OJB™

Online Journal of Bioinformatics©

Volume 9 (2): 130-143, 2008

In silico analysis of PCR amplified DOF (DNA binding with one finger) transcription factor domain and cloned genes from cereals and millets

Kushwaha H, Gupta N, Singh VK, Kumar A, Yadav D

Department of Molecular Biology and Genetic Engineering, College of Basic Sciences and Humanities, G.B. Pant University of Agriculture and Technology, Pantnagar (UttaraKhand)

ABSTRACT

Kushwaha H, Gupta N, Singh VK, Kumar A, Yadav D, In silico analysis of PCR amplified DOF (DNA binding with one finger) transcription factor domain and cloned genes from cereals and millets, Online Journal of Bioinformatics, 9 (2): 130-143, 2008. The PCR amplified DOF domain of different cereals and millets were gel eluted, sequenced and subjected to homology search, multiple sequence alignment, phylogenetic tree construction and motif analysis. The presence of four cysteine residues associated with a typical zinc like finger and a conserved motif with multilevel consensus sequence QPRHFCKSCQRYWTA was uniformly observed across different cereals and millets. The putative DOF genes of *Eleusine coracana* along with *Oryza sativa, Triticum aestivum, Hordeum vulgare* and *Zea mays* were PCR amplified, gel eluted, cloned and assigned accession numbers EU760631, EU586268, EU586269, EU586267 and EU586266 respectively after sequencing. The homology search confirmed their identity to PBF (Prolamin-box Binding Factor) DOF genes except for *O. sativa*. Multiple sequence alignment of these PBF DOF genes revealed the presence of conserved DOF domain with four cysteine residues in *E. coracana* while partial sequences were observed in other cloned genes. The presence of conserved motifs involved with regulation of seed storage protein genes further confirmed their identity to PBF DOF genes. Phylogenetic tree constructed based on protein sequences of cloned DOF genes resulted in distinct clusters for PBF DOF genes of monocots and dicots.

Keywords: DOF, Transcription factor, Cereals, Millets, PBF, Multiple sequence alignment, Motif, Domain.

INTRODUCTION

Plant gene expression involves classes of transcription factors that have specifically evolved to regulate plant-specific genes and/or to mediate a variety of plant-specific signals. The DOF (DNA binding with One Finger) family is a typical example of Zinc-finger transcription factors extensively reviewed (Takatsuji, 1998; Yanagisawa, 2002; 2004). The DOF family is one of the well characterized plant specific transcription factor having diverse roles. The proteins of DOF family comprises of 200–400 amino acids having a highly conserved domain (DOF domain) of 50-52 amino acids including a C2C2-type zinc-finger motif at N-terminal end. The typical structure of DOF domain as elucidated from sequenced clone of *Eleusine coracana* is shown in **Figure-1**. The DOF domain is known to be a bi-functional domain that mediates not only DNA-binding but also protein-protein interactions, including interaction with another class of transcription factor (basic domain-leucine zipper protein, bZIP protein) and non-histone nuclear HMG protein (Yanagisawa, 1995, 1996, 1997; Zhang, 1995 and Diaz et al., 2002). The variations at C-terminal end of the DOF protein might be associated with diverse functions. Multiple DOF genes have been reported in many plants conferring to its diverse roles exclusively for plants like the expression of genes associated with carbon assimilation**,** phytochrome signalling, seed maturation and germination, the auxin response, the salicylic acid response, involved with the function of the stomata guard cells, photoperiodic flowering and biosynthesis of glucosinolates (Yanagisawa, 2002).

The DOF family of transcription factor generally recognize the *Cis*-motifs 5'-T/AAAAG-3' or 5'-CTTTT/A-3' in the promoters of genes regulated by them.

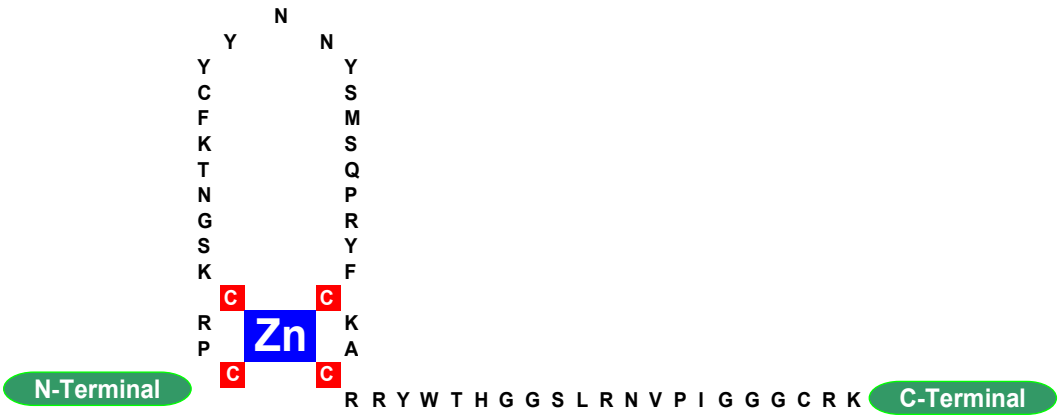

Figure-1: Structure of the DOF domain of *E. coracana* based on its amino acid sequence

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 seed storage protein genes is known as PBF (Prolamin-box Binding Factor) DOF transcription (**Yamamoto et al., 2006**). These genes have been well characterized in maize (Vicente-Carbajosa et al., 1997**),** barley (Mena et al., 1998 and Diaz et al.,2002; 2005**),** rice (Yamamoto et al., 2006**)** and wheat (Ravel et al., 2006 and Dong et al., 2007).

An attempt has been made to reveal the genome-wide comparative analysis of the rice and arabidopsis DOF gene families (Lijavetzky et al., 2003) based on the availability of complete genome sequence of rice (**http://rgp.dna.affrc.go.jp/IRGSP/**) and arabidopsis (**http://www.arabidopsis.org**) representing monocots and dicots respectively. In cereals DOF genes have been studied in maize, barley, wheat and rice. There exists great diversity of DOF genes in cereals with 30 and 24 different DOF genes being reported in rice and barley respectively. (Lijavetzky et al., 2003 and Moreno-Risueno et al., 2007).

The origin and evolution of the DOF transcription factor family based on phylogenetic analysis of DOF sequences across the representative organisms belonging to green unicellular algae to vascular plants has been attempted.

A total of 116 DOF genes representing green unicellular alga (*Chlamydomonas reinhardtii)*, moss *Physcomitrella patens*, fern (*Selaginella moellendorffii*), gymnosperm (*Pinus taeda)*, dicotyledoneous angiosperm (*Arabidopsis thaliana)*, monocot (*Oryza sativa and Hordeum vulgare)* were analyzed by various bioinformatics tools. The phylogenetic tree constructed revealed the existence of six major clusters of orthologous and paralogous genes that probably originated by gene duplication events from a paraphyletic basal grade (Moreno-Risueno et al., 2007**)**.

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**)**. These cereals comprises of more than 10,000 species belonging to a family poaceae with several subfamilies like oryzoideae, pooideae, panicoideae and chloridoideae (Kellogg and Birchler, 1993**).** 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, 1998; Devos and Gale, 2000; Feuillet and Keller, 2002; Paterson et al., 2004, 2005 and Caetano-Anolles, 2005**).**

This paper reports *in silico* analysis of PCR amplified DOF domain of five cereals namely rice (*Oryza sativa*), wheat (*Triticum aestivum*), sorghum (*Sorghum bicolor*), barley (*Hordeum vulgare*) and oat (*Avena sativa*), and six millets *viz.* finger millet (*Eleusine coracana*), barnyard millet (*Echinochloa frumentacea*), proso millet (*Panicum milliaceum* Linn.), little millet (*Panicum antidotale*), kodo millet (*Paspalum scrobiculatum* Linn.) and foxtail millet (*Setaria italica* Beauv.). Further cloned putative DOF genes of finger millet, rice, wheat, barley and maize (*Zea mays*) were subjected to homology search, multiple sequence alignment, phyogenetic tree construction and motif analysis.

MATERIALS AND METHODS

The nucleotide sequences of different DOF genes were downloaded from NCBI and DRTF and subjected to multiple sequence alignment using Mega align module of DNASTAR (Burland et al., 2000). Based on consensus sequences different sets of primers specific for DOF domain and genes as listed in **Table-1** were designed using Primer-3 (Untergasser et al., 2007) and DNASTAR.

Table-1: List of primers used for PCR amplification of DOF domains and genes from different cereals and millets

Sl. No.	Source (Accession Number)	Sequence (5'-3')	Expected amplicon size (bp)	DOF Domain/ Gene
DOF-001	Rice (OsIBCD001074)	F=GAAGCCAGATAAGATTCTCCCTTG R=ATCAGTAAGTGACGGTAGTGCAAC	210	Domain
DOF-008	Rice (OsIBCD001074 & OsIBCD024636)	F= AGATTCTGCCTTGTCCTCGTTG R=GAGCTCTTACTCTTGCGCCTAC	172	Domain
DOF-011	Rice (OsIBCD001074)	F=GCCCTCGGTGCAACAGTATGGACA R=CTGCTTGGCCTTCGCTCTCACCTT	1065	Gene
DOF-024	Wheat (AJ012284)	F=GCCCTCGGTGCAAGTCTGG R=ATCACCGCCGCTGTCGTTATT	785	Gene
DOF-025	Maize (U82230)	F=GTGGCAGTGGGGAGCGTAAG R=TGAGGCCCCACCAGTTGTCT	715	Gene
DOF-031	Barley (AJ000991)	F=GCCCTCGGTGCAAGTCTGGTA R=CGCCGCCGTTATTGTTGTTCT	783	Gene

The genomic DNA of different cereals and millets were isolated by standard method (Murray and Thompson, 1980), quantified and analyzed on agarose gel electrophoresis (Maniatis et al., 1989). PCR amplification was performed as per the standard protocol using 50-100 ng of template DNA with variable annealing temperature based on Tm value of the primers. The amplified products were analyzed on agarose gel and the expected size

amplicons were gel eluted using QIAquick Gel Extraction Kit (Qiagen, USA) and cloned in pGEM-Teasy vector (Promega, USA) as per the kit instructions. Eluted PCR amplicons of DOF domains of different cereals and millets were directly sequenced using DOF domain specific primer. Putative cloned DOF genes were sequenced using M13 universal primer present in pGEM-T easy vector. The list of assigned accession number of PCR amplified DOF domain and putative DOF genes of different cereals and millets are provided in **Table-2a and b.**

Table-2a: List of assigned accession number of sequenced DOF genes and domains from cereals and millets

Name of Crop(s) (Scientific name)	Cultivar	Sub-family	Sequence length (bp)	DOF Gene / Domain	Accession Number	Protein ID
Rice (*Oryza sativa*)	Pusa sughandha	Oryzoideae	939	Gene	EU586268	ACC59771
			112	Domain	EU586263	ACC59766
			106	Domain	EU 586265	ACC59768
Wheat (*Triticum aestivum*)	Punjab PBW 373	Pooideae	724	Gene	EU586269	ACC59772
			94	Domain	EU760640	ACF06726
Sorghum (*Sorghum bicolor*)	Pant chari 5	Panicoideae	116	Domain	EU760632	ACF06718
Barley (*Hordeum vulgare*)	VLB56	Pooideae	712	Gene	EU586267	ACC59770
			121	Domain	EU760633	ACF06719
Oat (*Avena sativa*)	UP0212	Pooideae	334	Domain	EU760634	ACF06720
Maize (*Zea mays*)	DQPMC4W	Panicoideae	644	Gene	EU586266	ACC59769
Finger millet (*Eleusine coracana*)	VL315	Chloridoideae	589	Gene	EU 760631	ACF06717
			409	Domain	EU 586262	ACC59765
			137	Domain	EU 586264	ACC59767
Name of Crop(s) (Scientific name)	**Cultivar**	**Sub-family**	**Sequence length (bp)**	**DOF Gene / Domain**	**Accession Number**	**Protein ID**
Barnyard millet (*Echinochloa frumentacea*)	PRB401	Panicoideae	116	Domain	EU 760639	ACF06725
Proso millet (*Panicum milliaceum* Linn.)	405	Panicoideae	334	Domain	EU760635	ACF06721
Little millet (*Panicum antidotale)*	Local	Panicoideae	110	Domain	EU760636	ACF06722
Kodo millet (*Paspalum scrobiculatum* Linn.)	Local	Panicoideae	108	Domain	EU760637	ACF06723
Foxtail millet (*Setaria italica* Beauv.)	PRK1	Panicoideae	70	Domain	EU760638	ACF06724

Table-2b (above): List of assigned accession number of sequenced DOF genes and domains from cereals and millets

The sequenced DOF domain and genes were subjected to homology search with both general and transcription factor specific databases as mentioned in **Table-3** using BLASTN, TBLASTX and discontinuous MEGABLAST (Altschul et al., 1990, 1997**)**.

Table-3: List of Databases used for homology/similarity search

Database	Description	Website	Reference
NCBI	National Center for Biotechnology Information	http://www.ncbi.nlm.nih.gov	(Pruitt et al., 2007)
DRTF	Database of Rice transcription factor	http://drtf.cbi.pku.edu.cn/	(Gao et al., 2006)
DDBJ	DNA Data Bank of Japan	http://www.ddbj.nig.ac.jp/Welcomee.html	(Tateno et al., 2002)
PGD	Plant Genome Data base	http://www.plantgdb.org/	(Dong et al., 2005)
RICE-GAAS	Rice Genome Automated Annotation System	http://ricegaas.dna.affrc.go.jp	(Sakata et al., 2002)
GRAINGENES	genome database for small-grain crops	http://www.graingenes.org	(Matthews et al., 2003)
GRAMENE	cereal genome database	http://www.gramene.org/	(Jaiswal et al., 2006)
DBD	DNA-binding transcription Factors database	http://dbd.mrc-lmb.cam.ac.uk/DBD/index.cgi?Home	(Kummerfeld and Teichmann et al., 2006),
RARTF	RIKEN Arabidopsis transcription factor	http://rarge.gsc.riken.jp/rartf/	(Iida et al., 2005)
DATF	Database of *Arabidopsis* transcription factors	http://datf.cbi.pku.edu.cn/	(Guo et al., 2005)
PlnTFDB	plant transcription factor database	http://plntfdb.bio.uni-potsdam.de.	(Riano-Pachon et al., 2007)

The sequences were further subjected to bioinformatics softwares namely GENESCAN (Stormo, 2000**)** and FGENESH (Solovyev et al., 2006**)** for fishing out the probable genes and the putative CDS and were translated to protein sequence using translation tool (**http://ca.expasy.org/tools/dna.htm**l). The translated DOF protein sequences were subjected to protein functional analysis using PFAM version 23.0 (Finn et al., 2006**)**, PROSITE version 20.37 (Castro et al., 2006**)** and INTERPROSCAN version 4.4 (Quevillon et al., 2005**)**. These sequences of DOF proteins from different cereals and millets were aligned using ClustalW (Thompson et al., 1994**)** and phylogenetic tree was constructed using UPGMA method**.** A tree was inferred by Bootstrap phylogenetic inference using MEGA3.1 (Tamura et al., 2007). The conserved motifs present in these sequences were analyzed using BLOCKS and MEME (Multiple EM for Motif Elicitation) software version 3.5.7 (Bailey et al., 1998, 2006). For motif analysis of DOF domain, the selection of maximum number of motifs was set to 5 with minimum width of 15 amino acids while for genes maximum number of motifs was set to 10 with minimum width of 15 amino acid residues while other factors were of default selections. These selections were made in order to minimize the 'E-value' of the given parameter based on the probability of finding an equally well conserved pattern in set of sequences. Motifs involved with regulatory region of putative DOF genes and domain were analysed using PLACE (Higo et al., 1999**)**.

RESULTS AND DISCUSSION

*In silico analysis of DOF domain from different cereals and millets:T*he DOF domain from different cereals and millets were PCR amplified using domain specific primer (i.e. DOF-008) designed from rice DOF gene, gel eluted, sequenced and assigned accession number as listed in **Table-2a and b.** A variable length of DOF domain sequence varying from 70 to 409 bp was obtained for different cereals and millets though the PCR amplicon of expected size i.e. 172 bp was uniformly observed with all the templates except maize as shown in **Figure-2**. This variation in sequence length might be due to variable length of sequencing by domain specific primer. These domain sequences were subjected to BLAST search for deducing similarity with sequences available in different databases as mentioned in **Table-3**. The nucleotide sequences were translated to respective protein sequences using translation tool and subjected to protein Blast to reveal the similarity at protein level with other existing DOF domain proteins. Further the deduced protein sequences so translated were confirmed by subjecting the nucleotide sequences to gene finding software namely GENESCAN and FGENESH. The sequences showing similarity were subjected to multiple sequence alignment using ClustalW as shown in **Figure-3**. The presence of four cysteine residues associated with typical zinc like finger of DOF family of proteins was observed in domain of *S. bicolor*, *P. scrobiculatum*, *E. frumentacea, A. sativa, P. milliaceum* and *O. sativa.*

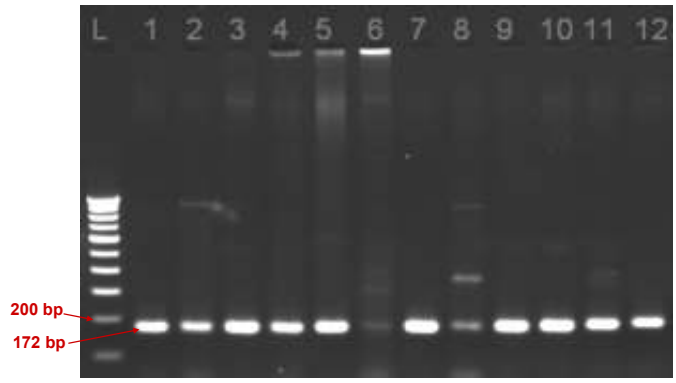

Figure-2: PCR amplification of DOF domain of different cereals and millets using DOF-008 primer. Lane L: 100 bp ladder (Bangalore Genei, India); Lane 1: Rice, Lane 2: Wheat, Lane 3: Sorghum, Lane 4: Barley, Lane 5: Oat, Lane 6: Maize, Lane 7: Finger millet, Lane 8: Barnyard millet, Lane 9: Proso millet, Lane 10: Little millet, Lane 11: Kodo millet, Lane 12: Foxtail millet.

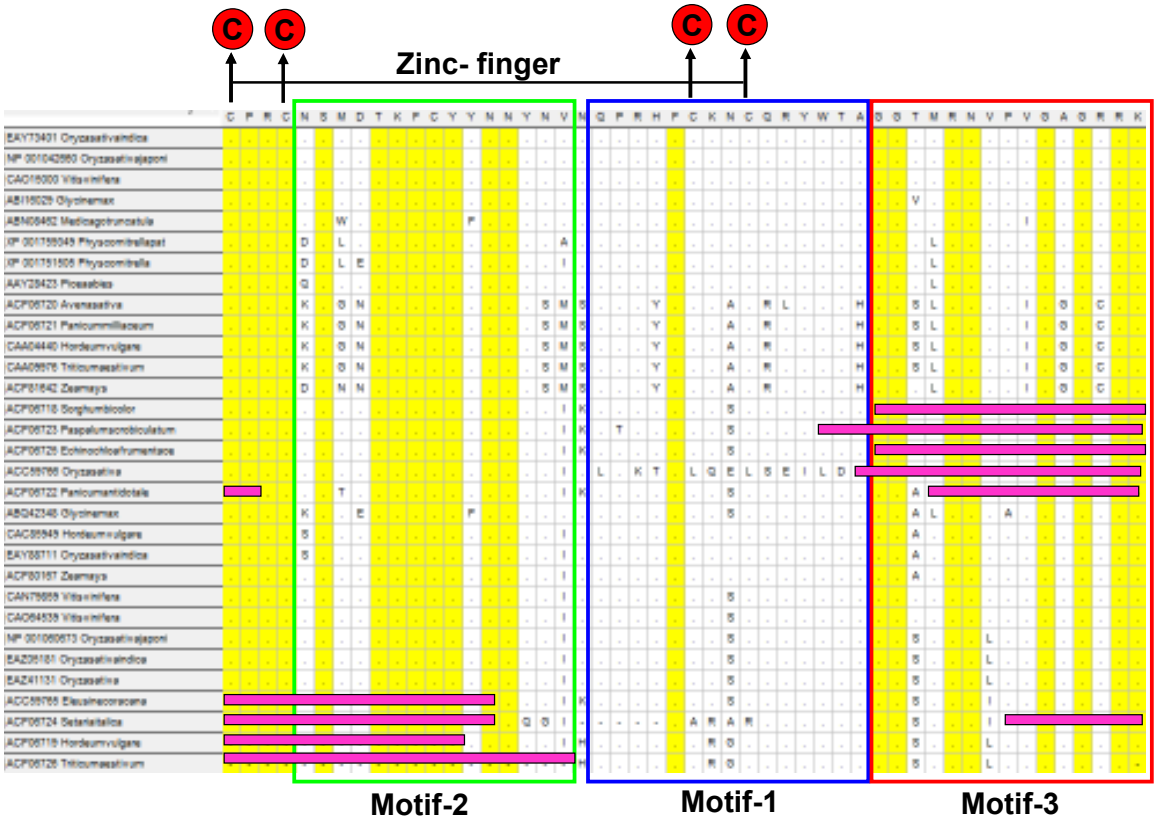

Figure-3: Variations in amino acid residues of the PCR amplified DOF domains of cereals and millets subjected to Multiple Sequence Alignment with other DOF domains. The three conserved motifs 1, 2 and 3 are highlighted. The yellow dotted portion indicates the conserved amino acid residues among different DOF domains while pink marked portion indicates the partial DOF domain sequences.

In case of *E. coracana, H. vulgare, T. aestivum, S. italica* and *P. antidotale* partial DOF domain with missing cysteine residue was observed. This sequence level variation of DOF domain is neither of any evolutionary

significance nor species specific and might be generated due to partial sequencing. Protein functional analysis of these domains using PFAM, PROSITE, and INTERPROSCAN confirmed there identity to DOF like proteins.

The sequenced DOF domains along with 20 other published sequences from NCBI subjected to phylogenetic tree construction using UPGMA method revealed two distinct clusters. These sequences were further subjected to MEME program for motif analysis.

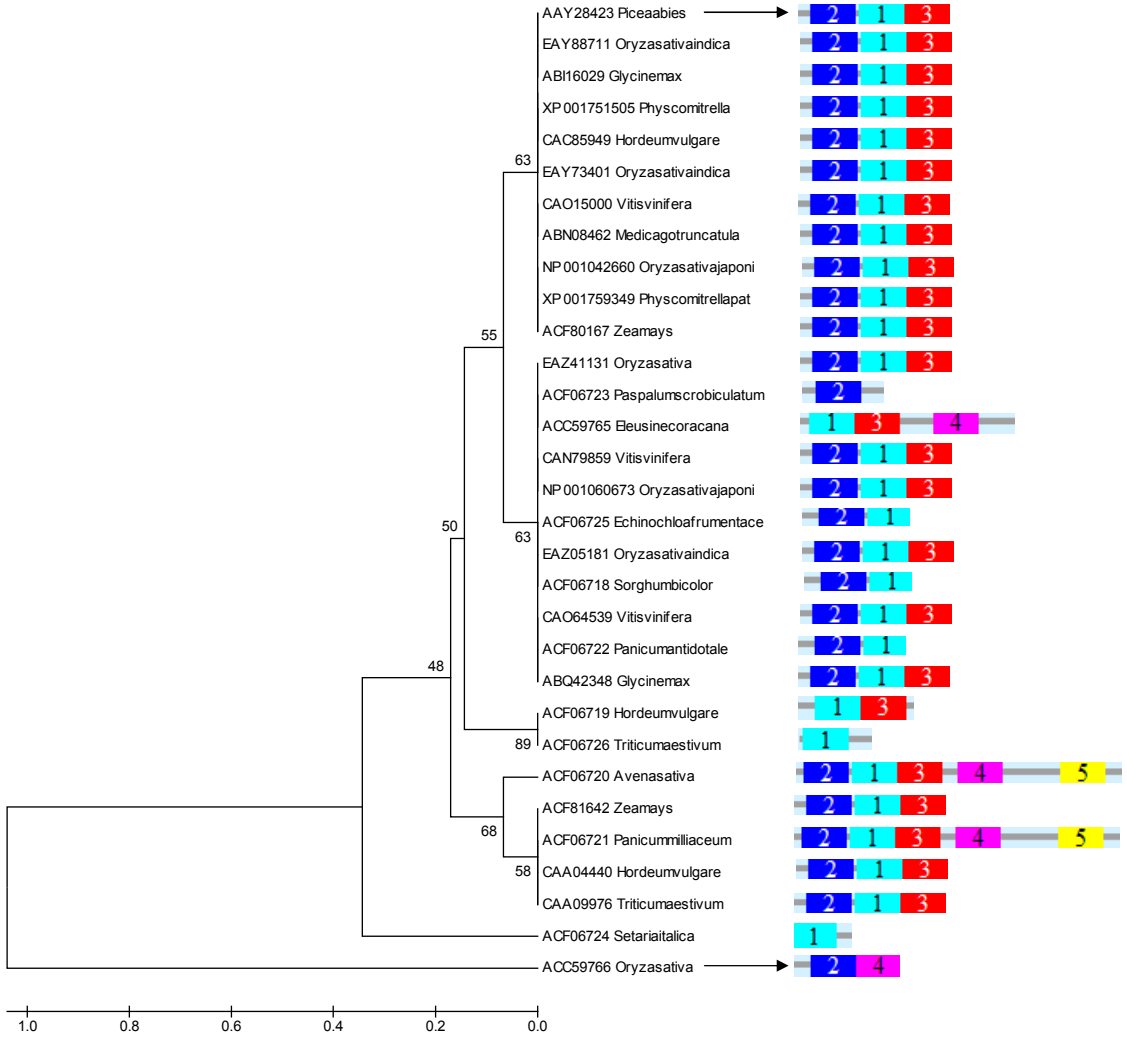

Figure-4: Phylogenetic tree constructed based on 31 DOF domain sequences and schematic distribution of respective conserved motifs identified by means of MEME software. Multilevel consensus sequences for the MEME defined motifs are listed in Table-4**.**

A total of five conserved motifs were observed in different DOF domain sequences. The distribution of conserved motifs in different accessions is provided in **Figure-4.** The overall multilevel consensus sequences associated with each of the five motifs is provided in **Table-4**. The multilevel consensus sequence corresponding to the motif is an aid in remembering and understanding the motif. It is calculated from the motif position-specific probability matrix. MEME motifs are represented by position-specific probability matrices that specify the probability of each possible letter appearing at each possible position in an occurrence of the motif.

Table-4: Multilevel consensus sequences for the MEME defined motifs observed among different DOF domains

Motif	No. of amino acids	E-value	Multilevel consensus sequence
1	15	2.7e-340	QPRHFCKSCQRYWTA
2	15	1.2e-285	DSMDTKFCYYNNYNI
3	15	1.0e-239	GGTMRNVPVGAGRRK
4	15	8.3e-008	TLDAHELEEASEIEP
5	15	1.1e-005	TFMSGGFDIQSSLSL

The motif-1 with multilevel consensus sequence QPRHFCKSCQRYWTA is uniformly observed among all DOF domain sequences except in *O. sativa* and *P. scrobiculatum*, which might be due to its partial sequence. Motif-1 and 2 is involved in regulation of phosphoenolpyruvate carboxylase gene. Motif-3 is involved in regulation of flavonoid biosynthesis. Motif 4 and 5 lacks any specified functions.

Cereals and millets have been subjected to comparative genomics studies based on the availability of the whole genome sequence of rice. Comparative analyses revealed a high level of colinearity between *E. coracana* and *O. sativa* genomes belonging to two different sub families namely chloridoideae and Oryzoideae (Srinivasachary et al., 2007**)**. Based on this fact primers designed from rice DOF genes successfully amplified the DOF domain of different millets as elucidated by the present *in silico* studies of these sequences. The DOF domain characterized in these millets could further be used for investigating the diversity of DOF genes among different millets using amplified DOF domain as probe. The *in silico* studies of DOF domain of different millets have further confirmed the conserved 50-52 amino acid residues at its N-terminal region as identified first in *Z. mays* **(**Yanagisawa, 1995**)** followed by *A. thaliana* (Zhang et al., 1995**),** *N. tobaccum* (De Paolis et al., 1996**)**, *Z. mays* (Vicente-Carbajosa et al.,1997**)** and *O. sativa* (Lijavetzky et al., 2003**)**.

*In silico analysis of cloned DOF Genes***:** The putative DOF genes of *O. sativa*, *H. vulgare*, *Z. mays*, *T. aestivum* and *E. coracana* were PCR amplified with primers DOF-011, DOF-031, DOF-025, DOF-024 and DOF-031 respectively and cloned in pGEMTeasy vector. The nucleotide sequence of cloned partial / full length DOF genes with assigned accession numbers as mentioned in **Table-2** were subjected to *in silico* analysis. The sequence of all the cloned DOF genes except of *O. sativa* showed maximum similarity with available PBF (Prolamin-box binding factor) DOF genes when subjected to BLAST. The correct frame coding for PBF protein was then deduced by subjecting it to translation tool. The nucleotide sequence along with translated protein sequence of putative PBF DOF gene of *E. coracana* (EU760631) is shown in **Figure-5**.

```
tgccctcggtgcaagtctggtaacaccaagttctgctactacaacaactacagcatgtcc
 C  P  R  C  K  S  G  N  T  K  F  C  Y  Y  N  N  Y  S  M  S
cagccgcgctacttctgcaaggcctgccgccgctactggacccatggtggctccctccgc
 Q  P  R  Y  F  C  K  A  C  R  R  Y  W  T  H  G  G  S  L  R
aacgtccccatcggtggtggttgccgcaagcccaaacgcccggggacctctgacgcccac
 N  V  P  I  G  G  G  C  R  K  P  K  R  P  G  T  S  D  A  H
aagctcggcatggcctcctcgtcggaacccacgggtgtcgtgcccccctcgaactgcaca
 K  L  G  M  A  S  S  S  E  P  T  G  V  V  P  P  S  N  C  T
gggatgaactttgctaacgtcctcccgacgtttatgtctggtggctttgacatccaaagc
 G  M  N  F  A  N  V  L  P  T  F  M  S  G  G  F  D  I  Q  S
agcctctccctgacaacttttgggtcatcatcctcatccaacccgacggcgttgatgtcc
 S  L  S  L  T  T  F  G  S  S  S  S  S  N  P  T  A  L  M  S
ccgggggggaagacatcatttctggatgtgctgagaggtggcgcaggagggcttcttgat
 P  G  G  K  T  S  F  L  D  V  L  R  G  G  A  G  G  L  L  D
ggcagcctcggtccaaacaatggcttctactatggtgggcatgccaacggatcaagcatt
 G  S  L  G  P  N  N  G  F  Y  Y  G  G  H  A  N  G  S  S  I
gggatgttgatgactccgccagcggtgtcgtttggcattccaagtccgatgcaacaacat
 G  M  L  M  T  P  P  A  V  S  F  G  I  P  S  P  M  Q  Q  H
ggcggtctcgtggttggtggaaatggaataggtggcacaacttcttcaa
 G  G  L  V  V  G  G  N  G  I  G  G  T  T  S  S
```

Figure-5: Nucleotide sequence and its translated amino acid sequence of the PBF DOF of *E. coracana* (EU760631). The DOF domain in the N-terminus is underlined. The four cysteine residues involved in the formation of a typical zinc finger are marked with red colour**.**

The putative DOF gene had the conserved DOF domain with four cysteine residues similar to what has been observed as typical feature of DOF families of proteins available from different plants.

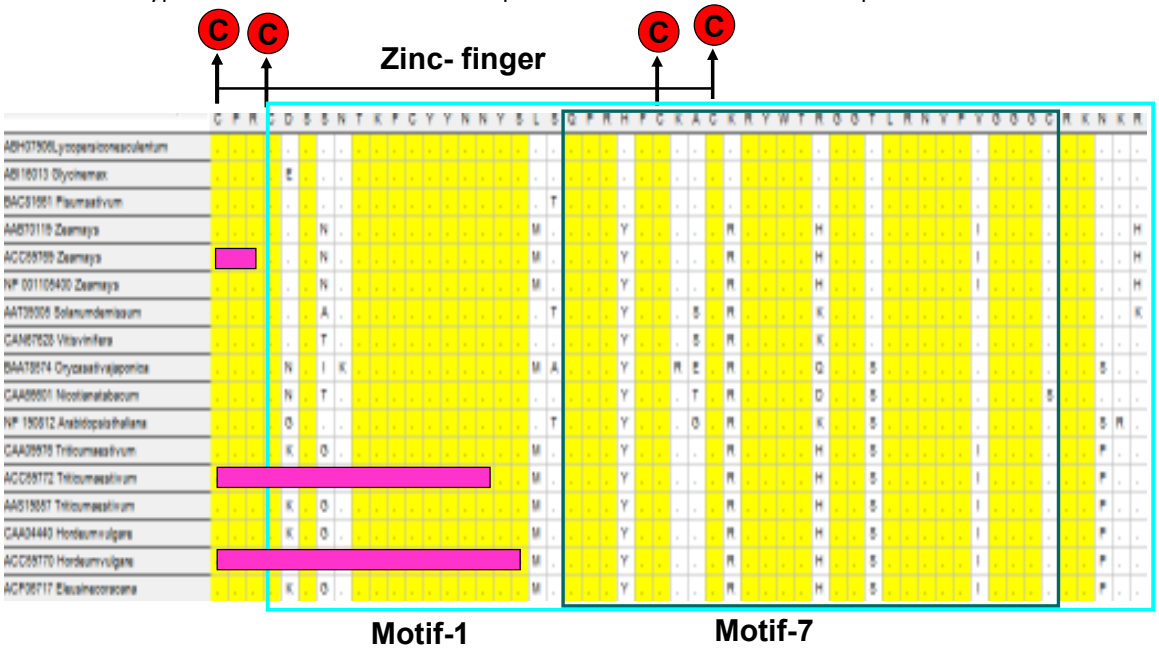

Figure-6: Variations in amino acid residues of domain region of the cloned DOF genes of *E. coracana* (ACF06717), *H. vulgare* (ACC59770), *T. aestivum* (ACC59772) and *Z. mays* (ACC59769) subjected to Multiple Sequence Alignment with other DOF genes. The two conserved motifs 1 and 7 are highlighted. The yellow dotted portion indicates the conserved amino acid residues among different DOF domains while pink marked portion indicates the partial DOF domain sequences. The four cysteine residues associated with the zinc-finger structure are indicated with arrows.

The DOF genes of *H. vulgare*, *T. aestivum* and *Z. mays* had only partial DOF domain. Further multiple sequence alignment of translated protein sequence of these cloned DOF gene reveals similarity with other published DOF gene as shown in **Figure-6** though partial variation within the DOF domain were also observed. The cloned PBF DOF genes along with 13 other PBF DOF genes subjected to phylogenetic tree construction using UPGMA method clearly revealed distinct clusters for monocots and dicots as shown in **Figure-7**.

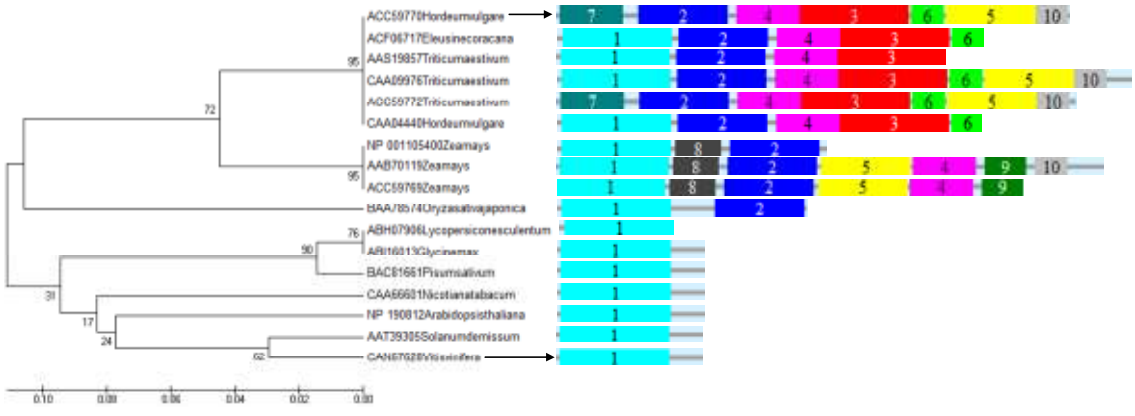

Figure-7: Phylogenetic tree constructed based on PBF DOF domain sequences and schematic distribution of respective conserved motifs identified by means of MEME software. Multilevel consensus sequences for the MEME defined motifs are listed in Table-5.

The translated protein sequence of cloned PBF DOF gene of *T. aestivum* (ACC59772), *H. vulgare* (ACC59770), *Z. mays* (ACC59769) formed cluster with existing sequences of PBF DOF genes of *T. aestivum, H. vulgare*, and *Z. mays*. The putative PBF DOF gene of *E. coracana* forms cluster with *T. aestivum* and *H. vulgare*.

The cloned PBF DOF genes along with the sequences showing maximum similarity as deduced in multiple sequence alignment were subjected to MEME program for motif analysis. A total of ten conserved motifs were observed. The distribution of conserved motifs in different PBF DOF accessions with respect to cluster is provided in **Figure-7.**

The motif-1 is most frequently observed among all PBF DOF genes owing to its function related with endosperm specific seed storage protein accumulation. In case of cloned PBF DOF genes of *T. aestivum* and *H. vulgare* the motif-1 was absent though an alternative motif related with seed storage protein accumulation i.e. motif-7 was observed as shown in **Figure-7**. Motif-7 is infact a part of motif-1 as elucidated in alignment as shown in **Figure-6.** The partial sequences of *T. aestivum* and *H. vulgare* PBF DOF genes might be one of the possible reasons for the absence of motif-1. The motifs 2 and 4 also have function related with expression of seed storage protein genes as motif 1 and 7 based on its interaction with *Cis*-regulatory sequences like CNAACAC, AAAG and TGHAAARK as observed in PLACE. Motif-3 has diverse functions related with hormone responsiveness, fermentative pathway, pathogenicity and defensive mechanism. Motif-5 is involved with auxin response and calmodulin binding sites. Motif-8 and 9 are associated with regulation of CBF and DREB genes. Motif-6 and 10 does not have any biological significance. The overall multilevel consensus sequences associated with each of the ten motifs is given in **Table-5**.

Table -5: **Multilevel consensus sequences for the MEME defined motifs among different PBF DOF sequences.**

Motif	No. of amino acids	E-value	Multilevel consensus sequence
1	50	1.3e-565	CDSINTKFCYYNNYSMSQPRYFCKACRRYWTHGGTLRNVPIGGGCRKNKR
2	41	7.0e-146	SDAHKLIVAHMMEPTTVVPPSPCTMMNFANVLPTFMSVGFE
3	50	2.7e-137	LRGGAGGLLDGSLGQNNGYYYGGHAIGSGIGMLMTPPAVSFGIPVPMQQH
4	29	1.0e-062	LSLTIFGSSSSSNTAAMMSPGGTTSFLDV
5	41	3.6e-038	FQFAMDEEHDDGMFTVMGLQWQPQVGNGAAGGVEAGGVHHE
6	15	7.3e-011	DLVVGGNGIGATTAS
7	29	2.2e-008	QPRYFCKACRRYWTHGGSLRNVPIGGGCR
8	21	4.8e-007	RFVLGSHTSSSSSATYAPLSP
9	19	3.2e+000	YGAMCTNGLSGSTTNDARQ
10	15	2.7e+002	GTGNNVTMGNSNINN

The available PBF DOF gene sequences of other crops have been used to design primers for amplifying the corresponding genes in finger millet. The cloned putative DOF gene of *E. coracana* when subjected to *in silico* analysis revealed the identity as PBF DOF and hence could be involved with the expression of seed storage protein genes like prolamin genes of *E. coracana*. Further the presence of motifs showing function related with regulation of endosperm specific seed storage protein genes during seed development further confirms the identity of cloned gene of *E. coracana* as PBF DOF. The availability of genetic map of *E. coracana* (Dida et al., 2007**)** could be used to investigate the diversity of DOF genes and its possible locations to specific chromosomes in near future. The cloned DOF genes of *T. aestivum*, *H. vulgare* and *Z. mays* showing homology with PBF DOF genes showed partial sequence with some of amino acid residues of N-terminal associated with typical DOF domain missing. The Phylogenetic tree constructed further revealed the greater degree of similarity of PBF DOF genes of monocots with the presence of motifs associated with regulation of endosperm specific seed storage proteins genes. The PBF DOF genes of dicots formed separate cluster.

CONCLUSIONS

The PCR based amplification of DOF domains of different cereals and millets using primers designed from the conserved DOF domain region of available DOF genes sequences has been *in silico* investigated after sequencing of respective PCR amplicons. The homology search, multiple sequence alignment, phylogenetic tree construction and motif analysis has clearly revealed the identity of these sequences as DOF domain of respective cereals and millets. The presence of conserved cysteine residues involved with the typical zinc finger like structure has been elucidated in most of the cereals and millets though there were some alterations, the possible reason of which may be the partial sequence of the domains obtained in some of the cereals and millets. Cloning of PCR amplified DOF domains and sequencing using universal primer might provide the uniformity in the length of sequence generated based on the expected size of conserved domain. Cloned DOF domain of rice and finger millet resulted in a sequence of 172 bp as expected based on the size of conserved domain. Cloning and sequencing of DOF domains of remaining cereals and millets might provide the uniformity of the sequences as expected based on amplicon size. *In silico* investigation of the cloned DOF genes of *E. coracana*, *H. vulgare*, *T. aestivum* and *Z. mays* revealed its identity to PBF DOF based on the presence of motifs related to regulation of endosperm specific seed storage protein genes. The PBF DOF gene of *E. coracana* is first reported based on *in silico* studies though it has already been reported in *T. aestivum*, *H. vulgare* and *Z. mays*. The temporal and spatial expression analysis of putative PBF DOF gene of *E. coracana* is in progress so as to further confirm their function related with regulation of seed storage protein genes.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the Department of Biotechnology, Govt. of India 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). H.K.'s, N.G.'s and V.K.S.'s work was supported by Junior Research Fellowship from the same grant. The facilities provided by DBT-funded SUB-DIC, Bioinformatics Centre, Pantnagar for carrying out *in silico* investigation is also thankfully acknowledged. The support provided by Director, Experiment Station, G.B. Pant University of Agriculture and Technology, Pantnagar is also thankfully acknowledged

REFERENCES

Altschul, S.F., Gish, W., Webb Miller Eugene W. Myers and Lipman, D.J. (1990). Basic Local Alignment Search Tool, Journal of Molecular Biology, 215,403-410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research, 25, 3389-3402.

Bailey, T.L. and Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches, Bioinformatics, 14, 48-54.

Bailey, T.L., Williams, N., Misleh, C. and Li, W.W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs, Nucleic Acids Research, 34, W369–W373.

Burland, T.G. (2000). DNASTAR's Lasergene sequence analysis software, Methods in Molecular Biology, 132, 71-91.

Caetano-Anolles. (2005). Evolution of Genome Size in the Grasses, Crop Science, 45, 1809-1816.

Castro, E. D., 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-W365.

De Paolis, A., Sabitini, S., De Pascalis, L., Costantino, P. and Capone, I. (1996). A rol B regulatory factor belongs to a new class of single zinc finger plant proteins, The Plant Journal, 10(2), 215-223.

Devos, K.M. and Gale, M.D. (2000). Genome Relationships: The Grass Model in Current Research, The Plant Cell, 12, 637–646.

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., Martinez, M., Moneda, I.I. and Carbonero, P. (2002). The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development, Plant Journal, 29, 453–464.

Dida, M.M., Srinivasachary., Ramakrishnan, S., Bennetzen, J.L., Gale, M.D. and Devos, K.M. (2007). The genetic map of finger millet, *Eleusine coracana*, Theoretical and Applied Genetics, 114, 321-332.

Dong, G., Ni, Z., Nie, X. and Sun, Q. (2007). 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.

Dong, Q., Lawrence, C.J., Schlueter, S.D., Wilkerson, M.D., Kurtz, S., Lushbough, C. and Brendel, V. (2005). Comparative Plant Genomics Resources at PlantGDB, Plant Physiology, 139, 610–618.

Feuillet, C. and Keller. B. (2002). Comparative genomics in the Grass families: Molecular characterization in the Grass genome structure and evolution, Annals of Botany, 89, 3-10.

Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., lassmann, T., **Moxon, S**., Marshall, M., Khanna, A., Durbin, R., Eddy, S.R., Sonnhammer, E.L.L. and Bateman, A. (2006). Pfam: clans, web tools and services, Nucleic Acids Research, 34, D247–D251.

Gao, G., Zhong, Y., Guo, A., Zhu, Q., Tang, W., Zheng, W., Gu, X., Wei, L. and Luo, j. (2006). DRTF: a database of rice transcription factors, Bioinformatics, 22(10), 1286–1287.

Guo, A., He, K., Liu, D., Bai, S., Gu, S., Wei, L. and Luo, J. (2005). DATF: a database of *Arabidopsis* transcription factors**,** Bioinformatics, 21(10), 2568–2569.

Higo, K., Ugawa, Y., Iwamoto, M. and Korenaga, T. (1999). Plant cis acting regulatory DNA elements (PLACE) database: 1999, Nucleic acid research, 27 (1), 297-300.

Iida, K., Seki, M., Sakurai, T., Satou, M., Akiyama, K., Toyoda, T., Konagaya, A. and Shinozaki, K. (2005). RARTF: Database and Tools for Complete Sets of Arabidopsis Transcription Factors, DNA Research, 12, 247–256.

Jaiswal, P., Ni, J., Yap, I., Ware, D., Spooner, W., Youens-Clark, K., Ren, L., Liang, C., Zhao, W., Ratnapu, K., Faga, B., Canaran, P., Fogleman, M., Hebbard, C., Avraham, S., Schmidt, S., Casstevens, T.M., Buckler, E.S., Stein, L. and McCouch, S. (2006). Gramene: a bird's eye view of cereal genomes, Nucleic Acids Research, 34, D717–D723.

Kellogg, E.A. (1998). Relationships of cereal crops and other grasses, Proceeding of National Academy of Sciences, USA, 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.

Kummerfeld, S.K. and Teichmann, S.A. (2006). DBD: a transcription factor prediction database, Nucleic Acids Research, 34, D74–D81.

Lijavetzky, D., Carbonero, P. and Vicente-Carbajosa, J. (2003). Genomewide comparative phylogenetic analysis of the rice and Arabidopsis DOF gene families, BMC Evolutionary Biology, 3, 17–28.

Maniatis, T., Sambrook, J. and Fritsch, E. F. (1989). Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Matthews, D.E., Carollo, V.L., Lazo, G.R. and Anderson, O.D. (2003), GrainGenes, the genome database for small-grain crops, Nucleic Acids Research, 31 (1), 183–186.

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, Plant Journal, 16, 53–62.

Moreno-Risueno, M.A., Martinez, 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.

Paterson, A.H., Bowers, J.E. and Chapman B.A. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics, Proceeding of National Academy of Sciences, USA, 101(26), 9903-9908.

Paterson, A.H., Freeling, M. and Sasaki, T. (2005). Grains of knowledge: Genomics of model cereals, Genome research, 15, 1643–1650.

Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Research, 35, D61–D65.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R. (2005). InterProScan: protein domains identifier, Nucleic Acids Research, 33, W116–W120.

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 and Integrative Genomics, 6, 310-321.

Riano-Pachon, D.M., Ruzicic, S., Dreyer, I. and Mueller-Roeber, B. (2007). PlnTFDB: an integrative plant transcription factor database, BMC Bioinformatics, 8:42

Sakata, K., Nagamura, Y., Numa, H., Antonio, B.A., Nagasaki, H., Idonuma, A., watanabe, W., Shimizu, Y., Horiuchi, I., Matsumoto, T., Sasaki, T. and Higo, K. (2002). RiceGAAS: an automated annotation system and database for rice genome sequence, Nucleic Acids Research, 30 (1), 98-102.

Solovyev, V., Kosarev, P., Seledsov, I. and Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters, Genome Biology, 7(1): S10

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 genome, Theoretical and Applied Genetics, 115, 489-499.

Stormo, G.D. (2000). Gene-Finding Approaches for Eukaryotes, Genome Research, 10, 394-397.

Takatsuji, H. (1998). Zinc-Finger transcription factor in plants, Cellular and Molecular Life Sciences, 54, 582-596.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0, Molecular Biology and Evolution, 24(8), 1596–1599.

Tateno, Y., Imanishi, T., Miyazaki, S., Fukami-Kobayashi, K., Saitou, N., Sugawara, H. and Gojobori, T. (2002) DNA Data Bank of Japan (DDBJ) for genome scale research in life science, Nucleic Acid Research, 30(1), 27-30.

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(22), 4673-4680.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. and Leunissen, J.A.M. (2007). Primer3Plus, an enhanced web interface to Primer3, Nucleic Acids Research, 35, W71–W74.

Varshney, R.K., Hoisington, A.D. and Tyagi, K.A. (2006). Advances in cereal genomics and applications in crop breeding, Trends in Biotechnology, 24, 1-10.

Vicente-Carbajosa, J., Moose, S.P., Parson, R.l. and Schmidt, R.J. (1997). A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator, Proceeding of National Academy of Sciences,USA, 94, 7685–7690.

Yamamoto, M.P., Onodera, Y., Touno, S.M. and Takaiwa, F. (2006). Synergism between RPBF DOF and RISBZ1 bZIP Activators in Regulation of Rice Seed Expression Genes, Plant Physiology, 141, 1694-1707.

Yanagisawa, S. (1995). A novel DNA-binding domain that may form a single zinc finger motif, Nucleic Acids Research, 23, 3403–3410.

Yanagisawa, S. (1996). DOF DNA binding proteins contain a novel zinc finger motif. Trends in Plant Science, 1, 213-214.

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 (12), 555-560.

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

Zhang, B. (1995). Interactions between distinct types of DNA binding proteins enhance binding to ocs element promoter sequences, Plant Cell, 7, 2241–2252.

.