International Journal of Agriculture, Environment and Biotechnology
Citation: IJAEB: 9(5): 773-785, October 2016
DOI: 10.5958/2230-732X.2016.00100.5

GENETICS AND PLANT BREEDING

PCR amplification and bioinformatics assessment of promoters of *PBF-DOF* (DNA binding with one finger) genes of finger millet

Chandra Prakash[1, 2], Hariom Kushwaha[1, 3], Vinay Kumar Singh[1, 4], Praveen K Singh[5], Anil Kumar[1] and Dinesh Yadav*[1, 6]

[1]Department of Molecular Biology and Genetic Engineering, College of Basic Sciences and Humanities, G.B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand 263 145, India
[2]National Research Centre on Plant Biotechnology, IARI, New Delhi 110012, India
[3]Institue of Microbial Technology (CSIR), Sector 39-A, Chandigarh 160036, India
[4]Centre for Bioinformatics, School of Biotechnology, Banaras Hindu University, Varanasi, Uttar Pradesh-221 005, India
[5]Division of Seed Science and Technology, Indian Agriculture Research Institute, New Delhi 110012, India
[6]Department of Biotechnology, D.DU Gorakhpur University, Gorakhpur, Uttar Pradesh-273 009, India

*Corresponding author: dinesh_yad@rediffmail.com

Paper No. 503 Received: 11-2-2016 Accepted: 21-9-2016

Abstract

The Dof (DNA binding with one finger) family is a plant specific transcription factor known to be involved with regulating diverse functions in plants and have been extensively studied in many crops. The Dof transcription factor regulating gene expression by interacting with *Cis*-regulatory elements namely prolamin box (P box), GCN4, AACA and ACGT motifs present in the promoters of seed storage protein genes is known as PBF (Prolamin-box Binding Factor) Dof transcription factor. A set of 15 primers were designed by considering approximately 1.5 kb upstream and 500bp downstream sequences of full length *Dof* genes from TSS of cereals like rice, wheat and sorghum available in databases. These primers were used for PCR amplification of putative promoters of *Dof* genes of finger millet along with few cereals and millets. Furthermore, based on the presence of expected size amplicon with different sets of primers tested, a total of 6 bands of expected size representing putative promoters of PBF-*Dof* genes of rice, sorghum, barnyard, finger millets (PRM-1 PRM-801, PRM-701) were eluted, sequenced and subjected to *in silico* investigation. The bioinformatics based characterization revealed uniform presence TSS and numerous seed storage protein specific motifs like DPBF motif, RY element, SKN1 motif, GCN4 motif, E-Box confirming the promoters of respective *PBF-Dof* genes of cereals and millets. Further, validation by cloning in promoter probe vector is required for confirmation of temporal and spatial expression associated with seed storage protein genes.

Highlights

- Primer were designed from rice, wheat and sorghum *Dof* gene promoters and PCR was attempted in finger millet, rice and sorghum.
- In silico analysis of cloned sequences revealed TSS and numerous seed storage protein specific motifs
- PRM-1, PRM-701 and PRM-801 showed presence of three DNA motifs which are binding sites for various well characterized transcription factors e.g. MYB, abi4, SOC, FLC, PBF, Dof 2 etc.
- Presence of these conserved motifs identifies cloned sequences as putative *Dof* promoter candidates.

Keywords: Cereals, millets, prolamin-box binding factor (PBF), Dof (DNA binding with one finger) transcription factor, *In-silico*, promoters

Transcription factors (TFs) play a key role in regulating transcriptional circuit to meet organism's developmental and adaptive requirements. Most TFs are modular proteins comprising of a DNA-binding domain that interacts with *cis*-regulatory elements of target gene promoters and a protein-protein interaction domain that facilitates oligomerization between TFs or other regulators (Wray *et al.*, 2003). Transcription factors are important regulators of gene expression comprising of at least four distinct domains, DNA-binding domain, nuclear localization signals (NLS), transcription activation domain, and oligomerization site, which functions together to regulate transcription initiation of many target genes of physiological and biochemical processes (Du *et al.*, 2009). TFs are classified into different families by comparing structural and functional information of conserved DNA-binding domains. TFs exists in gene families that show diversity in size and functional redundancy among organisms (Riechmann and Ratcliffe, 2000; Wray *et al.*, 2003). It is speculated that organismal complexity associates with an increase in the absolute number and the proportion of TFs in a proteome (Levine and Tjian, 2003) and accelerated expansion among plant TF genes and their tendency for parallel expansion suggest their adaption to selection pressure in higher plants (Shiu *et al.*, 2005).

A family of TFs putatively specific to plants is the DNA-binding with One Finger (DOF) family which has been extensively reviewed (Takatsuji, 1998; Liu *et al.*, 1999; Riechmann and Ratcliffe, 2000; Yanagisawa, 2002, 2004; Kushwaha *et al.*, 2011; Le Hir and Bellini, 2013; Noguero *et al.*, 2013; Gupta *et al.*, 2015). The Dof proteins are typically composed of 200-400 amino acids with a conserved DNA binding Dof domain of 52 amino acid residues structured as a Cys2/Cys2 Zn finger recognizing a *cis* regulatory element with the common core sequence 5′-AAAG-3′ (Yanagisawa and Schmidt, 1999*b*; Yanagisawa, 2002; Umemura *et al.*, 2004). The Dof domain is a bifunctional domain that mediates not only DNA-protein interaction but protein –protein interaction also (Yanagisawa, 1997; Krohn, 2002). There exists great diversity in terms of number of *Dof* genes in different crops. The number of Dof genes predicted in rice, barley, wheat, maize and sorghum is 30, 24, 31, 54 and 28 respectively using various computational tools (Lijavetzky *et al.*, 2003; Moreno-Risueno *et al.*, 2007; Libault *et al.*, 2009; Shaw *et al.*, 2009; Kushwaha *et al.*, 2011).

The DOF transcription factor participates in tissue differentiation, metabolism and seed development and regulates gene expression during seed development by binding with c*is*-regulatory elements namely prolamin box (P box), GCN4, AACA and ACGT motifs present in seed storage protein genes which is referred as PBF (Prolamin-box Binding Factor) DOF transcription factor (Yamamoto *et al.*, 2006). These genes have been well characterized in maize (Vicente-Carbajosa *et al.*, 1997; Marzabal *et al.*, 2008), barley(Mena *et al.*, 1998; Diaz *et al.*, 2002, 2005), rice (Yamamoto *et al.*, 2006) and wheat (Ravel *et al.*, 2006; Dong *et al.*, 2006).

Prolamines encoding genes are systematically expressed in the developing endosperm under spatial and temporal transcription control of *cis*-acting motifs in their promoters and *tans*-acting transcription factors. Several consensus sequences in gene promoters have been shown involved in imparting endosperm specificity in cereals (Mena *et al.*, 1998; Washida *et al.*, 1999; Wu *et al.*, 2000). The prolamin box (P-box) is a highly conserved 7-bp sequence element (5′-TGTAAAG-3′) found in the promoters of many cereal seed storage protein genes, approximately 300 nucleotides upstream of the start codon (Vicente-Carbajosa *et al.*, 1997; Xu and Messing, 2009). The P-box binding factor (PBF) interacts with the P-box as an endosperm-specific transcriptional activator that belongs to the Dof class of plant zinc-finger DNA-binding proteins (Forde *et al.*, 1985; Wu and Messing, 2012).

Several conserved *cis*-elements have been discovered within the promoters of the prolamin genes of cereals, including the endosperm box (EB) (Kreis *et al.*, 1985) and the ACAA motif (Takaiwa *et al.*, 1996; Diaz *et al.*, 2002). The EB consists of two distinct protein-binding sites the GCN4-like motif (GLM: 5′-ATGAG/CTCAT- 3′) and the prolamin box , also referred to as endosperm motif (Colot *et al.*, 1989). TFs from the basic leucine zipper (bZIP), DOF, and MYB classes bind to the GLM, the PB, and the ACAA *cis*-elements (Hammond-Kosack *et al.*, 1993; Suzuki *et al.*, 1998; Diaz *et al.*, 2002).An interaction network of *cis*-elements and their TFs in barley regulates seed storage protein genes (Rubio-Somoza

Table 1: List of primers for the PCR amplification of Dof domains

Sl. No	Primer	Locus/Source	Primer sequences forward 5' – 3'	Primer sequences reverse 5' – 3'	Amplicon size (in bp)
1	Dof-01	OsIBCD010967	AGGGGGACGGTGCAAGAGATG	GCGGCTGGCTGAGGCTGTAG	993
2	Dof-02	OsIBCD041244	TATCCTCCAGGTCTAGGCTGTC	GGCTGAGGCTGTAGTTGTTGTA	946
3	Dof-03	Os01g09720	GTACTAATCGCGGTTGCTCTAT	TAGTAGCAGAACTTGGTGTTGG	1112
4	Dof-04	Os01g48290	CTAGCTTTGCTCACGTTGTG	TGAAGTAGCAGAACTTGGTCTC	1263
5	Dof-05	Os03g60630	CTTTCTCTAGCTTCCCTCCTCT	GAAGTAGCAGAACTTGGTGTTG	1043
6	Dof-06	Os02g15350 (rpbf)	ACAAACACACACACACACAC	CGCCATGCTGTAGTTGTTGTA	1100
7	Dof-07	Consensus Oryza sativa	AAGAAGAGAAGAAGAAGGAGAT	CGAGTAGTTGTTGAAGTAGCAG	775
8	Dof-08	SBDOF1 (sorghum)	CCAAAACCCGGCAAAGGAATC	GCTGCGCCGTGTTGTAGTTGTT	1021
9	Dof-09	SBDOF2(sorghum)	ATGTTTGGCGGGTATGGGTGAC	GGCGTGGCTGGTTGATGTTGTA	2233
10	Dof-010	Sb08g000330.1 (sorghum)	GGCTACTTCCTACTTGGGTTAC	GACAGCCCTTGCAAAAGTAG	944
11	Dof-011	DOF2-chr3(sorghum)	CACTTTTCCGTGTTGCTTAG	CGCACGTTGTAGTTGTTGAA	1056
12	Dof-012	Uncharecterized DOF(sorghum)	GTGGTAGCCTTTGATTTGTAGG	TGAAGTAGCAGAACTTGGTGTC	1957
13	Dof-013	Sb chr01(sorghum)	AGGTCAGCCTTAGCAATTAAGC	GGGAGAGGGAGTAGTTGTTGT	1115
14	Dof-014	Sb chr04 g030420.1	GCTCTCTAGCGTATAGCGATCT	CGAGAGGCTGTAGTTGTTGTAG	2607
15	Dof-015	AF385139.1 (wheat)	GCTGAGGCGGCCAATAGTAGAGA	ACCGCAAAATGAGCAAGCCAGATA	1318

et al., 2006), and it is conserved in other cereals also (Verdier and Thompson, 2008; Xi and Zheng, 2011). RY element and Skn1motif is also present in promoters of several seed storage protein genes. Expression of seed storage protein genes of cereals is directly influenced by promoters with conserved seed specific motifs (Yadav *et al*., 2007; 2008). Since the single transcription factor i.e. Dof family is being involved with multifarious roles specific to plants, it is important to study the promoters of various Dof gene(s) so as to decipher the complexity of gene regulation in plants.

Cereals including rice, maize, wheat, barley, rye, sorghum, oats and millets are considered to be the most important group of cultivated plants in terms of food production and acreage covered, providing most of the calories and proteins requirement of our daily diet (Varshney *et al.*, 2006). Though there exists great diversity among cereals in terms of genome size, ploidy level and chromosome numbers, attempts have been made to reveal the existing synteny and colinearity on the basis of comparative genomics (Kellogg and Birchler, 1993; Kellogg, 1998; Devos, 2000; Feuillet and Keller, 2002; Caetano-Anollés, 2005; Li *et al.*, 2008; Paterson *et al.*, 2009; Chen and Cao, 2015).

This paper reports the PCR based amplification, sequencing and *in silico* characterization of promoters of *Pbf-Dof* genes of finger millet along with few cereals and millets.

Materials and Methods

The seeds of finger millets i.e. PRM-1 (Brown),PRM-801(White), PRM-701(Golden) and Barnyard millet (PRJ-1) were kindly provided by Dr. V. K. Yadav, Department of Genetics & Plant Breeding , Ranichauri Hill Campus, G.B.P.U.A & T, Pantnagar while seeds of Proso millet, Little millet, Kodo millet were provided by Dr. K.T. Gowda, project co-ordinator, Project Coordination cell, All India Co-ordinated Small Millets Improvement Project, ICAR, UAS, GKVK, Bangalore. Seeds of rice var. Sughadha and sorghum (SPV-462) were collected from College of Agriculture, G. B. P. U. A. & T, Pantnagar.

Table 2: PCR amplification profile showing no of bands and expected size amplicons obtained

Sl.No	Primer	Rice	Sorghum	PRM-1	PRM-801	PRM-701	Barnyard millet	Proso millet	Little millet	Kodo millet
1	Dof-01	3*	0	1	2	3	4*	4	4	4
2	Dof-02	3*	0	2*	3*	2*	1*	2*	0	3*
3	Dof-04	2*	0	4*	4*	3*	4	4*	10*	5*
4	Dof-07	2*	0	1	1	1	4	2	1	1
5	Dof-08	0	1*	2	2	2	0	0	0	1
6	Dof-11	0	2*	3	2	2	1	2	2	3
7	Dof-12	0	1*	3	3	3	7	4	6	4
8	Dof-13	0	2*	4	6	6	2	6	4	1
9	Dof-14	0	1*	0	0	0	4	1	0	1
10	Dof-15	0	5*	7	11	9	7	3	1	8

*Presence of expected size amplicon

Genomic DNA isolation, purification and quantification

The selected seeds of rice, sorghum & different millet varieties were germinated on the water soaked filter paper after surface sterilization using 0.1% $HgCl_2$ for 5 minutes. These were further incubated at 37°C for at least a week in the dark. The etiolated seedlings obtained were then harvested for standard DNA extraction method using CTAB buffer (Murray and Thompson, 1980). The spectrophotometric quantification of isolated genomic DNA and its analysis on 0.8% agarose gel was done by standard methods.

Primer Designing

Nucleotide sequences of different *Dof* genes promoter from rice, wheat and sorghum were retrieved from DRTF, NCBI and PHYTOZOME. Approximately 1000bp upstream from TSS and 500bp downstream from TSS containing CPRC region of Dof was taken. Primers upstream to TSS having nucleotide repeats were preferred because of conserved nature of repeats and therefore will increase probability of success in cross species amplification. To predict TSS and CPRC region Fgenesh online tool was used. Primers were made by selecting only significant sequences which confirmed the presence of seed storage regulatory motifs. e.g. GCN4, RY-element, O2 motif, SKN1 motif etc. Dof promoter specific primers was designed using BioEdit v.7.2.5 (Hall, 1999) and Primer3 (Untergasser *et al.*, 2012). The list of primers used in present study is shown in Table 1.

PCR amplification, gel elution and cloning

The standardization was carried out for PCR amplification of putative Pbf-Dof promoters with different sets of primers at different annealing temperatures. The concentration of template i.e. 100 ng was uniformly kept constant along with the concentration of primers i.e. 30 ng per reaction. The amplicons were analyzed on 1.5% agarose gel.

The expected size bands were gel eluted by Gel elution kit (EZ-10 spin column, Bio Basic Inc.) and quantified spectrophotometrically and also analyzed on agarose gel using standard DNA (Lambda *Hind* III marker DNA). The eluted product was then cloned in pJET1.2/blunt cloning vector using CloneJet™PCR cloning Kit (Fermentas) as per the instructions of the manufacturer. The cloned product was then subjected to $CaCl_2$ mediated transformation (Singh *et al.*, 2010). The recombinant colonies were analyzed for the presence of the cloned product by standard minipreparation of plasmid DNA (Sambrook *et al.*, 1989). Further confirmation was done by using the purified plasmid DNA isolated from the transformed colonies as template for PCR amplification using sequencing primers provided by the cloning kit and also by the primers used for amplification of the putative promoters of *Pbf-Dof* genes**.** The PCR amplicons were sequenced using respective primers.

***In silico* analysis**

Various online and offline Bioinformatics tools were used for analysis of sequenced products. CLUSTAL-W (Thompson *et al.*, 1994) software was

Table 3: Promoter Prediction by Neural network algorithm

Crop	Start	End	Score	Promoter sequence
Rice	26	76	1	TGTGATTATATATATATATAGCATTTTGGCAAGTGAATAA**A**ATTGTTCTC
	51	1	0.98	AAATGCTATATATATATATAATCACATGGAAATAAGCAGA**G**CACCGGAAA
Sorghum	53	3	0.78	GTCTCGGTTAACCTATAAAACCAAAGATTTCAGCCCAGAT**A**CTTAAGTCT
Barnyard	1	51	0.85	CCTTTTAAAAAAAAATCGAGCCATCTACCATGCATGTTCT**A**CGTGCAAGC
	194	244	0.77	CTCAAATAAATAAAAAAACTTCCGTATTAGGTTTGGAGTT**T**TGCTTGACC
	374	324	0.73	TTGTAACAATTGGTATCAGAGGTCGGGGTCCGGGCGGGGC**A**CGATGGAAT
	181	131	1	ACGTGAGAAATAAAAAAAGGGGGGACGTGTTGGTGTACCG**T**TTGGTACGT
PRM1	456	506	0.78	TCTTCGGGTGTAAATGAGTCTATGCCGTGAGTTGGAGGCT**A**GTCCCTACA
	97	47	0.74	ACAGCAATAACAGCACGGCTGCACAGTGCACAGGACGGGG**T**GGACAGTAG
PRM701	28	78	0.85	ACAGCTGCGCCACTGCAAAAGCAGCCACCCCGTCCTGTGC**A**CTGTGCAGC
	94	44	0.72	ACAGCAATAACAGCACGGCTGCACAGTGCACAGGACGGGG**T**GGCTGCTTT
PRM801	14	64	0.85	ACAGCTGCGCCACTGCAAAAGCAGCCACCCCGTCCTGTGC**A**CTGTGCAGC
	439	489	0.78	TCTTCGGGTGTAAATGAGTCTATGCCGTGAGTTGGAGGCT**A**GTCCCTACA
	80	30	72	ACAGCAATAACAGCACGGCTGCACAGTGCACAGGACGGGG**T**GGCTGCTTT

*Transcription start shown in larger font

used for multiple sequence alignment of sequences obtained with their respective source sequences from which primers were designed. NCBI blast was also done for homology search with other Dof sequences. PHYTOZOME (Goodstein *et al.*, 2012) blast was also used for homology search for sorghum sequence. Transcription start site (TSS) prediction is a first pre-requisite for any promoter analysis and therefore a neural network based algorithm (Reese, 2001) was used to predict promoter in our cloned sequences. For identification of *cis*-regulatory seed storage protein motifs online tools namely PLACE (Higo *et al.*, 1998) and PlantCARE (Lescot, 2002) were used. Motif discovery, Enrichment, scanning and comparison were done using MEME suit (Bailey *et al.*, 2009). MEME was used to discover motifs in 6 cloned promoter sequences, and then these motifs were searched in Gene Ontology for Motifs tool (GOMo) for gene ontology association with these DNA motifs. To search presence of similar motifs in other known plant database, TOMTOM (Motif Comparison Tool) was used. Motif Cluster Alignment and Search Tool (MCAST) was used to search for the presence of these conserved motifs in our 6 cloned and sequenced amplicons.

Results and Discussion

Millets, in general, have not been subjected to extensive genomics study due to the lack of complete genome information but it shows collinearity with cereals especially rice (Devos *et al.*, 1998; Srinivasachary *et al.*, 2007). *Dof* family is being involved with multifarious roles specific to plants, it is important to study the promoters of various *Dof* gene(s) so as to decipher the complexity of gene regulation in plants.

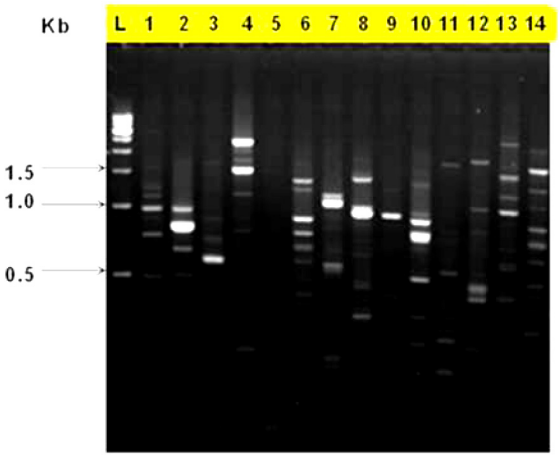

Fig. 1. PCR amplification of promoters of *Pbf-Dof* genes using 14 sets of primers. Lane L, 500 bp ladder; Lane1, rice (primer Dof-01); Lane 2, rice (primer Dof-02); Lane 3, rice (primer Dof-04); Lane 4, rice (primer Dof-05); Lane 5, rice (primer Dof-06); Lane 6, rice (primer-Dof-07); Lane 7, sorghum (primer Dof-08); Lane 8, sorghum (primer Dof-09); Lane 9, sorghum (primer Dof-10); Lane 10, sorghum (primer Dof-11); Lane 11, sorghum (primer Dof-12); Lane 12, sorghum (primer-Dof-13); Lane 13, sorghum (primer Dof-14) and Lane 14, wheat (primer Dof-15).

In the present study primers designed from *Dof* promoter region of rice, wheat and sorghum were

Table 4: Functions of different identified motifs

Motifs	Sequence	Function
5`UTR Py Stretch	TTTCTTCTCT	Cis-acting element conferring high transcription levels
AAGAA	GAAAGAA	Function not known
A-box	CCGTCC	Cis-acting regulatory element
ABRE	GCCGCGTGGC	Cis-acting element involved in the abscisic acid responsiveness
AC-1	CCCACCTACC	Function not known
ARE	TGGTTT	Cis-acting regulatory element essential for the anaerobic induction
Box S	AGCCACC	Function not known
BOXIII	CATTTACACT	Protein binding site
CAAT-Box	CAAAT	Common cis-acting element in promoter and enhancer regions
CCGTCC	CCGTCC	Cis-acting regulatory element related to meristem specific activation
Circadian	CAAAGATATC	Cis-acting regulatory element involved in circadian control
CTAG	ACTAGCAGAA	Function not known
DOF core	AAAG	Core site required for binding of Dof proteins
DPBF	ACACNNG	Embryo specific expression in carrot and Arabidopsis thaliana
E –Box	CANNTG	E-box of napA storage-protein gene of Brassica napus
G-Box	CACGTA	Cis-acting regulatory element involved in light responsiveness
GCN4	TGTGTCA	Cis-regulatory element involved in endosperm expression
HSE	AAAAAATTTC	Cis-acting element involved in heat stress responsiveness
MNF-1	GTGCCC(A/T)(A/T)	Light responsive element
Motif-1	gGTACGTGGCG	Cis-acting regulatory element root specific
MRE	AACCTAA	MYB binding site involved in light responsiveness
MYB Core	CNGTTR	Regulation of flavonoid biosynthesis
P- Box	GCCTTTTGAGT	Gibberellin-responsive element
Pyramidine Box	CCTTTT	Gibberellin-respons cis-element of GARE and pyrimidine box are partially involved in sugar repression
RY element	CATGCATG	Cis-acting regulatory element involved in seed-specific regulation
SKN 1	GTCAT	Cis-acting regulatory element required for endosperm expression
SP1	CC(G/A)CCC	Light responsive element
TAAAG	TAAAG	Target site for trans-acting StDof1 protein controlling guard cell-specific gene expression.
TATA-Box	TATAAA	Core promoter element around -30 of transcription start
TC rich repeat	ATTTTCTTCA	Cis-acting element involved in defense and stress responsiveness
TGA element	AACGAC	Auxin-responsive element
W box	TTGACC	Fungal elicitor responsive element
WRKY	TGAC	Binding site of rice WRKY71, a transcriptional repressor of the gibberellin signaling pathway
WUN motif	TCATTACGAA	Wound-responsive element
GT1	GGTTAA	Light responsive element

Source: Plant Care & PLACE

used in millets varieties, rice and sorghum with an aim to amplify various *Dof* promoters.

Amplification profiling with *Dof* primers in Millets, Rice and Sorghum

To check the amplification profile, a set of 15 primers were tested for PCR amplification of putative PBF-Dof promoters using isolated genomic DNAs of different millets, sorghum and rice as template DNA. Since the primers were designed from the available full length Dof genes of sorghum, rice and wheat, preliminary standardization was carried out

Table 5: Diversity of motifs present in sequences of putative promoters of PBF-Dof gene(s)

Sl. No	Motifs	Rice	Sorghum	Barnyard	PRM-1	PRM-701	PRM-801
1	5`UTR Py Stretch	0	0	0	1	1	1
2	AAGAA	0	0	0	1	1	1
3	A-box	0	0	0	1	1	1
4	ABRE	1	0	1	0	0	0
5	AC-1	0	0	1	0	0	0
6	ARE	0	1	0	0	0	0
7	Box S	0	0	0	0	0	1
8	BOXIII	0	0	0	1	0	1
9	CAAT-Box	0	1	1	1	1	1
10	CCGTCC	0	0	0	1	1	1
11	Circadian	0	1	1	0	0	0
12	CTAG	0	1	0	0	0	0
13	DOF core	1	1	1	1	1	1
14	DPBF	0	0	1	1	0	1
15	E –Box	1	0	1	1	1	1
16	G-Box	1	0	0	1	1	1
17	GCN4	0	0	0	1	0	1
18	HSE	0	0	1	0	0	0
19	MNF-1	0	0	0	1	1	1
20	Motif-1	0	0	1	0	0	0
21	MRE	0	0	1	0	0	0
22	MYB Core	1	1	1	1	1	1
23	P- Box	0	0	1	0	0	0
24	Pyramidine Box	0	0	1	1	1	1
25	RY element	0	0	1	0	0	0
26	RY repeat legumin Box	0	0	0	1	1	1
27	SKN 1	0	0	1	0	0	0
28	SP1	0	0	1	1	1	1
29	TAAAG	0	1	0	0	0	0
30	TATA-Box	1	1	1	1	1	1
31	TC rich	1	0	0	0	0	0
32	TC rich repeat	0	0	1	0	0	0
33	TGA element	0	0	0	1	0	1
34	W box	0	0	1	1	1	1
35	WRKY	0	0	1	1	1	1
36	WUN motif	1	0	0	0	0	0

with respective primers designed from the source organism using template DNA of sorghum, rice and wheat. The PCR amplification resulted in the multiple band formation with 14 sets of primers (Fig.1) and one primer Dof-03 didn't gave any amplification. Multiple bands indicated presence of multiple Dof genes in cereals (Kushwaha *et al.*, 2014) or it could be also due to repeats in our primers which resulted in length polymorphism. Many expected size amplicons were observed among multiple bands in different cereals. Dof-02 and Dof-04 gave maximum expected size amplicons across cereals under investigation (Table 2). Sorghum was mainly amplified by sorghum based primers and none of the rice based primers worked in sorghum. Based on the amplification profile only 10 sets

Table 6: Different TFs binding site on predicted DNA motifs and gene ontology predictions

Motifs	TFs binding site (TOMTOM)	Gene Ontology predictions for Motifs (GOMo)
Motif-1	AtMYB84, abi4, SOC1, ABF1, SMZ, ERF1, AtMYB15 etc.	Chloroplast, nucleus, transcription,factor activity, protein binding, plasma membrane
Motif-2	PI, SOC1, SVP, FLC, AGL15, HMGI/Y, Dof2, PBF etc.	Transcription factor activity, Nucleus, plasma membrane, regulation of transcription (DNAdependent), kinase activity.
Motif-3	AtMYB84, RAV1, ERF1, bZIP911, abi4 etc.	Chloroplast

AtMYB84-arabidopsis thaliana MYB84,abi4- abscisic acid-insensitive 4, SOC1- suppressor of constans1 overexpression1, ABF1- ABRE-binding factor 1, SMZ- Protein SCHLAFMUTZE, ERF1- ethylene-responsive transcription factor 1, PI- PISTILLATA, SVP- Short Vegetative Phase, FLC- Flowering locus C, AGL 15 - Agamous-like 15, HMGI/Y- High Mobility Group, DOF2-DNA Binding with one finger 2, PBF-Prolamin box binding factor, RAV1- RELATED TO ABI3/VP1 1and bZIP911- Basic leucine zipper 911

of primers were further analysed with different millets along with rice, wheat and sorghum. A total of 9 eluted product representing putative PBF-Dof promoters (Fig.2) of rice, sorghum, barnyard millet, finger millets (brown, white and golden), proso millet, and kodo millet were sequenced commercially with respective primers used for amplification. After sequencing only 6 out of 9 sequences were found satisfactory for further *In* silico analysis.

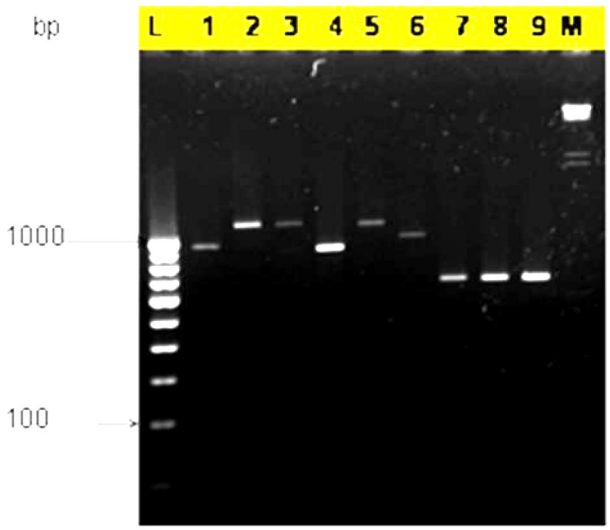

Fig. 2: 1.5% Agarose gel showing eluted amplicons. Lane L, 500 bp ladder; Lane1, sorghum; Lane 2, barnyard millet; Lane 3& 4, rice; Lane 5, proso millet; Lane 6, kodo millet; Lane 7, finger millet (PRM-1); Lane 8, finger millet (PRM-801); Lane 9, finger millet (PRM-701) and Lane M, Lambda DNA/*Eco*RI digest marker.

In silico analysis of TSS and *Cis*- acting elements

The cloned and sequenced amplicons of Rice, Sorghum, Barnyard, PRM1, PRM701 and PRM801 were 126,121,400,642,477 and 628 bases long respectively. RNA pol II binds to promoter element with the help of TATA and CAAT box and starts transcription at TSS. Thus, defining the TSS site in a promoter is a very essential step in promoter analysis. Cloned sequences were subjected to a neural network based tool and several TSS in both forward and reverse strands with high confidence scores of 0.73 to 1 (Table 3) were identified. All sequences showed TSS sites accompanied by TATA and CAAT boxes, corroborating our cloned sequences as putative *Dof* promoter.

In addition to RNA Pol II binding, transcriptional rate of a gene is also determined by binding of TFs to *cis*-elements in promoters, additional co-factors, and chromatin accessibility (Wasserman and Sandelin, 2004). To capture the structural and functional diversity of *cis*-elements(motifs), cloned sequences were further subjected to online bioinformatics tools namely PLACE and Plantcare. Sequence motifs are short, recurring patterns in DNA that are presumed to have a biological function (D'haeseleer, 2006). The promoter analysis identified ~36 diverse *cis*-acting elements associated with root, leaf, flower, seed, abiotic or biotic stress, and hormone (Table 4 & 5) occurring in the promoter regions. In cereal seed storage gene promoters, seed specific motifs like DPBF(Kim *et al.*, 1997), RY element, SKN1 motif, GCN4 motif, E-Box, etc. (Stalberg *et al.*, 1996; Bobb *et al.*, 1997; Reidt *et al.*, 2000; Fauteux and Strömvik, 2009) are generally present. In case of sorghum sequence (121 bases) different motifs like CAAT, ABRE, and TATA box were observed**.** In case of PRM-1(Brown),PRM-801(White) and PRM-701 (Golden) motifs like CAAT box, RY element, SKN1 motif,GCN4 motif, E-Box,

AAGAA, DOF core, G-Box, Pyramidine Box, W box, WRKY, etc. were observed. The G box or CACGTG motif, which function as an abscisic acid response element (ABRE) has been frequently reported in cereal promoters (Pla *et al.*, 1993; Choi *et al.*, 2000). The integrated map of TSS and seed storage protein specific *cis*-motifs (Fig.3) revealed high level of similarity among PRM1, PRM701 and PRM801 promoter sequences.

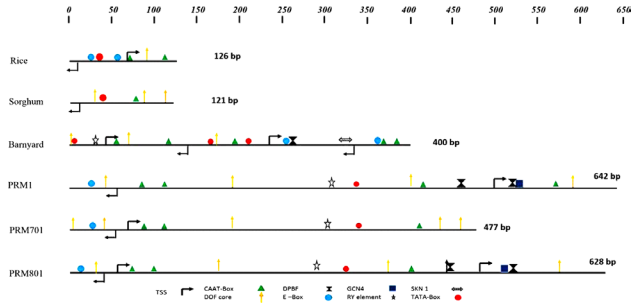

Fig. 3: Schematic diagram showing transcription start sites, TATA box, CAAT box and few important seed specific *cis*-elements in different cloned sequences

Core promoter element TATA box and common promoter and enhancer element CAAT box were uniformly dispersed around TSS across all promoters. DOFCOREZM (AAAG) which is the binding cite of Dof transcription factors (Yanagisawa and Schmidt, 1999) was most abundant *cis*-element present in all promoter region. Previously, various motifs in promoters of seed storage protein genes has already been found associated with a role in endosperm specific expression (Albani *et al.*, 1997; Wu *et al.*, 2000; Diaz *et al.*, 2005) (Yadav *et al.*, 2007; 2008).

Conserved DNA motif similarity search using MEME

MEME suit was utilized to do more intensive search of motifs in putative promoters, not captured by PLACE and PlantCARE. Three conserved DNA motifs of 15, 11 and 15 bases were identified (Fig. 4) from putative *Dof* promoters of PRM 1, PRM701 and PRM801. These motifs were absent in Rice, Sorghum and Barnyard. Motif 1 was more widely present in three promoters than Motif 2 & 3 (Fig 5). These motifs were further analysed in TOMTOM against a database of known motifs e.g., JASPAR CORE (2014) plants to produce an alignment for each significant match and then these significantly aligned motifs were searched against *Arabidopsis thaliana* database for gene ontology using GOMo. Predicted motifs has binding site for more than 20 well studied TFs which play role in Chloroplast, nucleus, transcription factor activity, protein binding, plasma membrane and kinase activity (Table 6) and are involved in different pathways (Duan *et al.*, 2005; Xue *et al.*, 2012). Motif-2 have binding sites for PBF and *Dof*-2 which has a direct correlation with PBF *Dof*. MYB, SOC, PI and FLC has also important role in plant reproductive stage development. In addition to common seed storage protein promoter elements, numerous other elements with diverse functions were observed in most of these clones. This provides a better option for selection of these promoters based on specific activity like endosperm specific expression, specificity towards abiotic stress etc.

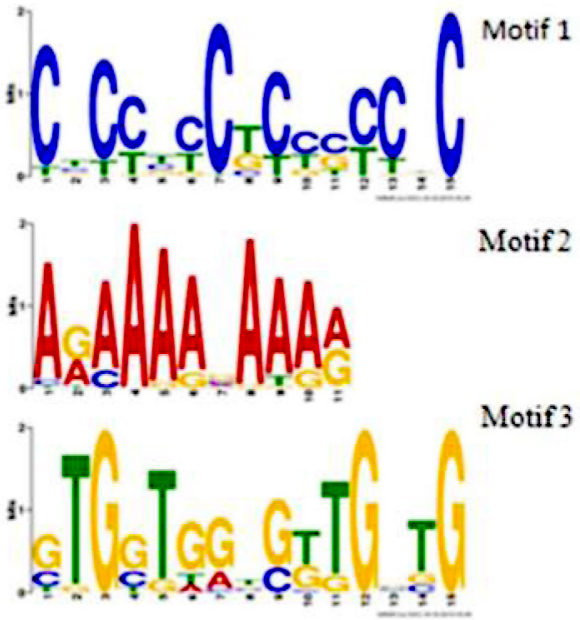

Fig. 4: Identified DNA motifs in sequenced products of PRM-1, PRM-701 and PRM-801

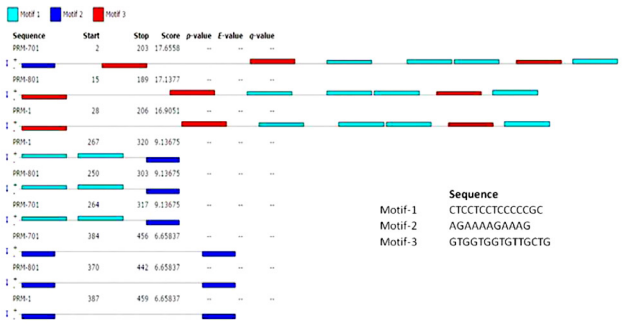

Fig. 5: Conserved DNA motifs in cloned sequences of PRM1, PRM701 and PRM801 analyzed by MCAST

The presence of TSS, TATA-box, CAAT-box, seed specific motifs like DPBF, RY element, SKN1 motif, GCN4 motif, E-Box, etc. and three conserved DNA motifs harbouring binding sites for well characterised TFs strongly suggests cloned sequences as candidate PBF-*Dof* gene promoters. These putative Dof promoters could further be assessed for qualitative and quantitative expression of targeted *Dof* genes by transgenic approaches.

Conclusion

The preliminary *in silico* investigation has provided the confirmation of these putative promoters to be of *Pbf-Dof* genes as numerous seed storage protein specific motifs like CAAT box, RY element, SKN1 motif, GCN4 motif, E-Box were uniformly present. The PCR amplification pattern of different varieties of finger millets (PRM-1 PRM-801, PRM-701) varying in protein content with different sets of primers gave more or less similar banding patter revealing the similarity at genomic level. PRM-1, PRM-701 and PRM-801 showed presence of three DNA motifs which are binding site for many well characterized transcription factors e.g. MYB, abi4, SOC, FLC, PBF, Dof 2 etc. Further validation by cloning in promoter probe vector is required for confirmation of temporal and spatial expression associated with seed storage protein genes.

Acknowledgments

The authors wish to acknowledge the Department of Biotechnology for providing financial support in the form of Programme Support for research and development in Agricultural Biotechnology at G.B. Pant University of Agriculture and Technology, Pantnagar (Grant No. BT/PR7849/AGR/02/374/2006). The support provided by Head, MBGE and Director, Experiment Station, GBPUAT, Pantnagar is also thankfully acknowledged.

References

Albani, D., Hammond-Kosack, M.C., Smith, C., Conlan, S., Colot, V., Holdsworth, M. and Bevan, M.W. 1997. The wheat transcriptional activator SPA: a seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. *The Plant cell* **9**: 171–184.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37**: W202–8.

Bobb, A.J., Chern, M.S. and Bustos, M.M. 1997. Conserved RY-repeats mediate transactivation of seed-specific promoters by the developmental regulator PvALF. *Nucleic Acids Research* **25**: 641–647.

Caetano-Anollés, G. 2005. Evolution of Genome Size in the Grasses. *Crop Science* **45**: 1809.

Chen, Y. and Cao, J. 2015. Comparative Analysis of Dof Transcription Factor Family in Maize. *Plant Molecular Biology Reporter* **33**: 1245–1258.

Choi, H.I., Hong, J.H., Ha, J.O., Kang, J.Y. and Kim, S.Y. 2000. ABFs, a family of ABA-responsive element binding factors. *Journal of Biological Chemistry* **275**: 1723–1730.

Colot, V., Bartels, D., Thompson, R. and Flavell, R. 1989. Molecular characterization of an active wheat LMW glutenin gene and its relation to other wheat and barley prolamin genes. *MGG Molecular & General Genetics* **216**: 81–90.

D'haeseleer, P. 2006. What are DNA sequence motifs? *Nature Biotechnology* **24**: 423–425.

Devos, K.M. 2000. Genome Relationships: The Grass Model in Current Research. *The Plant Cell Online* **12**: 637–646.

Devos, K.M., Wang, Z.M., Beales, J., Sasaki, Y. and Gale, M.D. 1998. Comparative genetic maps of foxtail millet (*Setaria italica*) and rice (*Oryza sativa*). *Theoretical and Applied Genetics* **96**: 63–68.

Diaz, I., Martinez, M., Isabel-LaMoneda, I., Rubio-Somoza, I. and Carbonero, P. 2005. The DOF protein, SAD, interacts with GAMYB in plant nuclei and activates transcription of endosperm-specific genes during barley seed development. *Plant Journal* **42**: 652–662.

Diaz, I., Vicente-Carbajosa, J., Abraham, Z., Martnez, M., Moneda II, La. and Carbonero, P. 2002. The GAMYP protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development. *Plant Journal* **29**: 453–464.

Dong, G., Ni, Z., Yao, Y., Nie, X. and Sun, Q. 2006. Wheat Dof transcription factor WPBF interacts with TaQM and activates transcription of an alpha-gliadin gene during wheat seed development. *Plant Molecular Biology* **63**: 73–84.

Du, H., Zhang, L., Liu, L., Tang, X.F., Yang, W.J., Wu, Y.M., Huang, Y.B. and Tang, Y.X. 2009. Biochemical and molecular characterization of plant MYB transcription factor family. *Biochemistry. Biokhimiia* **74**: 1–11.

Duan, K., Luo, Y.H., Luo, D., Xu, Z.H. and Xue, H.W. 2005. New insights into the complex and coordinated transcriptional regulation networks underlying rice seed development through cDNA chip-based analysis. *Plant Molecular Biology* **57**: 785–804.

Fauteux, F. and Strömvik, M.V. 2009. Seed storage protein gene promoters contain conserved DNA motifs in Brassicaceae, Fabaceae and Poaceae. *BMC Plant Biology* **9**: 126.

Feuillet, C. and Keller, B. 2002. Comparative genomics in the grass family: Molecular characterization of grass genome structure and evolution. *Annals of Botany* **89**: 3–10.

Forde, B.G., Heyworth, A., Pywell, J. and Kreis, M. 1985. Nucleotide sequence of a B1 hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. *Nucleic Acids Research* **13**: 7327–39.

Goodstein, D.M., Shu, S. and Howson, R. *et al*. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**: D1178–86.

Gupta, S., Malviya, N., Kushwaha, H., Nasim, J., Bisht, N.C., Singh, V.K. and Yadav, D. 2015. Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. *Planta* **241**: 549–562.

Hall, T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98.

Hammond-Kosack, M.C., Holdsworth, M.J. and Bevan, M.W. 1993. In vivo footprinting of a low molecular weight glutenin gene (LMWG-1D1) in wheat endosperm. *The EMBO Journal* **12**: 545–54.

Higo, K., Ugawa, Y., Iwamoto, M. and Higo, H. 1998. PLACE: a database of plant *cis*-acting regulatory DNA elements. *Nucleic Acids Research* **26**: 358–359.

Le Hir, R. and Bellini, C. 2013. The plant-specific dof transcription factors family: new players involved in vascular system development and functioning in Arabidopsis. *Frontiers in Plant Science* **4**: 164.

Kellogg, E.A. 1998. Relationships of cereal crops and other grasses. *Proceedings of the National Academy of Sciences* **95**: 2005–2010.

Kellogg, E.A. and Birchler, J.A. 1993. Linking Phylogeny and Genetics: *Zea Mays* as a Tool for Phylogenetic Studies. *Systematic Biology* **42**: 415–439.

Kim, S.Y., Chung, H.J. and Thomas, T.L. 1997. Isolation of a novel class of bZIP transcription factors that interact with ABA-responsive and embryo-specification elements in the Dc3 promoter using a modified yeast one-hybrid system. *The Plant Journal : for Cell and Molecular Biology* **11**: 1237–51.

Kreis, M., Forde, B.G., Rahman, S., Miflin, B.J. and Shewry, P.R. 1985. Molecular evolution of the seed storage proteins of barley, rye and wheat. *Journal of Molecular Biology* **183**: 499–502.

Krohn, N.M. 2002. Specificity of the Stimulatory Interaction between Chromosomal HMGB Proteins and the Transcription Factor Dof2 and Its Negative Regulation by Protein Kinase CK2-mediated Phosphorylation. *Journal of Biological Chemistry* **277**: 32438–32444.

Kushwaha, H., Gupta, S., Singh, V.K., Rastogi, S. and Yadav, D. 2011. Genome wide identification of Dof transcription factor gene family in sorghum and its comparative phylogenetic analysis with rice and Arabidopsis. *Molecular Biology Reports* **38**: 5037–5053.

Kushwaha, H., Jillo, K.W., Singh, V.K., Kumar, A. and Yadav, D. 2014. Assessment of genetic diversity among cereals and millets based on PCR amplification using Dof (DNA binding with One Finger) transcription factor gene-specific primers. *Plant Systematics and Evolution* 833–840.

Lescot, M. 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.

Levine, M. and Tjian, R. 2003. Transcription regulation and animal diversity. *Nature* **424**: 147–51.

Li, Z., Pinson, S.R.M., Park, W.D., Paterson, A.H. and Stansel, J.W. 2008. Mapping of quantitative trait loci and quantitative-trait-modifying factors affecting three grain yield components in rice (*Oryza sativa* L.). 624–630.

Libault, M., Joshi, T., Benedito, V.A., Xu, D., Udvardi, M.K. and Stacey, G. 2009. Legume Transcription Factor Genes: What Makes Legumes So Special? *Plant Physiology* **151**: 991–1001.

Lijavetzky, D., Carbonero, P. and Vicente-Carbajosa, J. 2003. Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC Evolutionary Biology* **3**: 17.

Liu, L., White, M.J. and MacRae, T.H. 1999. Transcription factors and their genes in higher plants: Functional domains, evolution and regulation. *European Journal of Biochemistry* **262**: 247–257.

Marzabal, P., Gas, E., Fontanet, P., Vicente-Carbajosa, J., Torrent, M. and Ludevid, M.D. 2008. The maize Dof protein PBF activates transcription of gamma-zein during maize seed development. *Plant Molecular Biology* **67**: 441–454.

Mena, M., Vicente-Carbajosa, J., Schmidt, R.J. and Carbonero, P. 1998. An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin-box of a native B-hordein promoter in barley endosperm. *The Plant Journal* **16**: 53–62.

Moreno-Risueno, M.Á., Martínez, M., Vicente-Carbajosa, J. and Carbonero, P. 2007. The family of DOF transcription factors: from green unicellular algae to vascular plants. *Molecular Genetics and Genomics* **277**: 379–390.

Murray, M.G. and Thompson, W.F. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* **8**: 4321–4326.

Noguero, M., Atif, R.M., Ochatt, S. and Thompson, R.D. 2013. The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. *Plant Science* **209**: 32–45.

Paterson, A.H., Bowers, J.E. and Bruggmann, R. 2009. The Sorghum bicolor genome and the diversification of grasses. *Nature* **457**: 551–556.

Pla, M., Vilardell, J., Guiltinan, M.J., Marcotte, W.R., Niogret, M.F., Quatrano, R.S. and Pagès, M. 1993. The *cis*-regulatory element CCACGTGG is involved in ABA and water-stress responses of the maize gene rab28. *Plant Molecular Biology* **21**: 259–266.

Ravel, C., Nagy, I.J., Martre, P., Sourdille, P., Dardevet, M., Balfourier, F., Pont, C., Giancola, S., Praud, S. and

Charmet, G. 2006. Single nucleotide polymorphism, genetic mapping, and expression of genes coding for the DOF wheat prolamin-box binding factor. *Functional & Integrative Genomics* **6**: 310–21.

Reese, M.G. 2001. Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. *Computers and Chemistry* **26**: 51–6.

Reidt, W., Wohlfarth, T., Ellerström, M., Czihal, A., Tewes, A., Ezcurra, I., Rask, L. and Bäumlein, H. 2000. Gene regulation during late embryogenesis: The RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. *Plant Journal* **21**: 401–408.

Riechmann, J.L. and Ratcliffe, O.J. 2000. A genomic perspective on plant transcription factors. *Current Opinion in Plant Biology* **3**: 423–434.

Rubio-Somoza, I., Martinez, M., Abraham, Z., Diaz, I. and Carbonero, P. 2006. Ternary complex formation between HvMYBS3 and other factors involved in transcriptional control in barley seeds. *The Plant Journal : for Cell and Molecular Biology* **47**: 269–81.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual. Cold Spring Harbor laboratory press*.

Shaw, L.M., McIntyre, C.L., Gresshoff, P.M. and Xue, G.P. 2009. Members of the Dof transcription factor family in Triticum aestivum are associated with light-mediated gene regulation. *Functional & Integrative Genomics* **9**: 485–498.

Shiu, S.H., Shih, M.C. and Li, W.H. 2005. Transcription factor families have much higher expansion rates in plants than in animals. *Plant physiology* **139**: 18–26.

Singh, M., Yadav, A., Ma, X. and Amoah, E. 2010. Plasmid DNA Transformation in Escherichia Coli : Effect of Heat Shock Temperature , Duration , and Cold Incubation of CaCl 2 Treated Cells. *Shock* **6**: 561–568.

Srinivasachary, Dida, M.M., Gale, M.D. and Devos, K.M. 2007. Comparative analyses reveal high levels of conserved colinearity between the finger millet and rice genomes. *Theoretical and Applied Genetics* **115**: 489–499.

Stalberg, K., Ellerstom, M., Ezcurra, I., Ablov, S. and Rask, L. 1996. Disruption of an overlapping E-box/ABRE motif abolished high transcription of the napA storage-protein promoter in transgenic Brassica napus seeds. *Planta* **199**.

Suzuki, A., Wu, C.Y., Washida, H. and Takaiwa, F. 1998. Rice MYB protein OSMYB5 specifically binds to the AACA motif conserved among promoters of genes for storage protein glutelin. *Plant & Cell Physiology* **39**: 555–9.

Takaiwa, F., Yamanouchi, U., Yoshihara, T., Washida, H., Tanabe, F., Kato, A. and Yamada, K. 1996. Characterization of common *cis*-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. *Plant Molecular Biology* **30**: 1207–21.

Takatsuji, H. 1998. Zinc-finger transcription factors in plants. *Cellular and Molecular Life Sciences* **54**: 582–596.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673–80.

Umemura, Y., Ishiduka, T., Yamamoto, R. and Esaka, M. 2004. The Dof domain, a zinc finger DNA-binding domain conserved only in higher plants, truly functions as a Cys2/Cys2 Zn finger domain. *The Plant Journal* **37**: 741–749.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Research* **40**: e115.

Varshney, R.K., Hoisington, D.A. and Tyagi, A.K. 2006. Advances in cereal genomics and applications in crop breeding. *Trends in Biotechnology* **24**: 490–499.

Verdier, J. and Thompson, R.D. 2008. Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant & Cell Physiology* **49**: 1263–71.

Vicente-Carbajosa, J., Moose, S.P., Parsons, R.L., Schmidt, Robert, J. 1997. A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 7685–7690.

Washida, H., Wu, C.Y., Suzuki, A., Yamanouchi, U., Akihama, T., Harada, K. and Takaiwa, F. 1999. Identification of *cis*-regulatory elements required for endosperm expression of the rice storage protein glutelin gene GluB-1. *Plant Molecular Biology* **40**: 1–12.

Wasserman, W.W. and Sandelin, A. 2004. Applied bioinformatics for the identification of regulatory elements. Nature reviews. *Genetics* **5**: 276–287.

Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V. and Romano, L.A. 2003. The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution* **20**: 1377–1419.

Wu, Y. and Messing, J. 2012. Rapid divergence of prolamin gene promoters of maize after gene amplification and dispersal. *Genetics* **192**: 507–519.

Wu, C.Y., Washida, H., Onodera, Y., Harada, K. and Takaiwa, F. 2000. Quantitative nature of the prolamin-box, ACGT and AACA motifs in a rice glutelin gene promoter: Minimal *cis*-element requirements for endosperm-specific gene expression. *Plant Journal* **23**: 415–421.

Xi, D.M. and Zheng, C.C. 2011. Transcriptional regulation of seed storage protein genes in Arabidopsis and cereals. *Seed Science Research* **21**: 247–254.

Xu, J.H. and Messing, J. 2009. Amplification of prolamin storage protein genes in different subfamilies of the Poaceae. *Theoretical and Applied Genetics* **119**: 1397–1412.

Xue, L.J., Zhang, J.J. and Xue, H.W. 2012. Genome-Wide Analysis of the Complex Transcriptional Networks of Rice Developing Seeds. *PLoS One* **7**, e31081.

Yamamoto, M.P., Onodera, Y., Touno, S.M. and Takaiwa, F. 2006. Synergism between RPBF Dof and RISBZ1 bZIP

Activators in the Regulation of Rice Seed Expression Genes. *Plant Physiology* **141**: 1694–1707.

Yadav, D., Singh, V.K. and Singh, N.K. 2007. *In silico* Cis-regulatory Elements Analysis of Seed Storage Protein Promoters Cloned from Different Cultivars of Wheat, Rice and Oat, *Online Journal of Bioinformatics* **8(1)**: 75-83.

Yadav Dinesh, Vinay Kumar Singh, Sarita, Naveen, C. Bisht and Nagendra K Singh. 2008. PCR based cloning of seed storage protein promoters of wheat, oat and rice and their *in silico* analysis for *cis*- regulatory elements. *Bioinformatics Trends* **3(2&3):** 67-74.

Yanagisawa, S. 1997. Dof DNA-Binding Domains of Plant Transcription Factors Contribute to Multiple Protein-Protein Interactions. *European Journal of Biochemistry* **250**: 403–410.

Yanagisawa, S. 2002. The Dof family of plant transcription factors. *Trends in Plant Science* **7**: 555–560.

Yanagisawa, S. 2004. Dof domain proteins: Plant-specific transcription factors associated with diverse phenomena unique to plants. *Plant and Cell Physiology* **45**: 386–391.

Yanagisawa, S. and Schmidt, R.J. 1999. Diversity and similarity among recognition sequences of Dof transcription factors. *Plant Journal* **17**: 209–214.