In silico analysis of motifs in promoters of Differentially Expressed Genes in rice (*Oryza sativa* L.) under anoxia

Ashutosh Kumar, Shuchi Smita and Neeti Sahu

Department of Biotechnology,
Birla Institute of Technology,
Mesra, Ranchi 835 215, Jharkhand, India
E-mail: kr.ashutosh@yahoo.com
E-mail: shuchi2803@yahoo.com
E-mail: neetisahu.ns@gmail.com

Vivekanand Sharma

Department of Plant Biology,
Marsh Life Sciences Building,
University of Vermont,
Burlington 05405 Vermont, USA
E-mail: vsharma@uvm.edu

Shankaracharya, Ambarish S. Vidyarthi and Dev Mani Pandey*

Department of Biotechnology,
Birla Institute of Technology,
Mesra, Ranchi 835 215 Jharkhand, India
E-mail: shankaracharya2222@gmail.com
E-mail: asvidyarthi@bitmesra.ac.in
E-mail: dmpandey@bitmesra.ac.in
*Corresponding author

Abstract: The aim of this study was to characterise the molecular mechanisms of transcriptional regulation of Differentially Expressed Genes (DEGs) in rice coleoptiles under anoxia by identifying motifs that are common in the promoter region of co-regulated genes. Un-changed DEGs (<2 fold and >–2), up-regulated DEGs (≥2 fold) and down-regulated DEGs (≤−2 fold) were separated in three different data sets. Their gene promoters were extracted from eukaryotic promoter database. Statistically significant consensus promoter motifs were detected by *in silico* method. A significant variation in the number of promoter motifs, consensus promoter motif and their sequences between UR-DEGs and DR-DEGs were detected that might be responsible for their related expression.

Keywords: anoxia; AREs; anaerobic response elements; DEGs; differentially expressed genes; *in silico* motif detection; microarray; *Oryza sativa*; consensus promoter motif.

Reference to this paper should be made as follows: Kumar, A., Smita, S., Sahu, N., Sharma, V., Shankaracharya, Vidyarthi, A.S. and Pandey, D.M. (2009) '*In silico* analysis of motifs in promoters of Differentially Expressed Genes in rice (*Oryza sativa* L.) under anoxia', *Int. J. Bioinformatics Research and Applications*, Vol. 5, No. 5, pp.525–547.

Biographical notes: Ashutosh Kumar is a PhD research scholar in the Department of Biotechnology at Birla Institute of Technology, Mesra, Ranchi, India. For the past two years he has been working on Bioinformatics approaches on Rice Functional Genomics under Anoxia. Previously, he was a Research Trainee in Bioinformatics and also gained his MSc (Bioinformatics) from BIT, Mesra, Ranchi, India. He has completed his BSc (Hons.) in Biotechnology from Ranchi University, Ranchi, India. His research interests are in the area of molecular biology, functional genomics, stress physiology and bioinformatics.

Shuchi Smita is working as senior research fellow at the Centre of Biotechnology at IARI, New Delhi, India. Last year she worked on Bioinformatics approaches on Rice Functional Genomics. Previously, she was Research Trainee (Bioinformatics) in the Department of Biotechnology at Birla Institute of Technology, Mesra, Ranchi, India and gained her MSc in Bioinformatics from BIT, Mesra, Ranchi, India. She has completed her BSc (Hons.) in Zoology from Jai Prakash University, Chapara, India. Her research interests are in the area of molecular biology, database designing, functional genomics and bioinformatics.

Neeti Sahu has completed her MSc in Bioinformatics from the Department of Biotechnology, BIT, Mesra, Ranchi, and BSc (Hons.) in Biotechnology from Ranchi University, Ranchi, India. Her research interests are in the area of functional genomics and bioinformatics. Neeti Sahu has completed her MSc in Bioinformatics from the Department of Biotechnology, BIT, Mesra, Ranchi, India. During her MSc Program she did her project work on Rice Functional Genomics under Anoxia with special emphasis on Bioinformatics. She has completed her BSc (Hons.) in Biotechnology from Ranchi University, Ranchi, India. Her research interests are in the area of functional genomics and bioinformatics.

Vivekanand Sharma is a PhD student in the Department of Plant Biology at University of Vermont, USA. Previously, he was a MSc (Bioinformatics) student at the Department of Biotechnology, BIT, Mesra, Ranchi, India. During his MSc Program he worked on Rice Functional Genomics under Anoxia with special emphasis on Bioinformatics. He gained his BSc in Sericulture from University of Agricultural Sciences, Bangalore, India. His research interests are in the area of signal transduction, functional genomics, and bioinformatics.

Shankaracharya is working as a Bioinformatician in the Department of Biotechnology, BIT, Mesra, Ranchi, India. He worked on Rice Functional Genomics under Anoxia. Previously, he gained his Advanced Diploma in Bioinformatics from Jawaharlal Nehru University, Delhi, India. He completed his Master of Science in Biotechnology. He completed his BSc (Hons.) in Biotechnology from Ranchi University, India. His research interests are in the area of database designing, computational biology and bioinformatics.

Ambarish Sharan Vidyarthi, FIE, is a Professor and Head, in the Department of Biotechnology at Birla Institute of Technology, Mesra, Ranchi, India. He holds a PhD Degree in Biochemical Engineering. His primary areas of scientific interest include biosensor and nanobiotechnology, plant biotechnology, environmental and food biotechnology. He is an active member of several scientific societies. His recent academic interests include the bioinformatics approaches in the development of various databases for medicinal and aromatic plants.

Dev Mani Pandey is Reader at the Department of Biotechnology, BIT, Mesra, Ranchi. He holds a PhD in Plant Physiology from CCS Haryana Agriculture University, Hisar, India. From 2002 to 2006, he was Postdoctoral Fellow in various Plant Molecular Biology and Functional Genomics Laboratories in South Korea and has presented his research findings in various international symposia. From May 2008 to November 2008, he worked as a Consultant at IRRI, Los Banos, Philippines. His research areas include rice functional genomics, stress physiology, molecular biology approaches on medicinal and aromatic plants, and development of databases through bioinformatics approach.

1 Introduction

Field-grown plants are exposed to various environmental stresses during their life cycle (Agarwal and Grower, 2005). Higher plants are aerobic organisms that rapidly die when oxygen availability is limited owing to soil flooding (Voesenek et al., 2006). Species originating from semiaquatic environments can survive complete submergence for weeks and some even have the capacity to grow vigorously and produce flowers and seeds. Rice (*Oryza sativa)* is the most important food crop as well as a model system in monocots (Cantrell and Reeves, 2002) that produce high yields even when it is grown in waterlogged rice paddies. Rice can tolerate partial submergence as paddy rice or deepwater rice. However, it was damaged when submerged for a relatively longer period (Agarwal and Grower, 2005).

Despite knowledge of adaptive mechanisms and regulation at the gene and protein level, our understanding of the mechanisms behind plant responses to anoxia is still limited. Even flood-intolerant species, such as *Arabidopsis thaliana*, switch on many genes in response to flooding stress (Branco-Price et al., 2005; Gonzali et al., 2005; Klok et al., 2002; Loreti et al., 2005). The response of plants to low oxygen stress comprises complex biochemical and genetic programmes that include the differential expressions of a large number of genes (Perata and Alpi, 1993; Sachs et al., 1996; Vartapetian and Jackson, 1997). Gene expression is altered under low oxygen stress, and the existence of Anaerobic Response Elements (AREs) along with their binding factors have already been reported (Dolferus et al., 1994; Hoeren et al., 1998; Klok et al., 2002; Olive et al., 1991; Paul et al., 2004). Besides those encoding ANPs, additional hypoxia-induced genes including transcription factors (de Vetten and Ferl, 1995; Hoeren et al., 1998), signal transduction components (Baxter-Burrell et al., 2002; Dordas et al., 2003) and those involved in ethylene biosynthesis (Nie et al., 2002; Olson et al., 1995; Vriezen et al., 1999) were identified.

Using a flood-tolerant (FR13A) and a flood-sensitive (PB1) rice genotype, a PCR-based hybridisation study was also conducted to isolate the genes associated with flooding tolerance in rice (Agarwal and Grower, 2005). Comprehensive understanding of the genetic components that underlie plant stress responses requires in-depth information on gene expression changes at the global level. Therefore, finding the detailed mechanisms of regulation of anoxia tolerance genes at the genome level becomes imperative. In this aspect, microarray is a powerful tool that is being used for gene expression profiling under different stimuli. The regulation of gene expression in response to oxygen deprivation has been described in *Zea mays* (Fennoy et al., 1998), *Arabidopsis thaliana* (Branco-Price et al., 2005; Klok et al., 2002; Liu et al., 2005; Loreti et al., 2005) and *Oryza sativa* (Guglielminetti et al., 2006; Lasanthi-Kudahettige et al., 2007; Mukhopadhyay et al., 2004; Nakazono et al., 2000; Pandey and Kim, 2006; Pandey et al., 2007; Tsuji et al., 2000; Xu et al., 2006). Furthermore, transcriptome profiling of plants subjected to heat, drought, cold, salt, high light, or mechanical stress revealed that very few genes respond in a similar manner to all of these stresses (Cheong et al., 2002; Davletova et al., 2005; Fowler and Thomashow, 2002; Kreps et al., 2002; Rizhsky et al., 2004; Rossel et al., 2002).

Anaerobically expressed genes are often characterised by the presence of AREs in their promoter regions (Walker et al., 1987). A core promoter contains the essential nucleotide sequences for the regulation of gene function having the TATA box and Transcription Start Site (TSS). Moreover, genes having similar expression patterns contain common motifs in their promoter regions (Vilo et al., 2000) that are the key signatures for a family of co-regulated genes (Wang et al., 2004). Tomato transcription factor Pti4, an Ethylene-Responsive Factor (ERF), interacts physically with the disease-resistant protein Pto and binds the GCC box *cis* element that is present in the promoters of many Pathogenesis-Related (*PR*) genes (Chakravarthy et al., 2003). Arabidopsis ethylene-responsive element binding factors (AtERFs) are factors that respond to extracellular signals to modulate GCC box-mediated gene expression positively or negatively (Fujimoto et al., 2000). Consensus motifs in the promoters of low oxygen stress induced genes have also been reported. Transcription factor AtMYB2 in Arabidopsis binding to the GT-motif (5′-TGGTTT-3′) is essential for the induction of *ADH1* during low oxygen stress (Hoeren et al., 1998). A preliminary study on motifs in promoters of 13 anaerobically induced genes in some plant (including rice) species has already been reported (Mohanty et al., 2005). Further, two consensus promoter motifs GGAG[A/G][G/A]G and GACGTGGCG taking into account of the 50 low oxygen stress up-regulated genes in rice seedlings have been reported (Pandey et al., 2007). Identification of regulatory elements in the promoter region that are responsible for activation and expression of co-regulated genes poses a great challenge. Fortunately, a remarkable work on transcript profiling of the anoxic rice coleoptiles has been reported recently (Lasanthi-Kudahettige et al., 2007). However, detailed study for finding the consensus promoter motif during anoxia in rice is scanty. Therefore, this study focuses *in silico* analysis of promoter motifs of DEGs in rice under anoxia. We have made an attempt to search the conserved promoter motifs for anoxia using microarray data set by Lasanthi-Kudahettige et al. (2007). These findings enhance the knowledge about transcriptional regulation of anoxia-induced genes and scores over the previous studies in terms of larger data set having differentially up-regulated, down-regulated and unaffected genes.

2 Methods

2.1 Extraction of DEGs under anoxia

In this study, we have used DEGs reported by Lasanthi-Kudahettige et al. (2007) from the microarray result of anoxic rice coleoptiles. DEGs containing NCBI accession numbers were directly used to extract the two different promoter lengths from the Eukaryotic Promoter Database (EPD) (Praz et al., 2002). DEGs having TIGR locus number, their genomic DNA were retrieved from TIGR rice genome database (www.tigr.org), were used to find their homologues by BLAST (Altschul et al., 1997) from NCBI having query coverage ≥80% and maximum identity ≥90%. Similarly, DEGs having predicted mRNA accession numbers were also used to retrieve their homologue from NCBI.

2.2 Data sets of DEGs under anoxia

DEGs were divided into three different data sets. Set-1 represents the DEGs that were up-regulated (UR-DEGs) by equal or more than two-fold (≥2X, Supplementary Table – data not shown). Set-2 represents DEGs that were down-regulated (DR-DEGs) by equal or less than −2 fold (≤−2X, Supplementary Table – data not shown). Similarly, Data Set-3 represents DEGs that were expressed in between <2X and >−2X marked as un-changed, UC-DEGs (Supplementary Table – data not shown).

2.3 Extraction of promoter sequences and arrangement of data sets

Rice plant promoter sequences were extracted from EPD (http://www.epd.isb-sib.ch/seq_download.html). He et al. (2002) reported that GAGA motif, present in three copies within downstream promoter sequences from +1 to +50, is involved in the regulation of rice tungro bacilliform virus promoter activity in protoplasts derived from cultured rice cells. In this study, we used the promoter lengths up to +50 and +100 in downstream of TSS. To increase the probability of occurrence of the regulatory elements or motifs in the promoter region, two different promoter lengths (−250 to +50 and −499 to +100) were retrieved from EPD for all the three sets, i.e., Set-1, -2, -3 (Supplementary Tables – data not shown).

2.4 Analysis of motifs

To find the promoter motifs from all sets (Set-1, Set-2 and Set-3, and each set with two different promoter lengths of −250 to +50 and −499 to +100), promoter sequences were compiled separately. Detection of motifs with the varying length of 2–18 nucleotides in all six data sets was done by MEME programme (Bailey and Elkan, 1994) considering the total number of possible motifs per session manually. MEME searches statistically significant motif by local, multiple sequence alignment, for repeated, ungapped sequence patterns that occur in the DNA or protein sequences (Bailey and Elkan, 1994).

MEME uses the statistical method for the prediction of motifs in a given data set but it does not separate the motifs according to their frequency of occurrence. Further, MEME judges smaller motifs that are included in larger motifs as a different entity. In our result, we noticed very large differences in the frequencies of promoter motifs in all data sets. And, it was very difficult to analyse them for finding the consensus promoter motifs. Therefore, we took the deviation from average values only for separating the data set into two subsets (high frequent and low frequent) according to their frequency of occurrence. The average of the motif frequencies occurred in each subset was calculated individually. The promoter motifs having frequency more than average were taken as high-frequency set for individual data subset that provides more weightage to the frequently occurred motif during MEME analysis. Similarly, promoter motifs having frequency less than the average were marked as low-frequency sets. Further, to get the consensus promoter motif in both highly frequent and less frequent motifs, all data subsets were further analysed by MEME.

3 Results

Total 13,427 rice promoter sequences were extracted from EPD (http://www.epd.isb-sib.ch/seq_download.html). From 13,427 rice promoter sequences, we found 6502 genes promoters in the promoter length of −250 to +50, whereas 7066 genes promoters in the promoter length of −499 to +100. Out of 6502 promoters in the range of −250 to +50, 1907 promoters (Data Set-1) belong to UR-DEGs, 1784 promoters (Data Set-2) to DR-DEGs and 2811 promoters (Data Set-3) to UC-DEGs. On the other hand, out of 7066 promoters in the range of –499 to +100, 2072 promoters (Data Set-4) belong to UR-DEGs, 1940 promoters (Data Set-5) to DR-DEGs and 3054 promoters (Data Set-6) to UC-DEGs (Table 1). All the above six data sets were used for motif analysis. Total 175 motifs (from 6502 promoters) and 306 motifs (from 7066 promoters) were detected in the promoter lengths of −250 to +50 and −499 to +100, respectively. Out of 175 motifs (from 6502 promoters), 55 motifs (Data Set-1) were obtained in UR-DEGs, 42 motifs (Data Set-2) in DR-DEGs and 78 motifs (Data Set-3) in UC-DEGs. On the other hand, out of 306 motifs (from 7066 promoters), 78 motifs (Data Set-4) were obtained in UR-DEGs, 112 motifs (Data Set-5) in DR-DEGs and 116 motifs (Data Set-6) in UC-DEGs (Table 1).

Table 1 Summarised table showing total number of Differentially Expressed Genes (DEGs) with their respective number of gene promoters, and promoter motifs (high and low frequency) in all six data sets (three different sets with each two different promoter length of – 250 to +50 and –499 to +100) during the anoxia stress

Summarised table					
DEGs in different promoter length	*Total No. DEGs*	*Total No. detected genes promoters*	*Promoter motifs recognised*		
			High frequency	*Low frequency*	*Total*
–250 to +50					
Up-regulated	6369	1907	16	39	55
Down-regulated	5633	1784	16	26	42
Un-changed	9479	2811	24	54	78
Total		*6502*	*56*	*119*	*175*

Table 1 Summarised table showing total number of Differentially Expressed Genes (DEGs) with their respective number of gene promoters, and promoter motifs (high and low frequency) in all six data sets (three different sets with each two different promoter length of – 250 to +50 and –499 to +100) during the anoxia stress (continued)

Summarised table					
DEGs in different promoter length	*Total No. DEGs*	*Total No. detected genes promoters*	*Promoter motifs recognised*		
			High frequency	*Low frequency*	*Total*
–499 to +100					
Up-regulated	6369	2072	18	60	78
Down-regulated	5633	1940	32	80	112
Un-changed	9479	3054	19	97	116
Total	*21481*	*7066*	*69*	*237*	*306*

We noticed very large differences in the frequencies of promoter motifs in our results. Therefore, we separated promoter motifs into two subsets (high frequent and low frequent) according to their deviation from average values only. Total 56 promoter motifs were screened as highly frequent motifs in the promoter length of −250 to +50. Out of 56 motifs, 16 motifs (Data Subset-1A) were obtained in UR-DEGs, 16 motifs (Data Subset-2A) in DR-DEGs and 24 motifs (Data Subset-3A) in UC-DEGs. In contrast, 119 promoter motifs were screened as less-frequent motif. Out of 119 motifs, 39 motifs (Data Subset-1B) were obtained in UR-DEGs, 26 motifs (Data Subset-2B) in DR-DEGs and 54 motifs (Data Subset-3B) in UC-DEGs. On the other hand, 69 promoter motifs were distinguished as highly frequent motifs in the promoter length of −499 to +100. Out of 69 motifs, 18 motifs (Data Subset-4A) belong to UR-DEGs, 32 motifs (Data Subset-5A) to DR-DEGs and 19 motifs (Data Subset-6A) to UC-DEGs. In contrast, 237 promoter motifs were distinguished as less-frequent motifs. Out of 237 motifs, 60 motifs (Data Subset-4B) belong to UR-DEGs, 80 motifs (Data Subset-5B) to DR-DEGs and 97 motifs (Data Subset-6B) to UC-DEGs (Table 1).

Nearly identical promoter motifs CGCCGCCGC, CGCCGCCGCC, CGCCGCCG CCG, and CGCCGCCGCCGC contributed maximum (~47.69%) whereas other two identical motifs AAAAAAAAA and AAAAAAAAAA also occurred more frequently (~10.11% frequency) in the Data Subset-1A. In Data Subset-2A, promoter motifs CTCC TCCTC, CTCCTCCTCC and CTCCTCCTCCTC were detected frequently (~45.15%) whereas the occurrence of CCGCCGCCGCC and CGCCGCCGCCGCC motifs also contributed more (~21%) collectively (Table 2).

Table 2 Promoter motifs with their high rate of occurrence (high frequency, HF) in the UR-DEGs (Data Subset-1A), DR-DEGs (Data Subset-2A), and UC-DEGs (Data Subset-3A) of promoter length −250 to +50 during anoxia

Promoter motifs with their High Frequency (HF) of promoter length –250 to +50					
UR-DEGs (*Data Subset-1A*)		*DR-DEGs* (*Data Subset-2A*)		*UC-DEGs* (*Data Subset -3A*)	
Promoter motif sequence	*HF*	*Promoter motif sequence*	*HF*	*Promoter motif sequence*	*HF*
AAAAAAAAA	523	[A/G]AGA[G/A]AA[G/A][A/G]A[G/A]AA	84	AAAAAAAA	719
AAAAAAAAAA	254	[G/A]AAAAA[A/G]AAAAA[G/A]AAA	75	AAAAAAAAA	461
CGCCGCCGC	1174	A[G/A]AAAAA[G/A]AAA	82	AAAAAAAAAA	172
CGCCGCCGCC	1026	AAAAAAAAAGAA	81	AGAAAAAAAAA	473
CGCCGCCGCC[G/A]	248	AGAAAAA[G/A]AA	92	CCGCCGCCGC	672
CGCCGCCGCCG	821	CCGCCGCCGCC	129	CCGCCGCCGCC	492

Table 2 Promoter motifs with their high rate of occurrence (high frequency, HF) in the UR-DEGs (Data Subset-1A), DR-DEGs (Data Subset-2A), and UC-DEGs (Data Subset-3A) of promoter length −250 to +50 during anoxia (continued)

Promoter motifs with their High Frequency (HF) of promoter length –250 to +50					
UR-DEGs (*Data Subset-1A*)		*DR-DEGs* (*Data Subset-2A*)		*UC-DEGs* (*Data Subset -3A*)	
Promoter motif sequence	*HF*	*Promoter motif sequence*	*HF*	*Promoter motif sequence*	*HF*
CGCCGCCGCCGC	643	CCTCCTCC[T/A]CCTCC	86	CCGCCGCCGCCGCC	193
CGCCGCCGCCGCC	326	CCTCCTCCTCCTC	167	CGCCGCCGC	857
CGCCGCCGCCGCCG	244	CGCCGCCGCCGCC	79	CGCCGCCGCC	667
GCCGCCGCCGCCG	404	CTCCTCCTC	398	CGCCGCCGCCGCC	375
GCGGCGGCGG	316	CTCCTCCTCC	330	CGGCGGCGGC	227
GCGGCGGCGGC	212	CTCCTCCTCCTC	255	CTCCTCCTCTC	411
GGCGGCGG	399	CTCCTCTCCTC	82	CTCTCTCTCTCT	262
GGCGGCGGCG	315	CTCTCTCTCTC	74	CTCTCTCTCTCTC	212
GGCGGCGGCGG	244	GGCGGCGGC	87	GAAAAAAAAAA	183
CCGCCGCCGCCG	531	TCCTCCTCCTCCTCC	76	GAGAGAGAGAGAG	289
				GAGAGAGAGAGAGA	220
				GCCGCCGCCGC	527
				GCCGCCGCCGCCG	281
				GGCGGCGG	341
				GGCGGCGG	309
				GGCGGCGGCG	210
				GGCGGCGGCGG	180
				TCTCTCTCTCTC	257

Three different groups of similar promoter motifs were detected in the Data Subset-3A (Table 2). Result depicted that the similar promoter motifs CGCCGCCGC, CGCCGCC GCC, CCGCCGCCGC, CCGCCGCCGCC, CGCCGCCGCCGCC occurred more frequently (~30%) compared with other two groups of motifs AAAAAAAA, AGAAA AAAAAA, AAAAAAAAA, GAAAAAAAAAA (~22.33%) and motifs CTCCTC CTCTC, CTCTCTCTCTCT, TCTCTCTCTCTC (~11.11%). No remarkable variation among the promoter motifs was noticed in the Data Subset-4A (Table 3). All the motifs of this set were almost similar with the higher rate of occurrence. The promoter motif GCCGCCGCC (~11.05%) occurred more frequently than CGCCGCCGC (~10.99%). The occurrence of promoter motif CGCCGCCGC was very high. We found a large number of promoter motifs in the Data Subset-5A (Table 3). The occurrence of identical promoter motifs CTCCTCCTC, CTCCTCCTCC, TCCTCCTCC, CCTCCTCCTC and CCTCCTCCTCC contributed (~22.25%) equally compared with other identical motifs CCGCCGCCG, CGCCGCCGC, CGCCGCCGCC, GCCGCCGCCG (~21.45%). Identical motifs GGCGGCGG, GGCGGCGGC and GGCGGCGGCG (~10.58%) also showed their vital occurrence. An interesting result was obtained in the Data Subset-6A (Table 3) where almost similar promoter motifs were detected with their varying frequencies. We noticed that the occurrence of motifs CGCCGCCG, CGCCGCCGC (~19.6%), GCCGCCGCC (~8.4%) and CCGCCGCCG (~8.27%) was maximum than other motifs among all the Data Subsets.

Although highly frequent motifs are more important, however in our study low-frequent motifs were also considered for promoter motif analysis (Tables 4 and 5).

Promoter motif AAAAAAAAAAA (~17.59%) occurred maximum than other motifs in the Data Subset-1B (Table 4). We found that the occurrence of CGCCGCCGCC[G/A]C, CGCCGCCGCC[G/A]CC motifs contributed more (~8.81%) compared with the other motifs [A/G]AAAAAAAAA[A/T/G]C, [A/G]A[A/G]AAAAAAAAA[A/T/G]C (~6.98%). The occurrence of motifs GAGAGAGAGAGA, GAGAGAGAGAGAGAG (~10%) was relatively higher compared with AAAAGAAAAAAAAA (~6.5%) motif (Data Subset-2B, Table 5). Interestingly, lowest occurrence of motif AGGAGG (~0.2%) was detected among all the subsets in our study. In our result (Data Subset-3B), similar promoter motifs CTCTCTCTCTCTCT, CTCTCTCTCTCTCTC and TCTCTCT CTCTCTCTCTC (~7.56%) occurred more frequently compared with other groups of promoter motifs CGCCGCCGCCGCCG, CGCCGCCGCCGCCGC, GCCGCCGCCG CCGCC (~6.47%) and GAGAGAGAGAGAGAGA, GAGAGAGAGAGAGAGAGA (~5.32%). In contrast, the motifs GGGG[G/T][G/A]GGGG (~0.36%), GGGGGGCGG (~0.39%) were less frequent (Table 4). In this study, numbers of different promoter motifs in the Data Subset-4B (Table 5) were three times more when compared with Data Subset-4A (Table 3). The motifs GGCGGCGGC, GGCGGCGGCG, CGGCGGCGGCG, GGCGGCGGCGGC, GCGGCGGCGGCG, CGGCGGCGGCGG and GCGGCGGCGGCGG were nearly identical in their sequence and showed maximum occurrence (~29.29%). Similarly, another group (Data Subset-4B, Table 5) of identical motifs CCGCCGCCGCCGCCG, CGCCGCCGCCGCCGCC and CCGCCGCCGCCG CCGC was also detected with low occurrence (~13.86%).

A large number of variations among promoter motifs were detected in Data Subset-5B (Table 5). A group of promoter motifs with identical sequence CTCCTCCTCCTCCTC, TCCTCCTCCTCCTCC, CTCCTCCTCCTCCTCC and CTC CTCCTCCTCCTCC[T/A/C]C showed maximum occurrence (~11.29%). Another group of motifs GGCGGCGGCGGCGGC, CGGCGGCGGCGGCGG and GGCGGCGGC GGCGGCG showed relatively higher occurrence (~7.34%) when compared with promoter motifs CGCCGCCGCCGCC[G/T]C and CGCCGCCGCCGCC[G/T]CC with low rate of occurrence (~3.11%). In the Data Subset-6B, we have detected maximum number of promoter motifs having large variation with reference to their sequence pattern and frequency. The promoter motifs GCGGCGGCG, CGGCGGCGG, GGCGGCGGC, GGCGGCGGCG, CGGCGGCGGCG, GGCGGCGGCGG and CGGCGGCGGCGG were detected with their maximum frequencies (~19.72%). Further, identical motifs CCGCCGCCGCCGCCG, GCCGCCGCCGCCGCC, CCGCCGCCGCCGCCGC, CG CCGCCGCCGCCGCC, GCCGCCGCCGCCGCCG and CCGCCGCCGCCGCCGCC also showed relatively higher (~18.79%) rate of occurrence. Other groups of motifs CC[G/A]CCGCCGCCGC, CC[G/A]CCGCCGCCGCC, CC[G/A]CCGCCGCCGCCG, CC[G/A]CCGCCGCCGCC[G/T] and CTCCTCCTC, CTCCTCCTCC, TCTCCT CCTCC, CTCCTCCTCCTC were detected with relatively lower frequency (Table 5).

The aim of this study was to identify the consensus promoter motifs from co-regulated genes during anoxia. After MEME, varying numbers of promoter motifs were detected in different sets (Tables 2–5). To find the consensus promoter motif, detected motifs were further analysed with MEME programme and presented in Table 6. For UR-DEGs in the promoter length of −250 to +50 (Data Subset-1A) and −499 to +100 (Data Subset-4A), two consensus promoter motifs (CGCCGCCGCCG and GCCGCCGC, respectively) were detected in high-frequency sets. Likewise, in low-frequency motif sets, two consensus promoter motifs GC[G/C]GC[G/C]GC[G/C]GC (Data Subset-1B) and CGCCG[C/G]C[G/T] (Data Subset-4B) were also detected. For DR-DEGs in the promoter length −250 to +50 (Data Subset-2A) and −499 to +100 (Data Subset-5A), two consensus promoter motifs (CTCCTCCTC and TCCTCCTC, respectively) were detected in high-frequency sets. Likewise, in the low-frequency motif sets, two

consensus promoter motifs CG[G/C]CG[G/C]CG[G/C]CG (Data Subset-2B) and C[G/T][C/G]CG[C/G]C[G/T][C/G]C (Data Subset-5B) were also detected. For UC-DEGs in the promoter length −250 to +50 (Data Subset-3A) and −499 to +100 (Data Subset-6A), two consensus motifs (AAAAAAAA and GCCGCCGC, respectively) were detected in high-frequency sets. Likewise, in the low-frequency motif sets, two consensus promoter motifs C[G/T]CCGCC[G/T]C (Data Subset-3B) and CGCCGCC[G/T]CCGC (Data Subset-6B) were also detected.

Table 3 Promoter motifs with their high rate of occurrence (High Frequency, HF) in the UR-DEGs (Data Subset-4A), DR-DEGs (Data Subset-5A), and UC-DEGs (Data Subset-6A) of promoter length −499 to +100 during anoxia

Promoter motifs with their High Frequency (HF) of promoter length –499 to +100					
UR-DEGs (Data Subset-4A)		*DR-DEGs (Data Subset-5A)*		*UC-DEGs (Data Subset-6A)*	
Promoter motif	*HF*	*Promoter motif*	*HF*	*Promoter motif*	*HF*
GCCGCCGCC	6981	CCGCCGCCG	1517	CCGCCGCCG	9409
CCGCCGCCGC	5328	CCGCCGCCGC	1160	CCGCCGCCGC	7667
CCGCCGCCGCC	4441	CCGCCGCCGCC	649	CCGCCGCCGCC	6034
CCGCCGCCGCCG	3081	CCGCCGCCGCCG	404	CCGCCGCCGCCG	4649
CCGCCGCCGCCGCC	1682	CCTCCTCCTC	1118	CCGCCGCCGCCGC	3300
CGCCGCCGC	6941	CCTCCTCCTCC	1026	CCGCCGCCGCCGCC	2271
CGCCGCCGCC	6222	CCTCCTCCTCCTC	637	CGCCGCCG	11316
CGCCGCCGCCG	4817	CCTCCTCCTCCTCC	527	CGCCGCCGC	10917
CGCCGCCGCCGC	3827	CGCCGCCGC	1615	CGCCGCCGCC	8876
CGCCGCCGCCGCC	2997	CGCCGCCGCC	1392	CGCCGCCGCCG	6859
CGCCGCCGCCGCCG	2098	CGCCGCCGCCG	965	CGCCGCCGCCGC	5096
CGCCGCCGCCGCCGC	1371	CGCCGCCGCCGC	492	CGCCGCCGCCGCC	4134
CGGCGGCG	1183	CGGCGGCGG	919	CGCCGCCGCCGCCG	2426
GCCGCCGCCG	5298	CGGCGGCGGCG	575	CGCCGCCGCCGCCGC	1328
GCCGCCGCCGCC	3110	CGGCGGCGGCGG	468	GCCGCCGCC	9551
GCCGCCGCCGCCGC	1578	CGGCGGCGGCGGCG	335	GCCGCCGCCG	7635
GCGGCGGCG	1042	CTCCTCCTC	1285	GCCGCCGCCGC	5789
GGCGGCGG	1136	CTCCTCCTCC	1120	GCCGCCGCCGCC	4483
		CTCCTCCTCCTC	847	GCCGCCGCCGCCGC	1913
		GCCGCCGCCG	1169		
		GCCGCCGCCGC	708		
		GCCGCCGCCGCC	434		
		GCCGCCGCCGCCG	364		
		GCGGCGGCGG	801		
		GCGGCGGCGGC	626		
		GCGGCGGCGGCG	498		
		GCGGCGGCGGCGG	448		
		GCGGCGGCGGCGGC	333		
		GGCGGCGG	1088		
		GGCGGCGGC	992		
		GGCGGCGGCG	729		
		TCCTCCTCC	1292		

Table 4 Promoter motifs with their low rate of occurrence (low frequency, LF) in the UR-DEGs (Data Subset-1B), DR-DEGs (Data Subset-2B), and UC-DEGs (Data Subset-3B) of promoter length −250 to +50 during anoxia

Promoter motifs with their Low Frequency (LF) of promoter length −250 to +50					
UR-DEGs (Data Subset-1B)		*DR-DEGs (Data Subset-2B)*		*UC-DEGs (Data Subset-3B)*	
Promoter motif sequence	*LF*	*Promoter motif sequence*	*LF*	*Promoter motif sequence*	*LF*
[A/G]A[A/G]AAAAAAAAA[A/T/G]C	79	[A/G]A[A/G]GAGAGAGAGAGA	31	[A/G]CGGCGGCGGCGG	30
[A/G]AAAAAAAAAA[A/T/G]C	166	[G/A]CG[GA]CGGCGGCG	39	[A/T/G]AAAAAAAAAA	72
[G/A]GCGGCGGCGGCGG	36	[G/A]GCGGCGGCGGCGG	32	[G/A]AAAA[A/G]AA[G/A]AG[A/G][G/A]A	88
[G/T]CG[GA]CGGCGG[C/A][G/A]GCGGCG	43	[G/C]CC[G/T]CCGCC[G/C]CCCGCC	45	[G/A]AAAAAAAAAAAGA	118
[G/T]GCGGCGGCGGCG	30	AAAAGAAAAAAAAA	66	A[A/G]AAAAAAAAAAAGAAAA	86
AAA[A/G]AA[A/G]AA[A/G]AAAA	80	AAGAAGAAG[A/G]AGAAG	54	AAA[A/G]AAAAAAAAAC	78
AAAAAAAAA[A/G]AA	91	AGAGAGAGAGAGA	33	AAAA[G/A]AAAAA[A/G/T]AAA	115
AAAAAAAAAA[A/G][A/G/T]A[A/G]AA	89	AGAGGGAG	32	AAAAA[G/A]A[A/G]A[G/A]A[G/A]AAA	94
AAAAAAAAAA[A/G]AAAAAAA	78	AGGAGG	2	AAAAAAAAAA[G/A]A[G/A]AAA	117
AAAAAAAAAA[G/A]AAAA	84	C[A/G]GGCGGCGGC	22	AAAAAAAAAA[A/G]AGAAA[A/G]	80
AAAAAAAAAA	617	CC[A/G]CC[A/G]CC[A/G/T]CCGCC	55	AAAAAGAA[G/A]AAAAAAAA	116
AAAAAAAAAA[A/G][A/G/T]A[A/G]AA	157	CCGCCGCCG[C/T]CGCCG	35	AAAAAGAAAA	100
AAAAAAAAAA[A/G]AAAAAAA	143	CCGCCGCCGCCGCC	38	AAAAGAAAAA[A/G]AAA	116
AAAAAAAAAA[G/A]AAAA	72	GGCGGCGGCGGCGGC[G/C][G/A]	35	AAAAGAAAAA[A/G/T]AA	108
AAAAAAAAAA[A/T/G]C	76	CTCTCTCTCTCTCTCTCT	34	AAGAAAAAAAAA	123
AAAAAAAAAA[A/G/T]C	87	G[A/T]G[AG]TGG	12	AGAAAA[G/A]AAAAA	96
C[G/T][G/T/C]C/G/T][G/A]GCGGCGGCGGCGG	31	G[G/A]CG[G/A]CGGCGCGG	40	AGAAAAAAAAAA	128
CAAAAAAAAAAAAA	77	GAGA[A/G]GA[G/A]GA[G/A]GA	41	C[A/G]CC[G/A/T]CCGCCGCCGCC[G/T]C	83
TCTCTCTCTCTCTCTCTC	48	GAGAGAGAGAGA	72	C[A/G/T]CC[G/A]CC[G/A]CC[G/T/A]CC[G/A/T]CC	87
TCTCTC[T/C]CTCTCTCTC	48	GAGAGAGAGAGAGAG	29	C[A/T/G]CC[T/G/A]CC[G/T/A]CC[A/G/T]C	88
CC[T/C/G]CGCCGCCGCC[G/T]CC	65	GAGGGAGT	41	C[G/T]CC[G/T/A]CCGCCGCCGC	79
CGCCGCC[G/A]CC[T/G]CC[T/G]C	77	GAGGGAGTAT	19	C[G/T]CCGCCGCC[G/A]CC	62
CGCCGCCGCC[G/A]C	189	GCGGCGG[A/C]GGCG[A/C/G]GGG	37	C[G/T]CCGCCGCCG	59
CGCCGCCGCC[G/A]CC	120	GCGGCGGCCG	65	CAAAAAAAAAA[A/G]AAAA	118
CGCCGCCGCC[G/A]CCG[T/C]C	53	GGCG[G/A][C/T]GGCGGCGG	33	CC[G/A]CCGCCGC	63
CGCCGCCGCC[G/T]CCGCC	87	GGCG[G/T]C[G/C]GCGGCGGC[G/C]	37	CCGCC[G/T]CCGCCGCCGC	61
CTCTCTCTC[C/T]TCTCTCTCT	42	GGCGGCGG[C/T]G[G/T]	36	CCTCTCTCTCTCTC	37

Table 4 Promoter motifs with their low rate of occurrence (low frequency, LF) in the UR-DEGs (Data Subset-1B), DR-DEGs (Data Subset-2B), and UC-DEGs (Data Subset-3B) of promoter length –250 to +50 during anoxia (continued)

Promoter motifs with their Low Frequency (LF) of promoter length –250 to +50					
UR-DEGs (Data Subset-1B)		*DR-DEGs (Data Subset-2B)*		*UC-DEGs (Data Subset-3B)*	
Promoter motif sequence	*LF*	*Promoter motif sequence*	*LF*	*Promoter motif sequence*	*LF*
G[G/A]CGGCGGC[G/A]GCGGCG	43			CGCC[G/T/A]CCGCCT[G]C	85
G[G/A]CGGCGGCGGCGG	40			CGCCGCCGCC[G/A]CC	69
GCCGCCGCCGCCGC	126			CGCCGCCGCCGCC[G/T]C	115
GCCGCCGCCGCCGCC	52			CGCCGCCGCCGCC[G/T]CCGC	63
GCG[G/A]CGGCGGC	36			CGCCGCCGCCGCCG	164
GCGGCGGC[G/T]G	28			CGCCGCCGCCGCCGC	70
GCGGCGGCGGCG	172			CGGCGGCGGCGG	78
GCGGCGGCGGCGG	70			CGGCGGCGGCGGCG	46
GGCGGCGGC[G/A]GCGGC	37			CTCTCTCTCTCTCT	157
GGCGGCGGCGGCG	70			CTCTCTCTCTCTCTC	110
GGCGGCGGCGGCCG[G/A]	32			GAAAAA[G/A]AAAA	90
GGCGGCGGCGGCGG	36			GAAAAAAAAAAA[G/A]A	84
				GAGAGAGAGAGAGA[G/A]	72
				GAGAGAGAGAGAGAGA	152
				GAGAGAGAGAGAGAGAGA	79
				GCC[G/T]CC[G/T/A]CCGCCGCCGCC	76
				GCCGCCGCCGCC[G/A]C	51
				GCCGCCGCCGCCGCC	47
				GGCGG[C/A/T]GGCGGCG	31
				GGCGGCGA	33
				GGG[G/A][G/A]GGGGG[G/T]GGGG	37
				GGGG[G/T][G/A]GGGG	16
				GGGGAG[A/G]GG	22
				GGGGGCGG	17
				GGGGGGGGGGGGGGG	74
				GGGGGGGGGGGGGGGGG	36
				TCTCTCTCTCTCTCTC	61

Table 5 Promoter motifs with their low rate of occurrence (Low Frequency, LF) in the UR-DEGs (Data Subset-4B), DR-DEGs (Data Subset-5B), and UC-DEGs (Data Subset-6B) of promoter length −499 to +100 during anoxia

Promoter motifs with their Low Frequency (LF) of promoter length −499 to +100					
UR-DEGs (Data Subset-4B)		*DR-DEGs (Data Subset-5B)*		*UC-DEGs (Data Subset-6B)*	
Promoter motif sequence	*LF*	*Promoter motif sequence*	*LF*	*Promoter motif sequence*	*LF*
C[C/T]CC[T/G]CC[T/G]CCTCC[T/G/C]CC	115	ACCACCCACCACC	114	AAAAGAAAAAAAAAA	93
C[G/A]CCGCC[G/T]CCGCCGCC	105	[C/A][CG]GCCGCCGCCGCC[G/C][C][C/T]G	81	[G/A]GCG[G/C]CGGCGGCGGCGGC	65
C[G/T/A]CC[G/A]CC[G/T]CC[G/T]CC[G/A/T]CC	94	[G/A][C/A]CGCCGCCGCCGCCG	67	[G/T]CCGCCGCCGCCG	180
C[G/T]CCGCCGCCGCCGCCGC	82	[G/A][G/C]CGGCGGCGGCGG	24	AAA[A/G]AAAAAAAA	94
C[G/T]CGCCGCCGCC[G/T]CC	115	[G/A]CGGCGGCGGCGGCGG	56	C[A/G]CCGCCGCC[G/T/A]CCGCC	89
CC[G/A/T]CC[G/A/T]CC[G/A/T]CC[T/G/A]CC[G/A/T]	91	[G/C]CCGCC[G/A]CCGCC[G/A]CC	52	C[G/A/C/T]CCGCCGCCGCCGC	117
CC[G/A]CC[G/T/A]CC[G/T]CC[G/T/A]CC	89	[G/C]CGGCGGCGGCGG	45	C[G/A/T]CC[G/A/T]CC[G/A/T]CCGCC	123
CC[G/A]CCGCCGCC	419	[T/A]CC[T/A]CC[T/A]CC[T/A]CC[A/T]C	92	C[G/A/T]CCGCC[G/T]CC[G/T]CC[G/T]CC[G/T]C	112
CC[G/A]CCGCCGCCGC	315	C[A/G/T]CCACC[A/T]CC[A/T/G]CCAC	100	C[G/A]CC[G/A/T]CCGCCGCCGCC	109
CC[G/C/T]CGCCGCCGCC[G/T]CC	113	C[A/T/G]CCTCC[T/A/G]CCTCC	115	C[G/A]CCGCCGCC	940
CC[T/G]CC[T/G]CCTCC[T/G/A/C]CC	110	C[G/A]CC[G/A]CCGCCGCCG	71	C[G/A]CCGCCGCC[G/A/T]CC[G/T]C	90
CCCGCCGCCGCCGCCGCC	102	C[G/T/A]CC[A/T]CC[A/G/T]CC[A/T/C]CC	97	C[G/A]CCGCCGCCGCCG	310
CCCTCC[T/G]CCTCC[T/C]CC	109	C[G/T/A]CC[G/A/T]CCGCCGCCGC	81	C[G/T/A]CC[G/T/A]CC[G/T]CC[G/A]CC	107
CCGCC[G/A]CCGCC	101	C[G/T/A/C]CC[A/T/G]CC[A/T/G/C]CC[C/A/G/T]CC[A/T/G]CC	89	C[G/T]CCGCCGCCG	363
CCGCC[G/T/A]CCGCC[G/T/A]CCGCC	126	C[G/T/C]CC[A/T]CC[A/G/T]CC[A/C/T]CC[A/G]	91	C[G/T]CCGCCGCCGC	295
CCGCCGCC[G/A]CCG	253	C[T/A/C]CGCCGCCGCCG	76	C[G/T]CCGCCGCCGCC	103
CCGCCGCCGCCGCC[G/A]CCG	106	C[T/A]CC[T/A]CCTCC[T/A/G]CC	80	C[G/T/A]CC[G/A/T]CC[G/A/T]CC[G/A/T]CC[G/A/T]CC	102
CCGCCGCCGCCGCC	926	C[T/A]CCTCCTCCTCCTC	113	C[T/G/A]CC[T/A/G]CCGCC[A/G/T]CCGC	100
CCGCCGCCGCCGCCGC	302	C[T/A]CGCCGCCGCCGCC	88	CC[A/G]CC[G/T]CC[G/T/A]CC[A/T/G]CC	99
CCGCCGCCGCCGCCGCCG	96	CC[A/T/G]CC[G/A/T]CC[A/T]CC	67	CC[A/G]CCGCCGCC	189
CG[G/A]CGGCGG[C/A]GGCGGCGG[G/A]	59	CC[A/G]CC[A/G]CC[A/G/T]CC[A/G]	91	CC[C/A/T/G]CC[G/T]CCGCCGCCGC	85
CGCC[G/A/T]CC[G/A]CC[G/A/T]CCGC	114	CC[A/T]CCACC[A/T]CC[A/T]CCACC	85	CC[G/A/T]CC[A/G]CC[G/T/A]CC[A/G/T]CC	96
CGCC[G/A/T]CCGC	92	CC[G/A]CC[G/A]CCGCC[G/A]CC	51	CC[G/A/T]CC[T/G]CC[G/T/A]CC[G/T]CC[G/T/A]C	110

Table 5 Promoter motifs with their low rate of occurrence (Low Frequency, LF) in the UR-DEGs (Data Subset-4B), DR-DEGs (Data Subset-5B), and UC-DEGs (Data Subset-6B) of promoter length –499 to +100 during anoxia (continued)

Promoter motifs with their Low Frequency (LF) of promoter length –499 to +100

UR-DEGs (Data Subset-4B)		*DR-DEGs (Data Subset-5B)*		*UC-DEGs (Data Subset-6B)*	
Promoter motif sequence	*LF*	*Promoter motif sequence*	*LF*	*Promoter motif sequence*	*LF*
CGCC[G/A]CCGCC	416	CCACCACCACCACCAC	73	CC[G/A]CC[G/A]CCGCCGCCGCC[G/A]	108
CGCC[G/A]CCGCCGCCGCCGCCGC	206	CCGCC[G/A]CCGCCGCCGCCGC	71	CC[G/A]CC[G/T/A]CC[G/A/T]CC[G/T]CC[G/T]C	133
CGCC[G/T/A]CCGCC[G/A/T]CCG	113	CCGCCGCCGCCGCC	142	CC[G/A]CC[T/G]CC[G/T/A]CC[G/T/A]C	122
GGCGGCGGCG	886	CCTCC[A/T]CCTCC[T/G]CCTC	112	CC[G/A]CCGCCGCCGC	582
GGCGGCGGCGGC	432	CCTCCTCC[ATC]CCTCC	101	CC[G/A]CCGCCGCCGCC	512
GGCGGCGGCGGCG	238	CCTCCTCCTCC[C/T]	74	CC[G/A]CCGCCGCCGCCG[G/T]	99
GGCGGCGGCGGCGGC[C/T]GGC	69	CCTCCTCCTCCTCCTCCC	76	CC[G/A]CCGCCGCCGCCG	214
CGCC[G/T/A]CCGCC[G/T]CC	210	CG[G/T]CG[G/A/C]CGGCG[G/T]CG	75	CC[G/T]CCGCCGCCGCCGCCG	100
CGCC[G/T/A]CCGCC[G/T]CCGCC	116	CG[G/T/A/C]CGCCG[G/T/C]CG	89	CC[G/T]CCTCC[T/G]CC	123
CGCC[T/A/G]CCGCC[G/A/T]CC	115	CGCC[G/A]C/T]CGCCGCCGCCGCC	84	CC[T/A]CCTCCTCCTCC	127
CGCCGCC[G/A]CC	347	CGCC[G/A]CCGCC[G/A]CCGCCGC	69	CCG[C/G/T]CGCCGCCGC	95
CGCCGCCGCCGCC[G/T]C	324	CGCC[G/T]CCGCC[G/T]CC	70	CCGCC[A/G/T]CC[G/T]CC[G/T/A]CC	88
CGCCGCCGCC[G/T]CC	248	CGCCGCCGCC[A/G]CCG	76	CCGCC[G/T]CCGCC[G/A]CC	79
CGCCGCCGCCG[G/T]CCG	103	CGCCGCCGCCG[C/T]CG	48	CCGCCGCC[G/A]CCGCC	116
CGCCGCCGCCGCC[G/T/A]	98	CGCCGCCGCCG[T/G]CGC	66	CCGCCGCCGCCG[G/A]CC[G/T]C	90
CGCCGCCGCCGCC[G/T]CC	74	CGCCGCCGCCGCC[G/T]C	131	CCGCCGCCGCCG[G/T]CC	205
CGCCGCCGCCGCCGCC	704	CGCCGCCGCCGCC[G/T]CC	65	CCGCCGCCGCCG[G/T]CCGC	109
CGGCGGCGG[C/A/T]GGCG	60	CGCCGCCGCCGCCG	196	CCGCCGCCGCCGCCG	1123
CGGCGGCGGCG	700	CGCCGCCGCCGCCG[C/G]C	88	CCGCCGCCGCCGCCGC	706
CGGCGGCGGCGG	412	CGGCG[G/A]CGGCGGC	136	CCGCCGCCGCCGCCGC	486
CGGCGGCGGCGGCG	187	CGGCG[G/A]CGGCGGCGGC[G/C]G	67	CCGCCGCCGCCGCCGCCG	251
CTCC[T/G]CCTCC[T/G]CC	115	CGGCGGCG[G/A]CGGCG	76	CCTCC[T/G/C]CCTCCTCCTC	138
G[G/A]CGGCGGCGGCG[G/A][C/A][G/A]CG	131	CGGCGGCGGCG[G/A][C/A][G/A]CG	55	CGCC[G/A]CCGCC[G/A]CC[G/T]C	110
GCC[G/A/T]CCGCCGCCGCC	128	CGGCGGCGGCGG[C/A][G/A]CG[G/T][G/A]	55	CGCC[G/A]CCGCCGC	197
GCC[G/A]CCGCC[G/A]CCGC	117	CGGCGGCGGCGGCG[G/A]CG	43	CGCC[G/A]CCGCCGCC	179
GCC[G/T/A]CC[G/A]CC[G/A/T]CC	107	CGGCGGCGGCGGCGG	173	CGCC[G/A]CCGCCGCC[G/T]CC	106

Table 5 Promoter motifs with their low rate of occurrence (Low Frequency, LF) in the UR-DEGs (Data Subset-4B), DR-DEGs (Data Subset-5B), and UC-DEGs (Data Subset-6B) of promoter length –499 to +100 during anoxia (continued)

Promoter motifs with their Low Frequency (LF) of promoter length –499 to +100					
UR-DEGs (Data Subset-4B)		*DR-DEGs (Data Subset-5B)*		*UC-DEGs (Data Subset-6B)*	
Promoter motif sequence	*LF*	*Promoter motif sequence*	*LF*	*Promoter motif sequence*	*LF*
GCCGCCCGCC[G/A]	328	CGGCGGCGGG	53	CGCC[G/T]CC[G/A]CCGCC[G/T]C	136
GCCGCCCGCC[G/T]C	493	CGGCGGGCGG	23	CGCC[G/T]CC[T/A/G]CC[G/A/T]CC[G/T/A]CC	99
GCCGCCCGCC[G/T]CC	347	CTCCTCC[A/T]CC[T/A]CC	83	CGCC[G/T]CCGCCGCC[G/A]C	100
GCCGCCCGCCGCC[G/A]CC	174	CTCCTCCTCCTCCTC	258	CGCCGCC[A/G]CCGCCGCC[A/G]C	98
GCG[G/A]CGGCGGCGGCGGCG	67	CTCCTCCTCCTCCTCC	144	CGCCGCC[G/A/T]CC[G/A]CC[G/T]	86
GCGGCGGCGG[C/T/A/G]GGCG	62	CTCCTCCTCCTCCTCC[T/A/C]C	99	CGCCGCC[G/A]CC	445
GCGGCGGCGGCG	414	G[A/G]CGGCGGCGGCGG[G/A]C	58	CGCCGCC[G/A]CC[G/A]CCGCC	82
GCGGCGGCGGCGG	300	G[G/A]CGGC[G/T]GCGGCG	46	CGCCGCC[G/A]CCGC	279
GCGGCGGCGGCGGCG	61	G[G/A]CGGCGGCGGCGGCG	39	CGCCGCC[G/T/A]CCGCC	88
GGCGGC[G/A]GCCGGCGGCG	61	G[G/C]CGGCGGCG	62	CGCCGCCGCC[G/A]C	191
GGCGGCGGCC	938	GC[C/GT]GCGGCGGCG	36	CGCCGCCGCCGCC	603
		GCCGCCGG[C/T]CGCCG	56	CGCCGCCGCC[G/T]CCG	555
		GCCGCCGCCGCCGCC	73	CGCCGCCGCCGCC[G/A]C	126
		GCGG[C/A][G/A]GCGGCGG[C/T]GGCG	59	CGCCGCCGCCGCCGCC	90
		GCGGCGG[C/T]GGCG	46	CCC[A/T]CC[A/G/T]CC[T/G/A]CC[T/A]CC[G/T/A]CC	91
		GCGGCGGCGGCG[G/A][C/T][G/T]G	50	GCCGCCGCCGCCGCC	1113
		GG[C/A][G/A]GCGGCGG[C/T]G[G/A]C	49	CGCCGCCGCCGCCGCC[G/T]C	117
		GGAGGGAG	28	CGGCG[G/A]CGGCGGC[G/A]G	58
		GGC[G/A]GCGGCGGCGG	37	CGGCGGCGG	832
		GGC[G/C/T]GCGGCGGCG	33	CGGCGGCGGCG	549
		GGC[G/T]GCGGC[G/A]G	29	CGGCGGCGGCGG	490
		GGCG[A/T]CGGCG	34	CGGCGGCGGCGGC	368
		GGCG[G/T]C[G/C]GCGGCGGC[G/C]G	61	CGGCGGCGGCGG	298
		GGCGGCGGCG[G/A]C	50	CGGCGGCGGCGGCGG	175
		GGCGGCGGCG[G/T]	19	CGGCGGCGGCGGCGGCG	119

Table 5 Promoter motifs with their low rate of occurrence (Low Frequency, LF) in the UR-DEGs (Data Subset-4B), DR-DEGs (Data Subset-5B), and UC-DEGs (Data Subset-6B) of promoter length −499 to +100 during anoxia (continued)

Promoter motifs with their Low Frequency (LF) of promoter length –499 to +100					
UR-DEGs (Data Subset-4B)		*DR-DEGs (Data Subset-5B)*		*UC-DEGs (Data Subset-6B)*	
Promoter motif sequence	*LF*	*Promoter motif sequence*	*LF*	*Promoter motif sequence*	*LF*
		GGCGGCGGCGGCGGC	193	CTCC[G/T]CCTCC[G/T]CC	99
		GGCGGCGGCGGCGGCG	96	CTCC[T/C]CCTCCTCCTC[T/C]CC	136
		GGG[A/G]GGGGGGG	37	CTCCTCCTC	572
		GGGGGGGGG	26	CTCCTCCTCC	447
		TCCTCCTCCTCCTCC	210	CTCCTCCTCCTC	94
		CGCCGCCGCCGCC[G/A]C	70	GAGGCGGCG	29
				GCC[G/A]CCGCCGC	476
				GCC[G/A]CCGCCGCCGCC	210
				GCC[G/A]CCGCCGCCGCCGCC	111
				GCCGCC[G/A]CC[G/A]CC[G/A]CC	92
				GCCGCC[G/T]CCGC	379
				GCCGCC[G/T]CCGCC	200
				GCCGCC[T/G/A]CCGCC[G/T]CCGCC	103
				GCCGCCGCCGCCG[C/T]C	75
				GCCGCCGCCGCCGCCG	576
				GCG[G/A]CGGCGGCGGCGG	54
				GCGGCGGCG	834
				GGCGGCGGC	820
				GGCGGCGGCG	768
				GGCGGCGGCGG	544
				GGCGGCGGCGGCG	355
				GGCGGCGGCGGCGGCGGC	64
				TCTCCTCCTCC	121

Table 6 Consensus promoter motifs in all Differentially Expressed Genes (UR-DEGs, DR-DEGs, UC-DEGs) with their rate of occurrence (high and low frequency) in two different promoter lengths (−250 to +50 and −499 to +100) during the anoxia

Promoter length –250 to 50					
UR-DEGs		*DR-DEGs*		*UC-DEGs*	
High frequency (Data Subset-1A)	*Low frequency (Data Subset-1B)*	*High frequency (Data Subset-2A)*	*Low frequency (Data Subset-2B)*	*High frequency (Data Subset-3A)*	*Low frequency (Data Subset-3B)*
CGCCGCCGCCG	GC[G/C]GC[G/C]GC[G/C]GC	CTCCTCCTC	CG[G/C]CG[G/C]CG[G/C]CG	AAAAAAAA	C[G/T]CCGCC[G/T]C
Promoter length –499 to 100					
UR-DEGs		*DR-DEGs*		*UC-DEGs*	
High frequency (Data Subset-4A)	*Low frequency (Data Subset-4B)*	*High frequency (Data Subset-5A)*	*Low frequency (Data Subset-5B)*	*High frequency (Data Subset-6A)*	*Low frequency (Data Subset-6B)*
GCCGCCGC	CGCCG[C/G]C[G/T]	TCCTCCTC	C[G/T][C/G]CG[C/G]C[G/T][C/G]C	GCCGCCGC	CGCCGCC[G/T]CCGC

4 Discussion

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 or submergence. To determine the molecular mechanisms of stress-tolerant genes using microarray is very essential. In this context, transcript profiling of gene expression in response to low oxygen stress in rice have been reported (Lasanthi-Kudahettige et al., 2007; Pandey and Kim, 2006). It was shown that the core promoter, which can extend −35 bp upstream or downstream of this site, plays a central role in regulating the initiation of gene expression (Smale and Kodanaga, 2003).

We found that in DEGs, both promoter lengths of −499 to +100 and −250 to +50 contain repeated GCC sequence (GCC-box) in high- and low-frequent motifs (except Data Subset-2A, Data Subset-3A and Data Subset-5A, Table 6). Our result indicated the biological significance of the data by means that this DNA element binds with AP2/ERF domain containing proteins. This transcription factor superfamily also comprises the *Sub1A* gene that confers submergence tolerance in rice (Xu et al., 2006). Interestingly, a nearly identical consensus promoter motif GCGGCGGCG (identified as highly conserved motif in our analysis in Data Subset-1B, Table 6) has also been reported (Kumar et al., 2007) by taking account of 169 anoxia-inducible genes in rice coleoptiles that were validated by real time RT-PCR (Lasanthi-Kudahettige et al., 2007). In another study, two consensus promoter motifs GGAG[A/G][G/A]G and GACGTGGCG (in promoter lengths of −250 to +50 and −499 to +100, respectively) considering the 50 low oxygen stress inducible genes that were validated by RT-PCR in 5-day-old rice seedlings have also been reported by Pandey et al. (2007). Similar to our finding (Table 6), a nuclear-protein-binding sequence GCCG [G/C] CG motif, specifically

present in the promoter region of *RAmy3D* gene for rice α-amylase, was reported by Mitsunaga et al. (1994). An additional study on *RAmy3D* indicated that consensus sequences of G motif (TACGTA) and TATCCA T/C motif (GATA motif as its antisense sequence) are responsible for sugar repression (Toyofuku et al., 1998). A G-box 10 tetramer promoter motif (GCCACGTGCC) that exhibited high-level expression in transgenic monocot rice has also been reported (Ishige et al., 1999). We noticed that detected motifs of UR-DEGs (Data Subset-1A, Data Subset-1B, Data Subset-4A and Data Subset-4B) have slight variation in their nucleotide sequence (with less information content of nucleotide T; Bailey and Elkan, 1994).

In this study, a remarkable variation in number of promoter motifs and their sequences among the UR-DEGs, DR-DEGs and UC-DEGs were detected. Further analysis of these motifs exhibited a very interesting result. We noticed that consensus promoter motifs of DR-DEGs were different from UR-DEGs. The information content of nucleotide T among the consensus motif of DR-DEGs was high. On the basis of our findings (Table 6) between UR-DEG and DR-DEGs only, we can say that it might be a reason for their related expression. However, a number of other factors might also be responsible for their differential expression. In *Arabidopsis thaliana,* ethylene-responsive element binding protein (AtEBP) was able to bind with a 29-bp oligonucleotide probe containing the GCC box (TAAGAGCCGCCA) and the intensity of the bands directly correlated with the amount of AtEBP present. In contrast, the binding of AtEBP to the mutant GCC box (TAAGATCCTCCA) was abolished. Further, GST–AtEBP fusion protein was able to bind to the GCC box but not to the mutant GCC box (Büttner and Singh, 1997). Competitive gel retardation assays showed that DNA-binding activities are specific to the GCC box sequence in tobacco nuclear extracts and mutant GCC box containing two point mutations reduced GCC box activity *in vivo* (Ohme-Takagi and Shinshi, 1995). We found that either in high- or in low-frequent motifs GCC box (GCCGCC) was the conserved promoter motif for UR-DEGs. On the other hand, mutated GCC box (TCCTCC) was detected as conserved promoter motif for DR-DEGs. From the result, it might be postulated that transcription factor regulates the expression of anoxia-inducible genes directly via the GCC box (or a non-GCC box element) by interacting and activating their expression, and indirectly via other transcription factor as described in previous report (Chakravarthy et al., 2003). Electrophoretic Mobility Shift Assay (EMSA) and transient expression in Arabidopsis have elucidated the transcriptional regulation of AtERFs. AtERF1, AtERF2 and AtERF5 work as activators of GCC box-dependent transcription in Arabidopsis leaves. By contrast, AtERF3 and AtERF4 acted as repressors that down-regulated not only the basal transcription levels of a reporter gene but also the transactivation activity of other transcription factors (Fujimoto et al., 2000). A model suggests that the AtERFs are regulated differentially at both the transcriptional and the post-transcriptional levels that subsets of GCC box-containing genes might be regulated by different ERF proteins (Fujimoto et al., 2000). On the other hand, in our result for the DR-DEGs, the transcription factor might work as a suppressor by direct or indirect interaction with mutated GCC box (TCCTCC). AtERF genes are expressed in response to different stimuli; the AtERF proteins probably function as signal- and sequence-specific transcription factors by activating or repressing the expression of GCC box-containing stress-responsive genes dependent or independent of the ethylene signal (Fujimoto et al., 2000). Therefore, in this study, mutated GCC box (TCCTCC) directly or indirectly might be responsible for down-regulated genes expression under anoxia.

In our result, significant variation in sequences of consensus promoter motifs was detected among UR-DEGs, DR-DEGs and UC-DEGs in highly frequent motifs in the promoter length of −250 to+50. In contrast, similar motif was detected in UR-DEGs and UC-DEGs in highly frequent motifs in the promoter length of −499 to +100. Interestingly, in our result (Table 6), we did not notice any such difference in the nucleotide sequences of consensus promoter motif (except high frequency Data Subset-3A and -6A of UC-DEGs) despite the increase in promoter length from 301 bp (−250 to +50) to 600 bp (−499 to +100). Therefore, it might be suggested that there was no influence of promoter length over the variation of nucleotide usage pattern in consensus promoter motifs of high-frequency Data Subsets in UR-DEGs and DR-DEGs (Table 6). Our result (Table 6) suggested that either some transcription factor might interact with GCC box (GCCGCC) and AAAAAAAA motifs in such a way that their expression was not so much significant that they might be considered as UR-DEGs or DR-DEGs. Further, cooperative interaction of one transcription factor with the GCC box present in the promoter of another transcription factor gene that works as activator or suppressor on transcription factor might also be responsible for their low level of expression under anoxia. It is also very important to take into consideration that a single motif can also bind various transcription factors thereby bringing the genes under multiple regulatory controls (Jin and Martin, 1999; Yanagisawa, 2002).

Considering the UR-DEGs (Set 1), DR-DEGs (Set 2) and UC-DEGs (Set 3) of all the 21,481 genes in microarray result of Lasanthi-Kudahettige et al. (2007), we have done *in silico* analysis for finding the consensuses promoter motifs. In this study, we have used only 30% of total DEGs for consensus promoter motif analysis. Therefore, it is possible that results may vary during consensus promoter motif analysis while considering whole DEGs. Further, the average of the motif frequencies occurring in each subset was calculated individually. Therefore, in our result, frequencies of promoter motifs vary from subset to subset. Also, GCC box (GCCGCC) was the conserved promoter motif for UR-DEGs, while mutated GCC box (TCCTCC) for DR-DEGs. From our result, it might be postulated that significant variation in the consensus promoter motifs between UR-DEG and DR-DEGs might be a reason for their related expression. Further, there was no influence of promoter length over the variation of nucleotide usage pattern in consensus promoter motifs of high-frequency data subsets in UR-DEGs and DR-DEGs. These findings enhance the knowledge about transcriptional regulation of DEGs during anoxia. Experimental validation of biologically important consensus promoter motifs and regulatory elements involved in the activation and expression of DEGs during anoxia needs to be carried out.

Acknowledgement

Bioinformatics facility provided by DBT, India, through BTISnet SubDIC to Department of Biotechnology, Birla Institute of Technology, Mesra-Ranchi, India, is gratefully acknowledged.

References

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.*, Vol. 96, pp.831–844.

Altschul, S.F., Madden, T., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Res.*, Vol. 25, pp.3389–3402.

Bailey, T.L. and Elkan, C. (1994) 'Fitting a mixture model by expectation maximization to discover motifs in biopolymers', *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, Vol. 2, pp.28–36.

Baxter-Burrell, A., Yang, Z., Springer, P.S. and Bailey-Serres, J. (2002) 'RopGAP4-dependent Rop GTPase rheostat control of Arabidopsis oxygen deprivation tolerance', *Science*, Vol. 296, pp.2026–2028.

Branco-Price, C., Kawaguchi, R., Ferreira, R.B. and Bailey-Serres, J. (2005) 'Genome-wide analysis of transcript abundance and translation in Arabidopsis seedlings subjected to oxygen deprivation', *Ann. Bot*., Vol. 96, pp.647–660.

Büttner, M. and Singh, K.B. (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. USA*, Vol. 94, pp.5961–5966.

Cantrell, R.P. and Reeves, T.G. (2002) 'The cereal of the world's poor takes center stage', *Science*, Vol. 296, p.53.

Chakravarthy, S., Tuori, R.P., D'Ascenzo, M.D., Fobert, P.R., Després, C. and Martin, G.B. (2003) 'The tomato transcription factor Pti4 regulates defense-related gene expression via GCC box and non-GCC box *cis* elements', *Plant Cell*, Vol. 15, pp.3033–3050.

Cheong, Y.H., Chang, H-S., Gupta, R., Wang, X., Zhu, T. and Luan, S. (2002) 'Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis', *Plant Physiol.*, Vol. 129, pp.661–677.

Davletova, S., Schlauch, K., Coutu, J. and Mittler, R. (2005) 'The zinc-finger protein Zat12 plays a central role in reactive oxygen and abiotic stress signaling in Arabidopsis', *Plant Physiol.*, Vol. 139, pp.847–856.

de Vetten, N.C. and Ferl, R.J. (1995) Characterization of a maize G-box binding factor that is induced by hypoxia', *Plant J*., Vol. 7, pp.589–601.

Dolferus, R., Jacobs, M., Peacock, W.J. and Dennisn, E.S. (1994) 'Differential interactions of promoter elements in stress responses of the arabidopsis *Adh* gene', *Plant Physiol*., Vol. 105, pp.1075–1087.

Dordas, C., Hasinoff, B.B., Igamberdiev, A.U., Manac'h, N., Rivoal, J. and Hill, R.D. (2003) 'Expression of a stress-induced hemoglobin affects NO levels produced by alfalfa root cultures under hypoxic stress', *Plant J*., Vol. 35, pp.763–770.

Fennoy, S.L., Nong, T. and Bailey-Serres, J. (1998) 'Transcriptional and posttranscriptional processes regulate gene expression in oxygen-deprived roots of maize', *Plant J.*, Vol. 15, pp.727–735.

Fowler, S. and Thomashow, M.F. (2002) 'Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway', *Plant Cell*, Vol. 14, pp.1675–1690.

Fujimoto, S.Y., Ohta, M., Usui, A., Shinshi, H. and Ohme-Takagi, M. (2000) 'Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box–mediated gene expression', *Plant Cell*, Vol. 12, pp.393–404.

Gonzali, S., Loreti, E., Novi, G., Poggi, A., Alpi, A. and Perata, P. (2005) 'The use of microarrays to study the anaerobic response in Arabidopsis', *Ann. Bot.*, Vol. 96, pp.661–668.

Guglielminetti, L., Morita, A., Yamaguchi, J., Loreti, E., Perata, P. and Alpi, A. (2006) 'Differential expression of two fructokinases in *Oryza sativa* seedlings grown under aerobic and anaerobic conditions', *J. Plant Res.*, Vol. 119, pp.351–356.

He, X., Fütterer, J. and Hohn, T. (2002) 'Contribution of downstream promoter elements to transcriptional regulation of the rice tungro bacilliform virus promoter', *Nucleic Acids Res.*, Vol. 30, pp.497–506.

Hoeren, F.U., Dolferus, R., Wu, Y., Peacock, W.J. and Dennis, E.S. (1998) 'Evidence for a role for AtMYB2 in the induction of the arabidopsis alcohol dehydrogenase gene (*ADH1*) low oxygen', *Genetics*, Vol. 149, pp.479–490.

Ishige, F., Takaichi, M., Foster, R., Chua, N-H. and Oeda, K. (1999) 'A G-box motif (GCCACGTGCC) tetramer confers high-level constitutive expression in dicot and monocot plants', *Plant J.*, Vol. 18, pp.443–448.

Jin, H. and Martin, C. (1999) 'Multifunctionality and diversity within the plant MYB-gene family', *Plant Mol. Biol*., Vol. 41, pp.577–585.

Klok, E.J., Wilson, I.W., Wilson, D., Chapman, S.C., Ewing, R.M., Somerville, S.C., Peacock, J.W., Dolferus, R. and Dennis, E.S. (2002) 'Expression profile analysis of the low oxygen response in Arabidopsis root cultures', *Plant Cell*, Vol. 14, pp.2481–2494.

Kreps, J.A., Wu, Y., Chang, H-S., Zhu, T., Wang, X. and Harper, J.F. (2002) 'Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress', *Plant Physiol*., Vol. 130, pp.2129–2141.

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

Lasanthi-Kudahettige, R., Magneschi, L., Loreti, E., Gonzali, S., Licausi, F., Novi, G., Beretta, O., Vitulli, F., Alpi, A. and Perata, P. (2007) 'Transcript profiling of the anoxic rice coleoptile', *Plant Physiol*., Vol. 144, pp.218–231.

Liu, F., Vantoai, T., Moy, L.P., Bock, G., Linford, L.D. and Quackenbush, J. (2005) 'Global transcription profiling reveals comprehensive insights into hypoxic response in Arabidopsis', *Plant Physiol*., Vol. 137, pp.1115–1129.

Loreti, E., Poggi, A., Novi, G., Alpi, A. and Perata, P. (2005) 'A genome-wide analysis of the effects of sucrose on gene expression in Arabidopsis seedlings under anoxia', *Plant Physiol*., Vol. 137, pp.1130–1138.

Mitsunaga, S-I., Rodríguez, R.L. and Yamaguchi, J. (1994) 'Sequence-specific interactions of a nuclear protein factor with the promoter region of a rice gene for α-amylase', *RAmy3D. Nucleic Acids Res.*, Vol. 22, pp.1948–1953.

Mohanty, B., Krishnan, S.P.T., Swarup, S. and Bajic, V.B. (2005) 'Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species', *Ann. Bot*., Vol. 96, pp.669–681.

Mukhopadhyay, A., Vij, S. and Tyagi, A.K. (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*, Vol. 101, pp.6309–6314.

Nakazono, M., Tsuji, H., Li, Y., Saisho, D., Arimura, S-I., Tsutsumi, N. and Hirai, A. (2000) 'Expression of a gene encoding mitochondrial aldehyde dehydrogenase in rice increases under submerged conditions', *Plant Physiol*., Vol. 124, pp.587–598.

Nie, X., Singh, R.P. and Tai, G.C. (2002) 'Molecular characterization and expression analysis of 1-aminocyclopropane-1-carboxylate oxidase homologs from potato under abiotic and biotic stresses', *Genome*, Vol. 45, pp.905–913.

Ohme-Takagi, M. and Shinshi, H. (1995) 'Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element', *Plant Cell*, Vol. 7, pp.173–182.

Olive, M.R., Walker, J.C., Singh, K., Ellis, J.G., Llewellyn, D., Dennis, E.S. and Peacock, W.J. (1991) 'The anaerobic response element', *Plant Mol. Biol*., Vol. 2, pp.673–684.

Olson, D.C., Oetiker, J.H. and Yang, S.F. (1995) 'Analysis of LE-ACS3, a 1-aminocyclopropane-1-carboxylic acid synthase gene expressed during flooding in the roots of tomato plants', *J. Biol. Chem*., Vol. 270, pp.14056–14061.

Pandey, D.M. and Kim, S-R. (2006) 'Isolation of the low oxygen stress genes in rice (*Oryza sativa* L.)', *Proceedings of 4th International Symposium on Rice Functional Genomics (ISRFG' 2006)*, Montpellier, France, Poster 37, p.96.

Pandey, D.M., Deb, R., Kumar, A., Vidyarthi, A.S. and Kim, S-R. (2007) 'Study on the functional characterization of low oxygen stress inducible genes in rice (*Oryza sativa* L.)', *Proceedings of 5th International Symposium on Rice Functional Genomics (ISRFG'2007)*, Epochal, Tsukuba, Japan. Poster 110.

Paul, A.L., Schuerger, A.C., Popp, M.P., Richards, J.T., Manak, M.S. and Ferl, R.J. (2004) 'Hypobaric biology: Arabidopsis gene expression at low atmospheric pressure', *Plant Physiol.*, Vol. 134, pp.215–223.

Perata, P. and Alpi, A. (1993) 'Plant responses to anaerobiosis. Anaerobic gene expression and flooding tolerance in maize', *Plant Sci.*, Vol. 93, pp.1–17.

Praz, V., Perier, R., Bonnard, C. and Bucher, P. (2002) 'The Eukaryotic promoter database, EPD: new entry types and links to gene expression data', *Nucleic Acids Res*., Vol. 30, pp.322–324.

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S. and Mittler, R. (2004) 'When defense pathways collide: the response of Arabidopsis to a combination of drought and heat stress', *Plant Physiol*., Vol. 134, pp.1683–1696.

Rossel, J.B., Wilson, I.W. and Pogson, B.J. (2002) 'Global changes in gene expression in response to high light in Arabidopsis', *Plant Physiol*., Vol. 130, pp.1109–1120.

Sachs, M.M., Subbaiah, C.C. and Saab, I.N. (1996) 'Anaerobic gene expression and flooding tolerance in maize: glycolytic and XET genes and signal transduction', *J. Exp. Bot*., Vol. 47, pp.1–15.

Smale, S.T. and Kadonaga, J.T. (2003) 'The RNA polymerase II core promoter', *Ann. Rev. Biochem.*, Vol. 72, pp.449–479.

Toyofuku, K., Umemura, T. and Yamaguchi, J. (1998) 'Promoter elements required for sugar-repression of the *RAmy3D* gene for alpha-amylase in rice', *FEBS Lett.*, Vol. 428, pp.275–280.

Tsuji, H., Nakazono, M., Saisho, D., Tsutsumi, N. and Hirai, A. (2000) 'Transcript levels of the nuclear-encoded respiratory genes in rice decrease by oxygen deprivation: evidence for involvement of calcium in expression of the alternative oxidase 1a gene', *FEBS Lett.*, Vol. 471, pp.201–204.

Vartapetian, B.B. and Jackson, M.B. (1997) 'Plant adaptations to anaerobic stress', *Ann. Bot.*, Vol. 79, pp.3–20.

Vilo, J., Brazma, A., Jonassen, I., Robinson, A. and Ukkonen, E. (2000) 'Mining for putative regulatory elements in the yeast genome using gene expression data', *Proc. Int. Conf. Intel. Sys. Mol. Biol.*, Vol. 8, pp.384–394.

Voesenek, L.A.C.J., Colmer, T.D., Pierik, R., Millenaar, F.F. and Peeters, A.J.M. (2006) 'How plants cope with complete submergence', *New Phytol.*, Vol. 170, pp.213–226.

Vriezen, W.H., Hulzink, R., Mariani, C. and Voesenek, L.A.C.J. (1999) '1-Aminocyclopropane-1-carboxylate oxidase activity limits ethylene biosynthesis in *Rumex palustris* during submergence', *Plant Physiol*., Vol. 121, pp.189–196.

Walker, J.C., Howard, E.A., Dennis, E.S. and Peacock, W.J. (1987) 'DNA sequences required for anaerobic expression of the maize *Adh1* gene', *Proc. Natl. Acad. Sci. USA*, Vol. 84, pp.6624–6628.

Wang, Z., Dalkilic, M. and Kim, S. (2004) 'Guiding motif discovery by iterative pattern refinement', *ACM Sym. Applied Computing*, Nicosia, Cyprus, pp.162–166.

Xu, K., Xu, X., Fukao, T., Canlas, P., Maghirang-Rodriguez, R., Heuer, S., Ismail A.M., Bailey-Serres, J., Ronald, P.C. and Mackill, D.J. (2006) '*Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice', *Nature*, Vol. 442, pp.705–708.

Yanagisawa, S. (2002) 'The Dof family of plant transcription factors', *Trends Plant Sci.*, Vol. 7, pp.555–560.