4.1 PCR amplification of Dof domains/genes and studying the polymorphisms generated by different sets of primers for cereals and millets

PCR amplification of putative *Dof* genes and conserved Dof domain of cereals and millets has been attempted using primers designed from published sequences. The banding patterns obtained by different sets of Dof domain and gene specific primers were utilized for revealing the phylogenetic relationship among cereals and millets. These sets of primers showing polymorphism among different cereals and millets could be utilized as functional markers after extensive validation in other crops too.

4.1.1 Isolation, purification and quantification of genomic DNA

Good quality genomic DNA is a prerequisite for PCR amplification. Genomic DNA of cereals (rice, wheat, oat, sorghum, barley, maize) and millets (finger millet, barnyard millet, proso millet, little millet, kodo millet, foxtail millet) were isolated by standard CTAB method **(Murray & Thompson, 1980)**. The quantitative estimation and purity of isolated genomic DNA were monitored spectrophotometrically by measuring light absorbance (A) at 260 nm and 280 nm as shown in **Table 4.1**. The $OD_{260/280}$ values ranged from 1.77 to 2.0. The qualities of isolated genomic DNAs were further analyzed by 0.8 % agarose gel electrophoresis. The intact and discrete band of genomic DNAs was observed on agarose gel (**Fig. 4.1)**.

Table 4.1 Spectrophotometric quantification of genomic DNA of cereals and millets

Crop	**Cultivar**	**A_{260}**	**A_{280}**	**A_{260}/A_{280}**	**Concentration of DNA (ng/µl)**
Rice	Pusa sughandha	0.459	0.250	1.83	2295
Wheat	Punjab PBW 373	0.587	0.330	1.77	2930
Sorghum	Pant chari 5	0.710	0.383	1.85	3550
Barley	VLB56	0.862	0.483	1.78	4310
Oat	UP0212	0.102	0.050	2.0	511
Maize	DQPMC4W	0.441	0.230	1.91	2205
Finger millet	VL315	0.538	0.288	1.87	2690
Barnyard millet	PRB401	0.723	0.389	1.86	3615
Proso millet	405	0.678	0.365	1.86	3390
Little millet	Local	0.687	0.382	1.8	3435
Kodo millet	Local	0.809	0.452	1.79	4045
Foxtail millet	PRK1	0.569	0.303	1.88	2845

4.1.2 Primer designing for PCR amplification of Dof domain and *Dof* genes

Nucleotide sequences of different *Dof* genes of cereals (wheat, maize, rice and barley) reported in literature were retrieved from NCBI and DRTF databases. Multiple sequence alignment of different Dof domain and *Dof* genes were done by ClustalW program and primers were designed from resulting consensus sequences using DNASTAR, Primer3 and Gene Runner softwares. A total of 35 sets of primers were designed for PCR amplification of Dof domain and *Dof* genes of different cereals and millets as shown in **Table 3.4** and **3.5** in materials and methods section.

4.1.3 PCR amplification of Dof domain and *Dof* genes

Dof domain and *Dof* genes were subjected to PCR amplification using different sets of specific primers based on primer information as listed in **Tables 3.4** and **3.5**. PCR was performed as per the standard protocol and the resulting PCR products were analyzed on 1.5 % agarose gel and documented in the Gel Documentation System. The standard molecular weight markers (λ DNA *Hin*dIII digested DNA marker, 100 bp DNA ladder marker, 500 bp DNA ladder marker and 1 kbp DNA ladder marker) were used for analysis of fragment size. All amplification reactions were performed thrice to ensure consistency and reproducibility of finger prints generated using Dof domain and gene specific primers.

4.1.3.1 PCR amplification of Dof domains

A set of four primers, Dof-D1, Dof-D2, Dof-D3 and Dof-D4 as listed in **Table 3.4** were designed from flanking sequences of rice Dof domain. The PCR amplification using these set of primers resulted in multiple band formation across all the cereals and millets as shown in **Fig. 4.2**. The designed primers from flanking sequences of Dof domain were unable to produce single band of conserved Dof domain, hence another set of primers, Dof-1 and Dof-8 were designed from the conserved sequences for amplification of respective Dof domain. Using primers Dof-8, Dof domains of cereals namely rice, wheat, sorghum, barley and oat, and millets *viz.* finger millet, barnyard millet, proso millet, little millet, kodo millet and foxtail millet were PCR amplified and analyzed on 1.5 % agarose gel. The expected size bands of 172 bp were observed across all the crops except in case of barnyard millet where two bands was observed (**Fig. 4.3**). These bands were gel eluted and sequenced with Dof-8 primer.

4.1.3.2 PCR amplification of *Dof* genes

The *Dof* genes of rice, barley, maize, wheat and finger millet were PCR amplified with Dof-11, Dof-31, Dof-25, Dof-24 and Dof-31 primers, respectively (**Table 3.4**). The PCR amplification profile showed multiple band formation along with expected size amplicons i.e. 1065 bp (Dof-11), 785 bp (Dof-24), 715 bp (Dof-25) and 783 bp (Dof-31) in few of cereals and millets as shown in **Fig. 4.5b, 4.7b, 4.7c** and **4.8a**. These expected size bands were gel eluted, cloned and sequenced.

A total of four *Dof* genes of sorghum were PCR amplified using primer SbDof1, SbDof2, SbDof3 and SbDof4 designed from the *in silico* predicted *Dof* genes from the whole genome sequence of sorghum (**Table 3.5** of materials and methods section). The amplified products were analyzed on 1.5 % agarose gel and single band of expected sizes i.e. 937 bp (SbDof1), 966 bp (SbDof19), 2226 bp (SbDof23) and 1043 bp (SbDof24) were obtained **(Fig. 4.4)**. These bands were gel eluted, cloned and sequenced.

4.1.4 Comparative analysis of cereals and millets based on banding patterns generated by Dof domain and *Dof* genes-specific primers

The banding pattern generated by 24 sets of Dof domain primers (Dof-D1 to D4) and *Dof* gene-specific primers (Dof-10 to 13, Dof-15 to 22, Dof-24, Dof-25, Dof-28, Dof-31 and Dof-34 to 37) were utilized for studying the polymorphism among cereals and millets for revealing phylogenetic relatedness. The phylogenetic analysis was carried out using the NTSYS-pc version 2.11w software to calculate the similarity values and generate the phenogram. The similarity matrices were utilized to construct the UPGMA dendrogram. The PCR amplification pattern generated by different sets of Dof primers using genomic DNA of cereals and millets as template revealed various polymorphic and unique bands (**Fig. 4.2, 4.5, 4.6, 4.7** and **4.8**).

The existence of multiple bands using *Dof* gene specific primers in different cereals and millets might be related with the inherent diversity of *Dof* genes in the genomes of different crops. The amplified products were sized and then binary coded by 1 or 0 for their presence or absence. The variable size bands ranging from 0.1 to 3.0 kbp observed with different sets of primers along with polymorphic and monomorphic bands are presented in **Table 4.2**. The 24 sets of primers used in diversity analysis generated a total of 906 bands, out of which 894 were polymorphic and 12 were monomorphic in nature. The percentage of polymorphism ranged from 69.23 to 100 %. Very high degrees of polymorphism (100 %) were observed with all the primers except Dof-24 primer (69.23 %). The highest number of

bands (84) was obtained with primer Dof-15 (**Fig. 4.5c**), while the lowest number (25) was obtained with primer Dof-12. There exists considerable variation in the banding pattern among cereals and millets with respect to Dof domain and gene specific primers as visualized on 1.5 % agarose gel.

Table 4.2 The number of bands detected in different cereals and millets using 24 sets of Dof domain and gene-specific primers

S. No.	Primer code	Total no. of bands	Size range of bands (kbp)	Monomorphic bands	Polymorphic bands	Degree of polymorphism (%)
1	Dof-10	34	0.15-2.0	0	34	100
2	Dof-11	28	0.2-1.2	0	28	100
3	Dof-12	25	0.3-1.7	0	25	100
4	Dof-13	33	0.2-1.5	0	33	100
5	Dof-15	84	0.1-1.2	0	84	100
6	Dof-16	40	0.25-2.5	0	40	100
7	Dof-17	25	0.15-1.8	0	25	100
8	Dof-18	41	0.15-1.8	0	41	100
9	Dof-19	31	0.15-2.0	0	31	100
10	Dof-20	30	0.2-1.8	0	30	100
11	Dof-21	32	0.25-2.2	0	32	100
12	Dof-22	40	0.2-2.5	0	40	100
13	Dof-24	39	0.35-3.0	12	27	69.23
14	Dof-25	32	0.2-1.6	0	32	100
15	Dof-28	35	0.25-2.5	0	35	100
16	Dof-31	35	0.35-1.4	0	35	100
17	Dof-34	38	0.15-1.6	0	38	100
18	Dof-35	28	0.15-1.8	0	28	100
19	Dof-36	39	0.1-2.0	0	39	100
20	Dof-37	28	0.15-1.4	0	28	100
21	Dof-D1	25	0.45-2.1	0	25	100
22	Dof-D2	35	0.15-2.5	0	35	100
23	Dof-D3	57	0.2-2.3	0	57	100
24	Dof-D4	72	0.1-2.5	0	72	100
	Total	906		12	894	

Binary matrix generated based on the presence (1) and absence (0) of bands resulting from 24 sets of primers was subjected to NTSYS-pc software for deducing phylogenetic relationship with Jaccard's similarity coefficient values ranging from 0.0601 to 0.2118 in all cereals and millets (**Fig. 4.9**). Similarity matrices were utilized to construct the dendrogram using UPGMA method which showed two major clusters A and B. The major cluster A comprises of rice, sorghum, maize, finger millet, foxtail millet, barnyard millet and proso millet, while cluster B has wheat, barley, oat, little millet and kodo millet. The major clusters

A and B were further bifurcated into two sub-clusters as shown in **Fig. 4.10**. The distinct sub-clusters for rice and sorghum in major cluster-A along with close relationship of wheat, barley and oat in major cluster-B are in coherence with reports based on evolutionary patterns of genome rearrangement **(Kellogg, 1998)**.

Based on the preliminary results of PCR amplification of cereals and millets using Dof domain and *Dof* gene specific primers there is a possibility of developing these primers as functional markers as marker designed from genes encoding cytochrome P450, uridyldiphosphate glycosyltransferase (UGT) and 5 S rRNA **(Kumar *et al.*, 2007)**. The cytochrome P450 representing multi-gene family has been utilized as functional markers for genetic diversity studies in plants **(Somerville & Somerville, 1999; Tanksley & McCouch, 1997; Watanabe & Iwanaga, 1999)**. These sets of primers needs to be validated for its suitability as markers for diversity analysis by using various accessions/cultivars of different cereals and millets as performed with commonly used molecular markers like RAPDs, RFLPs, AFLPs, SSRs, ISSRs, etc **(Karp *et al.*, 1997; Powell *et al.*, 1996)**. Functional markers could provide universal tools for the assessment of genome-wide genetic diversity in diverse plants species for which relevant genetic markers and prior inheritance knowledge of traits are lacking.

The comparative studies on cereals and millets based on the banding pattern generated by *Dof* gene and domain specific primers might be utilized for developing functional markers after extensive validation. The cloning and sequencing of these bands might give certain insight into the largely unknown genome of millets through comparative genomics. This might have some implication regarding their syntenic relationships provided that all are sequenced and mapped onto respective genomes.

4.2 PCR based cloning, sequencing and *in silico* characterization of Dof domain and *Dof* genes of cereals and millet

The PCR amplified *Dof* genes of cereals and millets (rice, wheat, barley, maize, sorghum and finger millet) and conserved Dof domain of rice and finger millet were gel eluted and cloned in two different cloning vectors pGEM-T Easy and pBSK and subsequently sequenced using M13 universals. The PCR amplicons of Dof domain from rice, wheat, barley, sorghum, oat, finger millet, barnyard millet, proso millet, little millet, kodo millet and foxtail millet were eluted and directly sequenced with domain specific primer. The sequences of *Dof* genes and domains of cereals and millets were submitted to GenBank and assigned accession number and these were further subjected to extensive *in silico*

characterization for homology search, multiple sequence alignment, phylogenetic tree construction and motif analysis using different bioinformatics tools.

4.2.1 Gel elution of PCR products

The PCR amplification of *Dof* genes using different sets of primers resulted in multiple bands along with the expected size bands. The expected size band of PCR products were gel eluted and quality of purified amplicons were analyzed on 1.5 % agarose gel (**Fig. 4.11** and **4.12**). The gel eluted PCR amplicons were found to be of good quality and yield indicating efficient elution. The gel extracted PCR amplicons were also quantified by Nanodrop for ligation reaction setup (**Table 4.3**).

Table 4.3 Concentration of eluted PCR products

S. No.	Sample (primer name)	Concentrations (ng/µl)
1	Finger millet(Dof-1)	37.8
2	Rice (Dof-1)	71
3	Finger millet (Dof-8)	57
4	Rice (Dof-8)	92
5	Finger millet (Dof-24)	162
6	Finger millet (Dof-31)	12.5
7	Finger millet (Dof-11)	25
8	Sorghum (SbDof1)	90
9	Sorghum (SbDof19)	120
10	Sorghum (SbDof23)	170
11	Sorghum (SbDof24)	190
12	Barley (Dof-31)	199
13	Wheat (Dof-24)	171
14	Rice (Dof-11)	209
15	Rice (Dof-15)	20
16	Maize (Dof-25)	25

4.2.2 Cloning of Dof domain and *Dof* genes using pGEM-T Easy vector

The eluted and quantified PCR product of *Dof* genes of *O. sativa*, *H. vulgare*, *Z. mays*, *T. aestivum* and *E. coracana* and Dof domain of finger millet and rice (**Fig. 4.11** and **4.12a**) were ligated at MCS region of pGEM-T Easy vector. pGEM-T Easy vector is a convenient system for the cloning of PCR products amplified by certain polymerases like *Taq* DNA polymerase. Ligation reaction for eluted PCR products was set up in 10µl volume and incubated at 4ºC for overnight. Vector- insert molar ratio of 3:1 was used in the reaction as described in materials and methods section **3.2.7.1**.

The ligated products were transformed by $CaCl_2$ method (as described in material and methods section **3.2.7.3**) in *E. coli* DH5α strain. The initial screenings of

recombinants were made on the basis of blue-white selection on LB agar plates containing ampicillin/ isopropyl thiogalactoside (IPTG) and 5-bromo 4-chloro 3-indolyl β-D-galactoside (X-gal). Some of the recombinant white colonies were picked and inoculated in LB media containing ampicillin (50 µg/ml) and incubated overnight with shaking at 37°C. The overnight grown cultures were used for miniprepartion of plasmid DNA as mentioned in materials and methods section **3.2.7.5.1**.

The putative recombinant plasmids were isolated by alkali lysis method from some of the selected white colonies. The MCS region of pGEM-T Easy vector is flanked by *Eco*RI restriction sites. The restriction digestion of recombinant plasmid with *Eco*RI enzyme resulted in the release of expected size DNA insert as analyzed on agarose gel (**Fig. 4.13a** and **4.14)**.

4.2.3 Cloning of *Dof* genes of sorghum using pBSK vector

The gel eluted and quantified PCR products of *Dof* genes of sorghum were ligated at MCS region of pBSK cloning vector. The eluted product of *SbDof1* was ligated at *Eco*RI and *Hin*dIII sites, while *SbDof19, SbDof23* and *SbDof24* were ligated at *Sma*I site of pBSK vector based on the deliberate incorporation of these sites in the designed primers. Ligation reaction was set up in 10 µl volume and incubated at 16ºC for 16 h with vector- insert molar ratio of 4:1. Ligated products were then transformed in *E. coli* host (DH5α) using electroporation method as described in materials and methods section **3.2.7.4**. Transformed cells were plated on LB plate containing ampicillin (50 µg/ml) as selection marker.

The putative recombinant plasmids DNA were isolated by alkali lysis method from the transformed colonies. The putative recombinant clones were confirmed by restriction digestion of plasmid DNA with *Eco*RI/*Hin*dIII (*SbDof1*) and *Sma*I (*SbDof19, SbDof23* and *SbDof24*) restriction enzymes. The restriction digestion revealed the liberation of expected size insert of 937 bp (*SbDof*1), 966 bp (*SbDof*19), 2,226 bp (*SbDof*23) and 1,043 bp (*SbDof*24) as visualized on 1.2 % agarose gel (**Fig. 4.13b)**.

4.2.4 Sequencing of Dof domain and genes

A total of 60 eluted PCR products and recombinant plasmid DNAs were subjected to sequencing PCR reaction followed by processing (post reaction cleanup) of samples. The purified and dried samples were resuspended in HiDi formamide and denatured in thermal

cycler. Denatured samples were loaded in to 96 well plates and proceeded for capillary electrophoresis in Applied Biosystems sequencer for sequencing.

The PCR amplicons of Dof domain of cereals (rice, wheat, barley, oat, sorghum) and millets (finger millet, barnyard millet, proso millet, little millet, kodo millet, foxtail millet) were directly sequenced with Dof domain specific primers while putative *Dof* genes of finger millet, rice, wheat, maize, sorghum and barley and Dof domain of finger millet and rice were sequenced using M13 universal primer available in vector.

A total of 32 sequences of *Dof* genes/domains of different cereals and millets were submitted to GenBank and assigned accession numbers EU586262 to EU586269, EU760631 to EU760640, FJ854501 to FJ854504, GQ924919 to GQ924921, GQ352370 to GQ352372 and HQ540084 to HQ540087. The details about *Dof* genes and domain of various cereals and millets with assigned accession numbers are listed in **Table 4.4**.

Table 4.4 List of sequenced *Dof* genes and domains of cereals and millets

Crops (Scientific name)	**Cultivar**	**Sequence length (bp)**	**Primer used for sequencing**	**Dof gene/ domain**	**Accession number**
Rice (*Oryza sativa*)	Pusa sughandha	939	Dof-11	Gene	EU586268
		112	Dof-8	Domain	EU586263
		106	Dof-1	Domain	EU586265
		172	M13	Domain	FJ854504
		210	M13	Domain	FJ854502
Wheat (*Triticum aestivum*)	Punjab PBW 373	724	M13	Gene	EU586269
		94	Dof-8	Domain	EU760640
Sorghum (*Sorghum bicolor*)	Pant chari 5	116	Dof-8	Domain	EU760632
		937	M13	Gene	HQ540084
		966	M13	Gene	HQ540085
		2266	M13	Gene	HQ540086
		1043	M13	Gene	HQ540087
Barley (*Hordeum vulgare*)	VLB56	712	M13	Gene	EU586267
		121	Dof-8	Domain	EU760633
Oat (*Avena sativa*)	UP0212	334	Dof-8	Domain	EU760634
Maize (*Zea mays*)	DQPMC4W	644	Dof-25	Gene	EU586266
Finger millet (*Eleusine coracana*)	VL315	589	M13	Gene	EU760631
		409	Dof-8	Domain	EU586262
		137	Dof-1	Domain	EU586264
		367	Dof-59	Gene	GQ924921
		172	M13	Domain	FJ854503
		210	M13	Domain	FJ854501
Barnyard millet (*Echinochloa frumentacea*)	PRB401	116	Dof-8	Domain	EU760639
Proso millet (*Panicum milliaceum* Linn.)	405	334	Dof-8	Domain	EU760635
		143	Dof-19	Domain	GQ352370

Little millet (*Panicum antidotale*)	Local	110	Dof-8	Domain	EU760636
		396	Dof-19	Gene	GQ352371
Kodo millet (*Paspalum scrobiculatum* Linn.)	Local	108	Dof-8	Domain	EU760637
Foxtail millet (*Setaria italica* Beauv.)	PRK1	70	Dof-8	Domain	EU760638
		145	Dof-19	Domain	GQ352372

4.2.5 *In silico* characterization of sequenced Dof domains of cereals and millets

The sequences of PCR amplified Dof domains of different cereals and millets (**Table 4.4**) revealed a variability in sequence length ranging from 70 to 409 bp though the PCR amplicon of expected size i.e. 172 bp was uniformly observed with all the templates except maize as shown in **Fig. 4.3**. This variation in sequence length might be due to variable length of sequencing by domain specific primers. The PCR amplified Dof domain of rice and finger millet were cloned in pGEM-T Easy vector and sequenced with M13 universal primer giving exact length of sequences i. e. 210 and 172 bp resulting from PCR amplification with Dof-1 and Dof-8 primers, respectively (FJ854501-FJ854504).

The Dof domain sequences with assigned accession numbers as mentioned in **Table 4.4** were subjected to BLAST search for deducing similarity with available sequences in NCBI database which revealed identity with Dof domain sequences of other plants. The nucleotide sequences were translated to respective protein sequences using translation tool. Further the deduced protein sequences so translated were confirmed by subjecting the nucleotide sequences to gene finding software namely GENESCAN and FGENESH. The obtained protein sequences were subjected to BLASTP to reveal the similarity at protein level with other existing Dof domain proteins. The protein sequences showing similarity were retrieved and subjected to multiple sequence alignment using ClustalW as shown in **Fig. 4.15**. The presence of four cysteine residues associated with typical zinc like finger of Dof family of proteins was observed in domain of *P. scrobiculatum*, *E. frumentacea*, *A. sativa*, *P. milliaceum* and *O. sativa*.

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. Cloning of PCR amplified Dof domains and sequencing using universal primer might provide the uniformity in the length of sequence generated as the expected size of conserved domain. Sequencing of cloned Dof domain of rice and finger millet resulted in a sequence of 172 bp as expected based on the size of PCR amplified Dof

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. Protein functional analysis of these domains using PFAM, PROSITE and InterProScan further confirmed their identity to Dof like proteins.

4.2.5.1 Phylogenetic and motif analysis of sequenced Dof domains of cereals and millets

The sequenced Dof domains of cereals and millets along with 20 other published domain sequences from NCBI database were subjected to phylogenetic tree construction using UPGMA method, which revealed two distinct clusters (**Fig. 4.16)**.

The major cluster- I comprises of 30 Dof domain sequences while Cluster-II has only one member (*O. sativa*, ACC59766). The Major cluster-I was further divided into two sub-clusters A and B comprising of 29 and 1 sequences respectively. The sub-cluster A was further bifurcated into many small clusters. The PCR amplified Dof domain of rice (ACC59766) represent the single member of major cluster-I. The PCR amplified Dof domain of wheat (ACF06726) and barley (ACF06719) were closely placed in sub-cluster A of major cluster-I. Similarly the PCR amplified Dof domain of kodo millet (ACF06723), finger millet (ACC59765), sorghum (ACF06718), little millet (ACF06722) were grouped together in sub-cluster A. The Dof domain of foxtail millet (ACF06724) represents the single member of sub-cluster B, while Dof domain of oat (ACF06720) and little millet (ACF06721) were placed in sub-cluster A (**Fig. 4.16**). The Dof domain of dicots and monocots were grouped together revealing the highly conserved nature of the Dof domain and the variation if any might be due to variation in the sequence length of Dof domain as observed in case of the sequences of PCR amplified Dof domain of cereals and millets.

These 31 Dof domain sequences were further subjected to MEME tool for motif analysis. A total of five conserved motifs were observed in different Dof domain sequences. The distributions of conserved motifs in different accessions are shown in **Fig. 4.16**. The multilevel consensus sequences of each motif are provided in **Table 4.5**. 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 4.5 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 the 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 the regulation of flavonoid biosynthesis. Motif 4 and 5 lack any specified function.

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 co-linearity 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 *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)**, *T. aestivum* **(Chen *et al.*, 2005)** and *O. sativa* **(Lijavetzky *et al.*, 2003)**.

4.2.6 *In silico* characterization of cloned *Dof* genes

The nucleotide sequence of cloned *Dof* genes of *O. sativa* (EU586268), *H. vulgare* (EU586267), *Z. mays* (EU586266), *T. aestivum* (EU586269) and *E. coracana* (EU586270) were *in silico* analyzed for homology search, multiple sequence alignment, phylogenetic tree construction and motif 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 search. The correct open reading frame determination and translation of nucleotide sequences were then performed by translation tool. The nucleotide

sequence along with translated protein sequence of putative *Dof* genes of *E. coracana, O.sativa, Z. mays, T. aestivum, H. vulgare* and *S. bicolor* are shown in **Fig.4.17**.

The protein sequences showing similarity with these cloned genes on BLAST search were retrieved and taken in FASTA format followed by subjecting to multiple sequence alignment. This revealed maximum similarity with in Dof domain region of proteins, though partial variations within the Dof domain were also observed as shown in **Fig. 4.18**. The putative *Dof* gene had the conserved Dof domain with four cysteine residues similar to what has been observed as classical feature of Dof families of proteins available from different plants. The *Dof* genes of *H. vulgare, T. aestivum* and *Z. mays* had only partial Dof domain.

4.2.6.1 Phylogenetic and motif analysis

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 **Fig. 4.19**. 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* formed cluster with *T. aestivum* and *H. vulgare*.

The proteins sequences of cloned PBF *Dof* genes along with the sequences showing maximum similarity as deduced in multiple sequence alignment were subjected to MEME tools for motif analysis. A total of ten conserved motifs were observed. The distribution of conserved motifs in different Dof accessions is provided in **Fig. 4. 20**.

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 **Fig. 4.20**. Motif-7 is infact a part of motif-1 as elucidated in alignment as shown in **Fig. 4.18**. 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 do not have any biological significance. The overall multilevel consensus sequences associated with each of the ten motifs is given in **Table 4.6**.

Table 4.6 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	CDSINTKFCYYNNYSMSQPRYFCKACRRYWTHGGTLR NVPIGGGCRKNKR
2	41	7.0e-146	SDAHKLIVAHMMEPTTVVPPSPCTMMNFANVLPTFMS VGFE
3	50	2.7e-137	LRGGAGGLLDGSLGQNNGYYYGGHAIGSGIGMLMTPP AVSFGIPVPMQQH
4	29	1.0e-062	LSLTIFGSSSSSNTAAMMSPGGTTSFLDV
5	41	3.6e-038	FQFAMDEEHDDGMFTVMGLQWQPQVGNGAAGGVEA GGVHHE
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 *in silico* analysis of cloned putative *Dof* gene of *E. coracana* revealed the identity to PBF Dof proteins and hence could be involved in the expression of seed storage protein genes like prolamin genes of *E. coracana*. The presence of motifs showing functions 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 expression profiling of cloned *Dof* gene (EU760631) of finger millet in different tissues like root, stem, flag leaf (during vegetative growth) and developing spikes (S1, S2, S3, and S4 stages) showed its expression in all tissues. However, expression of PBF Dof is relatively higher in developing stages of spikes than in other tissues in all genotypes **(Gupta *et al.*, 2011)**.

The availability of genetic map of *E. coracana* **(Dida *et al.*, 2007)** could be used to investigate the diversity of *Dof* genes and their 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 exhibited partial sequence with some of amino acid residues at N-terminal associated with classical Dof domain were missing. The phylogenetic tree constructed further revealed the greater degree of similarity of PBF *Dof* genes of monocots

by forming separate cluster and PBF *Dof* genes from members of dicots also existed together in distinct cluster.

4.3 Genome wide identification and *in silico* characterization of *Dof* gene family of sorghum

Sorghum with a relatively small genome (730 Mbp) distributed among 10 chromosomes **(Kim *et al.*, 2005; Paterson *et al.*, 2009)** has long been an attractive model for advancing understanding of the structure, function, and evolution of cereal genomics. The recent release of whole genome shotgun sequence of sorghum has led to the prediction of whole set of *Dof* gene family as reported in case of rice and Arabidopsis using comparative genomics approach. In the present study, the *in silico* prediction followed by characterization for chromosomal location, gene structure prediction, phylogenetic analysis, motif analysis and *cis*-regulatory element analysis has been attempted.

4.3.1 *In silico* prediction of *Dof* gene family of sorghum

A total of 28 *Dof* genes were annotated from the recently released whole genome sequence of sorghum **(Paterson *et al.*, 2009)** by performing BLAST search using nucleotide and amino acid sequences of conserved *Dof* domain of rice. The predicted sequences of these 28 *Dof* genes of sorghum were designated as *SbDof1* to *SbDof28*. The predicted *SbDof* genes sequences were submitted to Third Party Annotation (TPA) section of the GenBank database and were assigned the accession numbers TPA:BK006983 to BK007006 and TPA:BK007079 to BK007082 (**Table 4.7**).

Table 4.7 *In silico* predicted *Dof* genes from whole genome sequence of *S. bicolor* (L) Moench

Annotated gene	Source	Assigned accession number	Chromosome number	Gene structure
SbDof1	ABXC01005623	TPA:BK007079	8	Intronless
SbDof2	ABXC01000338	TPA:BK007080	1	One intron
SbDof3	ABXC01000645	TPA:BK006983	1	Intronless
SbDof4	ABXC01000029	TPA:BK006984	1	Intronless
SbDof5	ABXC01000170	TPA:BK006985	1	Intronless
SbDof6	ABXC01001528	TPA:BK006986	2	Intronless
SbDof7	ABXC01001444	TPA:BK006987	2	Intronless
SbDof8	ABXC01001648	TPA:BK006988	2	Intronless
SbDof9	ABXC01000156	TPA:BK006989	1	Intronless
SbDof10	ABXC01000069	TPA:BK006990	1	Intronless
SbDof11	ABXC01002298	TPA:BK006991	3	Intronless
SbDof12	ABXC01002251	TPA:BK006992	3	Intronless

SbDof13	ABXC01002199	TPA:BK006993	3	Intronless
SbDof14	ABXC01001839	TPA:BK006994	3	One intron
SbDof15	ABXC01001828	TPA:BK006995	3	Two Introns
SbDof16	ABXC01002936	TPA:BK006996	4	Intronless
SbDof17	ABXC01002910	TPA:BK006997	4	One intron
SbDof18	ABXC01002897	TPA:BK006998	4	Intronless
SbDof19	ABXC01003020	TPA:BK006999	5	Intronless
SbDof20	ABXC01004404	TPA:BK007000	6	Intronless
SbDof21	ABXC01004473	TPA:BK007001	6	One intron
SbDof22	ABXC01005057	TPA:BK007002	7	Intronless
SbDof23	ABXC01000714	TPA:BK007003	1	One intron
SbDof24	ABXC01005079	TPA:BK007004	8	One intron
SbDof25	ABXC01006212	TPA:BK007005	9	Intronless
SbDof26	ABXC01005734	TPA:BK007006	9	Intronless
SbDof27	ABXC01001715	TPA:BK007081	3	Intronless
SbDof28	ABXC01005582	TPA:BK007082	8	Intronless

The presence of the conserved Dof domain in the predicted SbDof protein was a typical feature for consideration of a protein as a member of *Dof* family of transcription factor. Protein functional analysis of 28 putative SbDof proteins of *S. bicolor* using InterProScan has shown identity with InterPro accession number IPR003851. Those 28 SbDof proteins also showed similarity with signature accessions, namely PD007478 (ProDom database), PF02701 (Pfam database), PS01361 (PROSITE patterns database) and PS50884 (PROSITE profiles database) confirming their identity to Dof like proteins. The deduced protein sequences of *SbDof* genes were then subjected to motif scan database. Motif scan integrated with PeroxiBase profiles, PROSITE patterns, PROSITE profiles, HAMAP profiles, Pfam HMMs (local models), Pfam HMMs (global models) and PeroxiBase profile databases revealed the presence of glycine, threonine, alanine, serine, proline, histidine, glutamine, methionine and asparagine rich regions in different *SbDof* genes. A typical zinc finger Dof type profile was also observed in all *SbDof* genes.

The SbDof4 subjected to InterProScan for protein functional analysis revealed its identity with InterPro accession number IPR001412 aminoacyl-tRNA synthetase class I conserved site and PS00178 (PROSITE profiles database) indicating an additional function of t-RNA aminoacylation associated with protein translation. Only three of the SbDof proteins namely, SbDof1, SbDof7 and SbDof21 showed nuclear localization signal **(**NLS) when subjected to NUCPRED software. The NLS signal for SbDof1 was found to be *RKRPRPR* while SbDof7 and SbDof21 showed KRRRV as NLS signal.

4.3.2 Chromosomal locations

The predicted 28 *SbDof* genes were analyzed for its distribution on chromosomes of sorghum. Chromosome 1 and 3 have a maximum of 7 and 6 *Dof* genes respectively, while chromosomes number 2, 4 and 8 have 3 *Dof* genes each (**Table 4.7**). Chromosome 10 does not contain any *Dof* gene. The distribution of *Dof* genes among the 10 chromosomes of sorghum is shown in **Fig. 4.21**.

The distribution of *Dof* genes in rice and Arabidopsis chromosomes has been studied **(Lijavetzky *et al.*, 2003)**. In case of rice, 30 *Dof* genes were found to be organized on eleven out of the twelve chromosomes with maximum number of 6 *Dof* gene distributed on both chromosomes 1 and 3, while 36 *Dof* genes of Arabidopsis were distributed among all the five chromosomes with maximum 9 *Dof* genes on chromosome 1. The detailed distribution of *Dof* genes on the rice and Arabidopsis chromosomes is shown in **Table 4.8** and **4.9**, respectively.

Table 4.8 List of *Dof* genes of *O. sativa* (Source: **http://drtf.cbi.pku.edu.cn**)

Gene name	Locus	Gene model	Chromosome number	Gene structure
OsDof1	Os07g48570	Os07g48570.1	7	One intron
OsDof2	Os01g55340	Os01g55340.1	1	Intronless
OsDof3	Os01g15900	Os01g15900.1	1	One intron
OsDof4	Os01g17000	Os01g17000.1	1	One intron
OsDof5	Os03g16850	Os03g16850.1	3	One intron
OsDof6	Os03g07360	Os03g07360.1	3	One intron
OsDof7	Os10g26620	Os10g26620.1	10	One intron
OsDof8	Os05g36900	Os05g36900.1	5	Intronless
OsDof9	Os01g64590	Os01g64590.1	1	Intronless
OsDof10	Os07g32510	Os07g32510.1	7	Intronless
OsDof11	Os02g47810	Os02g47810.1	2	One intron
OsDof12	Os02g49440	Os02g49440.1	2	Intronless
OsDof13	Os03g38870	Os03g38870.1	3	Intronless
OsDof14	Os06g17410	Os06g17410.1	6	Intronless
OsDof15	Os01g09720	Os01g09720.1	1	Intronless
OsDof16	Os04g47990	Os04g47990.1 Os04g47990.2	4	One intron Intronless
OsDof17	Os02g45200	Os02g45200.1 Os02g45200.2 Os02g45200.3 Os02g45200.4	2	One intron Two introns Intronless One intron
OsDof18 (MNB1A)	Os08g38220	Os08g38220.1	8	Intronless
OsDof19	Os03g42200	Os03g42200.1	3	Two introns
OsDof20	Os01g48290	Os01g48290.1	1	Intronless
OsDof21	Os12g38200	Os12g38200.1	12	One intron
OsDof22	Os07g13260	Os07g13260.1	7	Intronless
OsDof23	Os02g15350	Os02g15350.1	2	One intron

(RPBF)				
OsDof24	Os05g02150	Os05g02150.1	5	One intron
OsDof25	Os04g58190	Os04g58190.1 Os04g58190.2 Os04g58190.3	4	One intron Two introns Two introns
OsDof26	Os10g35300	Os10g35300.1	10	One intron
OsDof27 (MNB1A)	Os12g39990	Os12g39990.1	12	Intronless
OsDof28	Os03g55610	Os03g55610.1	3	Intronless
OsDof29	Os03g60630	Os03g60630.1	3	Intronless
OsDof30 (MNB1A)	Os09g29960	Os09g29960.1	9	Intronless

Table 4.9 List of *Dof* genes of *A. thaliana* (Source: **http://datf.cbi.pku.cn**)

Gene name	Locus	Gene model	Chromosome number	Gene structure
AtDof1 (OBP2)	AT1G07640	AT1G07640.1 AT1G07640.2 AT1G07640.3	1	Intronless One intron One intron
AtDof2	AT1G21340	AT1G21340.1	1	Intronless
AtDof3	AT1G26790	AT1G26790.1	1	Two intron
AtDof4	AT1G28310	AT1G28310.1 AT1G28310.2	1	Intronless One intron
AtDof5 (COG1)	AT1G29160	AT1G29160.1	1	Intronless
AtDof6	AT1G47655	AT1G47655.1	1	Intronless
AtDof7 (ADOF1)	AT1G51700	AT1G51700.1	1	Intronless
AtDof8	AT1G64620	AT1G64620.1	1	One intron
AtDof9	AT1G69570	AT1G69570.1	1	One intron
AtDof10	AT2G28510	AT2G28510.1	2	One intron
AtDof11	AT2G28810	AT2G28810.1	2	One intron
AtDof12	AT2G34140	AT2G34140.1	2	Intronless
AtDof13	AT2G37590	AT2G37590.1	2	One intron
AtDof14 (DAG2)	AT2G46590	AT2G46590.1 AT2G46590.2	2	Intronless One intron
AtDof15 (ADOF2)	AT3G21270	AT3G21270.1	3	Intronless
AtDof16	AT3G45610	AT3G45610.1	3	One intron
AtDof17 (CDF3)	AT3G47500	AT3G47500.1	3	One intron
AtDof18 (OBP1)	AT3G50410	AT3G50410.1	3	Intronless
AtDof19	AT3G52440	AT3G52440.1	3	Intronless
AtDof20 (OBP3)	AT3G55370	AT3G55370.1 AT3G55370.2	3	Two introns One intron
AtDof21 (DAG1)	AT3G61850	AT3G61850.1 AT3G61850.2	3	Two introns One intron

AtDof22	AT4G00940	AT4G00940.1	4	Intronless
AtDof23	AT4G21030	AT4G21030.1	4	Intronless
AtDof24	AT4G21040	AT4G21040.1	4	Intronless
AtDof25	AT4G21050	AT4G21050.1	4	Intronless
AtDof26	AT4G21080	AT4G21080.1	4	Intronless
AtDof27	AT4G24060	AT4G24060.1	4	One intron
AtDof28	AT4G38000	AT4G38000.1	4	Intronless
AtDof29	AT5G02460	AT5G02460.1	5	One intron
AtDof30 (CDF2)	AT5G39660	AT5G39660.1 AT5G39660.2	5	One intron One intron
AtDof31	AT5G60200	AT5G60200.1	5	One intron
AtDof32 (OBP4)	AT5G60850	AT5G60850.1	5	Intronless
AtDof33 (CDF1)	AT5G62430	AT5G62430.1	5	Intronless
AtDof34	AT5G62940	AT5G62940.1	5	One intron
AtDof35	AT5G65590	AT5G65590.1	5	Intronless
AtDof36	AT5G66940	AT5G66940.1	5	Intronless

4.3.3 Phylogenetic relationships among Dof proteins of sorghum

The complete catalog of Dof protein in a single plant species is useful for viewing the existing structural and functional diversity associated with their diverse roles in plants. The evolutionary relationship between different Dof proteins were analyzed by subjecting the deduced amino acid sequences of the identified 28 *SbDof* genes for multiple sequence alignment. Alignment of Dof domain sequence of different SbDof proteins revealed a well conserved four cysteine residues putatively involved in the formation of the zinc finger structure (**Fig. 4.22**)

The Dof domain of sorghum revealed highly conserved sequences with 25 out of 52 amino acids being 100% conserved in all the 28 SbDof proteins while 9 amino acids showed variation in only two amino acid residues. The phylogenetic tree constructed using ClustalX2.0.10 software with bootstrapping revealed the existence of two major groups A and B. Group A was further divided in two subgroups A1 and A2. Group B was similarly divided into two subgroups B1 and B2 which were further arranged into two branches as shown in **Fig. 4.23**. Thus Dof family in sorghum exists in four subgroups comprising of six clusters. A very high bootstrap value suggested common origin for *Dof* genes of each sub-group.

4.3.4 Gene structure prediction

The putative gene structure of the predicted *SbDof* gene family in terms of intron/exon distribution pattern is shown in **Fig. 4.24**. A total of 21 out of 28 *SbDof* genes were found to be intronless. The *SbDof15* seems to have complex gene structure with the presence of two introns, while *Dof* genes with at least one intron are also observed in sorghum. Most members of *SbDof* genes with more than one exon represent sub groups A1 in the phylogenetic tree (**Fig. 4.23** and **4.24**). A total of 17 out of 36 *Dof* gene models in rice (**Table 4.8**) and 21 out of 43 *Dof* gene models in Arabidopsis (**Table 4.9**) were found to be intronless. There exists great similarity among *Dof* genes of rice, Arabidopsis and sorghum in terms of gene structure with a majority of *Dof* genes either lacking intron or having single intron **(Lijavetzky *et al.*, 2003)**. The analysis of intron/exon gene structures revealed that most introns have conserved positions and phases, providing the evidence for the intron-early theory, and that multiple independent intron loss events likely have occurred during the evolution of flowering plants. The hypothesis that genome-wide and tandem duplication contributed to the expansion of the *Dof* gene family across the plant kingdom seems to be applicable for sorghum too.

4.3.5 Evolutionary relationships of sorghum, rice and Arabidopsis with respect to *Dof* gene family

The predicted 28 SbDof proteins were subjected to multiple sequence alignment along with 36 Arabidopsis and 30 rice Dof proteins and a phylogenetic tree was constructed using software Clustalx2.0.10 with neighbor joining method and bootstrap analysis (1000 reiterations) (**Fig. 4.25**) A total of six major groups of Dof proteins designated as A, B, C, D, E and F were observed for sorghum similar to what has been reported earlier for Dof proteins of wheat, rice and Arabidopsis **(Shaw *et al.*, 2009)**.

The SbDof8, SbDof14 and SbDof15 of sorghum showed maximum similarity with AtDof17, AtDof30 and AtDof33 of Arabidopsis representing major group-A, which are basically CDF proteins associated with regulation of photoperiodic control of flowering **(Fornara *et al.*, 2009; Imaizumi *et al.*, 2005)**. The Arabidopsis Dof proteins namely AtDof23, AtDof24, AtDof25 and AtDof26 constitute a distinct major group-B similar to what has been reported earlier in Arabidopsis cluster C3 in MCOG Cc **(Lijavetzky *et al.*, 2003)**. These sets of *Dof* genes might be exclusively present in Arabidopsis as no apparent counterpart has been observed in rice or sorghum. In the major group-C, the SbDof22, SbDof7 and SbDof1 showed similarity with rice OsDof18, OsDof30 and OsDof27 which are

basically MNb1a protein. In group-D, SbDof21 showed maximum similarity with AtDof18 which is an Arabidopsis OBP1 protein involved in defense response **(Zhang *et al.*, 1995)**. The other members of group-D i.e. SbDof24 and SbDof19 showed similarity with rice OsDof23 (RPBF) which is associated with the regulation of seed storage proteins **(Washio, 2001; Washio, 2003; Yamamoto *et al.*, 2006)**. In group-F the SbDof16 and SbDof20 form cluster with AtDof14 and AtDof21, an Arabidopsis DAG proteins (Dof affecting germination proteins) indicating that these proteins might be involved in seed germination **(Gualberti *et al.*, 2002; Papi *et al.*, 2000)**.

The comparative phylogenetic analysis of sorghum *Dof* genes family clearly indicated their proximity with rice than Arabidopsis owing to the fact that sorghum and rice are members of monocot while Arabidopsis is a dicot. Further the presence of some similar groups and sub groups in comparative phylogeny of sorghum, rice and Arabidopsis revealed the conserved nature of some of the *Dof* genes during angiosperm evolution.The evolutionary relationship among the rice *Dof* genes family has been attempted based on their DNA binding domain sequences which revealed four major groups two of which were further divided in two subgroups resulting into a total of seven subgroups **(Lijavetzky *et al.*, 2003)**. Similarly phylogenetic analysis of a total of 116 genes in different organisms from green unicellular algae to vascular plants including 30 rice *Dof* genes also revealed seven subfamilies **(Moreno-Risueno *et al.*, 2007b)**. Thus increase in the number of subgroups of rice *Dof* genes as compared to sorghum might be due to the larger number of *Dof* genes in rice.

4.3.6 Motif analysis

The 28 Dof proteins of sorghum along with the 66 Dof proteins of Arabidopsis and rice were analyzed for the presence of conserved motifs by means of MEME software. A total of 50 conserved motifs were observed in all the 94 Dof proteins (**Fig. 4.26**). The motif-1 uniformly observed in all the Dof proteins except AtDof33 of Arabidopsis represented the conserved Dof domain of 50 amino acids. The lack of motif-1 in AtDof33 might be due to absence of some sequences at the N-terminus of Dof domain. There were number of common motifs observed for sorghum, rice and Arabidopsis. A total of 12 motifs designated as motif-15, 23, 25, 26, 27, 30, 33, 35, 39, 40, 43 and 47 were common for sorghum and rice, while motif-31 was observed common for rice and Arabidopsis. Motifs exclusively present in Arabidopsis are motif-14, 20, 22, 34, 38, 45, 46 and 50. In sorghum motif-24, 28 and 48 are unique as these were not observed in rice or Arabidopsis. The presence of unique motifs in a particular crop might be associated with specific functions.

Based on the groups indicated from the phylogenetic tree of Dof proteins and motifs through MEME analysis, a schematic distribution of conserved motifs among the defined gene groups is shown in **Fig. 4.26**. Each sub group has similar number and arrangement of motifs. Some motifs were located in the proteins of specific subgroups, for instance motif-2, 3, 4, 13, 19, 26, 38, 39, 41 and 47 for group-A, motif-14, 22 and 34 for group-B, motif-27 and 40 for group-C, motif-24 and 28 for group-D, motif-8 and 31 for group-E and motif-17, 35, 15, 50, 16, 11 and 23 for group-F. These similarities in motif patterns might be related to similar functions of each subgroup. The multilevel consensus sequence for the identified motifs is provided in **Table 4.10**.

Table 4.10 Multilevel consensus sequences for the MEME defined motifs observed among different Dof proteins from sorghum, rice and Arabidopsis

Motif	No. of aa	E value	Multilevel consensus sequence
1	50	4.8e-4482	CPRCDSTNTKFCYYNNYNLSQPRHFCKTCRRYWTKGG ALRNVPVGGGCRK
2	38	4.3e-253	EDEKREKKVWVPKTLRIDDPDEAAKSSIWTTLGIKPDD
3	25	1.5e-102	IIETLPVLQANPAALSRSQTFQETT
4	28	1.1e-095	DEKSEEEKTEGDESEQEKKLKKPDKILP
5	20	1.1e-092	MSMAERARLAKIPLPEPGLK
6	20	1.1e-085	DDPGIKLFGKTIPVPEDIEC
7	20	2.0e-038	ISIWCSDSNSFTLGKHSRDE
8	26	1.2e-033	SSIESLSFINQDLHWKLQQQRLATMF
9	20	1.9e-033	QQAAGGTERRARPQKEKALN
10	20	5.2e-032	AIGLEQWRVQQQQQQQPQQQ
11	23	1.2e-031	KNPKLMHEGAQDLNLAFPHHGGI
12	20	4.2e-031	NKRSKSSAAAAAAAAASAAA
13	28	5.8e-028	NETVLKFGPDVPLCESMASVLNIKEQNI
14	28	7.3e-027	HRLDFHDESFEQDYYDVGSDDLIVNQQI
15	47	4.2e-026	IVPPLTWWEDTKYDPFDSFPDDAMSLHDIMIGDEDHW WSVDCCQLEE
16	20	1.6e-025	DDEAAGRLLFPFEDLKPPVS
17	20	2.6e-024	DGSTIDLALLYAKFLNHHPD
18	28	8.9e-039	YIPIPPAVWWYWPCWPPGAWNAPWIRQN
19	39	2.9e-024	VDKEIWPYHGNCVMHPIPYYPGPPYMYPWNPAWNIVP VM
20	39	1.4e-023	KRPKIDQPSVAQMVSVEIQPGNHQPFKNVQENIDFVGSF
21	20	6.1e-023	KVSIVSGLITQLASVKMEDH
22	20	6.6e-018	MDNLNVFANEDNQVNDVKPP
23	23	1.2e-017	VGALSAMELLRSTGCYMPLQQPM
24	49	2.7e-015	KMSINTQPMMVPNMMMPTPTMTGLFPNVLPTLMSTGE GGEFNFTMDNQH
25	20	1.8e-013	QPPEFPAFPSLESSSVCNPG
26	46	8.1e-011	KDDEIKVDVPQEEEDNEMKVDAPQEEKDDEMKIDVQE EKKDEEMEV
27	23	2.8e-010	FEWPSGCDLGTYWPTAVFADTDP

28	48	1.0e-009	TMPSFLEMLRKGLLHGSSSYDTGLVMSDGNNEMDMSF PLPAYGAMHGH
29	25	1.5e-009	YHMNQVDQFKWNQSFNNTMNMNYNN
30	20	1.3e-008	HRFPGPVRPDMILEGMVGNP
31	20	1.7e-015	MVFSSIPIYLDPPNWWQQQH
32	20	1.7e-008	MWTHWWQQLIVVKPMEEIIT
33	28	2.2e-008	YADQAFALGEFFAPPPPPPHPVLTFALD
34	27	5.1e-008	VTAVGNHFGYLSEIHGIMVTNPIPTFT
35	49	3.2e-006	TTATTTMLCTDVSVQAAFGELNFAMDQSCFDSLGLPTD DVVGGLLSSWC
36	34	1.0e-005	HHHHHDHHKEQDGGGGVIGGHETPGFWNGMIIGN
37	20	2.9e-008	SGWPWEDLSGFNSSSSGNVL
38	22	1.8e-005	HGGFRHDFPMKRLRCYTDGQSC
39	27	6.7e-005	LHYRQLLMAPDCMMGSRVEISKSMNPE
40	45	1.4e-004	HMDSQLGMGPLGQHDVLSSLGLKLPPPASSPPAASYYS DQLHAVV
41	27	1.7e-004	TDPKDQENTVQDSTDPPPQPEVVDTED
42	28	9.9e-006	MQEFQKIPGLAGRLFGGAAAADIVRAQE
43	39	2.9e-003	QLPFLASLHHPLGGGDHYSTGASRLGFPGLSSLDPVDY Q
44	20	7.0e-005	NSSPREFLGLPGNLQFWGGG
45	20	2.0e-005	NEQKQEMDPTRVLWGFPWQM
46	20	1.9e-003	HGHVDQIDSGREIWTNMVYI
47	38	2.5e-003	NQKEKVTAEKSPKIVQHPCMNGVAMWPFGCAPPPACY T
48	33	8.9e-004	TTDDARQLVGTQQGMNTDGGFVGSTRVQEEEEE
49	28	4.5e-003	YWGNGGIGGAAAAWPDLANCGSSIATLF
50	20	1.1e-002	DLNLLSFPVMQDQHHHHIEM

4.3.7 *Cis*-regulatory elements analysis

The diverse functions attributed to *Dof* genes are due to interaction of these transcription factors with different conserved sequences in the promoter regions of the respective genes. The *cis*-regulatory element analysis was carried out by retrieving 500 bp upstream sequences from the initiation codon of only 26 putative *SbDof* genes. *SbDof1* gene could not be analyzed for *cis*-regulatory elements as the 500 bp upstream sequences could not be retrieved due to the unavailability of sequences after 124 bases upstream from initiation codon. The *SbDof16* was also not studied in detail due to lack of important *cis*-regulatory elements. A large number of *cis*-regulatory elements were found when the 26 *SbDof* genes were subjected to PlantCARE database. The *cis*-regulatory elements related with five important physiological phenomena i.e. light responsiveness, endosperm-specific gene expression, hormone responsiveness, meristem-specific expression and stress responsiveness were found to be distributed among these *Dof* genes as shown in **Table 4.11**.

Table 4.11 Identified *cis*-regulatory elements of *SbDof* genes of sorghum. The positions are shown in parenthesis

Annotated gene	Light responsive elements	Elements for endosperm-specific gene expression	Hormone responsive elements	Elements for meristem-specific expression	Stress responsive elements
SbDof2	I-box (-382) Sp1 (-248)	Nil	Nil	Nil	ARE (- 194) TCA-element (-11)
SbDof 3	TCCC-motif (-428)	O2-site (-210)	ABRE (-388)	Nil	Nil
SbDof 4	Gap-box (-341) I-box (-447) MNF1 (-204) box II (-123)	Nil	Nil	Nil	MBS (-366)
SbDof 5	ATC-motif (-449) Box 4 (-417) G-Box (-178) MNF1 (-217) Pc-CMA2c (-381) Sp1(-341)	RY-element (-409)	Nil	CCGTCC-box (-238)	CGTCA-motif (-181) TGACG-motif (-427)
SbDof 6	G-box (-56) Sp1(-200) TCCC-motif (-193)	Nil	Nil	OCT (-134)	GC-motif (-272)
SbDof 7	MNF1 (-148) Sp1(-273)	Nil	Nil	CAT-box (-237)	MBS (-332) TGACG-motif (-232)
SbDof 8	ACE (-61) TCT-motif (-89)	Nil	Nil	Nil	Box-W1 (-125) MBS (-451)
SbDof 9	G-box (-473) GATA-motif (-481)	RY-element (-425)	ABRE (-473)	CAT-box (-46)	GC-motif (-121) ARE (-309) MBS (-361) CGTCA-motif (-68)
SbDof 10	G-box (-245) I-box (- 84) Sp1(-131)	Nil	Nil	Nil	ARE (-398) GC-motif (-463)
SbDof 11	G-Box (-446) Sp1(-66) TCCC-motif (-174)	Nil	ABRE (-19)	Nil	GC-motif (-439) CGTCA-motif (-294)
SbDof 12	G-box (-8) Sp1 (-337)	GCN4_motif (-129)	P-box (-322) TGA-element (-164)	CCGTCC-box (-220)	TGACG-motif (-31)
SbDof 13	Sp1 (-45)	Nil	AuxRR-core (-326)	Nil	Nil
SbDof 14	Sp1(-304) TCCC-motif (-132) rbcS-CMA7a (-259)	Nil	Nil	CCGTCC-box (-94)	MBS (-385) GC-motif (-233) TCA-element (-269)
SbDof 15	ATCC-motif (-213) GT1-motif (-386)	Nil	Nil	Nil	Nil
SbDof 17	Sp1 (-490)	RY-element (-141) Skn-1_motif (-359)	Nil	Nil	Nil
SbDof 18	G-Box (-42) Sp1 (-120)	Nil	Nil	Nil	Nil
SbDof 19	ATCT-motif (-387) G-Box (-226)	GCN4_motif (-422) Skn1_motif (-419)	TGA-element (-283)	Nil	ARE (-431)
SbDof 20	Pc-CMA2c (-155) Sp1 (-194)	O2-site (-61)	Nil	Nil	TCA-element (-346)

SbDof 21	CATT-motif (-235) Sp1(-392)	Nil	TGA-element (-358)	CAT-box (-242) CCGTCC-box (- 35)	CGTCA-motif (-462) TGACG-motif (-286)
SbDof 22	Box I (-165) G-box (-69) Sp1 (-77)	Nil	ABRE (-30) TGA-box (-364)	CCGTCC-box (-176)	CCAAT-box (-5) C-repeat/DRE (-324) CGTCA-motif (-67) TGACG-motif (-364)
SbDof 23	Box II (-310) G-Box (-309) Sp1 (-293) box II (-311)	Nil	ABRE (-308)	CCGTCC-box (-250)	GC-motif (-290) CGTCA-motif (-462)
SbDof 24	Box 4 (-64) TCT-motif (-52)	GCN4_motif (-305) Skn-1_motif (-302)	Nil	Nil	ARE (-164)
SbDof 25	G-Box (-303) G-box (-199) L-box (-306)	Skn-1_motif (-342)	ABRE (-363) P-box (-225)	CCGTCC-box (-24)	CGTCA-motif (-343)
SbDof 26	Sp1 (-205)	RY-element (-177)	Nil	Nil	MBS (-407)
SbDof 27	G-box (-341) Sp1 (-115) TGG-motif (-395)	Nil	Nil	Nil	MBS (-77)
SbDof 28	Box 4 (-32)	Nil	Nil	Nil	MBS (-472) ARE (-256)

The presence of many light responsive elements in the promoter region of *SbDof* genes strongly suggests the involvement of these *Dof* genes in the regulation of photoperiodic control of flowering. The frequency of light responsive elements in *Dof* gene promoter region ranges from 1 to 6. The maximum number of light responsive elements i.e. 6 was found in the promoter region of *SbDof5* gene. The presence of *cis*-regulatory elements like Skn-1_motif, GCN4_motif, O2-site and RY element confers endosperm-specific gene expression **(Mena *et al.*, 1998; Vicente-Carbajosa *et al.*, 1997)**. The *SbDof19* and *SbDof24* contains GCN4_motif and Skn-1_motif and formed cluster with *OsDof23* (*RPBF*) associated with the regulation of seed storage proteins **(Yamamoto *et al.*, 2006)**. Similarly *SbDof3*, *SbDof5*, *SbDof9*, *SbDof12*, *SbDof17*, *SbDof19*, *SbDof17*, *SbDof24*, *SbDof25* and *SbDof26* also revealed endosperm specific gene expression. The majority of the *SbDof* genes when subjected to *cis*-regulatory element analysis indicated their putative role in abiotic and biotic stress owing to the presence of stress responsive elements except in *SbDof3*, *SbDof13*, *SbDof15*, *SbDof17* and *SbDof18*. The involvement of *SbDof* gene in regulation of meristem-specific and hormone-specific expression has also been observed. The list of identified *cis*-regulatory elements among the 26 *Dof* genes is shown in **Table 4.11** and the putative functions of these elements are represented in **Table 4.12**. The diverse *cis*-regulatory elements present in *Dof* gene family might contribute to their multifarious roles.

Table 4.12 Functions of identified *cis*-regulatory elements

S. No.	*Cis*-regulatory elements	Functions/description
1	ARE	*Cis*-acting regulatory element essential for the anaerobic induction
2	Box-W1	Fungal elicitor responsive element
3	CCGTCC-box	*Cis*-acting regulatory element related to meristem-specific activation
4	GC-motif	Enhancer-like element involved in anoxic-specific inducibility
5	GCN4_motif	*Cis*-regulatory element involved in endosperm-expression
6	MBS	MYB Binding Site
7	MBSII	MYB binding site involved in flavonoid biosynthetic genes regulation
8	Skn-1_motif	*Cis*-acting regulatory element required for endosperm expression
9	ABRE	*Cis*-acting element involved in the abscisic acid responsiveness
10	CGTCA-motif	*Cis*-acting regulatory element involved in the MeJA-responsiveness
11	TGACG-motif	*Cis*-acting regulatory element involved in the MeJA-responsiveness
12	GARE-motif	Gibberellin-responsive element
13	Sp1	Light responsive element
14	Box 4	Part of a conserved DNA module involved in light responsiveness
15	I-box	Light responsive element
16	O2-site	*Cis*-acting regulatory element involved in zein metabolism regulation
17	MNF1	Light responsive element
18	ATC-motif	Part of a conserved DNA module involved in light responsiveness
19	G-Box	*Cis*-acting regulatory element involved in light responsiveness
20	Pc-CMA2c	Part of a light responsive element
21	RY-element	*Cis*-acting regulatory element involved in seed-specific regulation
22	OCT	*Cis*-acting regulatory element related to meristem-specific activation
23	CAT-box	*Cis*-acting regulatory element related to meristem-expression
24	ACE	*Cis*-acting element involved in light responsiveness
25	TCT-motif	Part of a light responsive element
26	GATA-motif	Part of a light responsive element
27	TCCC-motif	Part of a light responsive element
28	P-box	Gibberellin-responsive element
29	TGA-element	Auxin-responsive element
30	AuxRR-core	*Cis*-acting regulatory element involved in auxin responsiveness

31	TCA-element	*Cis*-acting element involved in salicylic acid responsiveness
32	rbcS-CMA7a	Part of a light responsive element
33	ATCC-motif	Part of a conserved DNA module involved in light responsiveness
34	GT1-motif	Light responsive element
35	ATCT-motif	Part of a conserved DNA module involved in light responsiveness
36	Pc-CMA2c	Part of a light responsive element
37	Box I	Light responsive element
38	C-repeat/DRE	Regulatory element involved in cold and dehydration responsiveness
39	CCAAT-box	MYBHv1 binding site
40	Box II	Part of a light responsive element
41	L-box	Part of a light responsive element
42	TGG-motif	Part of a light responsive element

4.4 Secondary and tertiary structure prediction of SbDof proteins of sorghum

4.4.1 Secondary structural prediction

The sequences of cloned *SbDof* genes from sorghum having accession numbers HQ540084 (SbDof1), HQ540085 (SbDof19), HQ540086 (SbDof23) and HQ540087 (SbDof24) were considered for structural classification and prediction of secondary and tertiary structure. Molecular weights and isoelectric points of these four SbDof proteins were analyzed by PROTEAN module of DNASTAR as shown in **Table 4.13**.

Table 4.13 Physical properties of translated proteins of cloned *SbDof* genes

Cloned gene name	Gene size (bp)	Gene structure	Protein length (Amino acid)	Molecular weight (kDa)	Isoelectric point (pI)
SbDof1	937	Intronless	275	28.03	8.32
SbDof19	966	Intronless	321	34.23	9.08
SbDof23	2226	One intron	424	45.17	8.32
SbDof24	1043	One intron	333	35.88	9.23

The deduced protein sequences of these *SbDof* genes were subjected to motif scan search. Motif scan was integrated with PeroxiBase profiles, PROSITE patterns, PROSITE profiles, HAMAP profiles, Pfam HMMs (local models) and Pfam HMMs (global models) databases. The results revealed the presence of glycine and alanine rich profile for SbDof1 while proline rich region profile for SbDof23. Asparagine, methionine and serine rich profile were observed for SbDof19 and SbDof24 protein. These two proteins also revealed high sequence conservation and could be a result of recent duplication event in sorghum. A typical

zinc finger Dof type profile was also observed for these four Dof proteins as shown in **Fig. 4.27**.

The two-dimensional structure predictions of putative SbDof proteins were carried out by PDBsum server (**http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html**). The results revealed the presence of helices, sheets, turns, coils and hairpin loops. Dof domain region has turns as dominant feature though sheets, coils and helices were also observed. In case of SbDof1 protein alanine rich (107-119 aa) and glycine rich (142-216 aa) regions were observed in motif scan output (**Fig. 4.27a**). The alanine rich region mainly consisted of sheet and β-turn while glycine rich region consisted of β-turn, sheet and helix. The SbDof19 protein revealed serine rich (97-128 aa), methionine rich (127-146 aa) and asparagines rich (295-316 aa) regions in motif scan analysis (**Fig. 4.27b**). Serine rich region consisted of helix and β-turn, methionine rich region consisted of helix, β-turn and sheet while asparagine rich region consist of β-turn and γ-turn. The SbDof23 protein contained proline rich region (259-303 aa) with β-turn, γ-turn, sheet and β-hairpin (**Fig. 4.27c**). The SbDof24 protein showed serine rich (95-125 aa), methionine rich (124-143 aa) and asparagine rich (294-328 aa) regions in motif scan analysis (**Fig. 4.27d**). Serine rich region consisted of β-turn, γ-turn, sheet and β-hairpin, methionine rich region consisted of β-turn, sheet and β-hairpin while asparagine rich region consisted of β-turn, helix, sheet and β-hairpin. The overall predicted secondary structures of four SbDof proteins are shown in **Fig. 4.28**.

4.4.2 Tertiary structural prediction

Three-dimensional structures provide valuable insight into the identification of molecular functions and putative active site residues. To our information three-dimensional structure of Dof proteins has not been predicted so far as is evident from no entries of Dof proteins in Protein Data Bank (PDB). An attempt was made to predict 3D structures by multiple threading from online available I-TASSER server. 3D models for these 4 SbDof proteins were successfully predicted using different threading templates as given in **Table 4.14**.

Table 4.14 List of top ten templates used by I-TASSER for 3D structure prediction of sorghum Dof proteins

Protein name	PDB IDs
SbDof1	1zlgA,1m11R, 1ticB, 1pqvA, 1w0rA, 1tiaA, 1k82A, 2eyzA, 1pdiR, 1tfiA
SbDof19	1tiaA, 1w0rA, 1tiaA, 3h0gA, 2xd8A, 1ticB, 1twfA, 1m11R, 1pdiR, 2epsA
SbDof23	3chnS,3h0gA,1worA, 1w0sA, 3h0gA, 3gavA, 3h0gA, 1n7dA, 1w0rA, 3h0gA
SbDof24	1w0rA, 1tiaA, 3h0gA, 2xd8A, 2fl9R, 3po3A, 3k6sA, 1ticB, 1twfA, 1m11R

The final results of function prediction are deduced from the consensus of top structural matches with the function scores calculated based on the confidence score (C-score) of the I-TASSER structural models. C-score estimates quality of predicted models and it is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. The structural similarity between model and templates were evaluated by template modeling score (TM-score) and the sequence identity in the structurally aligned regions were determined **(Zhang, 2008)**.

A total of 5 structure models were predicted for each SbDof proteins by I-TASSER server. Only most appropriate predicted structures were chosen for each proteins based on maximum C-score and maximum number of decoys for evaluation and verifications. The selected models have expected TM score ranging from 0.32±0.11 to 0.49±0.15 and RMSD (root mean square deviation) score lying between 11.3±4.5 Å to 15.3±3.4 Å for the these four SbDof proteins from I-TASSER server (**Table 4.15)**. The selected models were found to be in correct topology based on C-score, TM-score and RMSD value.

Table 4.15 Model evaluation data for the predicted structures of the four Dof proteins of sorghum

Protein name	C-score	Exp. TM-score	Exp. RMSD	No. of decoys	Cluster density
SbDof1	-3.01	0.37±0.13	13.2±4.1Å	3000	0.0507
SbDof19	-3.40	0.34±0.11	14.7±3.6 Å	1819	0.0341
SbDof23	-1.85	0.49±0.15	11.3±4.5 Å	600	0.0896
SbDof24	-3.58	0.32±0.11	15.3±3.4 Å	1105	0.0274

The chosen models from I-TASSER server were then subjected to energy minimization using Swiss-PdbViewer software for stabilizing their stereochemical properties. The energy values before and after energy minimization for each SbDof protein are shown in **Table 4.16**.

Table 4.16 Comparison of energy values of four SbDof proteins before and after energy minimization

Protein name	Energy KJ/mole (before energy minimization)	Energy KJ/mole (after energy minimization)
SbDof1	2672.395	-5886.063
SbDof19	-587.184	-11498.337
SbDof23	-7395.831	-12317.017
SbDof24	-1743.353	-10377.256

The energy refined final models of SbDof proteins and Dof domain were submitted to the public domain PMDB database (**http://www.caspur.it/PMDB/**) and assigned PMDB IDs i.e. PM0077395 to PM0077398 for SbDof1, SbDof19, SbDof23 and SbDof24, respectively and PM0076448 for Dof domain. The final modeled structures of SbDof proteins are shown in **Fig. 4.29**.

4.4.2.1 Validation of the predicted 3D structure

The stereochemical qualities of the predicted models PM0077395 to PM0077398 of four SbDof proteins were validated by subjecting PDB files to PDBsum server and assessed using PROCHECK server. The Ramachandran plot statistics (% of the residues in the core region, allowed regions and disallowed region) for the four SbDof models are shown in **Table 4.17** and **Fig. 4.30**. The shading on the plot represents different regions; the red areas correspond to the core regions representing the most favorable combinations of phi-psi values. Maximum likelihood of finding residues of protein (>90 %) in the core regions suggests better stereochemical quality **(Morris *et al.*, 1992)**. The results of the PROCHECK analysis indicated that a relatively low percentage of residues have phi/psi angles in the disallowed regions suggesting the acceptability of Ramachandran plots for Dof proteins. The percentage of residues in the core region were found to be 78.5 %, 75.2 %, 56.9 % and 68.9 % for SbDof1, SbDof19, SbDof23 and SbDof24, respectively. The stereo chemical quality of the predicted model was found to be satisfactory. The Ramachandran plots of the modeled Dof proteins are shown in **Fig. 4.30**.

Table 4.17 Ramachandran plot statistics of four SbDof proteins

Protein name	Residues in most favored regions (%)	Residues in additional allowed regions (%)	Residues in generously allowed regions (%)	Residues in disallowed regions (%)
SbDof1	78.5	18.2	1.9	1.4
SbDof19	75.2	19.5	3.8	1.5
SbDof23	56.9	32.9	6.9	3.4
SbDof24	68.9	25.6	4.4	1.1

The qualities of 3D structure of the four SbDof proteins were further estimated by subjecting QMEAN server (**http://swissmodel.expasy.org/qmean/cgi/index.cgi**). The QMEAN Z score were found to be 0.097 (Z-score: -7.24), 0.117 (Z-score: -7.49), 0.127 (Z-score: -7.61) and 0.262 (Z-score: -5.88) for SbDof1, SbDof19, SbDof23 and SbDof24, respectively. The presence of significant QMEAN Z score for these four SbDof proteins suggested the predicted model quality to be acceptable.

4.4.2.2 Superposition of the Dof domain with predicted 3D structure of SbDof proteins

Protein structural comparison and alignment of all predicted models were done using the combinatorial extension (CE) method. The weighted root mean square deviation (RMSD) of Cα trace between the template (Dof domain) and the final refined models were found to be 5.5Å with a significant Z-score of 1.2 for SbDof1, 5.1Å with Z-score of 1.2 for SbDof19, 5.8Å with Z-score 1.6 for SbDof23 and 4.9Å with Z-score 2.3 for SbDof24. The predicted models showed structurally highly conserved region with in domain region (**Fig. 4.31**).

4.4.2.3 Active site identification, metal detection and interaction of Dof domain structure

The binding sites of the target proteins were predicted using Q-Site finder active site prediction tools **(Laurie & Jackson, 2005)**. A total of 10 active binding sites were predicted in 3D structure of Dof domain (PM0076448). Only top 5 prominent active sites are shown in **Fig. 4.32a**, along with the active site residues in **Table 4.18**.

Table 4.18 List of active site residues in predicted top five sites of Dof domain 3D structure

Sites	Active site residues
1	Pro2, Arg3, Tyr14, Tyr17, Asn18, Thr19, Phe25, Leu39, Arg40, Asn41, Val42, Pro43, Val44, Gly46
2	Ser6, Asp8, Thr9, Lys10, Cys26, Arg27, Ala28
3	Arg3, Cys4, Ala28, Cys29, Arg31, Tyr32
4	Cys1, Pro2, Cys12, Tyr13,Tyr14, His24, Phe25, Cys26
5	Trp33, Thr34, Gly37, Ser38, Leu39

The metal binding region within Dof domain was identified by metal detector tool **(Lippi *et al.*, 2008)**. It has been observed that cysteine residues were actively involved in the coordination with metal on the basis of greater metal score and lesser free score value with the presence of disulphide bridges **(Table 4.19)**. The Dof domain was truly functioning as Cys2/Cys2-type zinc finger proteins **(Umemura *et al.*, 2004)** and docking study of Dof domain also suggested the involvement of cysteine residues in the interaction with zinc cofactor as shown in **Fig. 4.32b**.

Table 4.19 Dof domain residues involved in metal interaction

Residue	Positions	Predictions	Metal bonded score	Free score	Disulphide bonded score
C	1	M	0.58	0.03	0.39
C	4	M	0.77	0.05	0.18
C	12	M	0.42	0.20	0.38
C	26	M	0.63	0.06	0.30
C	29	M	0.48	0.11	0.41

Thus an attempt has been made to predict the three-dimensional structures of SbDof proteins for the first time using various bioinformatics tools. The expression profiling and crystal structure identification of these proteins will provide a better insight in to the function of these cloned *Dof* genes in sorghum. Cloning of the remaining predicted *Dof* genes of sorghum along with an extensive *in silico* characterization for homology search, multiple sequence alignment, phylogenetic tree construction, motif scan, *cis*-regulatory element analysis, two-dimensional and three-dimensional structural predictions is under progress in our laboratory.