

<https://doi.org/10.48047/AFJBS.6.13.2024.5371-5383>



African Journal of Biological Sciences

Journal homepage: <http://www.afjbs.com>



Research Paper

Open Access

Genome wide analysis of DOF (DNA-binding with one finger) gene family in *Stevia rebaudiana* by an insilico approach

Asim Ali¹, Saman Zulfiqar², Jawad Ali¹, Ansa Batool³, Samiya Rehman¹, Fouzia Tanvir⁴, Fatima Ali¹, Ayesha Aslam¹, Aqeela Nawaz⁴, Yasir Nawaz^{4*}, Maryyum Amin¹, Memmona Ghulam Rabbani¹

1 Department of Biochemistry, University of Okara, Okara, Pakistan

3 Department of Biological Sciences, Superior University Lahore, Pakistan

4 Department of Zoology, Faculty of Life Sciences, University of Okara, Okara, Pakistan

2 Department of Botany, The Government Sadiq College Women University Bahawalpur, Pakistan

Corresponding author: Royyasirawaz@gmail.com

Volume 6, Issue 13, August 2024

Received: 15 Jun 2024

Accepted: 25 July 2024

Published: 15 August 2024

[doi:10.48047/AFJBS.6.13.2024.5371-5383](https://doi.org/10.48047/AFJBS.6.13.2024.5371-5383)

Abstract

Transcription factors act as the master regulators and control the Cellular processes by decoding the DNA sequence. Plant growth and development is significantly influenced by the DOFs transcription factors. The current study used the DOF genes from *Oryza Sativa* to conduct a genome-wide investigation of DOF family in *Stevia rebaudiana*. Different bioinformatics methods were utilized to perform phylogenetic analysis, exon and intron structure, gene structure, chromosomal distribution of various DOFs genes, conserved domain analysis, and subcellular localization prediction. The genome databases 52 DOF genes in *Stevia rebaudiana* were extracted. Chromosome 4 contained the most SrDOF genes (8 members), followed by chromosome 2 & 11 (7 SrDOF). The majority of SrDOF genes had a one-exon gene structure while mostly DOF genes were intron less. The motif composition and gene structure of SrDOFs from same sub-group were similar. More than 15 conserved motifs were observed in all these proteins. The Zf DoF Superfamily domain was found in all DOF proteins except SrDOF 17 & 20. The molecular weights of 52 SrDOF proteins ranges from 16.7 to 52.16 KDa with average molecular weight 32.6 KDa. Iso electric point of SrDOF proteins range from 4.79 to 9.72 with average PI 7.89. To conclude, this will help to understand the functions of DOF family in evolutionary and physiological aspects of *Stevia rebaudiana* and provides future direction for novel work to improve crop productivity. Identifying single-gene families in multiple plant species are best to enhance crop productivity, growth, and development.

Keywords: Transcription factors, Bioinformatics, Rice, *Oryza Sativa*, *Stevia rebaudiana*, Genome-wide analysis

Introduction

The proteins involved in the transcription and capable of changing the expression level of genes are termed as the Transcription factors in modern times (1). Transcription Factors being “master regulators” and “selector genes” controls the cell process and interpret the genome by decoding the DNA sequence (2). Transcription factors have capability to drive and control the differentiation (3) and trans-differentiation (4). DNA binding with one finger (DOF) protein belong to the Zinc Finger Superfamily (5). DOF transcription factor have been widely known for the Contribution in the process of the germination in the plants (6). DOF is a plant-specific transcription factor family that regulates vascular tissue development Xylem and Phloem, Nonliving stress factors (abiotic), cell cycle, photoperiodregulation, floral organ abscission, redox homeostasis, and secondary metabolite generation,among other biological activities. The N-terminal DNA binding domainof the DOF protein is highly conserved. This DOF domain is a 50–53 amino acid Cys2Cys2(C2C2) type zinc finger motif that has remained constant throughout evolution (7).

Apart from the highly conserved DOF domain, this protein'stranscriptional regulatory C-terminal amino acid sequence is highly variable. DOF-TFs bind to a core sequence of (AT)/AAAG in the cis-regulatory region of their target gene. DOF-TF,like other zinc fingers has the ability to bind both DNA and proteins, making it a bifunctionalprotein. DOF-TF and other physiologically relevant proteins are known to interact with severalDOF proteins. The intrinsic property of DOF-TF to homo- and heterodimerize and itsinteraction with other transcription factors (like bZIP, Myb-TF, and TCP-TF) reflect its complex regulatory network (8).

First Zn DOF transcription factor was identified in the Maize (7). Various Studies have identified different number of DOF TFs in different species like Arabidopsis Thalina contain 37 (7), Oryza sativa 30 (9), 36 watermelon (10), 22 in (11) and 24 in rose (12). In the present study we collected the genes related to the Zn DOF family in Stevia rebaudiana and then performed various bioinformatics analysis to gain insights into their Gene structure, Motif structure, evolutionary relationship among their genes, location of genes on the Chromosomes and properties of the genes and proteins were determined (9).

Materials and Methods

Dataset

The whole genome of *S. rebaudiana* was downloaded from Figshare database

(<https://figshare.com/>) as reported in literature (13). The amino acid sequence of the 30 *Oryza sativa* (Rice) DOFs were used as blast queries against the *S. rebaudiana* local proteome database by using Tb tools Software V1.095 (14).

All the output genes were analysed using the NCBI-CDD batch database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) to confirm the conserved DOF domains. Genes without conserved DOF domains were discarded. The molecular weights and isoelectric points of SrDOF sequences were analyzed by ExPASy Compute pI/Mw tool (https://web.expasy.org/compute_pi/).

The full-length genomic and coding SrDOF sequences were extracted from Genomic and CDS file of *S. rebaudiana* respectively by Tbttools software (<https://bio.tools/tbttools>) and physical positions of SrDOF genes were obtained from the *S. rebaudiana* GFF3 file of *S. rebaudiana* genome by Tbttools software and were used to construct the chromosomal distribution map of SrDOF proteins with Tbttools software (14).

Gene structure, and Phylogenetic Analysis

To characterize the phylogenetic relationships SrDOF amino acid sequences were used. Multiple sequence alignment of SrDOFs was performed by Clustal Omega online (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). Maximum likelihood phylogenetic tree with bootstrap value 100 was constructed by MEGAXI software (15). Genestucture of SrDOF proteins was made by using the Gene Structure Display Server (<http://gsds.gao-lab.org/>) online to identify exon-intron organizations. The amino acid sequences of SrDOF proteins were analysed using MEME tools (16).

Conserved Motif analysis

The server (<https://meme suite.org/meme/tools/meme>) with default parameters was used to identify the conserved motifs. Different motifs were represented using different colored boxes (17).

Results and discussion

Identification and Characterization of the DOF Genes *Stevia rebaudiana*:

To identify *Stevia rebaudiana* SrDOF transcription factor genes, we used the amino acid sequence of *Oryza sativa* (OsDOF) transcription factor blast queries against the *S. rebaudiana* proteome database to search for candidate genes.

NCBI CDD-Batch programs were used for analysis of output genes to confirm the conserved DOF domains, and a total of 52 *S. rebaudiana* DOF genes were identified. *S. rebaudiana* DOF genes were named SrDOF01 to SrDOF52 according to their chromosomal positions. Many characteristics of SrDOF proteins were analyzed, such as DOF gene ID, chromosome number along with start and end position on chromosome, amino acid number, Molecular Weight, Isoelectric pH and coding strand. SrDOF08 has maximal coding region with length 19176 bp coding 199 aa while SrDOF41 has 4544 bp encoding 481 amino acids followed by the SrDOF encoding 481 aa with 2510 bp. This is shown in table 1.

Gene identification and chromosomal distribution analysis

Chromosomal distribution analysis revealed that SrDOF genes were unevenly distributed among chromosomes (Fig. 1, Table 1). Chromosome 4 contained the most SrDOF genes (8 members), followed by chromosome 2 & 11 (7 SrDOF). While chromosome 3, 7, 10 contain only 2 SrDOF genes on them. SrDOF 50, 51 & 52 genes do not lie on any of 11 chromosomes rather they are located on separate Scaffold DNA contig03291, contig03842 and contig05355 respectively. Chromosome 2 is largest chromosome having length almost 170 Mb while chromosome 3 is the smallest chromosome having length 65 Mb. Cluster of 6 SrDOF genes located on chromosome 2 (SrDOF8-13) and chromosome 11 (SrDOF43-49). This is shown in table 1.

Proteomic analysis of GRF gene family

For understanding the biological functioning Molecular Weight, Isoelectric point was determined. The molecular weights of 52 SrDOF proteins range from 16.7 to 52.16 kDa with average molecular weight 32.6 kDa. Isoelectric point (pI) of SrDOF proteins range from 4.79 to 9.72 with average pI 7.89. Amino acid number of SrDOF proteins range from 131 to 472 with average 292 amino acids. 27 SrDOF genes lie on forward strand and 25 SrDOF genes lie on reverse strand as shown in table 1.

Table 1: Detailed information on the Genomic and physiochemical properties of SrDOF Family

Our ID	ID	Gene Length	Protein Length	Chromosome	Mw (Kda)	PI	Start Position	End Position	Coding Strand
SrDOF01	Streb.1G012110.1	1044	297	1	32.6	7.61	25924028	25925071	+
SrDOF02	Streb.1G025010.1	1613	353	1	38.38	6.56	95579602	95581214	+
SrDOF03	Streb.1G025090.1	1596	216	1	24.79	4.94	95828084	95829679	+
SrDOF04	Streb.1G030170.1	1558	216	1	24.74	4.94	107310233	107311790	+
SrDOF05	Streb.1G046430.1	2378	178	1	20.14	6.49	135366603	135368980	-
SrDOF06	Streb.1G048490.1	1874	285	1	31.8	8.97	139034517	139036390	-
SrDOF07	Streb.2G014770.1	2207	199	2	20.72	9.01	37458123	37460329	-
SrDOF08	Streb.2G045920.1	19176	199	2	20.69	9.01	154768195	154787370	+
SrDOF09	Streb.2G049930.1	1701	207	2	23.19	6.52	161342006	161343706	+
SrDOF10	Streb.2G055760.1	2046	347	2	38.37	7.02	170158968	170161013	+
SrDOF11	Streb.2G056380.1	2030	331	2	35.89	9.05	171025581	171027610	-
SrDOF12	Streb.2G056910.1	774	334	2	36.14	8.89	171814313	171815086	+
SrDOF13	Streb.2G056930.1	2120	327	2	35.06	9.45	171849814	171851933	+
SrDOF14	Streb.3G022590.1	2120	306	3	33.21	9.16	33066265	33068384	-
SrDOF15	Streb.3G022640.1	2188	341	3	36.62	9.12	33174999	33177186	-
SrDOF16	Streb.4G004390.1	753	290	4	32.07	8.19	56096332	56097084	+
SrDOF17	Streb.4G004400.1	825	269	4	29.58	9.16	56111181	56112005	-
SrDOF18	Streb.4G005030.1	513	383	4	43.21	8.58	58982613	58983125	-
SrDOF19	Streb.4G010210.1	972	337	4	36.65	6.1	76831968	76832939	+
SrDOF20	Streb.4G016810.1	654	223	4	24.83	8.79	92233783	92234436	+

SrDOF21	Streb.4G0 16830.1	1470	257	4	28.3	6.07	9226326 6	9226473 5	+
SrDOF22	Streb.4G0 21950.1	867	347	4	37.1	6.58	1034132 07	1034140 73	+
SrDOF23	Streb.4G0 22520.1	1785	388	4	42.8	7.55	1043496 81	1043514 65	-
SrDOF24	Streb.5G0 16170.1	2270	472	5	51.22	5.24	7983709 4	7983936 3	+
SrDOF25	Streb.5G0 16230.1	2261	250	5	27.14	8.95	7995833 8	7996059 8	+
SrDOF26	Streb.5G0 29130.1	967	274	5	30.5	7.63	1067326 57	1067336 23	-
SrDOF27	Streb.6G0 17550.1	1642	170	6	19.16	8.92	7341379 9	7341544 0	-
SrDOF28	Streb.6G0 30190.1	1257	323	6	35.68	7.69	9389327 2	9389452 8	-
SrDOF29	Streb.6G0 32520.1	929	217	6	22.83	7.68	9773196 0	9773288 8	+
SrDOF30	Streb.6G0 32880.1	2118	392	6	43.1	9.62	9836532 4	9836744 1	+
SrDOF31	Streb.6G0 33230.1	1235	198	6	22.16	8.63	9886622 2	9886745 6	+
SrDOF32	Streb.7G0 19570.1	2510	380	7	42.3	8.02	8646107 5	8646358 4	+
SrDOF33	Streb.7G0 22060.1	2100	342	7	37.96	6.07	9216191 6	9216401 5	-
SrDOF34	Streb.8G0 09390.1	1740	340	8	37.88	5.99	2172728 1	2172902 0	+
SrDOF35	Streb.8G0 09670.1	1743	193	8	21.58	9.63	2238748 8	2238923 0	+
SrDOF36	Streb.8G0 15810.1	789	313	8	34.33	8.72	5045501 9	5045580 7	-
SrDOF37	Streb.8G0 21070.1	618	241	8	27.43	9.43	7603118 3	7603180 0	-
SrDOF38	Streb.9G0 19520.1	789	295	9	32.5	9.11	3715161 6	3715240 4	-
SrDOF39	Streb.9G0 19870.1	1709	310	9	34.27	8.86	3831447 5	3831618 3	-
SrDOF40	Streb.9G0 22230.1	1751	264	9	29.96	8.84	4737623 3	4737798 3	-
SrDOF41	Streb.10G 007210.1	4544	481	10	52.16	5.33	1564065 8	1564520 1	+
SrDOF42	Streb.10G 029830.1	1975	481	10	52.16	5.33	9576453 6	9576651 0	-
SrDOF43	Streb.11G 010900.1	2726	335	11	36.1	9.72	7307636 0	7307908 5	-

SrDOF44	Streb.11G 010920.1	1652	335		11	36.1	9.72	7325617 5	7325782 6	-
SrDOF45	Streb.11G 011370.1	762	262		11	29.39	4.79	7547643 9	7547720 0	-
SrDOF46	Streb.11G 012520.1	1208	131		11	14.67	9.71	7908446 2	7908566 9	-
SrDOF47	Streb.11G 014700.1	600	262		11	29.53	8.76	8590932 3	8590992 2	-
SrDOF48	Streb.11G 014810.1	600	290		11	32.4	8.74	8640245 5	8640305 4	+
SrDOF49	Streb.11G 023590.1	981	261		11	29.05	7.65	1039386 47	1039396 27	-
SrDOF50	Streb.Cont ig03291G 000010.1	1696	335	Contig03291		36.1	9.72	8346	10041	+
SrDOF51	Streb.Cont ig03842G 000020.1	2510	481	Contig03842		52.16	5.48	61036	63545	+
SrDOF52	Streb.Cont ig05355G 000010.1	617	161	Contig05355		18.46	8.59	14294	14910	+
			291.5			32.6	7.9			

Chromosomal map chart

The chromosomal map chart is shown in figure 1.

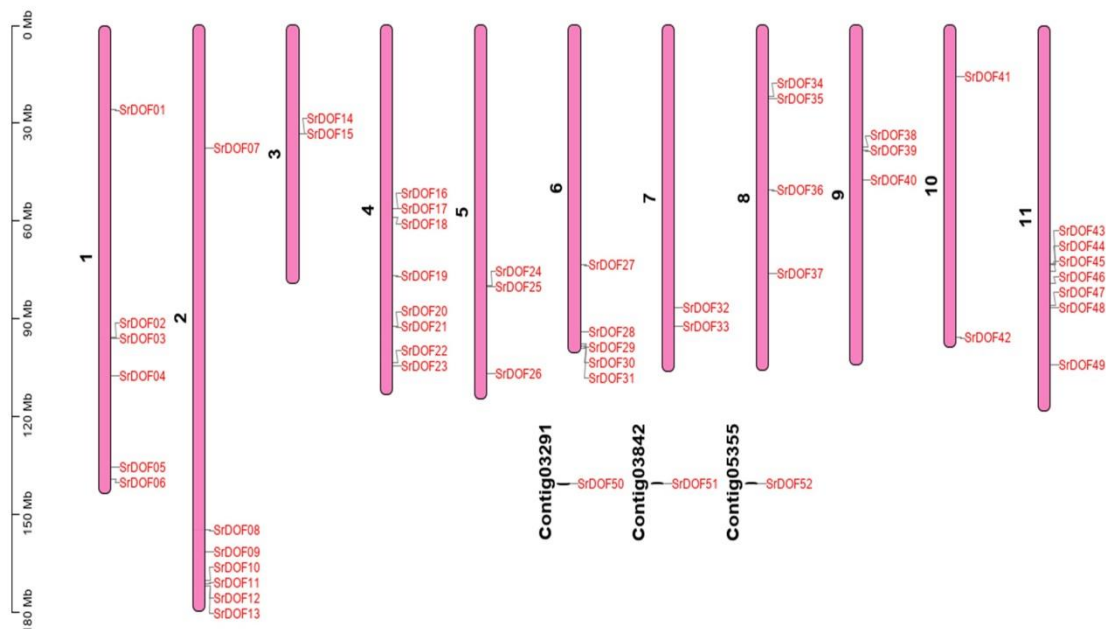


Figure 1: Chromosome Map of 52 SrDOF Gene

Phylogenetic Tree Analysis

Phylogenetic tree analysis was used to explain the evolutionary History of DOF genes. Phylogenetic tree analysis of 52 SrDOF proteins clustered into 4 major groups as shown in figure. Group1 contain highest number of 18 sequences. SrDO18 & SrDOF20 are closely clustered in Group1 and located on same chromosome4. SrDOF41 & SrDOF42 occur under same parent node and located on same chromosome10.

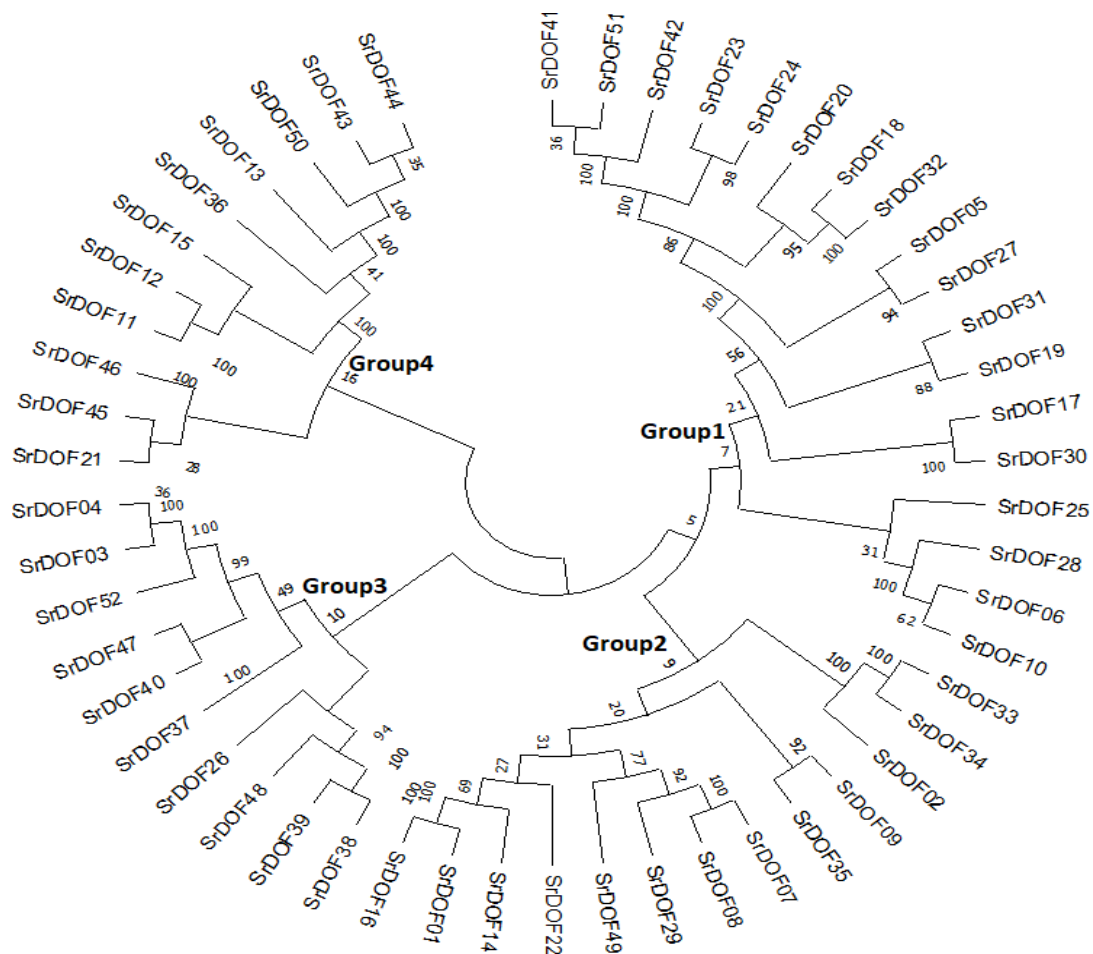


Figure 2: Phylogenetic Tree of 52 SrDOF Protein Sequences

SrDOF 51 is not located on any chromosome but on a separate scaffold (contig03842) DNA also clustered in Group1. Group2 contain 13 SrDOF transcription factors. SrDOF07 & SrDOF08 comes under same parent node and distantly located on same chromosome2. Group3 contain

lowest number of SrDOF genes i.e., 10. SrDOF38 & SrDOF39 are closely located on chromosome9 and clustered as sister clades under same parent nodes. SrDOF03 and SrDOF04 comes under same parent node and located on same chromosome1. SrDOF52 gene is not clustered on any of the *S. rebaudiana* chromosome but on a separate scaffold (contig05355). Group4 contain 11 SrDOF sequences. SrDOF11-13 and SrDOF43-46 gene clusters that are located on chromosome2 and 11, comes under Group4. SrDOF50 that is located on separate scaffold (contig03291) also comes under group4.

Gene Structure of SrDOF Gene Family

Gene Structure Display Server program was used to analyze the intron-exon organization using genomic and coding sequences of the SrDOF gene (18). The gene structure of 31 SrDOF genes is shown in figure. The majority of the SrDOF genes had a one-exon gene structure, which is consistent with DOF gene structures found in other plants (19). Majority of SrDOF genes are intron less genes. Only SrDOF05, 40, 50, 51 and 52 genes contain an intron in their gene structures. All SrDOF genes are less than 5kb but Only SrDOF08 is of 19kb Figure.3.

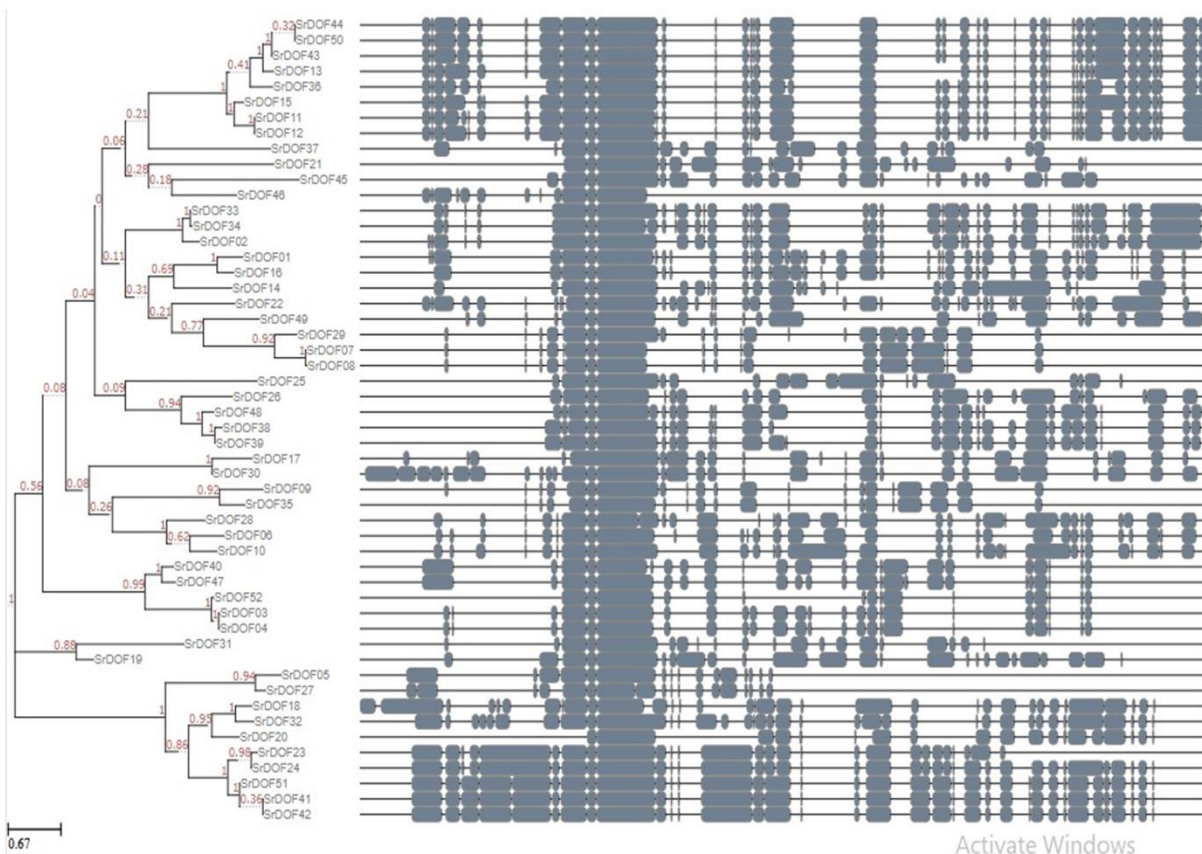


Figure 3: Alignment view of 52 SrDOF Transcription Factors

The gene structure has been shown in figure 4.

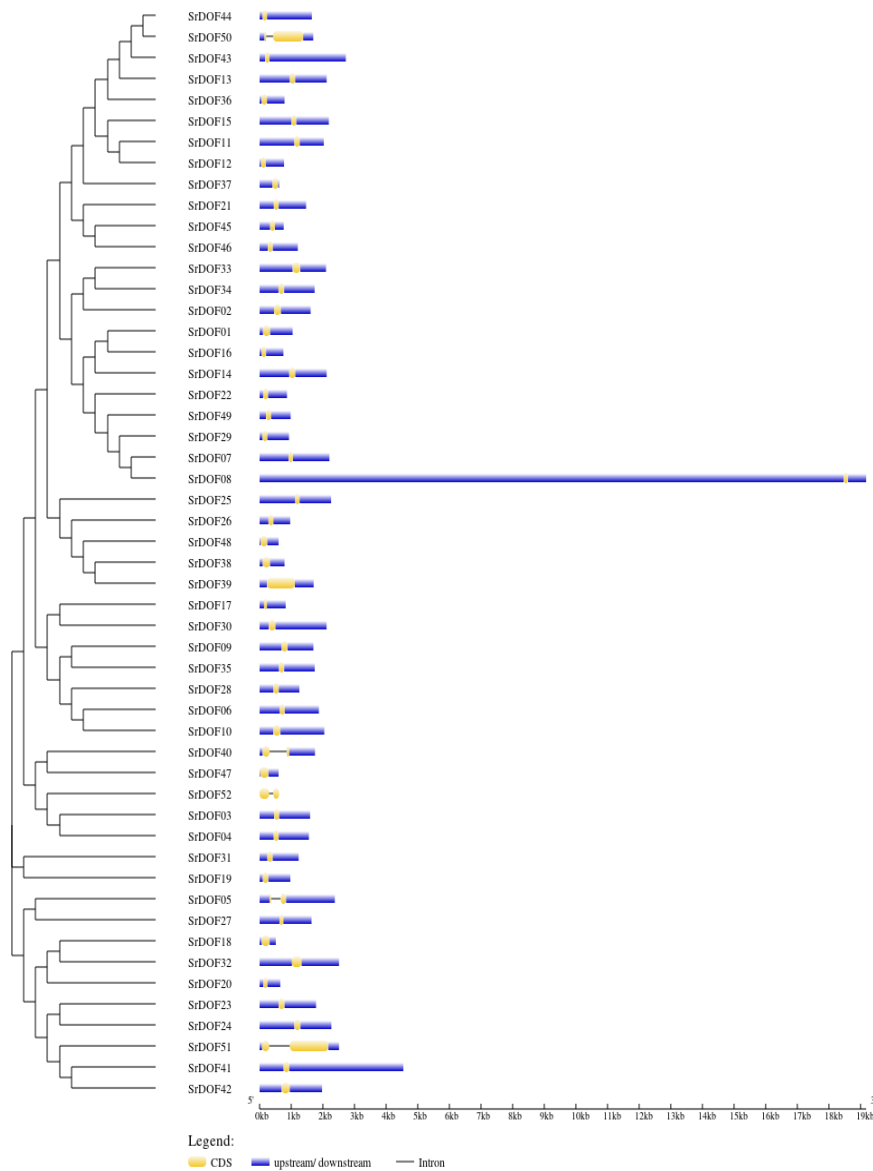


Figure 4: Gene Structure of 52 SrDOF Genes

Motif Analysis of 52 SrDOF Transcription factors

In addition, 15 conserved motifs were observed in all these proteins. Most of the SrDOF members in the same subfamily shared some conserved motifs, reflecting the functional conservation of these proteins. Motif1 and 2 present in most of the SrDOF proteins in dimer form. SrDOF17 & SrDOF20 do not contain motif 1. The motif analysis is given in the Figure 5.

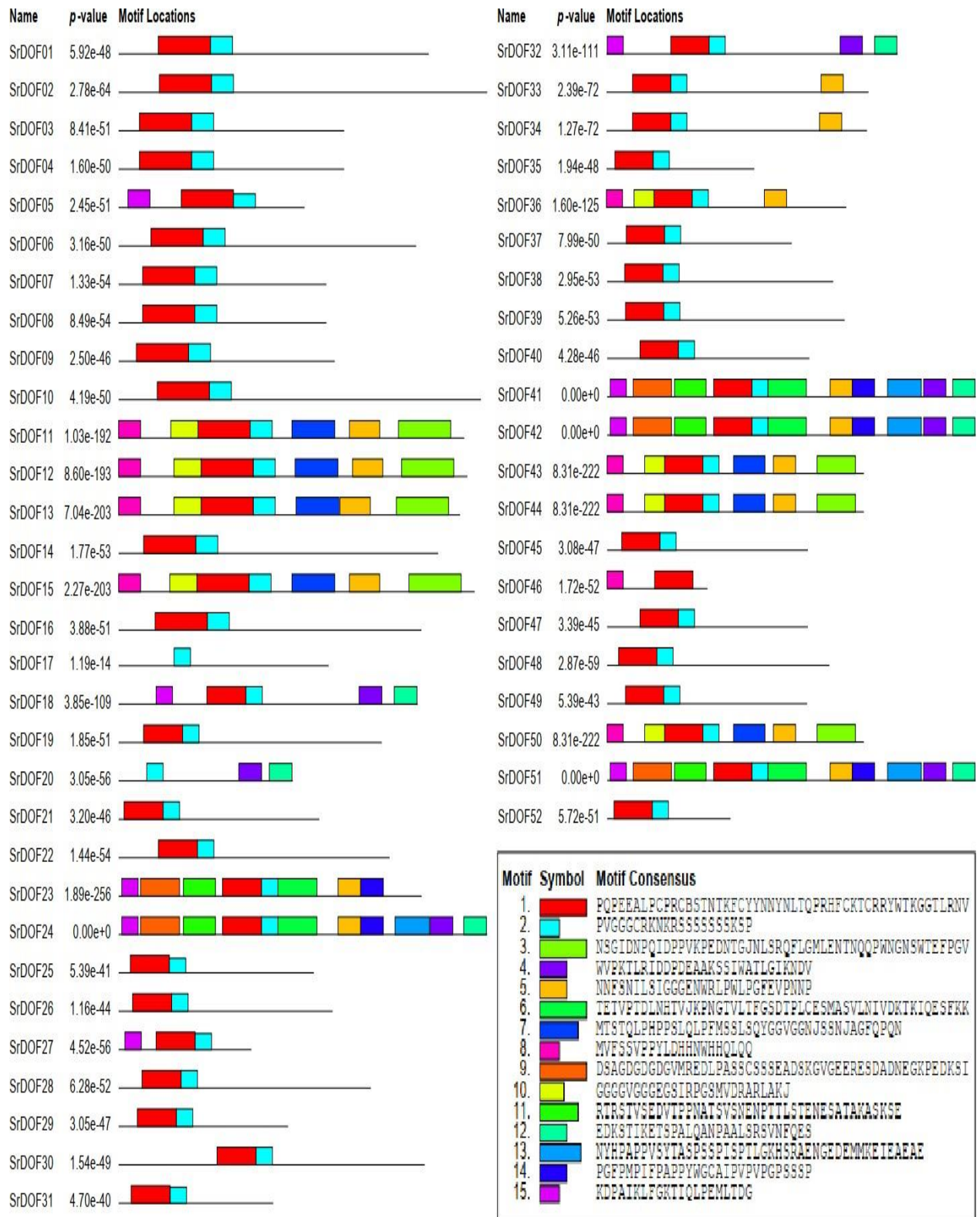


Figure 5: Motif Analysis of 52 SrDOF Family

Subcellular Location of the DOF Transcription Factor Proteins:

As predicted by the WoLFPSORT (<https://wolfpsort.hgc.jp/>) tools most of the proteins are localized inside the nucleus while other exist in the extracellular space and chloroplast respectively as shown in the Figure 6.

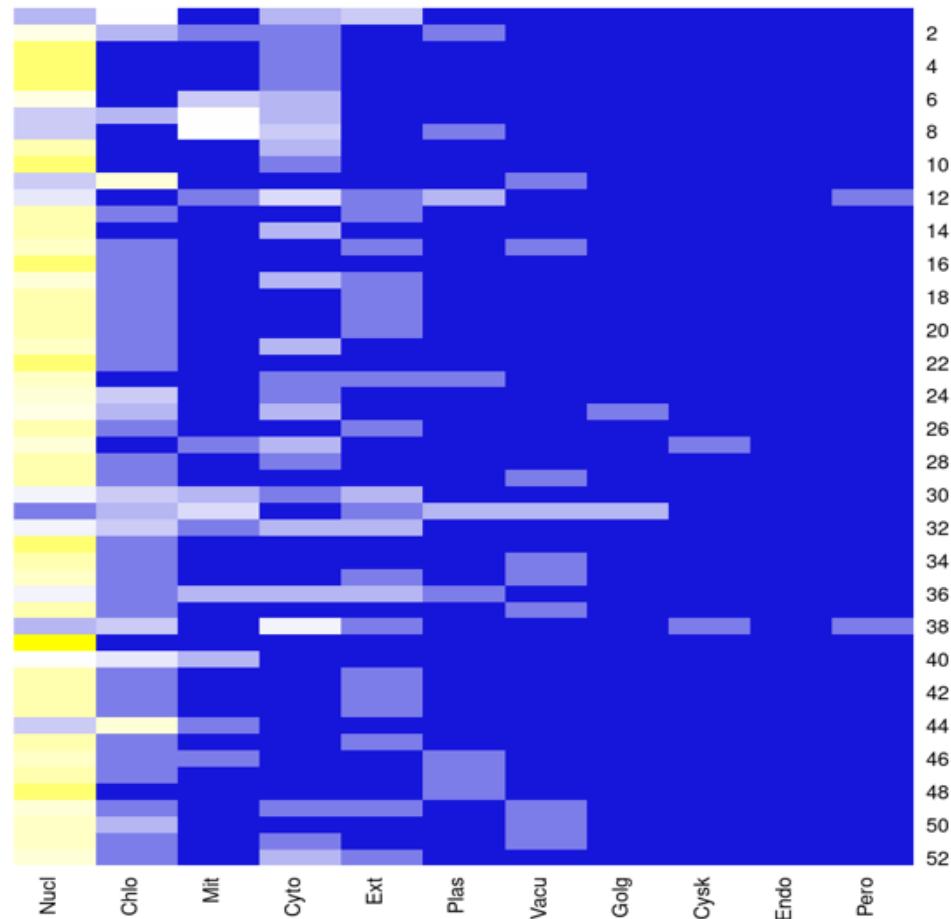


Figure 6: Subcellular Localization probability of DOF Transcription Factor Proteins

Conclusion

It is concluded that, we identified 52 DOF transcription factor genes and their physical position on the chromosomes, evolutionary history, gene structure, encoded proteins, subcellular location and motif were identified. Results discovered that gene structure and motif distribution characteristics were generally maintained across subfamilies. More than 15 conserved motifs were observed in all these proteins. The Zf Dof Superfamily domain is found in all DOF proteins except SrDOF 17 & 20.

Acknowledgment

Authors are thankful to the department for completing the work.

Author's contribution

All authors contributed equally in the manuscript.

Funding

None

Conflict of interest

None

References:

1. Fulton DL, Sundararajan S, Badis G, Hughes TR, Wasserman WW, Roach JC, et al. TFCat: the curated catalog of mouse and human transcription factors. *Genome biology*. 2009;10:1-14.
2. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. *Cell*. 2013;152(6):1237-51.
3. Fong AP, Tapscott SJ. Skeletal muscle programming and re-programming. *Current opinion in genetics & development*. 2013;23(5):568-73.
4. Takahashi K, Yamanaka S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nature reviews Molecular cell biology*. 2016;17(3):183-93.
5. Yang Y, He Z, Bing Q, Duan X, Chen S, Zeng M, et al. Two Dof transcription factors promote flavonoid synthesis in kumquat fruit by activating C-glucosyltransferase. *Plant Science*. 2022;318:111234.
6. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular plant*. 2020;13(8):1194-202.
7. Yanagisawa S. The Dof family of plant transcription factors. *Trends in plant science*. 2002;7(12):555-60.
8. Gupta S, Malviya N, Kushwaha H, Nasim J, Bisht NC, Singh V, et al. Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. *Planta*. 2015;241:549-62.
9. Lijavetzky D, Carbonero P, Vicente-Carbajosa J. Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC evolutionary biology*. 2003;3:1-11.
10. Zhou M, Guo S, Tian S, Zhang J, Ren Y, Gong G, et al. Overexpression of the watermelon ethylene response factor CIERF069 in transgenic tomato resulted in delayed fruit ripening. *Horticultural Plant Journal*. 2020;6(4):247-56.
11. Yu H, Ma Y, Lu Y, Yue J, Ming R. Expression profiling of the Dof gene family under abiotic stresses in spinach. *Scientific reports*. 2021;11(1):14429.
12. Nan H, Ludlow RA, Lu M, An H. Genome-wide analysis of Dof genes and their response to abiotic stress in rose (*Rosa chinensis*). *Frontiers in Genetics*. 2021;12:538733.
13. Xu X, Yang Y, Zhang T, Zhang Y, Tong H, Yuan H. Systematic analysis of the R2R3-MYB transcription factor gene family in *Stevia rebaudiana*. *Industrial Crops and Products*. 2024;210:118123.
14. Chen M, Liu X, Huan L, Sun M, Liu L, Chen X, et al. Genome-wide analysis of Dof family genes and their expression during bud dormancy in peach (*Prunus persica*). *Scientia horticulturae*. 2017;214:18-26.
15. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*. 2021;38(7):3022-7.
16. Guo A-Y, Zhu Q-H, Chen X, Luo J-C. GSDS: a gene structure display server. *Yi chuan= Hereditas*. 2007;29(8):1023-6.
17. Hardigan MA, Feldmann MJ, Lorant A, Bird KA, Famula R, Acharya C, et al. Genome synteny has been conserved among the octoploid progenitors of cultivated strawberry over millions of years of evolution. *Frontiers in plant science*. 2020;10:1789.
18. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 2015;31(8):1296-7.
19. Corbin C, Drouet S, Markulin L, Auguin D, Lainé É, Davin LB, et al. A genome-wide analysis of the flax (*Linum usitatissimum* L.) dirigent protein family: from gene identification and evolution to differential regulation. *Plant molecular biology*. 2018;97:73-101.