Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Molecular Cloning, Expression and Insilco Analysis of Drought Stress Inducible MYB Transcription Factor Encoding Gene From C4 Plant Eleusine Coracana

MEGHA BHATT (megha4bhatt@gmail.com)
GBPant University of Agriculture & Technology: Govind Ballabh Pant University of Agriculture & Technology
https://orcid.org/0000-0003-4364-7093

Research Article

Keywords: Abiotic stress, Transcriptional regulation, MYB-transcription factor cis-regulatory elements, In- silico analysis, finger millet

Posted Date: November 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-955905/v1

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

---

Abstract

Drought is one of the key abiotic stresses that critically influences the crops by restraining their growth and yield potential. Being sessile, plant tackle the detrimental effects of drought stress via modulating the cellular state by changing the gene expression. Such alteration of gene expression is essentially driven by the transcriptional syndicate. Transcription factors (TF) are the key regulatory protein that controls the expression of their target gene by binding to the cis-regulatory elements present in the promoter region. Myb-TF subiquitously present in all eukaryotes belong to one of the largest TF family, and play wide array of biological functions in plants including anthocyanin biosynthesis, vasculature system, cell signaling, seed maturation and abiotc stress responses. In the present study, isolation, and molecular cloning of full length Myb TF from *Eleusine corocana* has been performed. The isolated full-length coding sequence has 1053 bp and 350 aa was submitted to NCBI (Accession number MT312253). The transcript level of EcMYB increases under different abiotic stress treatment including dehydration, salinity, and high temperature stress. The promoter region of *EcMyb1* was found to be enriched in stress-responsive cis-regulatory elements such as DRE, HSE, ABRE etc. In phylogenetic analysis, *EcMyb1* appeared to have high homology with its monocot orthologs particularly *Sateria italica, Hordeum vulgare, Saccharum barberi and Oryza sativa*. The three-dimension protein structure was generated based on homology modeling and structural aspects were discussed. Further, Insilco analysis was conducted to explore the physiological properties, subcellular localization, potential post-translational modification sites (phosphorylation and glycosylation sites), and molecular and biological function of full-length protein. Overall, the expression profiling and Insilco analysis of *EcMyb1* strongly indicated its potential role in abiotic stress response in *Eleusine corocana*.

Introduction

Drought stress is one of the major challenges encountered by plants due to the dramatic increase in climate change. The sharp rise in population and the anthropogenic activities further aggravated the situation. Greenhouse emission rise has led to erratic precipitation, increase in arid land area, desertification and ultimately diminution of crop productivity (Yang et al. 2010). Owing to their molecular plasticity, plants have the ability to adapt and survive these changes in their environment. The molecular plasticity of plants is driven by the cell signaling cascade that follows the stress sensing, signal perception and adequate respond towards it. The signaling pathway is initiated by signaling molecules that detect the stress and relay the signaling pathway by activating secondary signaling molecules which further amplify the signaling and trigger the stress response. Among different secondary signaling molecules, calcium, cAMP, ROS, NO3, phosphorylation cascade etc. play important role. Moreover, the stress signaling is largely governed by the phytohormonal as well as transcriptional regulation. Interplay of phytohormones holds a remarkable role in the activation of TF. Activated transcription factors regulate the expression of stress responsive gene by limiting the stress induced cellular damage. TF, being a master regulator of gene expression are gaining more attention/ importance in context of target of crop improvement. In Arabidopsis, about 6% (or more than 1500 genes) of total number of genes are assigned to encode TF's only. TF essentially binds to the *cis*-regulatory elements presents in the promoter region of gene and modulate its transcription. MYB (Myeloblastosis), is one of the largest TF family, found ubiquitously in eukaryotes with varying functional role in plants. Like all TF, MYB-TF is also characterized by the presence of a highly conserved DNA-binding domain. Majority of MYB proteins contain variable numbers of N-terminus conserved MYB repeats (R). The MYB proteins categorization is carried out on the basis of repeat numbers present in sequences that may range from 1 to 4.A composition of 52 amino acids forms each repeats that ultimately creates 3 α-helices.Of the three α-helices, the second and third helices in every repeat create a helix turn helix (HTH) domain (Ogata et al. 1996). Further, MYB protein family is categorized into four different subfamily1R, R2R3, 3R and 4RMYB proteins respectively on the basis of MYB domain (Dubos C et al. 2010). Anumber of responses are regulated by plants through MYB proteins that are associated with R2R3 MYB subfamily (Lippold et al. 2009, Segarra et al. 2009). Large number of MYB transcription factors has been identified and have been reported to function in many important physiological and biochemical processes involving anthocyanin accumulation, cell cycle and cell development, metabolism, hormone biosynthesis and signal transduction as well as drought stress responses (Allan et al. 2008, Ambawat et al. 2013, Dubos et al. 2010). Previously we have reported a drought stress induced MYB transcription factor from *Eleusine corocana.* The expression level increases *EcMyb1* significantly enhanced with the progression of severity of drought stress (Salvi et al. 2012, Jadhav et al. 2018). Henceforth, it is essential to carry out research on drought responsive gene Myb besides their protein structure and promoter. Finger millet a member of poeceae, is a hardy crop and shows tolerance towards abiotic stresses specifically drought (Gupta et al.2017).Finger millet can survive in harsh environmental stress conditions due to their potent alleles which show drought resistant characteristics. Such crops will be of great significance in meeting challenges. Understanding the mechanisms involved in the response of plants to adverse environmental conditions will

help to generate crops with high tolerance to these stresses (Sanchita et al. 2013). Therefore, isolate full length coding sequence of Myb gene from drought tolerant variety of *Eleusine coracana* (PRM 6107), sequencing and cloning. Also, *Insilco* analysis was conducted to annotated the sequence-structure-function relational ship using bioinformatics tools and transcript analysis through real time PCR under during abiotic stress like drought, heat and salt stress.

Material And Methods

Plant material and growth conditions

Seeds of *Eleusine coracana* (finger millet) genotype -6107, drought tolerant genotype, were obtained from the Molecular biology department, GB Pant University. Seeds of finger millet were sown in pots filled with soil, peat moss and vermiculite in 3:1:1 proportion in polyhouse and allowed to grow up to seedling stage. Further plants were imposed to drought stress by withholding watering for 11 days.

RNA Isolation and cDNA preparation

After 11 days of drought treatment leaf samples of finger millets were collected and total RNA was extracted using the RNA-Xpress^{TM} Reagent (Himedia) as described in the manufacturer's instructions. Using RNA as template, the first-strand cDNA synthesis was performed with oligo dT primers and avian myeloblastosis virus reverse transcriptase enzyme using revert first strand cDNA synthesis kit (Thermo Scientific India Pvt. Ltd., Mumbai) according to manual instructions.

PCR Amplification, Cloning and sequencing of EcMyb1 **cDNA**

To obtain full-length of *EcMyb1* cDNA, two primers, 5′TCAACTAATGGTAGCCCTTCCCTCT-3′ (sense) and 5′-GATATTCTCAAAAGACAGTTGCATTCT -3′ (antisense) were designed based on the partial sequence of *EcMyb1* gene which has already been submitted with Genbank accession number JN107890.1 BLAST with complete *Eleusine coracana* whole genome sequence (GenBank: LXGH01099917.1) using bioinformatic tool nucleotide blast (nBLAST). After nBLAST the sequence that shows similarity with partial *EcMyb1* gene sequence was taken and further processed for ORF search using ORF finder tool (NCBI). The lateral flanking sequence of ORF was used for primer designing. Both the primers corresponded to complete fragment of *EcMyb1* gene. PCR reactions were carried out in Himedia Thermal Cycler under controlled conditions: 30 sec at 98°C, 30 cycles of 3 min at 98°C, 30 sec at 55°C, 40 sec at 72°C and final extension step for 5 min at 72°C. Fragment of putative *EcMyb1* gene was then cloned into pGEM-T easy Vector system (Invitrogen) according to the manual instructions. Additionally, selection of positive clone was performed on the basis of colony PCR which further proceeded for sequencing. After attaining full length sequence of *EcMyb1* gene, the sequence was submitted to NCBI GenBank with accession number MT312253.

Sequence analysis of cloned DNA fragment using bioinformatics approach

For the sequence analysis, first the cloned DNA fragment's nucleotide sequence converted to amino acid sequence using bioinformatics translate tool, ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem). This tool allows the translation of a nucleotide (DNA/RNA) sequence to amino acid sequence (Gasteiger et al. 2005). Further the deduced amino acid sequences were subjected to protein blast (pBLAST) to find sequence homology with other plant sequences. The isoelectric point (pI), molecular weight,, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index (AI) and grand average hydropathy (GRAVY) parameters were predicted by using protparam software (https://web.expasy.org/compute_pi/) (Bjellqvist et al..1993). The protein sequence of Myb domain was annotated by pfam (http://pfam.xfam.org/search/sequence) (Finn et al.2015). NetPhos2.0 and NetNGly 1.0 server was used for computing potential post translational modification sites. The domains present in the functional region which are responsible for the activity of *EcMyb1* gene were identified using PROSITE (http://expasy.hcuge.ch/sprot/prosite.html) (Sigrist et al. 2012). Location of Motifs was identified using the MEME suite (Motif-based sequence analysis tools) (Bailey et al. 2011). Phylogenetic tree was constructed by using software named as Neighbor joining BIONJ program of MEGA version 7.0 which follows Maximum Likelihood method for evolutionary analysis (Kumar et al.2016). Secondary structure of *EcMyb1* protein such as α-helix, β-sheet, and turn etc were predicted using self-optimized prediction method with alignment (SOPMA) (Geourjon et al.1995). 3-D structures *EcMyb1* protein was characterized by Swiss model web server (https://swissmodel.expasy.org/), through homology template approach. These are bioinformatics tools method for predicting three-dimensional structure model of protein molecules from amino acid sequences (Yang et al.2015).

Analysis of promoter and cis-regulatory elements

Nucleotide sequence of cloned DNA fragment further BLAST with *Eleusine coracana* whole genome sequence (GenBank: LXGH01099917.1) using bioinformatics tool nucleotide blast (nBLAST). Upstream sequences in the vicinity of the gene were selected and subjected to promoter analysis. Web-based bioinformatics tools such as, PLACE (http://www.dna.affrc.go.jp/htdocs/PLACE/) (Higo et al. 1999), PlantCARE (http://bioinformatics.psb .ugent.be/webtools/ plantcare/html/ (Lescot et al. 2002) and PlantPAN (http://plantpan2.itps.ncku.edu.tw/promoter.php (Chow et al. 2015) have been used for the analysis of the cis regulatory elements present in the upstream region of *EcMyb1* gene.

Differential expression analysis by quantitative real-time reverse transcription–PCR of EcMyb1 **gene for different Abiotic stresses**

Stress treatment

Finger millet genotype PRM 6107 was used for stress treatment. Seeds were surface sterilized and germinated in half strength MS media. For germination bottles were kept in dark for 2 days. After 2 days bottles were transferred to growth chamber under controlled conditions (light/dark regime of 18/6 h at 25°C, and light intensity of 200 μmol photons m–2 s–1. The seedlings were allowed to grow for 15 days. After 15 days seedlings were subjected to different abiotic stresses. For water stress seedlings were transferred to bottles containing MS media with three different concentration of PEG-6000. Three concentrations were 100ml of 5% (P1), 10% (P2), 15% (P3).

For salt stress, seedlings were allowed to grow on MS media with 100mM, 150mM and 200mM NaCl. Additionally heat stress was also given to the seedling. For heat stress germination bottles were kept in incubator at 35° C (H1), 40° C (H2) and 45°C (H3) temperature. All these treatment were carried out with control (in water) for 6 hours.

Analysis of EcMyb1 **gene through Real time PCR**

After 6 hours total RNA was extracted from the treated and control leaves samples using the RNA-Xpress^{TM} Reagent (Himedia) as described in the manufacturer's instructions and converted to cDNA using revert first strand cDNA synthesis kit (Thermo Scientific India Pvt. Ltd.) according to manual instructions. Real time PCR (Realplex; Applied Biosystem) was carryout in duplicates by applying SYBR green (PowerUp™ SYBR™ Green; Applied Biosystems, Thermo Fisher Scientific) using gene specific primers. Cycle conditions were 94° C for 3 min, 40 cycle of denaturation at 94°C for 45 min, 55° C for 30 sec and 72°C for 40 min followed by final extension of 7 min at 72°C. To check the relative fold in expression of *EcMyb1* gene for all stress treatments in comparison to control were calculated through ΔΔCT method. Two different housekeeping genes actin and tubulin were used as endogenous control.

Statistical analysis

For statistical analysis independent sample t-test was applied and mean values represented the measurements. The t-test results showed a significant effect of abiotic stressors on the relative gene expression at $P<0.05$.

Results And Discussion

Isolation and cloning of full-length gene encoding EcMyb1

In our previous study we have identified and isolated a partial gene (CDS) encoding MYB-TF from *Eleusine* coracana which was found to induced after drought stress exposure. To further understand the molecular role and regulation of the *EcMyb1*, we sought out to clone full length sequence of the gene. In order to clone the full-length gene, we isolated the RNA from the from *Eleusine coracana* leaves and reverse transcribed it to cDNA. The cDNA was used to amplify the full-length gene amplicon of *EcMyb1* using gene specific primers, and subsequently cloned in pGEM-T easy vector. The cloned amplicon was confirmed by sequencing. The *EcMyb1* full length was found to have 1053bp (Figure 1) and was translated using expasy tool to 350aa protein.

Characterizing EcMyb1 **sequence by homology search**:

Homology search is a very good method for characterizing newly identified sequences. The sequences that show considerable similarity are stated as homologous. Thus, the similarity searching helps to identify similar sequences with common ancestry in a reliable way. The extensively used bioinformatics tool for homology search is nBLAST for nucleotide sequence and pBLAST for protein sequence. The nBLAST and pBLAST tool were used to find out similarity of *EcMyb1* gene with myb gene sequences of other members of poaceae family.

It was observed that *EcMyb1* gene had 86.87 % sequence similarity with *Oryza sativa japonica* group myb16, 89.55% sequence similarity with *Sataria italica* transcription factor myb 61. Additionally, EcMYB 1 protein had 86.67% similarity with *Hordeum vulgare* transcription factor MYB86, 88.63% similarity with *Oryza sativa Japonica* Group transcription factor MYB61 and 86 % sequence similarity with *Triticum aestivum* MYB88 protein sequence (Table1). This result indicates that *EcMyb1* gene shows good similarity with Myb genes of different members of poaceae family which are directly or indirectly contributing in generation of drought tolerance in plants (Jadhav et al. 2018).

Table 1
Homology of *EcMyb1* gene with the other nucleotide and protein sequences using nBLAST and pBLAST tool

S.No.	Name	Accession number	Function	Query cover	E-value	% similarities	References
1	*Satariaitalica* transcription factor myb61	XM_004960378	Involved in regulation in stomatal movement	97%	0.0	89.55%	Liang, 2005
2	*Oryza sativa japonica* group myb16	AJ495784	Regulates cell morphogenesis	97%	0.0	86.87 %	Katiyar et al., 2012
3	*Saccharumbarberi*myb 16	HF546403	Regulates cell morphogenesis	97%	0.0	89.36 %	Katiyar et al., 2012
4	*Saccharumarundinaceum*myb 18	HF546406	control hypocotyl elongation responding to far-red light	97%	0.0	89.35 %	Zhang et al., 2018
5	*Satariaviridis*myb transcription factor 61	XM_034728679	involved in regulation in stomatal movement	97%	0.0	89.35%	Baldoni, 2015
6	*Sorghum bicolor*myb 61	XR_002447626	involved in regulation in stomatal movement	97%	0.0	89.16 %	Baldoni, 2015
7	*Saccharumofficinarum*myb 18	HF546401	Plays an important role in drought stress by regulating membrane biosynthesis, antioxidant regulation and osmolyte synthesis	97%	0.0	88.58	Shingote, 2017
8	*Hordeumvulgare* transcription factor MYB86	KAE8806364	Mainly associated with lateral root growth regulation.	97%	0.0	86.67	Oh, 2011
9	*Oryza sativa Japonica* Group transcription factor MYB61	XP_025881490	involved in regulation in stomatal movement	97%	0.0	88.63	Baldoni, 2015
10	*Triticumaestivum* MYB88	AFH08282	Genes which are associated with abiotic stress tolarence,positively regulated by MYB 88 TF.	97%	0.0	86.38	Xie, et al., 2010

Phylogenetic analysis

The homology of coding DNA and encoded amino acids sequences along with the extent of conservation of intron positions, protein and domain structure are compared to deduce the gene evolution and divergence. Additionally, by assimilating these parameters, phylogenetic tree can be constructed which shows homology with other gene or protein sequences. A phylogenetic tree using MEGA version 7.0 which is based on Neighbor joining BIONJ program was prepared with *EcMyb1* nucleotide and protein sequence as well as other similar sequences were obtained by nBLAST and pBLAST (Figure 2a and b). The bootstrapping value shows the number of times the same branch was displayed during the repetition of phylogenetic reformation on data re-sampling for phylogenetic tree. For *EcMyb1* gene 1000 bootstrap was used. It was predicted that *EcMyb1* nucleotide sequence was adjacent to *Triticum aestivum* followed by *Sorghum bicolor, Oryza sativa, Zea mays* and *Sateria italica.* Further EcMYB1 protein was nearest

to *Sateria italica* preceded by *Zea mays*, *Oryza sativa* and *Sorghum bicolor* and others as they show more recent common ancestors.

Motif Analysis

Spatial arrangements of enclosed structures that could probably be an adjacent sequence and are conserved short segments of protein 3D structure are stated as structural motifs. Motifs play a functional or structural role that help in recognizing new likely members of extant family or superfamily. The structural categorization of protein can be carried out by analyzing several structural motifs and their spatial arrangement for a given query protein structure (Voet and Voet 2006). Using ExPASY PROSITE motif search program it was observed that in EcMYB1 protein HTH conserved motif is present. It is a major characteristic of MYB transcription factor protein (Figure 2a) (Peters et al. 1987; Biedenkapp et al. 1988). Comparison of the expression of Myb genes in *Arabidopsis* and *Oryza* by genome wide comparative analysis revealed that HTH domain plays vital role in response to stress (Ogata et al.,1996). The MYB transcription factors recognize and bind specific sequence of DNA by HTH domain (Figure 3a). Various secondary metabolites are produced by phenylpropanoid pathway during the onset of drought stress. The pathway of phenylpropanoid is regulated by various members of R2R3-type MYB transcription factors. Moreover, MEME suite (Motif-based sequence analysis) was used for identification of location of motif for EcMYB1 protein (Figure 3b). It was predicted that EcMYB1 protein shows similar location of motifs as shown by other MYB transcription factor proteins of *Zea mays*, *Sateria italica* or *Saccharum barberi.*

Physiochemical properties of EcMYB1 protein

Expasy-protpraram program is routinely used for protein identification and analysis by researchers. It plays a pivotal role in detection of proteins through two-dimensional (2-D) gel and mass spectrometry. The procedure applied for protein identification involves matching of specific empirically attained information of the protein with the existing protein database. This helps in ascertaining whether the matched protein is a novel protein or a preexisting one (Gasteiger et al. 2005).

Various physiochemical properties of a protein such as molecular weight, isoelectrical point, total number of negatively and positively charged amino acids, atomic composition, extinction coefficients, estimated half life, instability index, aliphatic index and Grand average of hydropathicity (GRAVY) can be analyzed using Expasy-protpraram tool. Physiochemical characterization of EcMYB1 protein was carried out using this tool and results are presented in (Table 2).

Table 2
Physiochemical characterization of EcMYB1 protein using protparam

S.No	Physiochemical properties	Values
1	Molecular weight	38262.53 or 38.26 kDa
2	Isoelectrical point,	5.18
3	Total number of negatively charged amino acids of a protein,	(Asp + Glu) : 35
4	Total number of positively charged amino acids of a protein,	(Arg + Lys) : 26
4	Atomic composition,	$C_{1661}H_{2576}N_{464}O_{542}S_{17}$
5	Extinction coefficients,	47815M^{-1} cm^{-1}
6	Estimated half life,	30 hours (mammalian reticulocytes, *in vitro*). >20 hours (yeast, *in vivo*). >10 hours *(Escherichia coli*, *in vivo*).
7	Instability index,	57.76
8	Aliphatic index	61.94
9	Grand average of hydropathicity (GRAVY).	-0.586

The molecular mass of EcMYB1 protein was found to be 38.26 KDa. EcMYB1 protein was found to be made up of 5260 atoms. Extinction coefficient is a parameter to find out the quantity of light which is absorbed by a protein at a specific wavelength. The value of extinction coefficient for EcMYB1 protein range from 47440 M^{-1} cm^{-1} wavelength (where all Cys residues get reduced), to 47815 M^{-1} cm^{-1} (where all pairs of Cys residue are in cystines form) was reported by the software during analysis. Isoelectric point of EcMYB1 protein reported as 5.18 indicates that the protein is slightly acidic in nature. The protein contains a total of 26 positively charged amino acids (Arginine and Lysine) and 35 negatively charged amino acids (Aspartic acid + Glutamic acid). The result indicates that the overall charge on EcMYB1 protein is negative. Using ProtParam tool half life for a protein can be predicted by observing the N-terminal amino acid sequence of a protein. The importance of using N-terminal sequence is that these sequences play a major role in proteolytic degradation by ubiquitine. This is called 'N-end rule' which determines the stability of a protein (Ciechanover et al. 1989). This tool gives its prediction based on three organisms that are human (*in vitro*), yeast and *E.coli.(in vivo).* Half life of 30hours in mammalian reticulocytes (*in vitro*), >20 hours for yeast (*in vivo*) and >10 hours for *Escherichia coli* (*in vivo*) was predicted for EcMYB1 protein. Instability index for a protein gives an approximate idea of its stability in a test tube. Those proteins having instability index smaller than 40 are designated as stable proteins. However, proteins that have index greater than 40 are known as unstable proteins. Statistically, it was reported that there is significant difference in stable and unstable proteins. This specific difference between the two proteins is due to di-peptides that are involved in degradation of a protein. In unstable proteins the amount of these dipeptides is high as compared to stable ones. As per the findings, 12 unstable and 32 stable proteins have been discovered and MYB protein is one of the unstable proteins (Guruprasad, 1990). The instability index for EcMYB1 protein was predicted to be 57.76 which depicts that EcMYB1 is an unstable protein. The protein will be produced when it is needed for regulation of expression of drought tolerance gene and will be degraded when its role is fulfilled. Extant research shows that AtERF53 transcription factor expression increased under drought stress. Subsequently, it was also shown that after the activation of drought responsive genes, the ERF53 protein is subjected to ubiquitine mediated degradation (Cheng et al.2012; Hsieh et al.2013).

The relative volume occupied by aliphatic side chains constituting of alanine, valine, isoleucine, and leucine is known as aliphatic index of a protein. This is considered as positive factor which increases the thermostability of globular proteins. The aliphatic index of EcMYB1 protein was computed to be 61.94. Characteristically, higher the aliphatic index more will be the thermostability of a protein. For a protein or a peptide the Grand Average of Hydropathy (GRAVY) is the total hydropathy of all amino acids which is divided by total number of residues present in the sequence. A negative score of -0.586 was recorded for EcMYB1 protein which predicts that protein is soluble in nature and is an important characteristic of all transcription factors (Kyte et al. 1982; Katiyar et al. 2012). This is important as EcMYB1 protein is expressed in drought condition that is usually accompanied by high temperature.

Post Translational modification sites:

When DNA gets transcribed into RNA and translated into protein then some chemical alteration may occur in the amino acid chain of proteins known as post translational modifications or PTMs. This modification includes covalent joining of specific chemical groups like phosphates, sulphates etc. or lipid or carbohydrate moieties on proteins. PTMs can bring about diversification of the structure and function of a protein (Burkle 2013; Gui et al.2019). In protein phoshorylation, phosphate group (PO_4) is added to the polar head of amino acids. Due to this addition protein changes from hydrophobic (apolar) to hydrophilic (polar) which allows protein to interact with other molecules. Mostly phoshorylation occurs on serine (86.4%), threonine (11.8%) and is lowest in tyrosine (1.8%) (Ardito et al.2017).

Phosphorylation and protein-protein interaction plays major role in regulating the activity of plant MYB proteins. It can regulate transcription factor by altering their mechanism such as DNA binding capacity, stability and their interaction with other regulatory proteins (Kirchler et al. 2010). In EcMYB1 protein various phoshorylation sites were observed using NetPhos server (Figure. 4a). Total 42 sites for serine, 21 sites for threonine and 6 sites for tyrosine were computed. The sites for serine phosphorylation are mostly located at position 56-70, 225-250 and 290-300 in aminoacid sequence while the sites for threonine and tyrosine phosphorylation could be found at position 26,50, 60-65, 124,180 and at 220-224, 236, 300-305 respectively. It has been reported that Thr_{126} and Thr_{131} in Myb75of *Arabidopsis thaliana* are sites for MAP kinase phoshorylation. The phosphorylation event activates Myb75 which regulates stress response and secondary metabolism in *Arabidopsis* (Kreynes 2018). Past studies conducted also show that phosphorylation on serine 236 in PtMYB4 regulated the activity of the transcription factor (Morse et.al. 2009).

Other types of posttranslational modifications in proteins include O-linked glycosylation and N-linked glycosylation. (Steen 2008; Aebi 2013). In EcMYB1 protein, 20 O-linked glycosylation sites (Figure 4b) and 4 potential sites for N- linked glycosylation (Figure 4c) were predicted using NetNGlyc server. The process of Glycosylation is very important for cell viability because it plays a major role in protein folding as it attaches core N-glycans into membrane proteins and secreted glycoproteins in endoplasmic reticulum. It has been reported that O –linked glycosylation of transcription factor play an important role in regulation of RNA polymerase II activity (Jackson et al. 1988). It may be thought that o-linked glycosylation at different position in EcMYB1 protein may help in modulating the activity of RNA polymerase II involved in transcription of drought responsive genes.

Secondary Structure Annotation of EcMYB1 protein

When backbone atoms of a polypeptide chain interact with each other and form local folded structure, it leads to formation of secondary structure of a protein. For the prediction of EcMYB1 protein secondary structure SOPMA (self-optimized prediction method) was used (Geourjon et al. 1995). It was observed that EcMYB1 protein contains 28% alpha helix, 5.43% extended strand, 2.57% beta turn and 64% random coil (Figure 5). This result suggests that in EcMYB1 protein is not a compact globular protein but a protein containing alpha helix, extended strands and beta turns held together by random coils. The presence of alpha helix is important as the protein interacts with DNA via helix turn helix domain. Further, in a study on transcription factor Pdx1 using NMR revealed that the protein adopted a random coil structure in solution. This structure was found to be helpful in interaction of the transcription factor with other protein partners (Bastidas 2015).

EcMYB 1 protein 3-D structure prediction

For understanding the function of a protein it is necessary to collect knowledge about its tertiary or three dimensional structure (3-D). 3D structure of EcMYB1 protein was characterized by online tool SWISS MODEL which is based upon homology modeling approach. SWISS MODEL gives protein structure on the basis of similarity in sequence between template and query. In EcMYB1 protein 3D structures Myb type helix turn helix domain is seen (Figure 6a). EcMYB1 protein shows 52.78 % sequence similarity with template sequence which is R2R3 type MYB transcription factor. Ramachandran plot helps in visualizing energetically favored regions in protein structure. The plot gives an idea about the feasibility of secondary structure on the basis of Cα-C (psi ψ) and N-Cα (phi ϕ) angle (Figure 6b). Analysis of bond geometry in a protein provides a good estimate of existence of the protein. The feasibility of secondary structure of EcMYB1 protein was predicted using SWISS MODEL Mol Probity tool. Mol Probity is a web based service that provides structure validation and evaluation of model quality for proteins (Chen et al. 2010). Many studies have reported that a good protein model will usually contain 90% of its amino acid residues in the favorable regions of a Ramachandran plot (Laskowski 1993; Pramanik et al. 2018). The molprobity score for EcMYB1 protein was 1.05. Mol Probity score, which should be as low as possible, is a combined protein quality score that reflects the crystallographic resolution at which quality could be studied. The clash score which is generated due to overlapping of any two non bonding atoms in protein structure was zero. There were no bad angles or bad bonds in the structure (Figure 6). It can be concluded that EcMYB1 protein lies in energetically favored region and therefore could exist in nature.

QMEAN (Qualitative Model Energy Analysis) is a composite scoring function that describes almost all the major geometrical features of the structure of protein. The QMEAN Z-score gives a good estimate of the "degree of nativeness" of the structural features observed in the model on a global scale (Benkert et al.2011). The score provides a comparative understanding of the derived structure with the experimental structure of similar size. An approximate QMEAN- Z score of zero stipulate the higher quality agreement between the experimental and modeled structure. However the score of – 4.0 or below shows a low quality model. The QMEAN Z-score value for EcMYB 1 protein showed -0.38 (figure 6c) which indicates that the proposed homology model has good reliability, shows adequate fit and is acceptable.

Cellular localization of EcMYB1 protein

CELLO2GO makes use of BLAST homology searching approach to look for homologous sequences for a query protein sequence. The obtained homologous sequences are GO (gene ontology) annotated in a data base derived from the UniProt Knowledge Base database. CELLO2GO predicts molecular function, cellular compartment and biological process of the query protein and display it as pie chart (Yu et al. 2014).

The CELLO2GO functional annotation result for EcMYB1 protein project that the protein is a transcription factor/ regulatory protein involved in transcriptional regulation of genes. EcMYB1 protein which is thought to be a transcription regulatory protein will exhibit its role by binding to DNA sequences specifically for recognition of cis-acting elements. It may bind DNA sequences to support the specific interaction of DNA and protein (Figure 7).

Promoter analysis of EcMyb1 **gene**

Promoter of a gene is located upstream of the coding regions of the gene and facilitates optimum expression of the gene by enabling binding of various transcription factors with their *cis*- acting regulatory sequences. Promoters also help in binding of RNA polymerase to initiate transcription. The promoters may be active throughout the life cycle of an organism in tissues (constitutive promoter) or active in some specific tissues. It can be developmental stage of the organism (tissue specific promoter) or promoter may require some stimulus for activation (inducible promoter) (Biłas et al. 2016). The *cis*-regulatory sequences of the promoter need to be identified and studied in order to understand the expression of the gene. The present study helps in determining and designing expression cassette for exogenous expression of the gene.

Analysis of cis**-regulatory elements**:

In NCBI whole genome sequence of *Eleusine corocana* is already present with the Accession number: GenBank: LXGH01099917.1 which was further BLAST with Nucleotide sequence of cloned DNA fragment using bioinformatics tool nucleotide blast (nBLAST). Subsequently, the upstream region of the sequence which was determined through similarity between aforementioned sequences was selected and subjected to promoter analysis. Nearly 2 kb nucleotides were scanned using various softwares for determining the regulatory elements in the promoter region.

Analysis of CpG/CpNpG island and Tandem repeats by PlantPAN

Plant PAN is an important platform for analysis of plant promoters and helps in creating transcriptional regulatory networks. Chemical modifications like DNA methylation, chromatin rearrangements and histone remodeling are affected by the epigenetic alterations that are hereditary in gene expression. CpG islands are non-methylated DNA sequences present in plant genome that are rich in G and C nucleotides. The characteristic feature of CpG island is that most of them are sites of transcription initiation. Presence of CpG island in the promoter region of a gene may make the gene transcriptionally active by facilitating nucleosome remodeling and recruiting other regulatory proteins at the site of transcription initiation (Deaton et al. 2011). They are uniquely associated with genes, encode more RNA polymerase II binding sites than other promoters and are present in a tissue-specific manner compared to other CpG islands promoters (Elango et al. 2011). PlantPAN employs following criteria to characterize CpG islands:

(1) GC content should be above 50%

(2) Length of CpG/CpNpG region should be greater than 200 bp

(3) Ratio of observed to expectedCpG dinucleotide number should be above 0.6.

Table 3 shows the CpG/CpNpG analysis by PlantPAN tool that shows the presence of CpG/CpNpG islands in *EcMyb1* gene towards the 3' end in a distal promoter region.

Table 3
Location of CpG islands in the promoter region of *EcMyb1* gene

Begin site	End site	Length	G+C Frequency	CpG o/e ratio
707	1266	554	0.5	0.67

Presence of CpG island in the promoter region of *EcMyb1* gene indicate that the gene is transcriptionally active and the level of expression is not intermediate.

Tandem repeats occur in DNA when a pattern of one or more nucleotides is repeated and the repetitions are directly adjacent to each other. On the basis of repeating unit DNA tandem repeats can be classified into following three sets:

1. Microsatellite with repeat unit less than 9 nucleotides in length
2. Minisatellite with 6-100 bp repeats mostly around 15 bp

(iii)Megasatellite tandem repeats with longer units, length is more than 135nucleotides (Mehrotra et al. 2014). Tandem repeat in the promoter region of a gene is associated with transcriptional regulation of the gene. Sometimes, the tandem repeats are binding sites of transcription factors and bring about increased expression of the gene (Richard et al. 2009). Analysis of tandem repeats in promoters revealed that repeating unit likely plays a role in transcription and regulation of *gene* expression (Tian et al. 2017).

As represented by Table 4, the presence of DNA tandem repeats is shown in the upstream promoter region of *EcMyb1* gene. The repeat is 278 nucleotides in length and repeats twice in the promoter region. It may play a role in regulating the conditional expression of the *EcMyb1* gene.

Table 4
List of Cis regulatory motifs in promoter region of *1 EcMyb1* gene by PLACE software

Location	Period size	Copy number	Consensus size	Percent match	Nucleotides				Entropy (0-2)
					A	T	C	G	
1-518	278	1.9	278	98	29	18	28	23	1.98

Analysis of Regulatory Elements

The transcription of a gene is controlled by the *cis* regulatory elements which are non-coding regions of the DNA. These elements function as molecular switch by being present in promoter sequence and controlling the transcriptional regulation (Banu et al. 2014). Further, the program PlantCare and PLACE was utilized to scan the promoter sequence upto 2 Kbp upstream from translation commencement site of *EcMyb1* gene of *Eleusine coracana*. This process helped in analyzing 13 *cis* acting regulatory elements, as shown through Figure 8. The length of *cis* acting regulatory elements varied from 4- 10 bp in *EcMyb1* gene. This scanning led to the identification of various *cis* acting elements such as ABA responsive elements (CACT GG), stress responsive elements (AAGG GG), Myb recognition sites (CAACAG), Myb binding sites (CAAC TG) etc.

ABA responsive elements

Accumulation of abscisic acid (ABA) in plant cells is one of the rapid responses of drought stress (Hsiao 1973) which modulates ABA inducible gene expression (Yamaguchi-Shinozaki et al. 2006) and stomatal closure for shrinking water loss due to transpiration (Schroeder et al. 2001**)**. In *Arabidopsis thaliana* many downstream ABA signaling components were identified. ABA binds to the ABA receptor regulatory components of ABA receptor (RCAR) further, RCAR inactivates type 2C protein phosphatase (PP2Cs) such as ABSCISIC ACID INSENSITIVE (ABI 1). PP2Cs and RCAR together work as co-receptor and creates high affinity ABA binding site. Also inactivation of PP2Cs causes suppression of PP2C-mediated de-phosphorylation of Sucrose non-fermenting Kinase-1-Related protein kinase 2s (SnRK2s), which are important positive regulators of ABA signaling. As a result, activated SnRK2s target ABA-dependent gene expression and ion channels. Phosphorylated SnRK2s subsequently phosphorylate the ABA-responsive element Binding Factors (ABFs) which are basic leucine zipper transcription factors that bind to ABA-Responsive Elements (ABRE), the major *cis*-element in the promoter region of downstream genes that are induced by ABA (Fernando et al. 2016). Presence of abscisic acid binding site in the promoter region of a gene indicates that the gene is expressed in response to ABA signaling pathway. Promoter analysis of *EcMyb1* gene by PlantCARE (Figure 8) and PLACE softwares (Table 5) revealed presence of ABA-Responsive Elements (ABRE) in the promoter. Presence of ABRE in *EcMyb1* promoter indicates that some protein of ABA signaling pathway binds to this element and upregulates the expression of *EcMyb1* gene at drought condition. This is supported by the study in which the expression of *EcMyb1* gene takes place not only in the drought tolerant genotype but also in the drought sensitive genotype of *Eleusine coracana* on exogenous application of abscisic acid (Kumari et al. 2017).

Table 5
Prediction of various regulatory elements present in *EcMyb1* gene promoter using PLACE tool

Motifs	**Reference Species**	**Sequence**	**Functions**	**References**
TATA	*Ipomoea batatasPhaseolus vulgaris*	TATATAA	cis elements and trans-acting factors affecting regulation of a nonphotosynthetic light-regulated gene .	(Grace et al. 2004)
CAAT	*Pisum sativum*	CAAT	Sequences responsible for the tissue specific promoter activity of a gene.	(Shirsat et al. 1989)
GATA	*Petunia hybrida Arabidopsis thaliana, Oryza sativa*	GATA	Responsible for the tissue specific promoter activity.	(Teakle et al.2002)
MYB core element	*Arabidopsis thaliana*	CAGTTG	involved in regulation of genes that are responsive to water stress	(Urao et al. 1993)
MYb recognition site	*Arabidopsis thaliana*	CAACTAG	function as transcriptional activators in abscisic acid signaling.	(Abe et al. 2003)
ASF-1 binding site" in CaMV 35S promoter	Arabidopsis thaliana	TGACG	TGACG motifs are found in many promoters and are involved in transcriptional activation of several genes by auxin and/or salicylic acid; May be relevant to light regulation; Binding site of tobacco TGA1a; TGA1a and b show homology to CREB; TGA6 is a new member of the TGA family; Abiotic and biotic stress differentially stimulate "as-1 element" activity;	(Redman et al. 2002)
Binding site for MYC	Arabidopsis thaliana	CACATG	dehydration-resposive gene, rd22; MYC binding site in rd22 gene of Arabidopsis thaliana; ABA-induction	(Abe et al. 1997)
MYC recognition site	Arabidopsis thaliana	CAGTTG	MYC recognition site found in the promoters of the dehydration-responsive gene rd22 and many other genes in Arabidopsis; Binding site of ATMYC2 (previously known as rd22BP1); MYC recognition sequence in CBF3 promoter; Binding site of ICE1 (inducer of CBF expression 1) that regulates the transcription of CBF/DREB1 genes in the cold in Arabidopsis; ICE1 This sequence is also known as RRE (R response element)	(Chinnusamy et al. 2004, Abe et al. 2003)

Stress responsive elements:

The primary and specific metabolism gene expression at the transcription stage is regulated by stress through explicit *cis*-elements binding. A probable molecular link is arranged through the patterns of binding by transcription factor in assorted metabolic pathways. It is also due to the existence of different *cis*-element signatures upstream to varying stress responsive genes (Sheshadri et al. 2016). During the onset of drought stress there is production of important metabolic proteins which are involved in the synthesis of compatible solutes and other regulatory proteins involved in signal transduction pathways. The transcriptional genes are directly controlled by a network of TFs and transcription factor binding sites (Ciarmiello et al. 2011). In the promoter region of *EcMyb1* gene, Myb binding sites (CAAC TG), Myb core element (CAGTTG), Myb recognition sites (CAACTAG), Myc recognition sites (CAGTTG), binding sites for Myc (CACATG) etc were identified along with the usual TATA and CAAT box (Table 5 and Figure 8). These elements are associated with genes responsive to water stress. Presence of these elements in the promoter region of the *EcMyb1* gene also indicates that the gene is a drought responsive gene. Expression of the gene in different water stress condition may be facilitated by binding of either a specific Myb or a Myc or both proteins.

Expression of EcMyb1 gene in response to abiotic stresses in seedling stage

Drought stress

For imposing drought stress finger millet seedlings were grown in PEG medium at three different concentrations. In present study PEG-6000 was used, molecular weight more than 3000 cannot enter the cell wall space therefore, does not show any harmful effects on plants growth and creates significant water stress (Meher et al., 2018). It was observed that as the concentration of PEG increases relative expression fold also increases as compared to control plant. The expression fold increases from 12 to 28 fold as compare to control (Figure 9).

Salt stress

In plants NaCl causes osmotic imbalance due to generation of reactive oxygen species. Present study indicated that as the salt stress increases the expression of *EcMyb1* gene also increases. In higher salt concentration (200mM) the expression fold increases 29 fold as compared to minimum salt concentration (100mM) (Figure 9).

Heat stress

Heat stress causes detrimental effect on plant activities including seed germination, growth development, photosynthesis and reproduction. Heat stress causes direct accumulation of toxic compounds such as reactive oxygen species besides it also disturbs cellular homeostasis. *EcMyb1* gene also shows enhanced expression during heat stress (Figure 9). Relative expression of *EcMyb1* gene enhanced from 19 to 46 fold. This indicates that *EcMyb1* gene was induced in response to all stresses. In other words gene expression is induced with the onset of drought or other abiotic stresses.

Conclusion

Abiotic stress tolerance is a multi-genic quantitative trait involving complex genetic control. At present none of the research have been successfully carried out that develop tolerant cultivars by targeting a single gene. Therefore this research holds high significance and adds to the existing literature of master regulators that modulate expression of related genes. The master regulators are the transcription factors that are induced during stress conditions and bind specifically to the promoters of downstream genes associated with stress tolerance. They build strong association with general transcription factor at promoter of target gene and regulate their expression. Association of transcription factors with each other developed in response to intracellular signals leads to activation or repression of target genes. In this study, we have cloned a full-length coding sequence of abiotic stress inducible gene encoding Myb- transcription factor and submitted to NCBI (Accession number MT312253). The expression of the gene increased manifolds during abiotic stress treatment indicating it to be a stress-responsive gene. The stress inducible expression of the gene well correlated with the presence of different stress responsive *cis*-regulatory elements in its promoter region. Expression profiling and Insilco analysis of *Ecmyb1*gene strongly indicated its role in abiotic stress response in *Eleusine corocana* and recommends it as a potential candidate gene for producing abiotic stress tolerance in plants.

References

1. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 15:63–78
2. Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, Shinozaki K (1997) Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell 9:1859–1868
3. Aebi M (2013) N-linked protein glycosylation in the ER. Biochimica et BiophysicaActa. Molecular Cell Research 1833:2430–2437
4. Allan AC, Hellens RP, Laing WA (2008) *MYB* transcription factors that colour our fruit. Trends Plant Science 13:99–102
5. Ambawat S, Sharma P, Yadav NR, Yadav RC (2013) *MYB* transcription factor genes as regulators for plant responses: an overview. Physiology Molecular Biology of Plants 19:3307–3321
6. Ardito F, Giuliani M, Perrone D, Troiano G, Muzio LL (2017) The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy. Int J Mol Med 40:271–280
7. Bailey TL (2011) DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics 27(12):1653–1659
8. Baldoni E, Genga A, Cominelli E (2015) Plant MYB transcription factors: their role in drought response mechanisms. Int J Mol Sci 16(7):15811–15851
9. Banu SA, Huda KM, Tuteja N (2014) Isolation and functional characterization of the promoter of a DEAD-box helicase Psp68 using Agrpobacterium-mediated transient assay. Plant Signalling amp: Behavior 91(1):8–9
10. Bastidas M (2015) Structural and functional characterization of the transcription factor Pdx1 and its interactions with DNA and proteins. (Doctoral dissertation, The Pennsylvania State University)
11. Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27(3):343–350
12. Biedenkapp H, Borgmeyer U, Sippel AE, Klempnauer KH (1988) Viralmyb oncogene encodes a sequence-specific DNA-binding activity. Nature 335:835–837
13. BiłasR SK, Hnatuszko-Konka K, Kononowicz AK (2016) Cis-regulatory elements used to control gene expression in plants. Plant Cell Tissue Organ Culture(PCTOC) 127:2269–2287
14. Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences, 14. Electrophoresis, pp 1023–1031
15. Burkle A, Uversky VN (2013) Posttranslational Modification, 3. Elsevier, pp 425–430
16. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. ActaCrystallographica Section D: Biological Crystallography 66(1):12–21
17. Cheng MC, Hsieh EJ, Chen JH, Chen HY, Lin TP (2012) Arabidopsis RGLG2, Functioning as a RING E3 Ligase, Interacts with AtERF53 and Negatively Regulates the Plant Drought Stress Response. Plant Physiol 158:363–375
18. Chinnusamy V, Schumaker K, Zhu JK (2004) Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. J Exp Bot 55:225–236
19. Chow C, Zheng H, Wu N, Chien C, Huang H, Tzong-Yi Lee T (2015) PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. Nucl Acids Res 44:1154–1160
20. Ciarmiello LF, Woodrow P, FuggiA, Pontecorvo G, Carillo P (2011)Plant Genes for Abiotic Stress:Mechanisms and Adaptations in Plant Genes for Abiotic Stress.Intech Publisher,Londonpp 283–308
21. Ciechanover A, Schwartz AL (1989) How are substrates recognized by the ubiquitin-mediated proteolytic system? Trends Biochem Sci 14:483–488
22. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes &Development,25:101010-22
23. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYBtranscription factors in ArabidopsisTrends. in Plant Science 15:10573–10581
24. ElangoN,Soojin VY (2011) Functional relevance of CpG island length for regulation of gene expression Genetics187:4 1077-83
25. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL (2015) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285

26. Gasteiger E, Gattiker A, Hoogland C, IvanyiI., Appel RD, Bairoch A (2005)Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook Humana press 571–607
27. Geourjon C, Deleage G (1995) SOPMA: Significant improvements in proteinsecondary structure prediction by consensusprediction from multiple alignments. Oxford University Press 11:681–684
28. Grace ML, Chandrasekharan MB, Hall TC, Crowe AJ (2004) Sequence and spacing of TATA box elements are critical for accurate initiation from the beta-phaseolin promoter. Journal of Biology Chemistry 279:8102–8110
29. Gui J, Luo L, Zhong Y, Sun J, UmezawaT, Li L (2019) Phosphorylation of LTF1, an MYB Transcription Factor in Populus, Acts as a Sensory Switch Regulating Lignin Biosynthesis in Wood Cells. Mol Plant 12:1325–1337
30. Gupta SM, Arora S, Mirza N, Pande A, Lata C, Puranik S, Kumar J, Kumar A (2017) Finger Millet: A"Certain" Crop for an "Uncertain" Future and a Solution to Food Insecurity and Hidden Hunger under Stressful Environments. Frontiers in plant science 8:643
31. Guruprasad K, Reddy BVB, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng 4(2):155–161
32. Higo K, Ugawa Y, Iwamoto M, KorenagaT (1999) Plant cis-acting regulatory DNA elements (PLACE) database. Nucl Acids Res 27:1297–1300
33. Hsieh EJ, Cheng MC, Lin TP (2013) Functional characterization of an abiotic stress-inducible transcription factor AtERF53 in Arabidopsis thaliana. Plant Mol Biol 82:223–237
34. Jackson SP, Tjian R (1988) O-glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation, 55. Cell, pp 125–133
35. JadhavP, Salvi P, Bhatt M, Lohani P (2018) Expression ofEcMYB Transcription Factor Gene Under Different Abiotic Stress Conditions in *Eleusine coracana*. International Journal of AgricultureEnvironment Biotechnology 11:5 799–806
36. Katiyar A, Smita S, Lenka SK, Rajwanshi R, Chinnusamy V, Bansal K (2012) Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. BMC Genomics 13(1):544
37. Kirchler T, Briesemeister S, Singer M, Schutze K, Keinath M, Kohlbacher O, Carbajosa JV, Teige M, Harter K, Chaban C (2010) The role of phosphorylatable serine residues in the DNA-binding domain of Arabidopsis bZIP transcription factors. Eur J Cell Biol 89:175–183
38. Kreynes AE (2018) Phosphorylation of MYB75 Transcription Factor by MAP Kinases in *Arabidopsis thaliana*
39. Kumar S, StecherG,Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular BiologyEvol 33:1870–1874
40. Kyte J, Russell F, Doolittle A (1982) Simple Method for Displaying the Hydropathic Character of a Protein. J Mol Biol 157:105–132
41. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26(2):283–291
42. Lescot M, Dhais P, Thijs G, Marchal K, Moreau Y (2002) PlantCARE a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. NucleicAcids Res 30:1325–1327
43. Liang YK, Dubos C, Dodd IC, Holroyd GH, Hetherington AM, Campbell MM (2005))AtMYB61, an R2R3-MYB transcription factor controlling stomatal aperture in *Arabidopsis thaliana*. Curr Biol 15(13):1201–1206
44. Lippold F, Sanchez DH, Musialak M, Schlereth A, Scheible WR, Hincha DK, Udvardi MK (2009))AtMyb41 regulates transcriptional and metabolic responses to osmotic stress in Arabidopsis. Plant Physiol 149:1761–1772
45. Mehrotra S, Goel S, Raina SN, Rajpal VR) (2014) Significance of satellite DNA revealed by conservation of a widespread repeat DNA sequence among angiosperms,Applied. Biochemistry Biotechnolog 173:71790–71801
46. Ogata K, Kanei-Ishii C, Sasaki M, Hatanaka H, Nagadoi A, EnariM,Nakamura H, Nishimura Y, Ishii S, Sarai A (1996) The cavity in the hydrophobic core of Myb DNA binding domain is reserved for DNA recognition and transactivation. Nat StructBiol 3:178–818
47. Oh JE, Kwon Y, Kim JH, Noh H, Hong SW, Lee H (2011) A dual role for MYB60 in stomatal regulation and root growth of Arabidopsis thaliana under drought stress. Plant Mol Biol 77(1-2):91–103

48. Peters CWB, Sippel AE, Vingron K (1987) MKHDrosophila and vertebrate myb proteins share two conserved regions, one of which functions as a DNA-binding domain. EMBO J 6:3085–3090
49. Pramanik K, Kundu S, Banerjee S, Ghosh PK, Maiti TK (2018) Computational-based structural, functional and phylogenetic analysis of Enterobacterphytases. 3 Biotech 8(6) 262
50. Redman J, Whitcraft J, Johnson C, Arias J (2002) Abiotic and biotic stress differentially stimulate as-element activity in Arabidopsis. Plant Cell Rep 21:180–185
51. Salvi P, Arora A, Lohani P (2012) Expression of MYB Transcription Factor in Eleusinecoracana: towards making of drought tolerant plants.LAP. Lambert Academic Publishing USA
52. Sanchita, Blessy BM, 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 Journal 6:4278–4285
53. Schroeder JI, KwakJM, Allen GJ (2001) Guard cell abscisic acid signalling and engineering drought hardiness in plants.Nature 410: 6826327
54. Segarra G, Vander Ent S, Trillas I, Pieterse CMJ (2009) MYB72, a node of convergence in induced systemic resistance triggered by a fungal and a bacterial beneficial Microbe. Plant Biol 11:90–96
55. Sheshadri SA, NishanthMJ,Simon B (2016) Stress Mediated cisElement Transcription Factor Interactions Interconnecting Primary and Specialized Metabolism in planta.Frontiers in. Plant Sci 7:1–23
56. Shingote PR, Kawar PG, Pagariya MC, Rathod PR, Kharte SB (2017) Ectopic Expression of SsMYB18, a Novel MYB Transcription Factor from *Saccharum spontaneum* Augments Salt and Cold Tolerance in Tobacco. Sugar Tech 19(3):270–282
57. Shirsat A, Wilford N, Croy R, Boulter D (1989) Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. Mol Gen Genet 215:326–331
58. Sigrist CJ, De Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A (2012) New and continuing developments at PROSITE. Nucleic Acids Res 41:D344–D347
59. Steen PVD, Rudd PM, Dwek RA, Opdenakker G (2008) Concepts and Principles of O-Linked Glycosylation. Critical Reviews in Biochemistry Molecular Biology33(3):151–208
60. Teakle GR, Manfield IW, Graham JF, Gilmartin PM (2002) Arabidopsis thaliana GATA factors: organisation, expression and DNA-binding characteristics. Plant Mol Biol 50:43–57
61. Tian SL, Li Z(2017) Analysis of tandem repeat units of the promoter of capsanthin/capsorubin synthase (Ccs) gene in pepper fruit.PhysiolMolBiol Plants23: 3685–691
62. Urao T, Yamaguchi-Shinozaki K, Urao S, Shinozaki K (1993) An Arabidopsis myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. Plant Cell 5:1529–1539
63. Voet D, Voet JG (2006) Protein Structural Motifs: Identification, Annotation and Use in Function Prediction. Via and Tramontano 145–171
64. Xie Z, Li D, Wang L, Sack FD, Grotewold E (2010) Role of the stomatal development regulators FLP/MYB88 in abiotic stress responses. Plant J 64(5):731–739
65. Yamaguchi-Shinozaki K, Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol 57:781–803
66. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein structure and function prediction. Nature methods 12(1):7–8
67. Yang S, Vanderbeld B, Wan J, Huang Y2010). Narrowing down the targets: towards successful genetic engineering of drought-tolerant crops.Mol Plant3469–490
68. Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, Lu CH (2014) CELLO2GO: a web server for protein subCELlularLOcalization prediction with functional gene ontology annotation. PloS one 9(6):.e99368
69. Zhang T, Zhao Y, Wang Y, Liu Z, Gao C (2018) Comprehensive analysis of MYB gene family and their expressions under abiotic stresses and hormone treatments in *Tamarix hispida*. Front Plant Sci 9:1303

Figures

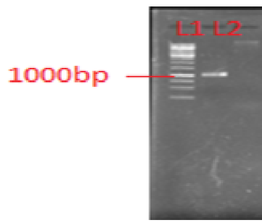


Figure 1

PCR amplicon using EcMyb1gene specific primers

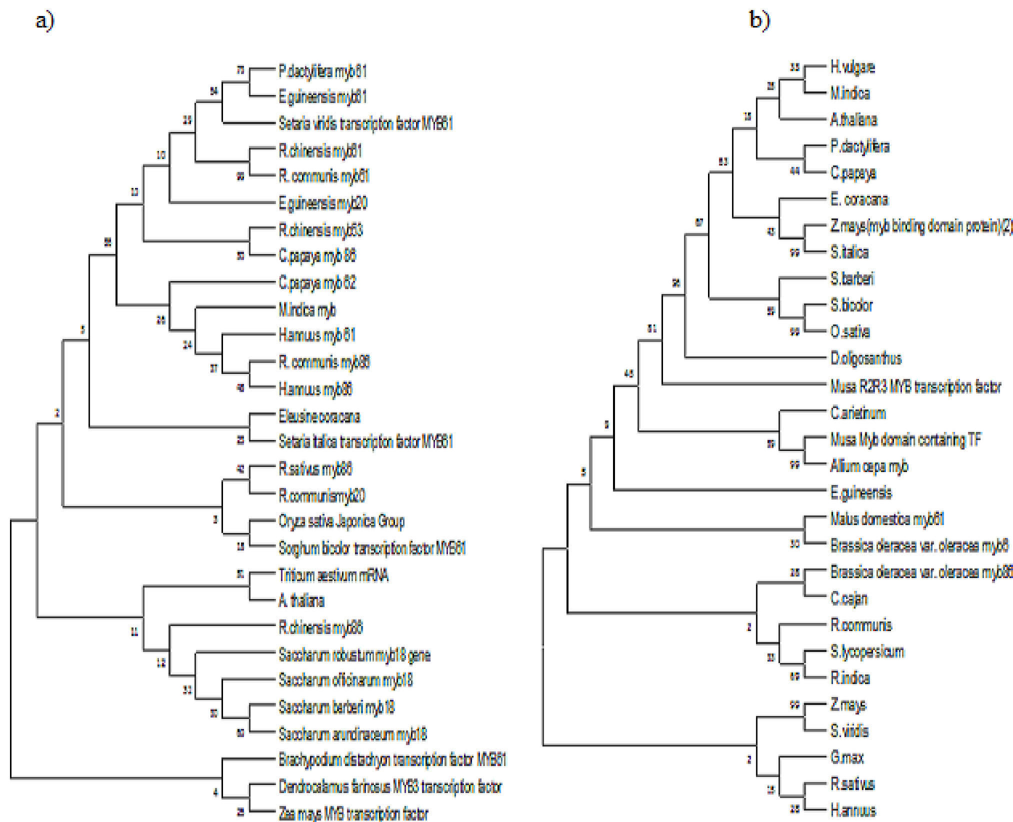
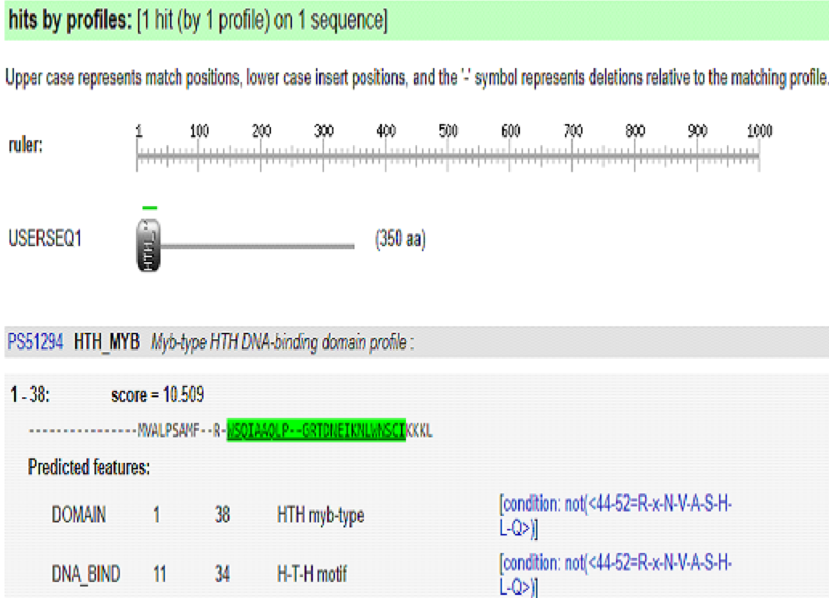


Figure 2

a) Phylogenetic tree of EcMyb1gene using Neighbor joining BIONJ program of MEGA version 7.0.with sequences obtained from nBLAST. Arabidopsis thaliana mybtranscription factor (TF),Carica papaya mybTF62, Carica papaya TFmyb86, Elaeisguineensismyb 20, Elaeisguineensismyb TF 61, Helianthus annuusTF myb61, Helianthus annuusmyb TF 86, SetariaitalicaTF MYB61, Zea mays MYB TF, Saccharumbarberimyb TF 18, Magniferaindicamyb TF, Phoenix dactyliferamyb TF 61, Ricinuscommunismyb TF 61, Ricinuscommunismyb TF 86, Rosa chinensismyb TF53, Rosa chinensismyb TF 61, Rosa chinensismyb TF86, Setariaviridis TF MYB61, Saccharumarundinaceummyb TF 18, Sorghum bicolor transcription factor MYB TF 61, Ricinuscommunismyb TF 20, Raphanussativusmyb TF 86Saccharumofficinarummyb TF18, Oryza sativa Japonica Group myb TF, Triticumaestivum mRNA, Brachypodiumdistachyon TF MYB61, Saccharumrobustummyb TF 18 gene, Dendrocalamusfarinosus MYB3 TF. b) Phylogenetic tree of putative EcMYB1 protein with different MYB Transcription factors of Dichantheliumoligosanthes, Zea mays, Oryza sativa Japonica Group, Hordeumvulgare, Arabidopsis thaliana, Sorghum bicolor, Allium cepa, Setariaitalica, Saccharumbarberi, Elaeisguineensis, Phoenix dactylifera, Setariaviridis, Carica papaya, Raphanus sativus, Rosa indica, Solanumlycopersicum, Helianthus annuus, Ricinus communis, Brassica oleracea var. oleraceamyb 6, Cicerarietinum, Brassica oleracea var. oleraceamyb 86, Malusdomestica myb61, Malusdomesticamyb 86, MusaMyb domain containing TF, Musa R2R3 type

Myb TF, Glycine max, Cajanuscajan, Mangifera indica and Putative MYB DNA-binding domain superfamily protein Zeamayswas constructed using Neighbor joining BIONJ program of MEGA version 7.0.

a)



b)

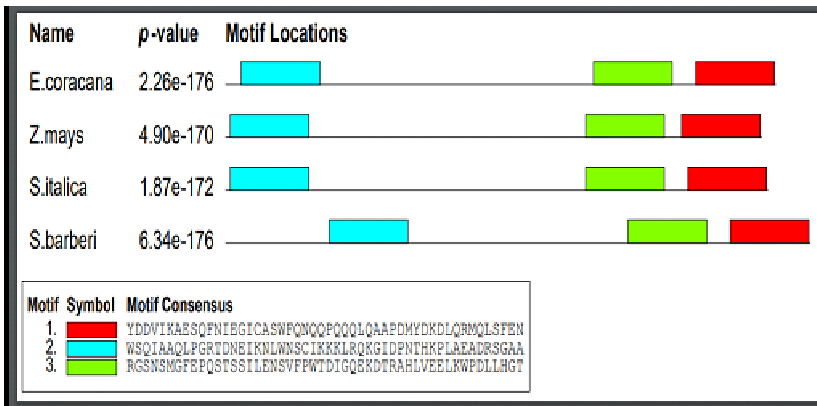
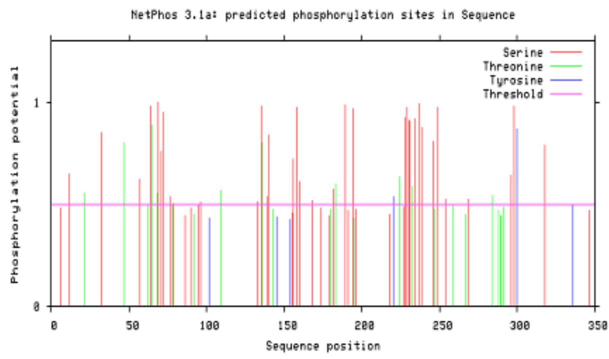


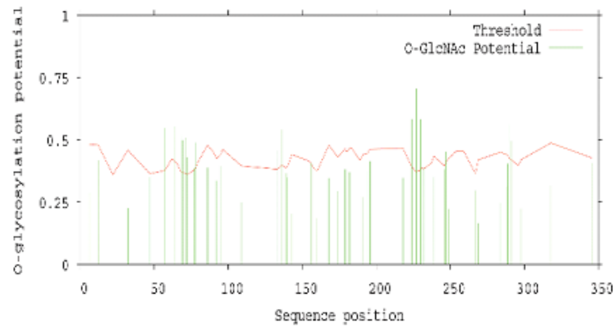
Figure 3

a) ExPASY PROSITEmotif search program showed the presence of Myb type HTH (Helix turn Helix) DNA binding domain in EcMYB1 protein.b) Location ofMotifs were identified using the MEME suite (Motif-based sequence analysis tools) observed that the translated sequences of MYB protein ofE.coracana, Zea mays, Sateriaitalica and Saccharumbarbericonsisted of similar motifs location.

a)



b)



c)

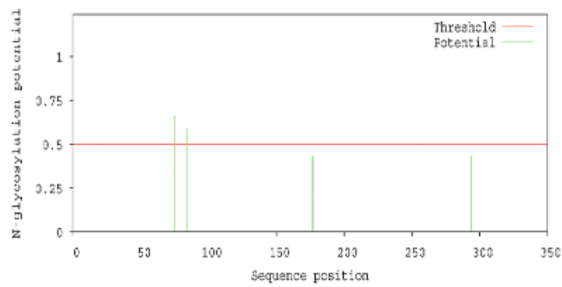


Figure 4

Prediction of post translational sites for EcMYB1 protein using NetPhos server and NetNGlyc 1.0 server (A) Phosphorylation sites (B) O-linked Glycosylation sites (C) N-linked Glycosylation sites

Alpha helix	(Hh) :	98 is	28.00%
3_{10} helix	(Gg) :	0 is	0.00%
Pi helix	(Ii) :	0 is	0.00%
Beta bridge	(Bb) :	0 is	0.00%
Extended strand	(Ee) :	19 is	5.43%
Beta turn	(Tt) :	9 is	2.57%
Bend region	(Ss) :	0 is	0.00%
Random coil	(Cc) :	224 is	64.00%
Ambiguous states (?)	:	0 is	0.00%
Other states	:	0 is	0.00%

Figure 5

Secondary structure of EcMYB1 protein computed using SOPMA web based tool

a) b)

c)

MolProbity Results

MolProbity Score	1.05
Clash Score	0.00
Ramachandran Favoured	90.00%
Ramachandran Outliers	0.00%
Rotamer Outliers	0.00%
C-Beta Deviations	0
Bad Bonds	0 / 267
Bad Angles	0 / 358

Results obtained using MolProbity version 4.4

Figure 6

(a) 3D structure of EcMYB1 protein by SWISS model b) Ramchandran plot of EcMYB1 protein indicating 90% residues in favored region c) Quality estimate of EcMYB1 protein based upon QMEAN value quality score

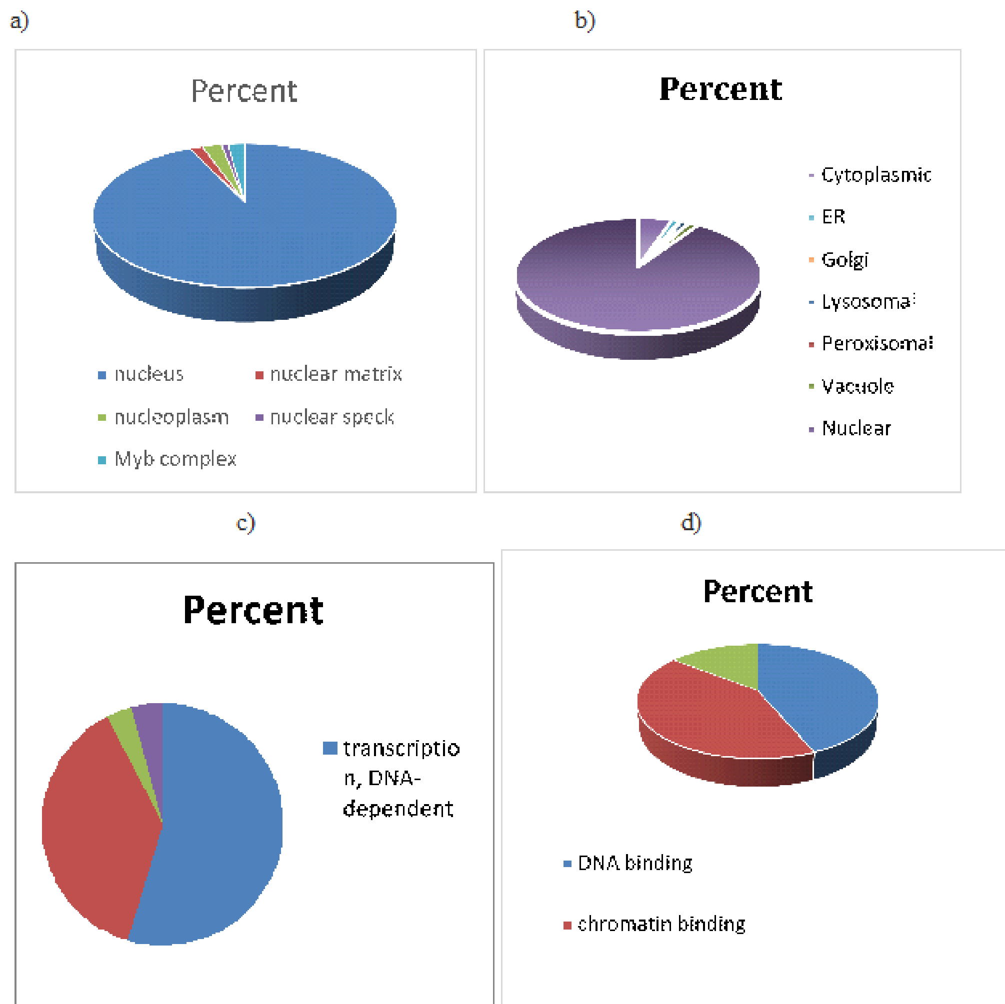


Figure 7

CELLO2GO tool shows subcellular localization of EcMYB1 protein a) cellular components b) localization of MYB protein c) EcMYB1 protein involved in many biological processes d) Molecular functions performed by EcMYB1 protein

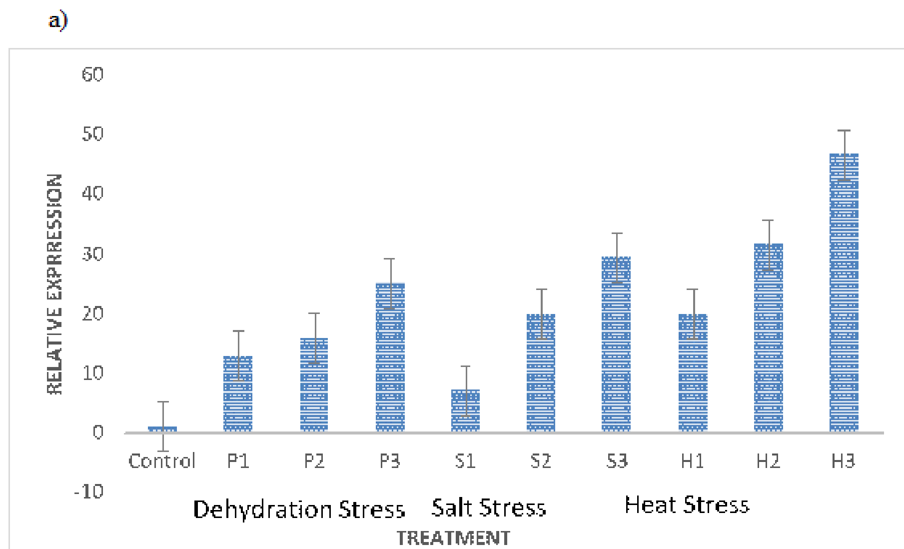
>PlantCARE_10938
+ CACTGCAATG ATCAGTGGCT CCATTGACGC TAGTGACAGT CAATCAGACA AAGCATA ATA CCTTGTACCT
+ GATCAACCGC CTGTTCAATT ATAGGAAGCC CAATTCTTAA CCCCGGGCCT GCAACAT GCC TCTTGTTTGA
+ GTGGAGAAGT AAAGATTTCT GATTAGAATT GCTAGCTAGC AACCTAGCTG CCCGAGA TAA ACAAAAGGAA
+ GATGCCCTCT GTCCTGGAAT ACGAATATCT GATGCCACCT TTTTTCCCCA GGAC
MYB binding site(+267)
CCCAAC TGTGATGATT
+ GACGAAGGCT TTTCCTAGAG AGAAGTGATC CAGTGAGTTA CCTGAAAAGC AAGAGG GAAA AGGGACAACA
+ CTTAGGTCTG ACACATAAAA TCTTGCAGGT GGGAAGAAAA TGAGGTGCAC ATGCACG CAG GAACAGCAGC
MYB-Recognition-site(+443)
+ TAACAAAACG AGTTTTAATT GGCAACAGAA GAAAGGTGGA CAACTATTAC CAACTGT GTG CTTTGCCCTG
+ CATGGAGCAC CCGTGATGCA GAAATAGCTG CAGAAAAGAA GGTTTCTAGT GGCAGT GAAC TAGTTCTGGA
+ CCACAGTAAT TCGCTATGAA TTGGGAGCAA GAGAATGGAT CCATTTCCCA ATTATGT ATG AGCAGATGGA
ABA-responsive-element(+952)
+ CCCGTTTCAC TTCACTGGCT TGTGCTCGAA CAAATACACT GGGTACGGTC GGA
Stress responsive element(+966)
TGAAGG GGAATGTGGG
+ AAGAAATAAC AGGTTCTAGG CAACAGAAGA AAGGTGGACA ACTATTACCA ACTGTG TGCT TTGCCCTGCA

TATA BOX(+141)
+ ACAGTTGCTC AATTCTCGTC CTAGCTGCAG CCCAATCATT CACCGATATA TAAATCAC TA CAGCCTCTGC
CAATBOX(+41)
+ ATTTTCCGCC CCTTCAGCGA TCAATCCTTC AAGCAGATTT GTGCTACCCC

Figure 8

EcMyb1 gene promoter sequence analysis by PlantCARE

a)



b)

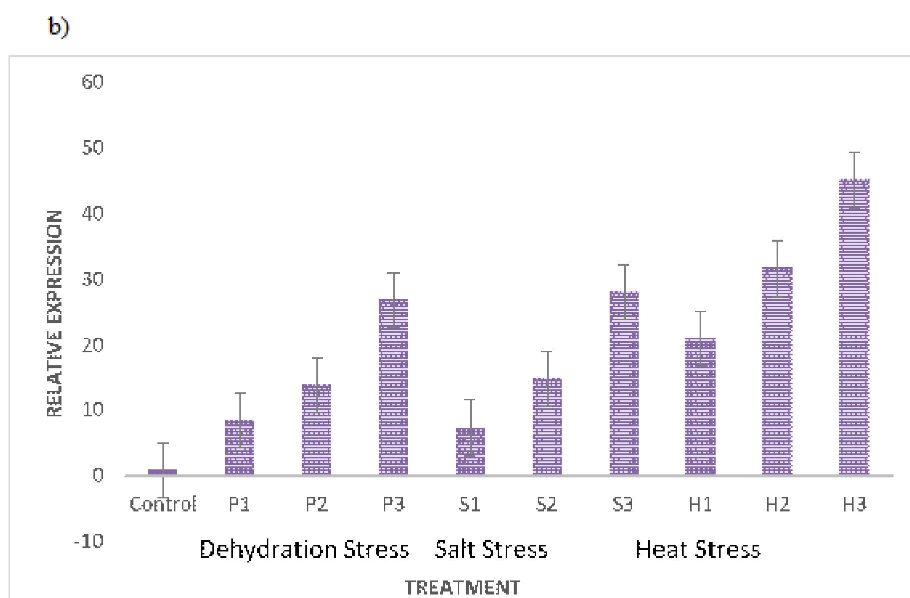


Figure 9

Quantitative real-time PCR analyses showing relative expressionof EcMyb1 gene under control and different abiotic stress conditions. P1-5% PEG, P2-10%PEG, P3-15%PEG, S1-100mM, S2- 150mM, S3-200mM NaCl solution and H1-35̊C, H2-40̊C and H3-45̊C temperature. The expression of EcMyb1 mRNA was normalized usingtwo endogenous controls, Tubulin (a) and actin (b) and calculated using the ΔΔCT method. Two replicates for each sample were analyzedby real-time PCR.