I J T A
© Serials Publications
National Academy of Agricultural Science (NAAS) Rating : 3.03

Sequencing and Analysis of ABA-inducible bHLH TF gene in Banana

*Resmi L. and *Ashalatha S. Nair

ABSTRACT: *Identification of candidate genes has been a major objective of the transgenic approach for inducing drought tolerance in crop plants. One of these genes conferring drought tolerance is ABA inducible bHLH transcription factor. The present investigation isolated genomic DNA sequences of ABA inducible bHLH transcription factor genes from two banana cultivars namely Monthan (drought tolerant) and Nendran (susceptible) cultivars. The isolated sequences were deposited in NCBI database with accession numbers KR262885 and KR262886. Bioinformatic analysis of the sequences showed the presence of putative conserved HLH domain with E-box/N-box specificity site and dimerization interfaces on the HLH domain.Signal scan search of the translated sequence resulted in the identification of major cis-acting elements acting as the core binding sites of several drought responsive genes.*

Key words*: Banana, bHLH,* Musa*, transcription factor*

INTRODUCTION

Bananas and plantains (*Musa* spp.) are large annual monocotyledonous herbaceous plants found in tropical and subtropical climates, and is one of the most popular fresh fruits enjoyed worldwide. Abiotic stresses are the most limiting factor of crop productivity and it is estimated to be more than 50% decline in the average yields of major crops worldwide. Abiotic tolerance is a genetically complex trait that involves multiple genes. Abiotic stresses solely associated with physiological and developmental changes in plants, which are due to changes in plant genes expression. Because banana has shallow roots and a permanent green canopy, it is especially sensitive to conditions that lead to water deficit [1]. A better understanding of the mechanisms employed by banana plants to tolerate abiotic stresses will be helpful for increasing crop production and quality of this economically valuable fruit. There are some transcription factor(s) that regulate the expression of several genes related to stress. ABA-inducible BHLH-type transcription factor encodes a nuclear localized BLH domain containing transcriptional activator involved in response to ABA. It acts as a transcription activator by regulating positively abscisic acid (ABA) response and confers drought tolerance and sensitivity to ABA. The present study indicates the sequencing and analysis of ABA-inducible bHLH-type transcription factor genes from two banana cultivars.

MATERIAL AND METHOD

Genotypes used

Banana cultivars namely Monthan (*Musa* ABBgroup, drought tolerant) and Nendran (*Musa* AAB group, susceptible) were selected for the present study. One month old tissue culture plantlets were purchased from Model Floriculture and Biotechnology Centre, Kazhakkuttom, Thiruvananthapuram, Kerala which were further maintained in green house conditions at Department of Botany, University of Kerala, Kariavattom.

Primer design

Nucleotide sequences of predicted ABA-inducible BHLH-type transcription factor gene from *Musa acuminata* sub species *malaccensis* were retrieved from Genbank (http://www.ncbi.nlm.nih.gov/Genbank) database of NCBI and used for primer designing. The predicted primers were subjected to check for various properties namely hairpin loops, primer dimer, Tm (temperature), GC%. The specificity of both forward and reverse primers as well as product size was

* Department of Botany, University of Kerala, Kariavattom, Thiruvananthapuram, Kerala

checked using primer Blast program of NCBI database against our retrieved sequence.

DNA extraction and PCR analysis

From fresh leaf tissue DNA was isolated using NucleoSpin® Plant II Kit (Macherey-Nagel) according to manufacturer's protocol and kept at -20°C until further use. The quality of the isolated DNA was checked using agarose gel electrophoresis. PCR amplification reactions were carried out in a 20 µl reaction volume which contained 1X Phire PCR buffer (contains 1.5 mM $MgCl_2$), 0.2mM each dNTPs (dATP, dGTP, dCTP and dTTP), 1 µl DNA, 0.2 µl Phire Hotstart II DNA polymerase enzyme, 0.1 mg/ml BSA and 3% DMSO, 0.5M Betaine, 5pM of forward 5'-GCTGTTCAGAGCCTGGTAT-3'and reverse primers 5'-TCGCTTGGCCAATTCCAGAT-3'. PCR amplification cycle consisted of an initial denaturation at 98 °C for 30 sec, followed by 40 cycles of denaturation at 98 °C for 5 sec, annealing at 54 °C for 10 sec, elongation at 72 °C for 15 sec and a final extension at 72 °C for 60sec. The PCR products were resolved in 1.2% agarose gels prepared in 0.5X TBE buffer containing 0.5 µg/ml ethidium bromide.

PCR product purification and sequencing

Five micro litres of PCR product is mixed with 2 µl of ExoSAP-IT and incubated at 37°C for 15 minutes followed by enzyme inactivation at 80°C for 15 minutes. Sequencing reaction was done in a PCR thermal cycler (GeneAmp PCR System 9700, Applied Biosystems) using the BigDye Terminator v3.1 Cycle sequencing Kit (Applied Biosystems , USA) following manufactures protocol.

Nucleotide sequence data and Computational protein analysis

After sequencing, sequenced amplicons of two cultivars were subjected Pairwise Sequence Alignment using Clustal W program. We made BLAST [2] searches to align our isolates with the already existing sequences in the genebank. These two sequences of ABA-inducible bHLH-type transcription factor gene were verified at protein level by using Blast-x (*http://blast.ncbi.nlm.nih.gov/blast.cgi*) programme from NCBI and Expasy (*http://web.expasy.org/cgi-bin/translate/dna_aa*) server. Conserved domains were identified using Conserved domain search service (NCBI) and Interpro scan [3]. Domain annotation of the sequence as well as homology modeling of the domain was done using Swiss model workspace [4]. The prediction and identification of protein sequence was performed through FGENESH Program[5]. Pfam available online [6] was used to find matching protein families, while isoelectric point stability, amino acid length was analyzed by Protparam in Expasy Website [7]. HNN program [8] in Expasy (*www.expasy.org*) was used to predict the secondary structure of protein sequences. To determine the cellular location of protein and terminal peptide signal, PSORT program [9] was used. Protein sequences were submitted to PLACE (Plant cis-acting regulatory DNA elements) to identify all possible promoter/cis-element [10].

RESULTS AND DISCUSSION

Sequence analysis

A 667 bp fragment was amplified from two banana cultivars (*Musa* AAB group cultivar Nendran and *Musa* ABB group cultivar Monthan) using the primer pair which on sequencing yielded 613 bp long nucleotides. Nucleotide BLAST search showed that our sequences are highly similar (99% identity) to predicted *Musa acuminata* subspecies *malaccensis* ABA inducible bHLH type like transcription factor gene. In addition, sequence analysis indicated that deduced protein belonged to HLH transcription factor superfamily (Fig. 1) and contained a putative conserved HLH domain with E-box/N-box specificity site and dimerization interfaces on the HLH domain (Fig. 2). Search using Pfam database also showed significant match of the protein sequence with HLH family with an E value 1.1e-10. The Pfam database is a large collection of protein families, each represented by *multiple sequence alignments* and *hidden Markov models (HMMs)*.

Protein-encoding nucleotide sequence (accession number KR262885 & KR262886 in Genebank) was identified by FGENESH, which locates at position from 24 bp to 590 bp in the nucleotide sequence, encoding a 189 amino acid sequence (aa) in length. This protein is similar to predicted ABA inducible bHLH type transcription factor gene from *Musa acuminata* subspecies *malaccensis* with accession number XP_009409474. In biochemical characteristics, protein sequence that has molecular weight and isoelectric point is 21785.6 and 6.46 respectively. The essential amino acid present in the protein sequence is valine (7.1%), followed by Isoleucine (5.1%). The amount of negative protein residues (Asp and Glu) is 31 while that of positve protein residues (Arg and Lys) is 30. iPSORT predicted the protein sequence as having a chloroplast tansit peptide. iPSORT is a

subcellular localization site predictor for N-terminal sorting signals. Given a protein sequence, it will predict whether it contains a Signal Peptide (SP), Mitochondrial Targeting Peptide (mTP), or Chloroplast Transit Peptide (cTP). The secondary structure of protein is composed of 41.33% helices. Extended strands form 3.06% of the residues while the remaining 55.61% comprise the random coil. Phylogenetic analysis (Fig. 3) shows the position of *Musa* ABB group cultivar Monthan in the dendrogram constructed using deduced protein sequence based on NJ method with some other bHLH transcription factor sequences of monocot by BLAST software. The two nucleotide sequences were used further for DNA polymorphism analysis in which 612 invariable (monomorphic) and 1 variable (polymorphic) sites were identified including a single mutation and two haplotypes. The haplotypes (gene) diversity was 1.00; the variance and standard deviation of haplotypes diversity were determined 0.025 and 0.5 respectively. The secondary structure of the domain was annotated using Swiss model work space and shown in Fig. 4. Homology modeling of the domain revealed two alpha helices and a loop as shown in Fig 5.

Analysis of cis-acting regulatory elements

The identification of cis-acting regulatory elements can provide important key for understanding spatial and temporal expression of protein sequences. It can also provide insights on the physiological and molecular processes that involved the action of drought-induced genes. . The protein sequences were used as queries for signal scan search for and identified the following elements. MYCCONSENSUSAT is the MYC recognition site found in the promoters of the dehydration-responsive gene rd22 and many other genes in *Arabidopsis*. DPBFCOREDCDC3 is core sequence for binding a novel class of bZIP transcription factors, DPBF-1 and 2 (Dc3 promoter-binding factor-1 and 2)in carrot. Dc3

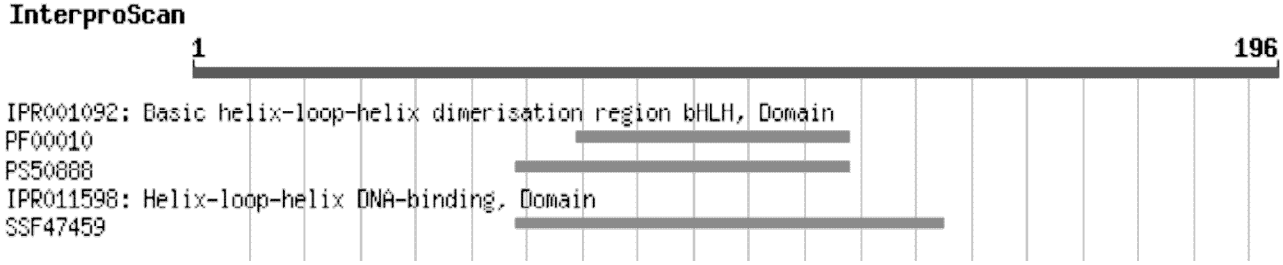

Figure 1: Result of InterproScan of deduced protein sequence

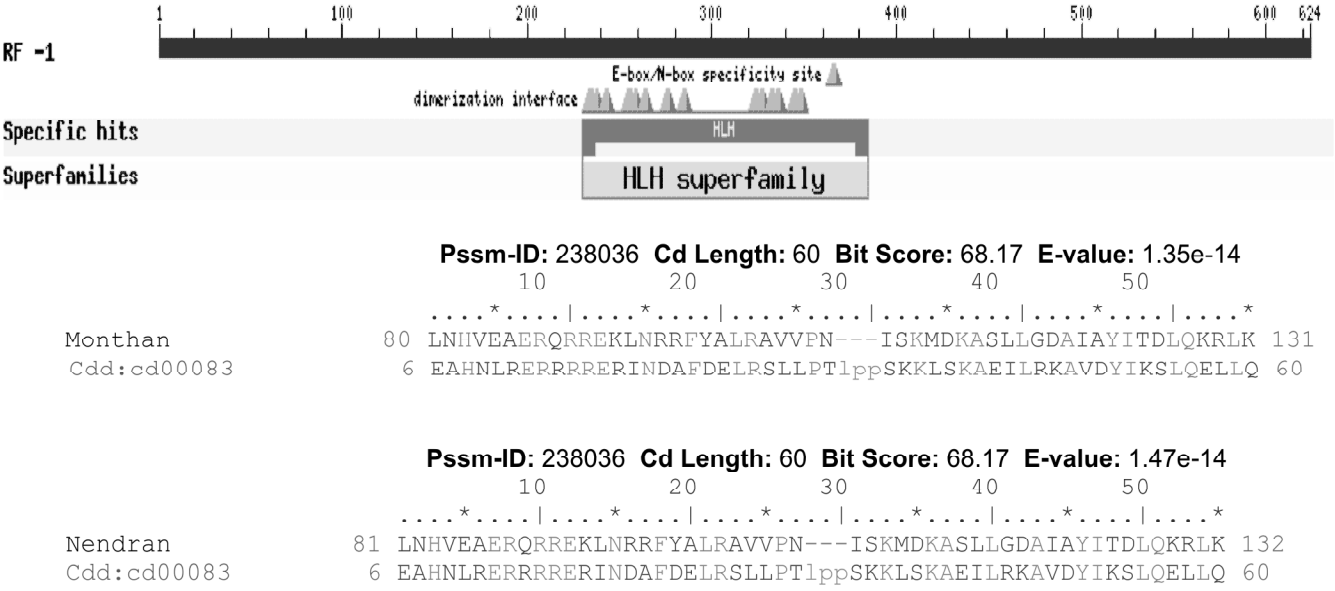

Figure 2: Identification of putative conserved domain in the deduced protein sequence

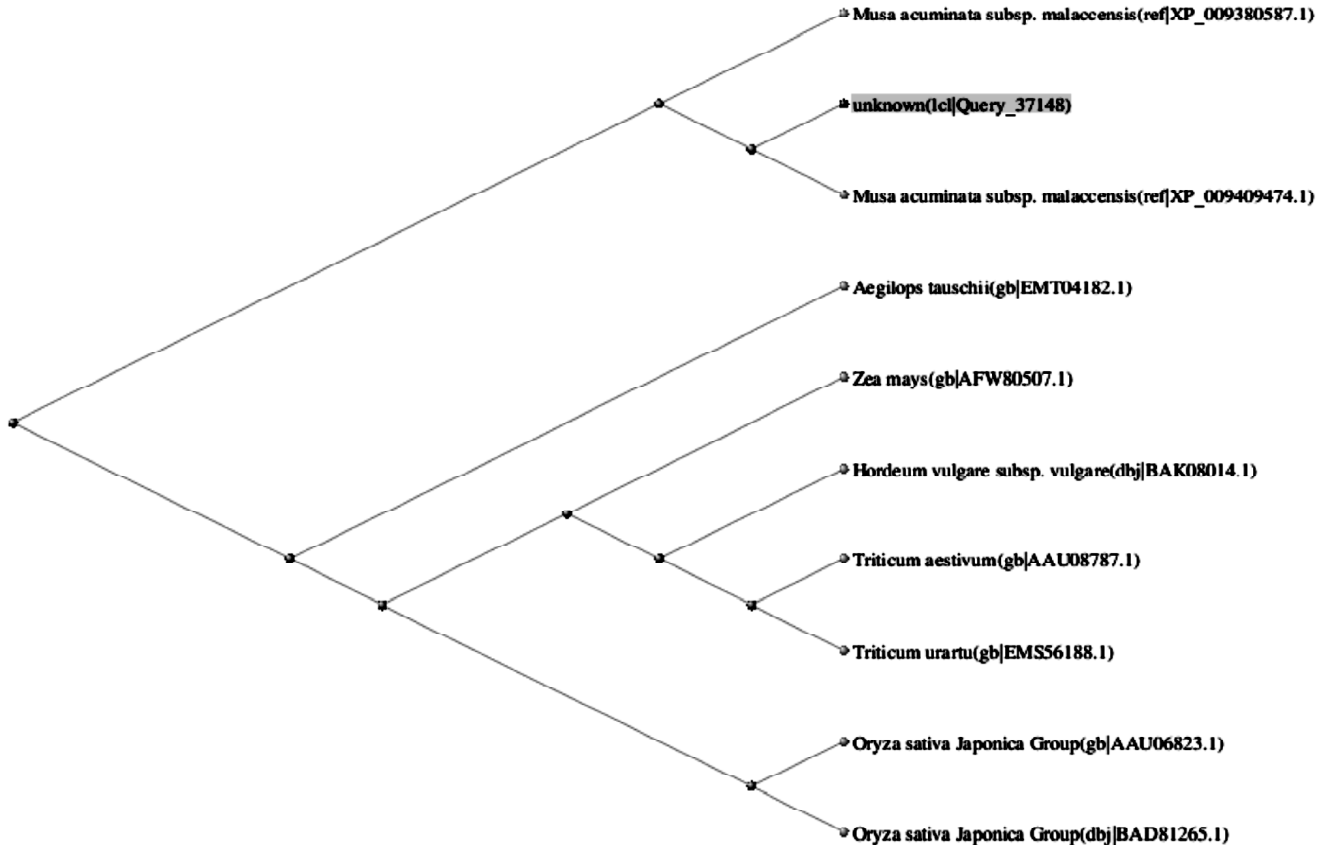

Figure 3: Position of *Musa* ABB group cultivar Monthan (unknown/ Query) in the NJ dendrogram derived based on protein sequences of bHLH transcription factor genes of monocots

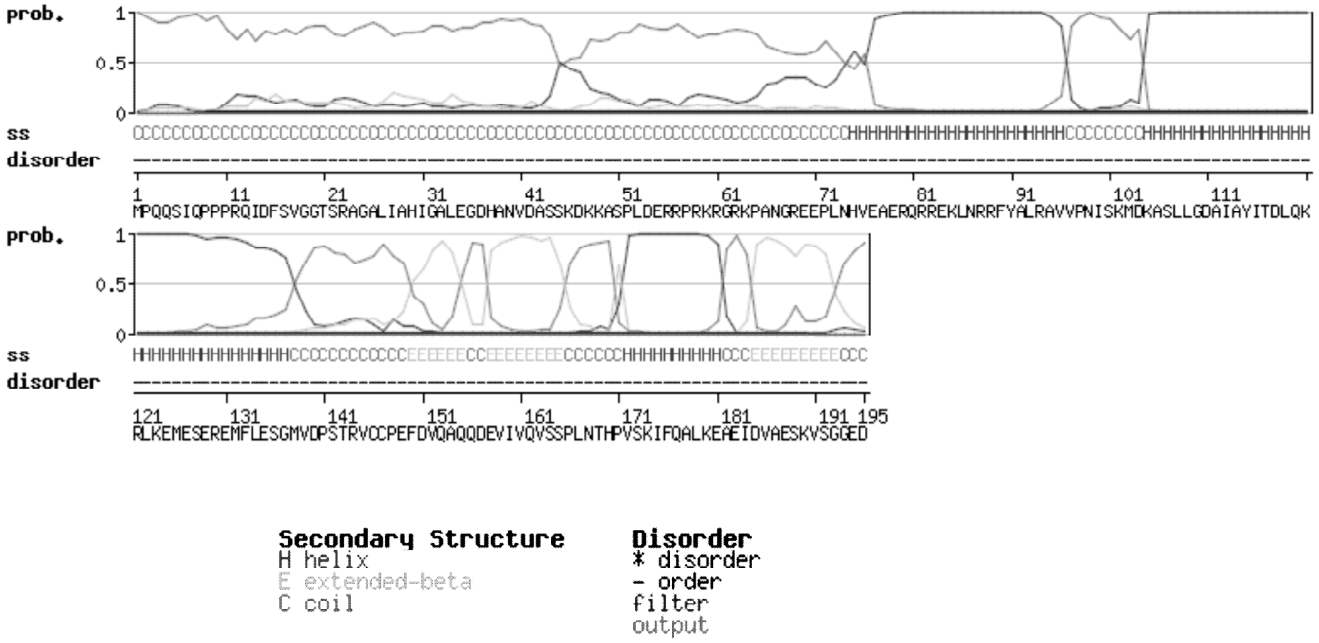

Figure 4: Domain annotation of deduced protein sequence using Swiss model workspace

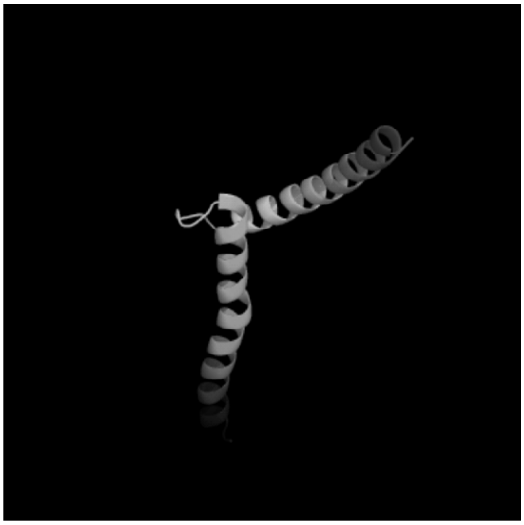

Figure 5: Homology model of the domain constructed using Swiss model workspace

expression is normally embryo-specific, and also can be induced by ABA. DRE1COREZMRAB17 is DRE1 (Drought responsive element 1) core found in maize) rab17 gene promoter. MYBCORE is the binding site for all animal MYB and at least two plant MYB proteins ATMYB1 and ATMYB2, both isolated from *Arabidopsis*. ATMYB2 is involved in regulation of genes that are responsive to water stress in *Arabidopsis*. MYCATRD22 is the binding site for MYC (rd22BP1) in Arabidopsis dehydration-resposive gene, rd22. DOFCOREZM is the Core site required for binding of Dof proteins in maize Dof proteins are DNA binding proteins, with presumably only one zinc finger, and are unique to plants.

CONCLUSION

Present study sequenced and analysed the ABA inducible bHLH transcription factor gene from banana. The partial sequence information for two banana cultivars have been deposited in the NCBI nucleotide database with accession numbers KR262885 and KR262886. The study revealed the potential of the gene to be considered as a candidate gene for drought tolerance and future works are designed to analyse the expression of the same in banana under drought conditions.

ACKNOWLEDGEMENT

First author acknowledges University Grants Commission, New Delhi for granting post doctoral fellowship under UGC Research Award scheme 2012-14.

REFERENCES

Sreedharan S., Shekhawat U. K., Ganapathi T. R. (2013), Transgenic banana plants overexpressing a native plasma membrane aquaporin MusaPIP1;2 display high tolerance levels to different abiotic stresses. *Plant Biotechnol J.* 11: 942–952.

Altschul, SF., Madden, TL., Schafer, AA., Zhang, J., Zhang, Z., Miller, W. and Lipman, DJ. (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25: 3389-3402.

Zdobnov E. M. and Apweiler R. (2001), InterProScan - an integration platform for the signature-recognition methods in InterPro Bioinformatics 17: 847-848.

Arnold K., Bordoli L., Kopp J., and Schwede T. (2006), The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22,195-201.

Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol*. 2006,7, Suppl 1: P. 10.1-10.12.

R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L.L. Sonnhammer, J. Tate, M. Punta. ThePfam protein families database. Nucleic Acids Research (2014) Database Issue 42:D222-D230.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; *Protein Identification and Analysis Tools on the ExPASy Server;* (In)Walker JM (ed). The Proteomics protocols Handbook, Humana Press (2005). Pp. 571-607.

Combet, C., Blanchet, C., Geourjon, C. and Deleage, G. (2000), NPS@: network protein sequence analysis. Trends Biochem. Sci. 25, 147-150.

Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., and Miyano, S., *Extensive feature detection of N-terminal protein sorting signals", Bioinformatics,* 18(2) 298—305, 2002.

Prestridge, D.S. (1991), SIGNAL SCAN: A computer program that scans DNA sequences for eukaryotic transcriptional elements. CABIOS 7, 203-206.