www.ijbtjournal.com
International Journal of Biological Technology (2012) 3(1):5-10.
ISSN: 0976 – 4313

Original Article

In silico prediction of DNA binding with one finger (*Dof*) transcription factor genes in *Lycopersicum esculentum*

Sunaina Kashyap, Neeraj Kumar, Gaurav Kumar, Vandana Kashyap, Jonnada A.V.P. Rao*

Neutrogenomics and Molecular Biology Laboratory, Department of Biotechnology,
D.D.U. Gorakhpur University, Gorakhpur- 273 009, Uttar Pradesh, India.

***To whom correspondence may be Email address: jonnada_avpr@rediffmail.com; jonnada_avpr@yahoo.co.uk.**

Published: 15, April, 2012; Vol. No. 3(1):5- 10; Online:www.ijbtjournal.com/documents/ijbt150412102.

Abstract

The *Dof* (DNA binding with one finger) is a typical example of plant specific transcription factor involved with regulation of diverse genes associated with multifarious roles. Attempts are being made to characterize the *Dof* gene family of the important vegetable crops namely tomato (*Lycopersicum esculentum)* using available ESTs retrieved from the sequence information generated by the efforts of the international SOL consortium associated with the whole genome sequencing, assembling and annotating these genomes. Studies have revealed the presence of at least 23 *Dof* genes in tomato based on the available ESTs. The size of predicted 23 Le *Dof* (*Lycopersicum esculentum Dof)* related ESTs ranged from 384-804 bp with varying position of *Dof* domain region. Predicted 23 Le*Dof gene were* subjected to multiple sequence alignment, phylogenetic tree construction, and motif analysis. The multiple sequence alignment of these predicted *Dof* gene family clearly revealed the presence of conserved four cysteine residues. Dendrogram construction revealing three major group for *L. esculentum*.. Motif analysis shows that Motif 1 is conserved in all *Dof* EST sequences except EST sequence with accession no. AI486973.1a nd AW648310.1. Protein family analysis of these motifs shows there resemblance to *Dof* Domain, Zinc Finger (Trascription Factor) family.

Keywords: Tomato, *Lycopersicum esculentum*, Solanaceae, Dof gene, genome sequencing, *Arabidopsis.*

Introduction

Many biological processes are strictly regulated through transcriptional control of particular genes in plants. Thereby, plants appear to need a large number of transcription factors governing proper and strict transcriptional regulation in response to developmental programs and environmental changes. Some classes of plant transcription factors share similar DNA-binding domains with animal transcription factors, whereas other classes of transcription factors appear to have classes of transcription factors specific to plant only (Yeangisawa *et al*., 2004). A family of TFs putatively specific to plants is the Dof (DNA-binding with One Finger) of 200-400 amino acids (Yeangisawa *et al*., 2002). Dof protein share a conserved DNA binding domain of 52 amino acid residues referred as Dof domain that is structured as a Cys2/Cys2 Zn finger (Umemura *et al*., 2004) recognizes Cis regulatory element containing the common core 5'-AAAG-3 (Yeangisawa *et al*., 1999).

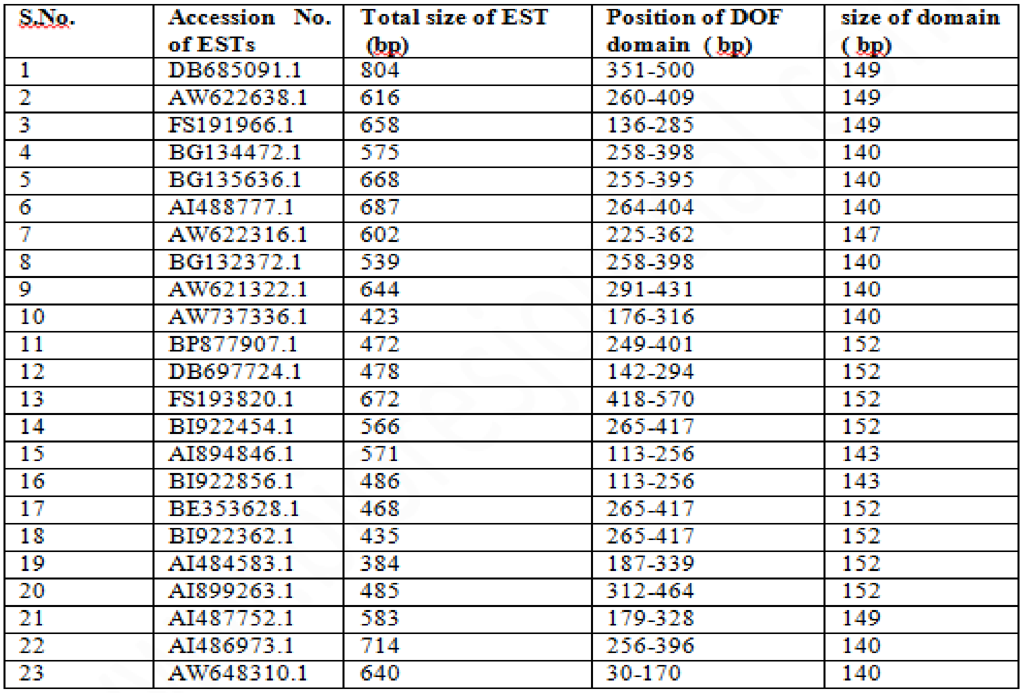
Dof proteins contain only a single copy of Dof domain fond in their N terminal regions and a domain for transcriptional regulation regions on their C-terminus regions. Serine stretches that are frequently located immediately downstream of the Dof domains might be the molecular hinges linking two domains (Yeangisawa *et al.*, 2002). Dof transcription factors controls many vital processes in plants such as photosynthetic carbon assimilation, light regulated gene expression, accumulation of seed storage proteins, germination, dormancy, response to phytohormones, flowering time, and guard cell-specify gene expression (Lijavetzky *et al*., 2003) . The Dof domain is a bifunetional domain that mediates not only DNA-protein interaction but protein-protein interaction also. Dof domains interacted with maize HMG proteins and different HMG proteins enhanced DNA binding of Dof proteins to different extents. (Yeangisawa *et al*., 1997; Krohn *et al*., 2002) In cereal crops like maize rice, barley and wheat different *Dof* genes have been reported. In rice, 30 divergent *Dof* genes have been predicted through genome analysis (Lijavetzky *et al.,* 2003) and 24 different *Dof* genes are reported in barley (Moreno-Risueno *et al*., 2007). In case of wheat 31 different *Dof* gene sequences containing the *Dof* domain were identified thorough a survey of the presently available nucleotide sequence databases (Shaw *et al*., 2009). In maize 54 diverse *Dof* genes have been predicted on the basis of protein sequence databases by comparing them with pfam signature domains (Libault *et al*.,2009) Phylogenetic relationships between rice and Arabidopsis Dof proteins have been previously established and four major clusters of orthologus genes (MCOGs) identified. Phylogenetic analyses have been also described for other plant TF families such as the WRKY, R2R-MYB, bZIP, MADS, GATA, etc (Eulgem *et al*., 2000; Martin *et al*., 1997; Jakoby *et al*., 2002; Parenicova *et al*., 2003; Reycs *et al*., 2004). The origin and evolution of the Dof transcription factor family based on phylogenetic analysis of *Dof* sequences across the representative organisms belonging to green unicellular algae to vascular plants has attempted. A total of 116 *Dof* genes representing green unicellular alga (*Chlamydomonas reinhurdtii*), moss (*Physcomitrulla patens*), fern (*Selaginella moellendorffii*), gymnosperm (*Pinus taeda*), dicotyledoneous angiosperm (*Arabidopsis thaliana*) and monocot (*Oryza sativa* and *Hordeum vulgare*) were analyzed by various bioinformatics tools. The phylogenetic tree constructed revealed the existence of six major clusters of orthologous and paralogous genes that probably originated by gene duplication events from a paraphyletic basal grade(Moreno-Risueno *et al*., 2007)

Tomato (*Lycopersicum esculentum*) represent important food crop of Solanaceae family having essential nutrients as carbohydrate, proteins, fat, vitamins and minerals and having medium size plant genome. Efforts are in progress to sequence the whole genome of *S. tuberosum* & under *Lycopersicum esculentum* International Solanaceae Genome Programme (SOL). International tomato genome sequencing project (ITGSP) has released information that 53% sequencing of tomato genome is completed out of 950 million bp sequences. Availability of sequence information along with ESTs of these crops provides an opportunity to reveal the diversity of *Dof* gene(s). The predicted *Dof* gene can further be PCR amplified, cloned in suitable vector, sequenced and characterized *in silico*. The cloning of putative *Dof* genes can further be investigated for its association with different functions so that it can be manipulated for addressing problem associated with these crop mainly the susceptibility for various abiotic (drought, high temperature) and biotic, as leaf curl disease of tomato.

Materials and Methods

The non redundant *Dof* proteins of Arabidopsis were retrieved from two different sources namely DATF (http://datf.cbi.pku.cn) database and NCBI, (http://www.ncbi.nlm.nih.gov). The nucleotide and deduced amino acid sequences Dof domain of *Arabidopsis* were used to search the potential *Dof* genes in the genome of *L .esculentum* through BLASTN, TBLASTN and discontinuous MEGABLAST (Altschul *et al*., 1990) at the NCBI database (http://www.ncbi.nlm.nih.gov) (Pruitt *et al*., 2007). The annotated *Dof* genes of *L .esculentum* (Table - 1), were further subjected to bioinformatics tools namely ORF finder for fishing out putative CDS and best suitable CDS was used for sequences Alignments by the Clustal X 2.0.10 (Thompson *et al*., 1998). The phylogenetic tree was inferred by bootstrap (1000 reiteratations), NJ, phylogenetic inference using clustal X 2.0.10. The deduced protein sequences of the putative 23 *Dof* genes of *L .esculentum*. were analyzed by means of the MEME (Multiple EM for Motif Elicitation) (http://www.meme.sdse.edu) software version 4.3.0 default parameters were used with the following exceptions: the occurrence of a single motif was set to any number of repetitions then maximum number of motifs to be set to 25 and the minimum width of each motif was set to five amino acid residues. The putative *Dof* protein sequences were subjected to protein functional analysis using pfam (Finn *et al*., 2006) version 24.0 (http://pfam.sanger .ac.uk/).

Table-1: *In silico* predicted *Dof* sequences using ESTs of *L .esculentum*



S.No.	Accession No. of ESTs	Total size of EST (bp)	Position of DOF domain (bp)	size of domain (bp)
1	DB685091.1	804	351-500	149
2	AW622638.1	616	260-409	149
3	FS191966.1	658	136-285	149
4	BG134472.1	575	258-398	140
5	BG135636.1	668	255-395	140
6	AI488777.1	687	264-404	140
7	AW622316.1	602	225-362	147
8	BG132372.1	539	258-398	140
9	AW621322.1	644	291-431	140
10	AW737336.1	423	176-316	140
11	BP877907.1	472	249-401	152
12	DB697724.1	478	142-294	152
13	FS193820.1	672	418-570	152
14	BI922454.1	566	265-417	152
15	AI894846.1	571	113-256	143
16	BI922856.1	486	113-256	143
17	BE353628.1	468	265-417	152
18	BI922362.1	435	265-417	152
19	AI484583.1	384	187-339	152
20	AI899263.1	485	312-464	152
21	AI487752.1	583	179-328	149
22	AI486973.1	714	256-396	140
23	AW648310.1	640	30-170	140

Results and Discussion

There was 23putative *Dof* gene were predicted in *L .esculentum* based on availability of EST sequences and the complete catalog of Dof protein in a single plant species is useful for viewing the structural and functional diversity among Dof proteins as a whole. In order to evaluate the evolutionary relationship between different Dof proteins, and to predict the homologous relations among them, the deduced amino acid sequences of the Dof proteins identified were completely aligned. The sequence alignment of Dof domains (Fig. 1) of different LeDof proteins showed well conserved four cysteine residue putatively responsible of the zinc finger structure. The Dof domain in *L. esculentum* revealed highly conserved sequences with 30 out of 52 amino acids beings 100% conserved in all 18 out of 23 proteins. Phylogenetic tree revealed the existence of three major groups A, B & C in *L. esculentum*. Group A was divided into two subgroups A1 & A2. Group B & C was divided into two subgroup B1 & B2 and C1 & C2. Further A1 & C2was divided in A1.1 & A1.2 and C2.1 & C2.2 respectively as shown in Fig. 2. As a result the *Dof* family in. *L .esculentum* exists in six subgroups constituting eight clusters. The bootstrap value for all the subgroups was very high, suggesting that the genes in each subgroup might share a similar origin. The phylogenetic tree of *Dof* gene family in *L. esculentum* suggests that the branching of *Dof* genes might produce functional diversity. Motifs were identified by means of the MEME software using the 23 LeDof proteins of. *L. esculentum* A total of twenty five conserved motifs were observed. The distribution of conserved motifs in different Dof accessions with respect to cluster is provided in table 2. The motifs are conserved with respect to their sequences. In this analysis both types of motifs were identified those which are present only in closely related genes and those which are commonly found in all genes according to phylogenetic tree. The motif 1with is frequently observed in among all *Dof* genes which reflect the conserved Dof domain. Motif next to motif I in frequent occurrence is motif 2. This motif is present in sixteen proteins out of twenty one. In spite of this, motifs, 3, 4, 5,6,7,8, 9, and 10 are present only in the closely related ESTs.A detailed information about motif 1 given in table-2. The shown Consensus sequence follows the criteria of (Joshi *et al*., 1997).

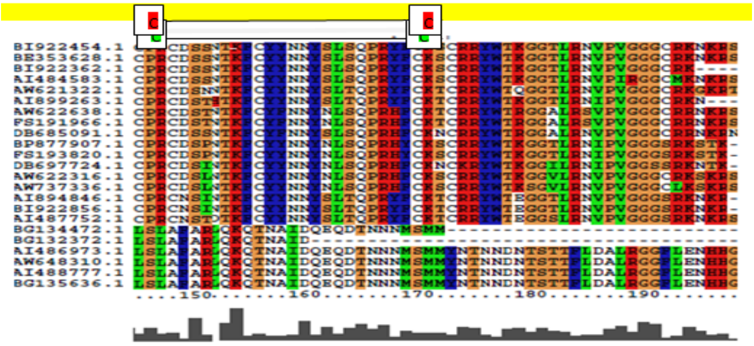


Figure 1: Multiple sequence alignment of *Dof* Domain of *L .esculentum*

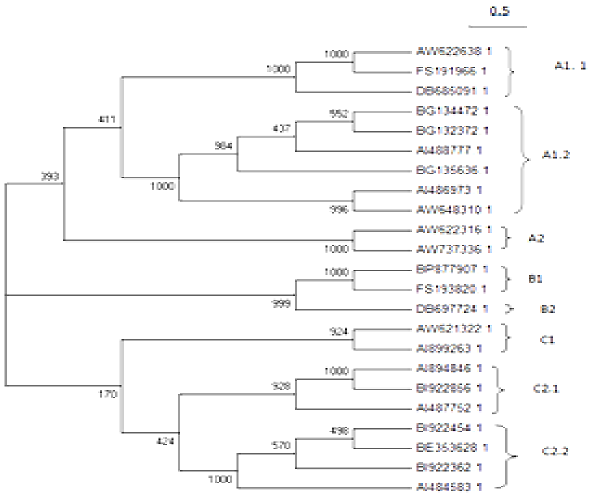
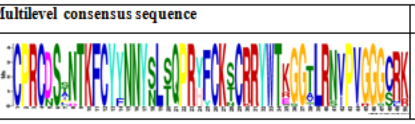


Figure 2: Phylogenetic tree of *L .esculentum* constructed based on amino acid sequences of predicted *Dof* gene using NJ method.

Table-2: Schematic representation of conserved motif in *L . esculentum*

Accession No.	**Motif**																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
DB685091.1	■				■				■		■						■			■	■		
AW622638.1	■				■		■		■														
FS191966.1	■				■		■				■	■									■		
BG134472.1	■	■	■									■											
BG135636.1	■	■	■	■																			
AI488777.1	■	■	■	■																			
AW622316.1	■							■								■		■					
BG132372.1	■	■													■								
AW621322.1	■														■		■				■		■
AW737336.1	■			■																			
BP877907.1	■					■																	
DB697724.1	■			■		■						■							■				
FS193820.1	■					■																■	
BI922454.1	■	■		■														■					
AI894846.1	■	■		■						■				■		■							
BI922856.1	■	■																					
BE353628.1	■	■		■																			
BI922362.1	■	■		■																			
AI484583.1		■																					
AI899263.1	■											■	■										
AI487752.1	■							■						■					■	■			
AI486973.1		■	■																				
W648310.1		■	■										■										

Table- 3: Multilevel consensus sequence of the motif 1

Motif	Length	E-value	Multilevel consensus sequence	Family
1	50 amino acids	2.6e-883		Zf-Dof

Conclusion

There was 23 putative *Dof* gene were predicted in *L .esculentum* based on availability of EST sequences. The Dof domain in *L .esculentum* revealed highly conserved sequences with 30 out of 52 amino acids being well conserved in all 17 proteins sequence .Another 22 amino acids were also highly conserved with variation in only three amino acid residues. The Dof family in *L. esculentum* exists in six subgroups and eight clusters The bootstrapping of tree suggesting that the genes in each subgroup might share a similar origin. Motif analysis shows that Motif 1 is present in all Dof ESTs and significantly matched with Dof Domain (Zinc Finger) in signature Domains of pfam .Further characterization of these predicted *Dof* genes for chromosomal location, gene structure and Cis-regulatory element analysis for putative functions can provide based on this information. PCR based cloning and sequencing of predicted *Dof* genes of potato followed by expression profiling will elucidate the exact role being played by this important transcription factor.

Acknowledgment

The authors acknowledge thanks to Head, Department of Biotechnology, D.D.U. Gorakhpur University, Gorakhpur for providing the infrastructural facilities.

References

Yeangisawa, S. 2004. DOF Domain Proteins: Plant-Specific Transcription Factors Associated with Diverse Phenomena Unique to Plants. *Plant Cell Physiology*, 45 (4): 386-391.

Yeangisawa, S.2002. The DOF family of plant transcription factors. *Trends Plant Science*, 7:555-560.

Lijavetzky, D., Carbonero, P. and Vicent Carbajosa, J. 2003. Genome wide comparative phylogenetic analysis of the rice and Arabidopsis DOF gene families. *BMC Evol. Biol.*, 3:17-28.

Umemura, Y., Ishiduka, T., Yamamoto, R., Esaka, M. 2004. The DOF domain a zine whger DNA binding domain conserved only in higher plants, truly functions as a Cys2/Cys2 Zn wnger domain. *Plant J.*, 37: 741-749.

Yeangisawa, S. and Schmidt, R.J.1999. Diversity and similarity among recognition sequences of DOF transcription factors. *Plant J.*, 17:209-214.

Yeangisawa, S. 1997. DOF DNA binding domains of plant transcription factors contribute to multiple protein-protein interactions. *Eur. J. Biochem.*, 250: 403-410.

Krohn, N.M., Yanagisawa, S. and Grasser, K.D.2002. Specificity of the stimulatory interaction between chromosomal HMBG proteins and the transcription factor DOF 2 and its negative regulation by protein kinase CK2 mediated phosporylation. *J. Biol. Chem.*, 277: 32438-32444.

Moreno-Risueno, M.A., Draz, Carrillo, L., Fuentes, R. and Carbonero, P. 2007. The HvDOF 19 transcription factor mediates the abscisic acid dependent repression of hydrolyase gene in germinating barley aleurone. *Plant J.*, 51: 352-365.

Shaw, L. M., Mclntyre, C. L., Gresshoff, P.M. and Xue, G.P.2009. Members of the *Dof* transcription factor family in *Triticum aestivum* are associated with light-mediated gene regulation. *Functional and Integrative Genomics*, DOI 10.1007/s 142-009-0130-2.

Libault, M., Joshi, T., Benedito, V.A., Xu, D., Udvardi M.K. *et al*. 2009. Legume transcription factor genes: What makes Legumes so special. *Plant Physiology*,151: 991-1001.

Eulgem, T., Rushton, P.J., Robatzek, S. and Somssich, I.E. 2000. The WRKY super family of plant transcription factors. *Trends Plant Sci*., 5: 199-206.

Martin, C. and Paz, J.1997. MYB transcription factors in plants. *Trends Genet.*, 13: 67-73.

Jakoby, M., Weisshaar, B., Droge-Laser, W., Vicente - Carbajosa J., Tiedemann, J., Tiedemann, J. *et al*.2002. bZIP transeription factors in *Arabidopsis*. *Trends Plant Sci.*, 7:106-111.

Parenicova, L., de Folter, S., Kie Ver M., Horner D.S. Favalli C. *et al*., 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*:

new openings to the MADS world. *Plant Cell*, 15: 1538-1551.

Reycs, J.C., Muro-Pastor, M. and Florencio, F.J.2004. The GATA family of transcription factors in Arabidopsis and rice. *Plant Physiol.*, 134:1718-1732.

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

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

Allschul, S.F., Gish, W., Webb Miller Eugene W. and Myers, D.J.1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215: 403-410.

Thompson, J.D., Higgins, D.G., Gibson, T.J. and Clustal, W.1994. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22:4673-4680.

Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V. *et al*. 2006. Pfam: clans web tools and services. *Nucleic Acids Res.*, 34: D247-D251.

Castro, E.D., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S. *et al*. 2006. Scan Prosite: detection of PROSITE signature matches and Pro Rule-associated functional and structural residues in proteins. *Nucleic Acids Res*., 34: W362-W365.

Quevillon, E., Sillventoinen, V., Pillai, S., Harte, N., Mulder, N., *et al*, Interpro Scan: protein domains identifier. *Nucleic Acids Res.*, 33: W116-120.

Joshi, C.P., Zhou, H., Huang, X. and Chiang ,VL.1997. Context sequences of translation initiation codon in plants. *Plant Mob. Biol.*, 35: 993-1001.

Manuscript Progress Date

Received : 28 .12.2011
Revised : 02.02.2012
Accepted : 30.03.2012