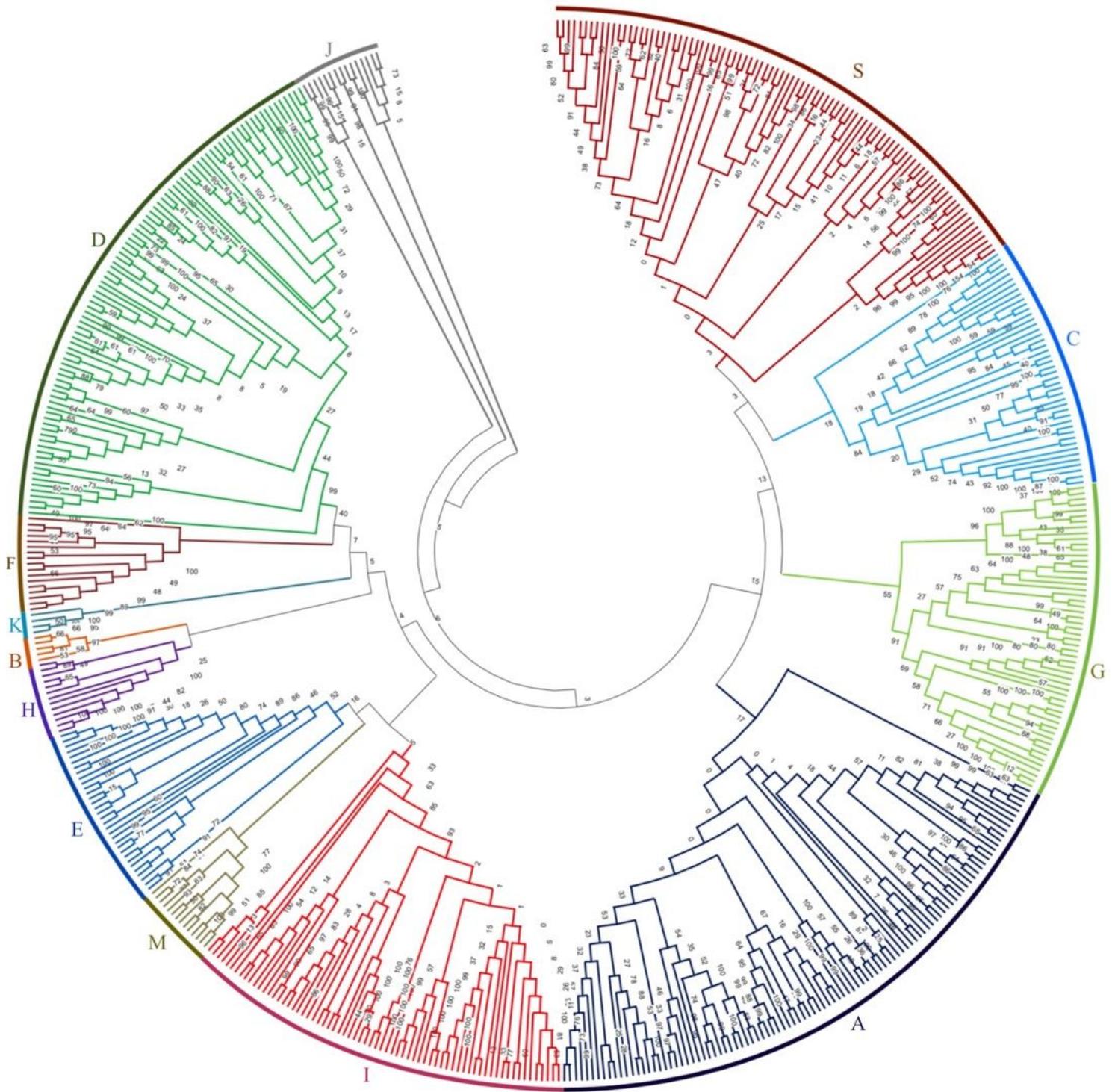


Figure 2

Phylogenetic analysis and group classification of bZIP proteins from alfalfa. 237 bZIP proteins and 312 proteins from *Arabidopsis thaliana* (78), *Lotus japonicus* (70), *Medicago truncatula* (75) and *Oryza sativa* (89) were used to create a neighbor joining tree with 1000 bootstraps. The bZIPs are grouped into 11 groups (A - K, M and S) based on tree topology results from *Arabidopsis thaliana* and *Medicago truncatula*. A detailed information of the tree including the genes from all the species mentioned above is presented in Genes expressed during stress conditions are distributed throughout the groups. Group D

contained genes highly expressed during drought stress conditions and ABA, while Group I and S contained genes that are upregulated during salt stress conditions. Although genes highly expressed during cold stress conditions were distributed over different groups, most of them were from group A and S. The detailed information with the gene is presented in **Supplement Figure (S6)**.

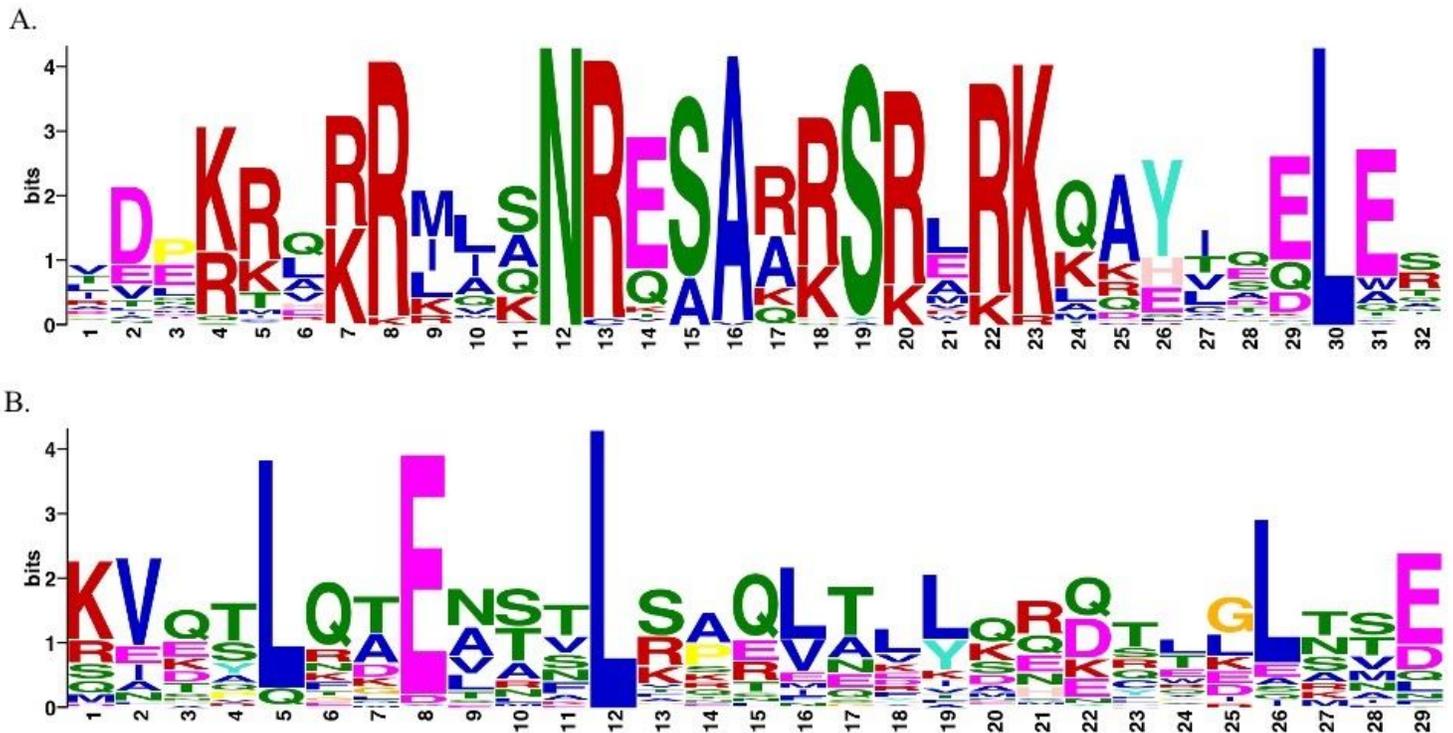


Figure 3

Conserved bZIP domain. (A) the conserved basic region of the bZIP motif (motif 12-29). The basic region is composed of an invariant motif (N-x7-R/K-x9) with 18 amino acid residues which can be observed. The basic region facilitates the sequence specific DNA binding. (B) The leucine zipper region that contains the heptad repeat of leucine (motif 5-19) or other bulky hydrophobic amino acids is represented by the figure. This region is important as it facilitates the dimerization specificity. This consensus sequence was generated using MEME suite 5.3.3. MEME (Multiple Em for Motif Elicitation) allows discovery of novel motifs in collection of nucleotides or protein sequences. The height of the character corresponds to how frequently the character occurs at that position in the motif.

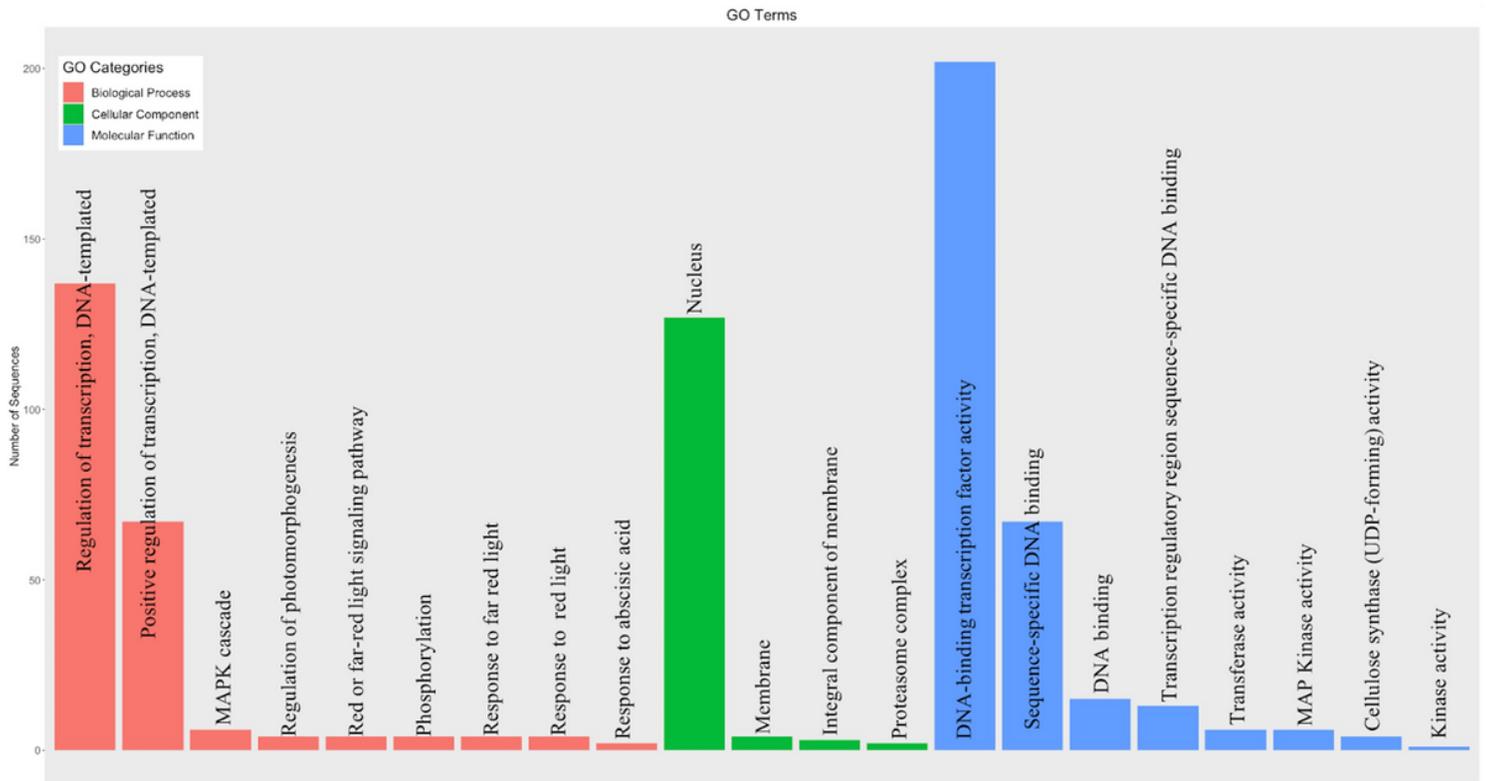


Figure 4

Functional annotation of bZIP proteins in alfalfa. Distribution of genes in different GO categories for Biological Process (pink), Cellular Component (green) and Molecular Function (blue). In the molecular function category, most of the genes were assigned to DNA binding transcription factor activity followed by sequence-specific DNA binding. Similarly, most of the biological process of these genes involves regulation of transcription and these genes are mostly located inside the nucleus as presented in the cellular component category.

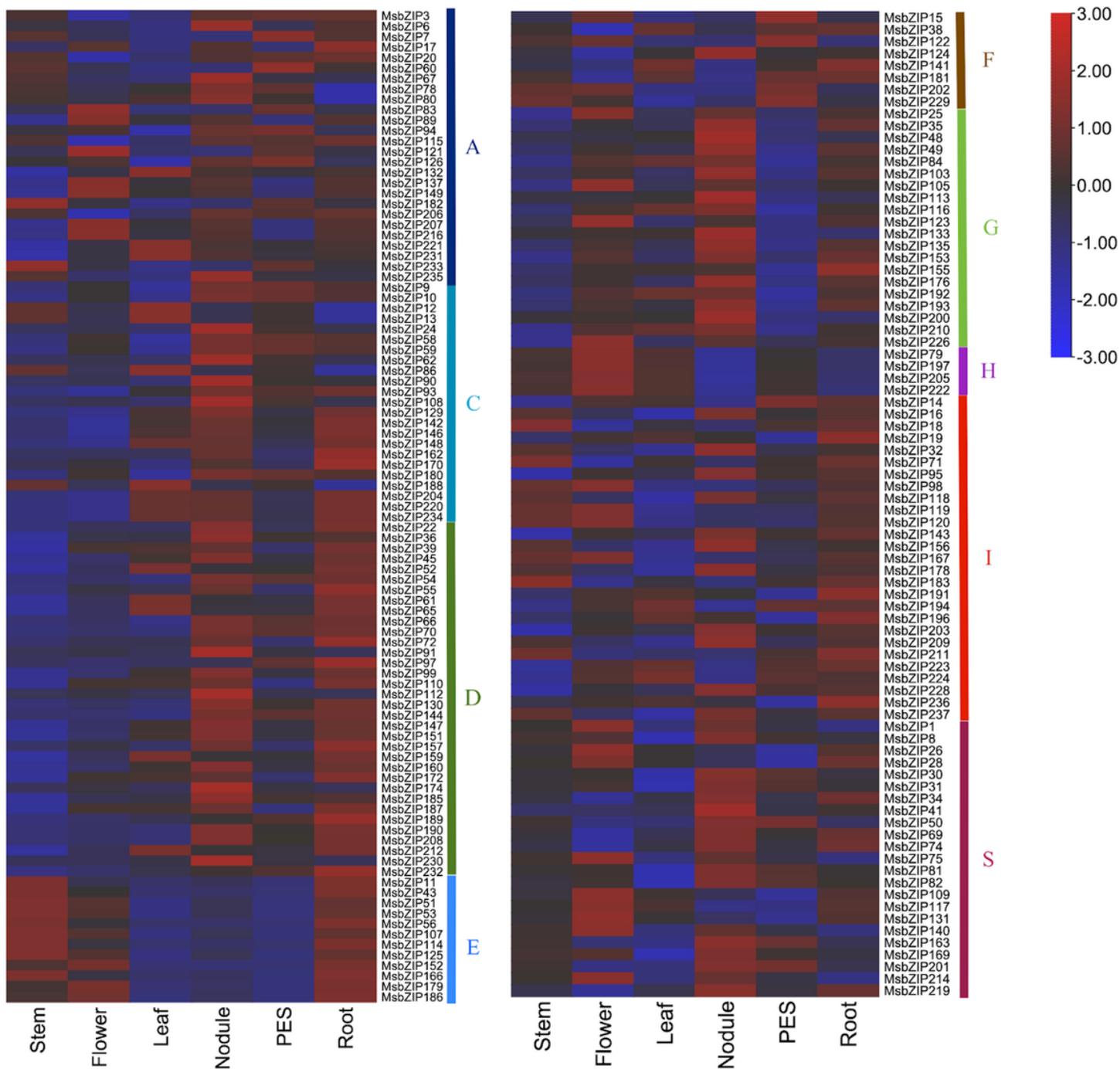


Figure 5

Expression profile of alfalfa bZIP genes among different tissues and organs. In the figure, PES is pre-elongated stem. Most of the genes were highly expressed in Stem, Flower, Leaf, Nodules, Pre-elongated stem (PES) and Root. The genes that were expressed in one tissue were not expressed in the other tissue indicating different genes may be required during different growth stages of alfalfa. Different groups are represented to the side of the genes by the respective group name along with the color as represented in the phylogenetic tree.

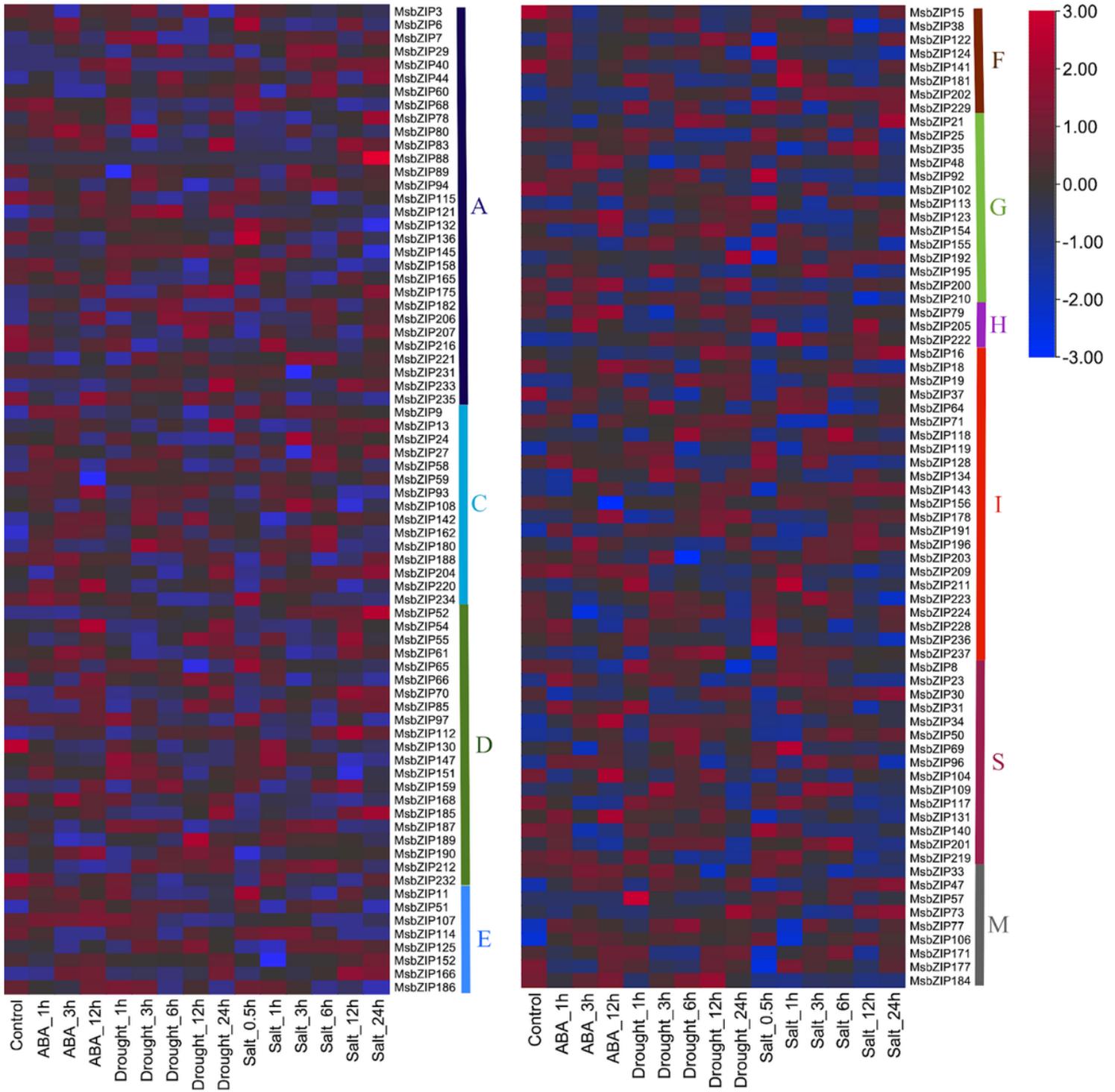


Figure 6

Expression profile of 146 alfalfa bZIP genes in response to ABA, drought, and salt stress. Most of the genes were highly expressed during initial treatment of salt stress from 0.5 to 3 hours. For salt stress, the genes that were highly expressed during the 0.5 hour of treatment were also actively expressed during 3 hours of treatment. For drought stress, gene expression levels were increased as the duration of drought stress was increased from 1 hour to 3 hours to 6 hour and more genes were expressed during 6 hour of drought stress. ABA treatment showed only a few genes that were expressed as the stress duration was

increased to 3 hours and 12 hours. Different groups are represented to the side of the genes by the respective group name along with the color as represented in the phylogenetic tree

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure1.png](#)
- [SupplementaryFigure4.pdf](#)
- [SupplementaryFigure5.pdf](#)
- [bZIPManuscriptalfalfaSupplementaryinformation.doc](#)