NJBAS

Available online at http://www.ajol.info/index.php/njbas/index
Nigerian Journal of Basic and Applied Sciences (December, 2022), 30(2):22-34
DOI: http://dx.doi.org/10.4314/njbas.v30i2.4

ISSN 0794-5698

Full Length Research Article0

In silico Analysis of *Dreb2a* Gene among Diploid *Oryza* Species Reveal Gene Retrogression in Cultivated Species

***[1]A. M. Gumi, [2]B. Kumar, [3]A. Pareek, [2]S. L. Singla-Pareek**
[1]Department of Biological Sciences, Usmanu Danfodiyo University, PMB 2346, Sokoto-Nigeria
[2]Plant Stress Biology Group, International Center for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, PMB 110067, New-Delhi, India
[3]Stress Physiology and Molecular Biology Laboratory, School of Life Sciences, Jawaharlal Nehru University, New Delhi, 110067, India
[*Corresponding Author: E-mail: muhammadag@yahoo.co.uk]

ABSTRACT
The dehydration-responsive element-binding (DREB) genes are types of plant-specific transcription factors that bind to DRE/CRT elements in response to abiotic stresses. It belongs to the AP2 superfamily and contains highly conserved APETELA2 (AP2) domain. In this study, comparative analysis of DREB2A gene from diploid *Oryza* species was performed using BLASTN approach. The *OglDREB2A* sequence (KU159742.1) was used as query against 11 diploid *Oryza* species in the Gramene database for identification of DREB2A orthologs. Molecular characterization, subcellular localization, conserved motifs and promoter analysis were also performed using bioinformatics tools. Molecular characterization of DREB2A protein in the diploid *Oryza* showed an AP2 conserved domain and nuclear localized protein which is similar to putative transcription factor involved in stress responses. Motif analysis using MEME suites identified 7 significant motifs across all the diploid *Oryza* species with motif 2, 3, 4 and 5 as the common motifs with ~50 amino acids length. Analysis of 1kb upstream of TSS of DREB2A of the 11 diploid *Oryza* species revealed many CAREs associated with stress (DRE, LTRE, ERE, G-box, TCC-motif and Sp1) and hormonal regulations (ABREs, MeJA, GARE, and GC-motif) with ABRE as the common element in all the diploid species. Additionally, wild species has more number of cis elements than the cultivated species suggesting a possible role in genome evolution. Overall, the 11 DREB2A orthologs are conserved in terms of sequence homology and protein structure but cultivated species has less number of transcripts than their progenitors suggesting gene retrogression in cultivated *Oryza* species.
Keywords: *In silico* Analysis, DREB2A, Rice, Gene retrogression, Transcription factor

INTRODUCTION

The genus *Oryza* is believed to have originated in an ancient land mass called Gondwanaland (from which Africa, South America, India and Australia drifted apart) before human dependent actions led to evolution of the 2 cultivated rice species (Chang, 1976). The cultivated species are believed to have unknown common ancestor, though domestication has occurred separately in Africa and Asia. In Africa, the weedy *O. longistaminata* distributed throughout Africa gave rise to wild *O. barthii* (endemic to only West Africa) which was independently domesticated into the African rice (*O. glaberrima*) about 3,000 years ago while in Asia, the wild species *O. rufipogon* and *O. nivara* were domesticated into the Asian rice (*O. sativa*) approximately 9,000 years ago (Khush, 1997; Bautista *et al.,* 2001; Cheng *et al.,* 2001). The genus *Oryza* consist of 2 cultivated (2n=24) and 22 wild species (2n=24, 48) distributed throughout the tropics and sub-tropics representing 10 genome types (AA, BB, CC, BBCC, CCDD, EE, FF, GG, HHJJ and HHKK) (Sanchez *et al.,* 2014). African rice (*Oryza glaberrima* Steud.) and Asian rice (*O. sativa* L.) are the only 2 cultivated species of rice with more than 3,000 accessions maintained by The International Rice Research Institute (IRRI) and The National Institute of Genetics' *Oryza* base.

DREB gene types (*DREB1* and *DREB2*), belongs to the ethylene responsive-element binding factor/APETELA2 (ERF/AP2) domain gene family that shares a conserved 58-59 amino acid domain that can bind to dehydration responsive (DRE) *cis*-elements (TACCGACAT) of RD29A promoter to regulate its induction in an ABA-independent response to drought, salinity and cold (Yamaguchi-Shinozaki and Shinozaki, 1994; Sakuma *et al.,* 2006). Although, both DREB1 and DREB2 proteins binds to DRE elements in the promoter regions of the stress responsive genes, DREB1 only regulates the expression of cold responsive genes while DREB2 specifically express drought and salinity responsive regulatory genes (Liu *et al.,* 1998). Genome wide conserved sequences analysis in rice indicates the existence of *DREB* homologs. Matsukura *et al* (2010) reported six (6) *DREB2* type homologs (*OsDREB2A*, 2B, 2C, 2D, 2E and OsAB14) in rice. Os*DREB2* types in rice are orthologous to *DREB2* genes in *Arabidopsis* which confer drought and salt tolerance (Liu *et al.,* 1998 and Nakashima *et al.,* 2000). The DREB2A gene in rice has been identified as important candidate gene in the drought responsive pathway and whose poor sensitivity for Abscisic Acid (ABA) suggests it may be involved in the ABA independent pathway (Dubouzet *et al.,* 2003). *Arabidopsis DREB2A* and rice DREB2A (*OsDREB2A*) show high levels of homology in the N-terminal region and the AP2/ERF domain suggests similarity in the expression and function of both genes.

Though there are numerous reports on *OsDREB* genes in rice (Dobouzet *et al.,* 2003; Nakashima *et al.,* 2009; Matsukura *et al.,* 2010; Mallikarjuna *et al.,* 2011; Filiz and

Tombuloglu, 2014), to date, little has been done on DREB2A in African rice (Gumi *et al.,* 2018). In this study, comparative analysis of DREB2A gene from African rice (*O. glaberrima*) and 10 diploid *Oryza* species was carried out in order to find out its conserved features and possible role of domestication on *DREB2A* gene evolution across the already sequenced diploid *Oryza* species.

MATERIALS AND METHODS

Sequence Retrieval

The complete coding sequence of *OglDREB2A* from African rice were retrieved from NCBI database (https://www.ncbi.nlm.nih.gov/) with accession number KU 159742.1 as reported in our previous study (Gumi *et al.,* 2018). This sequence was used as query to perform BLASTn search against each of the 11 *Oryza* genomes in Gramene database (http://gramene.org) and PlantEnsembl database (http://plants.ensembl.org/index.html) to identify the DREB2A orthologs using a cut off value of $E \leq 10^{-4}$ (Kersey *et al.,* 2016; Ganie *et al.,* 2017). Furthermore, details of each orthologs such as chromosomal location, CDS length, number of introns/exons, number of transcripts and amino acids were retrieved using the sequence export portal of the respective database used.

Distribution of *DREB2A* gene on chromosomes and its Subcellular Localization

The chromosomal location of each of the 11 DREB2A orthologs was derived from BLAST search of Gramene database (Kersey *et al.,* 2016) while their Sub cellular localization was predicted using plant subcellular localization integrative predictor (PSI-Predictor; http://bis.zju.edu.cn/psi) using P value of <0.05 according to methods of Liu *et al.* (2013).

Analysis of Domain Architecture, Conserved Motifs and Molecular Analysis of DREB2A Proteins

For protein domain analysis, Pfam (https://pfam.xfam.org/) and Simple Modular Architecture Research Tool (SMART) was used to reveal the exact domain architecture of all the 11 orthologs of DREB2A protein sequence in *Oryza* genomes (Letunic *et al.,* 2014; El-Gebali *et al.,* 2019) while conserved motifs were analyzed using Multiple Em for Motif Elicitation (MEME) software servers to reveal domain homology among the 11 diploid *Oryza* species (Bailey and Elkan, 1994). The 3D model of DREB2A protein sequences was generated using SWISS Model work space of Expasy translate tool. For higher reliability, built models were selected based on identity with templates, highest GMQE and QMEAN score (Biasini *et al.,* 2014). For each sequence, Expasy-Protparam (Gasteiger *et al.,* 2005) was used to compute the molecular weight, number of positive and negative residues, GRAVY, theoretical isoelectric point, instability index, aliphatic index and other physical and chemical properties.

Promoter Analysis

The promoter sequences of 1,000 bp upstream of the transcription start site (TSS) of each DREB2A orthologs in the diploid *Oryza* species were bulked retrieved from Gramene database (http://ensembl.gramene.org/Oryza_glaberrima/Gene/Sequence) through sequence export option by selecting 1,000 bp upstream of each DREB2A genomic sequence of the diploid species (Tello-Ruiz *et al.,* 2018). The putative *cis*- elements of each promoter sequence were identified and analyzed with Plant cis-acting regulatory elements (PlantCARE: http://bioinformatics.psb.ugent.be/webtools/plantcare/html) (Lescot *et al.,* 2002). Common *cis* elements (present in all the 11 diploid species), rare cis elements (present in >2 but less than 6 species) and unique *cis* elements (present in only single species) were identified and described for comparative analysis between diploid species.

RESULTS

Orthologous DREB2A Genes, Chromosome Location and its Subcellular Localization in Diploid *Oryza* species

Tables 1 and 2 present the results of BLASTN search of DREB2A orthologs in the 11 *Oryza* genomes of Gramene database revealed 11 sequences with higher similarity (>90%) to the query sequence. However, in all the 11 diploid *Oryza* species, the putative DREB2A sequences were located on chromosome 1. The number of DREB2A gene transcripts in diploid *Oryza* species revealed that *O. nivara* has the highest number of transcripts (5) while *O. punctata* and *O. glumipatula* have 4 and 3 transcripts respectively. The cultivated *O. sativa* ssp Japonica, *O. meridionalis* and *O. barthii* (progenitor of African rice) have 2 transcripts each while the African rice (*O. glaberrima*), *O. sativa* ssp Indica, *O. brachyantha*, *O. longistaminata* and *O. rufipogon* have 1 transcript each. Interestingly, all the 11 diploid *Oryza* species studied have a single gene copy number of the DREB2A gene with over 50% of the diploid *Oryza* species having multiple transcripts. The predicted sub cellular localization of DREB2A gene across the 11 diploid *Oryza* species based on reliable index of CELLO program in PSI revealed DREB2A proteins to be localized in nucleus of each of the diploid *Oryza* species.

.

Table 1: Physical properties of DREB2A gene sequences of 11 diploid *Oryza* species

S/N	Oryza species	Transcripts ID	Chromosomal Location	Transcript (bp)	CDS (bp)	Protein (AA)	Exons	Introns
1	*O. barthii*	OBART01G04210.1	Chr1:2,803,135-2,807,973	3726	825	274	3	2
		OBART01G04210.2	Chr1: 2,803,543-2,804,991	855	855	284	2	1
2	*O. brachyantha*	OB01G13930.1	Chr1:2,310,837-2,313,371	1990	825	274	4	3
3	*O. glaberrima*	ORGLA01G0037700.1	Chr1:2,588,134-2,588,979	846	846	281	1	0
4	*O. glumaepatula*	OGLUM01G04560.1	Chr1:3,482,420-3,487,502	3746	825	274	5	4
		OGLUM01G04560.2	Chr1:3,482,420-3,487,502	3355	825	274	4	3
		OGLUM01G04560.3	Chr1:3,482,420-3,487,287	3754	825	274	3	2
5	*O. longistaminata*	KN539827.1_FG005.1	KN539827.1:50,897-52,336	855	855	284	2	1
6	*O. meridionalis*	OMERI01G04110.1	Chr1:3,152,232-3,157,086	3748	825	274	3	2
		OMERI01G04110.2	Chr1: 3,156,333-3,157,103	183	183	60	2	1
7	*O. nivara*	ONIVA01G04630.1	Chr1:3,178,711-3,185,979	5842	825	274	6	5
		ONIVA01G04630.2	Chr1: 3,178,711-3,184,189	3195	825	274	4	3
		ONIVA01G04630.3	Chr1: 3,178,711-3,183,582	3177	825	274	4	3
		ONIVA01G04630.4	Chr1: 3,178,711-3,183,565	3748	825	274	3	2
		ONIVA01G04630.5	Chr1: 3,179,142-3,180,582	855	855	284	2	1
8	*O. punctata*	OPUNC01G03970.1	Chr1: 2,772,897-2,777,519	3417	825	274	4	3
		OPUNC01G03970.2	Chr1: 2,772,897-2,776,374	2156	825	274	4	3
		OPUNC01G03970.3	Chr1: 2,776,607-2,777,483	261	261	86	3	2
		OPUNC01G03970.4	Chr1: 2,772,936-2,774,370	855	855	284	2	1
9	*O. rufipogon*	ORUFI01G04360.1	Chr1:3,002,126-3,006,964	3732	825	274	3	2
10	*O. sativa Indica*	BGIOSGA002846.1	Chr1:3,781,747-3,783,187	855	855	284	2	1
11	*O. sativa Japonica*	Os01g0165000.1	Chr1: 3,356,383-3,358,426	1428	825	274	2	1
		Os01g0165000.2	Chr1: 3,356,461-3,361,204	3638	825	274	3	2

CDS = Coding sequence; AA = Amino acids; bp = Base pair

Table 2: Molecular characterization of the identified DREB2A proteins among diploid *Oryza* species

S/N	Oryza species	Subcellular Localization		Domain	Mol. Wt. (Da)	pI	GRAVY	Instability Index	Aliphatic Index	Stability
		Wolf PSORT	CELLO							
1	*O. barthii*	Nucleus	Nucleus	AP2	30684.28	5.77	-0.843	44.22	58.72	Unstable
		Nucleus	Nucleus	AP2	31846.7	5.72	-0.78	44.8	62.15	Unstable
2	*O. brachyantha*	Nucleus	Nucleus	AP2	30594.06	5.58	-0.849	47.11	55.88	Unstable
3	*O. glaberrima*	Chlo	Nucleus	AP2	31610.6	5.87	-0.679	45.17	65.23	Unstable
4	*O. glumaepatula*	Nucleus	Nucleus	AP2	30602.21	5.96	-0.815	42.58	61.2	Unstable
		Nucleus	Nucleus	AP2	30602.21	5.96	-0.815	42.58	61.2	Unstable
		Nucleus	Nucleus	AP2	30602.21	5.96	-0.815	42.58	61.2	Unstable
5	*O. longistaminata*	Nucleus	Nucleus	AP2	31842.69	5.88	-0.775	43.39	63.17	Unstable
6	*O. meridionalis*	Nucleus	Nucleus	AP2	30652.22	5.77	-0.835	43.37	59.78	Unstable
		Chlo	Nucleus	AP2	6688.42	4.85	-0.683	18.87	63.17	Stable
7	*O. nivara*	Nucleus	Nucleus	AP2	30652.22	5.77	-0.835	43.37	59.78	Unstable
		Nucleus	Nucleus	AP2	30652.22	5.77	-0.835	43.37	59.78	Unstable
		Nucleus	Nucleus	AP2	30652.22	5.77	-0.835	43.37	59.78	Unstable
		Nucleus	Nucleus	AP2	30652.22	5.77	-0.835	43.37	59.78	Unstable
		Nucleus	Nucleus	AP2	30652.22	5.77	-0.835	43.37	59.78	Unstable
		Nucleus	Nucleus	AP2	31814.64	5.72	-0.772	43.99	63.17	Unstable
8	*O. punctata*	Nucleus	Nucleus	AP2	30706.16	5.58	-0.868	39.38	56.93	Stable
		Nucleus	Nucleus	AP2	30706.16	5.58	-0.868	39.38	56.93	Stable
		Chlo	Nucleus/Cyto	AP2	9410.29	4.73	-0.979	26.33	38.49	Stable
9	*O. rufipogon*	Nucleus	Nucleus	AP2	31972.63	5.38	-0.82	40.36	58.38	Unstable
		Nucleus	Nucleus	AP2	30652.22	5.77	-0.835	43.37	59.78	Unstable
10	*O. sativa Indica*	Nucleus	Nucleus	AP2	31814.64	5.72	-0.772	43.99	63.17	Unstable
11	*O. sativa Japonica*	Nucleus	Nucleus	AP2	30665.22	5.77	-0.845	43.37	59.78	Unstable
		Nucleus	Nucleus	AP2	30665.22	5.77	-0.845	43.37	59.78	Unstable

pI = Isoelectric point; Mol. Wt. = Molecular weight; Da = Dalton

Analysis of Motifs and Domains of DREB2A Proteins

The DREB2A protein sequences of the diploid *Oryza* species were analysed for motif distribution using the MEME software of the MEME Suite 5.05 (http://meme-suite.org/index.html). We used the default setting with 10 expected motifs (Table 3). However, only 7 significant motifs were found in the identified DREB2A protein sequences with E-values less than 0.05. Motif 1 (with a size of 50 amino acids) contained an AP2 domain and is present in all the 11 diploid *Oryza* species and Motif 2, 3, 4 and 5 also had a size of 50 amino acids each (though with unknown domain) and occurred in all the 11 species except Motif 5 that is absent in *O. brachyantha.* Motif 7 (with 8 amino acids) is present in all the 11 species while motif 6 (with a size of 15 amino acids) was found in 9 species (absent in *O. Brachyantha* and *O. glumipatula*). The other three non-significant motifs (8, 9 and 10) with a size of 6 amino acids were found in only three DREB2A proteins of *O. Sativa* Indica, *O. glumipatula* and *O. brachyanta*. The motif distribution was almost same in all the proteins (Figure 1). The *p*-value of the motifs was zero in all the proteins except for *O. punctata* (4.21E-298), *O. glumipatula* (1.40E-296) and *O. brachyanta* (1.38E-240). Domain analysis using Pfam and SMART tools revealed the 11 orthologs DREB2A proteins to encode an AP2/ERF domain which has high homology to the DREB type proteins in *Arabidopsis thaliana*. Though the size of the domain was conserved across all the 11 species, the position of the domain was changed in *O. brachyanta* (76-127) from the conserved location of 81-132 amino acids (Table 3).

Secondary Structure Modelling and Molecular Analysis of DREB2A Proteins

The 3D structure models of all the 23 transcripts of the 11 DREB2A orthologs of diploid *Oryza* revealed many templates with structural similarities. For each species, the template with highest sequence similarity to the query sequence, highest QMEAN values and highest GMQE scores were selected. Based on the predicted structures, 21 protein transcripts with 274AA, 281AA or 284AA had identical 3D structures typical of DREB2A with distinct DNA binding domain and GCC box for protein functions and ligands interactions while the shortest transcripts; OMERI01G04110.2 – 60AA and OPUNC01G03970.3 – 86AA had a unique structure each that is devoid of any binding domain and GCC box (Figure 2).Additionally, all the transcripts derived from the 11 orthologs of DREB2A gene of diploid *Oryza* species revealed little variations in molecular weight of the protein.

Transcripts with higher number of amino acids revealed higher molecular weight proteins i.e. transcripts OPUNC01G03970.4 with 284AA has the highest molecular weight of 31972.63 Da while OMERI01G04110.2 transcript with 60AA has the lowest molecular weight of 6688.42 Da. Transcripts with 284AA has molecular weight of 31.8 KDa while those with 274AA has molecular weight of 30.5 -30.7 KDa. The transcripts with 274AA, 281AA and 284AA have similar PI values that range from 5.38 – 5.96 while the remaining two unique transcripts; OMERI01G04110.2 and OPUNC01G03970.3 with 60AA and 86AA have PI values of 4.85 and 4.73 respectively. The protein stability index revealed all the transcripts to produce unstable proteins that are typical of transcription factors except transcripts OMERI01G04110.2, OPUNC01G03970.1, OPUNC01G03970.2 and OPUNC01G03970.3 that have stable proteins.

Sequence Alignment and Phylogenetic Analysis

The DREB2A protein sequence of the 11 diploid *Oryza* species showed higher similarity in length i.e. 281 amino acid in all species except *O. brachyanta* that had a shorter protein with 276 amino acids. Though, little variation exists in terminal ends of the proteins, the phylogenetic analysis revealed higher similarities among the diploid species with 100% identical conserved AP2 domain. The constructed phylogenetic tree among the 11 diploid species displayed 2 clusters (*O. brachyanta* clade and other bigger clade with sub clusters accommodating the rest of the diploid species). Closed analysis of the bigger clade revealed 3 sub clusters with *O. punctata* on one sub cluster, *O. glumipatula* on another and the rest of the cultivated species (*O. glaberrima, O. sativa* Indica and *O. sativa* Japonica) and their progenitors (*O. rufipogon, O. nivara, O. barthii, O. longistaminata* and *O. meridionalis*) occupying the last sub cluster (Figure 3).

Table 3: Conserved DREB2A protein motifs identified in the 11 diploid Oryza species

Motif	E-value	Sites	Width	Amino Acid Sequence composition of motif	Pfam Domain
1	8.4e-480	11	50	CAYRGVRQRTWGKWVAEIREPNRGRRLWLGSFPTALEAAHAYDEAARAMY	AP2 domain
2	2.2e-477	11	50	VLHKEVNISYDYFNVHEVVEMIIVELSADQKTEVHEEYQEGDDGFSLFSY	Not found
3	1.7e-451	11	50	WVRKKRTRRKSDGPDSIAETIKWWKEQNQKLQEENSSRKAPAKGSKKGCM	Not found
4	1.1e-442	11	50	PTARVNFADNSTDANSGCTSAPSLMMSNGPATIPSDEKDELESPPFIVAN	Not found
5	2.1e-366	10	50	GPAVLYRPDKKDVLERVVPEVQDVKTEGSNGLKRVCQERKTMEVCESEGI	Not found
6	4.20E-109	9	15	MLFRFVSCNVQLCGI	Not found
7	3.60E-46	11	8	KGGPENSN	Not found
8	1.50E+05	2	6	GRRGDC	Not found
9	4.20E+06	2	6	ERKLMC	Not found
10	4.20E+06	2	6	RKWLGY	Not found

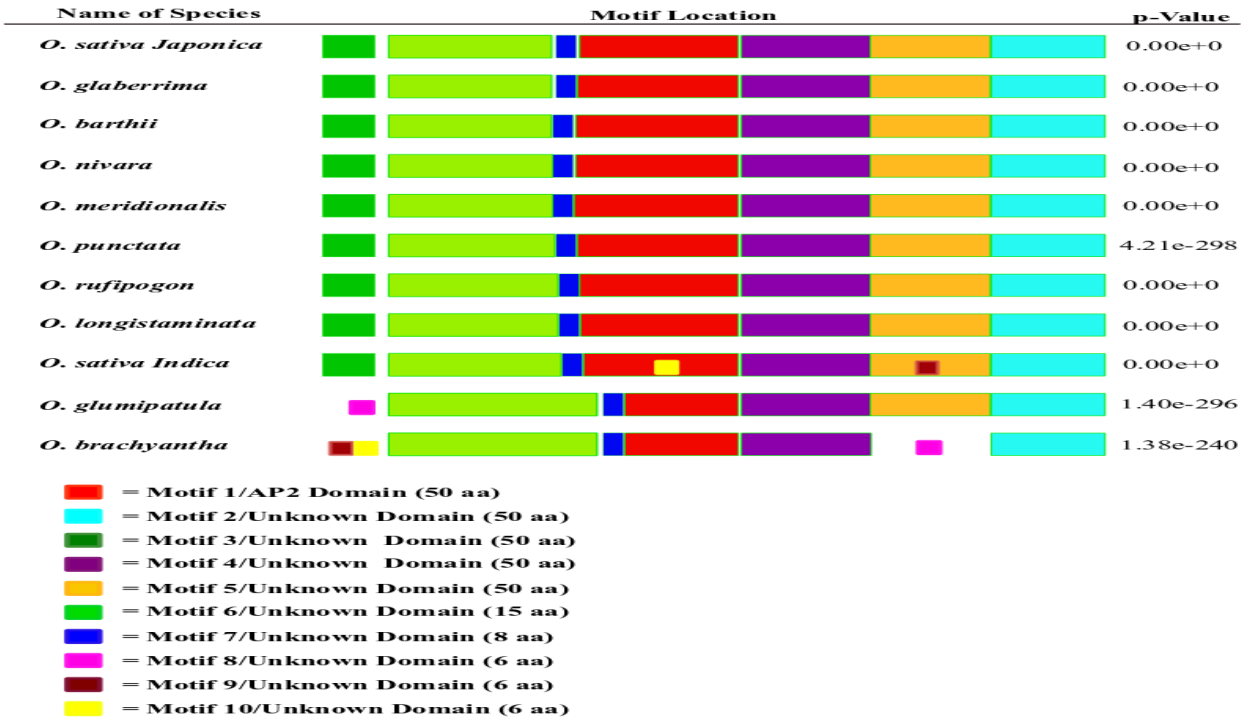

Figure 1: The conserved motifs of DREB2A proteins in diploid *Oryza* species. Each motif was represented in boxes with different colors as above. The number of amino acids of each motif is presented and p-value of each species is given by the right.

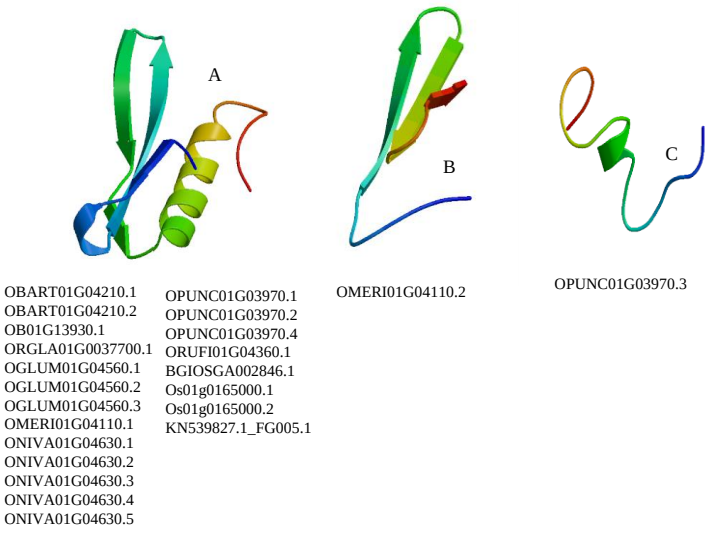

Figure 2: The predicted 3D model structure of DREB2A protein identical to the 11 diploid Oryza species using SWISS model tool of Expasy. (A) Predicted structure for 21 transcripts of 11 diploid Oryza species with 274, 281 and 284AA with distinct DNA binding domain and GCC box typical of DREB2A structure. (B) Predicted structure of OMERI01G04110.2 transcript of O. meridionalis with 60AA. (C) Predicted structure of OPUNC01G03970.3 transcript of O. punctata with 86AA.

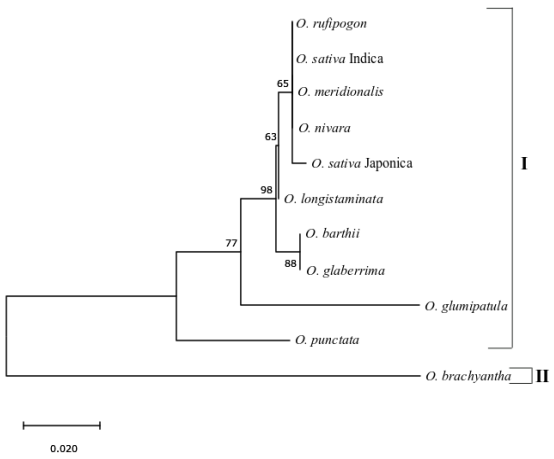

Figure 3: Phylogenetic tree based on amino acid sequences of DREB2A proteins of 11 diploid *Oryza* species.

Promoter Analysis

In silico analysis of the promoter regions (1kb upstream of TSS) of DREB2A identified a number of common, rare and unique *cis*-elements on both forward and reverse strands of the diploid *Oryza* species (Figure 4a). In total, 52% of the identified stress responsive elements are common to all the 11 diploid *Oryza* species and 36% are unique to individual species. Rare elements (present in only 2-3 species) accounts for only 12% of the stress responsive elements. The distribution and frequency of stress responsive cis elements across the diploid *Oryza* species revealed different elements responsible for drought, anoxic stress, light, methyl jasmonate, wound and biotic stress (Figure 4b). The frequency of occurrence of the different stress responsive elements at different position of the 1kb upstream of TSS (both on forward and reverse strand) of the diploid *Oryza* species showed reverse strand at 501-750bp region has the highest frequencies of stress elements while 251-500 bp region has the lowest frequencies of stress elements (Figure 5). In forward strand, the 501-750 bp regions have the highest frequency while 751-1,000 bp has the least frequencies of stress elements. Among the diploid species, *O. brachyantha* has the highest frequencies of occurrence of stress responsive elements (42) while the cultivated Asian species (*O. sativa* Indica and Japonica) has 21 each. African rice (*O. glaberrima*) and its progenitor *O. barthii* have 31 and 27 respectively while *O. nivara, O. meridionalis, O. rufipogon* and *O. punctata* have 26, 20, 20 and 25 respectively.

DISCUSSION

In this study, the CDS of DREB2A orthologs in 11 diploid *Oryza* had conserved sequence length (except in *O. brachyanta*) of 846 bp and located on chromosome 1 across all species. Previous studies on DREB2A sequence in *O. sativa* ssp Indica (Varshney *et al.,* 2009; Imesh De Silva *et al.,* 2014; Singh *et al.,* 2015) and *O. glaberrima* (Gumi *et al.,* 2018) affirmed the 846 bp sequence located on chromosome 1 of the cultivated species. In all the 11 diploid species, DREB2A proteins are predicted to be localized in cell's nucleus, a location that fits transcription factors. Many authors have reported that OsDREB2A and its homologs in grasses and *Arabidopsis* are localized in nucleus (Qin *et al.,* 2008; Matsakura*et al.,* 2010; Filiz and Tombuloglu, 2014; Herath, 2016). The DREB2A gene was reported to possess single gene copy across all the diploid species but however some species has multiple transcripts indicating alternative splicing has a role in their genome evolution. Variations in gene copy number affect gene dosage which in turn affects phenotypes of plants to variable extent (Zmienko *et al.,* 2013). In this context, the copy number has no role in any possible changes in DREB2A attributed functions but the variations in transcripts may contribute to regulation of DREB2A expression in the affected species. A number of reports have shown alternative splicing play a vital role in the differential regulation of stress responsive genes in plants at either protein or transcript level modifications (Resch *et al.,* 2004; Matlin *et al.,* 2005; Szakonyi and Duque, 2018). Our findings here suggests that alternative splicing may play a role in genome evolution of the diploid *Oryza* species considering DREB2A gene has single copy number in all the species but variant transcripts. Overall, the DREB2A gene and protein is conserved across the diploid species of rice both in size, chromosomal location and sub cellular localization

.

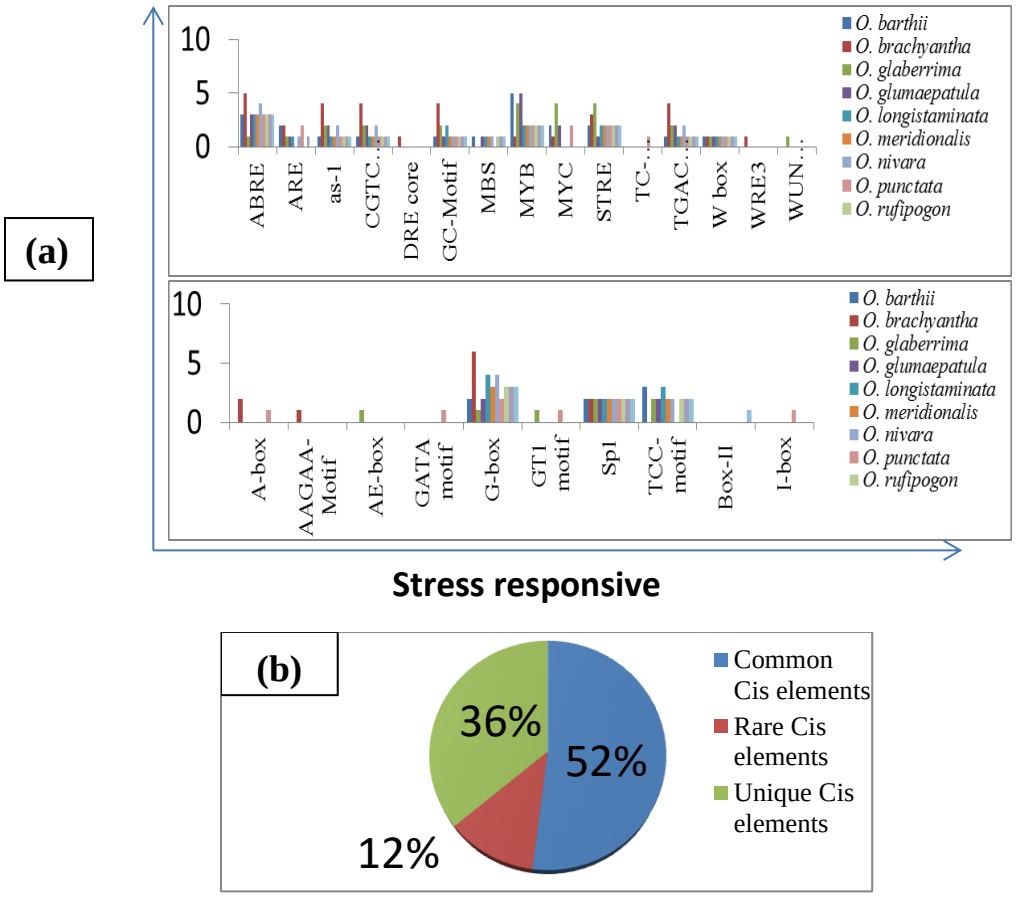

Figure 4: Distribution of cis-elements in the promoter region (1.5 kb) of DREB2A orthologs of 11 diploid *Oryza* species (a) Frequency of occurrence of cis-element motifs in the promoter region of DREB2A orthologs (b) distribution of cis-elements motifs of DREB2A orthologs based on biological functions.

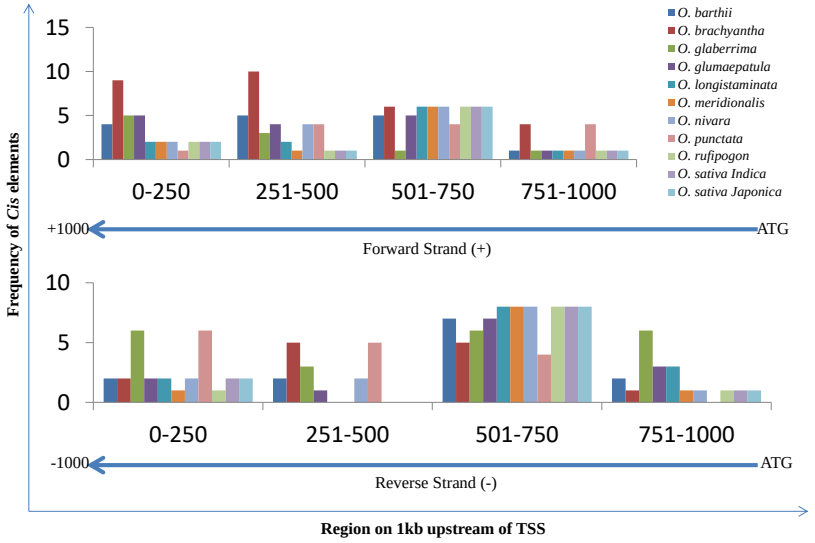

Figure 5: Distribution and frequency of occurrence of cis-elements in the forward and reverse promoter regions of DREB2A gene in 11 diploid *Oryza* species

Domain analysis of DREB2A protein revealed an AP2 type domain which was conserved across all the 11 diploid *Oryza* species. DREB family proteins belong to the AP2/ERF transcription factor family that contain one AP2/ERF binding domain that can bind to DRE *cis* element in the promoter of stress responsive genes (Yamaguchi-Shinozaki and Shinozaki, 1994; Sakuma *et al.,* 2006; Filiz and Tombuloglu, 2014). The conserved motifs derived from MEME suites indicated 7 significant motifs that are conserved and distributed in similar fashion among all the 11 diploid species but only motif1 had AP2 domain which is characteristic binding sites of DREB genes. The instability index suggests the unstable nature of the protein in all the diploid species while the negative Grand average hydropathicity index confirmed the hydrophilic nature of the DREB2A protein, which is a common feature of putative transcription factors. Generally, protein that has negative mean value for hydrophobicity have strong binding affinity for hydrophilic molecules such as DNA, thus suggest their DNA binding ability (Imesh De Silva *et al.,* 2014; Gumi *et al.,* 2018). In phylogenetic analysis, 11 DREB2A proteins from diploid *Oryza* species were used to understand the evolutionary relationship among rice species using neighbour joining method of MEGA 6. A tree with two main clusters was formed with *O. brachyanta* being the most diverged species which may be attributed to its size (833 bp as against 846bp in all diploid species). In the other cluster, *O. punctata* occupied separate sub cluster while the cultivated species along with their progenitor occupied a separate sub cluster. These results are in accordance with the phylogenetic tree of rice and its wild relatives (Ge *et al.,* 1999; Ammiraju *et al.,* 2010a, 2010b). The African rice (*O. glaberrima*) and *O. barthii* occupied same cluster which agrees with Sweeny and McCouch (2007) who reported *O. glaberrima* has been domesticated from its progenitor *O. barthii*. Additionally, the cultivated Asian rice (*O. sativa* Indica and *O. sativa* Japonica) are clustered on the same branch with *O. nivara* and *O. rufipogon* which also affirmed the reports of Chang (1976) that cultivated Asian rice were domesticated from *O. rufipogon* through *O. nivara.*

The analysis of 1kb promoter regions of the DREB2A in diploid *Oryza* species identified a number of stress and hormonal regulation *cis*-acting elements (such as DREs, G-box, GC motif, MYBs and ABREs elements) that are distributed differentially among the 11 diploid *Oryza* species. The DREB2A proteins are important plant transcription factors (TFs) that play a critical role in improving the abiotic stress tolerance of plants by interacting with a DRE/CRT cis-element present in the promoter region of many abiotic stress-responsive genes (Dubouzet *et al.,* 2003; Lata *et al.,* 2011). The ABREs has shown to be present in all the diploid *Oryza* species in variant proportions and play a key role in dehydration/drought response in rice that is dependent on ABA level. The DRE elements that solely interact with DREB genes were observed in only 3 wild species. A number of reports have shown that DREB genes in rice and *Arabidopsis* interact with DRE cis elements in the promoter region of stress responsive gene to regulate their expression in an ABA dependent pathway (Lata *et al.,* 2011). A number of cis elements involved in hormonal regulations are observed in the diploid species and are involved in response to light, abscisic acid, salicylic acid, methyl jasmonate, anaerobiosis and played a key role in ethylene induction (Yoshihara *et al.,* 1996; Siberil *et al.,* 2001). We identified a single unique *cis*-element (LTRE) of *O. nivara* that is associated with cold stress mitigation. Similar finding was observed in rice *SamDC* promoter region (Basu *et al.,* 2014).

Our findings showed wild species of diploid *Oryza* has high number of cis elements than the cultivated species, a trend that explains why wild species serve as reservoirs of important agronomic traits such as tolerance to stresses. Moreover, this high number of CAREs in wild species suggests that cis acting elements to be involved in genome evolution among the diploid *Oryza* species. In rice, DREB2A are known regulators of drought and salinity responses (Liu *et al.,* 1998; Sakuma *et al.,* 2006; Qin *et al.,* 2008; Matsakura *et al.,* 2010). Overall, the 1 kb promoter region of DREB2A in all the diploid species revealed a number of *cis* acting elements that are predominantly involved in abiotic stress response including elements of AP2/ERF, MYBs, ABA, DRE/CRT and bZIP transcription factors (Lee *et al.,* 2010; Kayum *et al.,* 2015; Baldoni *et al.,* 2015). Though unique *cis*-elements occurred in specific species, these might suggest why DREB2A transcription factors in the 11 diploid species are differentially regulated among species. Interestingly wild species had more number of unique *cis*-elements than the cultivated species thus, providing insight as to why wild relatives of rice are more tolerant to abiotic stress than cultivated species.

CONCLUSION

In this study, comparative *in silico* analysis of DREB2A protein was performed among 11 diploid *Oryza* species using orthologous sequence derived from BLASTN and BLASTP search of Gramene database. Based on our findings, (i) The AP2 domain and molecular properties of the DREB2A protein was conserved across all the diploid *Oryza* species with single gene copy but cultivated species has less number of transcripts, introns and exons than their progenitors suggesting gene retrogression in cultivated *Oryza* species and (ii) The identified common *cis*-element-ABRE in all the diploid species supports DREB2A as ABA dependent transcription factor and the high number of cis elements in wild species as compared to cultivated species suggests a possible role of cis elements in genome evolution of the diploid species.

REFERENCES

Bailey, T.L. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28-36, AAAI Press, Menlo Park, California.

Baldoni, E., Genga, A. and Cominelli, E. (2015). Plant MYB Transcription Factors: Their Role in Drought Response Mechanisms. *International Journal of Molecular Science,* **16**, 15811–15851. doi:10.3390/ijms160715811.

Bautista, N.S., Solis, R. and Ishii, T. (2001). RAPD, RFLP and SSLP analysis of phylogenetic relationships between cultivated and wild species of rice. *Genes & Genetic Systems,* **76**: 71–79.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L. and Schwede, T. (2014). SWISS-MODEL: Modeling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research,* **42:** W252–W258.

Chang, T.T. (1976). The origin, evolution, cultivation and diversification of Asian and African rices. *Euphytica,* **25**, 425–441.

Cheng, Z., Presting, G.G., Buell, C.R., Wing, R.A. and Jiang, J. (2001). High-resolution pachytene chromosome mapping of bacterial artificial chromosomes anchored by genetic markers reveals the centromere location and the distribution of genetic recombination along chromosome 10 of rice. *Genetics,* **157:** 1749–1757.

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B. H., Hong, X., Agarwal, M. and Zhu, J. K. (2003). ICE1: A regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. *Genes $ Development,* **17**: 1043–1054.

DeCastro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A. and Hulo, N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule associated functional and structural residues in proteins. *Nucleic Acids Research,* **34:** W362-5.

Dubouzet, J.G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E.D., Miura, S., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003). *OsDREB* genes in rice *Oryza sativa* L., encoded transcription activators that function in drought, high salt- and cold-responsive gene expression. *Plant Journal,* 33**:**751–763.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E. and Finn, R.D. (2019). The Pfam protein families database in 2019.*Nucleic Acids Research,* **47**, Database issue D427–D432.

Filiz, E. andTombuloglu, H. (2014). *In Silico* Analysis of DREB Transcription Factor Genes and Proteins in Grasses. Vol. *Applied Biochemistry and Biotechnology,***174**: 1272–1285.

Ganie, S.A., Debnath, A.B., Gumi, A.M. and Mondal,T.K. (2017).Comprehensive survey and evolutionary analysis of genome-wide miRNA from ten diploid *Oryza* species. *BMC Genomics,* **18:**711-726.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R. and Appel, R.D. (2005). Protein identification and analysis tools on the ExPASy server, In: *John M. Walker (ed): The Proteomics Protocols Handbook Humana, 571–607.*

Gumi, A. M., Guha, P. K., Mazumder, A., Jayaswal, P. and Mondal, T.K. (2018). Characterization of *OglDREB2A* gene from African rice (*Oryza glaberrima*), comparative analysis and its transcriptional regulation under salinity stress. *3 Biotech,* **8:** 91-96.

Haake, V., Cook, D., Riechmann, J.L., Pineda, O., Thomashow, M.F. and Zhang, J.Z (2002). Transcription factor CBF4 is a regulator of drought adaptation in Arabidopsis. *Plant Physiology,* **130**:639–648.

Herath, V. (2016). Small Family, Big Impact: In silico analysis of DREB2 transcription factor family in rice. *Computational Biology and Chemistry*, **10**, 12-25.

Higo, K., Ugawa, Y., Iwamoto, M. and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Research,* **27**: 297–300.

Imesh De Silva, W.S., Jayasekera, G.A.U., Fernandopulle, N.D. and Hettiarachchi, C. (2014). Identification, cloning and in-silico characterization of drought inducible OSDREB2A transcription factor from indica rice variety. *International Journal of Advanced Biotechnology and Research,* **5**(2), 117-125.

Kayum, M.A., Jung, H-J., Park, J-I., Ahmed, N.U., Saha, G., Yang, T-J.and Nou, I-S. (2015). Identification and expression analysis of WRKY family genes under biotic and abiotic stresses in Brassica rapa. *Molecular Genetics and Genomics,* **290**:79–95.

Kersey, P.J., Allen, J.E., Armean, I., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C. and Humphrey, J. (2016). Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Research,* **44 (D1)**: D574-D580.

Khush, G.S. (1997). Disease and insect resistance in rice. *Advances in Agronomy,* ***29***: 265-341.

Kizis, D. and Pages, M. (2002). Maize DRE-binding proteins DBF1 and DBF2 are involved in rab17 regulation through the drought-responsive element in an ABA-dependent pathway. *Plant Journal*, **30**:679–689.

Knight, H., Zarka, D.G., Okamoto, H., Thomashow M.F. and Knight, M.R. (2004). Abscisic acid induces CBF gene transcription and subsequent induction of cold-regulated genes via the CRT promoter element. *Plant Physiology,* **135:**1710–1717.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, **35**:1547-1549.

Lata, C., Bhutty, S., Bahadur, R.P., Majee, M. and Prasad, M. (2011). Association of an SNP in a novel DREB2-like gene SiDREB2 with stress tolerance in foxtail millet (*Setaria italica* L.). *Journal of Experimental Botany,* **62**: 3387–3401.

Lee, S-J., Kang, J-Y., Park, H-J., Kim, M.D., Bae, M.S., Choi, H-I. and Kim, S.Y. (2010). DREB2C interacts with ABF2, a bZIP protein regulating abscisic acid-responsive gene expression, and its overexpression affects abscisic acid sensitivity. *Plant Physiology,* **153**:716–727.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y. and Van De Peer, Y. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, **30**, 325–327.

Letunic, I., Doerks, T. and Bork, P. (2014). SMART: recent updates, new development and status in 2015. *Nucleic Acids Research,* **43:** D257-D260.

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 binding domain separate two cellular signals transduction pathways in drought- and low temperature responsive genes in *Arabidopsis*. *Plant Cell,* **10:** 1391-1406.

Liu, Q., Wang, H., Zhu, L., Hu, H. and Sun, Y. (2013). Genome-wide identification and analysis of miRNA related single nucleotide polymorphisms (SNPs) in rice. *Rice,* **6**: 10-14.

Mallikarjuna, G., Mallikarjuna, K., Reddy, M.K. and Kaul, T. (2011). Expression of *OsDREB2A* transcription factor confers enhanced dehydration and salt stress tolerance in rice (*Oryza sativa* L.). *Biotechnology Letters,* **33:** 1689-1697.

Matlin, A.J., Clark, F. and Smith, C.W.J. (2005). Understanding alternative splicing: towards a cellular code. *Nature Reviews Molecular Cell Biology*, **6**: 386–398.

Matsukura, S., Mizoi, J., Yoshida, T., Todaka, D., Ito, Y., Maruyama, K., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2010). Comprehensive analysis of rice DREB2-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. *Molecular Genetics and Genomics,* **283**: 185-196.

Nakashima, K., Ito, Y. and Yamaguchi-Shinozaki, K. (2009). Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. *Plant Physiology,* **149:** 88-95.

Nakashima, K., Shinwari, Z.K., Sakuma, Y., Seki, M., Miura, S., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2000). Organization and expression of two *ArabidopsisDREB2* genes encoding DRE-binding proteins involved in dehydration- and high-salinity responsive gene expression. *Plant Molecular Biology,* **42:** 657–665.

Qin, F., Kakimoto, M., Sakuma, Y., Maruyama, K., Osakabe, Y. and Tran, L. S. P. (2008). Regulation and functional analysis of ZmDREB2A in response to drought and heat stresses in Zea mays L. *Plant Journal*, **50**, 54–69.

Resch, A., Xing, Y., Modrek, B., Gorlick, M., Riley, R. and Lee, C. (2004). Assessing the impact of alternative splicing on domain interactions in the human proteome. *Journal of Proteome Research,.***3**:76–83.

Sakuma, Y., Maruyama, K., Osakabe, Y., Qin, F., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2006). Functional analysis of an *Arabidopsis* transcription factor, DREB2A, involved in drought-responsive gene expression. *Plant Cell,* **18:**1292-1309.

Sanchez, P.L., Wing, R.A. and Brar, D.S. (2014). The wild relatives of rice: Genome and Genomics. In: Zhang, Q. and Wing, R.A. (eds.), *Genetics and Genomics of Rice, Plant Genetics and Genomics: Crops and Models 5.* Springer science+ Business media, New York, pp. 193.

Siberil, Y., Doireau, P. and Gantet, P. (2001). Plant b ZIP G-box binding factors Modular structure and activation mechanisms. *European Journal of Biochemistry,* **268(22)**: 5655-5666.

Singh, N.K., Singh, B.P., Jayaswal, P.K., Singh, P.K., Singh, B., Kumar, V., Mishra, S., Singh, N. and Panda, K. (2015). Natural allelic diversity in OsDREB1F gene in the Indian wild rice germplasm led to ascertain its association with drought tolerance. *Plant Cell and Reproduction,* **15**:1760-1766.

Szakonyi, D. and Duque, P. (2018). Alternative Splicing as a Regulator of Early Plant Development. *Frontiers in Plant Science,* **9:**1174.

Tello-Ruiz, M.K., Naithani, S., Stein, J.C., Gupta, P., Campbell, M., Olson, A., Wei, S., Preece, J., Geniza, M.J., Jiao, Y., Lee, Y.K., Wang, B., Mulvaney, J., Chougule, K., Elser, J., Al-Bader, N., Kumari, S., Thomason, J., Kumar, V., Bolser, D.M., Naamati, G., Tapanari, E., Fonseca, N., Huerta, L., Iqbal, H., Keays, M., Munoz-Pomer, Fuentes, A., Tang, A., Fabregat, A., D'Eustachio, P., Weiser, J., Stein, L.D., Petryszak, R., Papatheodorou, I., Kersey, P.J., Lockhart, P., Taylor, C., Jaiswal, P. and Ware, D.(2018). Gramene 2018: unifying comparative genomics and pathway resources for plant research. *Nucleic Acids Research,* **4***,* 46(D1):D1181-D1189.

Varshney, R.K., Nayak, S.N., Balaji, J., Upadhyaya, H.D., hash, C.T., Kavi Kishor, P.B., Chattopadhyay, D., Rodriquez, L.M., Blair, M.W., Baum, M., McNally, K., This, D. and Hoisington, D.A. (2009). Isolation and sequence analysis of DREB2A homologues in three cereals and two legumes species. *Plant Science*, **177**, 460-466.

Yamaguchi-Shinozaki, K. and Shinozaki, K. (1994). A novel cis-Acting Element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *Plant Cell,* **6** (2): 251-264.

Yoshihara, T., Washida, H. and Takaiwa, F. (1996). A 45-bp proximal region containing AACA and GCN4 motif is sufficient to confer endosperm-specific expression of the rice storage protein glutelin gene *GluA-3*. *FEBS Letters,* 383.