PCR amplification and bioinformatics assessment of promoters of \textit{PBF-DOF} (DNA binding with one finger) genes of finger millet

Chandra Prakash\textsuperscript{1,2}, Hariom Kushwaha\textsuperscript{1,3}, Vinay Kumar Singh\textsuperscript{1,4}, Praveen K Singh\textsuperscript{5}, Anil Kumar\textsuperscript{1} and Dinesh Yadav*\textsuperscript{1,6}

\textsuperscript{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
\textsuperscript{2}National Research Centre on Plant Biotechnology, IARI, New Delhi 110012, India
\textsuperscript{3}Institute of Microbial Technology (CSIR), Sector 39-A, Chandigarh 160036, India
\textsuperscript{4}Centre for Bioinformatics, School of Biotechnology, Banaras Hindu University, Varanasi, Uttar Pradesh-221 005, India
\textsuperscript{5}Division of Seed Science and Technology, Indian Agriculture Research Institute, New Delhi 110012, India
\textsuperscript{6}Department of Biotechnology, D.DU Gorakhpur University, Gorakhpur, Uttar Pradesh-273 009, India

*Corresponding author: dinesh_yad@rediffmail.com

Abstract

The Dof (DNA binding with one finger) family is a plant specific transcription factor known to be involved in regulating diverse functions in plants and have been extensively studied in many crops. The Dof transcription factor regulating gene expression by interacting with \\textit{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 \textit{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.
- \textit{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, \textit{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, 1999b; 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 cis-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 trans-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 prolam in box (P-box) is a highly conserved 7-bp sequence element (5’-TGTAAGG-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 et al., 2008).
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 syntenies 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>
<thead>
<tr>
<th>Sl. No</th>
<th>Primer</th>
<th>Locus/Source</th>
<th>Primer sequences forward 5’ – 3’</th>
<th>Primer sequences reverse 5’ – 3’</th>
<th>Amplicon size (in bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dof-01</td>
<td>OsIBCD010967</td>
<td>AGGGGGACGGTGCAAGAGATG</td>
<td>GCGGCTGGCAGGGCTGTAGG</td>
<td>993</td>
</tr>
<tr>
<td>2</td>
<td>Dof-02</td>
<td>OsIBCD041244</td>
<td>TATCCTCCAGGTCTAGGCTG</td>
<td>GCGTGAAGCTGTAGTTTGGTA</td>
<td>946</td>
</tr>
<tr>
<td>3</td>
<td>Dof-03</td>
<td>Os01g09720</td>
<td>GTACTCTAGCTCTCTCGGTTCTAT</td>
<td>TAGTACAGAACCCTGTTGTTG</td>
<td>1122</td>
</tr>
<tr>
<td>4</td>
<td>Dof-04</td>
<td>Os01g48290</td>
<td>CTAGCTTTGGCTACGGTTG</td>
<td>TGAAGTAGCAGAACTTGGTCTC</td>
<td>1263</td>
</tr>
<tr>
<td>5</td>
<td>Dof-05</td>
<td>Os03g60630</td>
<td>CTTTCTCTAGCTCTCTTGTTCT</td>
<td>GAATAGCAGAACCCTGGTTG</td>
<td>1043</td>
</tr>
<tr>
<td>6</td>
<td>Dof-06</td>
<td>Os02g15350</td>
<td>ACAACACACACACACACACACAC</td>
<td>CGCAGCTCGTCTAGTTGTTGTA</td>
<td>1100</td>
</tr>
<tr>
<td>7</td>
<td>Dof-07</td>
<td>Consensus</td>
<td>AAGAAGAAGAAGAAGAAGAGAT</td>
<td>CGAAGTAGGTTGAGTTAGCA</td>
<td>775</td>
</tr>
<tr>
<td>8</td>
<td>Dof-08</td>
<td>SBDOF1</td>
<td>CCAAAACCCCGCAAAGGAAATC</td>
<td>GCTGCGCCGTTGTTAGTTGGA</td>
<td>1021</td>
</tr>
<tr>
<td>9</td>
<td>Dof-09</td>
<td>SBDOF2</td>
<td>ATGTTTGGCCGGTTATGTTG</td>
<td>GCGTGGGCTGTTGATGTTGTA</td>
<td>2233</td>
</tr>
<tr>
<td>10</td>
<td>Dof-10</td>
<td>SB8g003030.1</td>
<td>GTACTCTCTCTCTTTGTTTTGTT</td>
<td>GCACAGCCTTGGAAAACTTA</td>
<td>944</td>
</tr>
<tr>
<td>11</td>
<td>Dof-11</td>
<td>DOF2-chr3</td>
<td>CACTTTTCCGTGTGGTTTAG</td>
<td>GCACAGCCTGTTGTTGTA</td>
<td>1056</td>
</tr>
<tr>
<td>12</td>
<td>Dof-12</td>
<td>Uncharacterized DOF</td>
<td>GTCGAGCTCCTTTGATTGTAGG</td>
<td>TGAAGTAGCAGAACTTGGTTG</td>
<td>1957</td>
</tr>
<tr>
<td>13</td>
<td>Dof-13</td>
<td>Sb chr01</td>
<td>AGGTACCCCTATGCAATTAGC</td>
<td>GGGAGAGGAGTAGTTGTTG</td>
<td>1115</td>
</tr>
<tr>
<td>14</td>
<td>Dof-14</td>
<td>Sb chr04 g030420.1</td>
<td>GCTCTCTAGCTCGTTAGGTACTCT</td>
<td>CGAGACGGTCTAGTTGGTGA</td>
<td>2607</td>
</tr>
<tr>
<td>15</td>
<td>Dof-15</td>
<td>AF385139.1</td>
<td>GCTGAGCGGCCAACATAGTAAGAGA</td>
<td>ACCGCAAATAGGCAAAGCAGA</td>
<td>1318</td>
</tr>
</tbody>
</table>
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% HgCl₂ 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 CaCl₂ 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
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 prerequisite 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.
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.
PCR amplification and bioinformatics assessment of promoters of PBF-DOF (DNA binding with one finger)...

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

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

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,

<table>
<thead>
<tr>
<th>Motifs</th>
<th>TFs binding site (TOMTOM)</th>
<th>Gene Ontology predictions for Motifs (GOMo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motif-1</td>
<td>AtMYB84, abi4, SOC1, ABF1, SMZ, ERF1, AtMYB15 etc.</td>
<td>Chloroplast, nucleus, transcription, factor activity, protein binding, plasma membrane</td>
</tr>
<tr>
<td>Motif-2</td>
<td>PI, SOC1, SVP, FLC, AGL15, HMG1/Y, Dof2, PBF etc.</td>
<td>Transcription factor activity, Nucleus, plasma membrane, regulation of transcription (DNA dependent), kinase activity.</td>
</tr>
<tr>
<td>Motif-3</td>
<td>AtMYB84, RAV1, ERF1, bZIP911, abi4 etc.</td>
<td>Chloroplast</td>
</tr>
</tbody>
</table>

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

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, AGL15 - Agamous-like 15, HMG1/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
PCR amplification and bioinformatics assessment of promoters of PBF-DOF (DNA binding with one finger)...

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.

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

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 pattern 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


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


