

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

Sunayana Kashyap^{1*}, Vandana Kashyap^{2*}

¹Department of Zoology, Banaras Hindu University, Varanasi–221005, India

²Department of Biotechnology, D.D.U. Gorakhpur University, Gorakhpur- 273 009, Uttar Pradesh, India.

Corresponding author:

Sunayana kashyap

Phone No.: +917376001438.

Abstract

The *Dof* (DNA binding with one finger) is plant specific transcription factor which has been discovered in both monocots and dicots. It regulates diverse genes functionality which is coupled with multifarious roles. The complete genome sequence information of plant provides an opportunity to find out novel DOF genes through genome annotation and genome wide comparative analysis. Here afford has been made to characterize the *Dof* gene family of the important vegetable crops namely potato (*Solanum tuberosum*) using available ESTs retrieved from International Solanaceae Genome Programme (SOL) consortium. Studies have shown occurrence of 21 *Dof* genes in potato based on the available ESTs. The size of predicted 21 *StDof* related ESTs ranged from 368 -942 bp with varying position of *Dof* domain region. Predicted 21 *StDof* gene were subjected to multiple sequence alignment, phylogenetic tree construction, and motif analysis. The multiple sequence alignment analysis of these predicted *Dof* gene family open to the elements of conserved four cysteine residues. Dendrogram assembly gives information of three major clusters of *S. tuberosum*. Motif analysis confirms that Motif 1 is present in all *Dof* EST sequences except EST sequence with accession no. AW224680.1.

Index Terms : *In silico* prediction Potato *Dof*, , , *Solanum tuberosum*, Transcription factor.

I: Introduction

Gene expression is regulated by different mechanisms, although the most important one is transcriptional control. DNA binding transcription factors regulate the rate of transcription by interacting through their DNA binding domain with *cis* regulatory elements in the promoter of target genes. DNA binding domain is in general highly conserved and has been used to classify the transcription factor in to different families. Some of these families are common to most eukaryotic organisms but others are specific to a given taxonomic group. Some classes of plant transcription factors share similar DNA-binding domains with animal transcription factors, whereas other classes of transcription factors appear to precise for plant [1]. The dof family is one of the well characterized plant specific transcription factor having diverse roles. Dof proteins are typically composed of 200-400 amino acids [2]. 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 recognizes *cis* regulatory element containing the common core 5'-AAAG-3' [3],[4]. In spite of intensive homology in Dof domain, the rest of amino acid sequences in the proteins are divergent. Transcription factors sometimes contain multiple DNA binding domains, but in case of Dof proteins only a single copy of Dof domain can always be found in their N terminal regions. Dof proteins usually appear to consist of two major domains: an N-terminal conserved DNA binding domain and a C-terminal domain for transcriptional regulation. Serine stretches that are frequently located immediately downstream of the dof domains might be the molecular hinges linking two domains [2]. Substantial amount of work has been done in model crops like *Arabidopsis* & rice. In *Arabidopsis* & rice 37 & 30 divergent Dof genes have been predicted respectively, through whole genome analysis [5]. The Dof (DNA binding with one finger) family is a typical example of Zinc finger transcription factors extensively reviewed [1], [2]. The first protein to be identified with a Dof domain was MNB 1a/ZmDof1 of maize , which

can interact with the cauliflower mosaic virus 35S promoter [6]. Four cDNA clones encoding Dof proteins have already been isolated from maize [7],[8]. A partial cDNA encoding a Dof domain was also isolated from pumpkin [9]. Several Dof proteins have already been identified in both monocots and dicots, but not in yeast and animals [10]. Washio and coworker has reported the isolation of five cDNAs encoding DOF proteins (OsDof1-5) from a rice aleurone library by southwestern screening with a pyrimidine box probe [11]. Maize Dof 1 gene has been studied for its role in improved nitrogen assimilation and growth in low nitrogen condition [1], but only *OsDof3* was studied in detail [11]. Different bioinformatics and phylogenetic analysis tool have been used reveal origin and evolution of the DOF transcription factor family in green unicellular algae to vascular plants. A total of 116 DOF genes have also been studied in green unicellular alga (*Chlamydomonas reinhardtii*), moss (*Physcomitrella patens*), fern (*Selaginella moellendorffii*), gymnosperm (*Pinus taeda*), dicotyledonous angiosperm (*Arabidopsis thaliana*) and monocot (*Oryza sativa* and *Hordeum vulgare*). *Dof* genes have been also reported in different major cereal crops like wheat barley maize, and rice. there In 30, 24 31 putative Dof gene have reported in rice barley and wheat respectively [5] ,[12], [13]. In maize 54 divergent *Dof* genes have been predicted through genome analysis [14],[15]. Recently 23 *Dof* gene have been reported in tomato based on ESTs sequence [16].

Potato (*S. tuberosum*) is a one of most important food crop of Solanaceae family having essential nutrients as carbohydrate, proteins, fat, vitamin and minerals and having medium size plant genome. It is the fourth major crop of the world, coming after rice wheat and maize.

Efforts are in progress to sequence the whole genome of *S. tuberosum* under International Solanaceae Genome Programme (SOL). Potato genome sequencing consortium (PGSC) has released first draft genome of potato for public use *Solanum tuberosum*-RH.bac.tar.bz2(24MB), *Solanum tuberosum* RH89-03916.BACs (RHPOTKEY library) & *Solanum.tuberosum*-

RH.bac.tar.bz2(27MB) *Solanum tuberosum* RH89-03916BACs(RHPOTKEY library) out of 850 million bp sequences provide an opportunity to reveal the diversity of Dof gene(s). Availability of sequence information along with ESTs of these crops provides an opportunity to reveal the diversity of *Dof* gene(s).

II. EXPERIMENTAL STUDIES:

The non redundant DOF proteins of Arabidopsis were retrieved from two different sources namely DATF (<http://datf.cbi.pku.cn>) database and NCBI17, (<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 *Solanum tuberosum* and NCBI17, (<http://www.ncbi.nlm.nih.gov>) [17]. The nucleotide and deduced amino acid sequences DOF domain of Arabidopsis were used to search the potential DOF genes in the genome of *Solanum tuberosum* through BLASTN, TBLASTN and discontinuous MEGABLAST [18] at the NCBI database (<http://www.ncbi.nlm.nih.gov>) [17].

The annotated DOF genes of *S. tuberosum* were further subjected to bioinformatics tools namely ORF finder for fishing out putative CDS and best suitable CDS was used for Alignments of protein sequences that was done by the ClustalX2.0.10 [19]. The phylogenetic tree was inferred by bootstrap (1000 reiteratations), NJ phylogenetic inference using clustalX2.0.10. The deduced protein sequences of the putative 21 DOF genes of *S. tuberosum* were analyzed by means of the MEME (Multiple EM for Motif Elicitation).

(<http://www.meme.sdsc.edu/meme/meme.html>) 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 [20] version 24.0 (<http://pfam.sanger.ac.uk/>), PROSITE [21] version 20.58 (<http://www.expasy.ch/prosite/>) INTERPROSCAN [22] version 24.0 (<http://www.ebi.ac.uk/Tools/InterProScan/>) and MOTIFSCAN (http://myhits.isb-sib.ch/cgi-bin/motif_scan).

Table:1 *In silico* predicted *Dof* sequences using ESTs of *S. tuberosum*

Sr.No.	Accession number of ESTs of <i>Solanum tuberosum</i>	Total size of EST (bp)	Position of DOF domain (bp)	size of DOF domain(bp)
1	CK278292.1	942	242-391	149
2	BG888213.1	651	243-392	149
3	BQ510916.2	845	311-460	149
4	CK264965.1	862	122-274	152
5	DN906511.1	784	121-270	149
6	CK863864.1	636	121-270	149
7	BQ114477.2	624	143-295	152
8	BE340557.1	538	123-275	152
9	BF052550.1	368	143-295	152
10	BG595870.1	731	67-219	152
11	BQ116931.2	528	46-198	152
12	BG596035.1	751	237-389	152
13	BQ513434.1	670	164-307	143
14	BG592670.1	735	161-313	152
15	BQ116367.2	702	271-423	152
16	BE922578.1	546	271-423	152
17	BQ112887.2	685	271-423	152
18	BM111519.1	537	165-308	143
19	AW224680.1	486	249-413	164
20	CV470148.1	762	218-385	167
21	BM113542.1	656	248-388	140

III. FORMULATION OF EXPERIMENTS:

21 putative Dof genes were predicted in *S. tuberosum* 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 (Table I). 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 StDOF proteins showed well conserved four cysteine residues putatively responsible of the zinc finger structure. The dof domain in solanum revealed highly conserved sequences with 28 out of 52 amino acids being 100% conserved in all 21 proteins. Another 8 amino acids were also highly conserved with variation in only two amino acid residues. The phylogenetic tree constructed using clustal X 2.010 software with bootstrap value 1000 revealed the existence of two major groups A and B&C.

Group A was divided into two subgroups A1 and A2, Group B was divided into two B1 and B2, where as B1 subgroups were further subdivided into three clusters as shown in Fig.2. As a result the DOF family in sorghum exists in four subgroups constituting five 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 *S.tuberosum* suggests that the branching of DOF genes might produce functional diversity. Motifs were identified by means of the MEME software using the 21 StDOF proteins of *S.tuberosum*. A total of twenty one conserved motifs were observed. The distribution of conserved motifs in different DOF accessions with respect to cluster is provided in Table II. 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 I 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.

The overall multilevel consensus sequence associated with each of the 21 motifs is given in table III. The shown Consensus sequence follows the criteria of [23]. Putative functionality can provide based on Pfam analysis Table IV.

IV. CONCLUDING REMARKS:

21 putative Dof genes were predicted in *S. tuberosum* based on availability of EST sequences. The Dof domain in *S. tuberosum* revealed highly conserved sequences with 28 out of 52 amino acids being well conserved in all 21 proteins. Another 8 amino acids were also highly conserved with variation in only three amino acid residues. The Dof domain in *S. tuberosum* revealed highly conserved sequences with 28 out of 52 amino acids being well conserved in all 21 proteins. Phylogenetic tree constructed using clustal X 2.0.10 software with bootstrapping revealed the existence of three major groups A, B & C in *S. tuberosum*. Group A was divided into two subgroups A1 & A2. Group B was divided into two subgroups B1 & B2. So that as a result the Dof family in *S. tuberosum* exists in four subgroups and five clusters. Motif analysis shows that Motif 1 is present in all Dof ESTs. Except EST sequence with accession no. AW224680.1. 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.

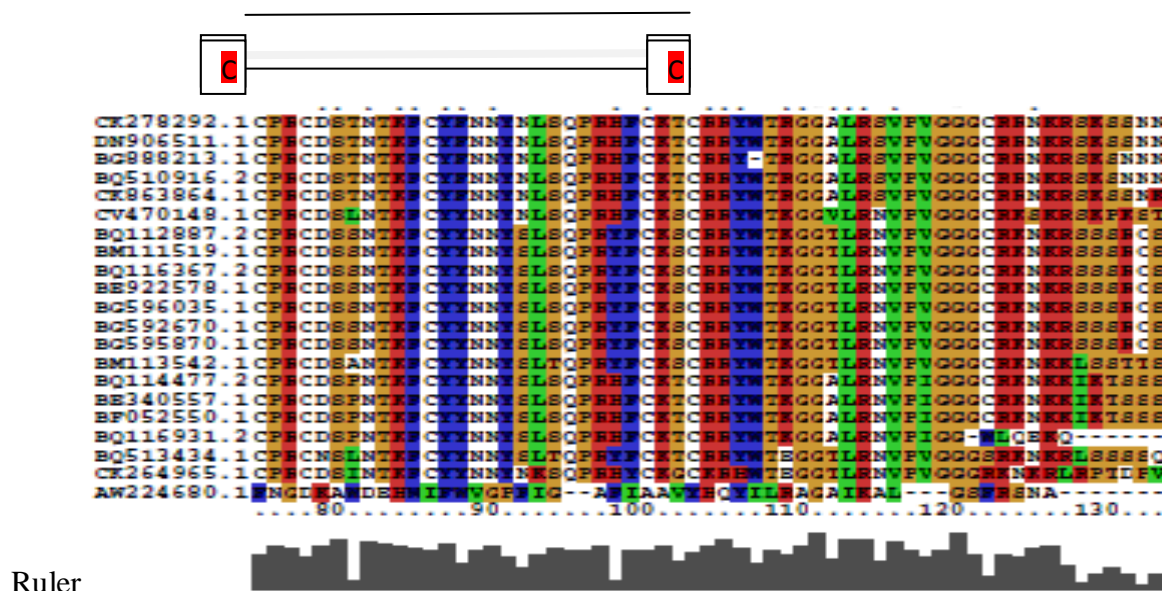


Fig. 1 Multiple sequence alignment of *Dof* Domain of *S. Tuberosum*.

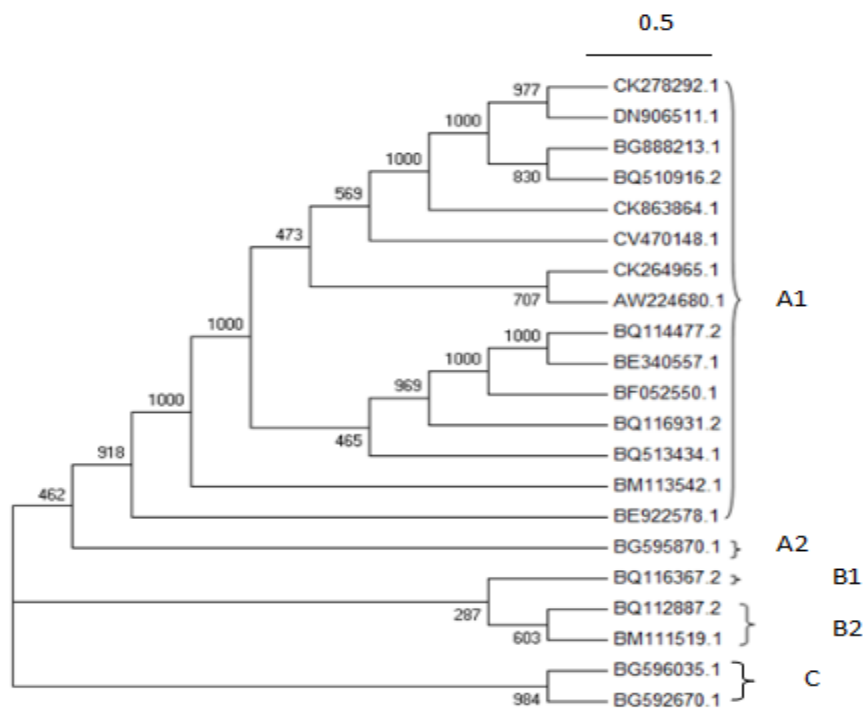


Fig.2 Phylogenetic tree of *S. tuberosum* constructed based on amino acid sequences of predicted *Dof* gene using NJ method.

Table:2 Schematic representation of conserved motif in *Solanum tuberosum*

ACCESSION No.	MOTIF																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
CK278292.1	█	█	█			█	█	█									█				
BG888213.1	█	█	█			█											█				
BQ510916.2	█	█	█			█	█					█									
CK264965.1	█		█							█	█	█	█			█					█
DN906511.1	█	█	█				█	█													
CK863864.1	█						█			█											
BQ114477.2	█	█						█	█												█
BE340557.1	█	█						█	█												
BF052550.1	█								█	█											
BG595870.1	█	█		█	█											█			█		
BQ116931.2	█									█											
BG596035.1	█	█	█	█	█																
BQ513434.1	█	█														█		█			█
BG592670.1	█	█	█	█	█																
BQ116367.2	█	█	█	█															█		
BE922578.1	█	█	█																		
BQ112887.2	█	█	█	█																	
BM111519.1											█			█							
AW224680.1	█	█											█			█		█			█
CV470148.1	█	█		█													█				
BM113542.1	█	█	█	█																	

Table:3 Multilevel consensus sequence of motifs

Motif	Length	E-value	Multilevel consensus sequence
1	39	3.4e-655	
2	39	8.4e-278	
3	39	1.7e-164	
4	31	1.9e-120	
5	50	1.2e-052	
6	45	7.7e-052	
7	20	3.4e-033	
8	49	5.8e-020	
9	28	6.9e-020	
10	20	3.7e-006	
11	20	6.1e+001	
12	20	3.1e+002	
13	20	3.7e+002	









14	20	1.4e+003	
15	20	2.2e+004	
16	20	6.4e+004	
17	20	3.6e+004	
18	20	4.9e+005	
19	20	1.9e+006	
20	20	2.5e+006	
21	20	5.8e+006	

Table: 4 Protein functional analysis of *S. tuberosum* using PFAM

S.N	Seq. No.	STATUS	FAMILY	DESCRIPTION
1	1	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger (Trascription Factor)
2	2	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
3	3	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
4	4	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
5	5	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
6	6	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
7	7	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
8	8	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
9	9	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
10	10	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
11	11	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
12	12	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
13	13	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
14	14	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)

15	15	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
16	16	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
17	17	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
18	18	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
19	19	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)
20	20	SIGNIFICANT	Zf-Dof	Dof Domain, Zinc Finger(Trascription Factor)

ACKNOWLEDGMENT

The authors wish to acknowledge, Department of Biotechnology, D.D.U. Gorakhpur University Gorakhpur for providing the infrastructural facilities and CSIR New Delhi for financial support of UGC, New Delhi.

Conflict of Interests: The authors declare no competing financial interest and sharing equal contribution*in paper.

V. REFERENCES:

- [1] S. Yeangisawa, "DOF Domain Proteins: Plant-Specific Transcription Factors associated with Diverse Phenomena Unique to Plants," *Plant Cell Physiology*, vol.45, pp.386-391, February 2004.
- [2] S. Yeangisawa, "The DOF family of plant transcription factors," *Trends Plant Science*,

vol.7, pp. 555-560, December 2002.

[3] Y. Umemura, T. Ishiduka, R. Yamamoto and M. Esaka, "The DOF domain a zinc finger DNA binding domain conserved only in higher plants, truly functions as a Cys₂/Cys₂ Zn finger Domain," *Plant J.*, vol. 37, pp.741-749, April 2004.

[4] S. Yeangisawa, and R. J. Schmidt. "Diversity and similarity among recognition sequences of DOF transcription factors," *Plant J.*, vol.17, pp. 209-214, January 1999.

[5] D. Lijavetzky, P. Carbonero, and J. Vicente Carbajosa. "Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis DOF gene families," *BMC Evol. Biol.*, vol. 3, pp.17-28, September 2003.

[6] S. Yanagisawa and K. Izui, "Molecular cloning of two DNA-binding proteins of maize that are structurally different but interact with the same sequence motif," *J. Biol. Chem.*, vol.268, pp.16028–16036, July 1993.

[7] S. Yanagisawa, "A novel DNA binding domain that may form a single zinc finger motif," *Nucleic Acids Res.*, vol. 23, pp.3403–3410, September 1995.

[8] J. Vicente-Carbajosa, "A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 94, pp.7685–7690, July 1997.

[9] Y. Kisu, "Characterization and expression of a new class of zinc finger protein that binds to silencer region of ascorbate oxidase gene," *Plant Cell Physiol.*, vol.39, pp.1054–1064, October 1998.

[10] S. Yanagisawa, "Dof DNA binding domains contain a novel zinc finger motif," *Trends Plant Sci.*, vol.1, pp.213–214, July 1996.

- [11] K. Washio, "Identification of Dof proteins with implication in the gibberellin- regulated expression of a peptidase gene following the germination of rice grains," *Biochim. Biophys. Acta*, vol.1520, pp. 54–62, July 2001.
- [12] M.A Moreno-Risueno, L. C. Draz, R. Fuentes, and P. Carbonero, "The HvDOF 19 transcription factor mediates the abscisic acid dependent repression of hydrolyase gene in germinating barley aleurone," *Plant J*, vol.51, pp. 352-365, August 2007.
- [13] L. M., Shaw, C. L., McIntyre, P.M. Gresshoff, and G.P Xue, "Members of the *Dof* transcription factor family in *Triticum aestivum* are associated with light-mediated Gene.regulation," *Functional and Integrative Genomics*, vol.9, pp.485-498, July 2009.
- [14] M. Libault, T. Joshi, V.A. Benedito, D. Xu, and M.K. Udvardi "Legume transcription factor.genes: What.makes.Legumes.so.special," *Plant.Physiology*, vol.151, pp.991-1001, November 2009.
- [15] D. Yadav, H., Kushwaha, S. Gupta, S. Kashyap, and V. Singh "In silico Characterization of DOF (DNA binding with one finger) transcription factor genes in different crops," *Journal of Computational Intelligence in Bioinformatics*, vol.4, pp.63-67, August 2011.
- [16] S. Kashyap, N. Kumar, G. Kumar, V. Kashyap, and A.V.P. R. Jonnada "In silico prediction of DNA binding with one finger (*Dof*) transcription factor genes in *Lycopersicon esculentum*," vol 3, pp.5-10, April 2012.
- [17] K.D., Pruitt, T. Tatusova, and D. R. Maglott NCBI reference sequences "(RefSeq): a curated non redundant sequence database of genomes, transcripts and proteins," *Nucleic Acids Res.*, vol.35, pp. D61-D65, January 2007.
- [18] S.F. Altschul, W. Gish, W. Webb Miller Eugene and D. J. Myers, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, pp.403-410, October 1990.

- [19] J.D. Thompson, D.G Higgins, T.J. Gibson, and W.Clustal, “Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice,” *Nucleic.Acids Res.*, vol.22, pp.4673-4680, November 1994.
- [20] R.D. Finn, J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, and V. Hollich, “Pfam: clans web tools and services,” *Nucleic Acids Res.*, vol.34, pp. D247-D251, January 2006.
- [21] E.D. Castro, C.J.A. Sigrist, A. Gattiker, V. Bulliard, and P.S. Langendijk-Genevaux, “Scan Prosite: detection of PROSITE signature matches and Pro Rule-associated functional and structural residues in proteins,” *Nucleic Acids Res.*, vol. 34, pp.W362-W365, July 2006.
- [22] E. Quevillon, V. Sillventoinen, S. Pillai, N. Harte, and N. Mulder, “Interpro Scan: protein domains identifier,” *Nucleic Acids Res.*, vol.33, pp.116-120, July 2005.
- [23] C.P. Joshi, H. Zhou, X. Huang, and V.L.Chian “Context sequences of translation initiation codon in plants,” *Plant Mob. Biol.*, vol.35, pp.993-1001, December 1997.

Sunayana Kashyap M.Sc. in Biotechnology from Department of Biotechnology, D.D.U. Gorakhpur University, Gorakhpur (U.P.). Published a paper on an *In silico* prediction of DNA binding with one finger (*Dof*) transcription factor genes in *Lycopersicum esculentum* and different crop. Currently pursuing Ph.D. from BHU, Varanasi,(U.P.).



Dr.Vandana Kashyap Ph.D. in Biotechnology from Department of Biotechnology, D.D.U. Gorakhpur University, Gorakhpur (U.P.) and Environmental Biotechnology Division, NEERI, Nagpur, (Maharashtra). Published a paper on an assessment of *in vitro* regeneration from seven agronomical cultivars of pigeon pea (*Cajanus cajan* (L) Millsp.) in Proceeding of International Conference on Food Security and Environmental Sustainability, Indian Institute of Technology Kharagpur, Dec. 2009. Also worked on *In silico* prediction of DNA binding with one finger (*Dof*) transcription factor genes in *Lycopersicum esculentum*.

