Final Progress Report

(April 2010 to July 2013)

Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) under Submergence

Submitted to

UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002, India

By

DR. DEV MANI PANDEY

(ASSOCIATE PROFESSOR)

PRINCIPAL INVESTIGATOR

DEPARTMENT OF BIOTECHNOLOGY

BIRLA INSTITUTE OF TECHNOLOGY

MESRA: RANCHI, INDIA

Project Title: Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) under Submergence

Introduction:

In lieu of changing environment standing crops faces various stresses during their life cycle resulting reduction in yield drastically (Agarwal and Grower, 2005; Wassmann et al. 2009) Although, some crops withstand the environment stresses by developing new features, while others unable to develop adaptation mechanisms and die. Importantly, rice has very antagonistic character about tolerance and susceptibility to abiotic stresses as compared to other crops (Yamauchi et al. 2000; Lafitte et al. 2004; Agarwal and Grower, 2005; Magneschi and Pierdomenico, 2009). Flooding is a major issue for plant survival in many regions of the world. Plant cells experience a deficit in cellular oxygen as a consequence of flooding and water logging, aerobic soil microbe blooms, and high rates of cellular metabolism or low levels of oxygen diffusion to internal tissues (Geigenberger, 2003). In plants flooding/water logging induced low oxygen stress stimulates the composite metabolic pathways and genetic programmes including the differential expression of a large number of genes (Vartapetian and Jackson, 1997).

A complex network of transcription factors orchestrates the response of plants to changes in environmental conditions (Chen et al. 2002). Gene expression studies revealed the up-regulation of genes coding for transcription factors (Liu et al. 2005) as well as signal transduction components (Baxter-Burrell et al. 2002). Study on gene expression in response to oxygen deprivation has already been described in *Zea mays* (Fennoy et al. 1998), *Arabidopsis thaliana* (Klok et al. 2002; Branco-Price et al. 2005) and *Oryza sativa* (Mukhopadhyay et al. 2004; Xu et al. 2006; Pandey and Kim, 2006, Pandey et al. 2007, Lasanthi-Kudahettige et al. 2007). Study on wide range of DEGs in anoxic rice coleoptiles using microarray analysis has also been conducted (Lasanthi-Kudahettigeet al. 2007). Isolation of anoxia induced differentially expressed genes in rice by 60k microarray analysis has also been reported (Pandey and Kim, 2012). Anaerobic response elements (AREs) with their binding sites (Paul et al. 2004); Apetala2/Ethylene Response Factor (AP2/ERFs) (Nie et al. 2002). Study on mRNA expression profiling in anoxic rice coleoptile has also been reported (Sadiq et al. 2011).

Anaerobically expressed genes are often characterized by the presence of anaerobic response elements (AREs) in their promoter regions. A core promoter contains the essential

nucleotide sequences for the regulation of gene function having the TATA box and transcription start site (TSS). Genes having similar expression patterns contain common motifs in their promoter regions (Vilo et al. 2000). Common promoter motifs are the key signatures for a family of co-regulated genes (Wang et al. 2004) and single motif can also bind various transcription factors thereby bringing the genes under multiple regulatory controls (Jin and Martin, 1999). Genes with common motifs in their promoter region have shown similar expression pattern (Vilo et al. 2000). Four ethylene-responsive element binding proteins (EREBPs) containing a novel DNA-binding domain that specifically bind to the GCC box have been isolated from tobacco (Ohme-Takagi and Shinshi, 1995). An ethylene-responsive factor (ERF) like transcription factor Pti4 in tomato binds the GCC box *cis*-element in the promoters of many pathogenesis-related (PR) genes (Chakravarthy et al. 2003). Cheong et al. (2003) also reported that transcription factor OsEREBP1 binds to GCC box element (AGCCGCC) in several PR-gene promoters when phosphorylated by MAPK. Arabidopsis ERFs (AtERFs) are reported to respond the GCC box-mediated gene expression positively or negatively in response to the extracellular signals (Fujimoto et al. 2000). Pandey et al. (2007) reported two consensus promoter motifs GGAG [A/G][G/A]G and GACGTGGCG by considering the 50 up-regulated genes during low oxygen stress in rice seedlings. Similarly DREB can identify and bind the *cis*-element (A/GCCGAC) in the promoters and regulate the expression of genes during environmental stresses (Sakuma et al. 2002). Hao et al. (1998) reported that ERF is involved in response to pathogenesis by recognizing the *cis*-acting element AGCCGCC (GCC box). Earlier Saleh et al. (2003) described AP2/ERF plant transcription factor genes that regulate developmental processes and involved in plant responses to various biotic and abiotic stresses. Jung et al. (2010) reported the association of APETALA2 (AP2)/ERF family transcriptional regulators with the Sub1A-1-mediated response upon submergence. Detailed gene structures, phylogeny, chromosome locations, and conserved motifs of ERF gene families in Arabidopsis have been described (Nakano et al. 2006). Another comprehensive study on differentially expressed genes under abiotic stress for 163 AP2/EREBP genes in rice has been described (Sharoni et al. 1996). Using publically available microarray data (Lasanthi-Kudahettige et al. 2007), consensus promoter motif having conserved GCC box (GCCGCC) in the promoter of up-regulated differentially expressed genes (UR-DEGs), while mutated GCC box (TCCTCC) in the promoter of down–regulated DEGs (DR-DEGs) in anoxic rice coleoptiles have been reported (Kumar et al. 2007, 2009).

Although existence of consensus/*cis*-elements in the promoters of genes have been reported. Various tools and databases are available for the analysis and retrievals of motifs/cis-elements. However, experimental validation towards availability of motifs/*cis*-elements in the promoter is little bit tedious and need ample expertise and experience. In this regard Molecular Beacon Probe (MBP) has been used to detect the presence of specific nucleic acids sequences in homogenous solutions (Tyagi and Kramer, 1996). Previously, MBP based Real-Time PCR amplification assay was used to detect nucleic acid sequences (Pas et al. 2005; Ye et al. 2009; Elsayed et al. 2003, 2006; Lata et al. 2009). MBP increases the sensitivity and precision over the convention PCR without post-reaction analysis for the detection and as well as quantification of target gene (Lata et al. 2009).

Since, promoter motifs/*cis*-regulatory elements are involved in the regulation of various cellular mechanisms during abiotic and biotic stresses. Therefore, identification of genes and mechanism of differential expression is great of interest. Presence of GCC box in UR-DEG and TCC box in the DR-DEG need to be validated by designing the sequence specific MBP, primers and MBP based Real-Time PCR analysis. This will also explain whether the GCC box actually present and has any role for UR-DEGs while TCC box in DR-DEGs. Therefore, in this study we aim to use MBP based Real Time PCR assay for rapid and accurate detection of GCC box in 17 UR-DEGs while mutated GCC box (i.e. TCC Box) in 15 DR-DEGs.

Objectives:

Molecular Beacon Based Real-Time PCR analysis towards validation of consensus promoter motif GCC box in the UR-DEGs while mutated GCC (TCC box) in DR-DEGs was carried out with the following objectives:

1. Screening of the highly expressed DEGs containing GCC box and mutated GCC box from our previous study (Kumar et al. 2009)
2. Designing the gene specific primer and specific Molecular Beacon probe
3. Validation of consensus promoter motifs by using above probes

Brief overview of the progress made:

Highly anoxia responsive UR-DEGs and DR-DEGs containing consensus promoter motif (GCC box i.e. GCCGCC or mutated GCC box i.e. TCCTCC) were sort listed from our previous study (Kumar et al. 2007, 2009). Promoter sequences of the selected highly anoxia responsive DEGs having length -499 to +100 were retrieved from the Eukaryotic Promoter Data Base (http://www.epd.isb-sib.ch/seq_download.html). Verification of Consensus promoter motifs of selected UR-DEGs and DR-DEGs were performed again by using online available software MEME (v 4.5.0) (http://meme.nbcr.net/meme4_5_0/cgi-bin/meme.cgi) for GCC box (GCCGCC) and Mutated GCC box (TCCTCC) sequences respectively. A total of 17 UR-DEGs and 15 DR-DEGs were short-listed for their validation experiment. *In-silico* interactions of SUB1A protein, CPuORF2 - conserved peptide uORF-containing transcript gene protein (LOC_Os09g13570) and B3 DNA binding domain containing gene protein (LOC_Os03g06850) with GCC box was studied. While bZIP transcription factor gene protein (LOC_Os02g52780) with TCC box was studied. Selected UR-DEGs and DR- DEGs promoter sequences were used for the designing of gene specific primers and Molecular Beacon Probes. Custom made consensus probe sequence for UR-DEGs and DR-DEGs were identified by using ClustalX software consecutively. These consensus probe sequences were further used for designing their respective primers and Molecular Beacon Probe. The software Beacon Designer7 was used for the designing of Molecular Beacon probe and their primers. The Beacon Designer7 designs the Molecular Beacon Probes and their specific primers by its specific parameters. It gives primer and probe sequence which were not forming dimers. The length of the primer was kept between 18-24 nucleotides long; amplicon size varying from 100 to 300 nucleotides in length and Tm of the target probe sequence having 8-10°C higher temperature than the annealing temperature of PCR primer. For Molecular Beacon Probes the GC content for stem sequence is 65-70% and Guanine is avoided at the 5' end next to the fluorophore. We designed target Molecular Beacon Probes with a stem sequence of 7 nucleotides long at both the ends, at 5' and 3' of target probe sequence for selected UR-DEGs and DR-DEGs promoter sequences. Primers of the selected UR-DEGs and DR-DEGs promoter sequences were designed in accordance with their target Molecular Beacon Probes. Rice (*Oryza sativa* L.) seedlings of Azucena cultivar were grown and total genomic DNA was isolated. DNA was treated with RNAse. Consensus promoter motifs providing binding sites to transcription factor in promoter region were validated by Real-time PCR. For this analysis total gDNA is used as template while Molecular Beacons as a probe. For the expression of UR-DEGs, GCC box (GCCGCC) specific Molecular Beacons probe were used, while mutated GCC box (TCCTCC) specific Molecular Beacons probe were used for DR-DEGs.

Material and Methods:

Validation of consensus promoter motif GCC box in the UR-DEGs while mutated GCC (TCC box) in DR-DEGs was carried out and described with following headings:

Screening of highly differentially expressed genes (DEGs): Publically available microarray analysis based differentially expressed genes in anoxic rice coleoptiles (Lasanthi-Kudahettige et al. 2007) were used for short listing the UR-DEGs, DR-DEGs and UC-DEGs. Promoter region of these sets of DEGs were used to find the consensus promoter motifs (GCC box i.e. GCCGCC or mutated GCC box i.e. TCCTCC) either in UR-DEGs or in DR-DEGs (Kumar et al. 2007, 2009).

Designing of gene specific probe and primer by Molecular Beacon designer: Promoter sequences of the selected DEGs having length -499 to +100 were retrieved from the TIGR (ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotationdbs/pseudomolecules/version_6.1/all.dir/all.utr). These promoter sequences were used for specific Molecular Beacon probe and primer designing with the help of software Beacon Designer 7. Specific Molecular Beacon probe were designed for the GCC box i.e. GCCGCC and mutated GCC box i.e. TCCTCC (Tyagi and Kramer, 1996).

Growing of rice seedling and DNA isolation: Publically available microarray data that was used to find the consensus promoter motif was based on the japonica cultivar of rice. Therefore, rice seeds of Azucena (Japonica *sp.*) were surface sterilized with 0.1% of $HgCl_2$ and incubated in dark for 48 h at 35° C. Sterilization and growing of seed were done in pot at room temperature. Genomic DNA was isolated from rice seedlings using CTAB (2X) method and subjected to RNase treatment. The concentration of genomic DNA was observed using Biophotometer (Eppendorf, USA) and DNA quality was checked in 0.8% agarose gel. A diagrammatic representation for sterilization and growing of seedlings as well as isolation of gDNA and RNase treatment has also been shown described.

Diagrammatic sketch of genomic DNA isolation from rice leaves by CTAB method

(Ratio of isolation buffer and sample, 100 µl buffer/0.1mg sample)

Grind young leaf tissues (0.15mg - 0.2mg)

Grinding in an autoclaved mortar and pestle with liquid nitrogen

Divide grinded leaf powder in two 1.5 ml eppendorf tubes

Add 0.750µl of 2x CTAB buffer (Extraction buffer) that was pre-warmed to 65˚C containing 20 µl β-mercaptoethanol/10ml of extraction buffer (mercaptoethanol added after cooling followed by transferring of the buffer into tube). Incubation was performed at 65˚C for 1h

Cool briefly at room temperature followed by addition of 750 µl chloroform-isoamyl alcohol (24:1). Shake gently for 10 min at room temperature and spin at 10000rpm for 15 min

Decant aqueous phase (top phase) into new 1.5 ml tube with wide bore tips

Add 0.6 volume chilled isopropanol and incubate at -20˚C for 1h

Spin at 10000rpm for 10 min at 4˚C (12000rpm for 5 min)

Decant isopropanol and wash pellet with 500-1000 µl 70% ethanol (gentle washing/taping). Spin at 10000rpm/5min/4˚C.

Drain the ethanol and keep it for air dry

Dissolve the pellet in 50µl of TE buffer

Storage at 4˚C

Diagrammatic sketch of RNase treatment of isolated genomic DNA

In dissolved DNA (total volume 100 µl) add (1.0-2.0 µl) RNase (10mg/ml) followed by incubation at 37˚Cfor 30 min

Add 1/10 volume sodium acetate (10 µl) and 2 volumes of absolute ethanol (200 µl)

Incubate at -20˚C for 1h or overnight (Mix gently about 5-10 times)

Centrifuge at 10000 rpm for 15 min and at 4˚C, drain and rinse pellet with 70% ethanol

Spin at 10000rpm for 5min and remove ethanol

Air dry dissolve the pellet in 50 µl TE (10:1)

Store stock solution in -20˚C and the working solution 4˚C

Identification/Validation of consensus promoter motifs using molecular beacon probe: Consensus promoter motifs providing binding sites to transcription factor in promoter regions. Identification of conserved promoter motif in the selected UR-DEGs was carried out by using MBP based Real-Time PCR. For the expression of UR- DEGs, GCC box (GCCGCC) specific MB probe was used, while mutated GCC box (TCCTCC) specific MB probe for DR-DEGs. For this analysis total gDNA was used as a template while Molecular Beacons as a probe. PCR amplification was performed in the total reaction volume of 15 µl (1X Taq buffer, 1 unit Taq polymerase, 0.2mM dNTPs, 3mM $MgCl_2$, 0.45µM primer, 3ng gDNA and 0.3µM MBP) at optimized PCR condition (95°C for 4 min; 40 cycles of 15 s at 95°C, 35 s at 60°C, and 45 s at 72°C). PCR amplification was carried out in Real Time PCR System (Applied Biosystems 7500 Fast Real-Time PCR Systems, USA). For the detection of GCC box in UR-DEGs, TCC box containing probe was used as a reference.

Gene Ontology classification to infer the biological role: To know the detailed biological role of above selected 17 UR-DEGs and 15 DR-DEGs, Gene Ontology classification (biological process, cellular component and molecular function) was carried out at http://www.ricearray.org/analysis/go_enrichment.php. Additional putative function was identified by finding orthologous gene models at TIGR (http://rice.plantbiology.msu.edu/cgi-bin/ORF_infopage.cgi).

Classification of anoxia responsive TFs families: Earlier we have shortlisted the UR-DEGs and DR-DEGs (from the publically available microarray data of Lasanthi-Kudahettige et al. 2007) and used for finding the consensus promoter motifs (Kumar et al. 2009). Here we further extended our work and tried to find the all TFs which belongs either UR-DEGs or DR-DEGs. Subsequently these TFs were further used for finding the domain/binding sites they have. Depending upon the type of domain/binding site these TFs were classified in various TFs families. These TFs belonging to particular families were again checked for their existence in Plant Transcription Factor Database (http://plntfdb.bio.uni-potsdam.de/v3.0). TFs listed in Plant Transcription Factor Database were only selected.

Promoter and protein sequence retrieval of selected DEGs: In this experiment presence of GCC box in the promoter of Ubiquinol Cytochrome C Chaperone gene (LOC_Os07g30790) was validated by MB probe based Real-Time PCR. Earlier it was reported that Sub1A is a gene (LOC_Os09g11480) playing important role to submergence tolerance. And introgressed rice lines using *Sub1A* gene has been developed having increased submergence tolerance by several folds. Therefore, it was interesting to see the interaction of GCC box promoter motif with Sub1A protein.

For this 1K Promoter sequence of Ubiquinol Cytochrome C Chaperone gene (LOC_Os07g30790) and Protein sequence of Sub1A was retrieved from the TIGR release version6.1 (ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.1/).

On the other hand some of the above selected anoxia responsive TFs belonging to UR-DEGs and DR-DEGs were selected for checking their involvement in the regulation of few DEGs having GCC box and mutated GCC box (TCC box) in their promoter region and already validated/identified using Molecular Beacon based Real-Time PCR. Among UR-DEGs, methyltransferase domain containing protein gene (LOC_Os06g05910) having GCC-box motif while in DR-DEGs, rhoGAP domain containing protein gene (LOC_Os12g05900) having TCC-box motif in their promoter region were selected. 1K Promoter sequence of methyltransferase domain containing protein gene and rhoGAP domain containing protein gene were retrieved. Similarly, among UR-DEGs highly up-regulated TFs (CPuORF2 - conserved peptide uORF-containing transcript gene (LOC_Os09g13570) protein and B3 DNA binding domain containing gene (LOC_Os03g06850) protein and DR-DEGs (bZIP transcription factor gene (LOC_Os02g52780) protein were selected and protein sequences were retrieved from TIGR release version6.1 (ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.1/).

3D structure prediction of various proteins: The 3D structure of Sub1A, CPuORF2 - conserved peptide uORF-containing transcript gene (LOC_Os09g13570) protein and B3 DNA binding domain containing gene (LOC_Os03g06850) protein and bZIP transcription factor gene (LOC_Os02g52780) protein were not available in PDB database, therefore their structures were predicted using I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). I-TASSER is a hierarchical protein structure modeling approach based on the secondary-structure enhanced Profile-Profile threading Alignment (PPA). I-TASSER (The iterative threading assembly refinement) server determines 3D structures of protein based on the sequence-to-structure-to-function paradigm algorithm. It predicts secondary structure, tertiary structure and functional annotations on ligand-binding sites, enzyme commission numbers and gene ontology terms. The accuracy of prediction is based on the confidence score of the modeling (Zhang, 2008, Zhang et al., 2012). C-score is a confidence score for estimating the quality of predicted models by I-TASSER. It is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations.

Construction of 3D DNA structure of promoter region having GCC-box motif: Construction of 3D DNA structure of promoter region having GCC-box motif in Ubiquinol Cytochrome C Chaperone gene (LOC_Os07g30790) and methyltransferase domain containing protein gene (LOC_ Os06g05910) was performed. Similarly 3D DNA structure in DR-DEGs like rhoGAP domain containing protein gene (LOC_Os12g05900) having TCC-box was performed as described by Pandey and Kumar (2013). To study protein-DNA interaction a 3D model of DNA fragment (promoter region of 25 bases having core GCCGCC motif) was required. Therefore, 3D-DART (3DNA-Driven DNA Analysis and Rebuilding Tool) server was used for generating custom 3D structural model of DNA and its PDB file. The promoter fragments of varying length having core GCCGCC motif of in selected genes were used for the 3D model of DNA. 3D-DART uses the DNA rebuild functionality of software package 3DNA (Lu and Olson, 2003) and extends its functionality with tools to change the global conformation of the DNA models from a sequence to a base-pair step parameter file (van Dijk and Bonvin, 2009).

In-silico protein-DNA interaction studies: Interactions between 3D DNA structure of promoter region and 3D structure of selected proteins were performed as described by Pandey and Kumar, (2013). For *in-silico* protein-DNA interactions studies between 3D structure of Sub1A, CPuORF2 - conserved peptide uORF-containing transcript gene protein, B3 DNA binding domain containing gene protein and bZIP transcription factor gene protein (I-TASSER generated 3D models) and 3D structure of DNA fragment having core GCC box motif (3D-DART generated models), HADDOCK web server (http://haddock.science.uu.nl/services/HdeADDOCK/haddockserver-easy.html) was used. HADDOCK (High Ambiguity Driven protein-protein Docking) is an information-driven flexible docking approach for the modeling of bimolecular complexes. HADDOCK distinguishes itself from *ab-initio* docking methods with the fact that it encodes information from identified or predicted protein interfaces in ambiguous interaction restraints (AIRs) to drive the docking process. These AIR files have information about active residues (directly involved in the interaction) in protein as well as in the DNA model. Result with the lowest HADDOCK score and Z-Score were considered as the best interaction between these molecules (de Vries et al. 2010).

Results:

Screening and retrieval of the highly differentially expressed genes (DEGs): Highly anoxia responsive UR-DEGs and DR-DEGs containing consensus promoter motif (GCC box i.e. GCCGCC or mutated GCC box i.e. TCCTCC) that were published previously (Kumar et al. 2007, 2009) were sort listed. Promoter sequences of the selected highly anoxia responsive DEGs having length -499 to +100 were retrieved from the Eukaryotic Promoter Data Base (http://www.epd.isb-sib.ch/seq_download.html). Verification of consensus promoter motifs of selected UR-DEGs and DR-DEGs were performed again by using online available software MEME (v 4.5.0) (http://meme.nbcr.net/meme4_5_0/cgi-bin/meme.cgi) for GCC box (GCCGCC) and Mutated GCC box (TCCTCC) sequences, respectively. A total of 17 UR-DEGs and 15 DR-DEGs were short-listed for their validation experiment (Table-1). Promoter sequences with the length of -499 to +100 from the above selected UR-DEGs and DR-DEGs were retrieved from the Eukaryotic Promoter Database (http://www.epd.isb-sib.ch/seq_download.html) and used for the designing of gene specific primers and MBPs with Beacon Designer7. As per BD7 protocol sequence GCCGCCGCCG indicated more Molecular Beacon compatibility score for probe and primers for anoxia responsive UR-DEGs (Figure - 1). In contrast sequence CTCCTCCTCCTCCTC indicated more Molecular Beacon compatibility score for probe and primers for anoxia responsive for DR-DEGs (Figure -2). In our MEME result information content (IC) of GCC box was 16.6 bits, E-value 1.4e-056, and width length of 11 nucleotides (Figure- 1). On the other hand IC of TCC box in DR-DEG was 17.9 bits, E-value 7.6e-089 and width length of 15 nucleotides (Figure- 2).

Table-1: Selected highly anoxia responsive UR-DEGs and DR-DEGs.

UR-DEGs		DR-DEGs	
TIGR locus Id	**NCBI Accession No.**	**TIGR locus Id**	**NCBI Accession No**
LOC_Os07g30790	AK068288	LOC_Os02g01240	AK103103
LOC_Os08g42920	AK058490	LOC_Os02g52130	AK103417
LOC_Os07g37280	AK063204	LOC_Os03g48970	AK069854
LOC_Os06g05910	AK064640	LOC_Os02g48110	AK100997
LOC_Os11g08940	AK108801	LOC_Os09g03620	AK065517
LOC_Os12g39520	AK120895	LOC_Os11g41150	AK099444
LOC_Os03g04140	AK067089	LOC_Os07g37100	AK102045
LOC_Os09g09650	AK073072	LOC_Os06g48590	AK071376
LOC_Os04g02310	AK121178	LOC_Os03g07480	AK100027
LOC_Os06g11720	AK063324	LOC_Os12g05900	AK067300
LOC_Os06g09560	AK111076	LOC_Os11g05190	AK073352
LOC_Os06g40040	AK121619	LOC_Os03g26210	AK068704
LOC_Os10g42150	AK121698	LOC_Os02g15550	AK065153
LOC_Os12g43100	AK059924	LOC_Os07g09190	AK100909
LOC_Os12g23780	AK071086	LOC_Os03g19390	AK073626
LOC_Os01g34060	Ak068565		
LOC_Os03g10460	AK102580		

Figure -1: Information content diagram (MEME v4.5.0) shows GCC box selected for specific Molecular Beacon and its Primer designed by Beacon Designer7.

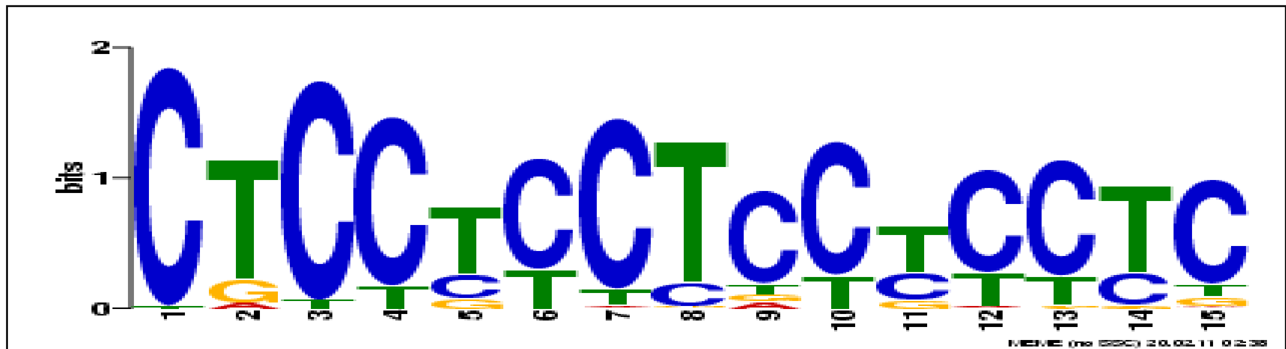


Figure - 2: Information content diagram (MEME v4.5.0) shows mutated GCC box (TCC box) selected for specific Molecular beacon and its Primer designed by Molecular Beacon designer7**.**

Depending upon Molecular Beacon compatibility score for probe and primers, selected anoxia responsive UR-DEGs having GCCGCCGCCG promoter motif sequence while DR-DEGs with CTCCTCCTCCTCCTC promoter motif sequence were selected for validation. Here it also is important to see the repetition of GCCGCCGCCG and CTCCTCCTCCTCCTC sequences in the promoters of above selected UR-DEGs and DR-DEGs that used for MEME (v4.5.0) analysis. Repeated sets of GCC-box having the CGCCGCCGCCG sequences were present in the promoter of selected UR-DEGs ranging from 200 to 600 bp of the promoter region. A snap shot of this has been shown in Figure -3. Similarly, repeated sets of mutated GCC-box (TCC box) with CTCCTCCTCCTCCTC sequences were present in the promoter of selected DR-DEGs ranging from 200 and 600 bp of the promoter region (Figure - 4).

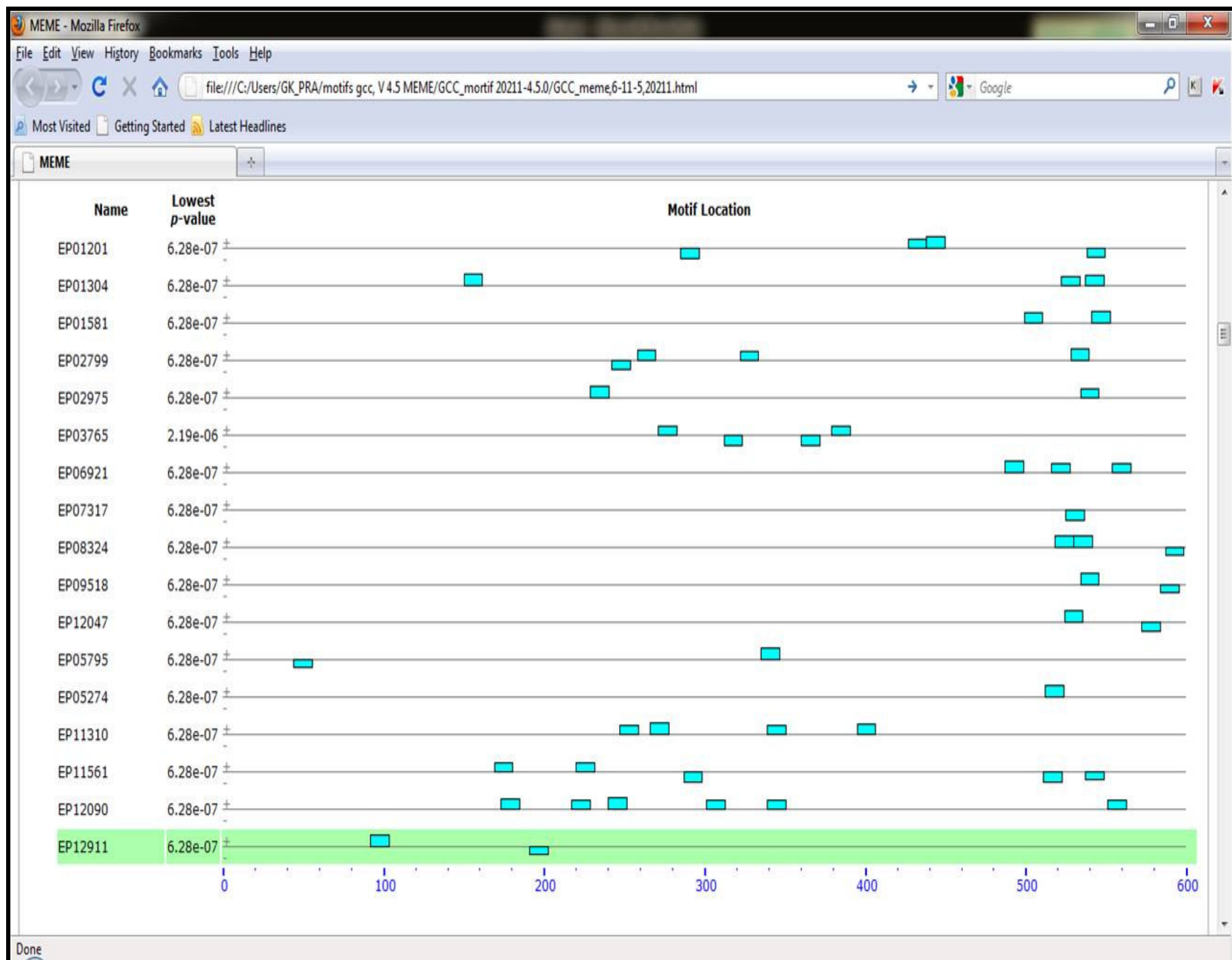


Figure -3: A snap shot of repeated sets of GCC-box having the CGCCGCCGCCG sequences that were present in the promoter of selected UR-DEGs ranging from 200 to 600 bp of the promoter region and analyzed by MEME (v4.5.0).

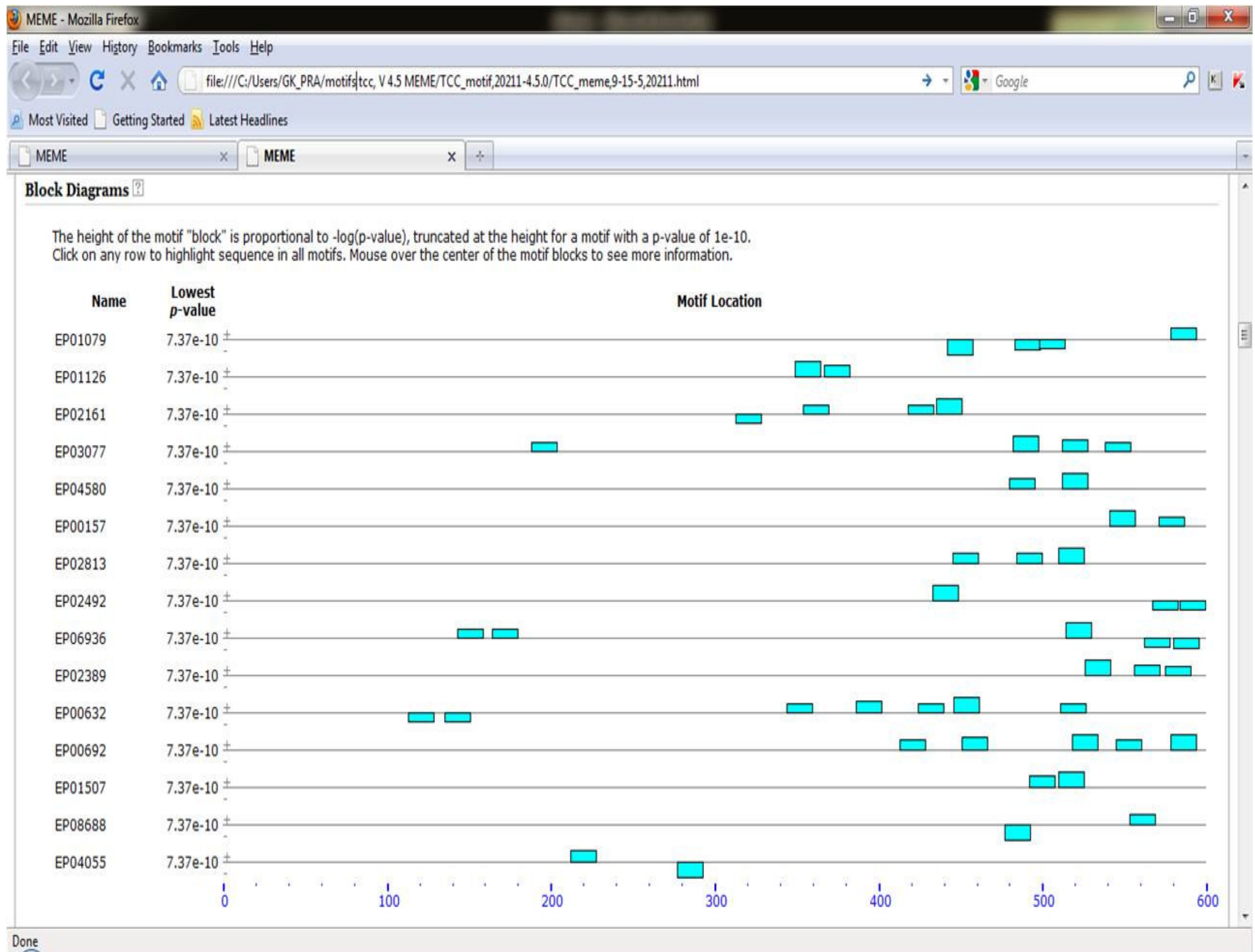


Figure-4: A snap shot of repeated sets of mutated GCC-box (TCC box) having CTCCTCCTCCTCCTC sequences that were present in the promoter of selected DR-DEGs ranging from 200 to 600 bp of the promoter region and analyzed by MEME (v4.5.0).

Designing of gene specific primers and Molecular Beacon probes: Selected UR-DEGs and DR-DEGs promoter sequences were used for designing the gene specific primers and MB Probes. Custom made consensus probe sequence for UR-DEGs and DR-DEGs were identified by using ClustalX software consecutively (Table-2). These consensus probe sequences were further used for designing their respective primers and Molecular Beacon Probe. The software Beacon Designer7 was used for the designing of Molecular Beacon probe and their primers (Tyagi and Kramer, 1996). The Beacon Designer7 designs the Molecular Beacon Probes and their specific primers by its specific parameters. It gives primer and probe sequence which were not forming dimers. The length of the primer was kept between 18-24 nucleotides long; amplicon size varying from 100 to 300 nucleotides in length and Tm of the target probe sequence having 8-10°C higher temperature than the annealing temperature of PCR primer. For Molecular Beacon Probes the GC content for stem sequence is 65-70% and Guanine is avoided at the 5' end next to the fluorophore. We designed target Molecular Beacon Probes with a stem sequence of 7 nucleotides long at both the ends, at 5' and 3' of target probe sequence for selected UR-DEGs and DR-DEGs promoter sequences (Table-2). Primers of the selected UR-DEGs (Table-3) and DR-DEGs (Table-4) promoter sequences were designed in accordance with their target Molecular Beacon Probes.

Table-2: Molecular Beacon probe sequence of UR-DEGs and DR-DEGs with stem sequences at 5' and 3' end (highlighted/underlined) designed by using software Beacon Designer7.

Differentially Expressed Genes	Length of Molecular Beacon probe (in bp)	Molecular Beacon probe with stem sequence (underlined)
Up -regulated	24	5'-**CGCGATC**GCCGCCGCCG**GATCGCG**-3'
Down-regulated	29	5'-**CGCGATC**CTCCTCCTCCTCCTC**GATCGCG**- 3'

Table-3: Primer sequences of UR-DEGs with their respective length designed by using software Beacon designer7.

Up-regulated DEGs					
S. No.	**Accession No.**	**Left Primer**	**Length of Left primer (in bp)**	**Right Primer**	**Length of right primer (in bp)**
1	AK068288	AAGAAACGGATGAACAAACAAAC	23	GAGGAGATGGAGCTGGTC	18
2	AK058490	GCCATAATAAGACGGTGAGA	20	CCGCTATCTCTACGCAAG	18
3	AK063204	TGAGGTTTGTATTGGTGAA	19	GCTGAGGTACATGACCAT	18
4	AK064640	CCTCCTAGTTCGTCCGTCAA	20	TCGAGCCTGGACTTCACC	18
5	AK108801	TGGGAGGATGGTAAACGGTAA	21	GAAGCAGCGCACTGGTAT	18
6	AK120895	GCCCATCTAAATAGTCCATCTAAA	24	ATCTTCCTCCTCGTCGTC	18
7	AK067089	GTTCCAGCACCAGACCAC	18	AGACGAGACGACGAGAGG	18
8	AK073072	CCTCTTTTCACTTCTCTGT	19	GTACTCCGATTGGATGTC	18
9	AK121178	CCTTCCGTCGCTGCTAAG	18	GACTTCTGTTCGGGATTGG	19
10	AK063324	GTGCCTCCTTTATCAATCAAT	21	AACTAGCCGACGTTGTAG	18
11	AK111076	CGTGAGTGAGTCTTCCGTGTCTTC	24	GCCACCGAGCACCTGTCC	18
12	AK121619	GCTTCCGAGTTCCGACCGA	19	GCGGCGCAAGAGGAATCG	18
13	AK121698	GTCATGGCGAGGCAGGTC	18	CCATCATTCCAGGTGAAGTCAGA	23
14	AK059924	GCCCACATAGCAACGCATA	19	GGTGAGCCAATCGAGTCC	18
15	AK071086	AATATAAACCGTCCACCCACTCAC	24	GAAGCCGCACGCCGATAC	18
16	Ak068565	CGACTAAATCCAGCCGCAAA	20	CACTCTCGCCTCCTCCTC	18
17	AK102580	GCACGAATCTCAAACCAATTCAATG	25	GTCCCGGAGGCACCTCAC	18

Table-4: Primer sequences of DR-DEGs with their respective length designed by using software Beacon designer7

Down-regulated DEGs					
S. No.	**Accession No**	**Left Primer**	**Length of Left primer (in bp)**	**Right Primer**	**Length of right primer (in bp)**
1	AK103103	GAGTGATCCGTTATATCTGTT	21	CTCTCCTTCCTTCCTTCT	18
2	AK103417	TCCTCCATCAGGTGTTAC	18	ACAGAACAGAGCAGGAATA	19
3	AK069854	CTGCTCACTGGTCAGTAC	18	CAACGCCAATCAGGATAA	18
4	AK100997	TTCAGCAGCAACGCACAA	18	GGAGAGAGCAGCGAAGGA	18
5	AK065517	ATGGCTTGATATTTCTCTCT	20	AAGGGTAGGATAGGGTAG	18
6	AK099444	GTGGGCTATACTCAATTTAG	20	TTACTCCCTCTACGCTTC	18
7	AK102045	AAAGAAACCCGAGAGATTC	19	CAAGGGGAGGAGAGGAAG	18
8	AK071376	TGTTCTCCTCTTCCTCTT	18	GCCCACCTATTTTGAAGA	18
9	AK100027	ACCACCCTTTATTATATTCC	20	GGAGATGGAAATGGAAAG	18
10	AK067300	CATCATTAGCGGAGGATT	18	CGGAGGTGGCTAAATAAC	18
11	AK073352	AATAGCCTCCACTACTTCTACTACT	25	GTTGTCGAAGGGCGAGAG	18
12	AK068704	GCGGCTCTCCTTCTTCTT	18	AAACAAGGGCAAACCTCAAA	20
13	AK065153	ATCCCTTGCTATATAATAATAATCC	25	GCGAGGGCATAAATAGAG	18
14	AK100909	AACTTGCTTGTCTGTTTGTTT	21	CAATGGCGATCGATGCTG	18
15	AK073626	TCAGACGCTGCGGAGATC	18	TTGACGCTGCGACCATTC	18

Growth of rice seedling: For the growth of rice seedlings, healthy seeds of Azucena (a japonica cultivar) were selected and surface sterilized with 0.1% of $HgCl_2$ for1 h. After 2-4 times washing with pure and sterilized water seeds were kept in incubator for 48 h at 37^0C in dark. Sterilized seeds (Figure-5A) were grown in pot at room temperature. Genomic DNA was isolated from rice seedlings (Figure-5B) using CTAB (2X) method and subjected to RNase treatment by standard protocol. The concentration of genomic DNA was observed using Biophotometer (Eppendorf, USA) and DNA quality was checked in 0.8% agarose gel (Figure-5C).

Figure - 5:(A) Surface sterlized Azucena (Japonica sp.) seeds; (B) Rice seedlings grown in tray; (C) RNase treated genomic DNA of Azucena leaves

Real Time PCR amplification: The identification of conserved promoter motif in the selected UR-DEGs and DR-DEGs was carried out by using MBP based Real Time PCR. Custom made MBPs were procured from Gene Link TM, (NY, USA). The PCR amplification was performed in the total reaction volume of 15 µl (1X Taq buffer, 1 unit Taq polymerase, 0.2mM dNTPs, 3mM $MgCl_2$, 0.45µM primer, 3ng gDNA and 0.3µM MBP). To get the optimum annealing temperature (Tm) for Real Time PCR analysis, a gradient PCR was performed. A optimized PCR condition (95°C for 4 min; 40 cycles of 15 s at 95°C, 35 s at 60°C, and 45 s at 72°C) with varying gradient temperature was followed. The gradient temperature was kept between 47.9˚C to 65˚C for selected UR-DEGs (Table-5) and for DR-DEGs 50˚C to 70˚C (Table-6). PCR amplification was carried out in Real Time PCR System (Applied Biosystems 7500 Fast Real-Time PCR Systems, USA). For the detection of GCC box in UR-DEGs, TCC box containing probe was used as a reference. On the other hand GCC box containing MBP was used as reference for the detection of TCC box in DR-DEGs. Importantly, optimization of annealing temperature was done for all UR-DEGs and DR-DEGs. Amplified product was run on 1.2% agarose gel followed by staining in EtBr solution. Stained gel was visualized by using Gel Documentation System with standard protocols. A representative gel doc picture having amplified product of predicted size has been shown for UR-DEGs (Figure- 6) and DR-DEGs (Figure-7), respectively.

Table-5: List of UR-DEGs with their respective primers and amplicon size.

S. No.	Gene Id	Primer sets (5'-3')	Amplicon size (bp)
1	AK068288	AAGAAACGGATGAACAAACAAAC (Fo)	139
		GAGGAGATGGAGCTGGTC (Ro)	
2.	AK064640	CCTCCTAGTTCGTCCGTCAA (Fo)	107
		TCGAGCCTGGACTTCACC (Ro)	
3.	AK108801	TGGGAGGATGGTAAACGGTAA (Fo)	147
		GAAGCAGCGCACTGGTAT (Ro)	
4.	AK111076	CGTGAGTGAGTCTTCCGTGTCTTC (Fo)	137
		GCCACCGAGCACCTGTCC (Ro)	

Table-6: List of DR-DEGs with their respective primers

S.No.	Gene Id	Primer sets(5'-3')	Amplicon size (bp)
1.	AK103103	GAGTGATCCGTTATATCTGTT (Fo)	200
		CTCTCCTTCCTTCCTTCT (Ro)	
2.	AK103417	TCCTCCATCAGGTGTTAC (Fo)	190
		ACAGAACAGAGCAGGAATA (Ro)	
3.	AK100997	TTCAGCAGCAACGCACAA (Fo)	173
		GGAGAGAGCAGCGAAGGA (Ro)	
4.	AK071376	TGTTCTCCTCTTCCTCTT (Fo)	121
		GCCCACCTATTTTGAAGA (Ro)	
5.	AK067300	CATCATTAGCGGAGGATT (Fo)	162
		CGGAGGTGGCTAAATAAC (Ro)	
6.	AK068704	GCGGCTCTCCTTCTTCTT (Fo)	123
		AAACAAGGGCAAACCTCAAA (Ro)	
7.	AK065153	ATCCCTTGCTATATAATAATAATCC (Fo)	152
		GCGAGGGCATAAATAGAG (Ro)	

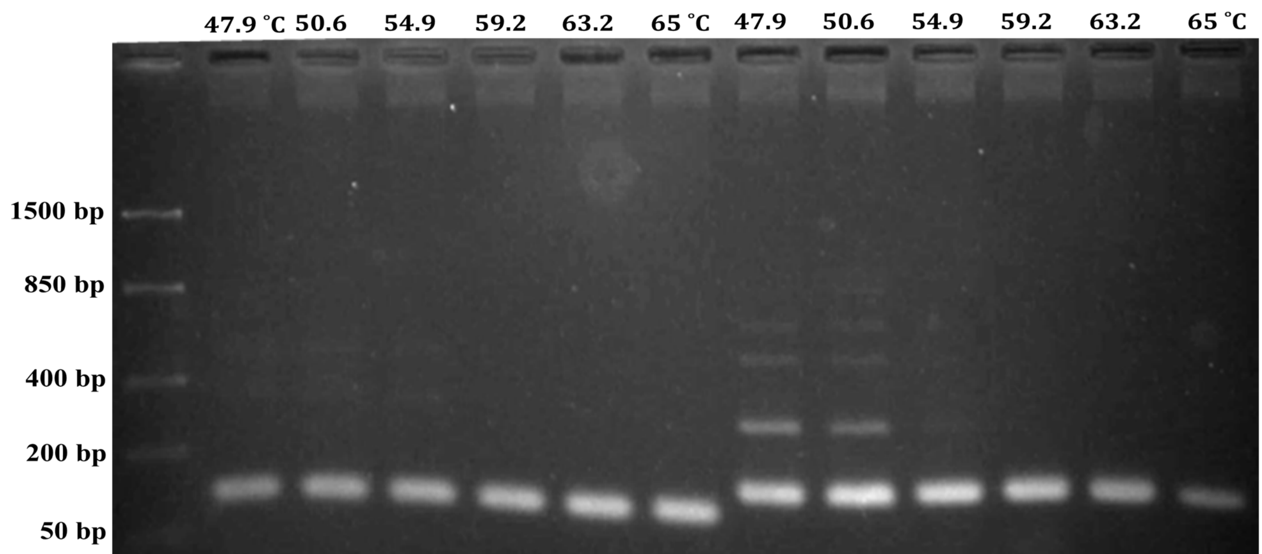


Figure - 6: Gradient PCR for annealing temperature optimization of two UR-DEGs (AK064640 and AK108801) with their respective primers.

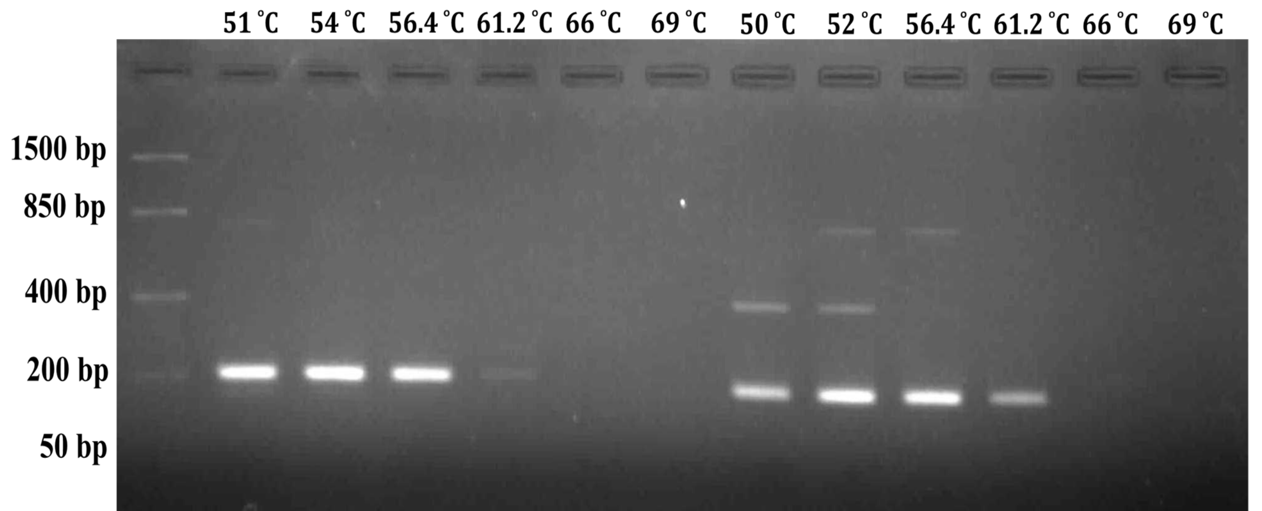


Figure - 7: Gradient PCR for annealing temperature optimization of two DR-DEGs (AK103103 and AK103417) with their respective Primers

Identification of GCC box in UR-DEGs: Identification of GCC box in UR-DEGs like ubiquinol-cytochrome C chaperone family protein gene (AK068288 or LOC_Os07g30790), methyltransferase domain containing protein gene (LOC_Os06g05910 or AK 064640), 60S ribosomal protein L7 gene (LOC_Os08g42920 or AK058490) and heat shock protein DnaJ gene (LOC_Os06g09560 or AK111076) were done by Real Time PCR assay. After completion of PCR graphs of Delta Rn vs cycle number for above four UR-DEGs were retrieved from the inbuilt Real Time PCR system software (Figures – 8A, 9A, 10A & 11A). In Rn vs cycle number graphs, a′ & b′ curves indicate the amplification of respective gene with GCC probe; c′ & d′ curves for genes having TCC probe and e′ & f′ for non template control, NTC (Figures – 8B, 9B, 10B & 11B). Similarly, dissociation curves of above UR-DEGs indicating the amplification of GCC and TCC box having only two specific products were also prepared. A dissociation curve (a″ & b″ depicting the two independent replications for the amplification with GCC probe while c″ & d″ for TCC probe of a representative UR-DEG only has been shown in Figures – 8C, 9C, 10C & 11C. Ct values chart for ubiquinol-cytochrome C chaperone family protein gene, methyltransferase domain containing protein gene, 60S ribosomal protein L7 gene and heat shock protein DnaJ gene were also prepared and enlisted in Table - 7. Ct values and Std dev Ct were obtained from inbuilt Real Time Software. To get the more authentic Real Time PCR amplification pattern each gene was analyzed in two replicates. Amplification was done with non template control (NTC) and with Template having Molecular Beacon probes specific to GCC box. To see the specific amplification of GCC box probe we also performed PCR amplification with non-specific TCC box probe. Result showed the good amplification pattern with Template having Molecular Beacon probes specific to GCC box only. This indicates that GCC box is present in the promoter of above selected UR-DEGs.

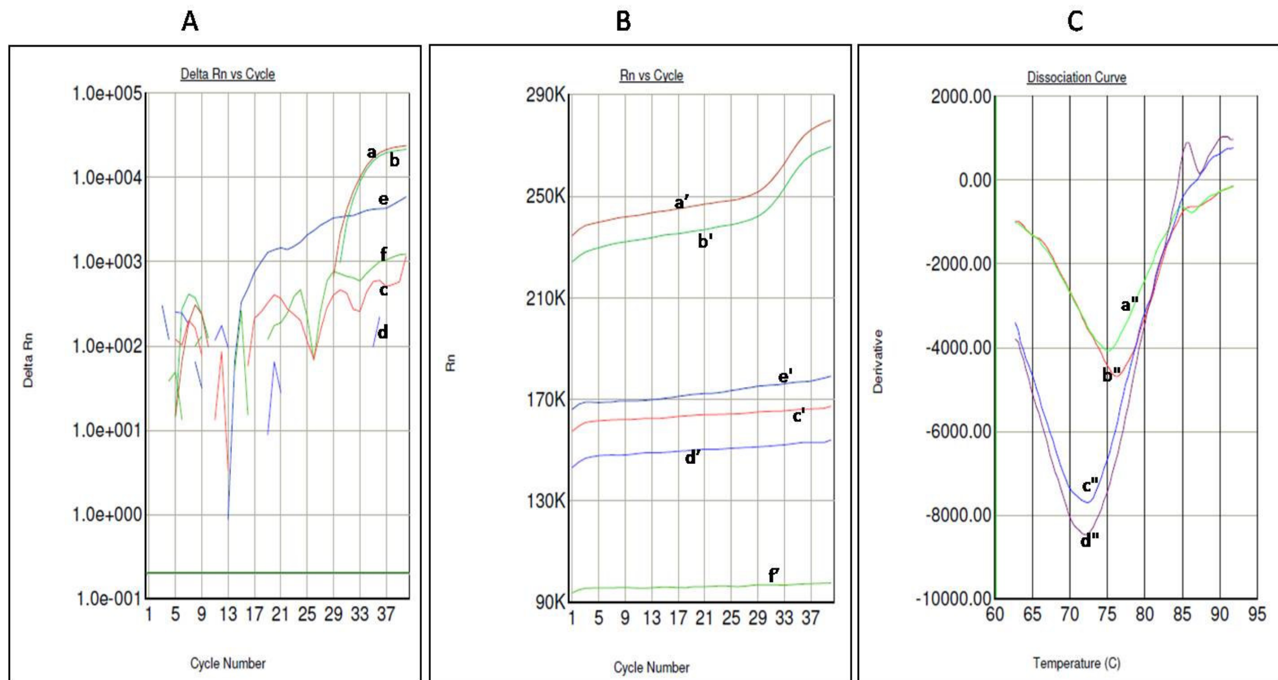


Figure - 8: Molecular Beacon based Real Time PCR result for Ubiquinol-cytochrome C chaperone family protein gene having GCC box in its promoter region (-499 to +100). (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of non template control, NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a"&b"* depict the dissociation curve of gene with GCC probe and *c"&d"* are dissociation curves of gene with TCC probe.

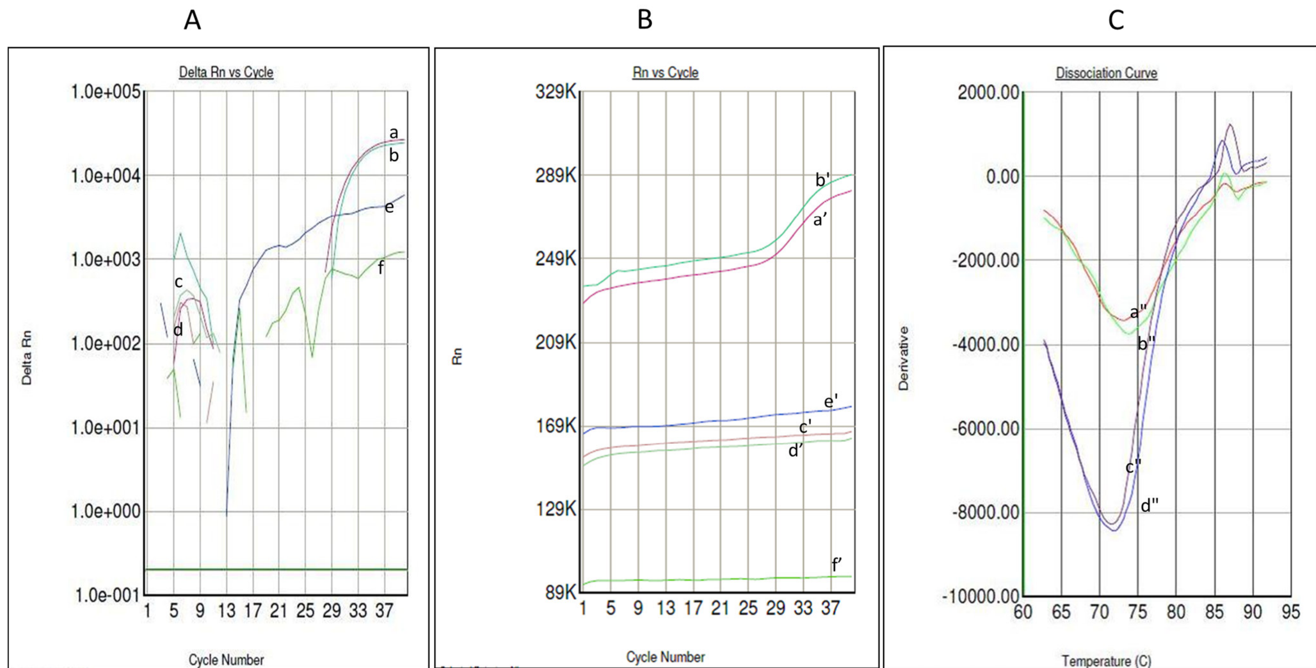


Figure - 9: Molecular Beacon based Real Time PCR result for methyltransferase domain containing protein gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depict amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

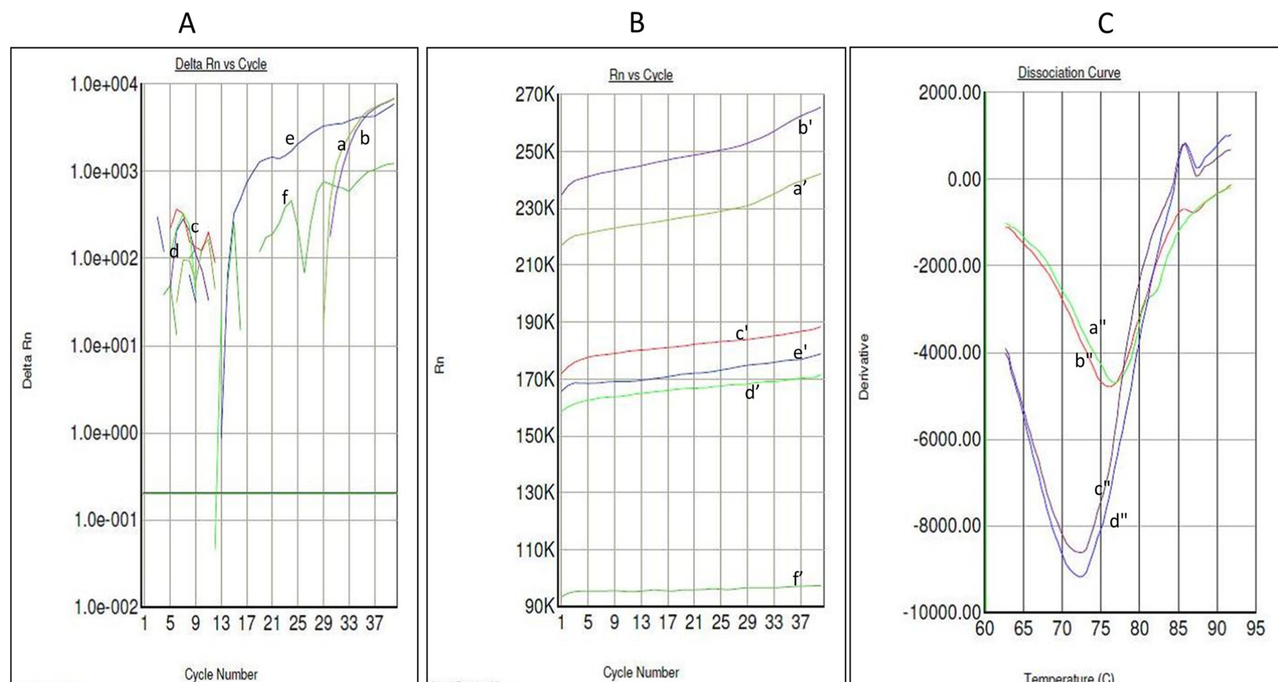


Figure - 10: Molecular Beacon based Real Time PCR result for 60S ribosomal protein L7 gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depict amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe

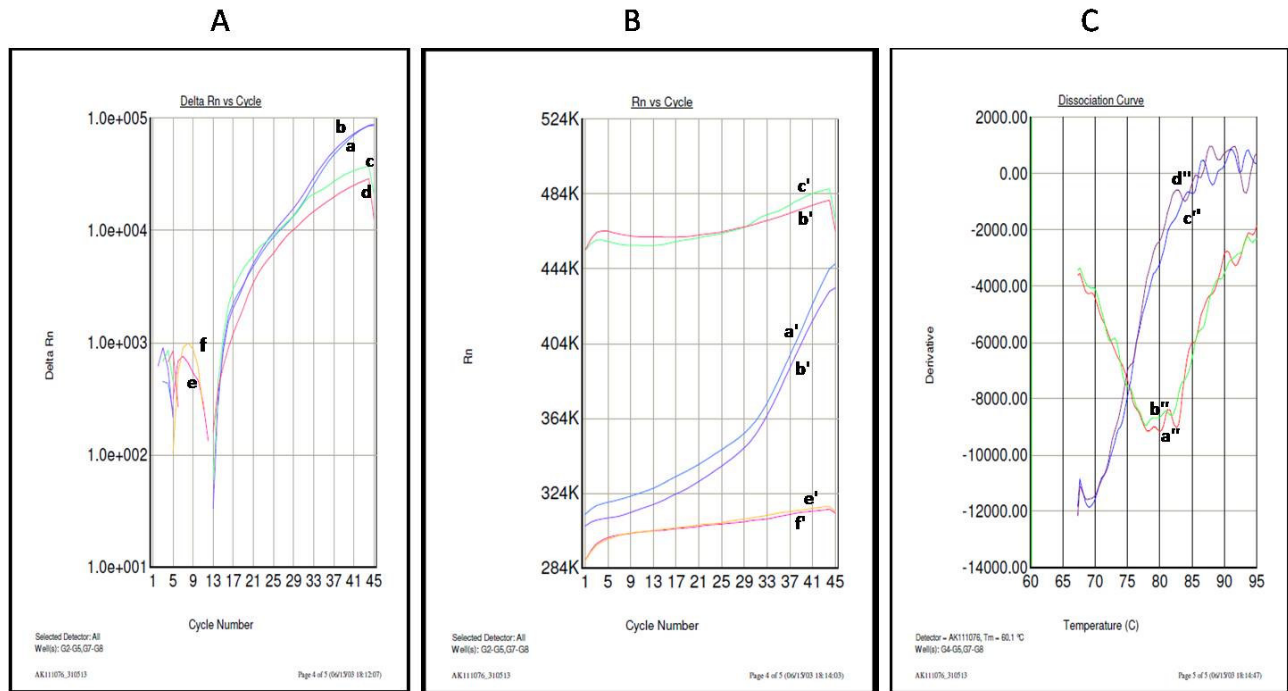


Figure - 11: Molecular Beacon based Real Time PCR result for Heat shock protein DnaJ gene (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depict amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

Table-7: Ct value chart of UR-DEGs like ubiquinol-cytochrome C chaperone family protein gene (AK068288 or LOC_Os07g30790), methyltransferase domain containing protein gene (LOC_Os06g05910 or AK 064640), 60S ribosomal protein L7 gene (LOC_Os08g42920 or AK058490) and heat shock protein DnaJ gene (LOC_Os06g09560 or AK111076) non template control, NTC and with Template having Molecular Beacon probes specific to GCC box or TCC box. PCR amplification for each gene was performed in two replicates. Ct values and Std dev Ct were obtained from inbuilt Real Time Software.

Gene ID	**Replicates**	**Template**	**Molecular Beacon**	**Ct value**	**Stddev Ct**
AK068288	R1	Template	GCC box	28.03	1.23
	R2	Template	GCC box	29.19	1.23
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-
AK064640	R1	Template	GCC box	28.64	1.23
	R2	Template	GCC box	27.35	1.23
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-
AK058490	R1	Template	GCC box	29.55	1.92
	R2	Template	GCC box	28.88	1.92
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-
AK111076	R1	Template	GCC box	13.34	9.033
	R2	Template	GCC box	12.61	9.033
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-

Identification of GCC box in other UR-DEGs like OsFBDUF66 - F-box and DUF domain containing protein gene (AK120895 or LOC_Os12g39520), Cytokinin-O-glucosyltransferase 2 gene (AK063324 or LOC_Os06g11720), Retrotransposon (Ty3-gypsy subclass) protein gene (AK073072 or LOC_Os09g09650), Retrotransposon (Ty1-copia subclass) protein gene (AK121178 or LOC_Os04g02310), Hypothetical protein gene (AK063204 or LOC_Os07g37280), RNA polymerases N 8 kDa subunit, protein gene (AK108801 or LOC_Os11g08940), Serine acetyltransferase protein gene (AK067089 or LOC_Os03g04140), Transposon protein unclassified gene (AK121698 or LOC_Os10g42150), Deoxyhypusine hydroxylase gene (AK059924 orLOC_Os12g43100), Protein of unknown function domain containing protein gene (AK121619 or LOC_Os06g40040), Expressed protein gene (AK071086 or LOC_Os12g23780), MYB family transcription factor gene (Ak068565 or LOC_Os01g34060), Expressed protein gene (AK102580 or LOC_Os03g10460) was also performed. After completion of PCR amplification graphs of Delta Rn vs cycle number for above mentioned UR-DEGs were also retrieved from the inbuilt Real Time PCR system software (Figures – 12A, 13A, 14A, 15A, 16A, 17A, 18A, 19A, 20A & 21A). In Rn vs cycle number graphs, a′ & b′ curves indicate the amplification of respective gene with GCC probe; c′ & d′ curves for genes having TCC probe and e′ & f′ for non template control, NTC (Figure – 12B, 13B, 14B, 15B, 16B, 17B, 18B, 19B, 20B & 21B). Similarly, dissociation curves of these UR-DEGs indicated the amplification of GCC and TCC box having only two specific products. A dissociation curve (a″ & b″ depicting the two independent replications for the amplification with GCC probe while c″ & d″ for TCC probe of UR-DEGs have been shown (Figure - 12C, 13C, 14C, 15C, 16C, 17C, 18C, 19C, 20C & 21C). Unfortunately, in our experiment three genes like expressed protein gene (AK071086 or LOC_Os12g23780), MYB family transcription factor gene (Ak068565 or LOC_Os01g34060), Expressed protein gene (AK102580 or LOC_Os03g10460) have not shown PCR amplification after several repeated set of the experiment. Ct values chart for OsFBDUF66 - F-box and DUF domain containing protein gene, Cytokinin-O-glucosyltransferase 2 gene, Retrotransposon (Ty3-gypsy subclass) protein gene, Retrotransposon (Ty1-copia subclass) protein gene, Hypothetical protein gene, RNA polymerases N (8 kDa subunit) protein gene, Serine acetyltransferase protein gene, Transposon protein unclassified gene, Deoxyhypusine hydroxylase gene and protein of unknown function domain containing protein gene were also prepared and enlisted in Table - 8. Ct values and Std dev Ct were obtained from inbuilt Real Time Software. To get the more authentic Real Time PCR amplification pattern each gene was analyzed in two

replicates. Amplification was done with non template control (NTC) and with Template having Molecular Beacon probes specific to GCC box. To see the specific amplification of GCC box probe we also performed PCR amplification with non-specific TCC box probe. Unfortunately, no good amplification pattern with Template having Molecular Beacon probes specific to GCC box was recorded.

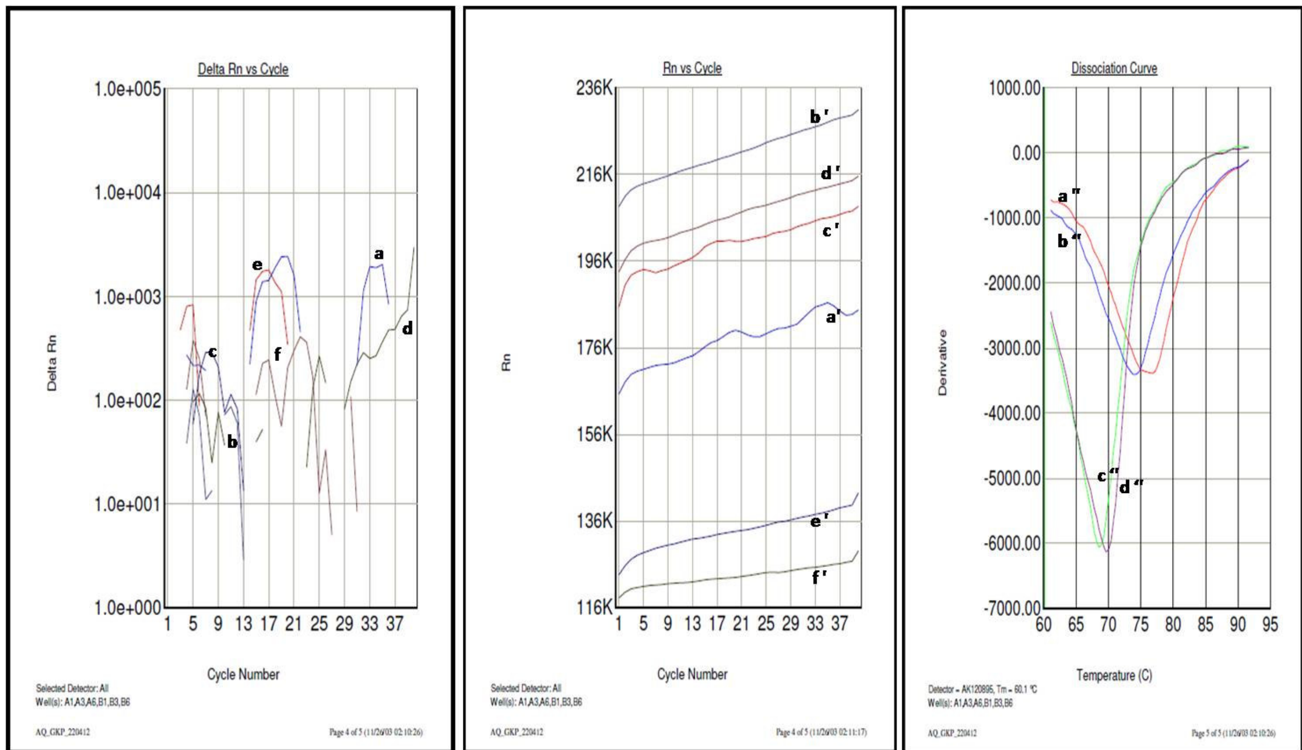


Figure – 12: Molecular Beacon based Real Time PCR result for OsFBDUF66 - F-box and DUF domain containing protein gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depict amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a"&b"* depict the dissociation curve of gene with GCC probe and *c"&d"* are dissociation curves of gene with TCC probe.

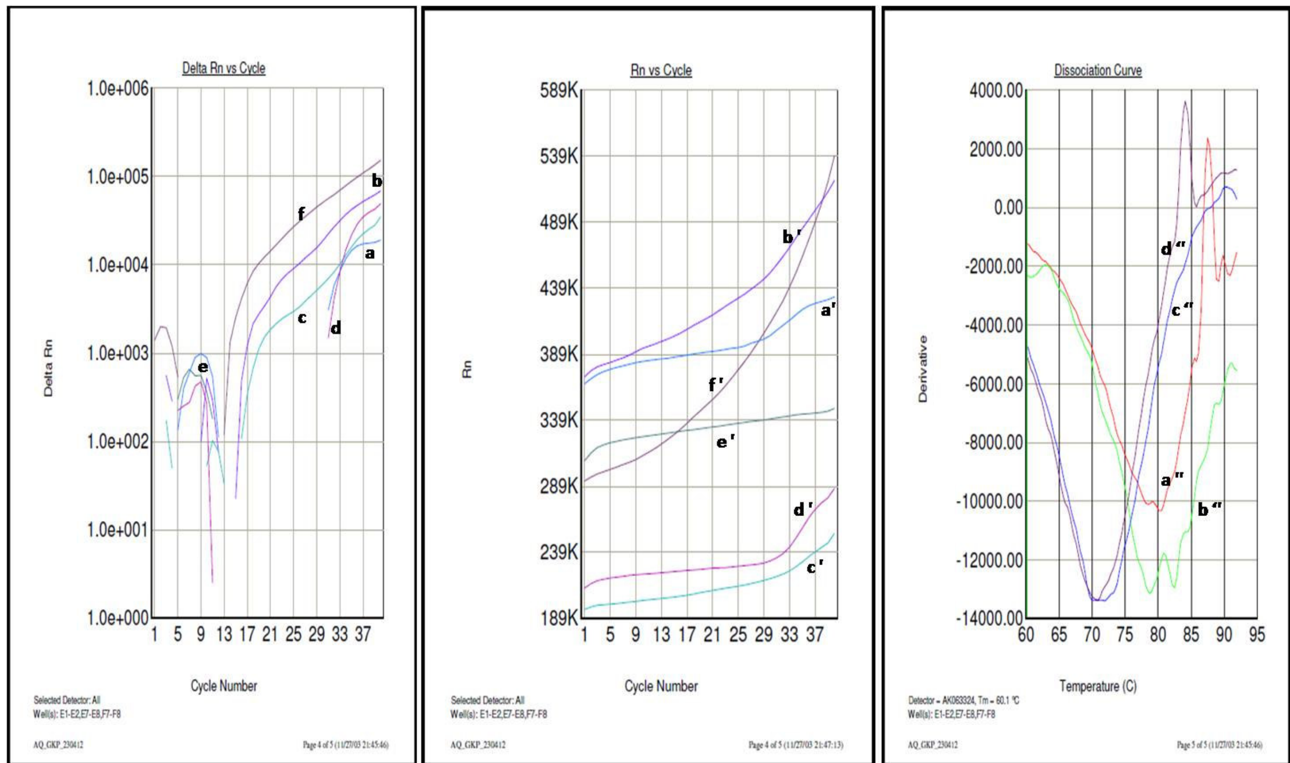


Figure – 13: Molecular Beacon based Real Time PCR result for Cytokinin-O-glucosyltransferase 2 gene. (A) Graph indicating the relation between Delta Rnvs cycle number. Curves *a&b* depict amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rnvs cycle number. Curves *a′&b′* depict amplification of gene with GCC probe. Curves *c′&d′* are the amplification of TCC probe. Curves *e′&f′* are amplification of NTC. (C) *a″&b″* depict the dissociation curve of gene with GCC probe and *c″&d″* are dissociation curves of gene with TCC probe.

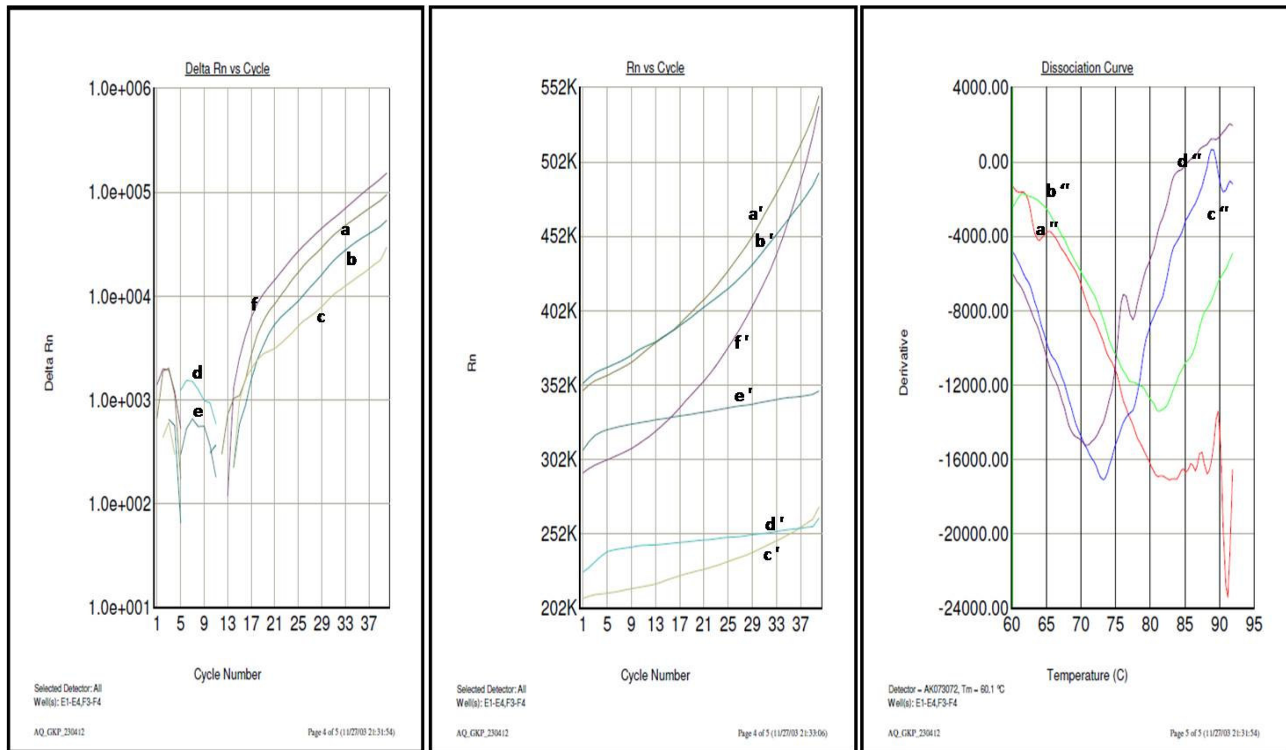


Figure – 14: Molecular Beacon based Real Time PCR result for Retrotransposon (Ty3-gypsy subclass) protein gene. (A) Graph indicating the relation between Delta Rnvs cycle number. Curves *a&b* depict amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

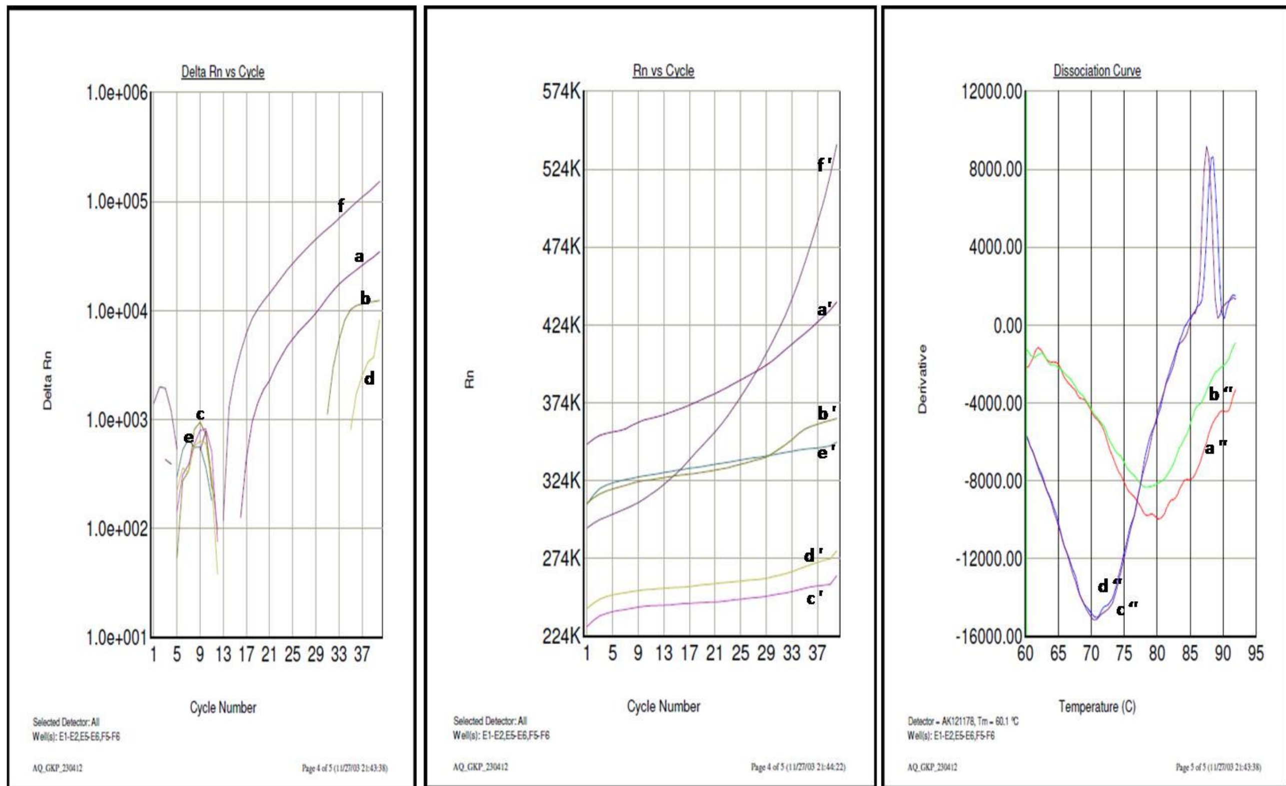


Figure – 15: Molecular Beacon based Real Time PCR result for Retrotransposon protein (Ty1-copia subclass) gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depict amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

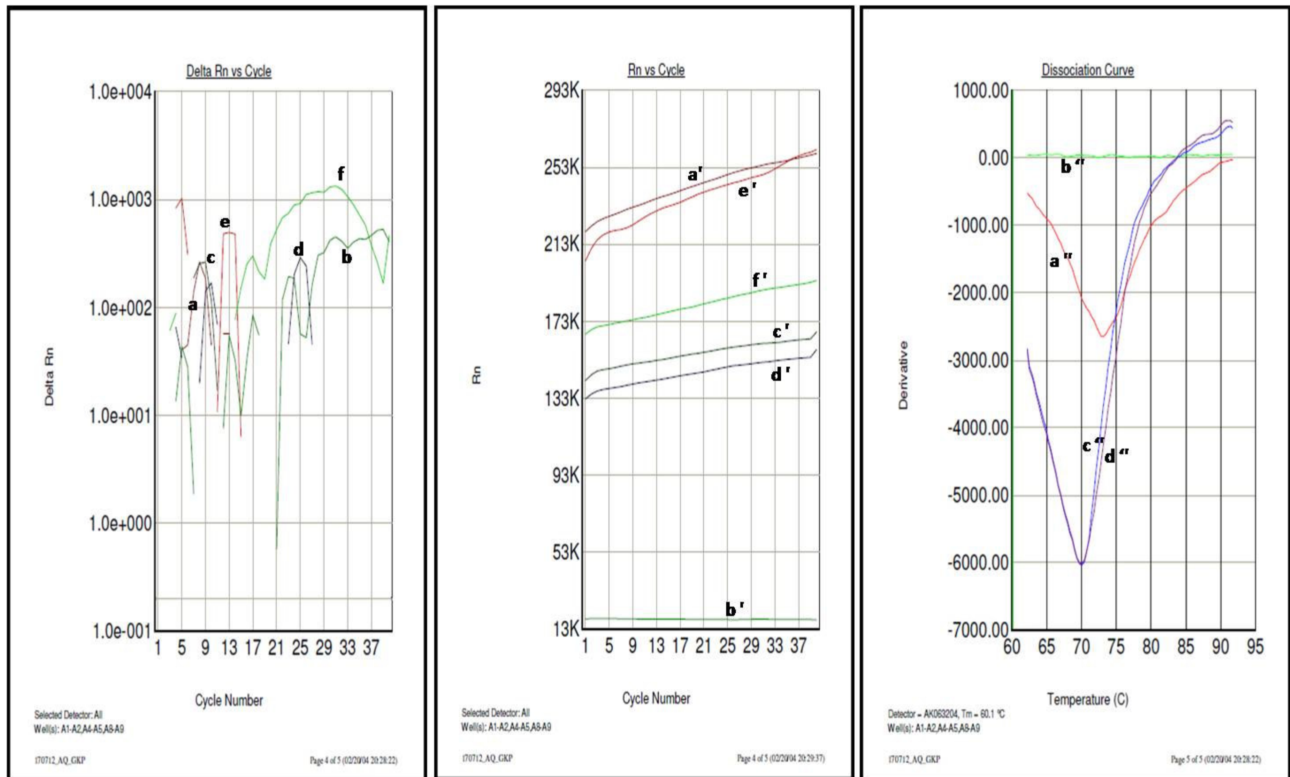


Figure – 16: Molecular Beacon based Real Time PCR result for Hypothetical protein gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

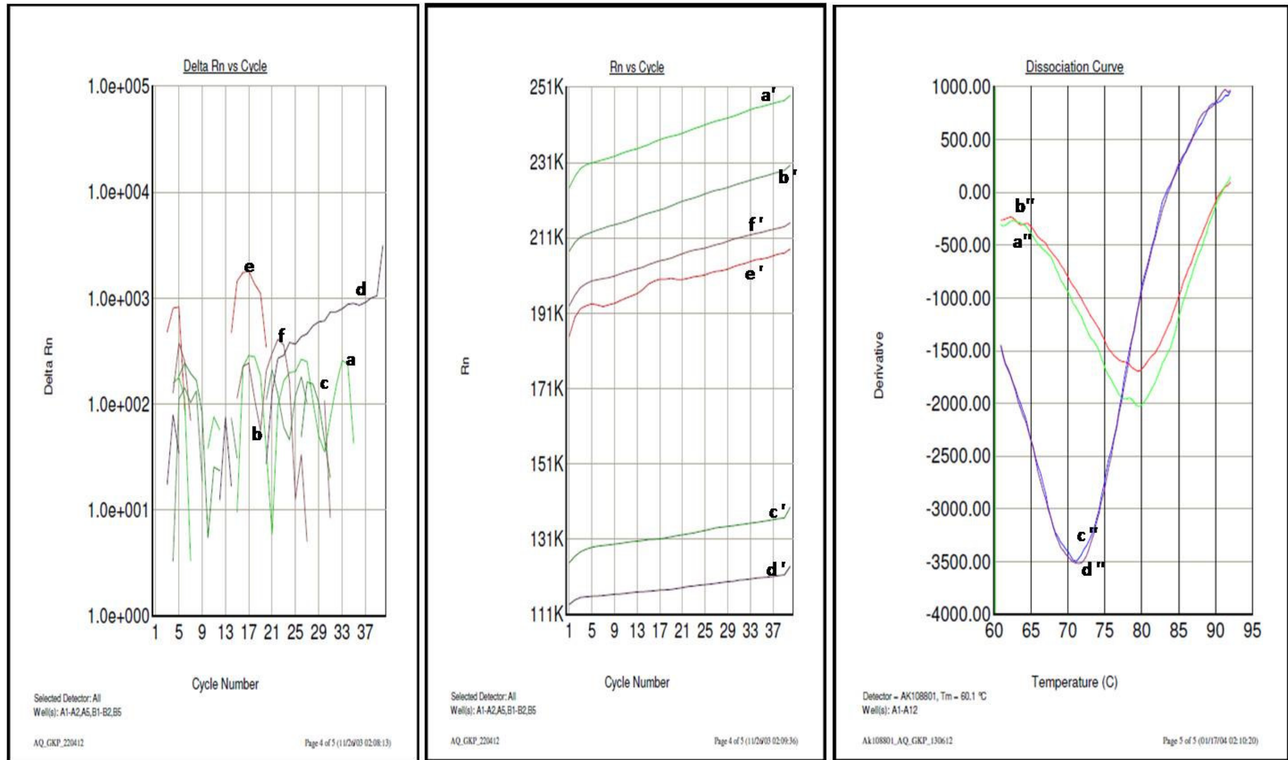


Figure – 17: Molecular Beacon based Real Time PCR result for RNA polymerases (N 8 kD a subunit) protein gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depict amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

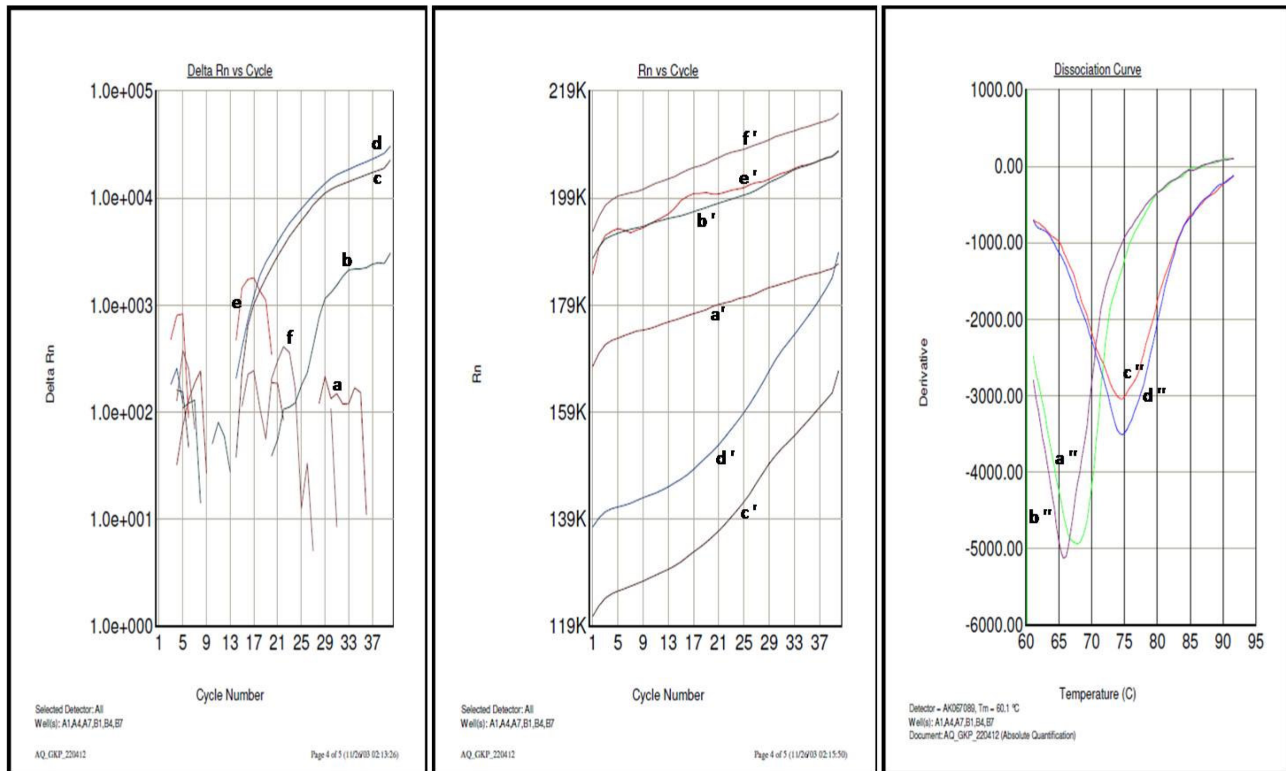


Figure – 18: Molecular Beacon based Real Time PCR result for Serine acetyltransferase protein gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a′&b′* depict amplification of gene with GCC probe. Curves *c′&d′* are the amplification of TCC probe. Curves *e′&f′* are amplification of NTC. (C) *a″&b″* depict the dissociation curve of gene with GCC probe and *c″&d″* are dissociation curves of gene with TCC probe.

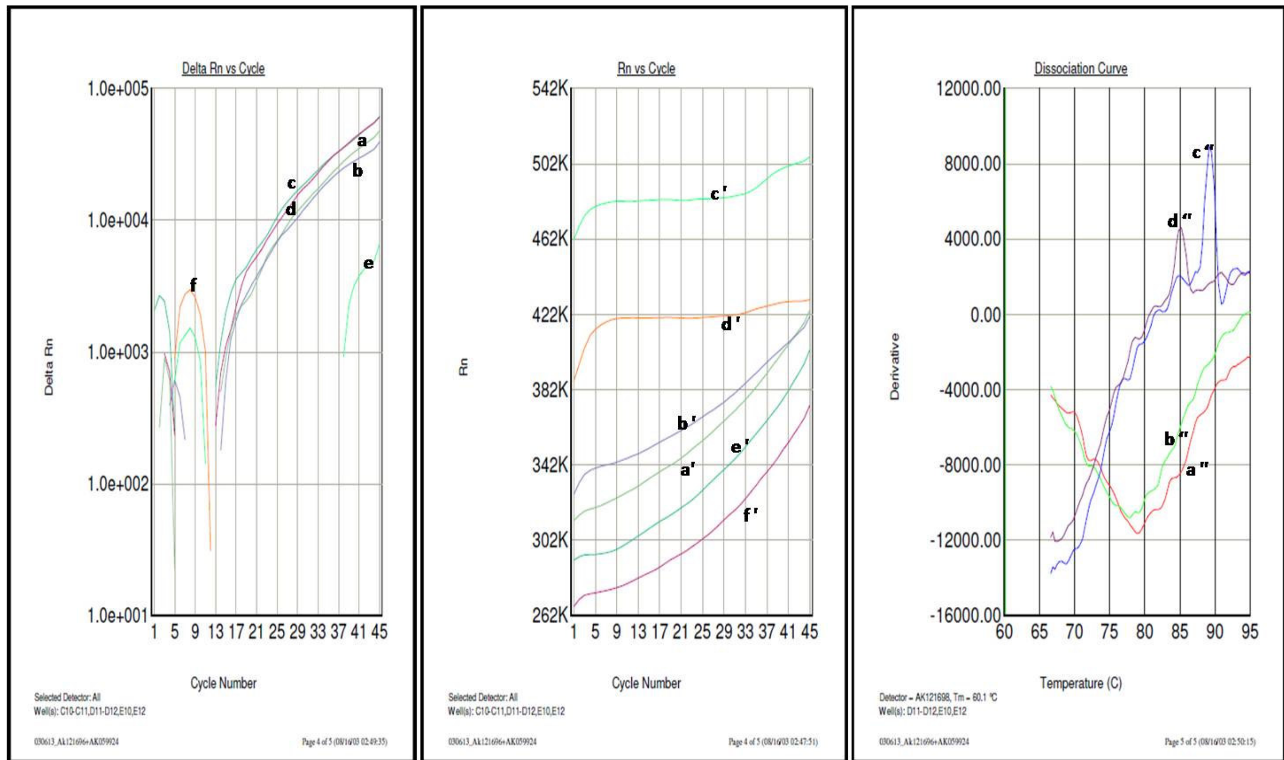


Figure – 19: Molecular Beacon based Real Time PCR result for Transposon (unclassified) protein gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a*&*b* depicts amplification of gene with GCC probe. Curves *c*&*d* are the amplification of TCC. Curves *e*&*f* are amplification of NTC. (B) Graph indicating relationship between Rnvs cycle number. Curves *a'*&*b'* depict amplification of gene with GCC probe. Curves *c'*&*d'* are the amplification of TCC probe. Curves *e'*&*f'* are amplification of NTC. (C) *a''*&*b''* depict the dissociation curve of gene with GCC probe and *c''*&*d''* are dissociation curves of gene with TCC probe.

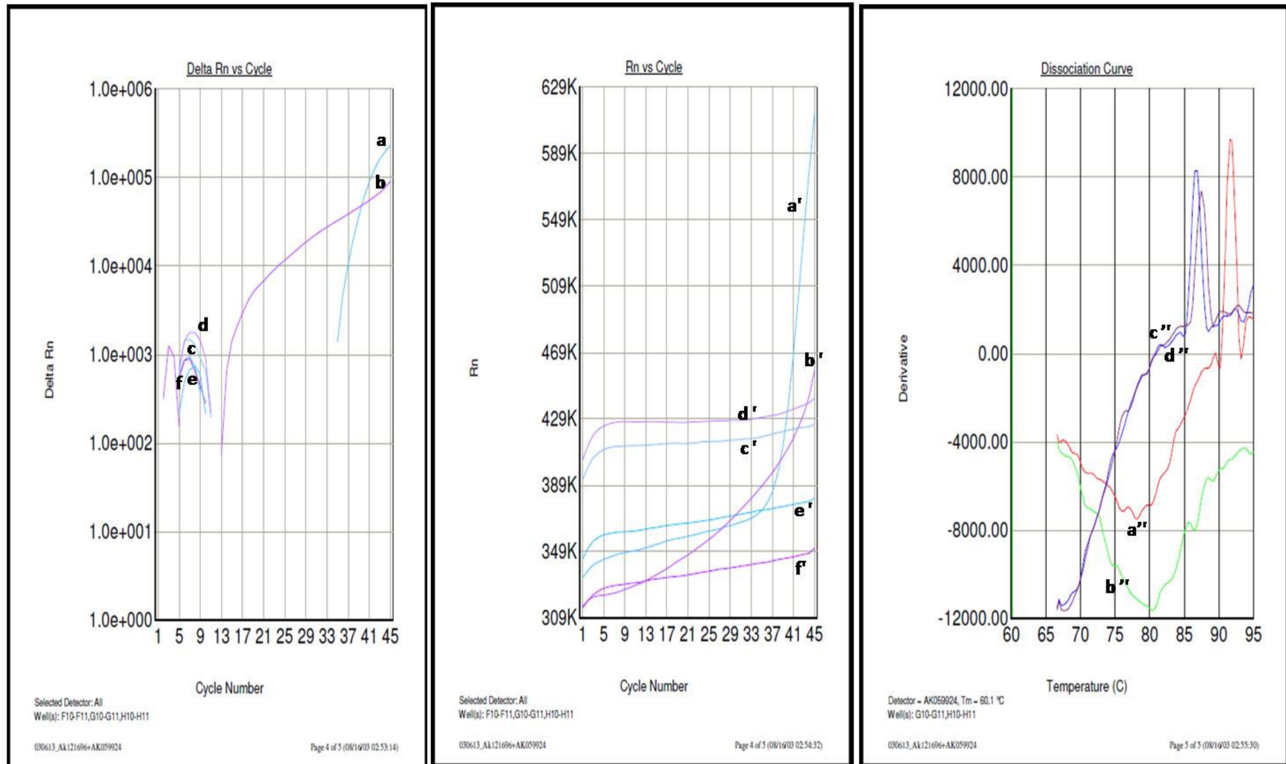


Figure – 20: Molecular Beacon based Real Time PCR result for Deoxyhypusine hydroxylase protein gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

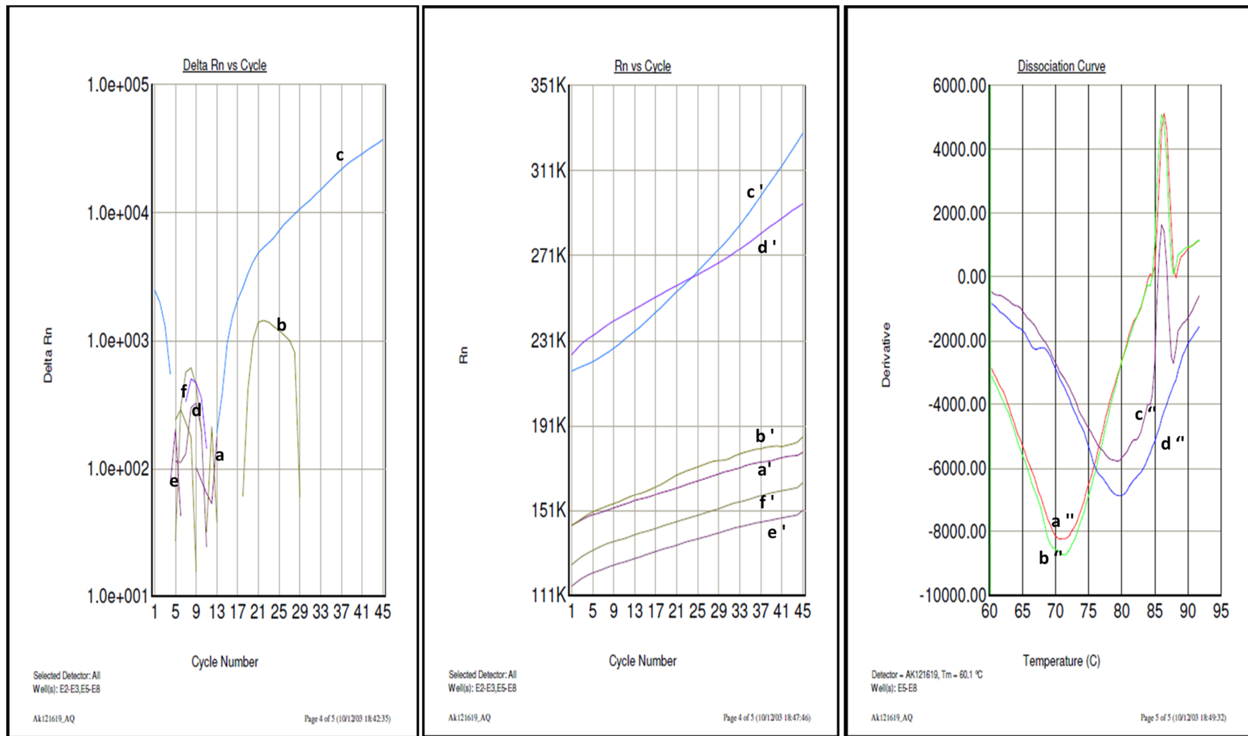


Figure – 21: Molecular Beacon based Real Time PCR result for Protein of unknown function domain containing protein gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depict amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rnvs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a'' &b''* depict the dissociation curve of gene with GCC probe and *c'' &d''* are dissociation curves of gene with TCC probe.

Table-8: Ct value chart of UR-DEGs like OsFBDUF66 - F-box and DUF domain containing protein gene (AK120895 or LOC_Os12g39520), Cytokinin-O-glucosyltransferase 2 gene (AK063324 or LOC_Os06g11720), Retrotransposon (Ty3-gypsy subclass) protein gene (AK073072 or LOC_Os09g09650), Retrotransposon (Ty1-copia subclass) protein gene (AK121178 or LOC_Os04g02310), Hypothetical protein gene (AK063204 or LOC_Os07g37280), RNA polymerases (N 8 kDa subunit) protein gene (AK108801 or LOC_Os11g08940), Serine acetyltransferase protein gene (AK067089 or LOC_Os03g04140), Transposon protein unclassified gene (AK121698 or LOC_Os10g42150), Deoxyhypusine hydroxylase gene (AK059924 or LOC_Os12g43100), Protein of unknown function domain containing protein gene (AK121619 or LOC_Os06g40040), non template control, NTC and with Template having Molecular Beacon probes specific to GCC box or TCC box. PCR amplification for each gene was performed in two replicates. Ct values and Std dev Ct were obtained from inbuilt Real Time Software.

Gene ID	**Replicates**	**Template**	**Molecular Beacon**	**Ct value**	**Std dev Ct**
AK120895	R1	Template	GCC box	30.76	15.33
	R2	Template	GCC box	9.08	15.33
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-
AK063324	R1	Template	GCC box	30.00	11.08
	R2	Template	GCC box	14.32	11.08
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-
AK073072	R1	Template	GCC box	11.58	1.44
	R2	Template	GCC box	13.62	1.44
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-
AK121178:	R1	Template	GCC box	15.08	10.81
	R2	Template	GCC box	30.38	10.81
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-
	R1	Template	GCC box	11.46	8.53

AK063204:	R2	Template	GCC box	20.99	8.53
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-
AK108801	R1	Template	GCC box	39.40	14.29
	R2	Template	GCC box	19.19	14.29
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-
AK067089	R1	Template	GCC box	19.15	0.32
	R2	Template	GCC box	19.62	0.32
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-
AK121698	R1	Template	GCC box	13.02	0.27
	R2	Template	GCC box	13.45	0.27
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-
AK059924	R1	Template	GCC box	16.91	13.79
	R2	Template	GCC box	12.83	13.79
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-
AK121619	R1	Template	GCC box	8.34	6.86
	R2	Template	GCC box	17.34	6.86
	R1	Template	TCC box	-	-
	R2	Template	TCC box	-	-
	R1	No	GCC box	-	-
	R2	No	GCC box	-	-

***Identification of TCC Box in DR-DEGs*:** On the other hand identification of TCC box in DR-DEGs like rhoGAP domain containing protein gene (AK067300 or LOC_Os12g05900), DnaK family protein gene (AK100997 or LOC_Os02g48110), CPuORF11 - conserved peptide uORF-containing transcript gene (AK103103 orLOC_Os02g01240), OsFBX61 - F-box domain containing protein gene (AK103417 or LOC_Os02g52130), CGMC_MAPKCMGC_2_SLT2y_ERK.2 - CGMC includes CDA, MAPK, GSK3, and CLKC kinases gene (AK071376 or LOC_Os06g48590), Helix-loop-helix DNA-binding domain containing protein gene (AK068704 or LOC_Os03g26210), spermidine synthase gene (AK065153 or LOC_Os02g15550) were also done by Real Time PCR. After completion of PCR graphs of Delta Rn vs cycle number for above seven DR-DEGs were retrieved from the inbuilt Real Time PCR system software (Figures – 22A, 23A, 24A, 25A, 26A, 27A &28A). In Rn vs cycle number graphs, a′ & b′ curves indicate the amplification of respective gene with GCC probe; c′ & d′ curves for genes having TCC probe and e′ & f′ for non template control, NTC (Figures – 22B, 23B, 24B, 25B, 26B, 27b &28 B). Similarly, dissociation curves of above DR-DEGs indicating the amplification of GCC and TCC box having only two specific products were also prepared. A dissociation curve (a″ & b″ depicting the two independent replications for the amplification with GCC probe while c″ & d″ for TCC probe of UR-DEGs has been shown in Figures – 22C, 23C, 24C, 25C, 26C, 27C &28 C. Ct values chart for rhoGAP domain containing protein gene, DnaK family protein gene, CPuORF11 - conserved peptide uORF-containing transcript gene, OsFBX61 - F-box domain containing protein gene,CGMC_MAPKCMGC_2_SLT2y_ERK.2 - CGMC includes CDA, MAPK, GSK3, and CLKC kinases gene, Helix-loop-helix DNA-binding domain containing protein gene, Spermidine synthase gene were also prepared and enlisted in Table - 9. Ct values and Stddev Ct were obtained from inbuilt Real Time Software. Here also to get the more authentic Real Time PCR amplification pattern each gene was analyzed in two replicates. Amplification was done with NTC and with Template having Molecular Beacon probes specific to TCC box. To see the specific amplification of TCC box probe we also performed PCR amplification with non-specific GCC box probe. Result showed the good amplification pattern with Template having Molecular Beacon probes specific to TCC box only. This indicates that TCC box is present in the promoter of above selected DR-DEGs.

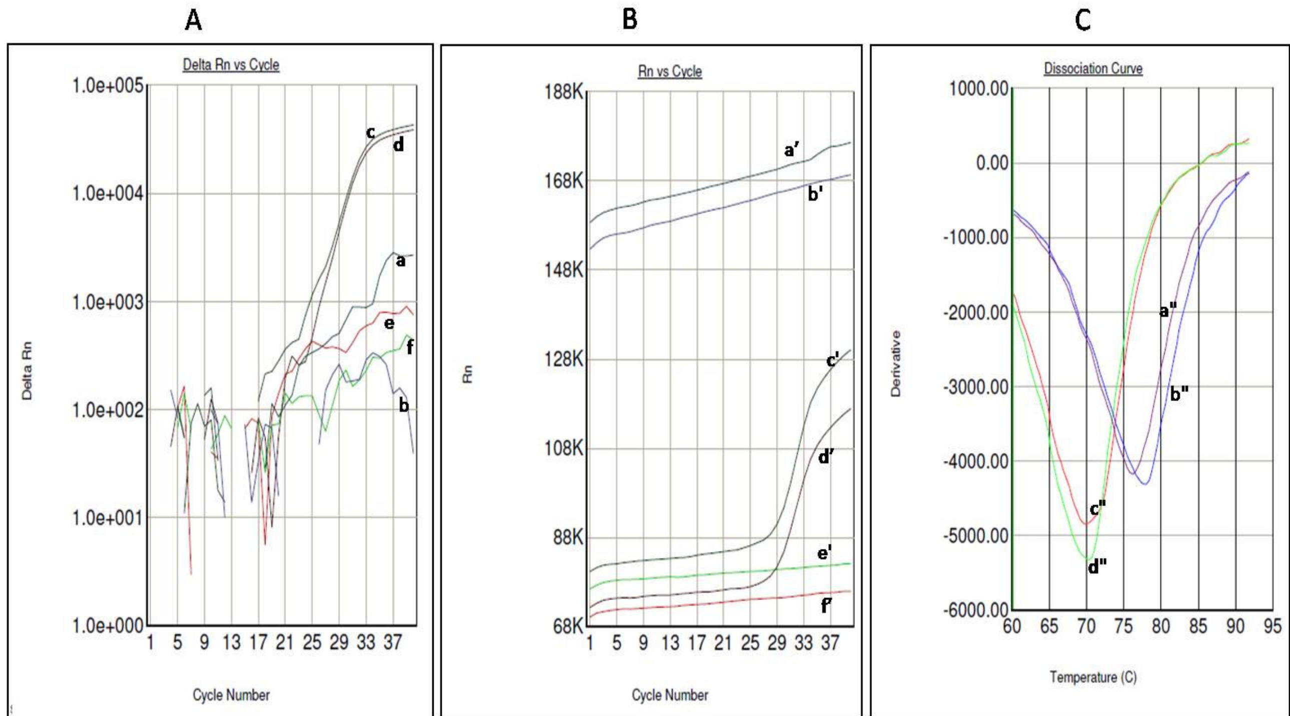


Figure - 22: Molecular Beacon based Real Time PCR result for rhoGAP domain containing protein gene having TCC box in its promoter region (-499 to +100). (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a"&b"* depict the dissociation curve of gene with GCC probe and *c"&d"* are dissociation curves of gene with TCC probe.

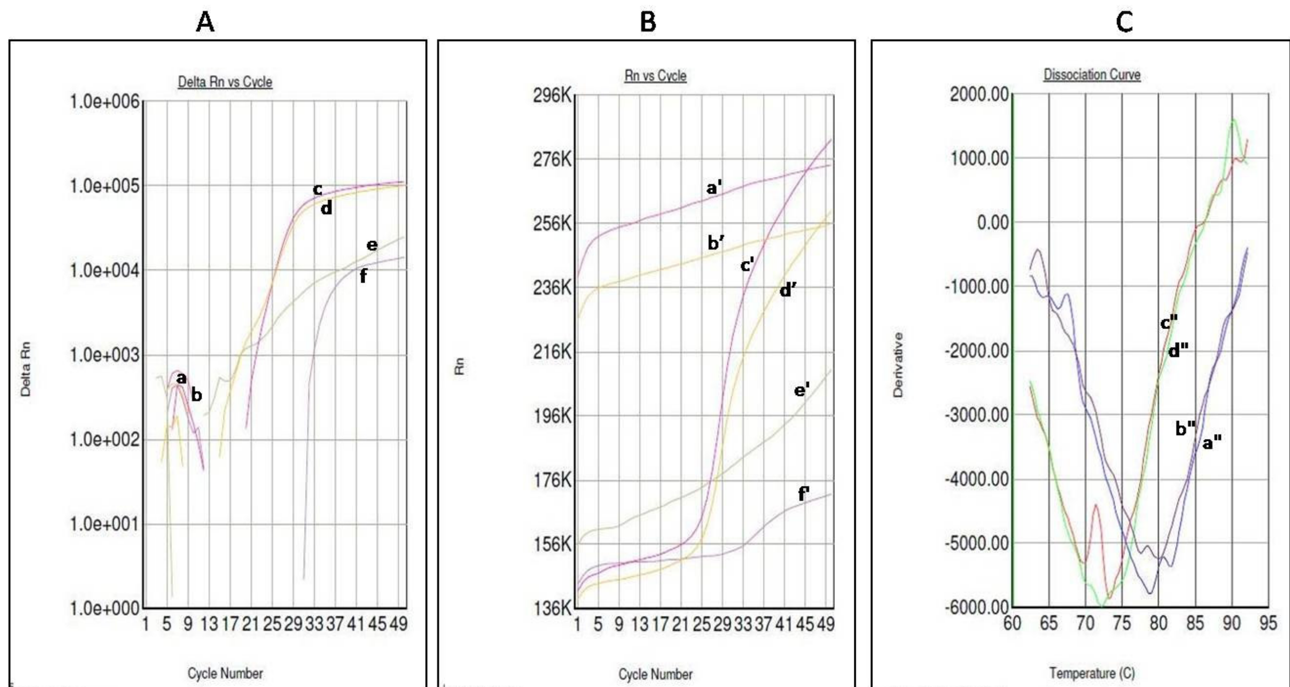


Figure - 23: Molecular Beacon based Real Time PCR result for DnaK family protein gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rnvs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

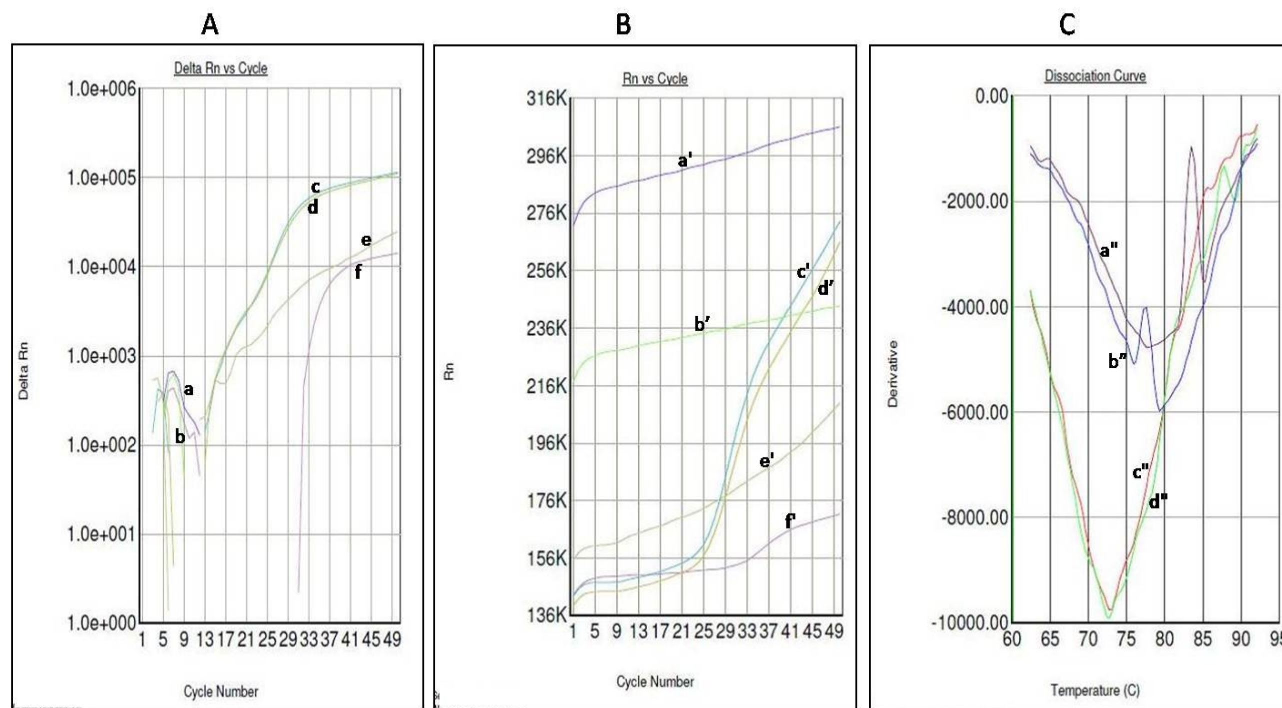


Figure - 24: Molecular Beacon based Real Time PCR result for CPuORF11 - conserved peptide uORF-containing transcript gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

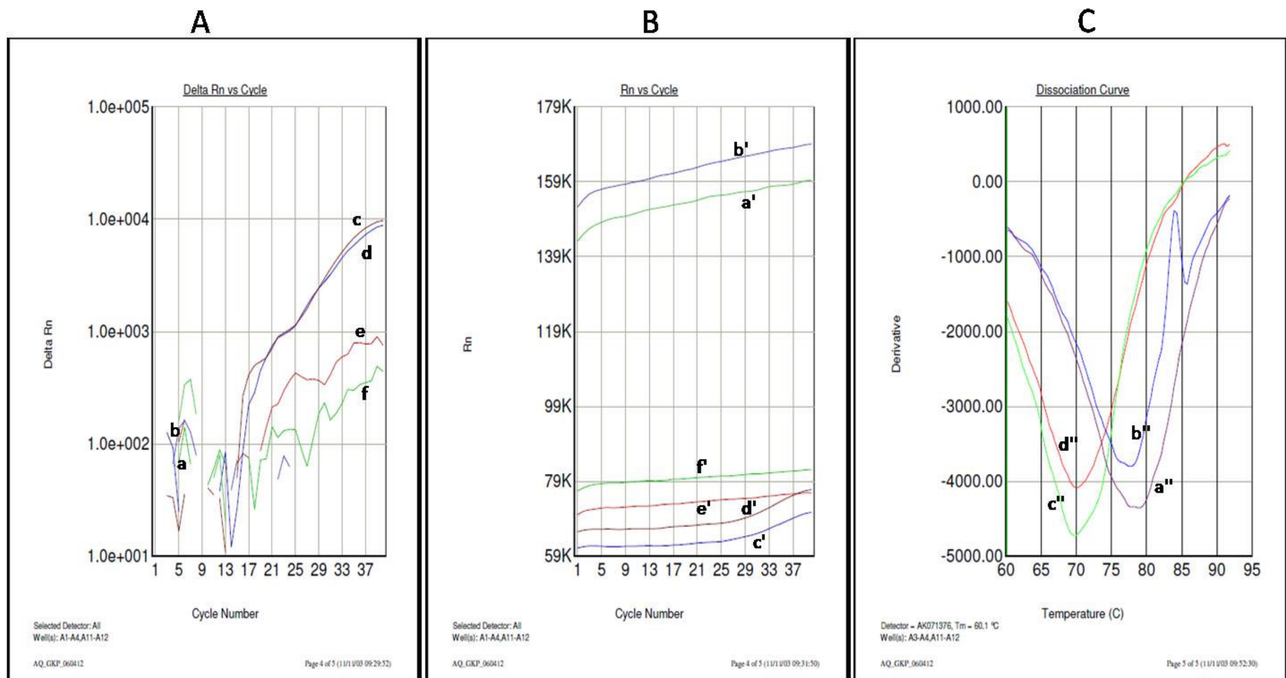


Figure - 25: Molecular Beacon based Real Time PCR result for CPuORF11 - conserved peptide uORF-containing transcript gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

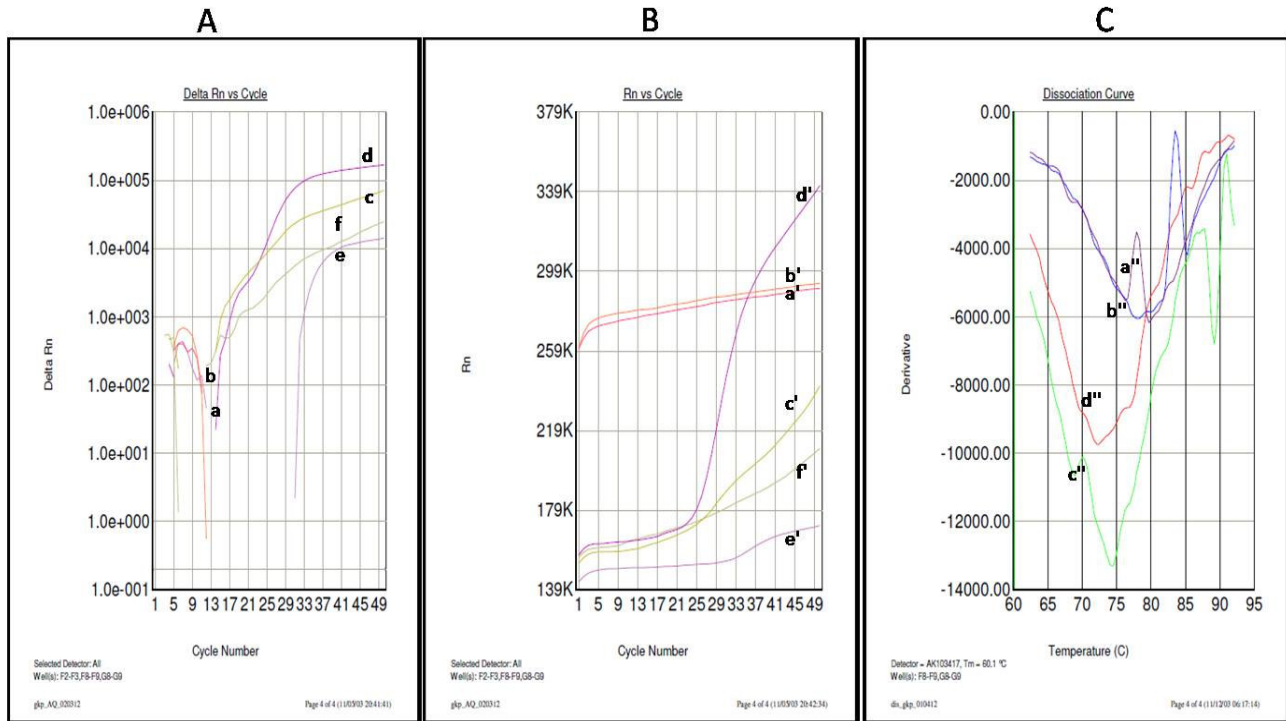


Figure - 26: Molecular Beacon based Real Time PCR result for OsFBX61 - F-box domain containing protein gene (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

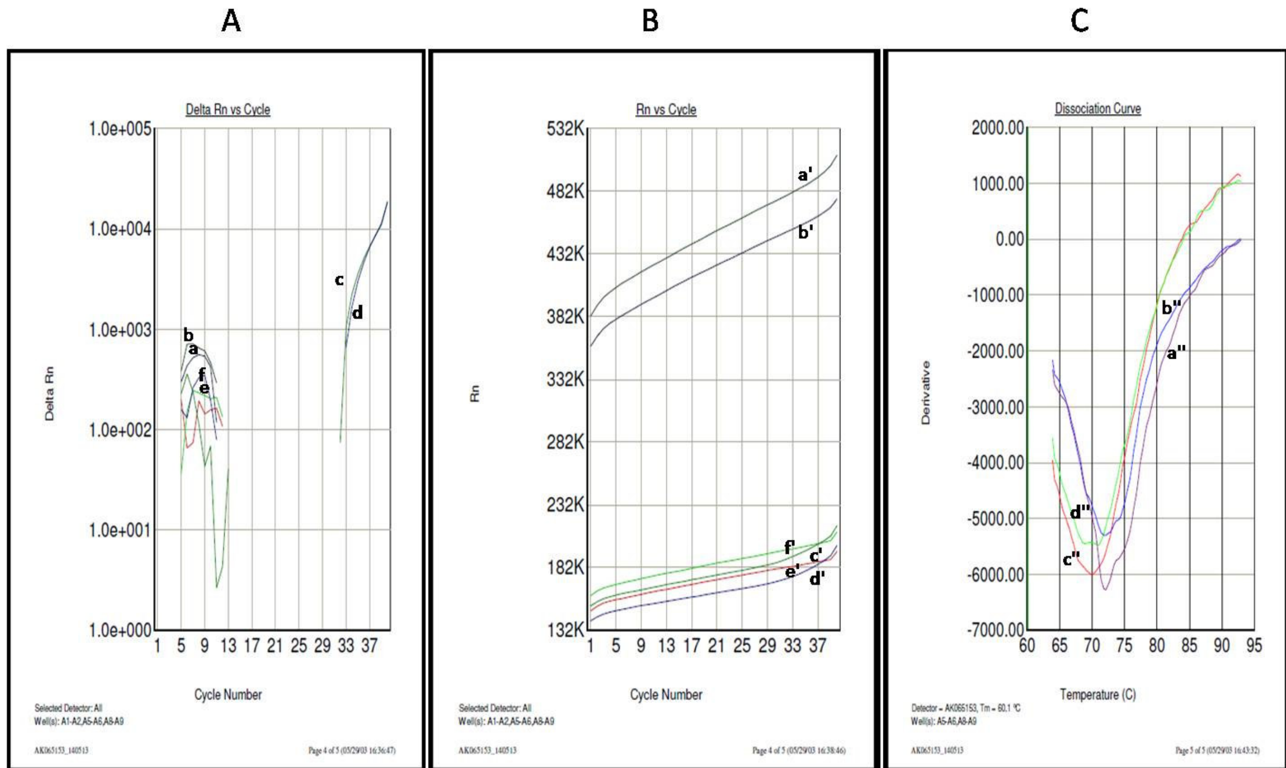


Figure - 27: Molecular Beacon based Real Time PCR result for Spermidine synthase gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

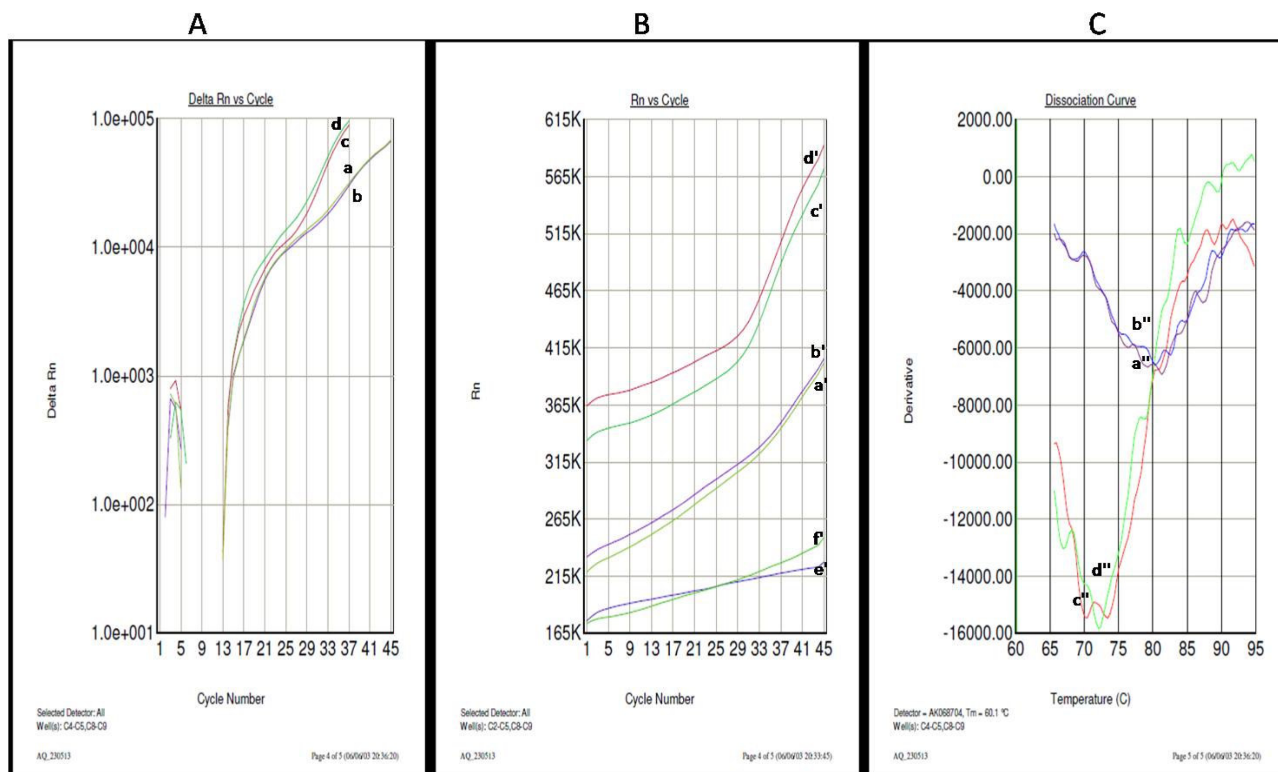


Figure - 28: Molecular Beacon based Real Time PCR result for Helix-loop-helix DNA-binding domain containing protein gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depict amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

Table-9: Ct value chart of DR-DEGs like rhoGAP domain containing protein gene (AK067300 or LOC_Os12g05900), DnaK family protein gene (AK100997 or LOC_Os02g48110), CPuORF11 - conserved peptide uORF-containing transcript gene (AK103103 orLOC_Os02g01240), OsFBX61 - F-box domain containing protein gene (AK103417 or LOC_Os02g52130), CGMC_MAPKCMGC_2_SLT2y_ERK.2 - CGMC includes CDA, MAPK, GSK3, and CLKC kinases gene (AK071376 or LOC_Os06g48590), Helix-loop-helix DNA-binding domain containing protein gene (AK068704 or LOC_Os03g26210), spermidine synthase gene (AK065153 or LOC_Os02g15550) non template control, NTC and with Template having Molecular Beacon probes specific to GCC box or TCC box. PCR amplification for each gene was performed in two replicates. Ct values and Std dev Ct were obtained from inbuilt Real Time Software.

Gene ID	**Replicates**	**Template**	**Molecular Beacon**	**Ct value**	**Stddev Ct**
AK067300	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	15.07	0.849
	R2	Template	TCC box	16.27	0.849
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-
AK100997	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	19.14	2.461
	R2	Template	TCC box	14.23	2.461
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-
AK103103	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	12.28	5.697
	R2	Template	TCC box	12.69	5.697
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-
AK0713076	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	10.76	2.446
	R2	Template	TCC box	14.22	2.446
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-
AK103417	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	13.74	2.461
	R2	Template	TCC box	13.39	2.461
	R1	No	TCC box	-	-

	R2	No	TCC box	-	-
AK065153	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	31.91	9.46
	R2	Template	TCC box	32.20	9.46
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-
AK068704	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	13.08	9.59
	R2	Template	TCC box	12.80	9.59
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-

Identification of TCC box in other DR-DEGs like nitrilase-associated protein gene (AK099444 or LOC_Os11g41150), wall-associated receptor kinase-like 20 precursor gene (AK065517 or LOC_Os09g03620), nucleoside transporter gene (AK102045 or LOC_Os07g37100), sucrose transporter gene (AK100027 or LOC_Os03g07480), phytosulfokines precursor gene (AK073352 or LOC_Os11g05190), transketolase gene (AK100909 or LOC_Os07g09190), TENA/THI-4 family protein gene (AK073626 or LOC_Os03g19390), were also done by Real Time PCR. After completion of PCR graphs of Delta Rn vs cycle number for above seven DR-DEGs were retrieved from the inbuilt Real Time PCR system software (Figures – 29A, 30A, 31A, 32A, 33A, 34A & 35A). In Rn vs cycle number graphs, a′ & b′ curves indicate the amplification of respective gene with GCC probe; c′ & d′ curves for genes having TCC probe and e′ & f′ for non template control, NTC (Figures – 29B, 30B, 31B, 32B, 33B, 34B & 35B). Similarly, dissociation curves of above DR-DEGs indicating the amplification of GCC and TCC box having only two specific products were also prepared. A dissociation curve (a″ & b″ depicting the two independent replications for the amplification with GCC probe while c″ & d″ for TCC probe of UR-DEGs has been shown in Figures – 29C, 30C, 31C, 32C, 33C, 34C & 35C. Ct values chart for Nitrilase-associated protein gene, Wall-associated receptor kinase-like 20 precursor gene, Nucleoside transporter gene, Sucrose transporter gene, Phytosulfokines precursor gene, Transketolase protein gene, TENA/THI-4 family protein gene were also prepared and enlisted in Table –10. Ct values and Stddev Ct were obtained from inbuilt Real Time Software. Here also to get the more authentic Real Time PCR amplification pattern each gene was analyzed in two replicates. Amplification was done

with NTC and with Template having Molecular Beacon probes specific to TCC box. To see the specific amplification of TCC box probe we also performed PCR amplification with non-specific GCC box probe. Unfortunately not very good amplification pattern with Template having Molecular Beacon probes specific to TCC box was recorded.

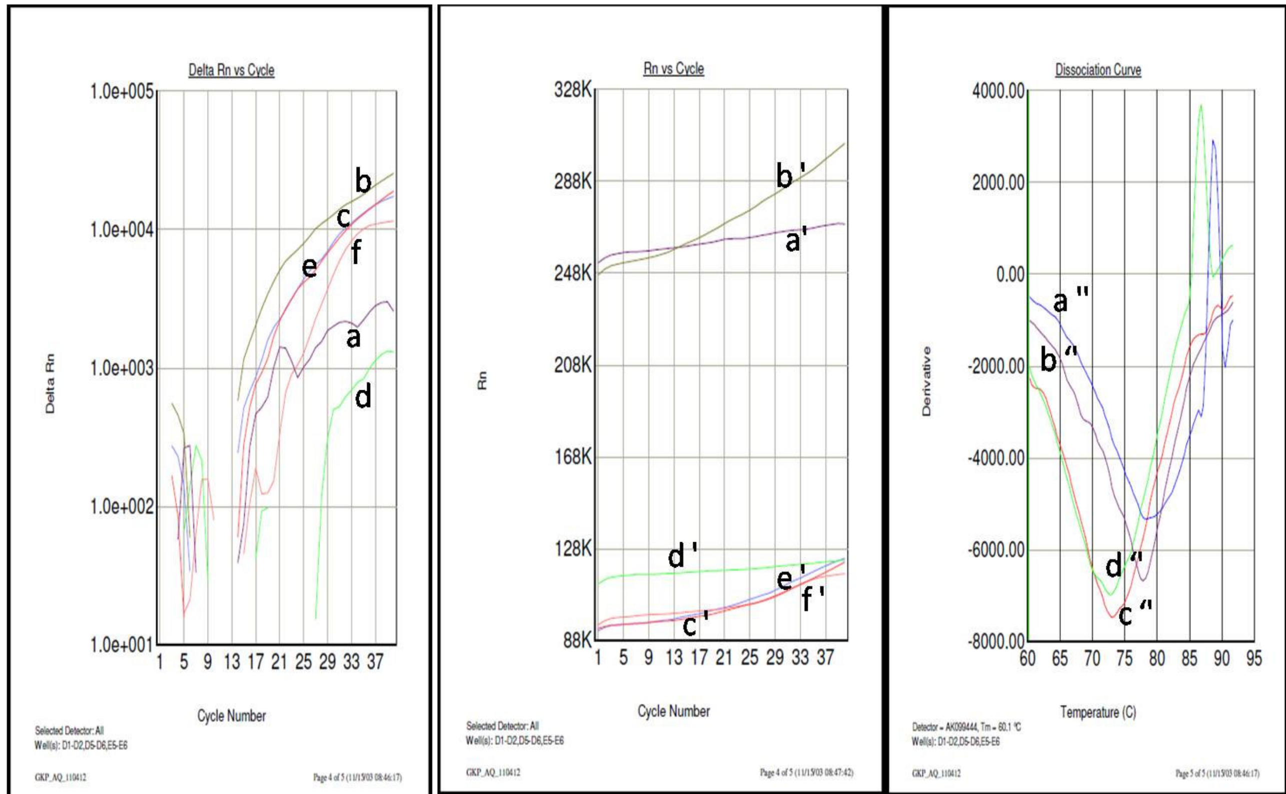


Figure - 29: Molecular Beacon based Real Time PCR result for nitrilase-associated protein gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. (C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

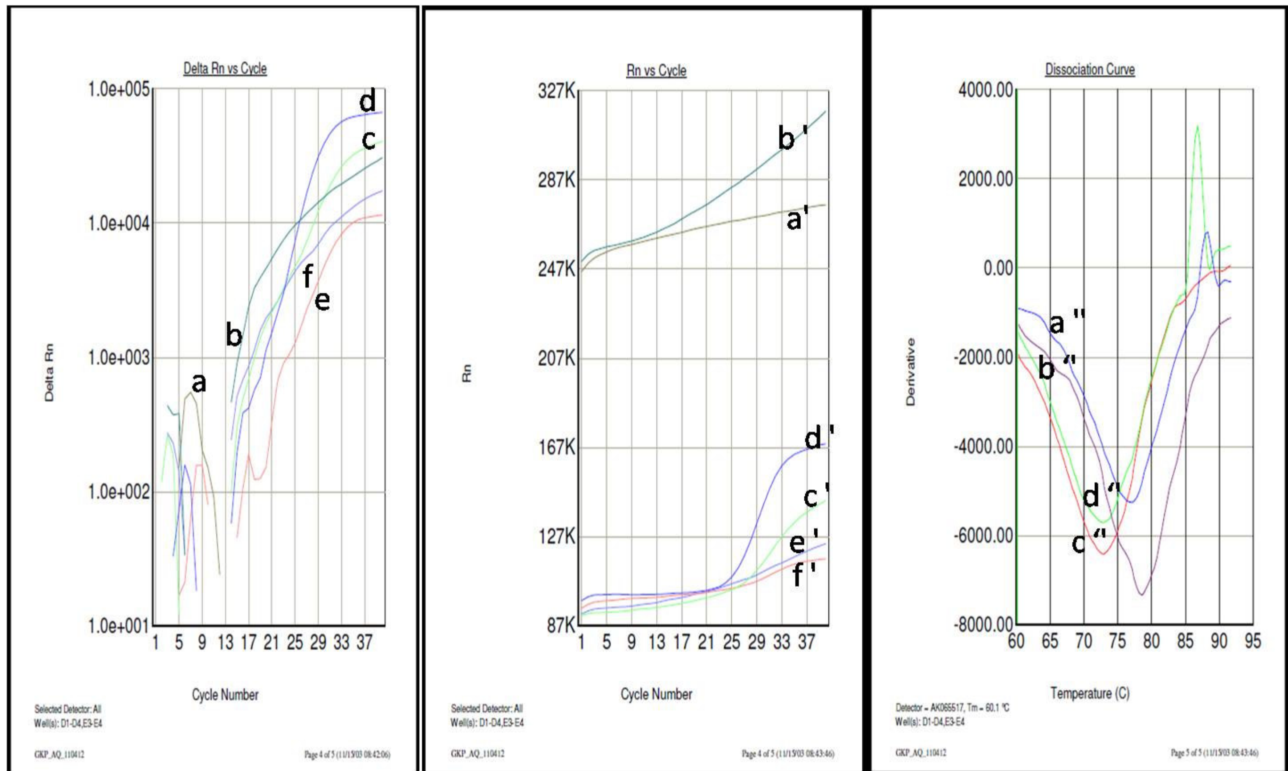


Figure - 30: Molecular Beacon based Real Time PCR result for wall-associated receptor kinase-like 20 precursor gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rnvs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

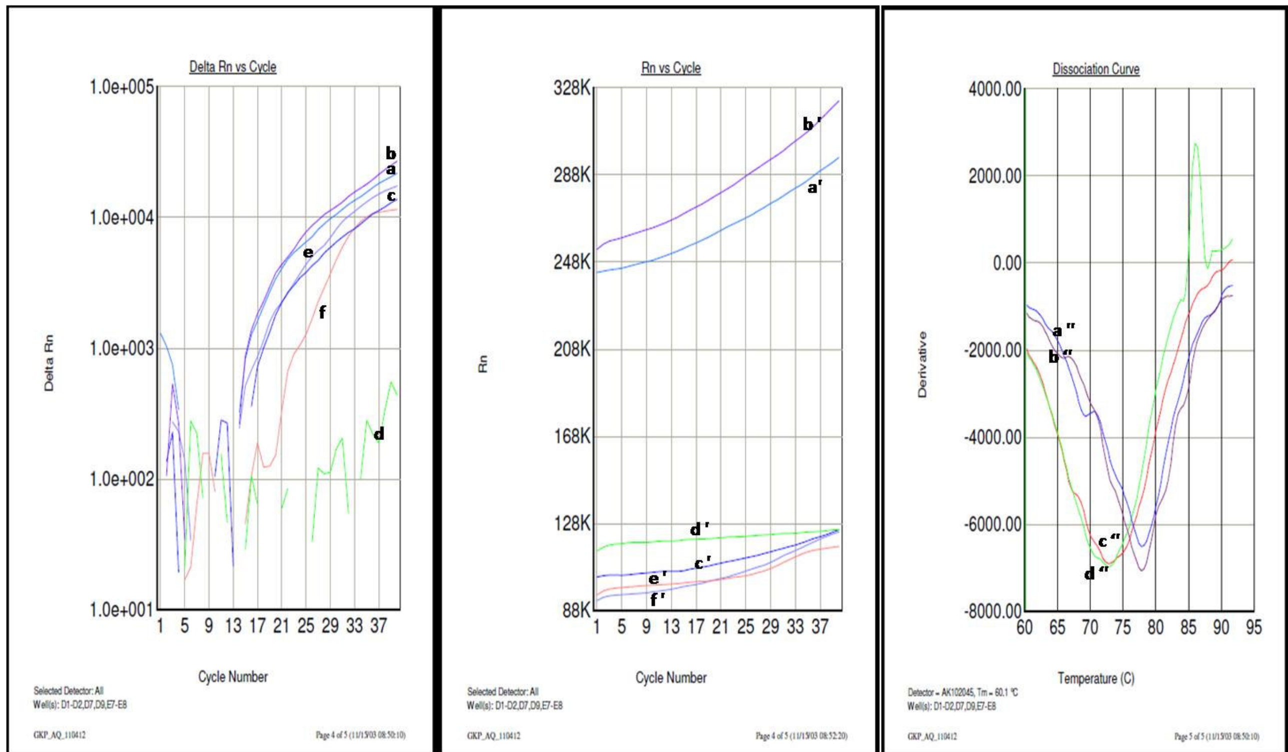


Figure - 31: Molecular Beacon based Real Time PCR result for nucleoside transporter gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

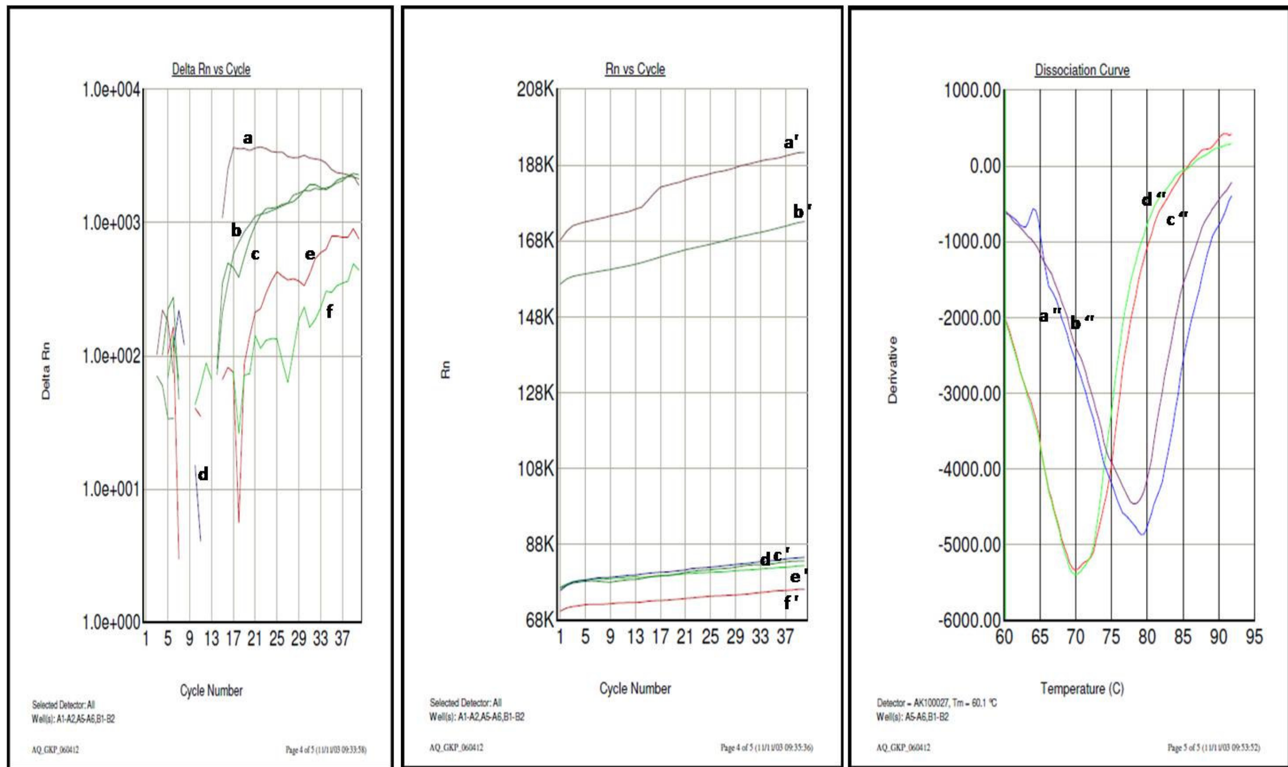


Figure - 32: Molecular Beacon based Real Time PCR result for sucrose transporter gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

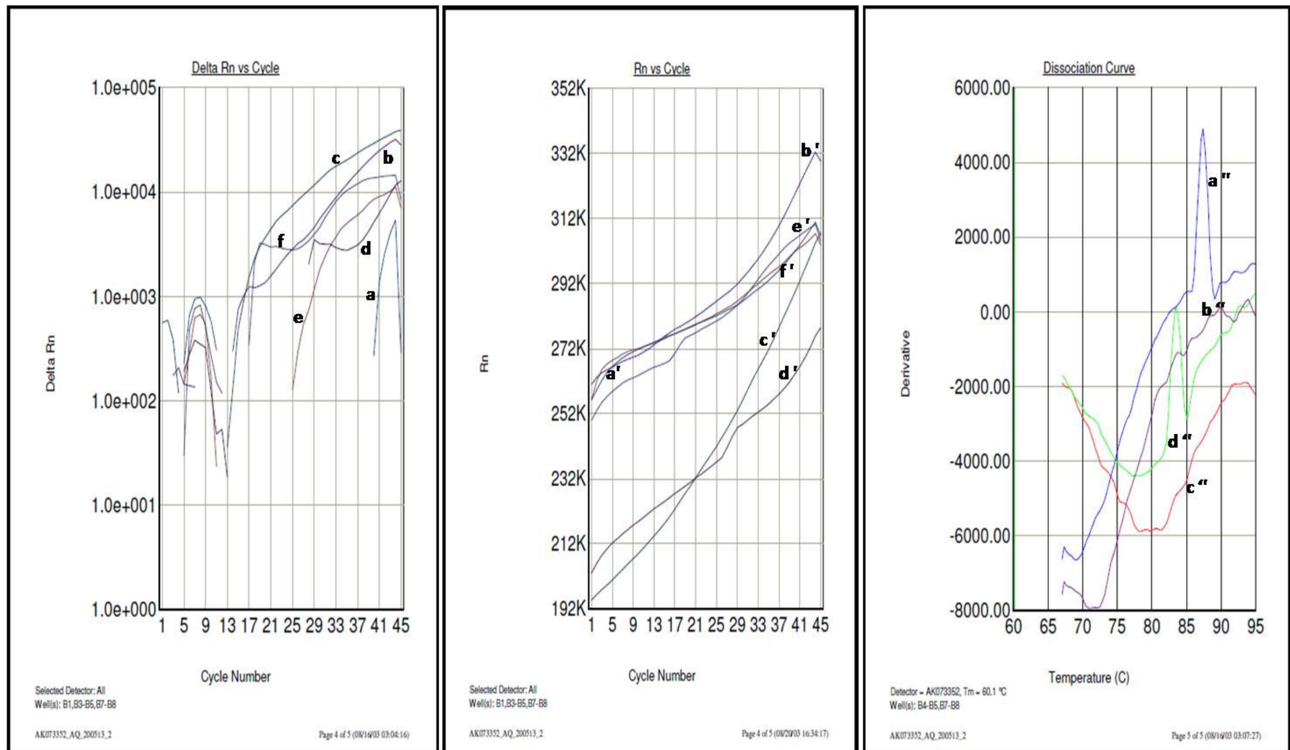


Figure - 33: Molecular Beacon based Real Time PCR result for phytosulfokines precursor gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depict amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

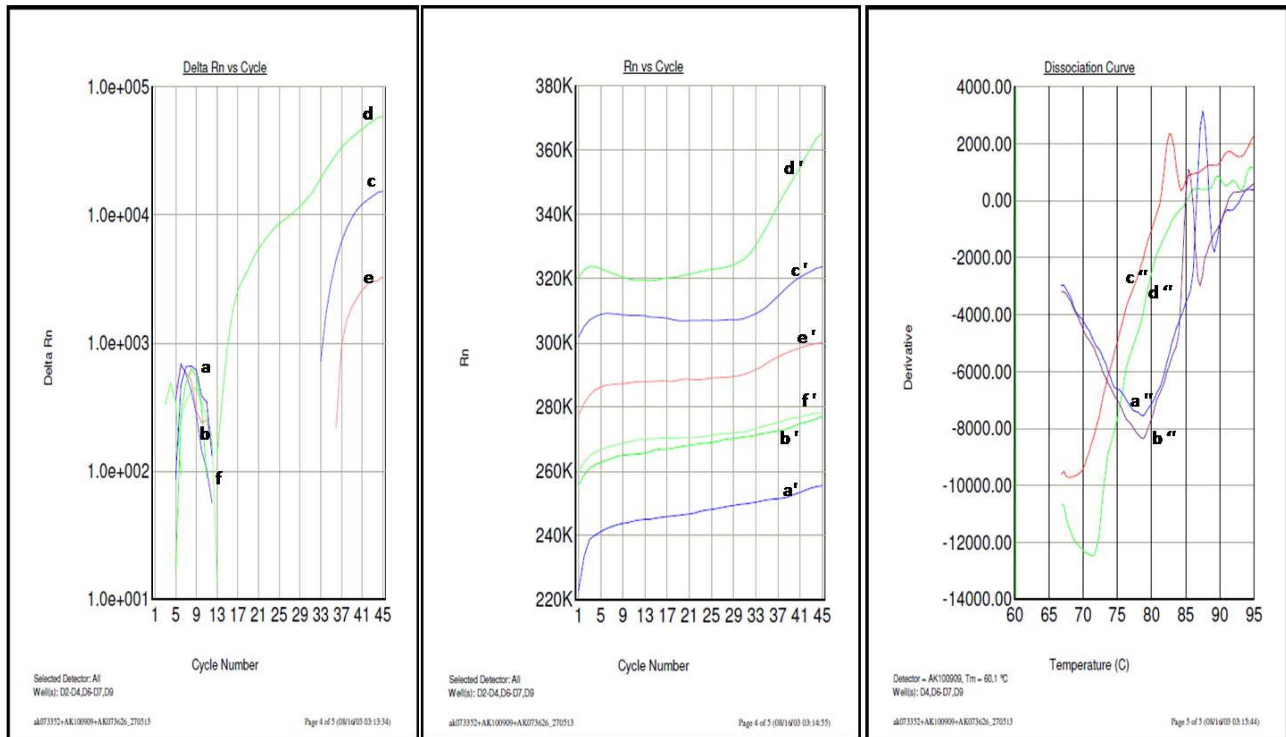


Figure - 34: Molecular Beacon based Real Time PCR result for transketolase gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depicts amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a'&b'* depict amplification of gene with GCC probe. Curves *c'&d'* are the amplification of TCC probe. Curves *e'&f'* are amplification of NTC. C) *a''&b''* depict the dissociation curve of gene with GCC probe and *c''&d''* are dissociation curves of gene with TCC probe.

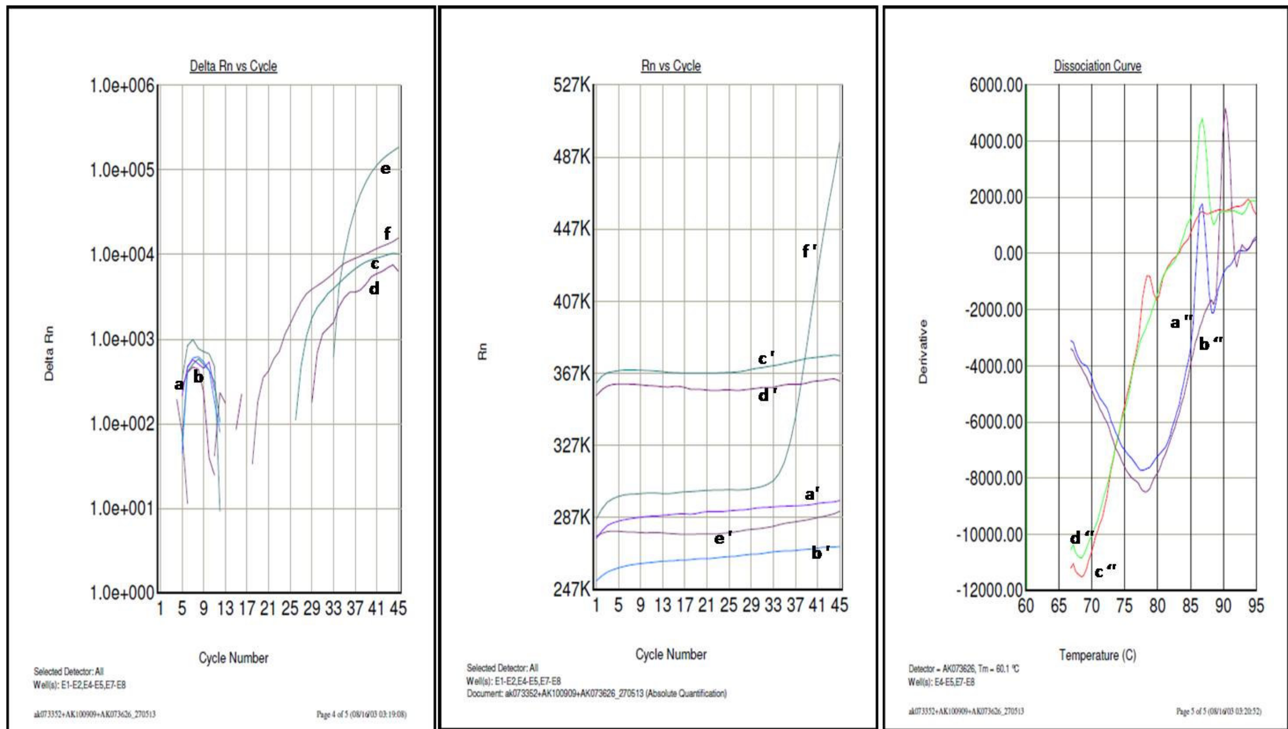


Figure - 35: Molecular Beacon based Real Time PCR result for TENA/THI-4 family protein gene. (A) Graph indicating the relation between Delta Rn vs cycle number. Curves *a&b* depict amplification of gene with GCC probe. Curves *c&d* are the amplification of TCC. Curves *e&f* are amplification of NTC. (B) Graph indicating relationship between Rn vs cycle number. Curves *a′&b′* depict amplification of gene with GCC probe. Curves *c′&d′* are the amplification of TCC probe. Curves *e′&f′* are amplification of NTC. C) *a″&b″* depict the dissociation curve of gene with GCC probe and *c″&d″* are dissociation curves of gene with TCC probe.

Table-10: Ct value chart of DR-DEGs like nitrilase-associated protein gene (AK099444 or LOC_Os11g41150), wall-associated receptor kinase-like 20 precursor gene (AK065517 or LOC_Os09g03620), nucleoside transporter gene (AK102045 or LOC_Os07g37100), sucrose transporter gene (AK100027 or LOC_Os03g07480), phytosulfokines precursor gene (AK073352 or LOC_Os11g05190), transketolase gene (AK100909 or LOC_Os07g09190), TENA/THI-4 family protein gene (AK073626 or LOC_Os03g19390) non template control, NTC and with Template having Molecular Beacon probes specific to GCC box or TCC box. PCR amplification for each gene was performed in two replicates. Ct values and Std dev Ct were obtained from inbuilt Real Time Software.

Gene ID	**Replicates**	**Template**	**Molecular Beacon**	**Ct value**	**Stddev Ct**
AK099444	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	13.51	2.01
	R2	Template	TCC box	16.36	2.01
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-
AK065517	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	13.42	0.02
	R2	Template	TCC box	13.39	0.02
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-
AK102045	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	15.23	3.71
	R2	Template	TCC box	20.49	3.71
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-
AK100027	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	13.33	2.70
	R2	Template	TCC box	9.51	2.70
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-
AK073352	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	12.70	11.18
	R2	Template	TCC box	27.01	11.18
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-
AK100909	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-

	R1	Template	TCC box	32.06	14.20
	R2	Template	TCC box	12.66	14.20
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-
AK073626	R1	Template	GCC box	-	-
	R2	Template	GCC box	-	-
	R1	Template	TCC box	25.31	8.41
	R2	Template	TCC box	28.64	8.41
	R1	No	TCC box	-	-
	R2	No	TCC box	-	-

Gene ontology classification for UR-DEGs: In the previously mentioned UR-DEGs (Table-7) we noticed that ubiquinol-cytochrome C chaperone family protein gene (AK068288 or LOC_Os07g30790) is associated with some biological process like response to freezing (GO: 0050826) and mitochondrion as cellular component (GO: 0005739). Orthologous gene AT5G51220 was found in Arabidopsis having similar putative function. Methyl transferase domain containing protein gene (AK064640 or LOC_Os06g05910) is involved in biological processes of embryonic (GO: 0009790) and post embryonic development (GO: 0009791), ubiquinone biosynthetic process (GO: 0006744) and metabolic process (GO: 0008152), molecular function as transcription regulator activity (GO: 0030528) and cellular function as component of ribosome (GO: 0005840). No orthologous gene and putative function has been reported. 60S ribosomal protein L7 gene (AK058490 or LOC_Os08g42920) is involved in structural molecular activity (GO: 0005198) and translation (GO: 0006412). Orthologous gene AT1G80750 was reported in Arabidopsis indicating putative function of ribosomal protein L30/L7 family protein. Heat shock protein DnaJ gene (AK111076 or LOC_Os06g09560) functions as heat shock protein binding as molecular function (GO: 0031072), and involved a protein metabolic process as biological function (GO: 0019538). Orthologous gene GRMZM2G070475 and Sb10g006350 shows similar putative function in maize and sorghum respectively. Similarly other UR-DEGs (AK063204, AK108801, AK120895, AK067089, AK073072, AK121178, AK063324, AK121619, AK121698, AK059924, AK071086, AK068565, AK102580) mentioned in Table-8 were also used for their Gene Ontology classification and detailed findings has been described in Table-11.

Table 11: Gene Ontology classification for UR-DEGs

UR-DEGs													
S. No.	Accession No.	LOC_ID		FOLD change	Gene ontology								
					Molecular function			Biological function			Cellular function		
					GO ID	GO Name	Hyper P value	GO ID	GO Name	Hyper P value	GO ID	GO Name	Hyper P value
1	AK068288	LOC_Os07g30790		2	GO:0050825	ice binding	0.0117	GO:0050826	response to freezing	0.0197		Genes unmapped by GO annotation	
2	AK058490	LOC_Os08g42920		7	GO:0003735	structural constituent of ribosome	0.173	GO:0006412	translation	0.2403	GO:0005622	intracellular	0.2934
					GO:0030528	transcription regulator activity	0.0929				GO:0005840	ribosome	0.177
											GO:0015934	large ribosomal subunit	0.0119
3	AK063204	LOC_Os07g37280		72	GO:0050825	ice binding	0.0117	GO:0042309	homoiothermy	0.0197		Genes unmapped by GO annotation	
								GO:0050826	response to freezing	0.0197			
4	AK064640	LOC_Os06g05910		15	GO:0008425	2-polyprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity	0.0006	GO:0006744	ubiquinone biosynthetic process	0.0030		Genes unmapped by GO annotation	
								GO:0008152	metabolic process	0.1338			

5	**AK108801**	**LOC_Os11g08940**		13	**GO:0003899**	**DNA-directed RNA polymerase activity**	0.0254	**GO:0006350**	**transcription**	0.2197		**Genes unmapped by GO annotation**	
6	**AK120895**	**LOC_Os12g39520**		16	**Genes unmapped by GO annotation.**			**Genes unmapped by GO annotation.**				**Genes unmapped by GO annotation**	
7	**AK067089**	**LOC_Os03g04140**		8	**GO:0008415**	**acyltransferase activity**	0.0405	**GO:0006535**	**cysteine biosynthetic process from serine**	0.0064	**GO:0005737**	**cytoplasm**	**0.1642**
					GO:0009001	**serine O-acetyltransferase activity**	0.0018	**GO:0008652**	**cellular amino acid biosynthetic process**	0.0201			
					GO:0016740	**transferase activity**	0.0621						
8	**AK073072**	**LOC_Os09g09650**		6	**Genes unmapped by GO annotation**			**Genes unmapped by GO annotation**				**Genes unmapped by GO annotation**	
9	**AK121178**	**LOC_Os04g02310**		17	**GO:0003676**	**nucleic acid binding**	0.2815	**Genes unmapped by GO annotation**				**Genes unmapped by GO annotation**	
10	**AK063324**	**LOC_Os06g11720**		11	**GO:0016740**	**transferase activity**	0.0621	**GO:0008152**	**metabolic process**	0.1338		**Genes unmapped by GO annotation**	
					GO:0016757	**transferase activity, transferring glycosyl groups**	0.1024						

					GO:0016758	transferase activity, transferring hexosyl groups	0.0723						
11	AK111076	LOC_Os06g09560		46	GO:0031072	heat shock protein binding	0.0458	Genes unmapped by GO annotation				Genes unmapped by GO annotation	
12	AK121619	LOC_Os06g40040		3	GO:0003676	nucleic acid binding	0.2815	Genes unmapped by GO annotation				Genes unmapped by GO annotation	
13	AK121698	LOC_Os10g42150		6	Genes unmapped by GO annotation			Genes unmapped by GO annotation				Genes unmapped by GO annotation	
14	AK059924	LOC_Os12g43100		7	GO:0016491	oxidoreductase activity	0.3113	GO:0008612	peptidyl-lysine modification to hypusine	0.0038		Genes unmapped by GO annotation	
					GO:0019135	deoxyhypusinemonooxygenase activity	0.0006	GO:0055114	oxidation reduction	0.2955			
					GO:0005506	iron ion binding	0.2048						
					GO:0005488	binding	0.2696						
					GO:0004497	monooxygenase activity	0.1218						
					GO:0046872	metal ion binding	0.3303						
15	AK071086	LOC_Os12g23780		17	Genes unmapped by GO annotation			Genes unmapped by GO annotation				Genes unmapped by GO annotation	
16	Ak068565	LOC_Os01g34060		6	GO:0050825	ice binding	0.0117	GO:0050826	response to freezing	0.0197	GO:0005634	nucleus	0.4078
17	AK102580	LOC_Os03g10460		6	Genes unmapped by GO annotation			Genes unmapped by GO annotation				Genes unmapped by GO annotation	

Gene ontology classification for DR-DEGs: Gene Ontology classification was also performed for DR-DEGs. In DR-DEGs (Table-9) we found that rhoGAP domain containing protein gene (AK067300 or LOC_Os12g05900) is involved in catabolic activity (GO: 0009056) and nucleo base, nucleoside, nucleotide and nucleic acid metabolic process (GO: 0006139), signal transduction as biological process (GO: 0007165). Orthologous gene AT5G22400 was noticed in Arabidopsis showing putative function of Rho GTPase activating protein with PAK-box/P21-Rho-binding domain. DnaK family protein gene (AK100997 or LOC_Os02g48110) indicated molecular function in nucleotide binding (GO: 0000166) and ATP binding (GO: 0005524), plastid as a cellular component (GO: 0009536). Orthologous gene AT4G16660 in Arabidopsis having putative function similar to heat shock protein 70 was noticed. CPuORF11 - conserved peptide uORF-containing transcript gene (AK103103 or LOC_Os02g01240) is involved in cellular activity (GO: 0009987) and mitochondrion as a cellular component (GO: 0005739). Orthologous gene AT5G07840 in Arabidopsis was reported that show putative function as Ankyrin repeat family protein. OsFBX61 - F-box domain containing protein gene (AK103417 or LOC_Os02g52130) is involved in molecular activity as GTP binding (GO: 0005525) and GTPase activity (GO: 0003924), protein binding (GO: 0005515) and biological function like response to endogenous stimulus (GO: 0009719), protein modification process (GO: 0006464), catabolic process (GO: 0009056). F-BOX WITH WD-40 2 found as putative function in Arabidopsis as Orthologous gene (AT4G08980). CGMC_MAPKCMGC_2_SLT2y_ERK.2 - CGMC includes CDA, MAPK, GSK3, and CLKC kinases, gene (AK071376 or LOC_Os06g48590) have molecular activity involved in signal transducer activity (GO:0004871), transferase activity (GO:0016740), kinase activity (GO:0016301), ATP binding (GO:0005524), biological function as protein amino acid phosphorylation (GO:0006468) and response to stress (GO:0006950) whereas no cellular activity was found. In Arabidopsis orthologous gene (AT1G10210) found and functioned as mitogen-activated protein kinase 1. Helix-loop-helix DNA-binding domain containing protein gene (AK068704 or LOC_Os03g26210) showed molecular activity as sequence-specific DNA binding transcription factor activity (GO: 0003700) and DNA binding (GO: 0003677), biological function as regulation of transcription (GO: 0045449), nucleobase, nucleoside, nucleotide and nucleic acid metabolic process (GO: 0006139), and as cellular component in nucleus (GO: 0005634) and plastid (GO: 0009536). Orthologous gene (AT3G47640) in Arabidopsis putatively functioned as basic helix-loop-helix (bHLH) DNA-binding superfamily protein. And Spermidine synthase gene

(AK065153 or LOC_Os02g15550) involved as catalytic activity (GO: 0003824), transferase activity (GO: 0016740) in molecular function, biosynthetic process (GO: 0009058), metabolic process (GO: 0008152) in biological processes, also involved in cytoplasm (GO: 0005737) as cellular component. Orthologous gene (AT5G53120) closely found in Arabidopsis have similar function as spermidine synthase 3. Similarly other DR-DEGs (AK069854, AK065517, AK099444, AK102045, AK100027, AK073352, AK100909, AK073626) mentioned in Table-10 were also used for their Gene Ontology classification and detailed findings has been listed in Table-12.

Table 12: Gene Ontology classification for DR-DEGs

DR-DEGs													
S. No.	Accession No	LOC_ID		FOLD	Gene ontology								
					Molecular function			Biological function			Cellular function		
					GO ID	GO Name	Hyper P Value	GO ID	GO Name	Hyper P Value	GO ID	GO Name	Hyper P Value
1	AK103103	LOC_Os02g01240		-3		Genes unmapped by GO annotation			Genes unmapped by GO annotation			Genes unmapped by GO annotation.	
2	AK103417	LOC_Os02g52130		-2	GO:0003924	GTPase activity	0.055		Genes unmapped by GO annotation			Genes unmapped by GO annotation	
					GO:0005525	GTP binding	0.1078						
3	AK069854	LOC_Os03g48970		-3	GO:0003677	DNA binding	0.2719				GO:0005634	nucleus	0.2932
											GO:0016602	CCAAT-binding factor complex	0.0030
4	AK100997	LOC_Os02g48110		-3	GO:0000166	nucleotide binding	0.0666		Genes unmapped by GO annotation		GO:0005783	endoplasmic reticulum	0.0616
					GO:0005524	ATP binding	0.1134						
5	AK065517	LOC_Os09g03620		-2	GO:0016301	kinase activity	0.0730	GO:0006468	protein amino acid phosphorylation	0.1190		Genes unmapped by GO annotation.	
					GO:0005524	ATP binding	0.1134						

					GO:0004872	receptor activity	0.1629						
					GO:0004713	protein tyrosine kinase activity	0.0996						
					GO:0004674	protein serine/threonine kinase activity	0.1030						
					GO:0004672	protein kinase activity	0.1053						
					GO:0000166	nucleotide binding	0.0666						
6	AK099444	LOC_Os11g41150		-5	GO:0050825	ice binding	0.3470	GO:0042309	homoiothermy	0.3585		Genes unmapped by GO annotation	
								GO:0050826	response to freezing	0.3585			
7	AK102045	LOC_Os07g37100		-2	GO:0005337	nucleoside transmembrane transporter activity	0.0036	GO:0017004	cytochrome complex assembly	0.0198	GO:0016020	membrane	0.3775
					GO:0015232	heme transporter activity	0.0132	GO:0015886	heme transport	0.0140			
								GO:0006810	transport	0.2816			
8	AK071376	LOC_Os06g48590		-2	GO:0016740	transferase activity	0.0962	GO:0006468	protein amino acid phosphorylation	0.1190		Genes unmapped by GO annotation.	
					GO:0016301	kinase activity	0.0730	GO:0006950	response to stress	0.0807			
					GO:0005524	ATP binding	0.1134						
					GO:0004872	receptor activity	0.1629						

					GO:0004713	protein tyrosine kinase activity	0.0996						
					GO:0004674	protein serine/threonine kinase activity	0.1030						
					GO:0004672	protein kinase activity	0.1053						
					GO:0000166	nucleotide binding	0.0666						
9	AK100027	LOC_Os03g07480		-23	GO:0008515	sucrose transmembrane transporter activity	0.0020	GO:0015770	sucrose transport	0.0021	GO:0005887	integral to plasma membrane	0.0040
10	AK067300	LOC_Os12g05900		-2		Genes unmapped by GO annotation.		GO:0007165	signal transduction	0.0793	GO:0005622	intracellular	0.3735
11	AK073352	LOC_Os11g05190		-10	GO:0008083	growth factor activity	0.0056	GO:0007275	multicellular organismal development	0.0421	GO:0005576	extracellular region	0.1890
								GO:0008283	cell proliferation	0.0043			
								GO:0030154	cell differentiation	0.0106			
12	AK068704	LOC_Os03g26210		-3	GO:0030528	transcription regulator activity	0.1197	GO:0045449	regulation of transcription	0.3373	GO:0005634	nucleus	0.2932

					GO:0003677	DNA binding	0.2719	GO:0045449	regulation of transcription				
13	AK065153	LOC_Os02g15550		-2	GO:0003824	catalytic activity	0.1582		Genes unmapped by GO annotation			Genes unmapped by GO annotation.	
					GO:0016740	transferase activity	0.0962						
14	AK100909	LOC_Os07g09190		-9	GO:0008661	1-deoxy-D-xylulose-5-phosphate synthase activity	0.0012	GO:0008152	metabolic process	0.3739	GO:0009536	plastid	0.2077
					GO:0003824	catalytic activity	0.1582	GO:0016114	terpenoid biosynthetic process	0.0043			
15	AK073626	LOC_Os03g19390		-29		Genes unmapped by GO annotation.			Genes unmapped by GO annotation			Genes unmapped by GO annotation.	

Classification of anoxia responsive TFs families: Plant genomes consist of considerable percentage of Transcription Factor Genes. Earlier we have shortlisted the UR-DEGs and DR-DEGs (from the publically available microarray data of Lasanthi-Kudahettige et al. 2007) and used for finding the consensus promoter motifs (Kumar et al. 2009). In this study we extended our work and tried to find the list of all TFs which belongs either UR-DEGs or DR-DEGs. Since structure of DNA-*binding domains* varies widely across TFs and TFs have classified on the basis of domain/binding site. Therefore, obtained TFs were further checked about domain/binding sites. Depending upon the type of domain/binding site these TFs were classified in various TFs families. These TFs belonging to particular families were again checked for their existence in Plant Transcription Factor Database. TFs listed in Plant Transcription Factor Database were only selected. A list of TFs families having varying number of TFs that were up-regulated by a factor of different fold and belonging to UR-DEGs are enlisted in Table-13 and Figure-36. On the other hand list of TFs families having varying number of TFs that were down-regulated by a factor of different fold and belonging to DR-DEGs are enlisted in Table-13 and Figure-37. It was found that AP2-EREBP TFs family members belonging to UR-DEGs show maximum up-regulation (2 to 82 fold) in the microarray data reported by Lasanthi-Kudahettige et al. (2007). On the other hand these AP2-EREBP TFs family members that also belong to DR-DEGs showed maximum expression (by -2 to -134 fold) in anoxic rice coleoptiles.

Table 13. Result showing the distribution of TFs family members in UR-DEGs and DR-DEGs and their fold expression in anoxic coleoptiles.

http://plntfdb.bio.uni-potsdam.de/v3.0/						
TF gene found in UR-DEGs				**TF gene found in DR-DEGs**		
TF family name	**Total No**	**Fold(nX)**		**TF family name**	**Total No**	**Fold(nX)**
ABI3VP1	5	2 to 10		ABI3VP1	1	-5
Alfin-like	7	2		Alfin-like	2	-2
AP2-EREBP	24	2 to 82		AP2-EREBP	29	(-134) to (-2)
ARF	5	2 to 6		ARF	12	(-7) to (-2)
BR/BPC	1	2		ARR-B	2	-2
bHLH	7	2 to 18		BES1	2	-2
BSD	1	2		bHLH	27	(-27) to (-2)
bZIP	6	3 to 16		BSD	2	-2
C2C2-CO-like	2	3 to 5		bZIP	8	(-7) to (-2)
C2C2-GATA	5	2 to 5		C2C2-CO-like	3	(-7) to (-3)
C2C2-YABBY	3	2 to 3		C2C2-Dof	3	(-12) to (-2)
C2H2 gene	8	2 to 5		C2C2-GATA	3	(-3) to (-2)
C3H gene	11	2 to 21		C2C2-YABBY	2	(-4) to (-2)
CCAAT	8	2 to 5		C2H2 gene	8	(-19) to (-2)
CPP	6	2 to 5		C3H	4	9-40) to (-2)
DBP	2	2		CCAAT	9	(-3)
E2F DP	3	2 to 3		CPP	4	-2
FAR1	10	2 to 3		DBP	2	(-5) to (-3)
FHA	4	2 to 4		EIL	3	-2
GRAS	6	2 to 4		FAR1	2	-2
GRF	4	2 to 5		G2-like	6	(-9) to (-2)
HB	8	2 to 8		GRAS	4	(-4) to (-2)
HSF	9	2 to 15		HB	5	(-7) to (-2)
MADS	3	2		HRT	1	-2
mTERF	11	2 to 5		HSF	3	(-7) to (-2)
MYB	19	2 to 18		LOB	4	(-7) to (-2)
MYB-related	2	2 to 18		MADS	2	(-7) to (-2)
NAC	6	2 to 4		MYB	17	(-65) to (-2)
Orphans	6	2 to 3		MYB-related	4	(-6) to (-2)
PLATZ	3	3		NAC	18	(-33) to (-2)
RWP-RK	2	2 to 8		Orphans	5	(_7) to (-2)
S1Fa-like	1	2		PLATZ	2	(-10) to (-2)
TAZ	4	2 to 3		RWP-RK	2	-2
Trihelix	3	3		SBP	2	(-3) to (-2)
VOZ family	1	3		TCP	2	-2
WRKY	7	2 to 12		Tify	11	(-64) to (-3)

				TIG	5	(-3) to (-2)
				Trihelix	6	(-3) to (-2)
				ULT	1	-2
				WRKY	19	(-92) to (-2)
				zf-HD	4	(-28) to (-2)

Figure 36. Result showing the distribution of TFs family member in UR-DEGs in anoxic coleoptiles.

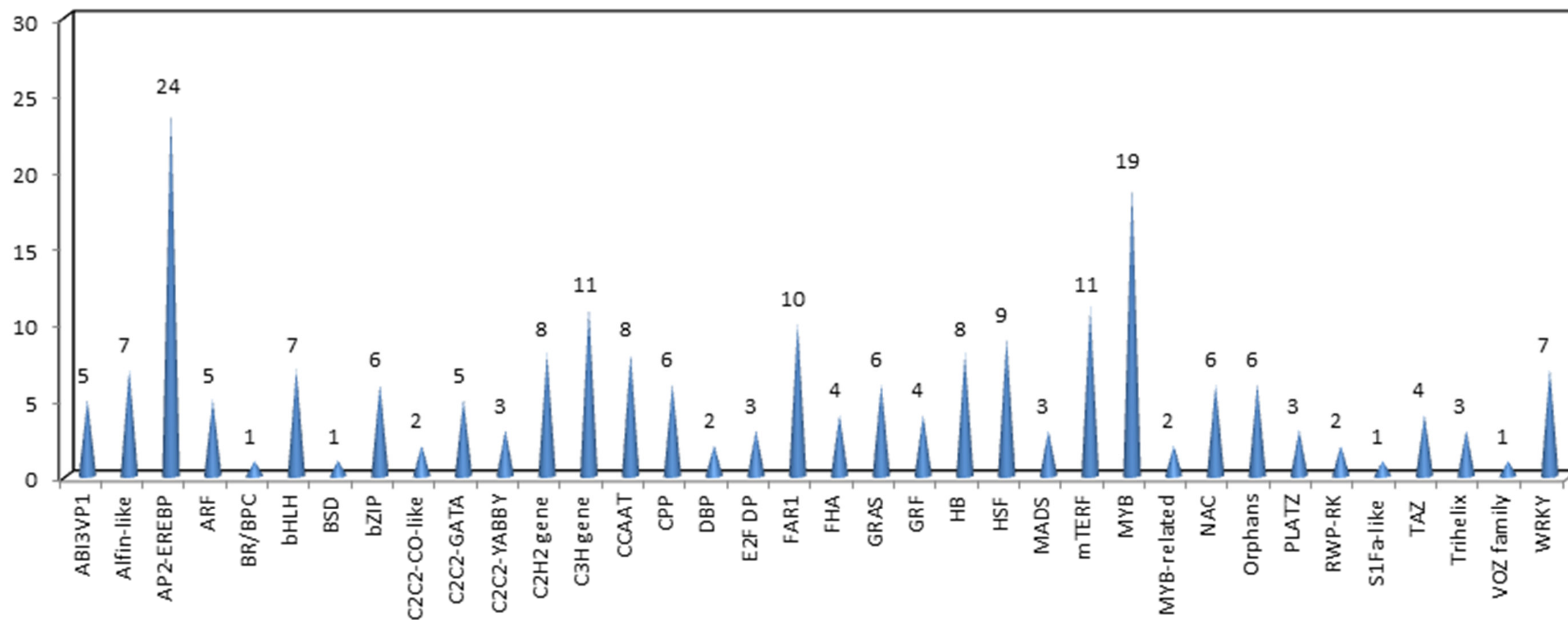
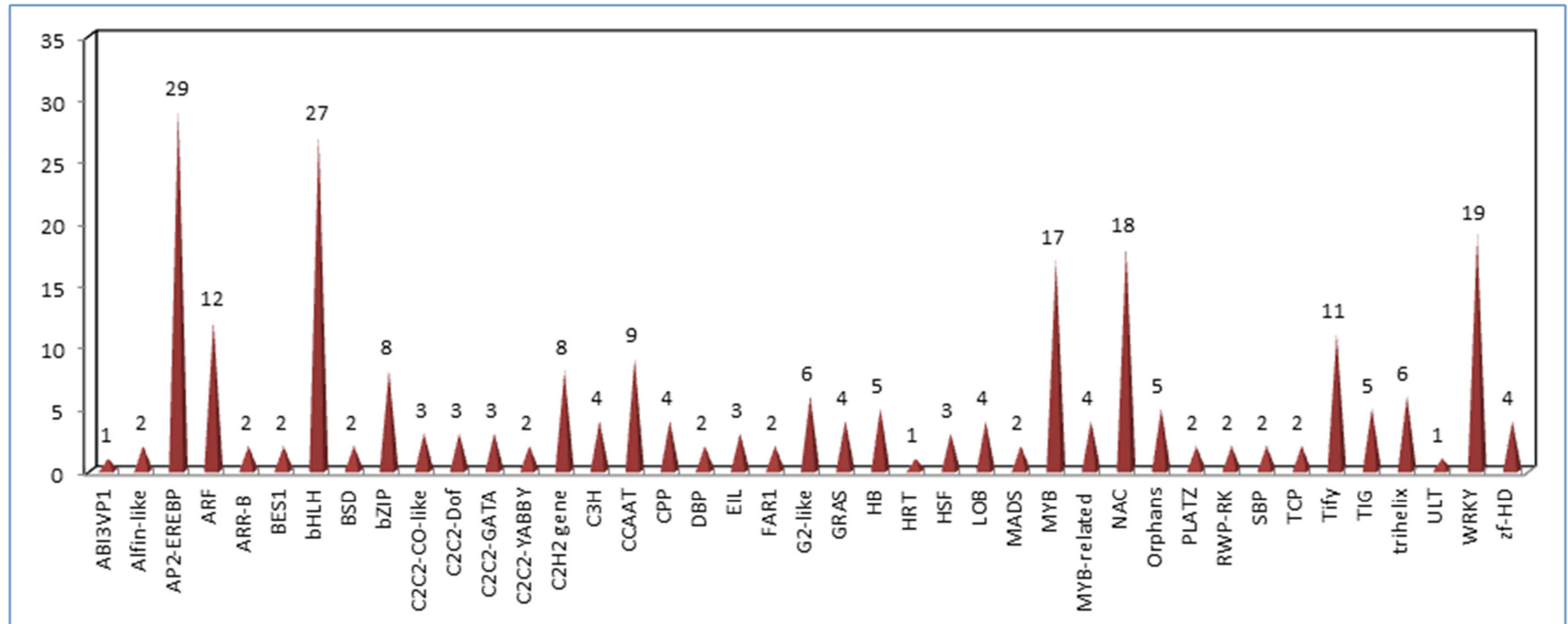


Figure 37. Result showing the distribution of TFs family member in DR-DEGs in anoxic coleoptiles.



***3D Structure prediction of Sub1A and TFs proteins*:** The 3D structure of Sub1A, CPuORF2 - conserved peptide uORF-containing transcript gene protein (LOC_Os09g13570) and B3 DNA binding domain containing gene protein (LOC_Os03g06850), bZIP transcription factor gene protein (LOC_Os02g52780) were not available in PDB database. Therefore their structures were predicted using I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). It is well known that *Sub1A* gene was responsible for the submergence tolerance in rice (Xu et al. 2006). And to predict the 3D structure protein sequences were required. The protein sequences were retrieved from the TIGR (v6.1) (ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.1/). I-TASSER server predicts and displays various features in different sections for best model studies. It was considered that the prediction and generation of the best model based on C-Score, their structural analogs and binding sites. The quality of the generated models are estimated based on a confidence score (C-score), ranges from -5 to 2 where a high value signifies a model with a high confidence and vice-versa. C-score is highly correlated with Tm score and RMSD. Therefore, TM-score and RMSD are known standards to measure the accuracy of structure modeling thereby measuring structural similarity between two protein structures. RMSD is an average distance of all residue pairs in two structures and is sensitive to local errors (i.e., a mis-orientation of the tail) which occurs in spite of the correct global topology hence, TM-score must be used for solving these errors A TM-score >0.5 indicates a model of correct topology. Roy et al. (2012) predicted the structures of three human GPCRs complexes using I-TASSER with a RMSD's 1.6A°, 2.27A° and 2.82A° to the crystal structures in the Trans membrane region. The models predicted by I-TASSER were based on the best 10 threading templates available on RCSB PDB. The best predicted models were selected on the basis of confidence score; TM-Score as well as RMSD value (Table -14). The C score value for the best predicted model which is model 1 of Sub1A was -4.06 and furthermore, highly similar structures in PDB (as identified by TM-align) were identified and listed in Table-15. Structure of Sub1A, CPuORF2 - conserved peptide uORF-containing transcript gene protein, B3 DNA binding domain containing gene protein and bZIP transcription factor gene proteins were predicted by I-TASSER and visualized by VMD tool has been shown in Figure-38, 39, 40, 41, respectively.

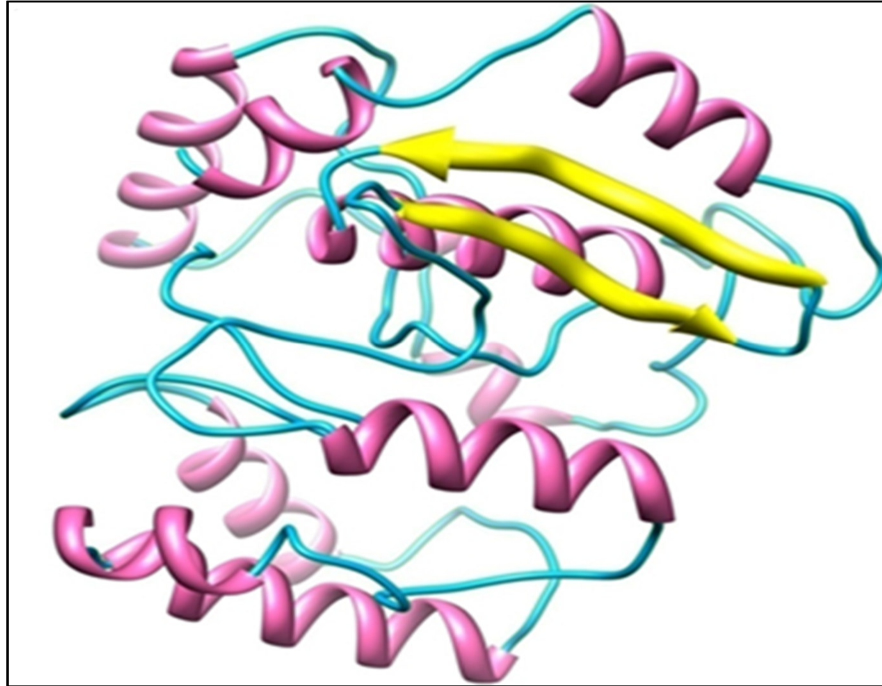


Figure 38. Result showing 3D structure of Sub1A protein predicted by I-TASSER. The coloring method is based on secondary structure. The pink color represents α-helix and yellow color represents β-strand and deep sky blue color represents the coil in the 3D structure.

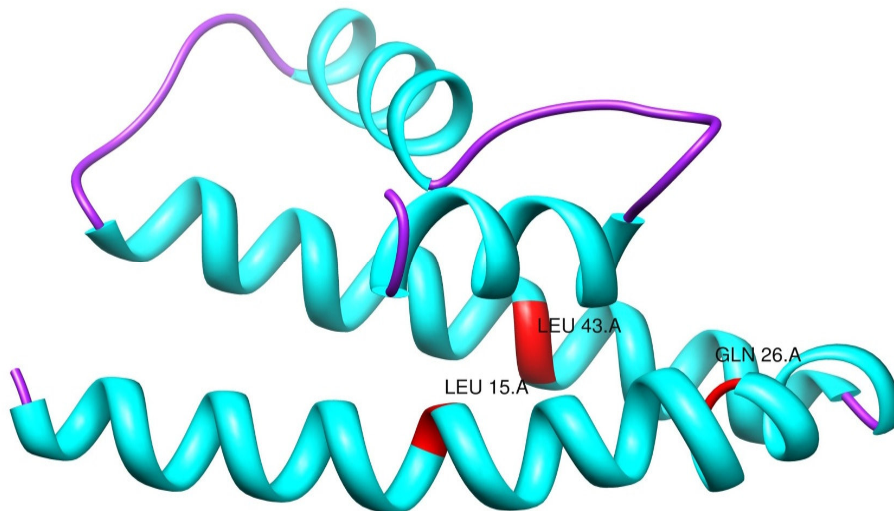


Figure 39. Result showing 3D structure of TF CPuORF2 (LOC_Os09g13570) predicted by I-TASSER. The coloring method is based on secondary structure. The red color represents residue LEU at 15, GLN at 26 and LEU at 43 positions in 3D structure.

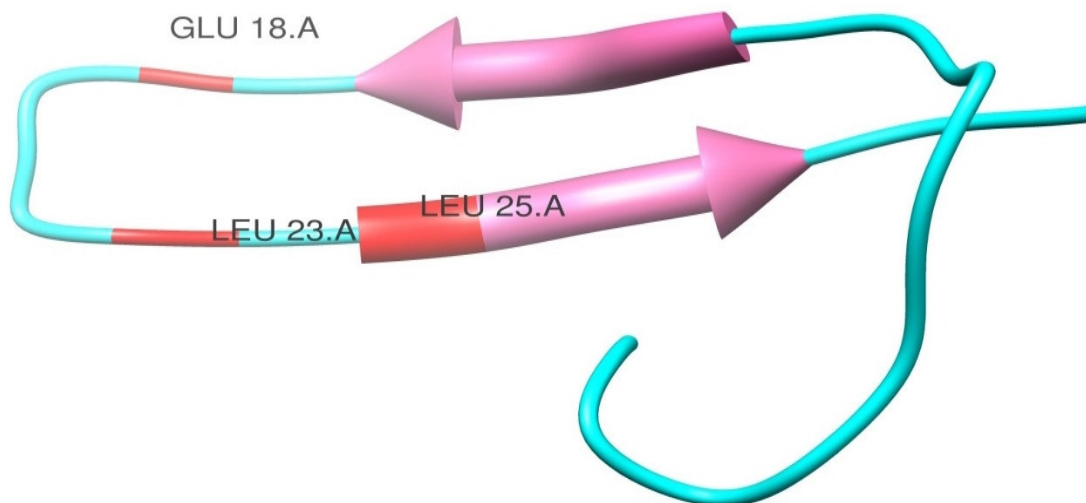


Figure 40: Result showing 3D structure of B3 DNA binding domain containing protein, predicted by I-TASSER. The coloring method is based on secondary structure. The red color represents residue GLU at 18, LEU at 23 and 25 positions in 3D structure.

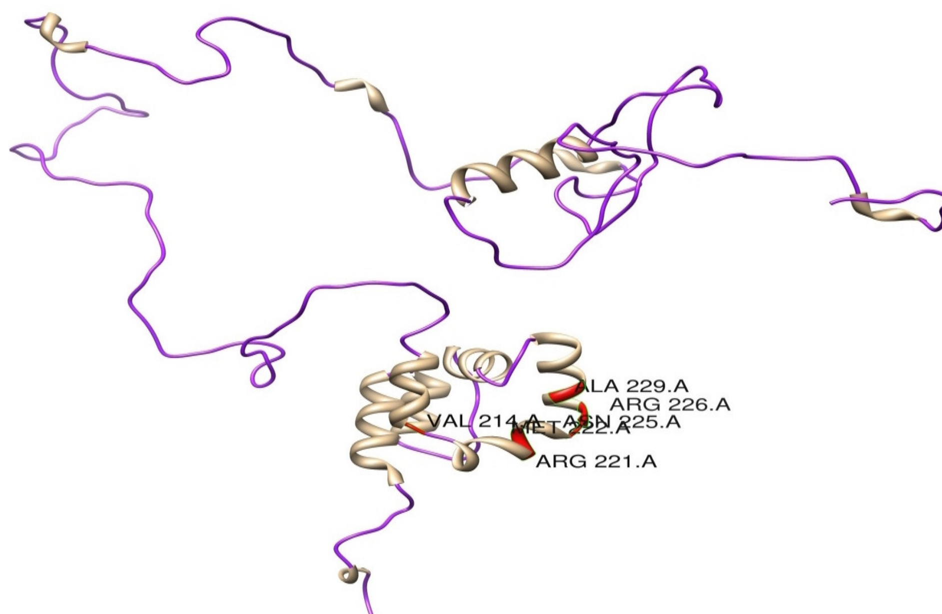


Figure 41: Result showing 3D structure of bZIP transcription factor protein predicted by I-TASSER. The coloring method is based on secondary structure. The red color represents residue VAL at 214 and ARG at 221, MET at 222, ASN at 225, ARG at 226 and ALA at 229 positions in 3D structure.

Table-14. Best predicted model with their C-Score, TM Score and RMSD value where C-Score is the confidence score for the predicted model, TM-score is a measure of global structural similarity between query and template protein and Root Mean Square Deviation is the RMSD between residues that are structurally aligned by TM-align.

Best Predicted Model					
Genes	**Best Model**	**Locus Id**	**C Score**	**TM score**	**RMSD value (Å)**
Sub1A	**Model 1**	Os09g11480	-4.06	0.28±0.09	15.8±3.2Å
UR-DEG	**CPuORF2 Model 1**	Os09g13570	-2.64	041.0±0.14	9.5 ± 4.6Å
	B3 Model 1	Os03g06850	-0.85	0.61±0.14	3.3 ± 2.3
DR-DEG	**bZIP Model 1**	Os02g52780	-3.53	0.33±0.11	14.8 ± 3.6

Table-15. Identified best two structural analogs of TFs in PDB where coverage represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein. Coverage represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Top 2 Identified structural analogs in PDB							
DEGs	**TF**	**LOC Id**	**PDB Hit**	**TM- Score**	**RMSD (Å)**	**IDEN[a]**	**Cov.**
Sub1A	**Model 1**	Os09g11480	4he8L	0.443	5.67	0.038	0.72
			3rkoL	0.441	5.75	0.046	0.724
UR-DEG	CPuORF2	Os09g13570	2htnG	0.688	2.7	0.07	0.914
			1dcnB	0.7	2.54	0.034	0.925
	B3 DNA Binding	Os03g06850	3rOgC	0.617	1.14	0.129	0.968
			3IiuA2	0.592	1.85	0.065	1
DR-DEGs	bZIP	Os02g52780	3einA	0.727	4.33	0.038	0.943
			3mx3A	0.684	4.48	0.066	0.923

Template proteins with similar binding sites for Sub1A are listed in Table -16. The best binding site is predicted on the basis of C score ^{LB} (Range = 0-1) and BS-Score (>1) values. A higher score C score indicates a more reliable ligand-binding site prediction and BS-score reflects a significant local match between the predicted and template binding site (Zhang, 2008, Roy et al. 2010). Qin and Zhou, (2011) suggested that binding site prediction is a useful tool for building structural models for protein-DNA complexes and for experimental design and validation. Two best predicted binding sites for Sub1A 3D was taken for further interaction studies (Table-16).

Table 16. Template proteins for similar binding sites of UR-DEGs and DR-DEG TFs. Binding sites represent the amino acid positions.

Template Protein with similar binding site										
DEGs	**TF**	**LOC ID**	**Cscore[LB]**	**PDB Hit**	**TM-score**	**RMSD[a]**	**IDEN[a]**	**Coverage**	**BS-score**	**Binding Site**
Sub1A	**Model 1**	Os09g11480	**0.31**	1gccA	0.236	1.5	0.629	0.248	1.67	102, 103, 104, 106, 108, 116, 118, 120, 141
			0.26	1gccA	0.236	1.5	0.629	0.248	1.76	106, 108, 110, 112, 118, 125, 127, 129, 130
UR-DEG	CPuORF2	Os09g13570	0.29	3aablP	0.733	2.47	0.056	0.957	0.61	15,18,19,46,47,51
			0.11	3asnC	0.733	2.48	0.056	0.0957	0.66	15,26,43
	B3	Os03g06850	0.05	3e6zX	0.322	2.71	0.214	0.839	0.68	18,23,25
			0.01	3adiA	0.555	2.81	0.069	0.935	0.49	15,19,23,24
DR-DEG	bZIP	Os02g52780	0.01	1e6j1	0.182	5.43	0.049	0.246	1.1	224, 225,226, 229,239, 242, 246
			0.01	1dh3A	0.077	1.6	0.019	0.081	1.09	214,221,222,225, 226,229

For *in-silico* protein-DNA interaction studies 3D structure of protein along with DNA was required. Consequently, 3D structure of DNA segment (25 nt long) containing identified and validated GCC Box promoter motif of Ubiquinol Cytochrome C Chaperone gene (LOC_Os07g30790) and methyltransferase domain containing protein gene (LOC_ Os06g05910) were generated by 3D-DART server. GCC-Box promoter motif positioned at 10-15 nucleotides have been shown in Figure-42 and 43. Similarly for rhoGAP domain containing protein gene (LOC_Os12g05900) TCC box containing 25 nucleotides were also determined. Construction of 3D DNA structure of promoter region having GCC-box motif in was performed. Similarly 3D DNA structure in DR-DEGs like having TCC-box was performed as described by Pandey and Kumar, (2013) Figure-44. To study protein-DNA interaction a 3D model of DNA fragment (promoter region of 25 bases having core GCCGCC or TCCTCC box motif) was required. Therefore, 3D-DART (3DNA-Driven DNA Analysis and Rebuilding Tool) server was used.

Figure 42. 3D structure of linear DNA segment of Ubiquinol Cytochrome C Chaperone gene generated by 3D-DART server of GCC-Box promoter motif positioned at 8-17 nucleotide. In DNA model red color represents Adenine and pink color represents Guanine and sea green color represents cytosine and gold yellow color represents Thymine.

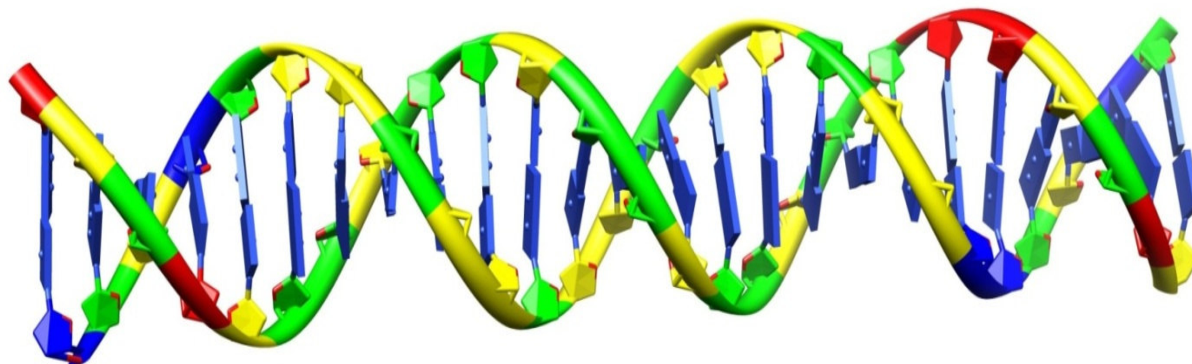


Figure 43: 3D structure of linear DNA segment of methyltransferase domain containing protein gene generated by 3D-DART server of GCC-Box promoter motif positioned at 9- 18 nucleotide. In DNA model red color represents Adenine and green color represents Guanine and yellow color represents cytosine and blue color represents Thymine.

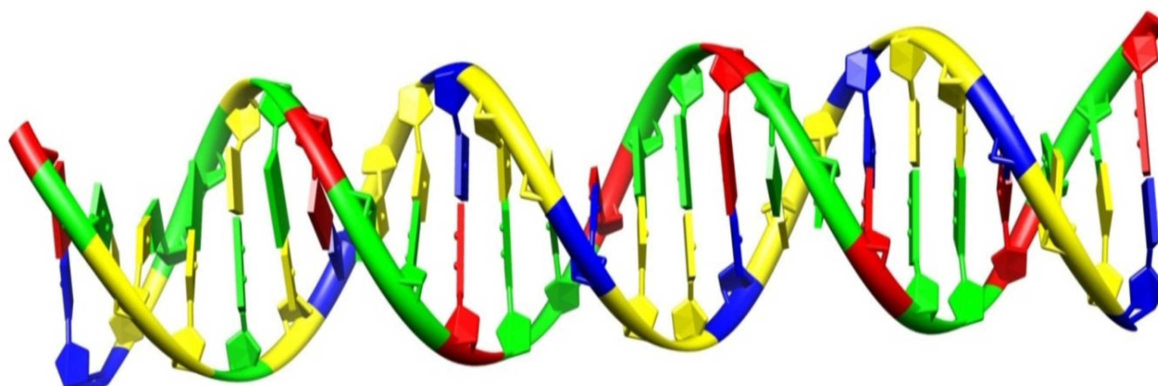


Figure 44: 3D structure of linear DNA segment of rhoGAP domain containing protein gene generated by 3D-DART server of TCC-Box promoter motif positioned at 6-20 nucleotide. In DNA model red color represents Adenine and green color represents Guanine and yellow color represents cytosine and blue color represents Thymine.

Protein-DNA interactions are the physical basis of gene expression and DNA modification for vital biological activities (Qin and Zhou, 2011). Because there is no simple mapping code between DNA base pairs and protein amino acids, the prediction of protein-DNA interactions is a challenging problem. Therefore, HADDOCK can make use of a broader array of restraints, including those derived from biochemical and biophysical data (Kobayashi et al. 2010). Determining the structure of protein-DNA complexes and clarifying the factors that regulating their interaction is essential to better understand many biological processes (Chou et al. 2010). A review describing the experimental strategies currently employed to solve structures of protein–DNA complexes and to analyze their dynamics has been published (Campagne et al. 2011). Protein–DNA interactions facilitate the fundamental functions of living cells and are universal in all living organisms (Sathyapriya and Vishveshwara, 2004).

To determine the protein-DNA interactions the Easy interface of HADDOCK web server was used (de Vries et al. 2010). Before going for docking AIR files were generated for both the interacting molecules having information about the active binding sites of various proteins as well as in the DNA model. The HADDOCK score is the weighted sum of van der Waals energy (negative indicating favorable interactions), electrostatic energy (negative indicating favorable interactions), distance restraints energy (only unambiguous and AIR (ambig-restraints), direct RDC restraint energy, inter vector projection angle restraints energy, diffusion anisotropy energy, dihedral angle restraints energy, symmetry restraints energy, buried surface area (negative weight indicate a better interface), binding energy, desolvation energy. Meanwhile, the solution structures are analyzed for their intermolecular hydrogen bonds and intermolecular hydrophobic contacts by HADDOCK, the solutions are clustered according to the interface ligand RMSDs. The Z-score indicates the standard deviations from the average of a particular cluster in terms of HADDOCK score. For the prediction of best interaction, different binding sites and models of four proteins were docked with 3D structure of DNA segment having promoter motif of selected genes (Table-16). Similarly protein-DNA interactions have also been shown in Figure-45, 46, 47.

Table 16. Protein-DNA interaction between Sub1A with promoter motif of *UCCC* gene, up-regulated CPuORF2 and B3 DNA BindingTFs with GCC-Box in the promoter of methyltransferase domain containing protein gene (Os06g05910) and down-regulated TF with TCC-Box of rhoGAP domain containing protein gene (Os12g05900) by HADDOCK server.

DNA	Protein	Model	HADDOCK SCORE	RMSD Value	Van der Waals Energy	Electrostatic Energy	Desolvation Energy	Restraints Violation Energy	Buried Surface Area	Z - Score
UCCC gene	**Sub1A** (LOC_Os09 g11480)	SAUGCM1-BS1	33.1 +/- 3.8	8.5 +/- 0.6	-55.0 +/- 7.3	-640.1 +/- 63.6	44.1 +/- 10.6	1719.7 +/-29.43	1645.3 +/- 95.6	-1.2
		SAUGCM1-BS2	12.4 +/- 5.7	16.6 +/- 0.6	-61.9 +/- 3.3	-613.6 +/- 30.2	32.0 +/- 6.8	650.2 +/- 19.23	1912.0+/- 117.0	-1.2
		SAUGCM2-BS1	27.8 +/- 3.6	8.7 +/- 0.2	-60.5 +/- 5.9	-574.1 +/- 24.4	34.4 +/- 2.5	1687.2 +/- 9.37	1741.7+/- 115.4	-1.2
		SAUGCM2-BS2	11.7 +/- 4.4	0.9 +/- 0.5	-56.2 +/- 3.0	-597.9 +/- 63.0	18.5 +/- 5.7	1689.9 +/-39.38	1804.5 +/- 70.5	-1.7
methyltransferase domain containing protein gene (Os06g05910)	CPuORF2 (LOC_Os09 g13570)	ILZMTGCM1-BS1	-62.6+/-8.6	16.2+/-0.3	-43.8+/-6.0	-261+/-14.4	22.1+/-2.9	112.7+/-14.33	1391.3+/-92.5	-2
		LZMTGCM1-BS2	-50.2+/-8.5	19.4+/-0.2	-35.0+/-4.8	-236.7+/-16.2	22.7+/-4.0	94.1+/-30.42	1253.0+/-88.2	-2.4
		ILZMTGCM2-BS1	-74.5+/-4.1	17.2+/-0.2	-45.9+/-1.8	-263+/-11.8	17.9+/-2.9	61.6+/-19.59	1375.8+/-50.7	-1.2
		ILZMTGCM2-BS2	-74.7+/-2.0	16.8+/-0.1	-51.8+/-4.3	-275.4+/-33.5	22.8+/-6.4	94.0+/-31.11	1430.4+/-67.7	-2
	B3 DNA Binding (LOC_Os03g 06850)	IABMTGCM1-BS1	-91+/-4.1	2.5+/-1.5	-43.7+/-7.2	-422.7+/-44.2	28.0+/-2.1	91.8+/-15.42	1299.1+/-141.3	-2.2
		IABMTGCM1-BS2	-80.7+/-9.7	2.0+/-1.4	-42.3+/-11.9	-359.0+/-34.8	27.9+/-4.1	55.6+/-15.47	1250.8+/-123.1	-1.1
		IABMTGCM2-BS1	-82.1+/-3.7	9.5+/-1.7	-37.2+/-2.0	-369.8+/-13.3	20.6+/-4.7	84.6+/-47.77	1197.1+/-84.5	-1.9
		IABMTGCM2-BS2	-84.5+/-6.9	5.0+/-1.4	-39.9+/-5.4	-377.2+/-42.9	21.2+/-3.3	91.6+/-40.46	1222.5+/-117.0	-1.5
rhoGAP domain containing protein gene (Os12g05900)	bZIP (LOC_Os12g 05900)	IBZRHOGTM1-BS1	-126.6+/-5.7	24.5+/-0.1	-82.5+/-1.9	-618.2+/-41.6	65.5+/-9.5	139.9+/-25.77	2082.2+/-139.6	-1.2
		IBZRHOGTM1-BS2	-131.5+/-8.1	4.3+/-1.7	-77.3+/-4.3	-604.6+/-44.0	51.0+/-6.0	157.7+/-35.13	1901.7+/-72.5	-1.8
		IBZRHOGTM2-BS1	-133.7+/-4.2	1.8+/-1.1	-64.4+/-1.2	-645.0+/-16.4	41.2+/-3.4	184.4+/-14.24	1686+/-13.5	-2.8
		IBZRHOGTM2-BS2	-154.1+/-9.7	3.3+/-2.2	-71.7+/-5.9	-691.6+/-25.3	44.0+/-6.1	119.4+/-37.85	1992.5+/-66.8	-1.9

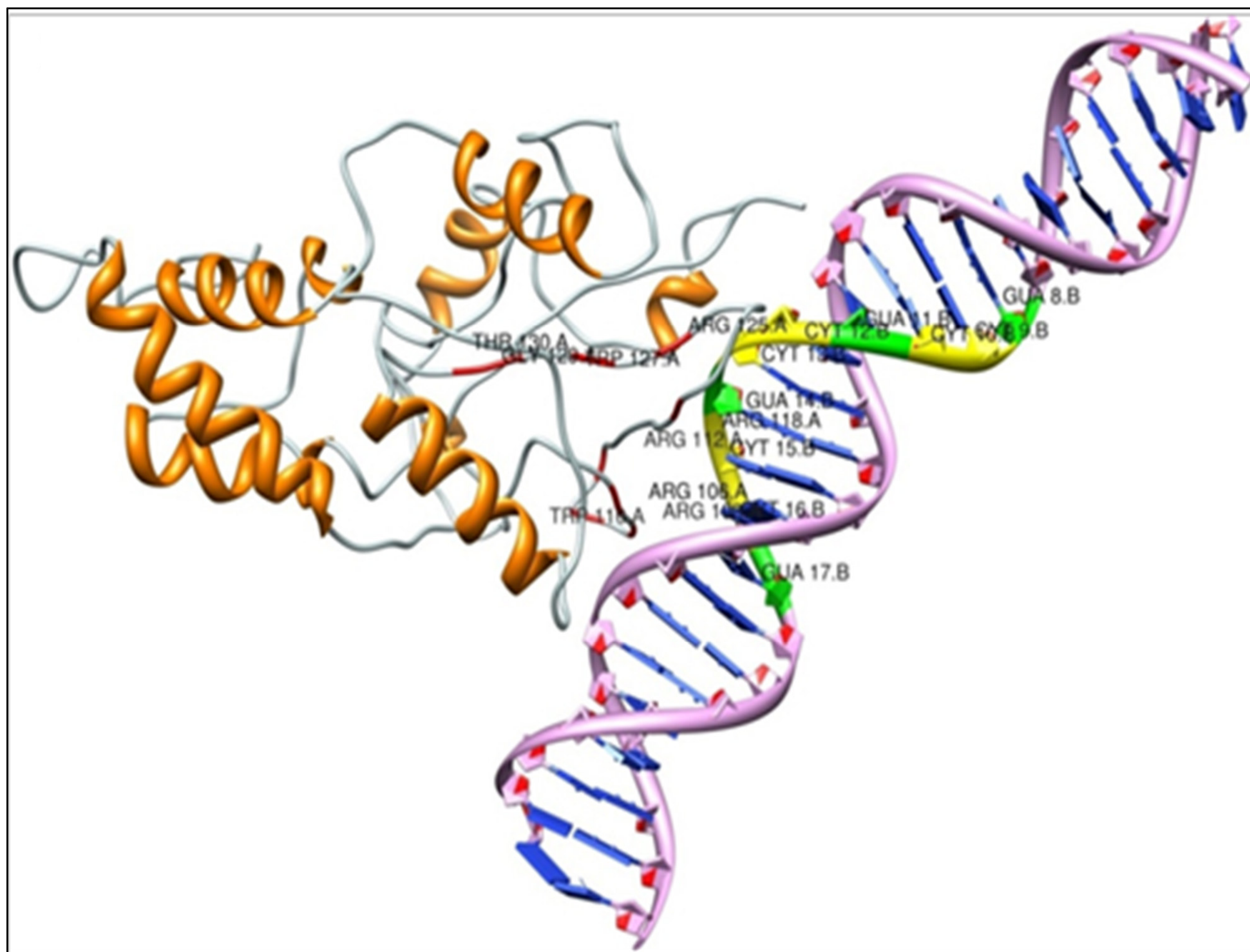


Figure 45. Result showing interaction between 3D structure of Sub1A protein predicted by I-TASSER and DNA model (promoter region having complementary GCCGCC core motif, CGGCGG) predicted by 3D-DART. In DNA model yellow color represents Cytosine at 9, 10, 12, 13, 15 and 16 position while green color represents Guanine at 8, 11, 14, and 17. Interacting residues (THR at 130 and TRP at 110, 127 and ARG at 106,108, 112, 118, 125 and GLY at 129 positions respectively) in protein model represented with red color.

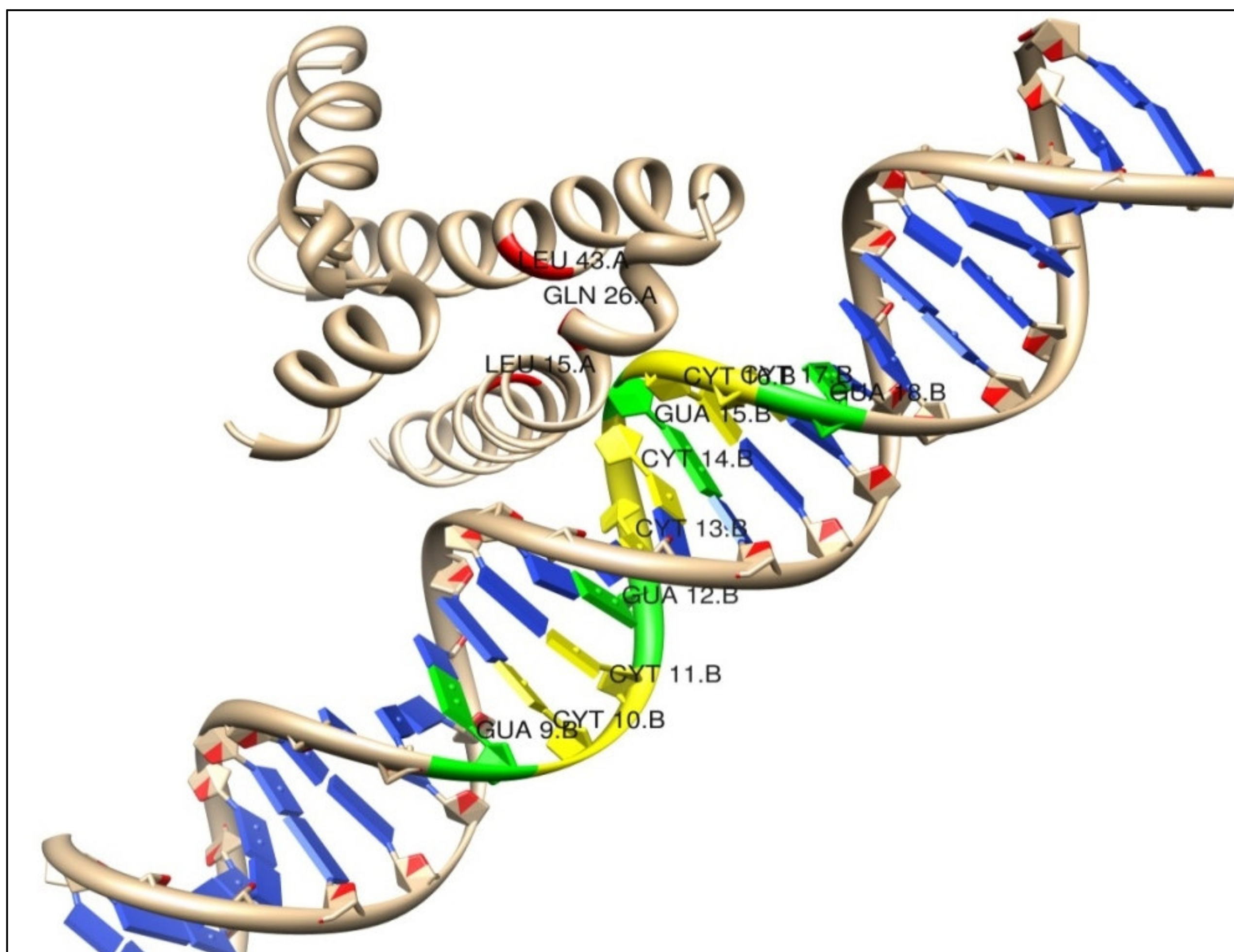


Figure 46: Result showing interaction between 3D structure of CPuORF2 predicted by I-TASSER and DNA model of methyltransferase domain containing protein gene (promoter region having GCC box motif) predicted by 3D-DART. In DNA model yellow color represents Cytosine at 10, 11, 13, 14, 16 and 17 position while green color represents Guanine at 9, 12, 15, and 18. Interacting residues (LEU at 15 and GLN at 26, and LEU at 43 positions respectively) in protein model represented with red color.

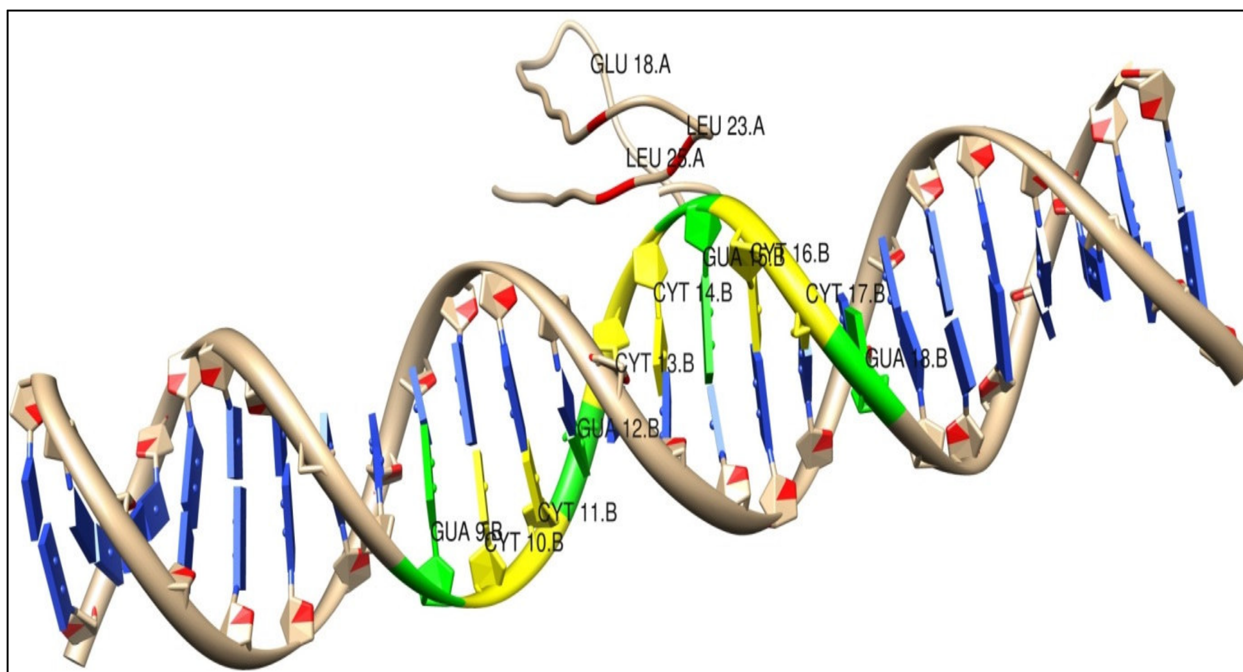


Figure 47: Result showing interaction between 3D structure of B3 DNA binding domain containing protein predicted by I-TASSER and DNA model (promoter region having GCC box motif) predicted by 3D-DART. In DNA model yellow color represents Cytosine at 10, 11, 13, 14, 16 and 17 positions, while green color represents Guanine at 9, 12, 15, and 18. Interacting residues (GLU at 18 and LEU at 23, LEU 25 positions respectively) in protein model represented with red color.

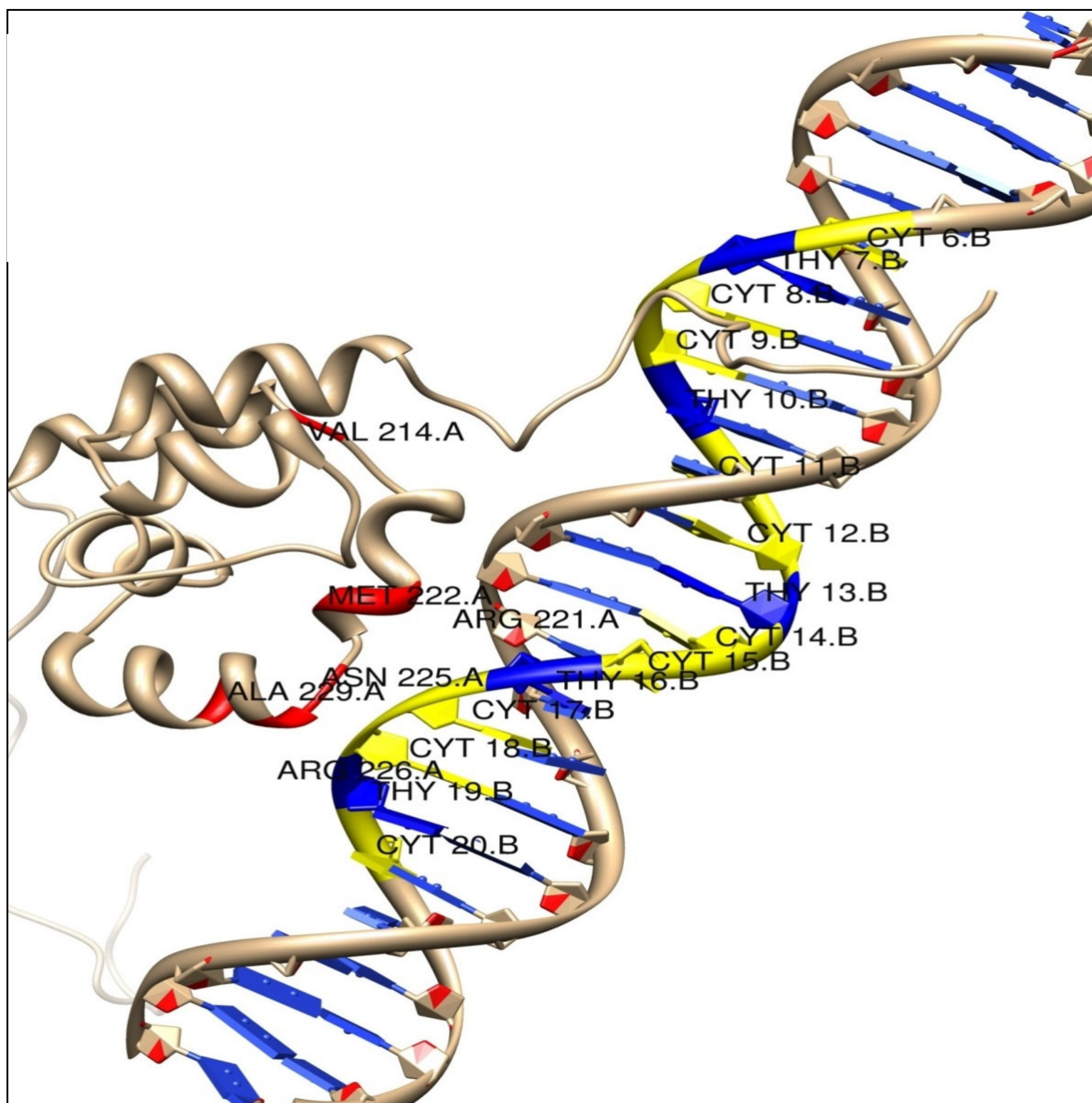


Figure 48: Result showing interaction between 3D structure of bZIP transcription factor protein predicted by I-TASSER and DNA model of rhoGAP domain containing protein gene (promoter region having TCCTCC core motif) predicted by 3D-DART. In DNA model yellow color represents Cytosine at 6,8,9,11,12,14,15,17,18 and 20 position while blue color represents thymine at 7, 10, 13, 16 and 19. Interacting residues (VAL at 214 and ARG at 221, MET at 222, ASN at 225, ARG at 226 and ALA at 229 positions respectively) in protein model represented with red color.

It is reported that Sub1A is a gene related to submergence tolerance. And introgressed rice lines using *Sub1A* gene has been developed having increased submergence tolerance by several folds. Presence of GCC box in the promoter of Ubiquinol Cytochrome C Chaperone gene (Os07g30790) also reported (Kumar et al. 2009). DNA–protein interactions are pivotal for many biological activities. These interactions are fundamental for gene expression and DNA modifications and their function in regulating (Qin and Zhou, 2011) and determining the structure of protein-DNA complexes have been understood many biological processes (Chou et al. 2010). Although, laboratory methods for protein-DNA interaction studies are very expensive and time-consuming, therefore by doing computational analysis possibilities of finding results in short time increases. Docking has been a powerful tool till date when protein-DNA interactions comes into play and there are many softwares available which can perform this type of study such as PISA, PROMOTIF, X3DNA, Read Out, DDNA, DCOMPLEX (Tomovic and Oakeley, 2007). It is very much essential to perform systematic docking for the prediction of protein-DNA complexes (Setny et al. 2012), which were well supported by different method/ techniques/program such as geometric hashing method (Banitt and Wolfson, 2011), Fast Fourier correlation techniques (Aloy et al. 1998), and HADDOCK program (van Dijk et al. 2006, 2010). Henceforth, protein-DNA docking has been studied using one of the advanced versions of HADDOCK software (Kobayashi et al. 2010). HADDOCK (High Ambiguity Driven protein-protein DOCKing) program which starts with a similar rigid body docking of the two partners based on interaction surface definition, a semi-flexible simulated annealing stage followed by a water-refinement step (van Dijk et al. 2006).

DNA protein interaction studies have been reported to establish the involvement of conserved sequence GCCGCC box motif of DNA with Protein (Pandey and Kumar, 2013). Protein–nucleic acid interactions therefore play a crucial role in central biological processes, ranging from the mechanism of replication, transcription and recombination to enzymatic events utilizing nucleic acids as substrates (Sundaralingam and Burkhart, 1997, Luscombe et al. 2000). Pandey and Kumar, (2013), reported the protein-DNA interaction of CCCH-type Zinc finger transcription factor gene and OsCCCH-Zn-1 protein using HADDOCK server. Chen et al. (2003), reported that OsBP-73, a rice gene, encodes a novel DNA-binding protein with a SAP-like domain and their results suggest that OsBP-73 may play an important role in the regulation of cell. Since, during abiotic and biotic stresses regulatory mechanism involve the promoter motifs/cis-regulatory elements in various cellular mechanisms.

Discussion:

Microarray analysis has been used for the differential expression of various genes in plants under various environmental perturbations. And *in silico* study on these microarray results has provided many new insights about the possible function of the different genes. However, validations of these results are very much required. Study on *in silico* motif analysis in various genes has been carried out earlier (Mohanty et al. 2005; Pandey et al. 2007; Kumar et al. 2007). Using *in silico* approach GCC box was detected as consensus promoter motif in 2072 genes promoter from UR-DEGs. On the other hand TCC box was consensus promoter motif detected in 1940 genes promoter from DR-DEGs (Kumar et al. 2007). Therefore, experimental detection of GCC box in UR-DEGs and TCC box in DR-DEGs becomes very much essential. In this study we have tried to validate *in silico* findings on promoter motif analysis by MBP based Real Time PCR that has already been used in nucleic acid sequence detection previously (Giesendorf et al. 1998; Tyagi et al. 1998; Lata et al. 2009). In one hand, presence of larger sequence of GCC box (GCCGCCGCCG) was analyzed in the promoter region (-499 to +100 bp) of UR-DEGs (Figure- 1). On the other hand, larger sequence of TCC box (CTCCTCCTCCTCCTC) was also analyzed in the promoter region (-499 to +100 bp) of DR-DEGs (Figure - 2). Significance (IC, E-value and width length) of GCC box and TCC box were also evaluated by MEME analysis (Figure-3 & 4). Using BD7 protocol Molecular Beacon compatibility score for probe and primers, UR-DEGs and DR-DEGs having sequences GCCGCCGCCG and CTCCTCCTCCTCCTC, respectively in their promoter region (-499 to +100) were selected for validation. Therefore, in present study some genes having more than 200/300 fold increased expression under anoxia (Lasanthi-Kudahettige et al. 2007) were not considered in present validation study. Interestingly, repeated occurrence of CGCCGCCGCCG as well as CTCCTCCTCCTCCTC sequences were seen in selected UR-DEGs and DR-DEGs, respectively (Figure-3& 4).

PCR amplification in terms of Delta Rn vs cycle number graph for UR-DEGs (Figure-8A to 21A), Rn vs cycle number graph for UR-DEGs (Figure-8B to 21B) while dissociation curve analysis for UR-DEGs (Figure-8C to 21C) has been obtained. Similarly, Delta Rn vs cycle number graph for DR-DEGs (Figure-22A to 35A), Rn vs cycle number graph for DR-DEGs (Figure-22B to 35B) while dissociation curve analysis for DR-DEGs (Figure-22C to 35C) has been noticed. Dissociation curve analysis for UR-DEGs and DR-DEGs showed only two curves representing two different amplified products indicating the success of the MBPs based Real Time PCR. Further,

being a fluorescent probe based Real Time PCR; detection of GCC box and TCC box in the promoters of above discussed genes is very much reliable. Literature on nucleic acid sequence detection, sensitivity, accuracy and reliability of MBP are available (Leone et al. 1998; Marras et al. 1999; Piatek et al. 1998; Pas et al. 2005; Elsayed et al. 2003, 2006; Ye et al. 2009). Hinz et al. (2010), reported that RAP2.2 (At3g14230, an APETALA2/ERF-type transcription factor) plays important role in providing resistance during hypoxia by inducing the genes involved in sugar metabolism and fermentation pathway enzymes, as well as ethylene biosynthesis genes. Association of APETALA2 (AP2)/ERF family transcriptional regulators with the Sub1A-1-mediated response under submergence (Jung et al. 2010) has also been reported. Chakravarthy et al. (2003) reported that tomato transcription factor Pti4 (an ERF) is involved in the regulation of gene expression by interacting with GCC box or non-GCC box *cis*-elements directly. AtEBP binding with an oligonucleotide probe containing a mutant GCC box (GCC box contained two points Mutations) eliminate the ability of a *cis*-regulatory element (47-bp fragment that contains two copies of the GCC box) to activate gene expression in an ethylene-dependent manner (Ohme-Takagi and Shinshi, 1995). It was reported that GCC-box work as an ethylene-responsive element that is essential as well as sufficient in some cases for the regulation of transcription (Ohme- Takagi and Shinshi, 1995). Hao et al. (1998) reported that numerous members of the ERF family interact specifically with AGCCGCC through the conserved ERF domain. Fujimoto et al. (2000) described that maltose binding protein –AtERF fusion proteins bind with GCC box sequence (AGCCGCC) and binding activity was abolished when both G residues within the GCC box were replaced by T residues (ATCCTCC). Similarly Buttner and Singh, (1997) also described that GST–AtEBP fusion protein was able to bind to the GCC box but not to the mutant GCC box. Similarly, using electrophoresis mobility shift assay (EMSA) Cheong et al. (2003) concluded that OsEREBP1 specifically binds to the GCC box (AGCCGCC) motif but not to the mutated GCC box (ATCCTCC). Additionally, ethylene response factors are also involved in regulating jasmonate-responsive gene expression by interacting with the GCC-box. And introduction of point mutations into GCC-box sequence substantially reduced jasmonate responsiveness (Brown et al. 2003). Xu et al. (2006) identified three genes (*Sub1A*, *Sub1B* and *Sub1C*) encoding putative ethylene responsive factors in Sub1 region of FR13-derived line. He further reported that *Sub1A* as the major determinant of submergence tolerance. Using EMSA and transient expression assay TiERF1 protein binds with GCC box and can enhance the transcripts of genes with the GCC box *cis*-element. And biochemical

assay study indicated that TiERF1 actually work as an activator-type ERF transcription factor, binds with GCC box *cis*-element and modulate the defense response by up-regulating transcripts of a subset of genes with the GCC box present in their promoters (Liang et al. 2008).

Amplification of DEGs by MBPs was also assessed by Ct values chart for UR-DEGs like ubiquinol-cytochrome C chaperone family protein gene (AK068288 or LOC_Os07g30790), methyltransferase domain containing protein gene (AK064640 or LOC_Os06g05910), 60S ribosomal protein L7 gene (AK058490 or LOC_Os08g42920) and heat shock protein DnaJ gene (AK111076 or LOC_Os06g09560) (Table – 7) successfully. Result showed the good amplification pattern with Template having Molecular Beacon probes specific to GCC box only. This indicates that GCC box is present in the promoter of above selected UR-DEGs and identified successfully. On the other hand OsFBDUF66 - F-box and DUF domain containing protein gene (AK120895 or LOC_Os12g39520), Cytokinin-O-glucosyltransferase 2 gene (AK063324 or LOC_Os06g11720) , Retrotransposon (Ty3-gypsy subclass) protein gene (AK073072 or LOC_Os09g09650), Retrotransposon (Ty1-copia subclass) protein gene (AK121178 or LOC_Os04g02310), Hypothetical protein gene (AK063204 or LOC_Os07g37280), RNA polymerases (N 8 kDa subunit) protein gene (AK108801 or LOC_Os11g08940), Serine acetyltransferase protein gene (AK067089 or LOC_Os03g04140), Transposon protein unclassified gene (AK121698 or LOC_Os10g42150), Deoxyhypusine hydroxylase gene (AK059924 or LOC_Os12g43100) and Protein of unknown function domain containing protein gene (AK121619 or LOC_Os06g40040) (Table – 8) were amplified but showed relatively poor amplification pattern with Template having Molecular Beacon probes specific to GCC box only compared to DEGs mentioned in Table-7. Unfortunately, three genes like expressed protein gene (AK071086 or LOC_Os12g23780), MYB family transcription factor gene (Ak068565 or LOC_Os01g34060), Expressed protein gene (AK102580 or LOC_Os03g10460) have not shown PCR amplification after 3-4 times repeated set of the experiment. This might be due to error in the designing of primer with MB7 and compatibility with Molecular Beacon. Ct values and Std dev Ct were obtained from inbuilt Real Time Software by doing in two replicates. Amplification was done with non template control (NTC) and with Template having Molecular Beacon probes specific to GCC box. To see the specific amplification of GCC box probe we also performed PCR amplification with non-specific TCC box probe.

On the other hand Ct values chart for DR-DEGs like rhoGAP domain containing protein gene (AK067300 or LOC_Os12g05900), DnaK family protein gene (AK100997 or

LOC_Os02g48110), CPuORF11 - conserved peptide uORF-containing transcript gene (AK103103 or LOC_Os02g01240), OsFBX61 - F-box domain containing protein gene (AK103417 or LOC_Os02g52130), CGMC_MAPKCMGC_2_SLT2y_ERK.2 - CGMC includes CDA, MAPK, GSK3, and CLKC kinases gene (AK071376 or LOC_Os06g48590), Helix-loop-helix DNA-binding domain containing protein gene (AK068704 or LOC_Os03g26210), Spermidine synthase gene (AK065153 or LOC_Os02g15550) were also assessed (Table – 9). Result showed the good amplification pattern with Template having Molecular Beacon probes specific to TCC box only. This indicates that TCC box is present in the promoter of above selected DR-DEGs.

Ct values chart for other DR-DEGs like Nitrilase-associated protein gene (AK099444 or LOC_Os11g41150), Wall-associated receptor kinase-like 20 precursor, gene (AK065517 or LOC_Os09g03620), Nucleoside transporter gene (AK102045 or LOC_Os07g37100), Sucrose transporter gene (AK100027 ort LOC_Os03g07480), Phytosulfokines precursor gene (AK073352 or LOC_Os11g05190), Transketolase protein gene (AK100909 or LOC_Os07g09190), TENA/THI-4 family protein gene (AK073626 or LOC_Os03g19390) were also prepared (Table – 10). Here also Real Time PCR amplification pattern of each gene was analyzed in two replicates. Amplification was done with NTC and with Template having Molecular Beacon probes specific to TCC box. To see the specific amplification of TCC box probe we also performed PCR amplification with non-specific GCC box probe. Our result (Table-10) showed the poor amplification pattern with Template having Molecular Beacon probes specific to TCC box only. Although presence of TCC box in the promoter of these DR-DEGs and its poor amplification pattern necessarily indicates that designing of primers and compatibility with MB7 was not up to the extent of DEGs mentioned in Table-9.

Ubiquinol-cytochrome C chaperone family protein is found in mitochondria and related to tissue respiration. Orthologous gene in Arabidopsis having similar putative function indicated its association in the regulation of different biological processes occurring in mitochondria. Methyltransferase domain containing protein gene is involved in embryonic and post embryonic development. However, up-regulation in rice coleoptiles under anoxia indicated its role in stress tolerance. 60S ribosomal protein L7 gene is associated with structural molecular activity and translation. Orthologous of this gene in Arabidopsis with putative function to ribosomal protein L30/L7 family protein has been reported. Heat shock protein DnaJ, gene functions as heat shock protein binding and involved a protein metabolic process. Orthologous of this gene shows similar

putative function in maize and Sorghum. Hence, up-regulation of DEGs in rice coleoptiles under anoxia indicated its association in regulation of structural genes and their mRNAs. Similarly other UR-DEGs (AK063204, AK108801, AK120895, AK067089, AK073072, AK121178, AK063324, AK121619, AK121698, AK059924, AK071086, Ak068565, AK102580) mentioned in Table-8 were also associated in various functions as described in Table-11. Result (Table-7& 11) indicated that UR-DEGs having GCCGCC motif in their promoter region might interact with some transcription factors that regulate their expression in rice coleoptiles under anoxia.

In DR-DEGs, rhoGAP domain containing protein gene is associated with metabolic processes. Its orthologous show putative function similar to Rho GTPase activating protein with PAK-box/P21-Rho-binding domain. DnaK family protein gene is associated with nucleotide binding and localization in plastid. Orthologous gene in Arabidopsis having putative function similar to heat shock protein 70 was noticed indicating its involvement in the regulation of nucleotide interactions in chloroplast. CPuORF11 - conserved peptide uORF-containing transcript gene is associated with cellular activity in mitochondria. Orthologous of this gene in Arabidopsis was noticed having putative function as Ankyrin repeat family protein indicating its involvement in the regulation of phytochromes. OsFBX61 - F-box domain containing protein gene is involved in molecular activities (GTPase activity and protein binding) and other biological function like response to endogenous stimulus, protein modification process and catabolic processes. F-BOX WITH WD-40 2 found as putative function in Arabidopsis as Orthologous gene. CGMC_MAPKCMGC_2_SLT2y_ERK.2 - CGMC includes CDA, MAPK, GSK3, and CLKC kinase gene have molecular activity involved in signal transduction and other biological function as protein amino acid phosphorylation and response to stress whereas no cellular activity was found. Helix-loop-helix DNA-binding domain containing protein gene has molecular activity as sequence-specific DNA binding transcription factor activity and involved in biological function for the regulation of transcription while as cellular component in nucleus and plastid. Spermidine synthase gene involved as catalytic activity and transferase activity in molecular function, biosynthetic and metabolic process in biological processes, also involved in cytoplasm as cellular component. Similarly gene ontology classification for other DR-DEGs (AK069854, AK065517, AK099444, AK102045, AK100027, AK073352, AK100909, AK073626) mentioned in Table-10 has also been done described in Table-12. Our MBPs based Real Time PCR analysis (Table-9 & 10) indicates that above selected DR-DEGs having TCCTCC box in their promoter region binds with other

transcription factor that decreased their expression in rice coleoptiles under anoxia. Preliminary findings related to the validation of GCC box in UR-DEGs and TCC box in DR-DEGs has been presented previously (Prajapati et al. 2011, Pandey et al. 2012).

Ubiquinol-cytochrome C chaperone family protein is found in mitochondria and related to tissue respiration. Orthologous gene in Arabidopsis having similar putative function indicated its association in the regulation of different biological processes occurring in mitochondria. Up-regulation in rice coleoptile under anoxia indicated its association in regulation of structural genes and their mRNAs. Result indicated that UR-DEGs having GCCGCC motif in their promoter region might interact with some transcription factors that regulate their expression in rice coleoptiles under anoxia. Results related to the validation of GCC box in UR-DEGs and TCC box in DR-DEGs has been presented previously (Prajapati et al. 2011, Pandey et al. 2012). Using Beacon Designer 7 we have successfully designed the MBP and specific primers for UR-DEGs and DR-DEGS. This is the first report where MBP based Real Time PCR was successfully used in the identification of GCC box and mutated GCC box (TCC box) in the promoter region of DEGs in rice. Analysis for above selected UR-DEGs having GCC box while DR-DEGs having TCC box in their promoters indicate that these motifs are present and might interact with transcription factor(s) that regulate the differential expression of these genes in rice. In this study we have identified the presence of GCC box in Ubiquinol Cytochrome C Chaperone (*UCCC*) Gene promoter and also tried to establish the relationship between the sub1A protein interactions with the GCC box of the *UCCC* gene. For that we generated 3D structure of Sub1A protein by I-TASSER. DNA model of *UCCC* Gene promoter sequence having core GCCGCC motif was generated by 3D-DART. Interaction of Sub1A and GCCGCC motif was studied by HADDOCK server. Eventually, the involvement in regulation by interacting with core GCC box motif with Transcription factor was found out.

Confirmation about the presence of GCC Box in the promoter region prompted to further study about their interaction with Transcription Factor during submergence. Chakravarthy et al. (2003), reported that tomato transcription factor Pti4 (an ERF) is involved in the regulation of gene expression by interacting with GCC box or non-GCC box *cis*-elements. AtEBP binding with an oligonucleotide probe containing a mutant GCC box (GCC box contained two points Mutations) eliminate the ability of a *cis* regulatory element (47-bp fragment that contains two copies of the GCC box) to activate gene expression in an ethylene-dependent manner (Ohme-Takagi and Shinshi, 1995). It was reported that GCC-box work as an ethylene-responsive element that is

essential in some cases for the regulation of transcription (Ohme-Takagi and Shinshi, 1995). Hao et al. (1998) reported that numerous members of the ERF family interact specifically with AGCCGCC through the conserved ERF domain. Fujimoto et al. (2000), described that maltose binding protein – AtERF fusion proteins bind with GCC box sequence (AGCCGCC) and binding activity was abolished when both G residues within the GCC box were replaced by T residues (ATCCTCC). Similarly, using electrophoresis mobility shift assay (EMSA) Cheong et al. (2003) concluded that OsEREBP1 specifically binds to the GCC box (AGCCGCC) motif but not to the mutated GCC box (ATCCTCC). Additionally, ethylene response factors are also involved in regulating jasmonate-responsive gene expression by interacting with the GCC-box. And introduction of point mutations into GCC-box sequence substantially reduced jasmonate responsiveness (Brown et al. 2003). Using EMSA and transient expression assay TiERF1 protein binds with GCC box and modulate the defense response by up-regulating transcripts of a subset of genes with the GCC box present in their promoters (Liang et al. 2008).

It is well known that *Sub1A* gene (LOC_Os09g11480) was responsible for the submergence tolerance in rice (Xu et al. 2006). Therefore the protein sequence and structure of Sub1A was required. The protein sequence of Sub1A was retrieved from the TIGR (v6.1) (ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.1/). The 3D structure of Sub1A was not available in PDB database, therefore its structure was predicted using I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). I-TASSER server predicts and displays various features in different sections for best model studies. Roy et al. (2012) predicted the structures of three human GPCRs complexes using I-TASSER with a RMSD's 1.6A°, 2.27A° and 2.82A° to the crystal structures in the trans membrane region. The models predicted by I-TASSER were based on the best 10 threading templates available on RCSB PDB. The best predicted model is selected on the basis of confidence score; TM-Score as well as RMSD value (Table -14). The C score value for the best predicted model which is model 1 of Sub1A was -4.06 and furthermore, highly similar structures in PDB (as identified by TM-align) were identified and listed in Table-15. Template proteins with similar binding sites for Sub1A are listed in Table -16. The best binding site is predicted on the basis of C score ^{LB} (Range = 0-1) and BS-Score (>1) values. A higher score C score indicates a more reliable ligand-binding site prediction and BS-score reflects a significant local match between the predicted and template binding site (Zang, 2008, Roy et al. 2010). Qin and

Zhou, (2011) suggested that binding site prediction is a useful tool for building structural models for protein-DNA complexes and for experimental design and validation. Two best predicted binding sites for Sub1A 3D was taken for further interaction studies (Table-16). Structure of Sub1A protein predicted by I-TASSER and visualized by VMD tool has been shown in Figure-45. For *in-silico* protein-DNA interaction studies 3D structure of protein along with DNA was required. Consequently, 3D structure of DNA segment (25 nt long) containing identified and validated GCC Box promoter motif of *UCCC* gene (LOC_Os7g30790) was generated by 3D-DART server. GCC-Box promoter motif positioned at 10-15 nucleotides have been shown in Figure-42.

Protein-DNA interactions are the physical basis of gene expression and DNA modification for vital biological activities (Qin and Zhou, 2011). Because there is no simple mapping code between DNA base pairs and protein amino acids, the prediction of protein-DNA interactions is a challenging problem. Therefore, HADDOCK can make use of a broader array of restraints, including those derived from biochemical and biophysical data (Kobayashi et al. 2010). Determining the structure of protein-DNA complexes and clarifying the factors that regulating their interaction is essential to better understand many biological processes (Chou et al. 2010). A review describing the experimental strategies currently employed to solve structures of protein–DNA complexes and to analyze their dynamics has been published (Campagne et al. 2011). Protein–DNA interactions facilitate the fundamental functions of living cells and are universal in all living organisms (Sathyapriya and Vishveshwara, 2004).

To determine the protein-DNA interactions the Easy interface of HADDOCK web server was used (de Vries et al. 2010). Before going for docking AIR files were generated for both the interacting molecules having information about the active binding sites of Sub1Aprotein as well as in the DNA model. The HADDOCK score is the weighted sum of van der Waals energy (negative indicating favorable interactions), electrostatic energy (negative indicating favorable interactions), distance restraints energy (only unambiguous and AIR (ambig-restraints), direct RDC restraint energy, inter vector projection angle restraints energy, diffusion anisotropy energy, dihedral angle restraints energy, symmetry restraints energy, buried surface area (negative weight indicate a better interface), binding energy, desolvation energy. Meanwhile, the solution structures are analyzed for their intermolecular hydrogen bonds and intermolecular hydrophobic contacts by HADDOCK, the solutions are clustered according to the interface ligand RMSDs. The Z-score indicates the standard deviations from the average of a particular cluster in terms of HADDOCK score.

For the prediction of best interaction, different binding sites and models of Sub1A were docked with 3D structure of DNA segment having promoter motif of *UCCC*. The result showed best model SAUGCM2-BS2 (Table-16) has HADDOCK score of 11.7 +/- 4.4 and Z-score of -1.7 when interacting with bended DNA segment (Figure-45). Similar protein-DNA interaction study performed with the other validated genes (UR and DR DEGs) and highly expressed TFs (Table-13). In UR-DEG the GCC box present in the promoter motif of validated methyltransferase domain containing protein gene (AK064640 or LOC_Os06g05910) showed the good HADDOCK score (Table-16) and interaction result with two highly expressed TFs protein i.e. CPuORF2 - conserved peptide uORF-containing transcript (AK064903 or LOC_Os09g13570) and B3 DNA binding domain containing protein (AK070845 or LOC_Os03g06850) in protein-DNA docking models. Highly expressed up- regulated TF CPuORF2 - conserved peptide uORF-containing transcript (Figure-39) and B3 DNA binding domain containing protein (Figure-40) docked with 3D structure of DNA promoter motif of methyltransferase domain containing protein gene (Figure-43). Result showed the best model ILZMTGCM2-BS2 have HADDOCK score -74.7+/-2.0 and Z-score of -2 (Table-16) when interaction between TF CPuORF2 - conserved peptide uORF-containing transcript with GCC box containing DNA segment of methyltransferase domain protein gene (Figure-46) and another best interaction model IABMTGCM1-BS1 have HADDOCK score -91+/-4.1 and Z-score of -2.2 (Table-16) when interaction between B3 DNA binding domain containing protein with GCC box containing DNA segment of methyltransferase domain protein gene (Figure-47). On the other hand in DR-DEG, TCC box present in the promoter of rhoGAP domain containing protein gene (AK067300 or LOC_Os12g05900) showed the good HADDOCK score (Table-16) and protein-DNA interaction result with highly down regulated bZIP transcription factor (LOC_Os02g52780). For interaction study 3D model of the highly down-regulated TF bZIP transcription factor (Figure-41) and promoter motif TCC box containing DNA model of rhoGAP domain containing protein gene (Figure-44) generated by on line I-TASSER and 3D DART tools. Best protein-DNA interaction model IBZRHOGTM2-BS2 have HADDOCK score -154.1+/-9.7 and Z-score of -1.9 (Table-16) when interaction between bZIP transcription factor protein with DNA segment of rhoGAP domain containing protein gene (Figure-48). This primarily report of protein–DNA interaction study of GCC and TCC box of DEGs with the highly expressed TFs indicated that the involvement of the GCC box and TCC box in the gene regulation with the help of TFs under anoxia or submergence induced anoxia.

Conclusion:

The validation/detection of consensus promoter motif that were identified by us in our *in silico* study (Kumar et al. 2009) were carried out by probe based (molecular beacon) Real Time PCR. Molecular beacons for GCC-Box and mutated GCC-Box were designed and procured respectively. A total of 32 DEGs (15 Up-regulated and 17 Down-regulated DEGs) were selected for this study. Result revealed the detection of promoter motif by Molecular Beacon Probe based Real Time PCR. GO classification was done to find the possible involvement of these DEGs. Among DEGs considerable number of TFs were also found and classified in different classes. Prediction of 3D structure of TF protein and promoter motifs as well as their interactions was successfully performed.

Year wise plan for achieving other objectives:

Objective	Year wise schedule	Work Progress
1.Screening of highly differentially expressed genes (DEGs)and Designing of gene specific primer and specific molecular beacon	**1st Year**	**Completed**
2.Plant Material and DNA isolation Real-Time PCR analysis	**2nd Year**	**Completed**
3. Analyzing the data of Real time PCR	**3rdYear**	**Completed**

References

1. Agarwal S and Grover A (2005). Isolation and transcription profiling of low-O_2 stress-associated cDNA clones from the flooding-stress-tolerant FR13A rice genotype. Ann Bot. 96:831.
2. Aloy P, Moont G, Gabb HA et al. (1998). Modelling repressor proteins docking to DNA. Proteins. 33(4): 535–549.
3. Banitt I and Wolfson HJ (2011). ParaDock: a flexible non-specific DNA–rigid protein docking algorithm. Nucleic Acids Res.39 (20):135.
4. Baxter-Burrell A, Yang Z, Springer PS, Bailey-Serres J (2002). RopGAP4-dependent RopGTPase rheostat control of Arabidopsis oxygen deprivation tolerance. Science. 296:2026.
5. Branco-Price C, Kawaguchi R, Ferreira RB, Bailey-Serres J (2005). Genome-wide analysis of transcript abundance and translation in Arabidopsis seedlings subjected to oxygen deprivation. Ann Bot. 96:647–660.
6. Brown RL, Kazan K, McGrath KC, Maclean DJ, Manners JM (2003). A Role for the GCC-Box in Jasmonate-Mediated Activation of the PDF1. 2 Gene of Arabidopsis. Plant Physiol. 132: 1020.
7. Buttner M and Singh KB (1997). *Arabidopsis thaliana* ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. Proc Natl Acad Sci U S A. 94(11): 5961–5966.
8. Campagne S, Gervais V, Milon A (2011). Nuclear magnetic resonance analysis of protein–DNA interactions. J. R. Soc. Interface. 8: 1065-1078.
9. Chakravarthy S, Tuori RP, D'Ascenzo MD, Fobert PR, Després C, Martin GB (2003). The tomato transcription factor Pti4 regulates defense-related gene expression via GCC box and non-GCC box cis elements. Plant Cell.15: 3033.
10. Chen J, Tang WH, Hong MM et al. (2003). OsBP-73, a rice gene,encodes a novel DNA-binding protein with a SAP-like domain and its genetic interference by double-stranded RNA inhibits rice growth. Plant Mol Biol. 52(3):579-90.
11. Chen W, Provart NJ, Glazebrook J et al. (2002). Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell.14: 559-574.

12. Cheong YH, Moon BC, Kim JK, Kim CY, Kim MC, Kim IH, Park CY, Kim JC, Park BO, Koo SC, Yoon HW, Chung WS, Lim CO, Lee SY, Cho MJ (2003). BWMK1, a Rice Mitogen-Activated Protein Kinase, Locates in the Nucleus and Mediates Pathogenesis-Related Gene Expression by Activation of a Transcription Factor. Plant Physiol.132:1961.

13. Chou CC, Rajasekaran M, Chen C (2010). An effective approach for generating a three Cys2His2zinc-finger-DNA complex model by docking. BMC Bioinformatics.11: 334.

14. de Vries SJ, van Dijk M, Bonvin AMJJ (2010). The HADDOCK web server for data-driven biomolecular docking.Nature Protocols. 5: 883-897.

15. Elsayed S, Chow BL, Hamilton NL, Gregson DB, Pitout JD, Church DL (2003). Development and validation of a molecular beacon probe-based real-time polymerase chain reaction assay for rapid detection of methicillin resistance in *Staphylococcus aureusm.* Arch Pathol Lab Med,127:845.

16. Elsayed S, Plewes K, Church D, Chow B, Zhang K (2006). Use of Molecular Beacon Probes for Real-Time PCR Detection of Plasmodium falciparum and Other Plasmodium Species in Peripheral Blood Specimens. J Clin Microbiol.44: 622.

17. Fennoy SL, Nong T, Bailey-Serres J (1998). Transcriptional and posttranscriptional processes regulate gene expression in oxygen-deprived roots of maize. Plant J.15:727-735.

18. Fujimoto SY, Ohta M, Usui A, Shinshi H, Ohme-Takagi M (2000). Arabidopsis ethylene-responsive element binding factors act as transcriptional activators and repressors of GCC box-mediated gene expression. Plant Cell. 12:393.

19. Geigenberger P (2003) Response of plant metabolism to too little oxygen. Current Opinion in Plant Biology. 6: 247-256.

20. Giesendorf BA, Vet JA, Tyagi S, Mensink EJ, Trijbels FJ, Blom HJ (1998). Molecular beacons: a new approach for semiautomated mutation analysis. Clin Chem. 44: 482.

21. Hao DY, Ohme-Takagi M, Sarai A (1998). Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. J Biol Chem. 273(41):26857-26861.

22. Hinz M, Wilson IW, Yang J, Buerstenbinder K, Llewellyn D, Dennis ES et al. (2010). Arabidopsis RAP2.2: an ethylene response transcription factor that is important for hypoxia survival. Plant Physiol.153:757.

23. Jin H and Martin C (1999). Multifunctionality and diversity within the plant MYB-gene family. Plant Molecular Biology.4:577-585.

24. Jung KH, Seo YS, Walia H, Cao P, Fukao T, Canlas PE, Amonpant F et al. (2010). The submergence tolerance regulator Sub1A mediates stress-responsive expression of AP2/ERF transcription factors. Plant Physiol.152:1674.

25. Klok EJ, Wilson IW, Wilson D, Chapman SC, Ewing RM, Somerville SC et al. (2002). Expression profile analysis of the low oxygen response in Arabidopsis root cultures. Plant Cell.14:2481-2494.

26. Kobayashi M, Eiso AB, Bonvin AMJJ et al. (2010). Structure of the DNA-bound BRCA1 C-terminal Region from Human Replication Factor C p140 and Model of the Protein-DNA Complex. The J Biological Chem. 285 (10):10087-10097.

27. Kumar A, Shankaracharya, Vidyarthi AS, Pandey DM (2007). *In-silico* analysis of promoter motif of anoxia inducible genes in rice (*Oryza sativa* L.). In: Proceedings of 5th International Symposium on Rice Functional Genomics (ISRFG' 2007), Epochal, Tsukuba, Japan, Poster 112.

28. Kumar A, Smita S, Sahu N, Sharma V, Shankaracharya, Vidyarthi A, Pandey DM (2009). *In Silico* analysis of motifs in promoters of differentially expressed genes in rice (*Oryza sativa* L.) under anoxia. Int J Bioinform Res & Appl.5:525.

29. Kumar A, VidyarthiAS, PandeyDM et al. (2009). *In Silico* analysis of motifs in promoters of differentially expressed genes in rice (*Oryza sativa* L.) under anoxia. Int J Bioinfo Res & App. 5(5): 525-547.

30. Lafitte HR, Ismail A, Bennett J (2004). Abiotic stress tolerance in rice for Asia: progress and the future."New directions for a diverse planet". In: Proceeding of the 4th International Crop Science Congress, 26 Sep – 1 Oct, Brisbane, Australia.

31. Lasanthi-Kudahettige R, Magneschi L, Loreti E, Gonzali S, Licausi F, Novi G, Beretta O, Vitulli F, Alpi A, Perata P (2007). Transcript profiling of the anoxic rice coleoptiles. Plant Physiol. 144:218-231.

32. Lata P, Ram S, Agrawal M (2009). Shanker R, Real Time PCR for the Rapid Detection of vanA Gene in Surface Waters and Aquatic Macrophyte by Molecular Beacon Probe. Environ Sci Technol.43:3343.

33. Leone G, van Schijndel H, van Gemen B, Kramer FR, Schoen CD (1998). Molecular beacon probes combined with amplification by NASBA enable homogeneous, real-time detection of RNA. Nucleic Acids Res. 26:2150.

34. Liang HX, Lu Y, Liu HX, Wang FD, Xin ZY, Zhang ZY (2008). A novel activator-type ERF of Thinopyrumintermedium, TiERF1, positively regulates defense responses. J Exp Bot. 59(11):3111.

35. Liu F, Vantoai T, Moy LP, Bock G, Linford LD, Quackenbush J (2005). Global transcription profiling reveals comprehensive insights into hypoxic response in Arabidopsis. Plant Physiol.137:1115.

36. Lu XJ and Olson WK (2003). 3DNA: a software package for the analysis rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res. 31: 5108-5121.

37. Luscombe N, Austin SE, Berman HM et al. (2000). An overview of the structures of protein–DNA complexes.Genome Biology. 1:1–37.

38. Magneschi L and Perata P (2009). Rice Germination and Seedling Growth in the Absence of Oxygen. Ann Bot. 103:181.

39. Marras SAE, Kramer FR, Tyagi S (1999). Multiplex detection of single-nucleotide variations using molecular beacons. Genet Anal Biomol Eng. 14:151.

40. Mohanty B, Krishnan SPT, Swarup S, Bajic VB (2005). Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species. Ann Bot. 96: 669.

41. Mukhopadhyay A, Vij S, Tyagi AK (2004). Over expression of a zinc-finger protein gene from rice confers tolerance to cold, dehydration, and salt stress in transgenic tobacco. Proc Nat Acad Sci USA.101:6309-6314.

42. Nakano T, Suzuki K, Fujimura T, Shinshi H (2006). Genome-wide analysis of the ERF gene family in Arabidopsis and Rice. Plant Physiol. 140:411.

43. Nie X, Singh RP, Tai GC (2002). Molecular characterization and expression analysis of 1-aminocyclopropane-1-carboxylate oxidase homologs from potato under abiotic and biotic stresses. Genome. 45:905.

44. Ohme-Takagi M and Shinshi H (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell. 7:173.

45. Pandey DM and Kim SR (2006). Isolation of the low oxygen stress genes in rice (*Oryza sativa L.)* Proceeding of 4th International Symposium on Rice Functional Genomics (ISRFG'2006) Montpellier, France.Poster. 37:96.
46. Pandey DM and Kim SR (2012) Identification and Expression Analysis of Hypoxia Stress Inducible CCCH-type Zinc Finger Protein Genes in Rice. J. Plant Biol. (2012) 55:489-497.
47. Pandey DM and Kumar A (2013). 3D structure prediction and protein-DNA interaction of CCCH-type zinc finger transcription factor gene in Rice (*Oryza sativa* L.) Int. J. Comp.Bioinfo. In Silico Modeling. 2(2):94-103.
48. Pandey DM, Deb R, Kumar A, Vidyarthi AS, Kim SR (2007). Study on the functional characterization of low oxygen stress inducible genes in rice (*Oryza sativa* L.) in proceedings of 5th *IntSym on Rice FuncGenom* (ISRFG'2007), Epochal, Tsukuba, Japan (2007) Poster 110.
49. Pandey DM, Kumar A, Prajapati GK, Kumari A, Wany A (2012). Computational and molecular biology approaches for plant research & improvement in *Intl Conf on Indus Biotech* (ICIB-2012) & 9th *Convention of The Biotech Research Society*, India held at Punjabi University, Patiala, India from November 21-23, 2012. IL-57.
50. Pas SD, Noppornpanth S, van der Eijk AA, de Man RA, Niesters HG (2005). Quantification of the newly detected lamivudine resistant YSDD variants of Hepatitis B virus using molecular beacons. J ClinVirol. 32:166.
51. Paul AL, Schuerger AC, Popp MP, Richards JT, Manak MS, Ferl RJ (2004). Hypobaric biology: Arabidopsis gene expression at low atmospheric pressure. Plant Physiol. 134:215.
52. Piatek AS, Tyagi S, Pol AC, Telenti A, Miller LP, Kramer FR, Alland D (1998). Molecular beacon sequence analysis for detecting drug resistance in *Mycobacterium tuberculosis*. Nat Biotechnol. 16:359.
53. Prajapati GK, Kumar A, Pandey DM (2011). Study on Expression Analysis of DEG with Respect to GCC box and TCC box in Rice (Oryza sativa L.) under Anoxia, poster presentation in *IntSym in Plant Biotech towards tolerence to stresses and Enhancing Crop Yield* (ISPB 2011) at B.I.T Mersra Ranchi, Jharkhand scheduled on September 28-October 1, 2011. Poster No-FG-03.
54. Qin S and Zhou HX (2011). Structural Models of Protein-DNA Complexes Based on Interface Prediction and Docking. Current Protein and Peptide Science. 12: 531-539.

55. Roy A, Kucukural A, Zhang Y (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols. 5:725-738.

56. Roy A, Yang J, Zhang Y (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucl Acids Res. 40: 471-477.

57. Sadiq I, Fanucchi F, Paparelli E, Alpi E, Bachi A, Alpi A, Perata P (2011). Proteomic identification of differentially expressed proteins in the anoxic rice coleoptiles. Journal of Plant Physiol. 168:2234.

58. Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002). DNA-binding specificity of the AP2/ERF domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. BiochemBiophy Res Commun. 290: 998.

59. Saleh A and Pagés M (2003). Plant AP2/ERF transcription factors. Genetika. 35:37.

60. Sathyapriya R and Vishveshwara S (2004). Interaction of DNA with clusters of amino acids in proteins.Nucl. Acids Res. 32 (14): 4109-4118.

61. Setny P, Bahadur RP, Zacharias M (2012). Protein-DNA docking with a coarse-grained force field. BMC Bioinformatics. 13: 228.

62. Sharoni AM, Nuruzzaman M, Satoh K, Shimizu T, Kondoh H, Sasaya T, Choi IR, Omura T, Kikuchi S (1996). Gene structures, classification and expression models of the AP2/EREBP transcription factor family in rice. Plant Cell Physiol. 52:344.

63. Sundaralingam M, Burkhart B (1997). Protein–nucleic acid interactions: general principles. In: Parthasarathy S and Glusker JP (eds) Aspects of Crystallography in Molecular Biology. 363–390. New Delhi: New Age International Publishers.

64. Tomovic A and Oakeley EJ (2007). Position dependencies in transcription factor binding sites. Bioinformatics. 23:933-941.

65. Tyagi and Kramer (1996). Molecular beacons: probes that fluoresce upon hybridization, NatureBiotechnology.14: 303-308.

66. Tyagi S, Bratu DP, Kramer FR (1998). Multicolor molecular beacons for allele discrimination. Nat Biotechnol. 16:49.

67. van Dijk M, van Dijk ADJ, Hsu V et al. (2006). Information driven protein- DNA docking using, HADDOCK: it is a matter of flexibility. Nucleic Acids Res. 34(11): 3317–3325.

68. van Dijk, M and Bonvin AMJJ (2010). Pushing the limits of what is achievable in protein-DNA docking: benchmarking HADDOCK's performance. Nucleic Acids Res. 38(17): 5634-5647.

69. vanDijk M and Bonvin AMJJ (2009). 3D-DART: a DNA structure modelling server. Nucl Acids Res. 37:235-239.

70. Vartapetian BB and Jackson MB (1997). Plant adaptations to anaerobic stress. Ann Bot. 79:3.

71. Vilo J, Brazma A, Jonassen I, Robinson A, Ukkonen E (2000). Mining for putative regulatory elements in the yeast genome using gene expression data. Proc. Int Conf Intel Sys Mol Biol8:384-394.

72. Wang Z, Dalkilic M, Kim S (2004). Guiding motif discovery by iterative pattern Refinement. ACM Sym.Applied Computing.Nicosia. Cyprus. 162-166.

73. Wassmann R, Jagadish SVK, Heuer S, Ismail A, Redona E, Serraj R, Singh R K, Howell G, Pathak H, Sumfleth K (2009). Production: The Physiological and Agronomic Basis for Possible Adaptation Strategies. In Donald L. Sparks, editor, Advances in Agronomy.Burlington: Academic. 101: 59.

74. Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S et al. (2006). Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature. 442:705-708.

75. Yamauchi M, Aragones DV, Casayuran PR, StaCruzJR PC, Asis CA, Cruz RT (2000). Seedling establishment and grain yield of tropical rice sown in puddled soil. Agron J. 92:275.

76. Ye Q, Zhuang H, Zhou C (2009). Detection of naphthalene by real-time immuno-PCR using molecular beacon, Mol Cell Probes. 23:29.

77. Zhang C, Zhang F, Zhou J et al. (2012). Overexpression of a phytochrome-regulated tandem zinc finger protein gene, OsTZF1, confers hypersensitivity to ABA and hyposensitivity to red light and far-red light in rice seedlings. Plant Cell Rep. 31:1333-1343.

78. Zhang Y (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 9: 40.

Publications

International Journals:

1. Prajapati GK, Kumar A, Pandey DM (2013) Molecular Beacon Probe–Based Real-Time Polymerase Chain Reaction Assay for the Identification of GCC box and TCC box in the Promoters of Differentially Expressed Genes in Rice (*Oryza sativa* L.). Indian Journal of Experimental Biology. (Submitted)
2. Prajapati G.K., Kashyap N., Kumar A., Pandey D.M. (2013) Identification of GCC box in the promoter region of ubiquinol cytochrome C chaperone gene using Molecular Beacon probe and its *in silico* protein-DNA interaction study in rice (*Oryza sativa* L.). Int J Comput Bioinfo In Silico Model. 2: 213-222.

Proceedings in Conference:

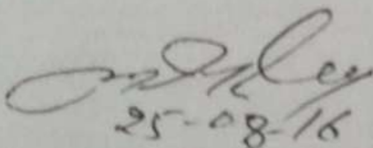
1. Prajapati, G.K. Kumar, A. and Pandey, D.M. (Abstract for Poster Presentation) Molecular Beacon probe based promoter motif detection & *in silico* protein-DNA interaction studies during submergence in rice (*Oryza sativa* L.). In: The 7th International Rice Genetics Symposium (RG7) at Dusit Thani Hotel Manila, Philippines scheduled from November 05 to 08, November 2013. Abstract no. **RICE-0451**.
2. Pandey DM, Kumar A, Prajapati GK, Kumari A, Wany A (Oral Presentation) Computational and Molecular Biology Approaches for Plant Research & Improvement. In: International Conference on Industrial Biotechnology (ICIB-2012) & 9th Convention of The Biotech Research Society, India held at Punjabi University, Patiala, India from November 21-23, 2012. IL-57, Pp 37
3. Pandey DM, Kumar A, Prajapati GK, Kumari A, Wany A (Oral Presentation) Unraveling of Integrated Mechanisms in OMICS Technology for Crop Improvement. In: 3rd World Congress on Biotechnology (Biotechnology-2012) held at Hyderabad International Convention Center, Hyderabad, India from September 13-15, 2012. *J. Biotechnol. Biomater.* 2(6)-166.
4. Prajapati, G.K. Kumar, A. and Pandey, D.M. (2011) 'Study on Expression Analysis of DEG with Respect to GCC box and TCC box in Rice (*Oryza sativa* L.) under Anoxia". In: Proceedings of International Symposium in Plant Biotechnology towards Tolerance to Stresses and Enhancing Crop Yield (ISPB 2011) organized by Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi, Jharkhand from September 28 to Octpober 01, 2011. Abstract no. FG03. Page no. 112.

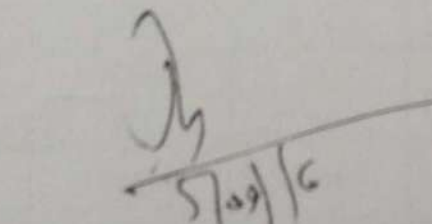
Annexure - IV

UNIVERSITY GRANT COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002

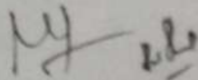
Utilization Certificate

It is certified that the grant of **Rs. 12,35,567/-** (Rupees Twelve Lakhs, Thirty Five Thousand and Five Hundred Sixty Seven Only) was sanctioned while a sum of **Rs. 13,34,020/-** (Rupees Thirteen Lakhs, Thirty Four Thousand and Twenty Only) including revised arrear of fellowship, HRA, etc. was received from the University Grant Commission under scheme of support for major research project entitled **"Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) Under Submergence"** vide UGC letter No. 37-113/2009(SR) dated 19th December, 2009; 15 October, 2012 and 28th August, 2015 vide RTGS UTR (No. CBINR52015082810002005). A sum of Rs. 2,864/- only was also earned as interest. Out of total received grant, only **Rs. 12,25,183/- (Rupees Twelve Lakhs, Twenty Five Thousand and One Hundred Eighty Three only)** has been fully utilized for the purpose for which it was sanctioned and in accordance with the terms and conditions laid down by the University Grant commission. A sum of **Rs 1,11,701/- (Rupees One Lakh, Eleven Thousand and Seven Hundred One only)** remaining unutilized is being refunded to UGC New Delhi vide DD no. 078422 dated 03.09.2016.



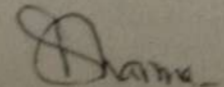


SIGNATURE OF THE PRINCIPAL INVESTIGATOR

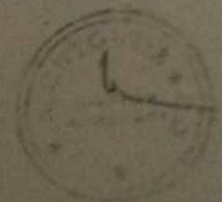


REGISTRAR/PRINCIPAL
Registrar
Birla Institute of Technology
Mesra: Ranchi

STAUTORY AUDITOR
For K. N. GUTGUTIA & CO.
Chartered Accountants
Firm Registration No.304153E



K. C. SHARMA
Partner
Membership No.50819



07 SEP 2016

Enclosure – II

Annexure - V

UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002

STATEMENT OF EXPENDITURE IN RESPECT OF MAJOR RESEARCH PROJECT

1. Name of Principal Investigator – **Dr. Dev Mani Pandey**

2. Deptt. of University/College –**Biotechnology**

3. UGC approval No. and Date – **37-113/2009 (SR); 19th December, 2009**

4. Title of the Research Project – **"Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) Under Submergence"**

5. Effective date of starting the project – 01/02/2010

6. a. Period of Expenditure – From 01/02/2010 to 31/03/2016

b. Details of Expenditure –

Sl. No	Expenditure against grant received	Sanctioned (Rs)	1st installment (Rs)	2nd installment (Rs)	3rd Installment (Rs)	Total amount received (Rs)	Total expenditure (Rs)	Grant to be reimbursed (Rs)
1	Books And Journals	Nil	-	-	-	-	-	-
2	Consumable	5,50,000/-	2,75,000/-	1,96,040/-	82,134/-	5,53,174/-	4,84,038/-	**69,136/-**
3	Fellowship	4,92,267/-	1,44,000/-	3,45,000/-	Nil	4,89,000/-	4,89,000/-	**0**
	HRA	Nil	Nil	Nil	98,546/-	98,546/-	98,453/-	**93/-**
4	Travelling Expenses	60,000/-	30,000/-	10,000/-	20,000/-	60,000/-	30,417/-	**29,583/-**
5	Contingency	45,000/-	22,500/-	10,000/-	12,500/-	45,000/-	34,975/-	**10,025/-**
6	Overhead	88,300/-	88,300/-	-	Nil	88,300/-	88,300/-	**0**
7	Equipment	Nil	Nil	Nil	Nil	Nil	Nil	**Nil**
8	Interest	Nil	Nil	Nil	Nil	2,864/-		**2,864/-**
	Total	12,35,567/-	5,59,800/-	5,61,040/-	2,13,180/-	13,36,884/-	12,25,183/-	**1,11,701/-**

- Due to receipt of revised arrear of fellowship, HRA etc. money received from UGC New Delhi is more compared to the sanctioned money.

c. Staff

Date of Appointment – **08/04/2010**

Expenditure Incurred	From	To	Amount Approved (Rs.)	Amount Received Expenditure Incurred (Rs.)	Expenditure Incurred (Rs.)
Honorarium to PI (Retired Teachers) Rs.12,000/- p.m.	NA	NA	NA	NA	NA
Post-Doctoral Fellow Fellowship @ Rs. 12,000/- p.m.	NA	NA	NA	NA	NA
Project Associate salary @ Rs.10,000/- p.m.	NA	NA	NA	NA	NA
Project Fellow salary @ Rs.14000/- p.m. for 2 years and Rs. 16000/-for 3rd Year	8/4/10	31/1/13	4,92,267	4,89,000	4,89,000/-

1. It is certified that the appointment(s) have been made in accordance with the terms and conditions laid down by the Commission.
2. It as a result of checks or audit objective, some irregularly is noticed, later date, action will be taken to refund, adjust or regularize the objected amounts.
3. Payment @ revised rates shall be made with arrears on the availability of additional funds.
4. It is certified that the grant of **Rs. 12,35,567**/- (Rupees Twelve Lakhs, Thirty Five Thousand and Five Hundred Sixty Seven Only) was sanctioned while a sum of **Rs. 13,34,020/-** (Rupees Thirteen Lakhs, Thirty Four Thousand and Twenty Only) including revised arrear of fellowship, HRA, etc. was received from the University Grant Commission under scheme of support for major research project entitled **"Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) Under Submergence"** vide UGC letter No. 37-113/2009(SR) dated 19th December, 2009; 15 October, 2012 and 28th August, 2015 vide RTGS UTR (No. CBINR52015082810002005). A sum of Rs. 2,864/- only was also earned as interest. Out of total received grant only **Rs. 12,25,183/- (Rupees Twelve Lakhs, Twenty Five Thousand and One Hundred Eighty Three only)** has been fully utilized for the purpose for which it was sanctioned and in accordance with the terms and conditions laid down by the University Grant commission. A sum of **Rs 1,11,701/- (Rupees One Lakh, Eleven Thousand and Seven Hundred One only)** remaining unutilized is being refunded to UGC New Delhi vide DD no. ---078422--------------- dated -03.09.2016.

25-08-16

SIGNATURE OF PRINCIPAL INVESTIGATOR

5/9/16

REGISTRAR/PRINCIPAL
Birla Institute of Technology
Mesra: Ranchi

Annexure -III

UNIVERSITY GRANTS COMMISSION

BAHADUR SHAH ZAFAR MARG

NEW DELHI – 110 002

Final Report of the work done on the Major Research Project

(Report to be submitted within 6 weeks after completion of each year)

1. Project report no. **Final**
2. UGC Reference No. **37-113/2009 (SR)**
3. Period of report: from **01/02/2010 to 31/07/2013**
4. Title of research project "Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) under Submergence".
5. (a) Name of the Principal Investigator: Dr. Dev Mani Pandey

   (b) Deptt. and University/College where work has progressed:

   Biotechnology & Birla Institute of Technology, Mesra, Ranchi, Jharkhand
6. Effective date of starting of the project: 01/02/2010
7. Grant approved and expenditure incurred during the period of the report:

a. Total amount approved: Rs. 12, 35,567/-Only

b. Total expenditure: Rs. 11,27,357/- Only

c. Report of the work done: (Please attach a separate sheet) **Annexure- I**

i. Brief objective of the project

1. Screening of the highly expressed DEGs containing GCC box and mutated GCC box from our previous study (Kumar et al., 2009)
2. Designing the gene specific primer and specific Molecular Beacon probe
3. Validation of consensus promoter motifs by using above probes

ii. Work done so far and results achieved and publications, if any, resulting from the work (Give details of the papers and names of the journals in which it has been published or accepted for publication _

Prajapati GK, Kashyap N, Kumar A and Pandey DM. (2013) Identification of GCC Box in the Promoter region of Ubiquinol Cytochrome C Chaperone Gene using Molecular Beacon Probe and its *In silico* Protein-DNA interaction study in Rice

(*Oryza sativa* L.). International Journal of Computational Bioinformatics and *In Silico* Modeling. 2(5):213-222.

iii. Has the progress been according to original plan of work and towards achieving the objective. if not, state reasons… Yes

iv. Please indicate the difficulties, if any, experienced in implementing the project - Designing and procuring of the Molecular beacon probe. Delay in receiving of 2nd Installment hence purchasing of required chemicals.

v. If project has not been completed, please indicate the approximate time by which it is likely to be completed. A summary of the work done for the period (Annual basis) may please be sent to the Commission on a separate sheet.

Project has been completed.

vi. If the project has been completed, please enclose a summary of the findings of the study. Two bound copies of the final report of work done may also be sent to the Commission __ please enclose a summary of the findings of the study.

- **Summary of the project:**

Rice is the primary staple food for more than half of the world's population but often suffers with various stresses that resulted reduction in productivity. During submergence, differential expression of genes (DEGs) is regulated by interaction of transcription factors (TF) with promoter motifs. However, the key signatures for co-regulated family genes are common promoter motifs having similar expression patterns. In co-expressed up-regulated and down-regulated genes, the promoter motifs GCC box and TCC box (mutated GCC box) were identified, respectively during submergence in rice. A report using computational approach indicated that GCC box (GCCGCC) is the consensus promoter motif of up regulated differentially expressed genes (DEGs), while mutated GCC box (TCCTCC) for down–regulated DEGs. However, no validation study has been done so far. Therefore, here we aim to validate/detect the GCC box in the up-regulated DEG and TCC box in down-regulated DEGs using molecular beacon probe and Real-Time PCR analysis. Molecular Beacon based qRT-PCR enabled to identify the GCC box in UR-DEGs while TCC box in DR-DEGs and obtained significant binding affinity for TF with these promoter motifs. A total of 32 DEGs (17 Up-regulated and 15 Down-regulated DEGs) were selected for

validation/detection. Result revealed the detection of promoter motif by Molecular Beacon Probe based Real Time PCR. Gene Ontology classification was done to find the possible involvement of these DEGs. Among DEGs considerable number of TFs were also found and classified in different classes. Prediction of 3D structure of TF protein and promoter motifs as well as their interactions was successfully performed.

- **Two bound copies of the final report of work done are being sent: Annexure- I**

vii. Any other information which would help in evaluation of work done on the project. At the completion of the project, the first report should indicate the output, such as (a) Manpower trained (b) Ph. D. awarded (c) Publication of results (d) other impact, if any

(a) Manpower trained:

Summer Trainee: Ms. Neha Kashyap

(b) Ph. D. awarded

Mr. Gopal Kumar Prajapati, a Junior Research Fellow in the said project, has been enrolled for Ph. D. program. He will submit his thesis soon.

(c) Publication of results —— Annexure – II

Prajapati GK, Kashyap N, Kumar A and Pandey DM. (2013) Identification of GCC Box in the Promoter region of Ubiquinol Cytochrome C Chaperone Gene using Molecular Beacon Probe and its *In silico* Protein-DNA interaction study in Rice (*Oryza sativa* L.). International Journal of Computational Bioinformatics and *In Silico* Modeling. 2(5):213-222.

(d) Other impact

NA

28/1/13

SIGNATURE OF THE INVESTIGATOR

REGISTRAR/ PRINCIPAL

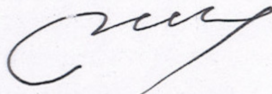
Registrar
Birla Institute of Technology
Mesra; Ranchi

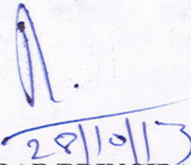
ANNEXURE IV

UNIVERSITY GRANT COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002

Utilization Certificate

It is certified that the grant of **Rs. 12, 35,567**/- (Rupees Twelve Lakhs, Thirty Five Thousand and Five Hundred Sixty Seven Only) was sanctioned and a total amount of **Rs. 1120840/-** (Rupees Eleven Lakhs, Twenty Thousand and Eight Hundred Forty Only) was received from the University Grant Commission under scheme of support for major research project entitled **"Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) Under Submergence"** vide UGC letter No. 37-113/2009(SR) dated 19th December, 2009 and 15 Oct, 2012. The total grant sanctioned (**Rs. 1120840/- only**) has been fully utilized for the purpose for which it was sanctioned and in accordance with the terms and conditions laid down by the University Grant commission. Hence, a total amount of **Rs 114727/-** is requested to be released. Kindly consider the same and do the needful.





SIGNATURE OF THE PRINCIPAL INVESTIGATOR

REGISTRAR/PRINCIPAL

Registrar
Birla Institute of Technology
Mesra; Ranchi

STAUTORY AUDITOR

Annexure - V

UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002

STATEMENT OF EXPENDITURE IN RESPECT OF MAJOR RESEARCH PROJECT

1. Name of Principal Investigator – **Dr. Dev Mani Pandey**

2. Deptt. of University/College –**Biotechnology**

3. UGC approval No. and Date – **37-113/2009 (SR); 19th Dec, 2009**

4. Title of the Research Project – **"Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) Under Submergence"**

5. Effective date of starting the project – 01/02/2010

6. a. Period of Expenditure – From 01/02/2010 to 31/07/2013

b. Details of Expenditure –

Sl. No	Expenditure against grant received	Sanctioned (Rs)	1st installment (Rs)	2nd installment (Rs)	Total amount received (Rs)	Total expenditure (Rs)	Grant remaining to be released (Rs)
1	Books And Journals	Nil	-	-	-	-	-
2	Consumable	5,50,000	275000	206040	481040	482396	**68960**
3	Fellowship	4,92,267	144000	345000	489000	489000	**3267**
4	Travelling Expenses	60000	30000	Nil	30000	30417	**30000**
5	Contingency	45,000	22500	10000	32500	37244	**12500**
6	Overhead	88, 300	88300	-	88300	88300	-
7	Equipment	Nil	Nil	Nil	Nil	Nil	
	Total	12,35,567/-	5,59,800	5,61,040	11,20840	1127357	**114727**

c. Staff

Date of Appointment – **08/04/2010**

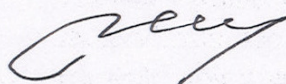
Expenditure Incurred	From	To	Amount Approved (Rs.)	Expenditure Incurred (Rs.)	Grant remaining to be released (Rs.)
Honorarium to PI (Retired Teachers) Rs.12,000/- p.m.	NA	NA	NA	NA	NA
Post-Doctoral Fellow Fellowship @ Rs. 12,000/- p.m.	NA	NA	NA	NA	NA
Project Associate salary @ Rs.10,000/- p.m.	NA	NA	NA	NA	NA
Project Fellow salary @ Rs.14000/- p.m. for 2 years and Rs. 16000/-for 3rd Year	8/4/10	31/1/13	4,92,267	489000	3267/-

1. It is certified that the appointment(s) have been made in accordance with the terms and conditions laid down by the Commission.

2. It as a result of checks or audit objective, some irregularly is noticed, later date, action will be taken to refund, adjust or regularize the objected amounts.

3. Payment @ revised rates shall be made with arrears on the availability of additional funds.

4. It is certified that the grant of **Rs. 12, 35,567/-** (Rupees Twelve Lakhs, Thirty Five Thousand and Five Hundred Sixty Seven Only) was sanctioned and a total amount of **Rs. 1120840/-** (Rupees Eleven Lacs, Twenty Thousand and Eight Hundred Forty Only) was received from the University Grant Commission under scheme of support for major research project entitled **"Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) Under Submergence"** vide UGC letter No. 37-113/2009(SR) dated 19th December, 2009 and 15 Oct, 2012. The total grant sanctioned (**Rs. 1120840/-**) has been fully utilized for the purpose for which it was sanctioned and in accordance with the terms and conditions laid down by the University Grant commission. Hence, a total amount of **Rs. 114727/-** is requested to be released. Kindly consider the same and do the needful.



SIGNATURE OF PRINCIPAL INVESTIGATOR

18/12/13

REGISTRAR/PRINCIPAL

Registrar
Birla Institute of Technology
Mesra; Ranchi

Annexure – VIII

UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002

PROFORMA FOR SUBMISSION OF INFORMATION AT THE TIME OF SENDING THEFINAL REPORT OF THE WORK DONE ON THE PROJECT

1. NAME AND ADDRESS OF THE PRINCIPAL INVESTIGATOR – **Dr. Dev Mani Pandey, Department of Biotechnology, BIT, Mesra, Ranchi, Jharkhand – 835215**

2. NAME AND ADDRESS OF THE INSTITUTION – **Birla Institute of Technology, Mesra, Ranchi, Jharkhand– 835 215**

3. UGC APPROVAL NO. AND DATE – **37-113/2009 (SR); 19th December, 2009**

4. DATE OF IMPLEMENTATION – **01/02/2010**

5. TENURE OF THE PROJECT – **3 years**

6. TOTAL GRANT ALLOCATED – **Rs. 12, 35, 567/-only**

7. TOTAL GRANT RECEIVED – **Rs. 11, 20,840/- only**

8. FINAL EXPENDITURE – **Rs. 11,27,357/- only**

9. TITLE OF THE PROJECT – "Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) Under Submergence".

10. OBJECTIVES OF THE PROJECT –

    a) Screening of the highly expressed DEGs containing GCC box and mutated GCC box from our previous study (Kumar et al., 2009)

    b) Designing the gene specific primer and specific Molecular Beacon probe

    c) Validation of consensus promoter motifs by using above probes

11. WHETHER OBJECTIVES WERE ACHIEVED –(GIVE DETAILS)-Yes, ANNEXURE -

12. ACHIEVEMENTS FROM THE PROJECT –

1. Enrollment of JRF in Ph. D. Program
2. Training for one student pursuing her M. Sc. Biotechnology
3. Publication of obtained research in International Journal as presentations of result in various International Conferences/Symposium

13. SUMMARY OF THE FINDINGS:

Rice is the primary staple food for more than half of the world's population but often suffers with various stresses that resulted reduction in productivity. During submergence, differential expression of genes (DEGs) is regulated by interaction of transcription factors (TF) with promoter motifs. However, the key signatures for co-regulated family genes are common promoter motifs having similar expression patterns. In co-expressed up-regulated and down-regulated genes, the promoter motifs GCC box and TCC box (mutated GCC box) were identified, respectively during submergence in rice. A report using computational approach indicated that GCC box (GCCGCC) is the consensus promoter motif of up regulated differentially expressed genes (DEGs), while mutated GCC box (TCCTCC) for down–regulated DEGs. However, no validation study has been done so far. Therefore, here we aim to validate/detect the GCC box in the up-regulated DEG and TCC box in down-regulated DEGs using molecular beacon probe and Real-Time PCR analysis. Molecular Beacon based qRT-PCR enabled to identify the GCC box in UR-DEGs while TCC box in DR-DEGs and obtained significant binding affinity for TF with these promoter motifs. A total of 32 DEGs (17 Up-regulated and 15 Down-regulated DEGs) were selected for validation/detection. Result revealed the detection of promoter motif by Molecular Beacon Probe based Real Time PCR. Gene Ontology classification was done to find the possible involvement of these DEGs. Among DEGs considerable number of TFs were also found and classified in different classes. Prediction of 3D structure of TF protein and promoter motifs as well as their interactions was successfully performed.

14. CONTRIBUTION TO THE SOCIETY –

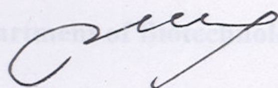
Providing Research Training to M. Sc. Biotechnology Student

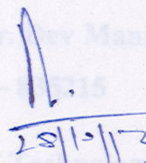
15. WHETHER ANY PH.D. ENROLLED/PRODUCED-OUT OF THE PROJECT

Project Fellow has enrolled for Ph.D.

16. NO. OF PUBLICATIONS OUT OF THE PROJECT – (PLEASE ATTACH RE-PRINTS)

Prajapati GK, Kashyap N, Kumar A and Pandey DM (2013) "Identification of GCC Box in the Promoter region of Ubiquinol Cytochrome C Chaperone Gene using Molecular Beacon Probe and its *In silico* Protein-DNA interaction study in Rice (*Oryza sativa* L.)". International Journal of Computational Bioinformatics and *In Silico* Modeling. 2(5):213-222





(PRINCIPAL INVESTIGATOR) **(REGISTRAR/PRINCIPAL)**

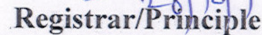
Registrar
Birla Institute of Technology
Mesra: Ranchi

Annexure X

UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002

MAJOR RESEARCH PROJECT COPY OF THE SPECIMEN OF HOUSE RENT FOR POST-DOCTORAL FELLOW / PROJECT ASSOCIATE / PROJECT FELLOW

Certified that Mr. Gopal Kumar Prajapati is eligible for House Rent Allowances of Rs. 2800/- per month for the duration 08.04.2010 to 31.03.12 and Rs. 3200/- per month for the duration 01.04.2012 to 31.01.13 @ 20% as per University Rules.



Registrar/Principle

(Signature with Seal)

Registrar
Birla Institute of Technology
Mesra; Ranchi

23236351, 23232701, 23237721, 23234116
23235733, 23232317, 23236735, 23239437

विश्वविद्यालय अनुदान आयोग
बहादुरशाह जफर मार्ग
नई दिल्ली-110 002
UNIVERSITY GRANTS COMMISSION
BAHADURSHAH ZAFAR MARG
NEW DELHI-110 002

PROFORMA FOR SUBMISSION OF INFORMATION AT THE TIME OF MID-TERM GROUP REVIEW WORKSHOP

(TO BE SUBMITTED IN TRIPLICATE)

1.	**UGC Reference No. & Date**	37-113/2009(SR), 19 December 2009
2.	**Name of the Principal Investigator**	Dr. Dev Mani Pandey
3.	**Address**	Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi-835215, Jharkhand, India **Residential:** Q.NO - CII/71 BIT Mesra, Ranchi-835215, Jharkhand, India
4.	**Department and University/ College where the project has undertaken**	Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi-835215, Jharkhand, India
5.	**Title of the Project**	"Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) under Submergence"
6.	**Date of Implementation**	February 01, 2010
7.	**Tenure of the project**	Three years from February 2010 to January 2013
8.	**Grants Received**	**1st Installment :** 5,59,800/- **2nd Installment :** Nil
9	**Objectives of the Project**	1) Screening of the highly DEGs from our previous study (Kumar et al., 2009) 2) Designing the gene specific Primer and specific Molecular Beacon Probe 3) Validation of consensus promoter motifs using above designed Probe
10	**Methodology**	1. Screening of highly differentially expressed genes (DEGs) 2. Designing of Gene specific Primer 3. Designing of specific probe for consensus promoter motifs 4. Plant Material and DNA isolation 5. Real-Time PCR analysis

11	**Work done so far (Please give details)**	1. Screening of the highly DECs from our previous study (Kumar et al., 2009)- **Completed** 2. Designing the gene specific primer and specific Molecular Beacon probe **Completed** 3. Validation of consensus promoter motifs using above designed probe – **In Progress**
12	**Work remains to be done (please give details)**	**Objective-3:** Validation of consensus promoter motifs by using designed probe
13	**Has the progress been according to original plan of work and towards achieving objectives if not, state reasons**	Yes, progress of the project is according to the original plan of work and towards achieving objectives
14	**Whether Project work was delayed. If yes, specify reasons**	Project was slightly delayed from February 2010 to April 2010 due to Advertisement and recruitment of Project Fellow
15	**Please indicate the approximate time by which the project work is likely to be completed**	One and half Year
16	**Please indicate the difficulties, if any, experienced in implementing the project**	We experienced difficulties in purchasing the Molecular Beacon Probe from UK that delayed the progress of project slightly
17	**Collaboration, if any (with Department, University, Industry etc,)**	No
18	**Ph. D Enrolled, if yes, details**	Yes, Phd/BT/1058/2010 (Enclosures)
19	**Details of the Publications resulting from the project work (please attach re-prints) letter of Acceptance of paper communicated.**	Preparation of Full Manuscript on the topic of Abstract presented in **ISPB-2011** is in progress
20	**Any other information which would help in evaluation of work done on the project**	Mr. Gopal Kumar Prajapati has attended and also presented research work in following international conferences/symposium: 1. Prajapati, G.K. Kumar, A. and Pandey, D.M (2011) Study on Expression Analysis of DE with Respect to GCC box and TCC box Rice (*Oryza sativa* L.) under Anoxia" abstra submitted in "International Symposium Plant Biotechnology towards tolerence stresses and Enhancing Crop Yeild (ISP 2011) at B.I.T Mersra Ranchi, Jharkhand, September 28 - October 1, 2011. Poster Nc FG03, Page No.: 112. 2. International Conference in "World Congre of Biotechnology" at International Conventic Centre (HICC), Hyderabad, (21-23 Marc 2011).

		3. International Symposium on Recent Advanc[illegible] in Cross-disciplinary Microbiology: Avenu[illegible] & Challenges and International Workshop [illegible] rRNA Sequencing, Phylogeny & Ne[illegible] Generation Genome Sequencing (AMI-201[illegible] at BIT Mesra, Ranchi, India on 14-1[illegible] December, 2010,

21. Financial Assistance Provided/ Expenditure incurred:

S. no.	Items	Amount Approved (1st Installment)	Expenditure incurred so far
1	Books & Journal	Nil	Nil
2	Equipments	Nil	Nil
3	Honorarium	Nil	Nil
4	Contingency	22,500/-	27,391/-
5	Travel/fieldwork	30,000/-	30,537/-
6	Chemicals & Glassware	2,75,000/-	2,71,841/-
6	Hiring Services	Nil	Nil
7	Overhead	88,300/-	88,300/-
8	Any other items (please specify)	Nil	Nil
9	Honorarium to Principal Investigator (from 2010 to 2013)	Nil	Nil
	Staff (date of appointment) (from 8 April 2010 to 7 April 2012) (please give details of staff appointed in the prescribed format annexure IX as per XI plan guidelines of Major Research Project)	1,44,000/-	1,42133/-
	Total	5,59,800/-	5,60,202/-

It is certified that the grant of **5,59,800/-(Five Lakhs Fifty Nine Thousand Eight hundred only)** received from the University Grants Commission under the scheme of support for Major Research Project entitled **"Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) under Submergence"** vide UGC letter No. **F. 37-113/2009(SR)** dated **19 Dec 2009** has been fully utilized for the purpose for which it was sanctioned and in accordance with the terms and conditions laid down by the University Grants Commission.

PRINCIPAL INVESTIGATOR
(SIGNATURES WITH SEAL)

REGISTRAR/PRINCIPAL
(SIGNATURES WITH SEAL)

Registrar
Birla Institute of Technology
Mesra; Ranchi

Dr. Ambarish S. Vidyarthi, FIE
Professor & Head
Deptt. of Biotechnology
Birla Institute of Technology
Mesra, Ranchi-835215

Annexure-IX

PROFORMA FOR SUPPLYING THE INFORMATION IN RESPECT OF THE STAFF APPOINTED UNDER THE SCHEME OF MAJOR RESEARCH PROJECT

UGC FILE NO. F.- 37-113/2009(SR) (HRP) **YEAR OF COMMENCEMENT- 2010**

TITLE OF THE PROJECT: Functional Genomics Approach towards Validation of Biologically Important Consensus Promoter Motifs in Rice (*Oryza sativa* L.) under Submergence

<table>
<tr><td>1.</td><td>Name Of the Principal Investigator</td><td>Dr. Dev Mani Pandey</td></tr>
<tr><td>2.</td><td>Name of the University/College</td><td>B.I.T Mesra, Ranchi Jharkhand</td></tr>
<tr><td>3.</td><td>Name of the Research Personnel appointed</td><td>Gopal Kumar Prajapati</td></tr>
<tr><td>4.</td><td>Academic qualification</td><td>
<table>
<tr><td>S. No.</td><td>Qualifications Year Marks %age</td><td>Year</td><td>Marks</td><td>%age</td></tr>
<tr><td>1</td><td>M.Sc.</td><td>2009</td><td>1547</td><td>64.45</td></tr>
<tr><td>2</td><td>M.Phil</td><td></td><td></td><td></td></tr>
<tr><td>3</td><td>Ph.D.</td><td></td><td></td><td></td></tr>
</table>
</td></tr>
<tr><td>5.</td><td>Date of joining.</td><td>8 April 2010</td></tr>
<tr><td>6.</td><td>Date of Birth of Research Personnel</td><td>10 August 1982</td></tr>
<tr><td>7.</td><td>Amount of HRA, if drawn</td><td>Nil</td></tr>
<tr><td>8.</td><td>Number of Candidate applied for the post</td><td>4</td></tr>
</table>

CERTIFICATE :

This is to certify that all the rules and regulations of UGC Major Research Project outlined in the guidelines have been followed. Any lapse on the part of the University will liable to terminate of said UGC project.

Principal Investigator | **Head of the Deptt.** | **Registrar/Principal**

Dr. Ambarish S. Vidyarthi, Ph.
Professor & Head
Deptt. of Biotechnology
Birla Institute of Technology
Mesra, Ranchi-835215

Registrar
Birla Institute of Technology
Mesra; Ranchi

International Journal of Computational Bioinformatics and In Silico Modeling
Vol. 2, No. 5 (2013): 213-222
Research Article
Open Access

aizeon publishers

ISSN: 2320-0634

Identification of GCC Box in the Promoter region of Ubiquinol Cytochrome C Chaperone Gene using Molecular Beacon Probe and its *in silico* Protein-DNA interaction study in Rice (*Oryza sativa* L.)

Gopal Kumar Prajapati[1], Neha Kashyap[1**], Ashutosh Kumar[1] and Dev Mani Pandey[1]*

[1] Department of Biotechnoogy, Birla Institute of Technology, Mesra, Ranchi-835215, Jharkhand, India
[**]Present Address: Department of Biotechnology, Ranchi University, Ranchi-834001, Jharkhand, India

*Corresponding author: Dev Mani Pandey; e-mail: dmpandey@bitmesra.ac.in

Received: 24 July 2013 Accepted: 03 August 2013 Online: 01 September 2013

ABSTRACT

Rice is the primary staple food for more than half of the world's population but very sensitive to various abiotic stresses (submergence) causing crop loss. During abiotic stresses various genes are differentially expressed to cope up with the stress conditions. The identification of *Sub1A* gene was a major breakthrough for the submergence tolerance which often regulates other genes by binding to their consensus promoter motifs such as GCC box. It was observed that Ubiquinol Cytochrome C chaperone (*UCCC*) gene was among many up-regulated differentially co-expressed genes having GCC box as a conserved motif. The primary role of *UCCC* gene is oxidative respiration but also has imperative secondary functions in plants. Therefore, *UCCC* gene was selected for the identification of GCC Box in the promoter region using Molecular Beacon Probe based Real Time PCR and their interaction with Sub1A protein. Real Time PCR analysis confirmed the presence of GCC box. Subsequently, the interaction of Sub1A with GCC box was studied through HADDOCK server. Protein-DNA interaction thus, suggested significant binding affinity of Sub1A towards GCC box present in the promoter region of *UCCC* gene.

Keywords: 3D-DART, Differentially expressed genes, HADDOCK, I-TASSER Oryza sativa, Sub1A, Ubiquinol Cytochrome C Chaperone gene

INTRODUCTION

Rice is a most important crop for human consumption with production in over 150 million hectares yielding almost 600 million megagrams annually [1]. It is a very important staple food that feeds more than half the world's population [2]. Rice thrives in water-logged conditions and can tolerate submergence at levels that would kill other crops [3] but is highly sensitive to salinity among cereals [4], and is susceptible to drought and cold [3]. Rice is a semi-aquatic plant and well adapted to survive partial flooding conditions. However, it is damaged when submerged for a relatively longer period of time [5-7].

A plant when in submerged condition inhibits aerobic respiration and photosynthesis, and stimulates a variety of responses that can enhance survival, acts as a switch from aerobic to anaerobic respiration [8]. The anaerobic stress stimulates the composite metabolic pathways by the differential expression of a large number of genes [9] including genes coding for transcription factors [10] and signal transduction components [11]. Studies on differential expression of genes have also been explained by microarray experiments in anoxic rice coleoptiles [12]. The regulation of gene expression in response to oxygen deprivation has been described in *Oryza sativa* [12-16].

Further, it was discovered that three ethylene response factors (ERFs) were identified within the *Sub1A* locus in tolerant rice varieties (e.g. FR13A) determining it as the major determinant of tolerance. Therefore, identification of the *Sub1* locus and the elucidation of its role in the adaptation of rice to submergence is a breakthrough in plant adaptation to anaerobiosis [6]. Xu et al. [16] reported that FR13A Sub1 region encodes three transcription factors (*Sub1A*, *Sub1B* and *Sub1C*) belonging to the B-2 subgroup of the ethylene response factors (ERFs)/ethylene-responsive element binding proteins (EREBPs)/apetala 2-like proteins (AP2). Study on molecular marker survey and expression analyses of *Sub1A* in rice has also been reported [17]. In Arabidopsis, gene expression study revealed the modulation of gene expression occurred positively or negatively by interaction of ERF-TF with GCC Box [18]. Chakravarthy et al. [19] described Tomato Ethylene-Responsive Factor (ERF) transcription factor Pti4, which binds the GCC box *cis*-element that is present in the promoters of many Pathogenesis-Related (PR) genes. Kumar et al. [20, 21] reported that consensus promoter motif GCC box (GCCGCC) was highly conserved in the promoter of up-regulated differentially expressed Genes (DEGs) in rice under anoxia. These transcriptional regulations of DEGs were facilitated by the interaction of TFs with promoter motifs/*cis*-regulatory elements which provided an insight to the vast molecular mechanisms of co-expressed/regulated genes during stress.

Cytochrome C is a small peripheral, nuclear encoded membrane protein located in inter-membrane space of mitochondria. It functions in the catalytic transfer of electrons between respiratory complexes III and IV [22]. The functional importance and unique intra-organellar position of cytochrome *c* molecule has been investigated in animals and is reported to contain cAMP response element (CRE) and nuclear respiratory factor (NRF)-binding sites [23]. CRE is a cytochrome c gene promoter involved in cAMP-dependent expression and NRF is involved in the coordinating activities of nuclear and mitochondrial genes [24]. However, little is known about the cytochrome *c* gene in plants with the exception that it has been cloned and sequenced in Arabidopsis and rice. The regulation of rice cytochrome *c* gene *OsCc1* (accession no. M63704) and its promoter activities in transgenic rice have been examined [25]. They demonstrated that *OsCc1* is expressed in most of the tissues, and its expression is particularly high in the non-photosynthetic parts of plants such as roots, calli, and suspension cells. A fusion gene was constructed for evaluating the *OsCc1* promoter with the *sgfp* gene and introduced into rice. The activity of this gene in various tissues and cell types of transgenic plants was analyzed in comparison with other promoters by conventional dot-blot hybridization techniques [25]. Chaperone genes are encoded proteins in both prokaryotes and eukaryotes that bind to nascent or unfolded polypeptides and ensure correct folding or transport [26]. It is well known that plant heat- shock proteins (HSPs) and other chaperones play major roles both in response to adverse environmental conditions and in various developmental processes [27], rather they are also responsible for protein folding, assembly, translocation & degradation [28].

DNA–protein interactions are pivotal for many biological activities. These interactions are fundamental for gene expression and DNA modifications and their function in regulating [29] and determining the structure of protein-DNA complexes have been understood many biological processes [30]. Although, laboratory methods for protein-DNA interaction studies are very expensive and time-consuming, therefore by doing computational analysis possibilities of finding results in short time increases. Docking has been a powerful tool till date when protein-DNA interactions comes into play and there are many softwares available which can perform this type of study such as PISA, PROMOTIF, X3DNA, ReadOut, DDNA, DCOMPLEX [31]. It is very much essential to perform systematic docking for the prediction of protein-DNA complexes [32] which were well supported by different method/ techniques/program such as geometric hashing method [33], Fast Fourier correlation techniques [34] and HADDOCK program [35,36]. Henceforth, protein-DNA docking has been studied using one of the advanced versions of HADDOCK software [37]. HADDOCK (High Ambiguity Driven protein-protein DOCKing) program which starts with a similar rigid body docking of the two partners based on interaction surface definition, a semi-flexible simulated annealing stage followed by a water-refinement step [35].

DNA protein interaction studies have been reported to establish the involvement of conserved sequence GCCGCC box motif of DNA with Protein [38]. Protein–nucleic acid interactions therefore play a crucial role in central biological processes, ranging from the mechanism of replication, transcription and recombination to enzymatic events utilizing nucleic acids as substrates [39, 40]. Pandey and Kumar [38] reported the protein -DNA interaction of CCCH-type Zinc finger transcription factor gene and OsCCCH-Zn-1 protein using HADDOCK server. Chen et al. [41] reported that OsBP-73, a rice gene, encodes a novel DNA-binding protein with a SAP-like domain and their results suggest that OsBP-73 may play an important role in the regulation of cell. Since, during abiotic and biotic stresses regulatory mechanism involve the promoter motifs/cis-regulatory elements in various cellular mechanisms. Hence, the presence of GCC box in Ubiquinol Cytochrome C Chaperone (*UCCC*) gene promoter needs to be validated using a suitable molecular technique. In present study, first time we have tried to validate the GCC box (GCCGCC) in UR-DEG *UCCC* gene by using MBP based Real-Time PCR amplification assay to detect nucleic acid sequences [42-45]. In this study we have identified the presence of GCC box in *UCCC* Gene and also tried to establish the relationship between the sub1A protein interactions

with the GCC box of the *UCCC* gene. For that we generated 3D structure of Sub1A protein by I-TASSER. DNA model of *UCCC* Gene promoter sequence having core GCCGCC motif was generated by 3D-DART. Interaction of Sub1A and GCCGCC motif was studied by HADDOCK server. Eventually, the involvement in regulation by interacting with core GCC box motif with Transcription factor was found out.

MATERIALS AND METHODS

Identification of consensus promoter motif and DEG

For the identification of DEG's, online available microarray result of anoxic rice coleoptiles reported by Lasanthi-Kudahettige et al. [12] was used, which was stated in our previous work [21]. From the microarray result, genes were shortlisted as UR-DEGs (up-regulated DEGs, expression increased by equal or more than two-fold, ≥2X). Promoter sequence of -499 to +100 bp was retrieved from Eukaryotic Promoter Database (EPD) for each shortlisted UR-DEGs, DR-DEGs and UC-DEGs. Promoter motifs as well as their consensus promoter motifs were analyzed through MEME (v 4.5.0). It was also observed that ubiquinol-cytochrome C chaperone family protein gene (AK068288 or Os07g30790) was up-regulated during anoxia and also have detected GCC box motif in their promoter region.

Designing of Molecular Beacon probes and gene specific primers

To identify the presence of the detected GCC box promoter motif in the up-regulated ubiquinol-cytochrome C chaperone family protein gene, Molecular Beacon probe (MBP) and its specific primers were designed. Promoter sequence of this gene having length -499 to +100 was retrieved from the Eukaryotic Promoter Database (http://www.epd.isb-sib.ch/seq_download.html). Considering the location of GCC box sequences, MBPs and their specific primers were designed using Beacon Designer 7 (BD7, PREMIER Biosoft, USA). As per BD7 protocol to maintain the optimum difference of annealing temperature (Tm) between probe and gene specific primers, MBP length was adjusted having the sequence of GCCGCCGCCG rather than the core sequence GCCGCC that was consensus motif in UR-DEGs [21]. Molecular Beacon compatibility score for probe and primers were also considered while its designing. Designed MBPs and primers were procured from GeneLink™ (USA) and Hysel India Pvt. Ltd, respectively.

Isolation of Genomic DNA from Rice

Rice seeds of Azucena (*japonica sp.*) were surface sterilized with 0.1% of $HgCl_2$ and incubated in dark for 48 h at 35° C. Sterilized seeds [Figure-1(a)] were grown in pot at room temperature. Genomic DNA was isolated from rice seedlings [Figure-1(b)] using CTAB (2X) method and subjected to RNase treatment by standard protocol. The concentration of genomic DNA was observed using Biophotometer (Eppendorf, USA) and DNA quality was checked in 0.8% agarose gel.

Amplification profiling through Real Time PCR

The identification of conserved promoter motif in the selected UR-DEGs was carried out by using MBP based Real Time PCR. The PCR amplification was performed in the total reaction volume of 15 µl (1X Taq buffer, 1 unit Taq polymarase, 0.2mM dNTPs, 3mM $MgCl_2$, 0.45µM primer, 3ng gDNA and 0.3µM MBP) at optimized PCR condition (95°C for 4 min; 40 cycles of 15 s at 95°C, 35 s at 60°C, and 45 s at 72°C). PCR amplification was carried out in Real Time PCR System (Applied Biosystems 7500 Fast Real-Time PCR Systems, USA). For the detection of GCC box in UR-DEGs, TCC box containing probe was used as a reference.

1K Promoter sequence and Protein/Genomic sequence retrieval

1K Promoter sequence of Ubiquinol Cytochrome C Chaperonee gene (Os07g30790) and Protein sequence of Sub1A was retrieved from the TIGR release version6.1 (ftp://ftp.plantbiology.msu.edu/pub/data /Eukaryotic_ Projects /o_sativa/annotation_dbs/ pseudomolecules/version_6.1/).

3D structure prediction of Sub1A protein

The 3D structure of Sub1A was not available in PDB database, therefore its structure was predicted using I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). I-TASSER is a hierarchical protein structure modeling approach based on the secondary-structure enhanced Profile-Profile threading Alignment (PPA). I-TASSER (The iterative threading assembly refinement) server determines 3D structures of protein based on the sequence-to-structure-to-function paradigm algorithm. It predicts secondary structure, tertiary structure and functional annotations on ligand-binding sites, enzyme commission numbers and gene ontology terms. The accuracy of prediction is based on the confidence score of the modeling [46,47]. C-score is a confidence score for estimating the quality of predicted models by I-TASSER. It is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations.

Construction of 3D DNA structure of promoter region having GCC-box motif

Construction of 3D DNA structure of promoter region having GCC-box motif was performed with slight modification as described by Pandey and Kumar [38]. To study protein-DNA interaction a 3D model of DNA fragment (promoter region of 25 bases having core GCCGCC motif) was required. Therefore, 3D-DART (3DNA-Driven DNA Analysis and Rebuilding Tool) server was used for generating custom 3D structural model of DNA and its PDB file. The promoter fragment of 25 nt long having core GCCGCC motif of *UCCC* gene was used for the 3D model of DNA. DNA was bended at angle of 40° with 5° tilt between 11-14 nucleotide (GCCG). 3D-DART uses the DNA rebuild functionality of software package 3DNA [48] and extends its functionality with tools to change the global

conformation of the DNA models from a sequence to a base-pair step parameter file [49].

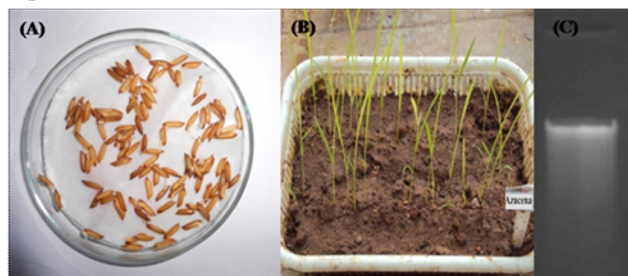


Figure 1. (A) Surface sterilized Azucena (*japonica sp.*) seeds; (B) Rice seedlings grown in tray; (C) RNase treated genomic DNA of Azucena leaves

In-silico protein-DNA interaction studies

Construction of 3D DNA structure of promoter region having GCC-box motif was performed with slight modification as described by Pandey and Kumar [38]. For *in-silico* protein-DNA interactions studies between 3D structure of Sub1A (I-TASSER generated 3D models) and 3D structure of DNA fragment having core GCC box motif (3D-DART generated models), HADDOCK web server (http://haddock.science.uu.nl/services/HdeADDOCK/haddockserver-easy.html) was used. HADDOCK (High Ambiguity Driven protein-protein Docking) is an information-driven flexible docking approach for the modeling of bimolecular complexes. HADDOCK distinguishes itself from ab-initio docking methods with the fact that it encodes information from identified or predicted protein interfaces in ambiguous interaction restraints (AIRs) to drive the docking process. These AIR files have information about active residues (directly involved in the interaction) in protein as well as in the DNA model. Result with the lowest HADDOCK score and Z-Score were considered as the best interaction between these molecules [50].

RESULTS AND DISCUSSION

Rice is the second largest produced cereal for half of the world population, posing itself as the staple food crop (http://www.irri.org/index.php). Harsh environmental conditions lead to partial to complete destruction of rice, among them flooding is considered as third major issue [51, 52]. Rice has a semiaquatic origin therefore; identification of traits associated with submergence tolerance with molecular techniques has been initiated and achieved. In this context, understanding the molecular mechanisms for submergence tolerance was very well supported by gene expression profiling through microarrays, quantitative Real-Time PCR analyses and Real-Time PCR based nucleic acid sequence detection [12, 14, 45 and 53-55] and transcriptome analysis using massively parallel signature sequencing [56, 57].

A large number of genes are differentially expressed during the submergence induced low oxygen stress involved in complex biochemical and genetic pathways [58, 9], ethylene biosynthesis [59] and enzymes encoded for sugar metabolism, glycolysis, and fermentation pathways in rice [60]. Plant promoters are the key component for studying the mechanism of transcriptional regulation. A core promoter contains the essential nucleotide sequences for the regulation of gene function known as transcriptional regulatory elements (*cis*- and *trans*- regulatory elements) and Transcription Start Site (TSS). The presence of transcriptional regulatory elements helps to regulate the function of transcription factors and their expression during normal and unfavorable (abiotic and biotic) conditions. Therefore, identification of these regulatory elements is very much essential. There are lots of reports about the identification and characterization of stress-responsive *cis*-regulatory elements [21 and 61-63].

In our previous study, consensus promoter motifs were identified using MEME that are common in their promoter region of differentially co-expressed genes in rice seedling under anoxia [20, 21]. The MEME detected GCC box (GCCGCC) as a consensus promoter motif in promoters of UR-DEGs with IC value of 16.6 bits, E-value 1.4e-056, and width length of 11 nucleotides (Figure-2). Similarly, Sharma et al. [54] identified 223 types of CREs associated with 40 rice sperm co-expressed genes by analysing 1-kb upstream regions with the help of MEME. Doi et al. [64] also identified 7514 motifs from 1-kb promoter of auxin-induced Arabidopsis prha homeobox gene using MEME.

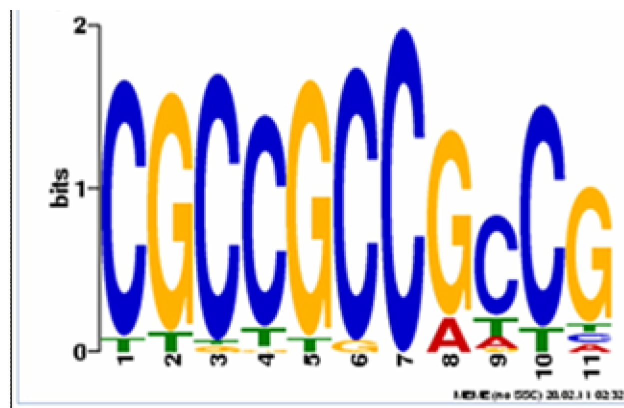


Figure 2. Identified consensus promoter motif GCC box in UR-DEGs with IC value by MEME (v 4.5.0)

Figure 3. Position of identified consensus promoter motif GCC box (green colored box) in UR-DEG ubiquinol-cytochrome C chaperone family protein gene of Eukaryotic Promoter Database Id

It was also observed that ubiquinol-cytochrome C chaperonee family protein gene was up-regulated during anoxia and also have detected GCC box motif in their promoter region (Figure-3). Gene ontology suggested that the ubiquinol-cytochrome C chaperonee family protein gene (Os07g30790) is associated with some biological process (GO:0008150) related to tissue respiration and found in mitochondrion as cellular component (GO:0005739). In principle mitochondria's

primary roles are the oxidation of organic acids through the tricarboxylic acid cycle and the synthesis of ATP coupled to the transfer of electrons from reduced NAD+ to oxygen via the electron transport chain. Beside these primary functions, it also has important secondary functions in plants, like synthesis of nucleotides, amino acids, lipids, and vitamins [65-67]. Undertaking transcription and translation [68] through their own genome [69], actively import proteins and metabolites from the cytosol [70], influence programmed cell death [71], and respond to cellular signals such as oxidative stress [72,73]. The mitochondrial respiratory chain (MRC) and oxidative phosphorylation (OXPHOS) system are composed of five enzymatic complexes (I to V) present in the inner mitochondrial membrane and two mobile electron carriers (ubiquinone and cytochrome c). Electrons are donated from reducing equivalents, NADH and FADH2, to complex I (NADH: ubiquinone oxidoreductase) and complex II (succinate: ubiquinone oxidoreductase), respectively, and flow down an electrochemical gradient in the MRC until complex IV (cytochrome c: oxygen oxidoreductase), which catalyzes the reduction of molecular oxygen, the final acceptor of electrons, to water. Complexes I, III (ubiquinol: cytochrome c oxidoreductase; cytochrome bc1 complex), and IV use the energy liberated by the electron flux to pump protons from the mitochondrial matrix to the intermembrane space, generating a proton gradient across the mitochondrial inner membrane that is used by complex V to drive the synthesis of ATP from ADP and inorganic phosphate [74, 75]. Therefore, presence and experimental detection of GCC box in UR-DEGs becomes very much essential, which was achieved through MBP based Real Time PCR, an accurate and advance molecular technique.

Consequently, MBPs and their specific primers were designed using BD7 (PREMIER Biosoft, USA). Depending upon the parameters of BD7, forward and reverse primers for Ubiquinol-cytochrome C chaperonee family protein gene were 5′-CCTCCTAGTTCGTCCGTCAA-3′ and 5′-TCGAGCCTGGACTTCACC-3′, respectively. MBP for the validation of GCC box containing UR-DEGs was 5'-[6-FAM]CGCGATCGCCGCCGCCGGATCGCG[BHQ-1]-3'.

Reporter dye 6-FAM (6 – Carboxyfluorescein) at 5′ and quencher BHQ1 (Black Hole Quencher@-1) at 3' was used for designing the MBPs.

Rice seedlings of Azucena (Japonica sp.) were grown and sampled for genomic DNA isolation using CTAB (2X) method and RNA contamination was removed by the standard protocol of RNase treatment. The concentration of genomic DNA was observed using Biophotometer (Eppendorf, USA) and DNA quality was checked at 0.8% agarose gel [Figure-1(c)].

Good quality of isolated gDNA was further used for the MBP based Real Time PCR assay with their specific primers for the identification of conserved GCC box promoter motif in up-regulated ubiquinol-cytochrome C chaperonee family protein gene. The PCR assay confirmed the presence of GCC box in the promoter region of the gene. The graph (Rn vs cycle number and Dissociation curve) generated by the inbuilt software suggested the amplification of gene with the GCC MBP (Figure-4) has been occurred having Ct values 28.03 and 29.19 for the two replicates respectively (Table-1). The Rn vs cycle number graph (Figure-4A) indicate the a' & b' curves for amplification of selected genes with GCC probe; curves c' & d' for the amplification of genes with TCC probe and e' & f' curves for non-template control (NTC). In this experiment the TCC probe was used as a reference which didn't showed any amplification suggesting the presence of GCC box (rather than TCC box) in the promoter of ubiquinol-cytochrome C chaperonee family protein gene. Similar works has also been reported on nucleic acid sequence detection, sensitivity, accuracy and reliability of MBP [43, 44 and 76-80]. Similarly, a dissociation curve (Figure-4B) for the amplification of GCC and TCC box was also prepared indicating the amplification of only two specific products. Dissociation curves a′′ & b′′ depict the two independent replications for the amplification of GCC probe while c′′ & d′′ for TCC probe (Figure-4B).

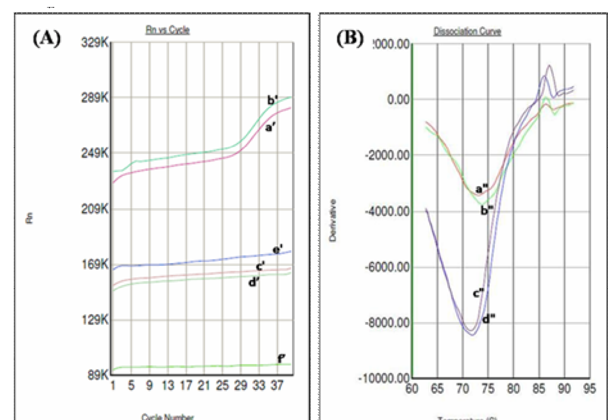


Figure 4. Molecular Beacon based Real Time PCR result for ubiquinol-cytochrome C chaperone family protein gene having GCC box in its promoter region (-499 to +100). (A) Graph indicating relationship between Rn vs cycle number. Curves a' and b' depict amplification of gene with GCC probe. Curves c' & d' are the amplification of TCC probe. Curves e' & f' are amplification of NTC. (B) a′′ & b′′ depict the dissociation curve of gene with GCC probe and c′′ & d′′ are dissociation curves of gene with TCC probe.

Confirmation about the presence of GCC Box in the promoter region prompted to further study about their interaction with Transcription Factor during submergence. Chakravarthy et al. [19] reported that tomato transcription factor Pti4 (an ERF) is involved in the regulation of gene expression by interacting with GCC box or non-GCC box cis-elements. AtEBP binding with an oligonucleotide probe containing a mutant GCC box (GCC box contained two points Mutations) eliminate the ability of a cis regulatory element (47-bp fragment that contains two copies of the GCC box) to

activate gene expression in an ethylene-dependent manner [81]. It was reported that GCC-box work as an ethylene-responsive element that is essential in some cases for the regulation of transcription [81]. Hao et al. [82] reported that numerous members of the ERF family interact specifically with AGCCGCC through the conserved ERF domain. Fujimoto et al. [18] described that maltose binding protein –AtERF fusion proteins bind with GCC box sequence (AGCCGCC) and binding activity was abolished when both G residues within the GCC box were replaced by T residues (ATCCTCC). Similarly, using electrophoresis mobility shift assay (EMSA) Cheong et al. [83] concluded that OsEREBP1 specifically binds to the GCC box (AGCCGCC) motif but not to the mutated GCC box (ATCCTCC). Additionally, ethylene response factors are also involved in regulating jasmonate-responsive gene expression by interacting with the GCC-box. And introduction of point mutations into GCC-box sequence substantially reduced jasmonate responsiveness [84]. Using EMSA and transient expression assay TiERF1 protein binds with GCC box and modulate the defense response by up-regulating transcripts of a subset of genes with the GCC box present in their promoters [85].

Table 1. Ct value chart of gene AK068288 with NTC and negative control

S. No	Replicates	Template	Molecular Beacon	Ct value
a	R1	Template	GCC box	28.03
b	R2	Template	GCC box	29.19
c	R1	Template	TCC box	-
d	R2	Template	TCC box	-
e	R1	No	GCC box	-
f	R2	No	GCC box	-

Table 2. Best predicted model with their C-Score, TM Score and RMSD value where C-Score is the confidence score for the predicted model, TM-score is a measure of global structural similarity between query and template protein and Root Mean Square Deviation is the RMSD between residues that are structurally aligned by TM-align

Best Predicted Model				
Best Model	**Locus Id**	**C Score**	**TM score**	**RMSD value (Å)**
Model 1	Os09g11480	-4.06	0.28±0.09	15.8±3.2Å

It is well known that Sub1A gene was responsible for the submergence tolerance in rice [16]. Therefore the protein sequence and structure of Sub1A was required. The protein sequence of Sub1A was retrieved from the TIGR (v6.1) (ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudo molecules/version_6.1/). The 3D structure of Sub1A was not available in PDB database, therefore its structure was predicted using I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). I-TASSER server predicts and displays various features in different sections for best model studies. It was considered that the prediction and generation of the best model based on C-Score, their structural analogs and binding sites. The quality of the generated models are estimated based on a confidence score (C-score), ranges from -5 to 2 where a high value signifies a model with a high confidence and vice-versa. C-score is highly correlated with Tm score and RMSD. Therefore, TM-score and RMSD are known standards to measure the accuracy of structure modeling thereby measuring structural similarity between two protein structures. RMSD is an average distance of all residue pairs in two structures and is sensitive to local errors (i.e., a mis-orientation of the tail) which occurs inspite of the correct global topology hence, TM-score must be used for solving these errors A TM-score >0.5 indicates a model of correct topology. Roy et al. [86] predicted the structures of three human GPCRs complexes using I-TASSER with a RMSD's 1.6A°, 2.27A° and 2.82A° to the crystal structures in the transmembrane region. The models predicted by I-TASSER were based on the best 10 threading templates available on RCSB PDB. The best predicted model is selected on the basis of confidence score; TM-Score as well as RMSD value (Table -2). The C score value for the best predicted model which is model 1 of Sub1A was -4.06 and furthermore, highly similar structures in PDB (as identified by TM-align) were identified and listed in Table-3. Template proteins with similar binding sites for Sub1A are listed in Table -4. The best binding site is predicted on the basis of C score LB (Range = 0-1) and BS-Score (>1) values. A higher score C score indicates a more reliable ligand-binding site prediction and BS-score reflects a significant local match between the predicted and template binding site [46,87]. Qin and Zhou [29] suggested that binding site prediction is a useful tool for building structural models for protein-DNA complexes and for experimental design and validation. Two best predicted binding sites for Sub1A 3D was taken for further interaction studies (Table-4). Structure of Sub1A protein predicted by I-TASSER and visualized by VMD tool has been shown in Figure-5.

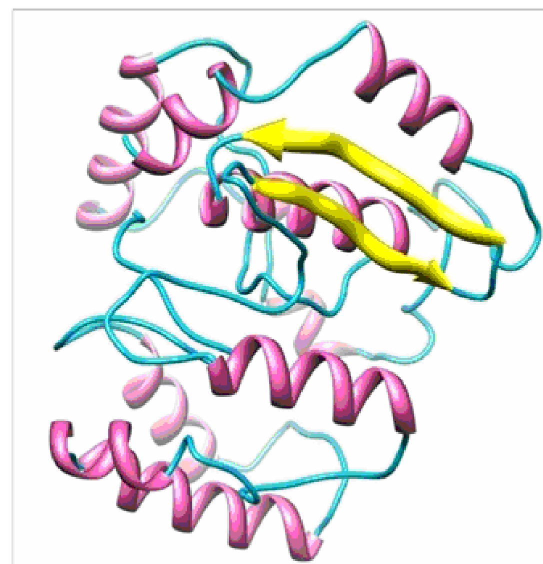


Figure 5. Result showing 3D structure of Sub1A protein predicted by I-TASSER. The coloring method is based on secondary structure. The pink color represents α-helix and yellow color represents β-strand and deep sky blue color represents the coil in the 3D structure.

Table 3. Identified best two structural analogs in PDB where coverage represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein. Coverage represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein

Top 2 Identified structural analogs in PDB				
PDB Hit	TM- Score	RMSD (Å)	IDEN[a]	Cov.
4he8L	0.443	5.67	0.038	0.72
3rkoL	0.441	5.75	0.046	0.724

For *in-silico* protein-DNA interaction studies 3D structure of protein along with DNA was required. Consequently, 3D structure of DNA segment (25 nt long) containing identified and validated GCC Box promoter motif of *UCCC* gene (Os7g30790) was generated by 3D-DART server. GCC-Box promoter motif positioned at 10-15 nucleotides have been shown in Figure-6. DNA was bended at angle of 40° with 5° tilt between 11-14 nucleotides (GCCG).

Figure 6. 3D structure of linear DNA segment generated by 3D-DART server of GCC-Box promoter motif positioned at 8-17 nucleotide. In DNA model red color represents Adenine and pink color represents Guanine and sea green color represents cytosine and gold yellow color represents Thymine.

Protein-DNA interactions are the physical basis of gene expression and DNA modification for vital biological activities [29]. Because there is no simple mapping code between DNA base pairs and protein amino acids, the prediction of protein-DNA interactions is a challenging problem. Therefore, HADDOCK can make use of a broader array of restraints, including those derived from biochemical and biophysical data [37]. Determining the structure of protein-DNA complexes and clarifying the factors that regulating their interaction is essential to better understand many biological processes [30]. A review describing the experimental strategies currently employed to solve structures of protein–DNA complexes and to analyze their dynamics has been published [88]. Protein–DNA interactions facilitate the fundamental functions of living cells and are universal in all living organisms [89].

To determine the protein-DNA interactions the Easy interface of HADDOCK web server was used [50]. Before going for docking AIR files were generated for both the interacting molecules having information about the active binding sites of Sub1A protein as well as in the DNA model. The HADDOCK score is the weighted sum of van der Waals energy (negative indicating favorable interactions), electrostatic energy (negative indicating favorable interactions), distance restraints energy (only unambiguous and AIR (ambig-restraints), direct RDC restraint energy, inter vector projection angle restraints energy, diffusion anisotropy energy, dihedral angle restraints energy, symmetry restraints energy, buried surface area (negative weight indicate a better interface), binding energy, desolvation energy. Meanwhile, the solution structures are analyzed for their intermolecular hydrogen bonds and intermolecular hydrophobic contacts by HADDOCK, the solutions are clustered according to the interface ligand RMSDs. The Z-score indicates the standard deviations from the average of a particular cluster in terms of HADDOCK score.

Table 4. Template protein for similar binding sites. Binding sites represents the amino acid positions

Template Protein with similar binding site							
Cscore[LB]	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Coverage	BS-score	Binding Site
0.31	1gccA	0.236	1.5	0.629	0.248	1.67	102, 103, 104, 106, 108, 116, 118, 120, 141
0.26	1gccA	0.236	1.5	0.629	0.248	1.76	106, 108, 110, 112, 118, 125, 127, 129, 130

Table 5. HADDOCK score of protein and UCCC gene (Os7g30790) segment containing GCC-Box depicted by HADDOCK server.

HADDOCK Table								
Model	HADDOCK Score	RMSD	Van der Waals Energy	Electrostatic Energy	Desolvation Energy	Restraints Violation Energy	Buried Surface Area	Z -Score
SAUGCM1-BS1	33.1 ± 3.8	8.5 ± 0.6	-55.0 ± 7.3	-640.1 ± 63.6	44.1 ± 10.6	1719.7 ±29.43	1645.3 ± 95.6	-1.2
SAUGCM1-BS2	12.4 ± 5.7	16.6 ± 0.6	-61.9 ±3.3	-613.6 ± 30.2	32.0 ± 6.8	650.2 ± 19.23	1912.0± 117.0	-1.2
SAUGCM2-BS1	27.8 ± 3.6	8.7 ± 0.2	-60.5 ± 5.9	-574.1 ± 24.4	34.4 ± 2.5	1687.2 ±9.37	1741.7±115.4	-1.2
SAUGCM2-BS2	11.7 ± 4.4	0.9 ± 0.5	-56.2 ± 3.0	-597.9 ± 63.0	18.5 ± 5.7	1689.9 ±39.38	1804.5 ± 70.5	-1.7

For the prediction of best interaction, different binding sites and models of Sub1A were docked with 3D structure of DNA segment having promoter motif of *UCCC* gene (Table-5). The result showed best model SAUGCM2-BS2 (Table-5) has HADDOCK score of 11.7 +/- 4.4 and Z-score of -1.7 when interacting with bended DNA segment (Figure-7).

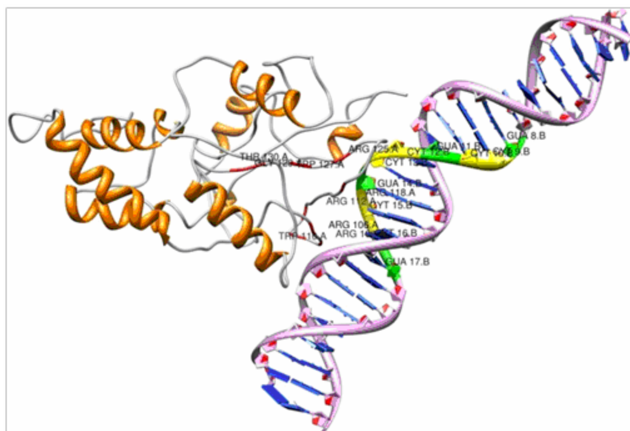


Figure 7. Result showing interaction between 3D structure of Sub1A protein predicted by I-TASSER and DNA model (promoter region having complementary GCCGCC core motif, CGGCGG) predicted by 3D-DART. In DNA model yellow color represents Cytosine at 9, 10, 12, 13, 15 and 16 position while green color represents Guanine at 8, 11, 14, and 17. Interacting residues (THR at 130 and TRP at 110, 127 and ARG at 106,108, 112, 118, 125 and GLY at 129 positions respectively) in protein model represented with red color.

CONCLUSION

The aim of the present study was to identify consensus promoter motifs and their interaction with Sub1A. Previous study suggested that the consensus promoter motif GCC box sequence was found in the up-regulated promoter of Ubiquinol the cytochrome c chaperone family gene. Identification of consensus promoter motif in the promoter of up-regulated gene was done by using MBP based real time PCR. Real time PCR result showed that the presence of GCC box in promoter sequence through amplification whereas no template controls (NTC) and negative control (TCC box in promoter sequence) showed no amplification. *In Silico* study on 3D structure prediction and protein-DNA interaction of *UCCC* gene in rice has been studied and revealed that core GCC box (GCCGCC) was identified in antisense strand of *UCCC* gene promoter. It predicted a good binding affinity of Sub1A protein with Core GCC box promoter motif. Thus it can be concluded that Computational analysis for Sub1A protein and interaction with promoter motif have been performed successfully. Furthermore, to confirm the above interaction validation study is very much required in future.

Contribution by Authors

Gopal Kumar Prajapati and Neha Kashyap performed the experimental and data analysis. Ashutosh Kumar helped in technical aspects of experimental and data analysis and drafting of the manuscript. Dr. Dev Mani Pandey helped in concept development, checking the updates at regular intervals, interpretation of result and editing of the manuscript. All authors read and approved the final manuscript.

Acknowledgement

This work was financially supported by University Grants Commission 10 (UGC), GOI, New Delhi for the major research project [F. No. 37-113/2009 (SR)]. Mr. Gopal Kumar Prajapati is working as Project Fellow and financially supported from the above mentioned UGC sponsored project. Use of Beacon designer7 software for designing of probe was generously provided by PREMIER Biosoft, USA. Central Rainfed Upland Rice Research Station (CRURRS), Hazaribagh-16 825301, Jharkhand, India is gratefully acknowledged for providing the Azucena rice seeds used in this experiment.

REFERENCES

1. Khush GS. (2005). What it will take to feed 5.0 billion rice consumers by 2030, Plant Mol.Biol. 59: 1-6.
2. Liu F, Xu W, Wei Q, et al. (2010). Gene Expression Profiles Deciphering Rice Phenotypic Variation between Nipponbare (Japonica) and 93-11 (Indica) during Oxidative Stress. PLoS ONE 5(1) doi:10.1371/journal.pone.0008632.
3. Lafitte HR, Ismail A, Bennet J. (2004). Abiotic stress tolerance in rice for Asia: progress and the future. In: Fischer T, Turner N, Angus J, McIntyre L, Robertson M, Borrell A, et al. eds. New Directions for a Diverse Planet: Proceedings for the 4th International Crop Science Congress. The Regional Institute Ltd. The proceedings are available online at: www.cropscience.org.au/ csc2004.
4. Munns R, Tester M. (2008). Mechanisms of salinity tolerance. Annu Rev Plant Biol. 59: 651–681.
5. Agarwal S, Grover A. (2005). Isolation and transcription profiling of low-O_2 stress-associated cDNA clones from the flooding-stress-tolerant FR13A rice genotype. Ann. Bot. 96: 831–844.
6. Fukao T, Xu K, Ronald PC, Bailey-Serres J. (2006). A variable cluster ethylene response factor-like genes regulates metabolic and developmental acclimation responses to submergence in rice. Plant Cell. 18: 2021–2034.
7. Hattori Y, Nagai K, Ashikari M. (2011). Rice growth adapting to deepwater. Current Opinion in Plant Biology.14:100–105.
8. Fukao T, Bailey-Serres J. (2004). Plant responses to hypoxia—is survival a balancing act? Trends Plant Science. 9: 449–456.
9. Vartapetian BB and Jackson MB. (1997). Plant adaptations to anaerobic stress. Ann. Bot. 79:3–20.
10. Liu F, Vantoai T, Moy LP et al. (2005). Global transcription profiling reveals comprehensive insights in to hypoxic response in Arabidopsis. Plant Physiol.137:1115-1129.
11. Baxter-Burrell A, Yang Z, Springer PS et al. (2002). RopGAP4-dependent Rop GTPase rheostat control of Arabidopsis oxygen deprivation tolerance. Science. 296:2026–2028.
12. Lasanthi-Kudahettige R, Magneschi L, Loreti E et al. (2007). Transcript profiling of the anoxic rice coleoptile. Plant Physiol. 144: 218–231.
13. Mukhopadhyay A, Vij S, Tyagi AK. (2004). Overexpression of a Zinc-finger protein gene from rice confers tolerance to cold, dehydration, and salt stress in transgenic tobacco, Proc. Natl. acad. Sci. Usa. 101:6309-6314.
14. Pandey DM, Kim SR. (2006). Isolation of the low oxygen stress genes in rice (*Oryza sativa L.)* Proceeding of 4th International Symposium on Rice Functional Genomics (ISRFG'2006) Montpellier, France.Poster .37: 96.
15. PandeyDM, Deb R, Kumar A et al. (2007) Study on the functional characterization of low oxygen stress inducible gemnes in rice((*Oryza sativa L)* Proceeding of 5th Intewrnational Symposium on Rice Functional Genomics (ISRFG'2007) Epochal, Tsukuba , Japan Poster 110.
16. Xu K, Xu X, Fukao T et al. (2006). Sub1A is an ethylene response-factor-like gene that confers submergence tolerance to rice. Nature. 442:705-708.
17. Singh N, Dang TTM, Vergara et al. (2010). Molecular marker survey and expression analyses of the rice submergence-tolerance gene *SUB1A*. Theor. Appl. Genet. 121: 1441-1453.
18. Fujimoto SY, Ohta M, Usui A et al. (2000). Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. Plant Cell. 12:394-404.
19. Chakravarthy S, Tuori RP, Ascenzo D et al. (2003). The tomato transcription factor Pti4 regulates defencse-related gene expression via GCC box and non-GCC box cis elements. Plant Cell. 15:3033-3050.
20. Kumar A, Shankaracharya, Vidyarthi AS et al. (2007). *In-silico* analysis of promoter motif of anoxia inducible genes in rice *Oryza sativa* L.). In: 5th International Symposium on Rice

Functional Genomics at Epochal, Tsukab, Japan scheduled on 15-17 October, 2007. Poster No-112.
21. Kumar A, Sharma V, Smita S et al. (2009). *In Silico* analysis of motifs in promoters of differentially expressed genes in rice (*Oryza sativa* L.) under anoxia. International J. Bioinfo. Res. and Application. 5(5): 523-547.
22. Kemmerer EC, Leim WuR. (1991). Isolation-ard Molecular evolutionary analysis of a cytochrome c genes: surprising implications for *Arabidopsis thaliana.*J Mol Evol. 32:227-237.
23. Gopalakrishnan L, Scarpulla RC. (1994). Differential regulation of respiratory chain subunits by a CREB-dependent signal transduction pathway. J Biol Chem. 1994: 269:105–113.
24. Evans MJ, Scarpulla RC. (1990). NRF-1: a *trans*-activator of nuclear-encoded respiratory genes in animal cells. Genes Dev.4:1023–1034.
25. Jang IC, Choi WB, Lee KH et al. (2002). High-level and ubiquitous expression of the rice cytochrome c gene OsCc1 and its promoter activity in transgenic plants provides a useful promoter for transgenesis of monocots. Plant Physiol. 129: 1473–1481.
26. Ellis Rj, Hartl FU. (1996). Protein folding in the cell: competing models of chaperonin function. FASEB J. 10: 20-26.
27. Vierling E. (1991) The role of heat shock proteins in plants. Annu Rev Plant Physiol Plant Mol Biol. 42:579–620.
28. Lindquist S, Craig EA. (1988). The heat-shock proteins. Annu. Rev. Genet. 22: 631-677.
29. Qin S, Zhou HX. (2011). Structural Models of Protein-DNA Complexes Based on Interface Prediction and Docking. Current Protein and Peptide Science. 12: 531-539.
30. Chou CC, Rajasekaran M, Chen C. (2010). An effective approach for generating a three Cys2His2zinc-finger-DNA complex model by docking. BMC Bioinformatics. 11: 334.
31. Tomovic A, Oakeley EJ. (2007). Position dependencies in transcription factor binding sites. Bioinformatics. 23: 933–941.
32. Setny P, Bahadur RP, Zacharias M. (2012). Protein-DNA docking with a coarse-grained force field. BMC Bioinformatics. 13: 228.
33. Banitt I, Wolfson HJ. (2011). ParaDock: a flexible non-specific DNA–rigid protein docking algorithm. Nucleic Acids Res. 39(20):135.
34. Aloy P, Moont G, Gabb HA et al. (1998). Modelling repressor proteins docking to DNA. Proteins. 33(4): 535–549.
35. van Dijk M, van Dijk ADJ, Hsu V et al. (2006). Information-driven protein- DNA docking using, HADDOCK: it is a matter of flexibility. Nucleic Acids Res. 34(11): 3317–3325.
36. van Dijk, M, Bonvin AMJJ. (2010). Pushing the limits of what is achievable in protein-DNA docking: benchmarking HADDOCK's performance. Nucleic Acids Res. 38(17): 5634–5647.
37. Kobayashi M, Eiso AB, Bonvin AMJJ et al. (2010). Structure of the DNA-bound BRCA1 C-terminal Region from Human Replication Factor C p140 and Model of the Protein-DNA Complex. The J Biological Chem. 285 (10):10087-10097.
38. Pandey DM, Kumar A.(2013). 3D structure prediction and protein-DNA interaction of CCCH-type zinc finger transcription factor gene in Rice (*Oryza sativa* L.) Int. J. Comp. Bioinfo. In Silico Modeling. 2(2):94-103.
39. Sundaralingam M, Burkhart B. (1997). Protein–nucleic acid interactions: general principles. In: Parthasarathy S and Glusker JP (eds) Aspects of Crystallography in Molecular Biology. 363–390. New Delhi: New Age International Publishers.
40. Luscombe N, Austin SE, Berman HM et al. (2000). An overview of the structures of protein–DNA complexes. Genome Biology. 1:1–37.
41. Chen J, Tang WH, Hong MM et al. (2003). OsBP-73, a rice gene, encodes a novel DNA-binding protein with a SAP-like domain and its genetic interference by double-stranded RNA inhibits rice growth. Plant Mol Biol. 52(3):579-90.
42. Tyagi S, Kramer FR. (1996). Molecular beacons: probes that fluoresce upon hybridization. Nature Biotechnology.14:303-308.
43. Pas SD, Noppornpanth S, van der Eijk AA et al. (2005). Quantification of the newly detected lamivudine resistant YSDD variants of Hepatitis B virus using molecular beacons. J. Clin. Virol. 32:166–172.
44. Ye Q, Zhuang H, Zhou C. (2009). Detection of naphthalene by real-time immuno-PCR using molecular beacon. Mol. Cell Probes. 23:29-34.
45. Lata P, Ram S, Agrawal M et al. (2009). Real Time PCR for the Rapid Detection of vanA Gene in Surface Waters and Aquatic Macrophyte by Molecular Beacon Probe. Environ. Sci.Technol. 43:3343–3348.
46. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9: 40.
47. Zhang C, Zhang F, Zhou J et al. (2012). Overexpression of a phytochrome-regulated tandem zinc finger protein gene, OsTZF1, confers hypersensitivity to ABA and hyposensitivity to red light and far-red light in rice seedlings. Plant Cell Rep. 31:1333–1343.
48. Lu XJ, Olson WK (2003). 3DNA: a software package for the analysis rebuilding and visualization of three-dimensional nucleic acid structures, Nucleic Acids Res. 31: 5108–5121.
49. van Dijk M, Bonvin AMJJ. (2009). 3D-DART: a DNA structure modelling server. Nucl. Acids Res. 37:235-239.
50. de Vries SJ, van Dijk M, Bonvin AMJJ. (2010). The HADDOCK web server for data-driven biomolecular docking. Nature Protocols. 5: 883-897.
51. Sarkar RK, Reddy JN, Sharma SG et al. (2006). Physiological basis of submergence tolerance in rice and implications for crop improvement. Current Science. 91: 899-906.
52. Wassmann R, Jagadish SVK, Heuer S et al. (2009). Production: The Physiological and Agronomic Basis for Possible Adaptation Strategies. In Donald L. Sparks, editor, Advances in Agronomy. Burlington: Academic. 101: 59-122.
53. Tyagi S, Baratu D, Kramer FR. (1998). Molecular beacons: probes that fluoresce upon hybridization. Nature Biotechnology. 16: 349-353.
54. Sharma N, Russell SD, Bhalla P et al. (2011). Putative cis-regulatory elements in genes highly expressed in rice sperm cells. BMC Res. Notes. 4: 319.
55. Mittal D, Madhyastha DA, Grover A. (2012). Genome-Wide Transcriptional Profiles during Temperature and Oxidative Stress Reveal Coordinated Expression Patterns and Overlapping Regulons in Rice. PLoS ONE 7(7): e40899. doi:10.1371/journal.pone.0040899.
56. Nakano T, Suzuki K, Fujimura T et al. (2006). Genome-wide analysis of the ERF gene family in Arabidopsis and Rice. Plant Physiol. 140: 411–432.
57. Nobuta K, Venu RC, Lu C et al. (2007). An expression atlas of rice mRNAs and small RNAs. Nature Biotechnol. 25: 473-477.
58. Sachs MM, Subbaiah CC, Saab IN (1996). Anaerobic gene expression and flooding tolerance in maize: glycolytic and XET genes and signal transduction, J. Exp. Bot. 47: 1-15.
59. Nie X, Singh RP, Tai GC. (2002). Molecular characterization and expression analysis of 1-aminocyclopropane-1-carboxylate oxidase homologs from potato under abiotic and biotic stresses. Genome. 45: 905–913.
60. Huang SB, Greenway H, Colmer TDet al. (2005). Protein synthesis by rice coleoptiles during prolonged anoxia: implications for glycolysis, growth and energy utilization. Ann. Bot. (Lond) 96: 703-715.
61. Narusaka Y, Nakashima K, Shinwari ZK et al. (2003). Interaction between two *cis*-acting elements, ABRE and DRE, in ABA-dependent expression of Arabidopsis rd29A gene in response to dehydration and high salinity stresses, Plant J. 34:137-148.
62. Yamaguchi-Shinozaki K and Shinozaki K (2005) Organization of *cis*-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci. 10: 88-94.
63. Xiang Y, Huang Y, Xiong L. (2007). Characterization of Stress-Responsive CIPK Genes in Rice for Stress Tolerance Improvement. Plant Physiol. 144: 1416-1428.
64. Doi K, Hosaka A, Nagata T et al. (2008). Development of a novel data mining tool to find *cis*-elements in rice gene promoter regions. BMC Plant Biology. 8: 20.
65. Rebeille, F, Macherel D, Mouillon JM et al. (1997). Folate biosynthesis in higher plants: Purification and molecular cloning of a bifunctional 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase/7,8-dihydropteroate synthase localized in mitochondria. EMBO J. 16:947–957.
66. Bartoli C, Pastori G, Foyer C. (2000). Ascorbate biosynthesis in mitochondria is linked to the electron transport chain between complexes III and IV. Plant Physiol. 123:335–343.

67. Gueguen V, Macherel D, Jaquinod M et al. (2000). Fatty acid and lipoic acid biosynthesis in higher plant mitochondria. J. Biol. Chem. 275: 5016–5025.
68. Brennicke A, Marchfelder A, Binder S. (1999). RNA editing. FEMS Microbiol. Rev. 23: 297–316.
69. Gray MW, Burger G, Lang BF. (1999). Mitochondrial evolution. Science 283: 1476–1481.
70. Lister R, Murcha MW, Whelan J. (2003). The Mitochondrial Protein Import Machinery of Plants (MPIMP) database. Nucleic Acids. Res. 31:325–327.
71. Balk J, Leaver CJ. (2001). The PET1-CMS mitochondrial mutation in sunflower is associated with premature programmed cell death and cytochrome *c* release. Plant Cell 13: 1803–1818.
72. Moller IM. (2001). Plant mitochondria and oxidative stress: Electron transport, NADPH turnover, and metabolism of reactive oxygen species. Annu. Rev. Plant Physiol. Plant Mol. Biol. 52: 561–591.
73. Sweetlove LJ, Heazlewood JL, Herald V et al. (2002). The impact of oxidative stress on Arabidopsis mitochondria. Plant J. 32: 891–904.
74. Richter OMH, Ludwig B. (2003). Cytochrome c oxidase – structure, function, and physiology of a redox-driven molecular machine. Rev Physiol Biochem Pharmacol. 147:47–74.
75. Fontanesi F, Soto IC, Horn D et al. (2006). Assembly of mitochondrial cytochrome c-oxidase, a complicated and highly regulated cellular process. Am J Physiol Cell Physiol. 291:1129–1147.
76. Leone G, van Schijndel H, van Gemen B et al. (1998). Molecular beacon probes combined with amplification by NASBA enable homogeneous, real-time detection of RNA. Nucleic Acids Res. 26: 2150-2155.
77. Marras SAE, Kramer FR, Tyagi S. (1999). Multiplex detection of single-nucleotide variations using molecular beacons. Genet. Anal. Biomol. Eng. 14:151-156.
78. Piatek AS, Tyagi S, Pol AC et al. (1998). Molecular beacon sequence analysis for detecting drug resistance in *Mycobacterium tuberculosis*. Nat. Biotechnol. 16: 359-363.
79. Elsayed S, Chow BL, Hamilton NL et al. (2003). Development and validation of a molecular beacon probe-based real-time polymerase chain reaction assay for rapid detection of methicillin resistance in *Staphylococcus aureus.* Arch. Pathol. Lab. Med. 127:845-849.
80. Elsayed S, Plewes K, Church D et al. (2006). Use of Molecular Beacon Probes for Real-Time PCR Detection of Plasmodium falciparum and Other Plasmodium Species in Peripheral Blood Specimens. J. Clin. Microbiol. 44 : 622-624.
81. Ohme-Takagi M, Shinshi H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell. 7: 173–182.
82. Hao DY, Ohme-Takagi M, Sarai A. (1998). Unique mode of GCC box recognition by the DNA- binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. J. Biol. Chem. 273: 26857–26861.
83. Cheong YH, Moon BC, Kim JK et al. (2003). BWMK1, a Rice Mitogen-Activated Protein Kinase, Locates in the Nucleus and Mediates Pathogenesis-Related Gene Expression by Activation of a Transcription Factor. Plant Physiol. 132:1961–1972.
84. Brown RL, Kazan K, McGrath KC et al. (2003). A Role for the GCC-Box in Jasmonate-Mediated Activation of the PDF1.2 Gene of Arabidopsis. Plant Physiol. 132: 1020–1032.
85. Liang HX, Lu Y, Liu HX et al. (2008). A novel activator-type ERF of Thinopyrum intermedium, TiERF1, positively regulates defense responses. J. Exp. Bot. 59(11):3111–3120.
86. Roy A, Yang J, Zhang Y. (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucl Acids Res. 40: 471-477.
87. Roy A, Kucukural A, Zhang Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols. 5: 725–738.
88. Campagne S, Gervais V, Milon A. (2011). Nuclear magnetic resonance analysis of protein–DNA interactions. J. R. Soc. Interface. 8: 1065-1078.
89. Sathyapriya R, Vishveshwara S. (2004). Interaction of DNA with clusters of amino acids in proteins. Nucl. Acids Res. 32 (14): 4109-4118.

© 2013; AIZEON Publishers; All Rights Reserved

This is an Open Access article distributed under the terms of the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*****

Innovare
Academic Sciences

International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 7, Issue 3, 2015

Original Article

MOLECULAR BEACON PROBE BASED PROMOTER MOTIFS VALIDATION IN ANOXIA RESPONSIVE DIFFERENTIALLY EXPRESSED GENES AND THEIR *IN SILICO* INTERACTION STUDIES WITH AP2/EREBP TF IN RICE (*ORYZA SATIVA* L.)

GOPAL KUMAR PRAJAPATI, DEV MANI PANDEY*

Department of Bio-Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand-835215
Email: dmpandey@bitmesra.ac.in

Received: 10 Dec 2014 Revised and Accepted: 30 Dec 2014

ABSTRACT

Objective: Progressive evolution in molecular biology revealed the differential expression of genes and their regulatory mechanism in rice under anoxia. In addition to that the consensus promoter motifs (GCC and TCC box) were identified in differentially expressed genes (DEGs) from microarray analysis through *in silico* study. These promoter motifs need to be validated and their interaction study with the transcription factors (TFs) are essential.

Methods: To unravel the regulatory mechanism in rice during anoxia, we identified and validated the promoter motifs through Molecular Beacon Probes (MBP) based Real Time PCR. *In silico* protein-DNA interaction was studied between highly up-regulated APETALA2/Ethylene-responsive element binding proteins (AP2/ERBP) TF under anoxia and validated promoter motifs through the HADDOCK and SiteMap module.

Results: It was identified that consensus promoter motif GCC and TCC box were present in highly up-regulated methyl-transferase domain containing protein gene (*MT*) and down-regulated RhoGAP domain containing protein gene (*RG*), respectively.

Conclusion: These promoter motifs were validated through MBP and further their interaction with AP2/ERBP shows the significant binding affinity towards GCC and TCC box present on *MT* and *RG*, respectively.

Keywords: DEGs, Anoxia, DEG, MBP, HADDOCK, SiteMap.

INTRODUCTION

Rice is one of the most important basic food crops. More than half of the world's population depends upon rice [1]. However, abiotic stress is a major limiting factor of rice productivity worldwide [2]. Abiotic stress like submergence, drought, salinity, cold, anoxia are the most prominent factors which affect the plant growth and development. Importantly, among natural hazard flooding plays most hazardous role for the standing crop leads to death during complete submergence for 1 to 2 weeks of most rice cultivars [3]. In consequence of submergence rice plant suffers from oxygen deficiency. Germination of the rice coleoptile under anoxia is highly infrequent characteristic by extending the coleoptile, above the water surface is a key feature of rice to sustain under anaerobic condition [4-6]. However, the low oxygen stresses regulate the different metabolic pathways and differential expression of genes [7]. Moreover, various studies in response to anoxia [8, 9], anaerobic response elements (AREs) with their binding sites [10] and factors regulating the wide range of differential expression of genes in anoxic rice coleoptile have also been reported [8]. However, under anoxia it is still largely unknown key regulatory mechanisms of rice coleoptile germination and elongation along with the differential expression of the genes.

During rice germination and coleoptile elongation TFs MYB, zip, AP2/ERF and ZnF play a potential role in controlling the transcription of sucrose metabolism and fermentation genes under anaerobic condition [6]. Moreover the differentially expressed genes (DEGs) and TF family, including AP2-EREBP, MYB, bHLH, WRKY, zip and NAC were identified which are involved in salinity and submergence stresses [11]. More specifically, AP2-EREBP/ERFs TF has been found to be involved in growth, development, metabolic regulation under biotic and abiotic responses [12]. This superfamily TF divided into subfamily TF AP2, RAV, CBF/DREB and ERF genes, which play a variety of roles throughout the plant life cycle and key regulator in various biotic and environmental stresses. TFs encoded by AP2/EREBP genes contain the highly conserved AP2/ERF DNA binding domain [13-14]. However, proteins encoded by ERF subfamily genes bind to the core motif AGCCGCC (GCC box) mainly a response to pathogenesis and wounding [15-17]. A similar study was reported on tomato, Ethylene- Responsive Factor (ERF) transcription factor Pti4, which binds the GCC box (*cis*-element) that is present in the promoters of many Pathogenesis-Related (PR) genes [18]. Whereas, CBF/DREB ERF subfamily gene TF also recognizes C-repeats *cis*-acting element, A/GCCGAC, which is often associated with ABA, drought and cold responsive genes [19, 20]. Additionally, in rice Submergence1 (*Sub1*) locus encoding three ERF transcriptional regulators. *Sub1* TF gene is a key breakthrough for the submergence tolerance, which often regulates other genes by binding to their consensus promoter motif, GCC box [21]. Similarly, it has been reported from the promoters of anaerobic stress responsive genes statistically significant, common and consensus promoter motifs are detected by *in silico* analysis in majority of promoters [22]. Further, *in silico* study of anoxia coleoptile rice microarray data [8] reveals that consensus promoter motif GCC box (GCCGCC) and the TCC box (TCCTCC) was highly conserved in the promoter of up-regulated differently expressed genes (DEGs) and down-regulated DEGs respectively [23]. Furthermore, GCC in the UR-DEG promoter of Ubiquinol Cytochrome C chaperone gene (*UCC*) identified by MEME (v 4.5.0; http: //meme. nbcr. net/meme 4_5_0/cgi-bin/meme. cgi) online tool and their validation done through MBP based on Real Time PCR [24].

Several probes based techniques have been reported like Molecular beacon, Minor groove binging (MGB) assays used to identify the specific sequences in the nucleic acids based on the RealTime PCR [24-28]. More specifically, MBP increases the sensitivity and precision over the conventional PCR without post-reaction analysis for the detection and as well as quantification of target genes [29].

The genes and their regulatory TF are central to the expression of the functional genes under abiotic condition. Remarkably, genes facilitate their expression by binding different transcription factor in the promoter region. Hence the interaction of the TFs with the *cis*-element are the key feature of the gene regulation and its expression [30]. In these complexes amino acids and nucleotide sequences have participated in the formation of the interactive structure of Protein-DNA complexes, which determine the many functional

characteristics [31]. On the basis of sequence and structural information several methods and software have been developed for prediction of the binding site and interacting residues in TFs [32,33]. Subsequently the prediction of complex structure and their active binding site through computational approach are becoming progressively important [34-36]. Moreover, several approaches like geometric hashing method [37], Fast Fourier, correlation techniques [38] and HADDOCK program [39-40] used for the Protein-DNA interaction study.

The intermolecular docking study reported the *Brassica Napus* DREB1 protein has a GCC binding domain which bind to six nucleotides GCC box (A/GCCGAC) [41]. Similarly, Haddock used for the interactive study of the CCCH-type Zinc finger transcription factor gene and OsCCCH-Zn-1protein [42]. Furthermore, the similar interaction study of *sub 1* gene protein with the GCC box promoter motif of *UCC* done through HADDOCK server in rice [24]. In Arabidopsis, *HARDY* (AtHRD) gene has Ap2/ERF domain had docked with the GCC box promoter motifs of several drought responsive genes [43]. Since the interaction of the regulatory protein and DNA involves the *cis*-element in the regulation of the various biological processes, hence these complex biological protein DNA structures need to be recognized. In this study, we validated and identify the GCC and TCC box promoter motif by using Molecular Beacon Probe (MBP) [25], in the *MT* gene (LOC_Os06g05910) and RG gene (LOC_Os12g05900) respectively, founded on our preliminary work [44] by the Real Time PCR. We carried out detailed *in silico* interaction study of validated promoter motifs with AP2/EREBP TF. We generated the 3D DNA model for the validated promoter motifs by 3D DART and protein model of AP2/EREBP TF Protein by using I-TASSER. A further interaction study was carried out through HADDOCK severs. Eventually, we examine the comparative interaction relation between the two promoter sequences with the respective TF family gene.

MATERIALS AND METHODS

Identification of consensus promoter motif in DEGs and designing of specific MBP and primers

Anoxic rice coleoptiles microarray result [8] used for the identification of DEGs. In our previous work, we identified the up-regulated differently (expression increased ≥2 fold) expressed genes (UR-DEGs) and down regulated differently (expression decreased ≥ -2 fold) expressed genes. The GCC and TCC box was found in the promoter of UR-DEGs and DR-DEGs respectively [23-24] after analysis through MEME (v4.5.0) (http: //meme. nbcr. net/meme/cgi-bin/meme. cgi). It was observed that promoter sequence of up-regulated *MT* gene (LOC_Os06g05910) expression increased 15 fold) also has GCC box likewise down-regulated RG gene (LOC_Os12g05900) (expression decreased -2 fold) has TCC box, which were further used to retrieve their promoter sequence from eukaryotic promoter database (EPD) (http: //www. epd. isbsib. ch/seq_download. html). Consensus promoter motifs GCC and TCC box found in the *MT* (UR-DEG) and *RG* (DR-DEG) respectively. For the validation of GCC and TCC box in the promoter of the *MT* and *RG* respectively, the specific MBP and primers [25] were designed by using Beacon Designer 7 (BD7, PREMIER Biosoft, USA) as reported in our previous study [24].

For designing of specific MBP and primer, the promoter sequences of the *MT* and *RG* of 600 nt length (-499 to +100) were retrieved from the EPD (http: //www. epd. isb-sib. ch/seq_download. html) and used to design the specific MBP with stem sequences at 5' and 3' end (highlighted/underlined) (Table-1) and primers (Table-2) by Beacon Designer7. Further validation analysis was carried out by using MBP based Real Time PCR (Applied Biosystems 7500 Fast Real-Time PCR Systems, USA).

Table 1: Molecular beacon probe sequence of *MT* (UR-DEG) and *RG* (DR-DEG)

DEGs	Length of MBP (nt)	MBP with stem sequence (underlined)
MT (Os06g05910)	24	5'-CGCGATCGCCGCCGCCGGATCGCG-3'
RG (Os12g05900)	29	5'-CGCGATCCTCCTCCTCCTCCTCGATCGCG- 3'

Table 2: Primer sequences of UR-DEG (*MT*) and DR-DEG (*RG*)

DEGs	Left primer (5'-3')	Right primer (5'-3')	Amplicon size
MT (Os06g05910)	CCTCCTAGTTCGTCCGTCAA	TCGAGCCTGGACTTCACC	107
RG (Os12g05900)	CATCATTAGCGGAGGATT	CGGAGGTGGCTAAATAAC	162

Isolation of genomic DNA from rice plant

Rice seeds of Azucena (Japonica *sp.*) were grown at room temperature (Fig. 1A) after surface sterilization (0.1% $HgCl_2$) and dark incubation (48 at 36 °C). The genomic DNA was isolated from rice seedlings using CTAB (2X) method and subjected to RNase treatment (fig. 1B). The quantity and quality checked in Biophotometer (Eppendorf, USA) followed by 0.8% agarose gel electrophoresis.

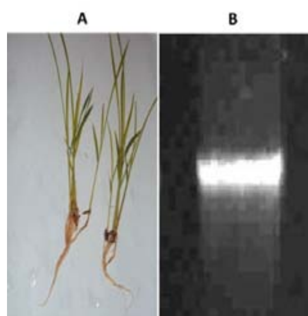


Fig. 1: (A) Fourteen days old rice seedling (B) RNase treated genomic DNA isolated from rice leaves

Validation of consensus promoter motifs using a molecular beacon probe based on real time PCR

For validation of the promoter motifs of the DEGs, specific MBP (table-1) and primers (table-2) used. Isolated genomics DNA was used as a template, whereas the GCC and TCC box containing MBP used as a probe for detection of consensus promoter motifs in DEGs. In PCR reaction volume of 15-20 µl (1X Taq buffer, 1 unit Taq polymerase, 0.2 mm dNTPs, 3-4 mM MgCl2, 0.45µM primer, 3-10ng g DNA and 0.3-0.8µM MBP) at optimized PCR condition (95°C for 4-10 min; 40-45 cycles of 10-15s at 95°C, 20-35s at 60°C, and 30- 45s at 72°C). PCR amplification was carried out in Real Time PCR System (Applied Biosystems 7500 Fast Real-Time PCR Systems, USA). For the detection of GCC box in the promoter of *MT* (LOC_Os06g05910) gene, probe (MBP) of GCC box, primers and target genomic DNA as well as with non-template control (NTC) and negative control having MB of TCC box were used in PCR amplification. Likewise, for TCC box promoter motif detection in *RG* (LOC_Os12g05900) gene promoter, TCC box MBP was amplified along with non-template control (NTC) and negative control having MB of GCC box using primers and target genomic DNA. The Ct value obtained from the Real Time PCR data.

In silico protein-DNA interaction studies

For the protein-DNA interaction study the UR-DEG AP2/EREBP (Loc_Os03g22170) TF (expression increased 29 fold), its protein

sequences retrieved from the TIGR (http: //rice. plantbiology. msu. edu/cgi-bin/ORF_infopage. cgi). Further, its structure was predicted by using I-TASSER (http: //zhanglab. ccmb. med. umich. edu/ITASSER/) which built the 3D models on multiple-threading alignments. The accuracy of protein prediction is based on the confidence score (C-score) of the model by I-TASSER [45]. The Best protein model generated by I-TASSER was run on Ramachandran Plot Analysis (RAMPAGE) (http: //mordred. bioc. cam. ac. uk/~rapper/rampage. php) [46] for evaluation of the stability of protein models. Moreover, for the construction of a 3D DNA structure of promoter motif DNA, 25 nt sequence of the promoter motif having GCC box and the TCC box required [24, 42]. 3D-DART (3DNA-Driven DNA Analysis and Rebuilding Tool) server (http: //haddock. science. uu. nl/dna/dna. php) was used for generating 3D custom made structural model of the validated promoter motif for both *MT* and *RG* DEGs.

The 3D model of DNA having varies bend angle ranging from the 0-40° for each constructed 3D DNA model. Hence five 3D Model generated for each *MT* and *RG* DEGs. Consequences for *in silico* interaction studies, both 3D DNA models of gene promoter motifs and AP2/EREBP TF model were run on the online HADDOCK web server (http: //haddock. science. uu. nl/services/HADDOCK/ haddock. php). Further for validation of the interacting protein-DNA complex molecular structure predicted by HADDOCK server, analyzed in SiteMap module of the Schrödinger Suite [36].

RESULTS AND DISCUSSION

Rice (*Oryza sativa* L.) represents the third most important food grain crop in the world behind wheat and corn [47]. However the various abiotic stresses reduce the crop productivity [48-49]. Abiotic stresses control the expression of the many genes and their product as well as TFs for their regulation [50]. In plants low oxygen stress stimulates the composite metabolic pathways and genetic programs, including the differential expression of a great number of genes [7]. The gene expression studies revealed the up-regulation of genes coding for transcription factors under low oxygen stresses [51]. Microarray analysis has been used to study differential expression of various genes in abiotic and biotic stresses. Moreover, *in silico* study of differentially expressed genes reveals many possible functions of the genes during the different stresses, hence the validation required for the obtained results. The gene promoter contains *cis*-elements which play a central role in genes regulation contains the essential nucleotide sequences and transcription start site.

In silico study about promoter motifs of the differentially expressed genes in anoxia have been reported [9,22,23]. However the anoxia responsive DEGs have the consensus promoter motifs (GCC and TCC box) in their promoter, reported using MEME analysis, in our previous study [23]. Further identification and validation study on a GCC box (GCCGCC) in the anoxia responsive differentially expressed *UCC* gene have been done [24]. We also identified the occurrence of the GCC box (CGCCGCCGCCG) in *MT* and the TCC box (CTCCTCCTCCTCCTC) in *RG gene* in their promoter motif (fig. 2).

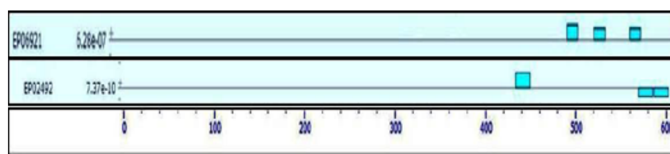


Fig. 2: A snapshot repeated set of GCC-box (CGCCGCCGCCG) and the TCC box (CTCCTCCTCCTCCTC) positioned in the promoter region of *MT* (EP06921) and *RG* (EP02492) DEG, respectively, which ranging from 200 to 600 bp of the promoter region and analyzed by MEME (v4.5.0)

To identify the promoter motifs in a set of sequences web-accessible bioinformatics tools are being used routinely by molecular biologists, such as MEME [52]. Similar studies on the identification of promoter motif reported in rice [23,24,53] and in *Zea Mays* [54] using MEME.

Several reports described methyltransferases involve in gene expression, genome stability and the DNA methylation in plants like in *Arabidopsis* [55], maize [56], rice [57] and in wheat [58]. It involves in embryonic development (GO: 0009790), metabolic process (GO: 0008152) and transferase activity (GO: 0016740) in plants. While *RG* has imperative secondary functions in plants like catabolic (GO: 0009056), metabolic process (GO: 0006139), signal transduction (GO: 0007165) and enzyme regulator activity (GO: 0030234). Moreover the RhoGAP is peripheral membrane proteins which control over the cell surface-associated actin cytoskeleton, contributing to the formation of social systems as diverse as lamellipodia and filopodia of animal cells, yeast buds, and plant root hairs and/or pollen tubes [59,60,61]. So, the presence and experimental detection of GCC and TCC box need to be validated in *MT* (UR-DEG) and *RG* (DR-DEG) becomes essential. The validation of the promoter sequences was achieved through MBP (table-1) and their specific primers (table-2) based on Real Time PCR. The genomic DNA extracted (fig. 1B) from the Rice seedlings of Azucena (japonica sp.) (fig. 1A) using CTAB (2X) method. DNA concentration and quality checked in Biophotometer (Eppendorf, USA) and 0.8% agarose gel electrophoresis, respectively.

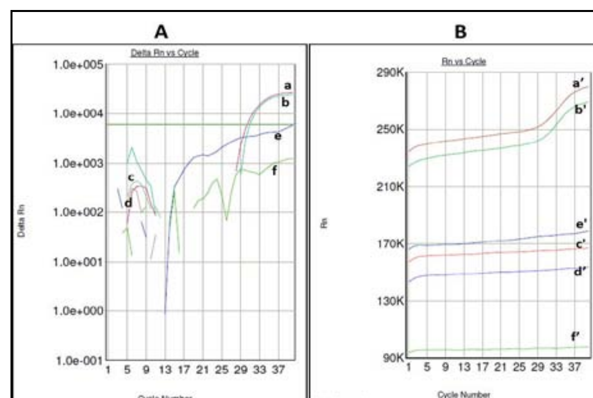


Fig. 3: Molecular beacon probe based real time PCR result for *MT* having a GCC box in its promoter region (-499 to +100). Graph indicating the relation between Delta Rn vs Cycle number (A) and Rn vs Cycle number (B). Curves *a′*&*b′* depict amplification of the gene with GCC probe. Curves *c′* & d′ are the amplification of TCC probe and *e′*&*f′* are amplifications of NTC (Non template control)

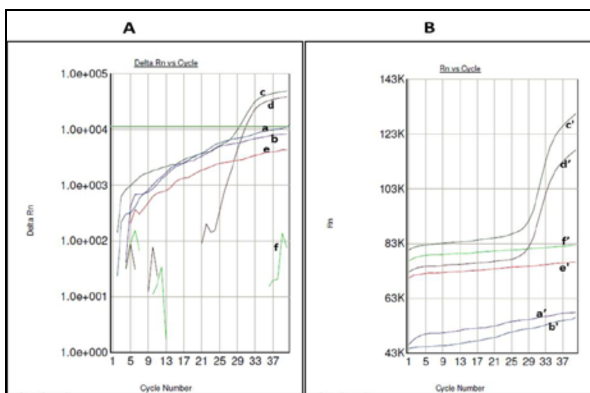


Fig. 4: Molecular beacon based real time PCR result for *RG* having a TCC box in its promoter region (-499 to +100). Graph indicating the relation between Delta Rn vs Cycle number (A) and Rn vs Cycle number (B). Curves *a′*&*b′* depict amplification of the gene with GCC probe. Curves *c′* & d′ are the amplification of TCC probe and *e′*&*f′* are amplifications of NTC (Non template control)

Further genomic DNA was used for the validation of Promoter motifs in DEG. The presence of GCC and TCC box promoter motif in the promoter region of the of *MT* and *RG* DEG validated through the designed MBP respectively, based on Real Time PCR assay. The graph (DRN vs cycle number (Figur-3A) and RN v's Cycle number (fig. 3B) generated by the inbuilt Real Time PCR software. The gene amplification of *MT* observed through the GCC box containing MBP (fig. 3) have avg. Ct values 30.62 (table-3). The GCC box containg MBP probe was detected during PCR amplification, however amplification of the NTC (Non template control) and the negative control TCC box containg MBP was undetected, confirming the GCC box presence in *MT* gene, after analysis in the Real Time PCR assay.

The presence of the TCC box detected in the promoter region after PCR amplification reaction of the *RG*, have 30.13 avg Ct value (table-3). The graph DRn vs cycle number (fig. 4A) and Rn vs Cycle number (fig. 4B) generated by the inbuilt software. The presence of the negative control GCC box containing MBP and NTC were undetected in a real time PCR assay, confirming the TCC box presence in *RG* gene. Similarly the GCC box validated in anoxia responsive differentially expressed *UCC* gene with MBP [24]. Moreover the sensitivity and accuracy of MBP have been reported earlier [29, 62]

Table 3: Ct value chart of *MT* and *RG* genes

DEGs	Replicate	Template	Molecular beacon	Ct value	Avg Ct value
MT (Os06g05910)	R1	Template	GCC box	30.9	30.62
	R2	Template	GCC box	30.34	
RG (Os12g05900)	R1	Template	TCC box	30.82	30.13
	R2	Template	TCC box	29.43	

Transcription factors (tFs) are the key regulator which controls the expression of clusters of genes through the specific binding site present in the genes promoter's site of the respective target genes [63]. Under biotic and abiotic responses AP2-EREBP/ERFs TF has been found to be involved in growth, development and metabolic regulation [12]. AP2/ERF superfamily proteins act as a transcriptional regulator plays a essential role in gene expression in response to the hormone, biotic and abiotic factors, symbiotic interactions, cell specialization, and stress signalling pathways in plants [64-65]. The rice ERF transcription factor OsERF922 binds specifically to the GCC box sequence, and acts as a transcriptional activator in rice plant cells [66]. However, in plants AP2/ERF superfamily TF interact specifically with widely conserved AGCCGCC motifs (GCC box). However base pair mutation decreases the binding affinity of the ERF TF [17]. Gene expression controlled by the AP2/ERF superfamily TF negatively or positively with the interaction of the GCC box promoter motif. However reduces its expression when G residue in GCC box replaced by T residue [67]. Moreover, in plant the mutation of the core sequence in the promoter region of GCC box reduces the binding activity of TF reported by several researchers [68-69]. In consequence, protein sequence and structure for the interaction study of the TFs with the promoter motif is needed to identify for understanding the regulation mechanism of various biological process. Hence, the 3D structure of the AP2/EREBP TF (Loc_Os03g22170) which is not available in the PDB database, generated from the I-TASSER (http: //zhanglab. ccmb. med. umich. edu/I-TASSER/) severe. I-TASSER predicted the five best models on the basis of the confidence score (C-score), the estimated TM-score and RMSD [45]. Best two predicted models (Model1 and Model 2) from the I-TASSER run in the RAMPAGE (http: //mordred. bioc. cam. ac. uk/~rapper/rampage. php) for evaluation of the stability of protein models. Protein model 2 of AP2/EREBP TF is a more stable structure having 68.5 % residues is in the most favorable region and 22.8% in allowed region (fig, 5). Hence the model 2 of AP2/EREBP TF used for the further interactive study.

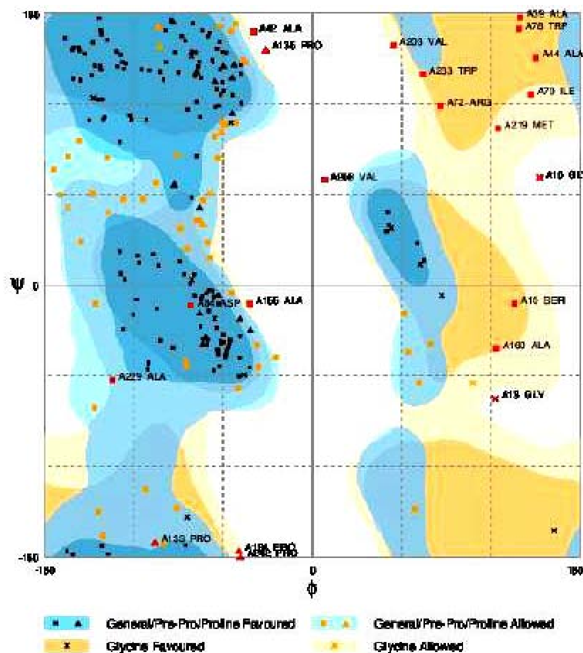


Fig. 5: Ramachandran plot assement of the AP2/EREBP (LOC_Os03g22170) TF (Protein model 2)

For *in silico* study of the protein-DNA interaction, the 3D model of protein as well as 3D model of DNA was required. The *MT* and *RG* gene promoter DNA of 25 nt length used for the generation of the 3D model (fig. 7) by the 3D DART server. Five 3D DNA models generated for each *MT* and *RG* gene promoter DNA having GCC and TCC box respectively.

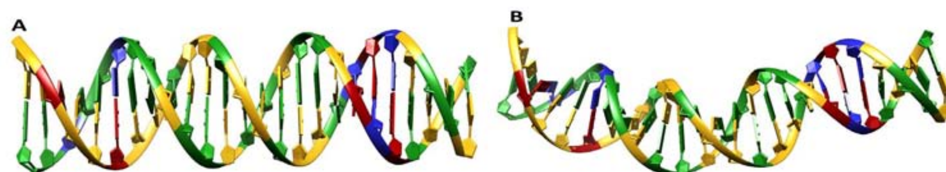


Fig. 7: 3D structure of the linear DNA segment of *MT* gene promoter motif. GCC-Box promoter motif positioned at 9- 18 nucleotide of 25 nt long sequences of the *MT* gene promoter DNA, generated by 3D-DART server. 3D DNA model structure varies between 0-40º of bending. Linear 3D model (Figure–7A) and bended 3D (40º) model (fig. 7B) represented for *MT* gene promoter DNA. In DNA model red color represents Adenine, green color represents guanine, gold color represents cytosine and the blue color represents thymine. The structure was generated using Chimera 1.9

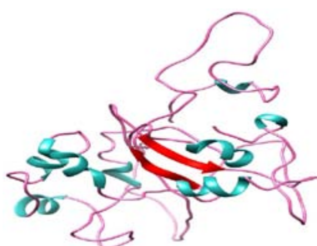


Fig. 6: 3D structure of AP2/EREBP (LOC_Os03g22170) TF (Model 2) predicted by I-TASSER. The coloring method is based on secondary structure. The red color represents strand, pink color is coiled and cyan color represents the helix on AP2/EREBP TF. The structure was generated using Chimera 1.9

Protein-DNA interaction proteins are central for many processes in living cells, especially transcriptional regulation and DNA modification. To understand the important biological process and working genomes, it is essential to understand the interaction at the macromolecular level [70-71]. Therefore, structural determination of the protein- DNA complexes and the factors that regulating interaction is essential [72]. In an interaction study of macromolecular complexes HADDOCK server was used [24,42,73]. The both DNA and protein, with their respective binding site run into the HADDOCK server. The HADDOCK result of the interacting molecules between *MT* gene Promoter DNA (containing GCC box) and AP2/EREBP TF (table-4) generated. Similarly, HADDOCK result generated for the interacting molecules between *RG* gene Promoter DNA (containing TCC box) and AP2/EREBP TF (table-5). Prediction of the best interaction model based on HADDDOCK score, which is based on the RMSD, van der wall energy, electrostatic buried surface area and Z-score. The lowest HADDOCK score showed the favorable interaction. The result showed the best interaction between the *MT* gene promoter DNA and AR2/EREBP TF in the model (IAPMTGM2-BS1) has HADDOCK score -112.9 +/- 9.0 (table-4), while the interaction model (IAPRGTBTM2-BS1) of *RG* gene promoter DNA and AP2/EREBP TF has HADDOCK score -120.7 +/- 5.9 (table-5).

Table 4: Protein-DNA docking Models of docked AP2/EREBP (LOC_Os03g22170) TF gene with DNA segment containing a GCC box of UR-DEG, *MT* (LOC_Os06g05910) by HADDOCK

Interaction	HADDOCK score	RMSD	Van der Waals energy	Electrostatic energy	Desolvation energy	Restraints violation energy	Buried Surface Area	Z-Score
IAPMTGM2-BS1	-112.9 +/- 9.0	11.1 +/- 0.2	-61.9 +/- 6.6	-438.1 +/- 22.6	25.6 +/- 9.4	110.1 +/- 27.36	2168.1 +/- 158.6	-2.4
IAPMTGATM2-BS1	86.1 +/- 14.6	19.1 +/- 0.5	-62.7 +/- 4.4	-273.7 +/- 43.3	5.5 +/- 3.9	1980.5 +/- 33.25	2110.6 +/- 148.6	-2.8
IAPMTGBTM2-BS1	94.9 +/- 6.9	26.0 +/- 0.4	-58.2 +/- 4.6	-351.4 +/- 44.9	21.8 +/- 6.5	2015.0 +/- 57.07	1834.5 +/- 149.1	-2.2
IAPMTGCTM2-BS1	97.1 +/- 13.4	25.4 +/- 0.2	-54.7 +/- 8.0	-317.3 +/- 31.1	3.2 +/- 5.0	2120.8 +/- 27.28	1551.4 +/- 106.5	-2
IAPMTGDTM2-BS1	119.2 +/- 16.0	15.5 +/- 0.6	-53.0 +/- 8.2	-149.6 +/- 59.4	1.2 +/- 11.0	2008.9 +/- 45.67	1373.1 +/- 195.1	-1.4

Keys: I- Interaction; AP- AP2/EREBP (LOC_Os03g22170) TF; MT- methyltransferase domain containing protein gene (LOC_Os06g05910), G- GCC box; (A/B/C/D) /T- 10-40º bend angle; M2- Protein model 2; BS1- Binding site.

Table 5: Protein-DNA docking Models of docked AP2/EREBP (LOC_Os03g22170) TF gene with DNA segment containing a TCC box of the DR-DEG, *RG* (LOC_Os12g05900) by HADDOCK server

Interaction	HADDOCK score	RMSD	Van der Waals energy	Electrostatic energy	Desolvation energy	Restraints violation energy	Buried Surface Area	Z-Score
IAPRGTM2-BS1	-102.7 +/- 10.9	4.2 +/- 2.6	-68.0 +/- 10.6	-289.1 +/- 32.1	12.0 +/- 6.0	110.5 +/- 32.04	2066.4 +/- 225.2	-1.7
IAPRGTATM2-BS1	-104.8 +/- 9.1	16.6 +/- 0.1	-73.3 +/- 7.5	-305.3 +/- 12.0	13.6 +/- 3.2	160.0 +/- 37.04	2306.6 +/- 123.9	-1.6
IAPRGTBTM2-BS1	-120.7 +/- 5.9	15.2 +/- 0.7	-68.4 +/- 2.5	-417.9 +/- 27.5	14.9 +/- 5.6	163.7 +/- 42.39	2232.8 +/- 44.4	-2
IAPRGTCTM2-BS1	-104.8 +/- 12.4	20.6 +/- 0.4	-76.4 +/- 10.8	-272.9 +/- 29.5	15.3 +/- 6.5	108.0 +/- 15.74	2204.6 +/- 167.7	-1.6
IAPRGTDTM2-BS1	-110.9 +/- 7.3	6.4 +/- 0.4	-73.9 +/- 4.0	-373.4 +/- 32.3	19.8 +/- 2.0	177.9 +/- 62.35	2231.4 +/- 68.0	-2.2

Keys: I- Interaction; AP- AP2-EREBP (LOC_Os03g22170) TF; RG- RhoGAP domain containing protein (LOC_Os12g05900); T-TCC box; (A/B/C/D) /T-10-40º bend angle; M2- Protein model2; BS1- Binding site. The structural visualization of these protein-DNA interaction models done by using Chimera 1.9 for the model IAPMTGM2-BS1 (fig. 8) and model IAPRGTBTM2-BS1 (fig. 9). In the complex protein-DNA model (IAPMTGM2-BS1), AP2/EREBP TF binds with the liner DNA segment (fig. 8A) whereas in the model (IAPRGTBTM2-BS1) it binds with the 20° bend 3D DNA model (Fig. 9A).

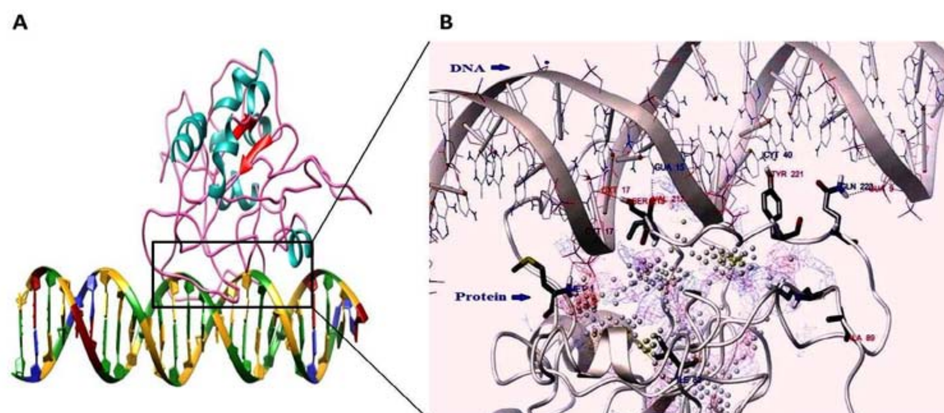


Fig. 8: (A) Protein-DNA docking model (IAPMTGM2-BS1) of the AP2/EREBP TF (LOC_Os03g22170) and DNA segment containing GCC box promoter motif of the UR-DEG, *MT* (LOC_Os06g05910) obtained from HADDOCK server. Interacting amino acid residues were represented in pink and cyan color at chain A. Whereas nucleotide Adenine in red, Guanine represented in forest green, Cytosine in a golden and thymine in blue color at chain B. The structure was generated using Chimera 1.9. (B): Enlarged molecular view of rectangle, area was generated using Site Map module of Schrödinger Suite

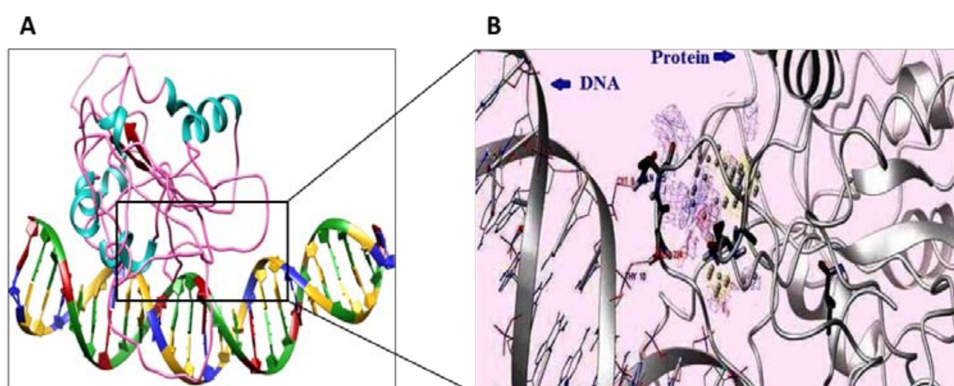


Fig. 9: (A) Protein-DNA docking model (IAPRGTBTM2-BS1) of the AP2-EREBP TF (LOC_Os03g22170) and DNA segment containing TCC box promoter motif of the DR-DEG, *RG* (LOC_Os12g05900) obtained from HADDOCK server. Interacting amino acid residues were represented in pink and cyan color at chain A. Whereas nucleotide Adenine in red, Guanine represented in forest green, Cytosine in a golden and Thymine in blue color at chain B. The structure was generated using Chimera 1.9 (B): Enlarged molecular view of rectangle, area was generated using the Site Map module of Schrödinger Suite

A further validation study of the HADDOCK generated complex protein-DNA model analyzed through SiteMap. The HADDOCK results for the model IAPMTGM2-BS1 (fig. 8A) and model IAPRGTBTM2-BS1 (fig. 9A) showing the binding interaction between the DNA and the protein. In a Site Map analysis of model IAPMTGM2-BS1 (fig. 8B) the binding residue SER213 and MET1 binds with the cytosine17, VAL212 to guanine15, GLN223 to guanine19 with Hydrogen bond. All binding residue is present in the protein active binding site. Similarly, in model IAPRGTBTM2-BS1 (fig. 9B). The binding residue SER224 bind with the thymine10 residue and GLN225 DNA binding with cytosine9 with H-bond. The residues present in the protein active site bind with the DNA sequence with H-bonding predicted by SiteMap for complex model IAPMTGM2-BS1 (fig. 8B) and IAPRGTBTM2-BS1 (fig. 9B). Similarly, the report on the active binding site prediction of the flexible loop PfRIO2 kinase (as plausible novel anti-malarial drug target) which can interact with appropriate ligands was identified computationally by SiteMap module [74]. The SiteMap program [36] can successfully suggest possible binding sites in protein. [75, 76]

CONCLUSION

The present study is proposed to show the relation between the TF and promoter motifs of anoxia responsive DEGs. The identification and validation the promoter motif sequences in *MT* (UR-DEG) and *RG* (DR-DEG) genes were done successfully through *in silico* study and MBP based Real Time PCR analysis, respectively. Anoxia responsive AP2/EREBP TF (LOC_Os03g22170) has shown the good interaction between the *MT* and *RG* genes. However, the result revealed that the AP2/EREBP TF binding affinity towards the TCC box in the *RG* gene promoter is more as compared to the GCC box promoter of *MT* gene. Hence the present study reveals the validation of the *in silico* study of the promoter motifs of *MT* and *RG* genes by MBP is reliable. Moreover, their interaction study with transcription factor shows that it might regulate the differential expression of these genes under rice in anoxia. Further the validation of interacting molecules will help to understand the molecular level of organization and their regulation mechanism.

ACKNOWLEDGEMENT

This work was financially supported by University Grants Commission (UGC), GOI, and New Delhi for the major research project [F. No. 37-113/2009 (SR)]. Mr. Gopal Kumar Prajapati worked as Project Fellow and financially supported from the above mentioned UGC sponsored project. Use SiteMap module of Schrödinger Suite facilities provided by Department of Pharmaceutical Sciences and Technology, BIT Mesra. Mr. Surendra Singh is kindly acknowledged for his concept in using SiteMap module of Schrödinger for this study. Dr. Ashutosh Kumar is kindly acknowledged for his concept in using Molecular Beacon for this study. Central Rainfed Upland Rice Research Station (CRURRS), Hazaribagh-825301, Jharkhand, India is gratefully acknowledged for providing the Azucena rice seeds used in this experiment

CONFLICT OF INTERESTS

Declared None

REFERENCES

1. Liu F, Xu W, Wei Q, Zhang Z, Xing Z, Tan L, *et al*. Gene expression profiles deciphering rice phenotypic variation between *nipponbare* (Japonica) and 93-11 (Indica) during oxidative stress. PLoS ONE 2010;5(1):e8632.
2. Mizoi J, Yamaguchi-Shinozaki K. Molecular approaches to improve rice abiotic stress tolerance. Methods Mol Biol 2013;956:269-83.
3. Perata P, Voesenek LA. Submergence tolerance in rice requires Sub1A, an ethylene-response-factor-like gene. Trends Plant Sci 2007;12:43–6.
4. Magneschi L, Perata P. Rice germination and seedling growth in the absence of oxygen. Ann Bot 2009;103:181.
5. Hattori Y, Nagai K, Ashikari M. Rice growth adapting to deepwater. Curr Opin Plant Biol 2011;14:100-5.
6. Lakshmanan M, Mohanty B, LimS-H, Ha S-H, Lee D-Y. Metabolic and transcriptional regulatory mechanisms underlying the anoxic adaptation of rice coleoptile. AoB Plants 2014;6:1-13.
7. Vartapetian BB, Jackson MB. Plant adaptations to anaerobic stress. Ann Bot 1997;79:3.
8. Lasanthi-Kudahettige R, Magneschi L, Loreti E, Gonzali S, Licausi F, Novi G, *et al*. Transcript profiling of the anoxic rice coleoptiles. Plant Physiol 2007;144:218-31.
9. Pandey DM, Deb R, Kumar A, Japan T. Study on the functional characterization of low oxygen stress inducible genes in rice((Oryza sativa L) Proceeding of 5th. Intewrnational symposium on rice functional genomics (ISRFG' Epochal; 2007. p. 110.
10. Paul AL, Schuerger AC, Popp MP, Richards JT, Manak MS, Ferl RJ. Hypobaric biology: arabidopsis gene expression at low atmospheric pressure. Plant Physiol 2004;134:215.
11. Garg R, Verma M, Agrawal S, Shankar R, Majee M, Jain M. Deep transcriptome sequencing of wild halophyte rice, Porteresia coarctata, Provides novel insights into the salinity and submergence tolerance factors. DNA Res 2013;1-16.
12. Hussain SS, Kayani MA, Amjad M. Transcription factors as tools to engineer enhanced drought tolerance in plants. Biotechnol Prog 2011;27(2):297-306.
13. Riechmann JL, Meyerowitz EM. The AP2/EREBP family of plant transcription factors. Biol Chem 1998;379:633-46.
14. Sharoni AM, Nuruzzaman M, Satoh K, Shimizu T, Kondoh H, Sasaya T, *et el*. Gene structures, classification and expression models of the AP2/EREBP transcription factor family in rice. Plant Cell Physiol 2011;52(2):344-60.
15. Buttner M, Singh B. *Arabidopsis thaliana* ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. Proc Natl Acad Sci USA 1997;94:5961–6.
16. Ohme-Takagi M, Shinshi H. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell 1995;7:173–82.
17. Hao DY, Ohme-Takagi M, Sarai A. Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. J Biol Chem 1998;273(41):26857-61.
18. Chakravarthy S, Tuori RP, D Ascenzo MD, Fobert PR, Després C, Martin GB. The tomato transcription factor Pti4 regulates defense-related gene expression via GCC box and non-GCC box cis elements. Plant Cell 2003;15:3033.
19. Yamaguchi-Shinozaki K, Shinozaki K. Organization of cis-acting regulatory elements in osmotic-and cold-stress-responsive promoters. Trends Plant Sci 2005;10:88-94.
20. Stockinger EJ, Gilmour SJ, Thomashow MF. *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcription activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc Natl Acad Sci USA 1997;94:1035–40.
21. Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, *et al*. Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nat 2006;442:705-8.
22. Mohanty B, Krishnan SPT, Swarup S, Bajic VB. Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species. Ann Bot 2005;96:669.
23. Kumar A, Sharma V, Smita S, Sahu N, Shankaracharya, Vidyarthi AS, *et al*. In Silico analysis of motifs in promoters of differentially expressed genes in rice (Oryza sativa L.) under anoxia. Int J Bioinfo Res Appl 2009;5(5):523-47.
24. Prajapati GK, Kashayp N, Kumar A, Pandey DM. Identification of GCC box in the promoter region of ubiquinol cytochrome c chaperone gene using molecular beacon probe and its *in silico* protein-DNA interaction study in rice (*Oryza sativa* L.). Int J Comp Bioinfo *In Silico* Mod 2013;2(5):213-22.
25. Tyagi S, Kramer FR. Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol 1996;14:303-8.
26. Andersen CB, Holst-Jensen A, Berdal KG, Thorstensen T, Tengs T. Equal Performance of taq man, MGB, Molecular beacon, and SYBR green-based detection assays in detection and quantification of roundup ready soybean. J Agric Food Chem 2006;54:9658-63.
27. Kamboj A, Pateriya AK, Mishra A, Ranaware P, Kulkarni DD, Raut AA. Novel molecular beacon probe-based Real-Time RT-PCR assay for diagnosis of crimean-congo hemorrhagic fever encountered in India. Bio Med Res Int 2014;27:496219.
28. Gohring J, Jacak J, Barta A. Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in Arabidopsis. Plant Cell 2014;26:754–64.
29. Lata P, Ram S, Agrawal M, Shanker R. Real Time PCR for the rapid detection of vanA gene in surface waters and aquatic macrophyte by molecular beacon probe. Environ Sci Technol 2009;43:3343-8.
30. Qin S, Zhou HX. Structural models of Protein-DNA complexes based on interface prediction and docking. Curr Protein Peptide Sci 2011;12:531-9.
31. Leung KS, Wong K-C, Chan TM, Wong MH, Lee KH, Lau CH, *et al*. Discovering protein–DNA binding sequence patterns using association rule mining. Nucl Acids Res 2010;38(19):6324-37.
32. Nagarajan R, Ahmad S, Gromiha MM. Novel approach for selecting the best predictor for identifying the binding sites in DNA binding proteins. Nucleic Acids Res 2013;41(16):7606–14.
33. Tomovic A, Oakeley EJ. Position dependencies in transcription factor binding sites. Bioinfo 2007;23:933-41.
34. Moreira IS, Fernandes PA, Ramos MJ. Computational alanine scanning mutagenesis-an improved methodological approach. J Comput Chem 2007;28(3):644–54.
35. Setny P, Bahadur RP, Zacharias M. Protein-DNA docking with a coarse-grained force field. BMC Bioinfo 2012;13:228.
36. Site Map-V-2.5, V 2.5, Schrödinger, LLC: New York, NY; 2011.
37. Banitt I, Wolfson HJ. Para dock: a flexible non-specific DNA–rigid protein docking algorithm. Nucleic Acids Res 2011;39(20):135.
38. Aloy P, Moont G, Gabb HA. Modelling repressor proteins docking to DNA. Proteins 1998;33(4):535–49.
39. Van Dijk M, van Dijk ADJ, Hsu V. Information driven protein-DNA docking using, HADDOCK: it is a matterof flexibility. Nucleic Acids Res 2006;34(11):3317–25.
40. van Dijk M, Bonvin AMJJ. Pushing the limits of what is achievable in protein-DNA docking: bench marking HADDOCK's performance. Nucleic Acids Res 2010;38(17):5634–47.
41. Qamarunnisa S, Hussain M, Jabbeen, Raza S, Khanani MH, Azhar A, *et al*. *In silico* studies on structure-fuction of DNA GCC-Box binding domain of *Brassica Nnapus* DREB1 protein. Pak J Bot 2012;44(2):493-500.
42. Pandey DM, Kumar A. 3D structure prediction and protein-DNA interaction of CCCH-type zinc finger transcription factor gene in Rice (*Oryza sativa* L.). Int J Comp Bioinfo In Silico Modeling 2013;2(2):94-103.
43. Shanker AK, Maddaala A, Kumar MA, Yadav SK, Maheswari M, Venkateswarlu B. In silico targeted genome mining and comparative modelling reveals a putative protein similar to an Arabidopsis drought tolerance DNA binding transcription factor in Chromosome 6 of Sorghum bicolor genome. Interdiscip Sci 2012;4(2):133-41.

44. Prajapati GK, Kumar A, Pandey DM. Poster presentation on molecular beacon probe based promoter motif detection & in silico protein-DNA interaction studies during submergence in rice (Oryza sativa L. In 7th International rice genetics symposium manila philippines; 2013. p. 85.
45. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinf 2008;9:40.
46. Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, *et al*. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 2003;50(3):437-50.
47. Binod P, Sindhu R, Singhania RR, Vikram S, Devi L, Nagalakshmi S, *et al*. Bioethanol production from rice straw: an overview. Bioresour Technol 2010;101:4767–74.
48. Rehman S, Harris PJC, Ashraf M. Stress environments and their impact on crop production. In: Ashraf M, Harris PJC, eds. Abiotic stresses: plant resistance through breeding and molecular approaches. Haworth Press: New York; 2005. p. 3-18.
49. Ashraf M, Athar HR, Harris PJC, Kwon TR. Some prospective strategies for improving crop salt tolerance. Adv Agron 2008;97:45–110.
50. Todaka D, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K. Toward understanding transcriptional regulatory networks in abiotic stress responses and tolerance in rice. Rice 2012;5:6.
51. Liu FL, Van Toai T, Moy LP, Bock G, Linford LD, Quackenbush J. Global transcription profiling reveals comprehensive insights into hypoxic response in Arabidopsis. Plant Physiol 2005;137:1115–29.
52. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, *et al*. MEME SUITE: tools for motif discovery and searching. Nucl Acids Res 2009;37:202-8.
53. Doi K, Hosaka A, Nagata T. Development of a novel data mining tool to find cis-elements in rice gene promoter regions. BMC Plant Biol 2008;8:20.
54. Liseron-Monfils C, Lewis T, Ashlock D, McNicholas PD, Fauteux F, Strömvik M, *et al*. Promzea: a pipeline for discovery of co-regulatory motifs in maize and other plant species and its application to the anthocyanin and phlobaphene biosynthetic pathways and the maize development atlas. BMC Plant Biol 2013;13:42.
55. Vongs A, Kakutani T, Martienssen RA, Richards EJ. *Arabidopsis thaliana* DNA methylation mutants. Sci 1993;260:1926–8.
56. Steward N, Kusano T, Sano H. Expression of ZmMET1, a gene encoding a DNA methyltransferase from maize, is associated not only with DNA replication in actively proliferating cells, but also with altered DNA methylation status in cold-stressed quiescent cells. Nucleic Acids Res 2000;28:3250–9.
57. Teerawanichpan P, Chandrasekharan M, Jiang Y, Narangajavana J, Hall T. Characterization of two rice DNA methyltransferase genes and RNA imediated reactivation of a silenced transgene in rice callus. Planta 2004;218:337–49.
58. Thomas M, Pingault L, Poulet A, Duarte J, Throude M, Faure S, *et al*. Evolutionary history of Methyltransferase 1 genes in hexaploid wheat. BMC Genomics 2014;15:922.
59. Rivero F, Cvrcková F. Origins and evolution of the actin cytoskeleton. Adv Exp Med Biol 2007;607:97-110.
60. Hall A, Rho GT. Pases and the control of cell behaviour. Biochem Soc Trans 2005;33:891-5.
61. Ridley AJ. Rho GT. Pases and actin dynamics in membrane protrusions and vesicle trafficking. Trends Cell Biol 2006;16:522-9.
62. Ye Q, Zhuang H, Zhou C. Detection of naphthalene by real-time immuno-PCR using molecular beacon. Mol Cell Probes 2009;23:29.
63. Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses, including drought, cold, and heat. Front Plant Sci 2009;5:170.
64. Song X, Li Y, Hou X. Genome-wide analysis of the AP2/ERF transcription factor superfamily in Chinese cabbage (*Brassica rapa* ssp. pekinensis). BMC Genomics 2013;14:573.
65. Rashid M, Guangyuan H, Guangxiao Y, Hussain J, Xu Y. AP2/ERF transcription factor in rice: genome-wide canvas and syntenic relationships between monocots and eudicots. Evol Bioinf Online 2012;8:321-55.
66. Liu D, Chen X, Liu J, Ye J, Guo Z. The rice ERF transcription factor OsERF922 negatively regulates resistance to Magnaporthe oryzae and salt tolerance. J Exp Bot 2012;63(10):3899-911.
67. Fujimoto SY, Ohta M, Usui A. Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. Plant Cell 2000;12:394-404.
68. Cheong YH, Moon BC, Kim JK. BWMK1, a rice mitogen-activated protein kinase, locates in the nucleus and mediates pathogenesis-related gene expression by activation of a transcription factor. Plant Physiol 2003;132:1961-72.
69. Sessa G, Meller Y, Fluhr R. A GCC element and a G-box motif participate in ethylene-induced expression of the PRB-lb gene. Plant Mol Biol 1995;28:145-53.
70. Donald JE, Chen WW, Shakhnovich EI. Energetics of protein-DNA interactions. Nucleic Acids Res 2007;35(4):1039-47.
71. Kauffman C, Karypic G. Computational tools for protein–DNA interactions. WIREs Data Mining Knowl Discov 2012;2:14–28.
72. Chou CC, Rajasekaran M, Chen C. An effective approach for generating a three Cys2His2zinc-finger-DNA complex model by docking. BMC Bioinf 2010;11:334.
73. deVries SJ, van Dijk M, Bonvin AMJJ. The HADDOCKweb server for data-driven Biomolecular docking. Nat Protoc 2010;5:883-97.
74. Parveen A, Chakraborty A, Konreddy AK, Chakravarty HC, Sharon A, Trivedi V, *et al*. Skeletal hybridization and PfRIO-2 kinase modeling for synthesis of a-pyrone analogs as anti-malarial agent. Eur J Med Chem 2013;70:607-12.
75. Nayal M, Honig B. On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. Proteins 2006;63:892-906.
76. Halgren TA. Identifying and characterizing binding sites and assessing drug ability. J Chem Inf Model 2009;49(2):377-89.