In silico analysis of putative transcription factor binding sites in the promoter region of drought responsive bZIP1 gene in rice and its orthologue in *Arabidopsis thaliana*

Manjesh Saakre*[1] and P. S. Abida[2]

[1]*Centre for Plant Biotechnology and Molecular Biology, College of Horticulture, Kerala Agricultural University, Thrissur 680656, Kerala, India.*
[2]*Regional Agricultural Research Station, Kerala Agricultural University, Pattambi, Palakkad 679306, Kerala, India.*

Received 04 January 2018; received in revised form 14 August 2018; accepted 26 October 2018

Abstract
The regulation of various biological processes in the plant systems, especially during different adverse climatic conditions are brought about by the change in the expression of different genes which in turn is the result of the binding of specific transcription factors in their respective transcription factor binding sites (TFBSs). In plants, basic region/leucine zipper motif (bZIP) transcription factors regulate processes including pathogen defense, light and stress signaling, seed maturation and flower development. In this study, an *in-silico* analysis was done to predict the TFBSs for the *bZIP1* gene of rice responses to drought. The AGI code of *bZIP1* was utilized for the prediction of TFBSs by performing orthologue search against *Arabidopsis* genome in RGAP (Rice Genome Annotation Project) database. Further, TFBSs were identified by AthaMap database, a genome-wide map of TFBSs in *Arabidopsis thaliana*, and STIFDB2 (Stress Responsive Transcription Factor Database V2.0) database which is a comprehensive collection of biotic and abiotic stress responsive genes in *Arabidopsis* and *Oryza sativa* L. The significant TFBSs were analyzed based on the parameters provided by databases and were cross validated. The results revealed that the MYB, HSF and WRKY Transcription Factor families and their respective TFBSs were predicted as functionally significant. These predicted TFBSs would be responsible for the change in expression of *bZIP1* gene under water stress. Such information will also help to understand the metabolic, physiological and cellular mechanisms implicated in such processes. These studies could help to engineer the plants for resistance to stress and achieve better yield in crop plants.

Key words: *Arabidopsis thaliana*, AthaMap, bZIP1, Drought, Rice, STIFDB2, TFBSs.

Introduction

Understanding the molecular mechanisms that underlie stress tolerance would be the first step in the generation of stress tolerant crops. Plants are often subjected to unfavorable environmental conditions – abiotic factors, causing abiotic stresses which play a major role in determining productivity of crop yields, and also the differential distribution of the plants species across different types of environment. A remarkable feature of plant adaptation to abiotic stresses is the activation of multiple responses involving complex gene interactions and crosstalk with many molecular pathways (Nguyen et al., 2016).

Abiotic stresses elicit complex cellular responses that have been elucidated by progress made in exploring and understanding plant abiotic responses at the whole-plant, physiological, biochemical,

*Author for correspondence: Phone: +91-8594052595, Email: manjesh.saakre565@gmail.com

cellular and molecular levels. One of the biggest challenges to modern sustainable agriculture development is to obtain new knowledge that should allow breeding and engineering plants with new and desired agronomical traits. The creation of stress tolerant crop either by genetic engineering or through conventional breeding covered almost all aspects of plant science, and is pursued by both public and private sector researchers (Sanghera et al., 2011).

To understand plant stress responses, disclosing the mechanisms of regulation of stress responsive genes assumes vital importance. Regulation of gene by Transcription Factors (TFs) is an important facet of stress responsive signal transduction cascades (Shameer et al., 2009). A transcription factor (TF) is a DNA binding protein that targets specific binding-sites (TFBSs) to regulate the transcript levels of its downstream genes (Yu and Li, 2017). Hence, identifying the TF-Transcription Factor Binding Sites (TFBSs) pairs is a crucial step in understanding the function of TFs and the regulatory network in an organism. A cis-regulatory element is a region of DNA that regulates the expression of genes located on that same molecule of DNA. A cis-element may be located upstream of the coding sequence of the gene it controls, in an intron, or downstream of the gene's coding sequence, in the untranscribed region (UTR) (Barrett et al., 2012). The transcription factors encoded by *basic leucine zipper (bZIP)* genes are concerned with directing various biological processes. The *bZIP* transcription factor family is one of the largest such families in plants. The members of this transcription factor family are mainly involved in stress responses as well as hormone signal transduction (Rodriguez and Connell, 2006). In rice, bZIP proteins have multiple biological processes including pathogen defense; responses to abiotic stresses; seed development and germination; senescence; and responses to salicylic acid, jasmonic acid, and abscisic acids. The expression of gene on its downstream is brought about by the bZIP transcription factors as a result of its interaction with ABA-responsive elements (ABREs) (Sheshadri et al., 2016). These *cis*-acting small DNA regions are present in the promoter region of ABA-inducible genes. Hence, the bZIP transcription factors are assigned as ABRE-binding factors (ABFs) or ABRE-binding proteins (AREBs) (Yamaguchi, 2005). The bZIP family consists of proteins with a DNA Binding Domain (BD). These domains are rich in basic amino acid residues and are present adjacent to a leucine zipper dimerization domain. Jakoby et al. (2002) reported 75 bZIPs in *Arabidopsis* whereas Correa et al. (2008) reported 92 bZIPs in rice.

Research groups have focused their research on elucidating the different components and molecular players underlying abiotic stress responses of a broad range of species both model and crop plants. Several attempts to engineer those species with improved abiotic stress traits were made and the response of genetically engineered plants was deeply studied after establishment of adequate physiological methods (Pandey et al., 2015). Large scale data integration from multiple experimental and bioinformatics resources will provide a robust platform to understand the major molecular players behind a biological problem. In the present study we developed a method for *in silico* identification of TFBSs on a selected gene. The TFBSs were predicted for differentially upregulated drought responsive *bZIP1* gene. The previous experiments had shown that the high levels of gene expression in plants lacking adequate water can be remarkably influenced by the drought and microbial colonization (Saakre et al., 2017). In this study, an *in-silico* analysis was done to identify TFBSs for the *bZIP1* gene of rice by developing a methodology to predict signal induced TFBSs**.** This method of prediction may useful before doing the *in vitro* experiments like Chromatin immuno Precipitation (ChIP) assay and DNA footprinting which will experimentally determine exact motifs of Transcription Factors.

Materials and Methods

Background work and sequence analysis of cDNA clone

In a previous study differential display technique was employed to understand gene expression patterns in rice plants that were susceptible to drought at reproductive stage that were subjected to stress by withholding water, *Pseudomonas fluorescens* strain (*Pf1*) treated plants subjected to drought stress by withholding water and control (well-watered). Differentially expressed cDNAs of six genes (*COX1, PKDP, bZIP1, AP2-EREBP, Hsp20* and *COC1*) were identified, cloned and sequenced. Real-time qPCR analysis showed that all the six genes were upregulated in drought-stressed plants treated with *Pf1*. This revealed that the remarkable influence of *Pf1* colonization led to drought tolerance at the reproductive stage. Among six genes gene *bZIP1* was found to have six fold increase in expression and highest among the genes. The cloned nucleotide sequence of *bZIP1* transcript was compared with rice genome database called RGAP (Rice Genome Annotation Project; http://rice.plantbiology.msu.edu/index.shtm) using BLASTX (nucleotide translated query vs. protein database) search to obtain Os (*Oryza sativa*) ID. The nucleotide sequence used in this study was a cloned sequence of differentially expressed gene under water stress in rice. The cloned nucleotide sequence used in this study showed homology with bZIP1 family protein and was considered as bZIP1 gene (Saakre et al., 2017). First hit from BLASTX results was considered as the best and gene ID (LOC_Os09g28310) was obtained. Orthologue search was done against *Arabidopsis thaliana* genome using RGAP database to obtain AGI code (AT3G19290) which required TFBSs identification.

Identification of TFBSs from AthMap database

TFBSs were identified using AthaMap database (http://www.athamap.de/; (Steffens et al., 2004)), a genome-wide map of TFBSs in *Arabidopsis thaliana*. AGI codes obtained from orthologue search were used to identify TFBSs. AthaMap provides two parameters to determine functionally significant transcription factors. The Matrix score e" 10 and Threshold e"5 were selected as potential TFBSs using "Search" tool in AthaMap (Steffens et al., 2004). The restriction to 20% by entering a "20" in the text field restriction was done to restrict the number of binding sites on the gene. The default upstream and downstream region of all genes to search was -500 and 50 bps respectively. The region of -500 bps already covered the area in which most of the regulatory sequences are found within the upstream region of *Arabidopsis thaliana* genes.

Identification of TFBSs from STIFDB2 V 2.0 database

STIFDB V2.0 (Stress Responsive Transcription Factor Database Version 2.0) (http://caps.ncbs.res.in/stifdb2/; is a comprehensive collection of biotic and abiotic stress responsive genes in *Arabidopsis thaliana* and *Oryza sativa* L. with options to identify plausible transcription factor binding sites in their promoters (Naika et al., 2013). The experimentally determined stress inducible TFBSs were identified from 1000bp data of query gene using the gene ID. Functionally significant transcription factors were identified and cis-regulatory elements were retrieved from the results. The database suggested that the algorithm Z-score for 100bp and its 5'UTR regions could be seen above 2.0 and for 1000bp and its 5'UTR regions could be seen above 1.5. This algorithm was validated with experimental data set of stress genes

.

Cross validation

AthaMap database provided the data about all categories of TFs involved in growth and development and also the TFs responses for stress signals. STIFDB2 contained the TFs responses predominantly for stress signals. The identified transcription factor binding sites from both AthaMap and STIFDB2 databases were cross validated by comparing the TF maps obtained by both databases to make prediction strong and also to screen stress inducible TFBSs from other binding sites. The detailed workflow is shown in Figure 1.

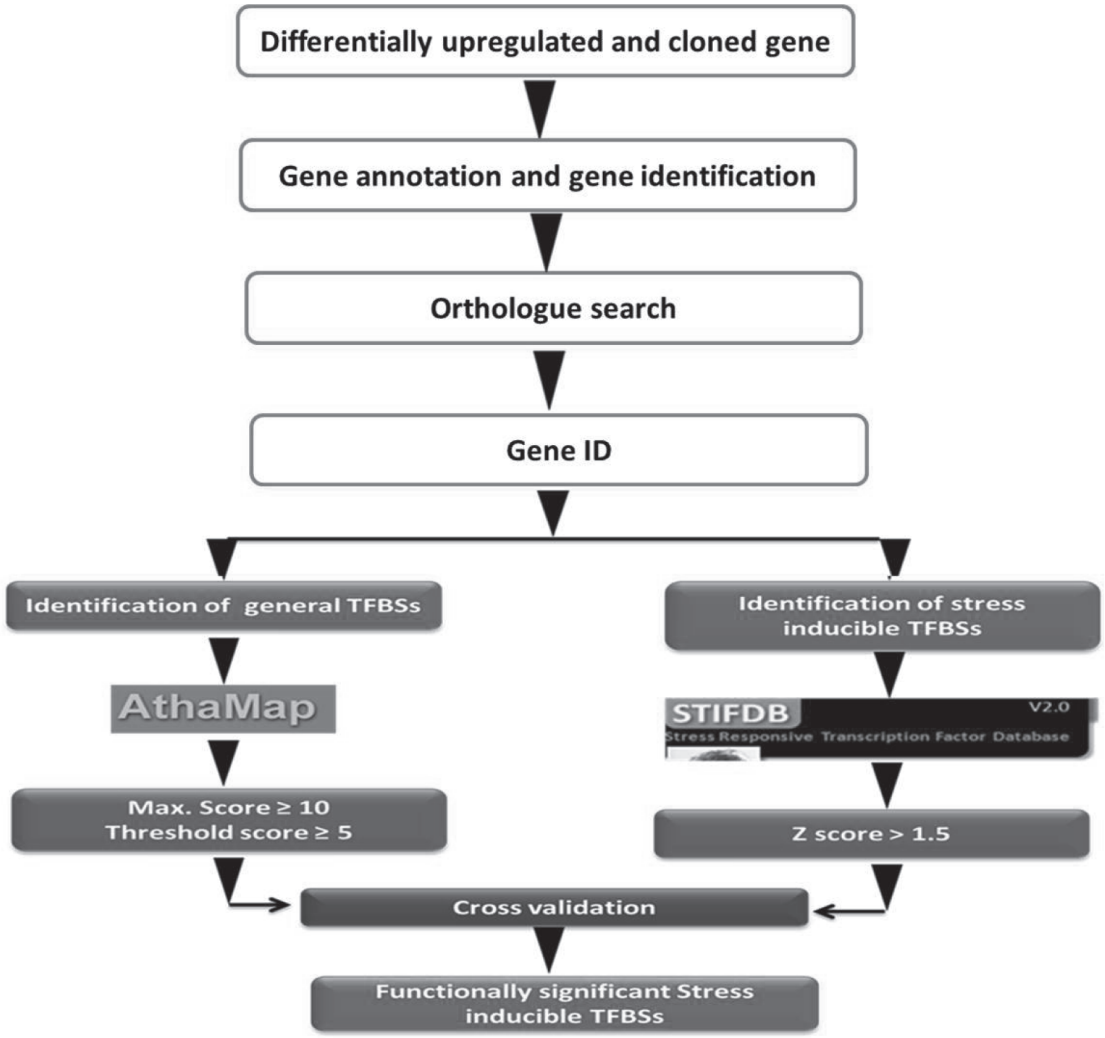

Figure 1. Schematic pipeline of the computational workflow involved in TFBS prediction in drought responsive bZIP1 gene.

Results and Discussion

The compiled data using AthaMap and STIFDB2 which are web based tools for database-assisted identification of transcription factor binding sites was cross validated for presence of common TFBSs at same regions or at least nearby with 100bps. A 1000bp map of promoter region was generated from the both AthaMap (Figure not shown) and STIFDB2 (Fig. 2) databases. The major transcription factor families such as MADS, C2H2(Zn), MYB, GARP, bZIP, MYB, AP2/EREBP, C2H2(Zn) and HSF identified using AthaMap are provided in Table 1. The data contains transcription factors (TFs), TF family for which the TFs belongs, Matrix score, thresholds and cis elements. Stress inducible, experimentally determined TFBSs were identified for *bZIP1* gene. The upstream region of *bZIP1* gene contained MYB bZIP, WRKY and HSF transcription factor family and MYB and bZIP transcription factors found on 5' UTR region. Details are given in Table 2 and Fig. 2. TFBSs obtained from both databases were cross validated, the transcription factor family MYB, HSF, bZIP and

Table 1. Predicted TFBs on promoter region of bZIP1 gene using AthaMap database

SL. No.	TF family	TF	Orientation	Matrix Score	Threshold	Cis element
1	MADS	AG	Reverse	14.56	6.32	tcaaccattttggaaag
2	C2H2(Zn)	ID1	Reverse	11.02	5.71	aagagacaaa
3	MYB	MYB55(2)	Reverse	12.9	6.36	caaccactc
4	MYB	MYB46(3)	Forward	10.94	5.14	caaccactc
5	GARP	ARR10	Reverse	12.04	5.23	cgaatcttcac
6	MYB	MYB52(2)	Reverse	11.47	5.2	aaacaaacg
7	bZIP	TGA1a	Reverse	10.45	5.21	ccacgtcgc
8	bZIP	bZIP60(2)	Forward	12.68	5.82	ccacgtcgc
9	bZIP	TGA1a	Reverse	10.45	5.21	ccacgtcgc
10	MYB	RVE1(1)	Reverse	10.28	5.95	ataatatct
11	AP2/EREBP	ATERF1(1)	Forward	12.97	5.42	ggcggcagc
12	AP2/EREBP	ATERF1(2)	Forward	12.97	5.42	aggcggca
13	AP2/EREBP	RAV1	Forward	10.65	5.01	cagcaacacac
14	C2H2(Zn)	ID1	Forward	11.02	5.71	ttgtcgtct
15	AP2/EREBP	WRI1	Forward	19.35	6.03	cttctctggagttttcgcagt
16	HSF	HSFB2a(2)	Forward	10.7	5.2	ttctggatc
17	HSF	HSFB2a(2)	Reverse	10.7	5.2	ttctggatc

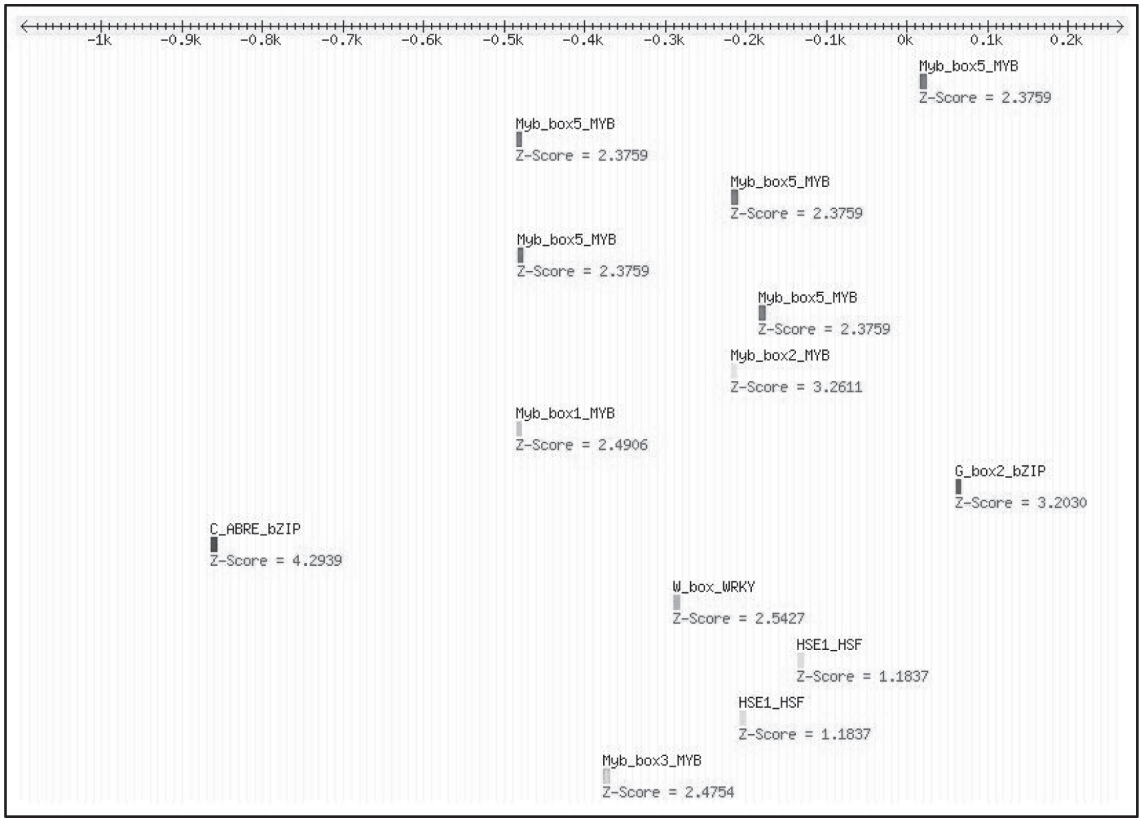

Figure 2. 1000bp TF map of bZIP1 gene obtained by STIFDB2 database

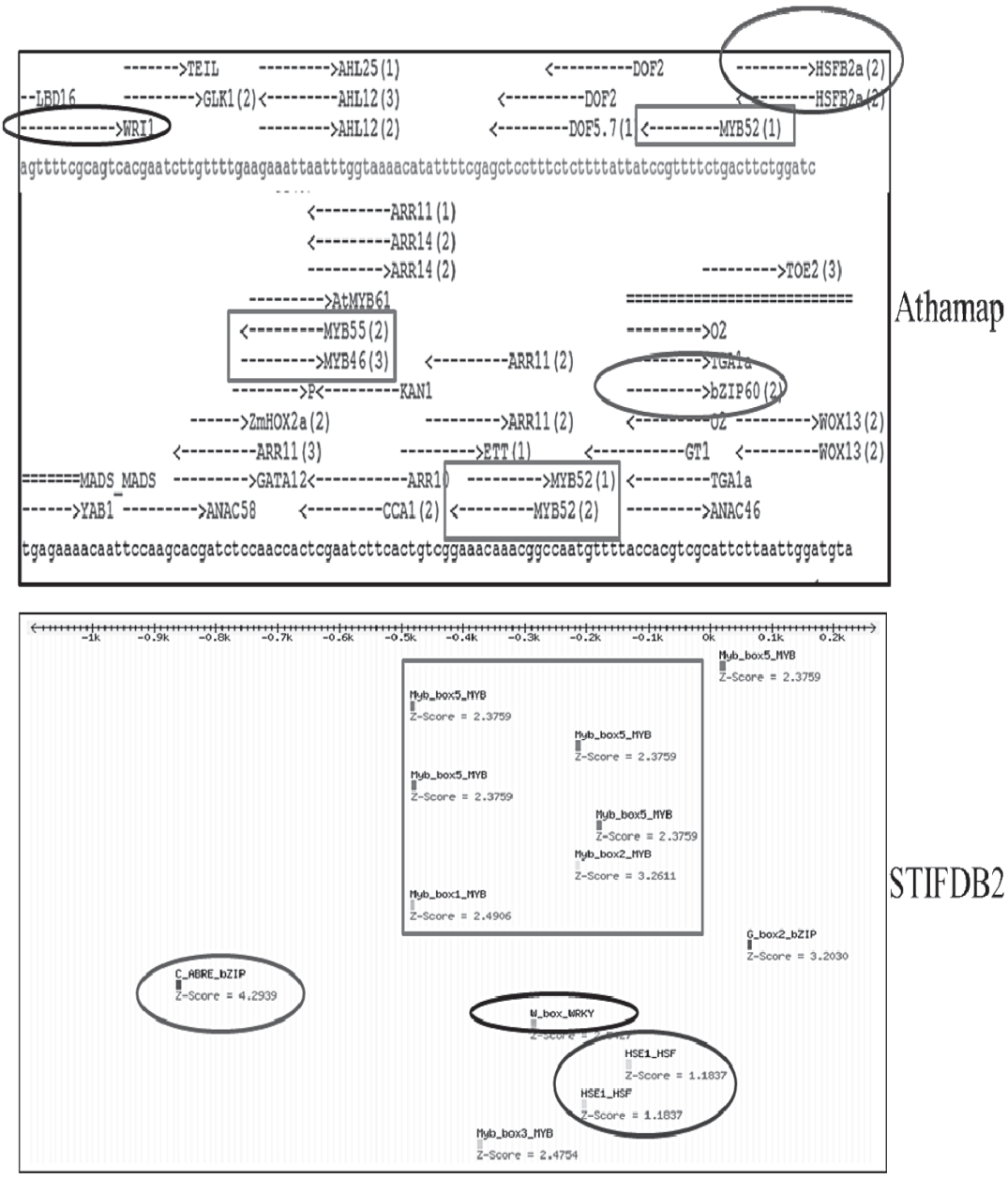

Figure 3. Common TFs found in *bZIP1* gene promoter; the transcription factors MYB, HSF, bZIP and WRKY were localized in both AthaMap and STIFDB2 databases

WRKY were found localized in both AthaMap and STIFDB2 databases (Fig. 3). It is important to understand the basic process and mechanism of the transcription factors and improve upon their stress response for better crop productivity. It is also be important to know the Transcription factors binding sites and help with combating stress using stress responsive adaptive mechanisms at molecular level. The regulation of various biological processes in the living systems, especially during different adverse conditions, are brought about by the change in the expression of different genes which in turn is the result of the binding of specific transcription factors in their respective TFBSs. Earlier Sanchita

Table 2. Predicted TFBSs identified for bZIP1 gene using STIFDB2 database

Sl. No.	Transcription Factor	Z - Score	Chromosome location	Orientation	Cis elements
1	Myb_box5_MYB	2.3759	5'UTR	Reverse	[C/T]AAC[A/T/G/C][A/G]
2	Myb_box5_MYB	2.3759	Upstream	Reverse	[C/T]AAC[A/T/G/C][A/G]
3	Myb_box5_MYB	2.3759	Upstream	Forward	[C/T]AAC[A/T/G/C][A/G]
6	Myb_box2_MYB	3.2611	Upstream	Reverse	CC[T/A]ACC
7	Myb_box1_MYB	2.4906	Upstream	Reverse	(T/C)AAC[G/T]G
8	G_box2_bZIP	3.2030	5'UTR	Reverse	TGACG[T/C]
9	C_ABRE_bZIP	4.2939	Upstream	Forward	CGCGTG
10	W_box_WRKY	2.5427	Upstream	Forward	(T)TGAC[C/T]
12	HSE1_HSF	1.1837	Upstream	Forward	TTC(A/T/G/C)(A,T,G,C)GAA,GAA(A/T/G/C)(A,T,G,C)TTC
13	Myb_box3_MYB	2.4754	Upstream	Reverse	TAACTG

et al. (2013) identified TFBSs using the data from seven plants of Solanaceae family having differentially expressed genes during different time periods of salt stress. In that study data was retrieved from the public domain and differential expressions of genes were revealed which might be due to binding of transcription factors in the promoter region. TFBSs were predicted by utilizing the promoter regions of differentially expressed genes. In an another study Mochida et al. (2011) identified and integrated all putative TFs from six grass species, *Brachypodium distachyon*, maize, rice, sorghum, barley, and wheat. In the present study we have targeted a specific gene bZIP1 which responds under drought and predicted TFBSs which may influence the expression of targeted gene. Stress Responsive Transcription Factor Database V.2.0 (STIFDB2) provided information on stress-responsive genes from *Arabidopsis thaliana, Oryza sativa* subsp. *japonica* and *O. sativa* subsp. *indica.* A total of 31 TFs were identified for 15 different stress signals (Naika et al., 2013). The compiled data using AthaMap and STIFDB2 was cross validated for common TFBSs. The *bZIP1* gene had transcription factor families. These TF families had high significant score and threshold in AthaMap.

Acknowledgements

The Authors thank the Department of Biotechnology (DBT), Government of India, and Bioinformatics Centre, College of Horticulture, Kerala Agricultural University, Thrissur, India, for providing facilities for current work.

References

Barrett, L. W., Fletcher, S and Wilton, S. D. 2012. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cell Mol Life Sci. 69(21): 3613–3634.

Correa, L. G., Pachon, R. D. M., Schrago, C.G., Santos. R. V., and Roeber, M. B. 2008. The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. PLoS One 3: e2944.

Jakoby, M., Weisshaar, B., Laser, D. W., Carbajosa, J., Tiedemann, J., Kroj, T., and Parcy, F. 2002. bZIP transcription factors in *Arabidopsis*. Trends Plant Sci. 7: 106-111.

Mochida, K., Yoshida, T., Sakurai, T., Shinozaki, K. Y., Shinozaki, K and Tran, L. S. 2011 *in silico* analysis of transcription factor repertoires and prediction of stress-responsive transcription factors from six major gramineae plants. *DNA Res*. 18, 321–332

Naika, M., Shameer, K., Mathew, O. K., Gowda. R., and Sowdhamini, R. 2013. STIFDB2: An updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in Arabidopsis and rice. Plant Cell Physiol. 54: 201-215.

Nguyen, D., Rieu I., Mariani, C., and Dam, N. M. V. 2016. How plants handle multiple stresses: hormonal interactions underlying responses to abiotic stress and insect herbivory. Plant Mol Biol. 91: 727–740.

Pandey, P., Ramegowda, V and Kumar S. M. 2015. Shared and unique responses of plants to multiple individual stresses and stress combinations: physiological and molecular mechanisms. Front Plant Sci. 2015; 6: 723.

Saakre, M., Baburao, T. M., Salim, A. P., Francies, R. M., Achuthan, V. P. Thomas, G and Sivarajan, R. S. 2017. Identification and characterization of genes responsible for drought tolerance in rice mediated by *Pseudomonas fluorescens*. Rice Sci. 2017, 24(5): 291"298.

Sanchita, Baby, B. M., and Sharma, A. 2013. *In silico* analysis of putative transcription factor binding sites in differentially expressed genes: Study of the turnover of TFBSs under salt stress responsiveness in solanaceae family. Plant Omics J. 6 (4):278-285.

Sanghera, G. S., Wani, S. H., Hussain, W., and Singh, N. B. 2011. Engineering cold stress tolerance in crop plants. Curr Genomics. 12(1): 30–43.

Shameer, K., Ambika, S., Mary, S., Varghese, Karaba, N., Udayakumar, N and Sowdhamini, R. 2009. STIFDB-Arabidopsis stress responsive transcription factor database. Int. J. Plant Genomics 8p.

Sheshadri, S. A., Nishanth, M. J. and Simon, B. 2016. Stress-mediated cis-element transcription factor interactions interconnecting primary and specialized metabolism in planta. Front Plant Sci. 7: 1725.

Steffens, N. O., Galuschka, C., Schindler, M., Bulow, L., and Hehl, R. 2004. AthaMap: an online resource for in silico transcription factor binding sites in the *Arabidopsis thaliana* genome. Nucleic Acids Res. 32: 368-372.

Rodriguez, U. L. and Connell, M. A., 2006. 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. 57: 1391-1398.

Yamaguchi, S. K. 2005. Organization of cis-acting regulatory elements in osmotic and cold-stress-responsive promoters. Trends Plant Sci. 10: 88-94.

Yu, C. P. and Li, W. H. 2017. Predicting transcription factor binding sites and their cognate transcription factors using gene expression data. Methods Mol Biol. 1629:271-282.