

Genome-wide identification, characterization and functional prediction of the Dof gene family in sesame (*Sesamum indicum* L.)

Md Nahid Hasan

Sher-e-Bangla Agricultural University, Dhaka, Bangladesh

Khandker Shazia Afrin

Genome Research Centre, Bangladesh Jute Research Institute (BJRI), Dhaka

Farjana Afroz Susmi

Sher-e-Bangla Agricultural University, Dhaka, Bangladesh

Tasmina Islam Simi

Sher-e-Bangla Agricultural University, Dhaka, Bangladesh

Md Abdur Rahim

rahimgpb@sau.edu.bd

Sher-e-Bangla Agricultural University, Dhaka, Bangladesh <https://orcid.org/0000-0001-7687-9347>

Research Article

Keywords: SiDof gene family, transcription factor, characterization, sesame

Posted Date: November 27th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-5529029/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Genome-wide identification, characterization and functional prediction of the *Dof* gene family in sesame (*Sesamum indicum* L.)

Md Nahid Hasan¹, Khandker Shazia Afrin², Farjana Afroz Susmi¹, Tasmina Islam Simi¹ and Md Abdur Rahim^{1*}

¹Department of Genetics and Plant Breeding, Sher-e-Bangla Agricultural University, Dhaka-1207, Bangladesh

²Genome Research Center, Bangladesh Jute Research Institute, Dhaka-1207, Bangladesh

*Corresponding author: rahimgepb@sau.edu.bd

Department of Genetics and Plant Breeding, Sher-e-Bangla Agricultural University, Dhaka-1207, Bangladesh

The DNA binding with one 'Zn²⁺' finger (*Dof*) is the plant-specific TF family that includes a highly conserved *Dof* domain. It plays a vital role in plant growth and development, and environmental stimuli. The *Dof* gene family is well characterized in many plants, including *Arabidopsis*, but has not yet been studied in sesame. In the present study, we reported for the first-time genome-wide identification of the *Dof* gene family in sesame, resulting in 32 *SiDof* genes which are located across 12 linkage groups (LGs). The *SiDof* genes comprised one to five exons and zero to four introns which is comparable with *Dof* genes in other species. Of 32 *SiDof* genes, about 71.88% genes were intronless. The *SiDof* genes were classified into six groups and exhibited evolutionary relationships with other species. Group III contained the highest number of *SiDof* genes (seven) while the lowest number in group V (three). There was a robust protein-protein interaction between NAC071 and two *Dof* proteins namely DOF5.3 and DOFF5.6. A total of 1668 'cis-regulatory' elements were detected in the promoter region of 32 *SiDof* genes. Among these, 733 were linked to the stress responses, 562 to the light responses, and 373 to the phytohormone response. The accessible transcriptome ('RNA-seq') data-based expression profiles shown that *SiDof* genes might be associated with various abiotic stresses, including drought and salinity in sesame. Nonetheless, further investigation on functionality analysis of candidate *SiDof* genes is required.

Keywords: *SiDof* gene family, transcription factor, characterization, sesame

Abbreviations

AA: Amino acid;

ABA: Absciscic acid

FPKM: Fragments per kilobase of transcript per million mapped reads

GRAVY: Grand average of hydropathicity

h: hour

Ka: Non-synonymous substitution rate

Ks: Synonymous substitution rate

MW: Molecular weights

LG: Linkage group

pI: Isoelectric points

RPKM: Reads per kilobase of transcript per million mapped reads

SiDof: *Sesamum indicum* DNA binding with one 'Zn²⁺' finger

TF: Transcription factor

Introduction

Transcriptional factors (TFs) play a vital role in gene expression and regulate many significant biological processes. Among these TFs, Dof (the DNA binding with one 'Zn²⁺' finger) TF family belongs to the 'Zn²⁺ finger' super-family group and is specific to plants (Yanagisawa, 1995). Dof proteins are consist of 200–400 amino acid (aa) residues with a highly conserved 50-52 aa residues that comprise a single 'C2C2-type zinc-finger' like structure (Yanagisawa and Sheen, 1998). They are consisted of two vital domains: a conserved 'DNA-binding' domain at the N-terminus and a C-terminal domain which is highly variable (Diaz et al., 2002). Furthermore, a bipartite nuclear localization signal (NLS) is present in Dof proteins which partially overlaps with the Dof domain, as well as a transcriptional activation domain located at the C-terminal (Yanagisawa, 2002). Many plant processes are influenced by the members of the Dof TF family, such as growth and development, seed germination, cell cycle, osmotic control, photosynthesis, leaf polarity, cryptochrome and phytochrome signaling, storage root growth, cell wall, nitrogen use efficiency, and defense responses (Yang et al., 2012) (Iwamoto et al., 2009; Kim et al., 2010; Kumar et al., 2009; Shuichi Yanagisawa, 2004; Skirycz et al., 2008; Tanaka et al., 2009; Wang et al., 2017; Yanagisawa, 2002; Yanagisawa and Sheen, 1998; Yang et al., 2011).

The members Dof TF family vary in different plant species. It came to light that maize contained the very first Dof domain protein (Yanagisawa and Sheen, 1998). To date, several

potential *Dof* genes have been reported in a wide range photosynthetic species. For example, there are 30, 36, 28, 76, 34, 26, 78, and 46 *Dof* genes in rice, *Arabidopsis*, sorghum, Chinese cabbage, tomato, barley, soybean, and maize, respectively (Cai et al., 2013; Chen and Cao, 2015; Kushwaha et al., 2011; Lijavetzky et al., 2003; Ma et al., 2015; Moreno-Risueno et al., 2007; Wang et al., 2007).

Sesame (*Sesamum indicum* L.; $2n = 26$) is one of the most important oil crops used and produced by humans around the world and a member of the Pedaliaceae family (Wei et al., 2022). It is cultivated in tropics, semi-tropics as well as temperate regions of the world (Abdiani et al., 2024). This crop has a resilient adaptation to water deficit conditions (Dossa et al., 2019). The oil content of sesame is very high (nearly 35-63%) compared to other oil-bearing seeds (Kadkhodaie et al., 2014). Besides, it consisted of various minerals, vitamins (like vitamin A, vitamin E, thiamine, riboflavin, niacin, pantothenic acid, folic acid, and ascorbic acid), fiber, protein (20-27%), carbohydrates (20-25%), and antioxidant lignans like sesamol and sesamin (Fasuan et al., 2018; Wei et al., 2022). Globally, around 70% of sesame is used for oil extraction, and 30% for food, beverage, and confectionary industries (Baraki et al., 2020). Sesame oil confers several health benefits, including lowering the cholesterol level, reducing the blood pressure, and anticancer agents (Abdiani et al., 2024; Anilakumar et al., 2010; Elleuch et al., 2011; Wei et al., 2022).

Recent studies have demonstrated that different *Dof* gene family members play an important role in stress responses (biotic and abiotic stress) in several plant species (Corrales et al., 2017; Ma et al., 2015; Nan et al., 2021; Yang et al., 2017; Zhang et al., 2018; Zhou et al., 2020). For example, overexpression of *AtCDF3* (a *Dof* gene) in *Arabidopsis* exhibited the enhanced resistance to osmotic stress, cold, and drought but delayed in flowering (Corrales et al., 2017). Moreover, overexpression of tomato *SICDF1* and *SICDF3* in *Arabidopsis* showed enhanced the salinity and drought tolerance, and delayed in flowering via modulation of flowering genes like CO and FT (Corrales et al., 2014). In watermelon, *CIDof* genes exhibited variations in their expression pattern under salt stress (Zhou et al., 2020). Likewise, *OsDof1* gene results in cold tolerance in rice (Liu et al., 2021). However, to date, *Dof* genes are not studied in sesame. Therefore, this study was undertaken to identify, characterize the potential candidate *Dof* genes along with their physicochemical properties, domains and motifs, syntenic association, duplication events, and expression patterns in sesame using an *in silico* approach.

Materials and methods

Identification of *Dof* genes in sesame genome

The aa sequences of *Dof* gene family members were retrieved from the ‘TAIR’ (<https://www.arabidopsis.org/>) and performed the ‘BLASTP’ search against the sesame genome in ‘Ensembl Plants’ (<https://plants.ensembl.org/index.html>) database. Further, the aa sequence of *Dof* proteins of different plant species published reports, including *Arabidopsis thaliana* (Lijavetzky et al., 2003), *Glycin max* (Guo and Qiu, 2013), *Zea mays* (Chen and Cao, 2015), *Solanum lycopersicum* (Cai et al., 2013), *Oryza sativa* (Liu et al., 2021) and *Medicago truncatula* (Shu et al., 2015) and ‘BLASTP’ searched against the sesame genome to identify candidate *Dof* genes. Furthermore, the putative sesame *Dof* genes were confirmed with the presence of *Dof* (DNA binding with one finger; pfam02701) domains individually with the ‘HMMER’ web tool which considers ‘Hidden Markov Models (HMM)’ (Potter et al., 2018) as well as ‘Pfam’ (<http://pfam.xfam.org>).

Phylogenetic analysis and protein sequence alignment of *SiDof* genes

A multiple sequence alignment was built and a circular phylogenetic tree was constructed with the aa sequences *Dof* genes of *S. indicum*, *A. thaliana*, *G. max*, *Z. mays*, *S. lycopersicum*, *O. sativa*, and *M. truncatula* by using MEGA11 (Tamura et al., 2021) to understand the evolutionary relationship among these species. The aa sequences of *Dof* genes of these species were aligned with ‘ClustalW’ and created a circular phylogenetic tree by the ‘Neighbor-Joining (NJ)’ method with 1000 bootstrap with default parameters. Finally, the circular tree was further improved with the tool ‘iTOL’ (<https://itol.embl.de/>).

Molecular characterization of *SiDof* genes

The gene length (bp), protein length (aa), and chromosomal location of the *SiDof* genes were predicted using the ‘Ensembl Plants’ database. Further, the ‘molecular weight’ (MW), ‘isoelectric point’ (pI) as well as ‘grand average of hydropathicity’ (GRAVY) of *SiDof* proteins were determined using the ‘ProtParam tool’ (<https://web.expasy.org/protparam/>). The subcellular localization of *SiDof* genes was predicted by the web-based tool ‘Plant-mPLoc’ (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>). The exon-intron structure of *SiDof* genes was predicted with the ‘Gene Structure Display Server (GSDS2.0)’ tool (Hu et al., 2015). Moreover, the ‘conserved domains’ and ‘motifs’ of *SiDof* proteins were analyzed using ‘CD-search’ tool of ‘Conserved Domain Database (CDD)’ (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) and ‘MEME’ (<https://meme-suite.org/tools/meme>), respectively. The genomic positions of *SiDof* genes were visualized

using ‘MapChart’ (Voorrips, 2002). The ‘3D structure’ of SiDof proteins was analyzed by the ‘SWISS-MODEL’ (<https://swissmodel.expasy.org/interactive#sequence>). The ‘cis-regulatory elements’ of *SiDof* genes were analyzed with ‘PlantCARE’ (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) within the 3.0-Kb upstream region of the transcription start site. The protein-protein interaction was performed by ‘STRING’ (<https://string-db.org/>). The synteny relationship of *SiDof* genes was analyzed with the ‘Circoletto’ web tool (Darzentas, 2010). The Ka and Ks values, and Ka/Ks ratio for *SiDof* gene pairs were determined using ‘TBtools’ (Yang and Nielsen, 2000).

Genetic variation, transcriptome data and expression pattern of *SiDof* genes

The *SiDof* genes were checked to know the presence of genetic variants like simple sequence repeat (SSR) with the ‘Web SSR Finder’ tool of ‘Plant Simple Sequence Repeat Database (PSSRD)’ (Song et al., 2021). The transcriptome data (FPKM/RPKM expression) was collected from ‘Gene Expression Omnibus (GEO)’ database of NCBI under the accession number ‘GSE164511’ (young leaf), ‘GSE148340’ (control vs PEG-treated root) and shoot under salinity condition (tolerant vs sensitive) (Zhang et al., 2020).

Results

Identification of putative *Dof* genes in sesame genome

The *Dof* gene family members in sesame were identified through the ‘BLASTP’ search. The protein sequences of *AtDof* genes were ‘BLASTP’ searched against the sesame genome in ‘Ensembl Plants’ database (<https://plants.ensembl.org/index.html>). Likewise, the reported *Dof* protein sequences of other plant species were also “BLASTP” searched against sesame genome and cross checked. We identified total 32 unique *SiDof* genes in sesame genome (Table 1 and Supplementary file 1). Besides, the protein sequences were validated with ‘HMMER’ and ‘Pfam’.

Phylogenetic analysis of *SiDof* genes

A phylogenetic tree (circular) was built (Fig. 1) using the *Dof* proteins from various plant species, including *S. indicum*, *A. thaliana*, *G. max*, *Z. mays*, *S. lycopersicum*, *O. sativa*, and *M. truncatula* to understand the relationships among the species throughout the evolutionary process. Total 32 *SiDof* proteins were distributed into six groups. Group III consisted of the

maximum seven *SiDof* proteins, including *SiDof2* (SIN_1003692), *SiDof4* (SIN_1004081), *SiDof5* (SIN_1004940), *SiDof7* (SIN_1005291), *SiDof9* (SIN_1005551), *SiDof18* (SIN_1011214), and *SiDof 29* (SIN_1025235) whereas group V contained minimum three *SiDof* proteins like *SiDof3* (SIN_1003703), *SiDof19* (SIN_1011668) and *SiDof24* (SIN_1019847). The groups such as group I, IV, and VI had six *SiDof* proteins in each. Group I included *SiDof6* (SIN_1005243), *SiDof8* (SIN_1005292), *SiDof20* (SIN_1013094), *SiDof21* (SIN_1013324), *SiDof30* (SIN_1027039) and *SiDof31* (SIN_1027065). Group IV contained *SiDof10* (SIN_1005860), *SiDof14* (SIN_1009212), *SiDof25* (SIN_1020428), *SiDof26* (SIN_1021128), *SiDof27* (SIN_1024161), and *SiDof28* (SIN_1024519). Group VI consisted *SiDof1* (SIN_1001405), *SiDof11* (SIN_1006007), *SiDof12* (SIN_1006579), *SiDof13* (SIN_1008018), *SiDof23* (SIN_1018780), and *SiDof32* (SIN_1027102). Group II contained four *SiDof* proteins like *SiDof 15* (SIN_1009420), *SiDof16* (SIN_1009530), *SiDof17* (SIN_1010375) and *SiDof22* (SIN_1017363).

Physicochemical properties, structure, conserved domain and motif analysis of *SiDof* genes

The different physicochemical properties of 32 *SiDof* genes, such as gene sequence length (bp), protein size (aa), ‘grand average of hydropathicity (GRAVY)’, ‘subcellular localization’ and ‘theoretical isoelectric point (pI)’ were presented in Table 1. The result showed that the length of 32 *SiDof* genes ranged from 354 bp to 12742 bp, whereas the length of the proteins varied from 117 aa to 719 aa. The minimum (117 aa) and maximum (719 aa) protein lengths were found for *SiDof10* (SIN_1005860) and *SiDof11* (SIN_1006007) genes, respectively. The maximum (10.4) and minimum (4.47) "isoelectric point" (pI) were observed for *SiDof24* (SIN_1019847) and *SiDof21* (SIN_1013324), respectively. The molecular weights of 32 *SiDof* proteins varied from 12583.67 kDa (*SiDof10*) to 78342.23 kDa (*SiDof11*). Likewise, the "GRAVY" values of 32 *SiDof* proteins exhibited negative which ranged from -1.03 (*SiDof24*) to -0.28 (*SiDof20*), indicating their hydrophilic properties. Moreover, the subcellular localization analysis revealed that all *SiDof* genes were localized in the nucleus.

The exon-intron structure of the 32 *SiDof* genes was predicted using the "GSDS2.0" algorithm. The results showed that exon number varied from one to five while introns number varied zero to four for *SiDof* genes (Fig. 2). Among 32 *SiDof* genes, maximum 23 genes had single exons such as *SiDof3* (SIN_1003703), *SiDof4* (SIN_1004081), *SiDof5* (SIN_1004940), *SiDof6* (SIN_1005243), *SiDof7* (SIN_1005291), *SiDof9* (SIN_1005551), *SiDof10* (SIN_1005860), *SiDof15* (SIN_1009420), *SiDof16* (SIN_1009530), *SiDof17* (SIN_1010375),

SiDof18 (SIN_1011214), *SiDof19* (SIN_1011668), *SiDof20* (SIN_1013094), *SiDof21* (SIN_1013324), *SiDof22* (SIN_1017363), *SiDof24* (SIN_1019847), *SiDof25* (SIN_1020428), *SiDof26* (SIN_1021128), *SiDof27* (SIN_1024161), *SiDof28* (SIN_1024519), *SiDof29* (SIN_1025235), *SiDof30* (SIN_1027039) and *SiDof32* (SIN_1027102). *SiDof11* (SIN_1006007) gene contained the maximum five exons whereas seven *SiDof* genes, including *SiDof1* (SIN_1001405), *SiDof2* (SIN_1003692), *SiDof8* (SIN_1005292), *SiDof12* (SIN_1006579), *SiDof13* (SIN_1008018), *SiDof14* (SIN_1009212), *SiDof23* (SIN_1018780) and *SiDof31* (SIN_1027065) had only two exons.

The "conserved domains" and "motifs" for SiDof proteins were determined using the "MEME" suite and the "CD-search" tool of the "National Center for Biotechnology Information (NCBI)" (Fig. 3-5 and Table 2-3). The results showed that a highly conserved "DNA-binding with one finger" (zn-Dof) domain protein with 52 aa residues was found in all SiDof proteins. These proteins are transcription factors (TFs) with a single "Cys2/Cys2 zinc finger" (pfam02701) present in plants and a highly conserved "DNA-binding domain. Furthermore, SiDof11 exhibited an additional domain "GDSL-like Lipase/Acylhydrolase" (pfam00657). Likewise, SiDof proteins exhibited 10 conserved "motifs" having more than 17 aa residues (Table 3 and Figure 4.) For all SiDof proteins, the Dof domains shared motif 1 which is the majority of these motifs (Fig. 5). Moreover, the sequences alignment of SiDof proteins revealed that all 32 proteins have motif 1 and conserved amino acid residues in the Dof domain (Fig. 6).

Analysis of the structure of SiDof proteins and protein-protein interactions

Utilizing the protein sequences that were retrieved from "Ensembl Plants", a three-dimensional structure was generated for each 32 SiDOF protein using an online tool "SWISS-MODEL" (Fig. 7). The results showed that the structure of each SiDOF protein was disordered. After a thorough analysis of the structures, it was found that practically all of the SiDOFs had prominent β -sheets while the α -helices were less noticeable. For instance, the percentage of β -sheets was considerably greater than that of α -helices from SiDOF1 to SiDOF9. Again, the α -helix was entirely missing in SiDOF10. Furthermore, the following six proteins - SiDOF20, SiDOF22, SiDOF23, SiDOF29, SiDOF30, and SiDOF31 showed a plethora of β -sheets but relatively few α -helices. Regarding SiDOF27 and SiDOF32, both proteins had an almost identical structure that was covered in numerous α and β -sheets. The remaining structures exhibited less accurate α -helices and more pronounced β -sheets. Further, using the *A. thaliana* protein information (Supplementary file 2) and the web program 'STRING' (<https://string->

db.org/), the SiDOF protein-protein interaction network was built (Fig. 8). The analysis indicates a robust interaction between NAC071 and two proteins namely DOF5.3 and DOF5.6. Once more, DOF2.2, DOF2.4, DOF5.1, and DOF5.8 are strongly correlated with another transcription factor, NAC028 (NAC domain-containing protein 28). Furthermore, MDC16.10 (Protein BIG GRAIN-like A) interacts with NAC028 and NAC086 in addition to other DOF proteins, including DOF2.2, DOF2.4 and DOF5.6. MDC16.10 is a member of the BIG GRAIN 1 plant protein family that controls the auxin signaling pathway (*Protein BIG GRAIN 1-like A*). Finally, a single interaction was observed between DOF3.1 and F13P17.18 proteins.

Position of *SiDof* genes in the sesame genome

The chromosomal locations of 32 *SiDof* genes were investigated (Table 1) and the "MapChart" tool was used to illustrate the results (Fig. 9). According to the results, the 32 *SiDof* genes were found to be distributed across 12 distinct 'linkage groups' (LGs). For example, *SiDof15* and *SiDof16* situated in LG1; *SiDof21* and *SiDof26* in LG2; *SiDof1*, *SiDof25*, *SiDof17*, and *SiDof22* in LG3; *SiDof23* and *SiDof10* in LG4, and the LG5 and LG13 contained only one *SiDof* gene in each *SiDof9* and *SiDof4*, respectively. Then there were two *SiDof* genes in LG7 (*SiDof14* and *SiDof19*). Furthermore, LG8 included a maximum of six *SiDof* genes (*SiDof11*, *SiDof24*, *SiDof12*, *SiDof30*, *SiDof31* and *SiDof32*). Again, *SiDof27* and *SiDof28* were located in LG9; *SiDof18* and *SiDof20* in LG11; *SiDof6*, *SiDof7* and *SiDof8* in LG12. Lastly, *SiDof2*, *SiDof3*, *SiDof5*, *SiDof13* and *SiDof29* were in LG15.

Syntenic relationship and gene duplication analysis of *SiDof* genes

The syntenic relationship of *SiDof* genes with *A. thaliana*, *O. sativa*, *G. max*, and *O. sativa* was analyzed using a web tool 'Circoletto' to understand the orthologous relationship and evolutionary conservation among these species (Fig. 10 and Supplementary file 3). The results showed that *SiDof* genes had synteny with 16 *A. thaliana* genes, 19 *G. max* genes, and 15 *O. sativa* genes. In addition, 32 *SiDof* genes were subjected to gene duplication analysis. A total of eleven pair of genes with high sequence similarity and within the same clade were identified through the phylogenetic analysis (Fig. 1) signifying the possibility of gene duplication events for these eleven pair of *SiDof* genes (*SiDof1*, SIN_1001405; *SiDof2*, SIN_1003692; *SiDof3*, SIN_1003703; *SiDof4*, SIN_1004081; *SiDof6*, SIN_1005243; *SiDof8*, SIN_1005292; *SiDof9*, SIN_1005551; *SiDof10*, SIN_1005860; *SiDof11*, SIN_1006007; *SiDof13*, SIN_1008018; *SiDof15*, SIN_1009420; *SiDof16*, SIN_1009530; *SiDof17*, SIN_1010375; *SiDof18*, SIN_1011214; *SiDof22*, SIN_1017363; *SiDof24*, SIN_1019847; *SiDof26*, SIN_1021128;

SiDof27, SIN_1024161; *SiDof28*, SIN_1024519; *SiDof29*, SIN_1025235; *SiDof30*, SIN_1027039; *SiDof31*, SIN_1027065). Moreover, the ‘non-synonymous divergence’ (Ka) and ‘synonymous divergence’ (Ks), and Ka/Ks ratio of duplicated *SiDof* gene pairs were analyzed with the ‘TBtools’ program (Table 5). Generally, the ratio of Ka/Ks <1, Ka/Ks = 1 and Ka/Ks >1 indicates purifying, neutral and positive selections, correspondingly (Nekrutenko et al., 2002). The Ka/Ks ratio of *SiDof* gene pairs varied from 0.1492 to 0.3323 suggested a purifying selection pressure for the evolution of *SiDof* genes.

Analysis of cis-acting elements in the promoter region of *SiDof* genes

The 3.0-Kb upstream promoter sequences for the 32 *SiDof* genes were obtained from ‘Ensembl Plants’ database. A total of 1668 *cis*-regulatory elements were found in the promoter region of *SiDof* genes using the web software ‘PlantCARE’ (Table 5 and Supplementary file 4). For the entire promoter area, three distinct categories have been identified such as light response, phytohormone response and stress response (Zhou et al., 2023). Of these, 733 were linked to the stress response (W BOX, WRE3, WUN-motif, ARE, and box S), 562 to the light response (AE-box, Box-4, G-Box, and GA-motif), and 373 to the phytohormone response (ABRE, P-box, TATC-box, and TGACG-motif). Among these, all *SiDof* genes contained Box-4 (associated with light response), and MYC (associated with stress response).

Genetic variants and expression pattern of *SiDof* genes with transcriptomic data

The microsatellite/simple sequence repeat (SSR) motifs were analyzed in the DNA sequence of 32 *SiDof* genes (Table 6). Out of 32 *SiDof* genes, only six such as *SiDof8* (SIN_1005292, *SiDof11* (SIN_1006007), *SiDof19* (SIN_1011668), *SiDof23* (SIN_1018780), *SiDof26* (SIN_1021128) and *SiDof31* (SIN_1027065) revealed SSR motifs. CT/GA, TA/AT, and GA/CT were the frequent SSR motifs ranging from mono- to trinucleotide repeats with 10 to 26 bp repeat size in the *SiDof11*. However, *SiDof8* (SIN_1005292) exhibited one di-nucleotide (TC/AG) SSR with 13 repeats while *SiDof19* (SIN_1011668) and *SiDof26* (SIN_1021128) had one tri-nucleotide CCA/GGT and ATC/TAG SSR with 5 repeats in each, respectively. *SiDof23* (SIN_1018780) had three di-nucleotides (GT/CA) SSR at different positions. Moreover, *SiDof31* (SIN_1027065) had one SSR with di-nucleotide (TC/AG) with 9 repeats and two SSRs with tri-nucleotide (TCC/AGG) with 6 repeats.

The expression pattern of 32 *SiDof* genes was analyzed in different organs of sesame-like wild type leaf and stressed conditions such as root (drought and waterlogging), shoot (salinity) with the transcriptome data (FPKM/RPKM) retrieved from ‘gene expression omnibus’ (GEO)

datasets and published articles (Fig. 11). The results showed that 27 *SiDof* genes out of 32 genes were expressed in the leaf of sesame. However, *SiDof10* (SIN_1005860), *SiDof11* (SIN_1006007), *SiDof15* (SIN_1009420), *SiDof25* (SIN_1020428), *SiDof26* (SIN_1021128) and *SiDof27* (SIN_1024161) showed higher expression in sesame leaf (Fig. 11a).

Further, the expression pattern of 32 *SiDof* genes was analyzed with the sesame root transcriptome data under PEG-treated drought stress (Fig. 11b). The results revealed that two genes like *SiDof4* (SIN_1004081) and *SiDof20* (SIN_1013094) were not expressed in roots both control and drought stress condition of both genotypes ('VEN-1' and 'TEX-1') of sesame while another two genes, including *SiDof3* (SIN_1003703) and *SiDof29* (SIN_1025235) were poorly expressed only in 'VEN-1' and 'TEX-1', respectively and their expression was very low. However, the expression of *SiDof1* (SIN_1001405), *SiDof11* (SIN_1006007), *SiDof12* (SIN_1006579), *SiDof13* (SIN_1008018), *SiDof18* (SIN_1011214), *SiDof22* (SIN_1017363), *SiDof23* (SIN_1018780), *SiDof27* (SIN_1024161), *SiDof28* (SIN_1024519) and *SiDof32* (SIN_1027102) genes was increased. Nonetheless, the expression of *SiDof1* was more than two-fold higher in drought stress conditions. On the contrary, expression of *SiDof2* (SIN_1003692), *SiDof8* (SIN_1005292), *SiDof21* (SIN_1013324), and *SiDof26* (SIN_1021128) genes was decreased in drought-stressed roots compared to the non-treated controls. However, the expression of *SiDof5* (SIN_1004940) was increased in PEG-treated roots of 'VEN-1' but decreased in 'TEX-1' as compared to controls. The *SiDof10* had maximum expression levels in both genotypes for both stressed and control conditions among all 32 *SiDof* genes. Nevertheless, its expression was decreased in and similar in both control and PEG-treated roots of 'TEX-1' in sesame.

Again, the expression of 32 *SiDof* genes was analyzed using another root transcriptome dataset of sesame under different time-point after salinity treatment as well as salinity tolerant and sensitive sesame genotypes (Fig. 11c). The result showed that the expression of *SiDof8* (SIN_1005292), *SiDof15* (SIN_1009420), *SiDof17* (SIN_1010375), *SiDof19* (SIN_1011668), *SiDof22* (SIN_1017363), *SiDof25* (SIN_1020428), *SiDof28* (SIN_1024519) genes was increased over time in roots of both sensitive and tolerant sesame genotypes. Conversely, the expression level *SiDof6* (SIN_1005243), *SiDof7* (SIN_1005291), *SiDof21* (SIN_1013324), *SiDof23* (SIN_1018780), *SiDof9* (SIN_1005551), *SiDof10* (SIN_1005860), *SiDof13* (SIN_1008018) and *SiDof30* (SIN_1027039) genes was decreased over time after (2h, 6h, 12h and 24h) salinity treatment in both sensitive and tolerant genotypes. Furthermore, the expression of *SiDof11* (SIN_1006007) increased at 6h after salinity treatment and then decreased in both genotypes. The expression of *SiDof1* (SIN_1001405) and *SiDof12*

(SIN_1006579) genes was increased at 2h while decreased at 6h and 12h, then again increased at 24h of salinity treatment of roots for both sensitive and tolerant sesame genotypes. The gene *SiDof20* (SIN_1013094) was not expressed in the root of both contrasting genotypes, but the expression of *SiDof3* (SIN_1003703), *SiDof4* (SIN_1004081), *SiDof5* (SIN_1004940) and *SiDof29* (SIN_1025235) was very low in both genotypes over time. However, *SiDof32* (SIN_1027102) gene revealed differences in expression levels between salinity sensitive and tolerant genotypes over period as higher in tolerant than in sensitive genotypes.

Discussion

The Dof proteins are plant-specific TFs in plants and play an important role in many plant processes, including plant growth and development, and responses to environmental stimuli (Shuichi Yanagisawa, 2004; Yanagisawa, 2002; Zou et al., 2019). *Dof* genes are characterized in several plant species including *Arabidopsis*, nevertheless, unexplored in sesame.

The members of *Dof* gene family varies species to species such as 30 in rice, 36 in *Arabidopsis*, 28 in sorghum, 76 in Chinese cabbage, 34 in tomato, 26 in barley, 78 in soybean and 46 in maize (Cai et al., 2013; Chen and Cao, 2015; Kushwaha et al., 2011; Lijavetzky et al., 2003; Ma et al., 2015; Moreno-Risueno et al., 2007; Wang et al., 2007). In the present study, we identified 32 *SiDof* genes in the sesame genome which were distributed across 12 distinct 'linkage groups' (LGs), including LG1, LG2, LG3, LG4, LG5, LG7, LG8, LG9, LG11, LG12, LG13 and LG15 (Table 1 and Fig. 9). Of which, LG8 comprises maximum number of *SiDof* genes such as *SiDof11*, *SiDof24*, *SiDof12*, *SiDof30*, *SiDof31* and *SiDof32*. The 32 members of *SiDof* genes identified in sesame are consistent with other plant species. The phylogenetic analysis has shown that 32 *SiDof* genes were categorized into six groups (Fig. 1) and showed a relationship with the *Dof* genes of *A. thaliana*, *G. max*, *Z. mays*, *S. lycopersicum*, *O. sativa*, and *M. truncatula* revealed their evolutionary conservation. Among these, group III exhibited the highest number of *SiDof* genes (seven) whereas group V contained the lowest number of *SiDof* genes (three). A similar result was reported in *Capsicum annuum* where 33 *Dof* genes were classified into seven groups (Wu et al., 2016). Further, 42 *M. truncatula* were categorized into four groups (Shu et al., 2015), 45 *Manihot esculenta* *Dof* genes were grouped into eight groups (Zou et al., 2019) and 117 *Brassica napus* *Dof* genes were classified into nine groups (Lohani et al., 2021).

The results of gene structure analysis showed that exon number varied from one to five whereas intron number varied from zero to four among 32 *SiDof* genes (Fig. 2). Among 32 *SiDof* genes, the majority of *SiDof* genes (71.88%) exhibited single exon and intronless.

Moreover, the maximum number of exons (five) was found in *SiDof11* (Fig. 2 and Table 1). It has been reported that 44.44% of *M. esculenta* (Zou et al., 2019), 75% of *Sorghum bicolor* (Kushwaha et al., 2011), 47.22% of *O. sativa* (Kushwaha et al., 2011) and 48.88% of *A. thaliana* (Kushwaha et al., 2011) *Dof* genes were without intron. The result shows that the diverse gene (exon-intron) structures of *SiDof* genes revealing the conservation of exon-intron structures. Moreover, the protein sequences of all 32 *Dof* genes contained a highly conserved N-terminal DNA-binding domain called ‘Dof’ domain. The result is also in agreement with the previously reported by (Kushwaha et al., 2011; Waschburger et al., 2024). The protein-protein network analysis revealed that there was a robust interaction between NAC071 and two *Dof* proteins namely DOF5.3 and DOFF5.6 (Fig. 8). As per findings, NAC071 (NAC domain-containing protein 71) is a transcription factor that is necessary for the division of pith cells during the reunion process and is involved in the tissue reunion of injured inflorescence stems (Asahina et al., 2011).

The results of synteny (Fig. 10) and phylogenetic analysis (Fig. 1) suggests that there were eleven pairs of *SiDof* genes that might be associated with gene duplication events. Among, those genes, *SiDof6*, *SiDof8*, *SiDof30*, and *SiDof31* were grouped into group I; *SiDof15*, *SiDof16*, *SiDof17*, and *SiDof22* were grouped into group II; *SiDof27* and *SiDof28* grouped into group IV. Moreover, *SiDof15* and *SiDof16* located on LG1; *SiDof7* and *SiDof22* located on LG3; *SiDof30* and *SiDof31* located on LG8; *SiDof27* and *SiDof28* located on LG9; *SiDof6* and *SiDof8* located on LG12 (Fig. 9). The results suggested that these *SiDof* genes might be resulting from duplication events. A similar duplication event was reported for *M. truncatula* *Dof* genes (Shu et al., 2015).

The promoter analysis of 32 *SiDof* genes showed that there were 733 were linked to the stress response (W BOX, WRE3, WUN-motif, ARE and box S), 562 to the light response (AE-box, Box-4, G-Box, and GA-motif), and 373 to the phytohormone response (ABRE, P-box, TATC-box, and TGACG-motif) (Table 5). Moreover, it was observed that *SiDof22* (SIN_1017363) had the greatest number of motifs (79) whereas *SiDof30* (SIN_1027039) had the least number of motifs. *SiDof31* (SIN_1027065) had the highest BOX-4 motifs (16), which is part of a conserved DNA module involved in light responsiveness. ABRE is an ABA-responsive element and *SiDof22* (SIN_1017363) had the highest of it (13) indicating that it may be a part of the ABA signaling pathway. Similarly, *SiDof13* (SIN_1008018) contained highest number of ARE motifs (6) which are essential for anaerobic induction.

Out of 32 *SiDof* genes, only six (*SiDof8*, *SiDof11*, *SiDof19*, *SiDof23*, *SiDof26*, and *SiDof31*) contained SSR motifs ranging from mono- to trinucleotides (Table 6). A similar SSR variants

for *Dof* genes were reported in *Cyperus esculentus* (Fu et al., 2024a) and *Camellia oleifera* (Fu et al., 2024b).

The transcriptome (RNA-seq) based expression profiles showed that the expression level of *SiDof1*, *SiDof11*, *SiDof12*, *SiDof13*, *SiDof18*, *SiDof22*, *SiDof23*, *SiDof27*, *SiDof28* and *SiDof32* genes was increased in drought stress condition compared to control while an opposite expression level was observed for *SiDof2*, *SiDof8*, *SiDof21* and *SiDof26* (Fig. 11b). The results indicated that these genes might be associated with drought stress. It has been reported that *CsDof10*, *CsDof13*, *CsDof14*, *CsDof15* and *CsDof16* were drought stress-responsive genes in *Camellia sinensis* (Yu et al., 2020). Gupta et al. also reported that the *Dof* gene was associated with drought stress in (Gupta et al., 2016). Likewise, 32 *SiDof* genes were analyzed with previously published sesame transcriptomic data (Zhang et al., 2020) under salinity stress and the results showed that *SiDof32* differs in expression levels between salinity sensitive and tolerant genotypes over time points as higher in tolerant than in sensitive genotypes suggesting this gene might be salinity responsive (Fig. 11c). Several previous research reports revealed that *Dof* genes were involved in salinity stress responses in various plant species like *S. bicolor* (Gupta et al., 2016), *C. annuum* (Wu et al., 2016), *GhDof1.7* in *Gossypium hirsutum* (Li et al., 2023) and *SiDof22* in *Solanum lycopersicum* (Cai et al., 2016).

Conclusion

The present study identified a total of 32 *SiDof* genes and characterized the identified genes. The available transcriptome data-based expression of *SiDof* genes showed differences in the expression levels in various abiotic stress conditions (such as drought and salinity) compared to controls in sesame. Therefore, the *SiDof* genes might have important role in drought and salinity.

CRedit authorship contribution statement

Conceptualization, M.A.R.; methodology, M.N.H., T.I.S., F.A.S., K.S.A. and M.A.R.; formal analysis, M.N.H., T.I.S., F.A.S., K.S.A.; investigation, M.N.H., T.I.S. and F.A.S.; resources, M.A.R.; data curation, M.N.H., T.I.S., F.A.S.; writing-original draft preparation, M.N.H., T.I.S., F.A.S., K.S.A. and M.A.R.; writing-review and editing, K.S.A and M.A.R.; supervision and project administration, M.A.R. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare they have no competing interests.

Acknowledgment

Not applicable

References

- Abdiani, N., Kolahi, M., Javaheriyani, M., Sabaeian, M., 2024. Effect of storage conditions on nutritional value, oil content, and oil composition of sesame seeds. *Journal of Agriculture and Food Research* 16, 101117. <https://doi.org/10.1016/j.jafr.2024.101117>
- Anilakumar, K.R., Pal, A., Khanum, F., Bawa, A.S., 2010. Nutritional, Medicinal and Industrial Uses of Sesame (*Sesamum indicum* L.) Seeds - An Overview. *Agriculturae Conspectus Scientificus* 75, 159–168.
- Asahina, M., Azuma, K., Pitaksaringkarn, W., Yamazaki, T., Mitsuda, N., Ohme-Takagi, M., Yamaguchi, S., Kamiya, Y., Okada, K., Nishimura, T., Koshihara, T., Yokota, T., Kamada, H., Satoh, S., 2011. Spatially selective hormonal control of RAP2.6L and ANAC071 transcription factors involved in tissue reunion in Arabidopsis. *Proceedings of the National Academy of Sciences* 108, 16128–16132. <https://doi.org/10.1073/pnas.1110443108>
- Baraki, F., Gebregergis, Z., Belay, Y., Berhe, M., Teame, G., Hassen, M., Gebremedhin, Z., Abadi, A., Negash, W., Atsbeha, A., Araya, G., 2020. Multivariate analysis for yield and yield-related traits of sesame (*Sesamum indicum* L.) genotypes. *Heliyon* 6, e05295. <https://doi.org/10.1016/j.heliyon.2020.e05295>
- Cai, X., Zhang, C., Shu, W., Ye, Z., Li, H., Zhang, Y., 2016. The transcription factor SlDof22 involved in ascorbate accumulation and salinity stress in tomato. *Biochemical and Biophysical Research Communications* 474, 736–741. <https://doi.org/10.1016/j.bbrc.2016.04.148>
- Cai, X., Zhang, Y., Zhang, C., Zhang, T., Hu, T., Ye, J., Zhang, J., Wang, T., Li, H., Ye, Z., 2013. Genome-wide analysis of plant-specific Dof transcription factor family in tomato. *J Integr Plant Biol* 55, 552–566. <https://doi.org/10.1111/jipb.12043>
- Chen, Y., Cao, J., 2015. Comparative Analysis of Dof Transcription Factor Family in Maize. *Plant Mol Biol Rep* 33, 1245–1258. <https://doi.org/10.1007/s11105-014-0835-9>
- Corrales, A.-R., Carrillo, L., Lasier, P., Nebauer, S.G., Dominguez-Figueroa, J., Renau-Morata, B., Pollmann, S., Granell, A., Molina, R.-V., Vicente-Carbajosa, J., Medina, J., 2017. Multifaceted role of cycling DOF factor 3 (CDF3) in the regulation of flowering time and abiotic stress responses in Arabidopsis. *Plant, Cell & Environment* 40, 748–764. <https://doi.org/10.1111/pce.12894>
- Corrales, A.-R., Nebauer, S.G., Carrillo, L., Fernández-Nohales, P., Marqués, J., Renau-Morata, B., Granell, A., Pollmann, S., Vicente-Carbajosa, J., Molina, R.-V., Medina, J., 2014. Characterization of tomato Cycling Dof Factors reveals conserved and new functions in the control of flowering time and abiotic stress responses. *Journal of Experimental Botany* 65, 995–1012. <https://doi.org/10.1093/jxb/ert451>
- Darzentas, N., 2010. Circoletto: visualizing sequence similarity with Circos. *Bioinformatics* 26, 2620–2621. <https://doi.org/10.1093/bioinformatics/btq484>
- Diaz, I., Vicente-Carbajosa, J., Abraham, Z., Martínez, M., Isabel-La Moneda, I., Carbonero, P., 2002. The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development. *Plant J* 29, 453–464. <https://doi.org/10.1046/j.0960-7412.2001.01230.x>
- Ding, L.-N., Li, M., Wang, W.-J., Cao, J., Wang, Z., Zhu, K.-M., Yang, Y.-H., Li, Y.-L., Tan, X.-L., 2019. Advances in plant GDSL lipases: from sequences to functional mechanisms. *Acta Physiologica Plant* 41, 151. <https://doi.org/10.1007/s11738-019-2944-4>

491 Dossa, K., Li, D., Zhou, R., Yu, J., Wang, L., Zhang, Y., You, J., Liu, A., Mmadi, M.A., Fonceka, D., Diouf,
492 D., Cissé, N., Wei, X., Zhang, X., 2019. The genetic basis of drought tolerance in the high oil
493 crop *Sesamum indicum*. *Plant Biotechnology Journal* 17, 1788–1803.
494 <https://doi.org/10.1111/pbi.13100>

495 Elleuch, M., Bedigian, D., Zitoun, A., 2011. Sesame (*Sesamum indicum* L.) Seeds in Food, Nutrition,
496 and Health, in: *Nuts and Seeds in Health and Disease Prevention*. Elsevier, pp. 1029–1036.
497 <https://doi.org/10.1016/B978-0-12-375688-6.10122-7>

498 Fasuan, T.O., Gbadamosi, S.O., Omobuwajo, T.O., 2018. Characterization of protein isolate from
499 *Sesamum indicum* seed: In vitro protein digestibility, amino acid profile, and some functional
500 properties. *Food Science & Nutrition* 6, 1715–1723. <https://doi.org/10.1002/fsn3.743>

501 Fu, C., Liao, Z., Jiang, N., Yang, Y., 2024a. Genome-wide identification and molecular evolution of Dof
502 transcription factors in *Cyperus esculentus*. *BMC Genomics* 25, 667.
503 <https://doi.org/10.1186/s12864-024-10565-y>

504 Fu, C., Xiao, Y., Jiang, N., Yang, Y., 2024b. Genome-wide identification and molecular evolution of Dof
505 gene family in *Camellia oleifera*. *BMC Genomics* 25, 702. [https://doi.org/10.1186/s12864-](https://doi.org/10.1186/s12864-024-10622-6)
506 [024-10622-6](https://doi.org/10.1186/s12864-024-10622-6)

507 Guo, Y., Qiu, L.-J., 2013. Genome-Wide Analysis of the Dof Transcription Factor Gene Family Reveals
508 Soybean-Specific Duplicable and Functional Characteristics. *PLOS ONE* 8, e76809.
509 <https://doi.org/10.1371/journal.pone.0076809>

510 Gupta, S., Arya, G.C., Malviya, N., Bisht, N.C., Yadav, D., 2016. Molecular cloning and expression
511 profiling of multiple Dof genes of *Sorghum bicolor* (L) Moench. *Mol Biol Rep* 43, 767–774.
512 <https://doi.org/10.1007/s11033-016-4019-6>

513 Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., Gao, G., 2015. GSDS 2.0: an upgraded gene feature
514 visualization server. *Bioinformatics* 31, 1296–1297.
515 <https://doi.org/10.1093/bioinformatics/btu817>

516 Iwamoto, M., Higo, K., Takano, M., 2009. Circadian clock- and phytochrome-regulated Dof-like gene,
517 *Rdd1*, is associated with grain size in rice. *Plant, Cell & Environment* 32, 592–603.
518 <https://doi.org/10.1111/j.1365-3040.2009.01954.x>

519 Kadkhodaie, A., Razmjoo, J., Zahedi, M., Pessarakli, M., 2014. Oil Content and Composition of
520 Sesame (*Sesamum indicum* L.) Genotypes as Affected by Irrigation Regimes. *Journal of the*
521 *American Oil Chemists' Society* 91, 1737–1744. <https://doi.org/10.1007/s11746-014-2524-0>

522 Kim, H.-S., Kim, S.J., Abbasi, N., Bressan, R.A., Yun, D.-J., Yoo, S.-D., Kwon, S.-Y., Choi, S.-B., 2010. The
523 DOF transcription factor Dof5.1 influences leaf axial patterning by promoting *Revoluta*
524 transcription in *Arabidopsis*. *Plant J* 64, 524–535. [https://doi.org/10.1111/j.1365-](https://doi.org/10.1111/j.1365-313X.2010.04346.x)
525 [313X.2010.04346.x](https://doi.org/10.1111/j.1365-313X.2010.04346.x)

526 Kumar, R., Taware, R., Gaur, V.S., Guru, S.K., Kumar, A., 2009. Influence of nitrogen on the
527 expression of TaDof1 transcription factor in wheat and its relationship with photo synthetic
528 and ammonium assimilating efficiency. *Mol Biol Rep* 36, 2209–2220.
529 <https://doi.org/10.1007/s11033-008-9436-8>

530 Kushwaha, H., Gupta, S., Singh, V.K., Rastogi, S., Yadav, D., 2011. Genome wide identification of Dof
531 transcription factor gene family in sorghum and its comparative phylogenetic analysis with
532 rice and *Arabidopsis*. *Mol Biol Rep* 38, 5037–5053. [https://doi.org/10.1007/s11033-010-](https://doi.org/10.1007/s11033-010-0650-9)
533 [0650-9](https://doi.org/10.1007/s11033-010-0650-9)

534 Li, Y., Tian, M., Feng, Z., Zhang, J., Lu, J., Fu, X., Ma, L., Wei, H., Wang, H., 2023. GhDof1.7, a Dof
535 Transcription Factor, Plays Positive Regulatory Role under Salinity Stress in Upland Cotton.
536 *Plants* 12, 3740. <https://doi.org/10.3390/plants12213740>

537 Lijavetzky, D., Carbonero, P., Vicente-Carbajosa, J., 2003. Genome-wide comparative phylogenetic
538 analysis of the rice and *Arabidopsis* Dof gene families. *BMC Evolutionary Biology* 3, 17.
539 <https://doi.org/10.1186/1471-2148-3-17>

- Liu, J., Meng, Q., Xiang, H., Shi, F., Ma, L., Li, Y., Liu, C., Liu, Y., Su, B., 2021. Genome-wide analysis of Dof transcription factors and their response to cold stress in rice (*Oryza sativa* L.). *BMC Genomics* 22, 800. <https://doi.org/10.1186/s12864-021-08104-0>
- Lohani, N., Babaei, S., Singh, M.B., Bhalla, P.L., 2021. Genome-Wide In Silico Identification and Comparative Analysis of Dof Gene Family in *Brassica napus*. *Plants* 10, 709. <https://doi.org/10.3390/plants10040709>
- Ma, J., Li, M.-Y., Wang, F., Tang, J., Xiong, A.-S., 2015. Genome-wide analysis of Dof family transcription factors and their responses to abiotic stresses in Chinese cabbage. *BMC Genomics* 16, 33. <https://doi.org/10.1186/s12864-015-1242-9>
- Moreno-Risueno, M.A., Martínez, M., Vicente-Carbajosa, J., Carbonero, P., 2007. The family of DOF transcription factors: from green unicellular algae to vascular plants. *Mol Genet Genomics* 277, 379–390. <https://doi.org/10.1007/s00438-006-0186-9>
- Nan, H., Ludlow, R.A., Lu, M., An, H., 2021. Genome-Wide Analysis of Dof Genes and Their Response to Abiotic Stress in Rose (*Rosa chinensis*). *Front. Genet.* 12. <https://doi.org/10.3389/fgene.2021.538733>
- Nekrutenko, A., Makova, K.D., Li, W.-H., 2002. The KA/KS Ratio Test for Assessing the Protein-Coding Potential of Genomic Regions: An Empirical and Simulation Study. *Genome Res* 12, 198–202. <https://doi.org/10.1101/gr.200901>
- Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., Finn, R.D., 2018. HMMER web server: 2018 update. *Nucleic Acids Research* 46, W200–W204. <https://doi.org/10.1093/nar/gky448>
- Protein BIG GRAIN 1-like A - *Arabidopsis thaliana* (Mouse-ear cress) | UniProtKB | UniProt [WWW Document], n.d. URL <https://www.uniprot.org/uniprotkb/Q9LVK2/entry> (accessed 3.14.24).
- Shen, G., Sun, W., Chen, Z., Shi, L., Hong, J., Shi, J., 2022. Plant GDSL Esterases/Lipases: Evolutionary, Physiological and Molecular Functions in Plant Development. *Plants* 11, 468. <https://doi.org/10.3390/plants11040468>
- Shu, Y.J., Song, L.L., Zhang, J., Liu, Y., Guo, C.H., 2015. Genome-wide identification and characterization of the Dof gene family in *Medicago truncatula*. *Genet Mol Res* 14, 10645–10657. <https://doi.org/10.4238/2015.September.9.5>
- Shuichi Yanagisawa, 2004. Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants. *Plant Cell Physiol* 45, 386–391. <https://doi.org/10.1093/pcp/pch055>
- Skirycz, A., Radziejewski, A., Busch, W., Hannah, M.A., Czeszejko, J., Kwaśniewski, M., Zanon, M.-I., Lohmann, J.U., De Veylder, L., Witt, I., Mueller-Roeber, B., 2008. The DOF transcription factor OBP1 is involved in cell cycle regulation in *Arabidopsis thaliana*. *Plant J* 56, 779–792. <https://doi.org/10.1111/j.1365-313X.2008.03641.x>
- Song, X., Yang, Q., Bai, Y., Gong, K., Wu, T., Yu, T., Pei, Q., Duan, W., Huang, Z., Wang, Z., Liu, Z., Kang, X., Zhao, W., Ma, X., 2021. Comprehensive analysis of SSRs and database construction using all complete gene-coding sequences in major horticultural and representative plants. *Hortic Res* 8, 1–17. <https://doi.org/10.1038/s41438-021-00562-7>
- Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* 38, 3022–3027. <https://doi.org/10.1093/molbev/msab120>
- Tanaka, M., Takahata, Y., Nakayama, H., Nakatani, M., Tahara, M., 2009. Altered carbohydrate metabolism in the storage roots of sweetpotato plants overexpressing the SRF1 gene, which encodes a Dof zinc finger transcription factor. *Planta* 230, 737–746. <https://doi.org/10.1007/s00425-009-0979-2>
- Umemura, Y., Ishiduka, T., Yamamoto, R., Esaka, M., 2004. 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* 37, 741–749. <https://doi.org/10.1111/j.1365-313x.2003.01997.x>
- Voorrips, R.E., 2002. MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *J Hered* 93, 77–78. <https://doi.org/10.1093/jhered/93.1.77>

- Wang, H., Zhao, S., Gao, Y., Yang, J., 2017. Characterization of Dof Transcription Factors and Their Responses to Osmotic Stress in Poplar (*Populus trichocarpa*). *PLoS One* 12, e0170210. <https://doi.org/10.1371/journal.pone.0170210>
- Wang, H.-W., Zhang, B., Hao, Y.-J., Huang, J., Tian, A.-G., Liao, Y., Zhang, J.-S., Chen, S.-Y., 2007. The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic *Arabidopsis* plants. *The Plant Journal* 52, 716–729. <https://doi.org/10.1111/j.1365-3113X.2007.03268.x>
- Waschburger, E.L., Filgueiras, J.P.C., Turchetto-Zolet, A.C., 2024. DOF gene family expansion and diversification. *Genet. Mol. Biol.* 46, e20230109. <https://doi.org/10.1590/1678-4685-GMB-2023-0109>
- Wei, P., Zhao, F., Wang, Z., Wang, Q., Chai, X., Hou, G., Meng, Q., 2022. Sesame (*Sesamum indicum* L.): A Comprehensive Review of Nutritional Value, Phytochemical Composition, Health Benefits, Development of Food, and Industrial Applications. *Nutrients* 14, 4079. <https://doi.org/10.3390/nu14194079>
- Wu, Z., Cheng, J., Cui, J., Xu, X., Liang, G., Luo, X., Chen, X., Tang, X., Hu, K., Qin, C., 2016. Genome-Wide Identification and Expression Profile of Dof Transcription Factor Gene Family in Pepper (*Capsicum annuum* L.). *Front. Plant Sci.* 7. <https://doi.org/10.3389/fpls.2016.00574>
- Yanagisawa, S., 2002. The Dof family of plant transcription factors. *Trends Plant Sci* 7, 555–560. [https://doi.org/10.1016/s1360-1385\(02\)02362-2](https://doi.org/10.1016/s1360-1385(02)02362-2)
- Yanagisawa, S., 1995. A novel DNA-binding domain that may form a single zinc finger motif. *Nucleic Acids Research* 23, 3403–3410. <https://doi.org/10.1093/nar/23.17.3403>
- Yanagisawa, S., Sheen, J., 1998. Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. *Plant Cell* 10, 75–89.
- Yang, J., An, D., Zhang, P., 2011. Expression Profiling of Cassava Storage Roots Reveals an Active Process of Glycolysis/Gluconeogenesis. *Journal of Integrative Plant Biology* 53, 193–211. <https://doi.org/10.1111/j.1744-7909.2010.01018.x>
- Yang, Q., Chen, Q., Zhu, Y., Li, T., 2017. Identification of MdDof genes in apple and analysis of their response to biotic or abiotic stress. *Functional Plant Biol.* 45, 528–541. <https://doi.org/10.1071/FP17288>
- Yang, Z., Nielsen, R., 2000. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Molecular Biology and Evolution* 17, 32–43. <https://doi.org/10.1093/oxfordjournals.molbev.a026236>
- Yu, Q., Li, C., Zhang, J., Tian, Y., Wang, H., Zhang, Y., Zhang, Z., Xiang, Q., Han, X., Zhang, L., 2020. Genome-wide identification and expression analysis of the Dof gene family under drought stress in tea (*Camellia sinensis*). *PeerJ* 8, e9269. <https://doi.org/10.7717/peerj.9269>
- Zhang, Yajuan, Li, D., Zhou, R., Liu, A., Wang, L., Zhang, Yanxin, Gong, H., Zhang, X., You, J., 2020. A collection of transcriptomic and proteomic datasets from sesame in response to salt stress. *Data in Brief* 32, 106096. <https://doi.org/10.1016/j.dib.2020.106096>
- Zhang, Z., Yuan, L., Liu, X., Chen, X., Wang, X., 2018. Evolution analysis of Dof transcription factor family and their expression in response to multiple abiotic stresses in *Malus domestica*. *Gene* 639, 137–148. <https://doi.org/10.1016/j.gene.2017.09.039>
- Zhou, W., Sheng, C., Koffi Dossou, S.S., Wang, Z., Song, S., You, J., Wang, L., 2023. Genome-wide identification of TPS genes in sesame and analysis of their expression in response to abiotic stresses. *Oil Crop Science* 8, 81–88. <https://doi.org/10.1016/j.ocsci.2023.03.004>
- Zhou, Y., Cheng, Y., Wan, C., Li, J., Yang, Y., Chen, J., 2020. Genome-wide characterization and expression analysis of the Dof gene family related to abiotic stress in watermelon. *PeerJ* 8, e8358. <https://doi.org/10.7717/peerj.8358>
- Zou, Z., Zhu, J., Zhang, X., 2019. Genome-wide identification and characterization of the Dof gene family in cassava (*Manihot esculenta*). *Gene* 687, 298–307. <https://doi.org/10.1016/j.gene.2018.11.053>

642
643

Table 1 Physicochemical characteristics of members of the *Dof* gene family in *S. indicum*

Gene name	Gene ID	Locus ID	Location	Gene length (bp)	Protein length (aa)	Sub-cellular localization	Molecular weight (kDa)	Isoelectric point (pI)	Gravy
<i>SiDof1</i>	SIN_1001405	LOC105155434	LG3: 2,688,274-2,691,024	2751	496	53022.55	Nucleus	5.55	-0.784
<i>SiDof2</i>	SIN_1003692	LOC105156034	LG15: 3,268,757-3,269,854	1098	333	34907.89	Nucleus	8.61	-0.312
<i>SiDof3</i>	SIN_1003703	LOC105157099	LG15: 3,345,250-3,346,231	702	233	24408.25	Nucleus	8.51	-0.285
<i>SiDof4</i>	SIN_1004081	LOC105157219	LG13: 3,816,279-3,817,304	1026	341	37095.26	Nucleus	8.96	-0.643
<i>SiDof5</i>	SIN_1004940	LOC105157727	LG15: 2,211,229-2,212,412	846	281	31711.16	Nucleus	5.97	-0.815
<i>SiDof6</i>	SIN_1005243	LOC105158041	LG12: 4,029,144-4,030,025	882	293	32239.93	Nucleus	7.61	-0.622
<i>SiDof7</i>	SIN_1005291	LOC105159975	LG12: 4,337,875-4,338,729	855	284	30783.5	Nucleus	9.81	-0.694
<i>SiDof8</i>	SIN_1005292	LOC105160319	LG12: 4,345,975-4,347,097	1123	315	34618.53	Nucleus	8.193	-0.748
<i>SiDof9</i>	SIN_1005551	LOC105161465	LG5: 793,746-794,336	591	196	20629.21	Nucleus	9.95	-0.389
<i>SiDof10</i>	SIN_1005860	LOC105162358	LG4: 8,931,010-8,931,363	354	117	12583.67	Nucleus	4.49	-0.475
<i>SiDof11</i>	SIN_1006007	LOC105163550	LG8: 14,949,214-14,961,955	12742	719	78342.23	Nucleus	5.31	-0.635
<i>SiDof12</i>	SIN_1006579	LOC105166214	LG8: 3,494,213-3,496,822	2610	497	53362.02	Nucleus	5.39	-0.679
<i>SiDof13</i>	SIN_1008018	LOC105166761	LG15: 1,006,747-1,009,488	2742	552	59895.47	Nucleus	5.72	-0.682
<i>SiDof14</i>	SIN_1009212	LOC105167467	LG7: 9,184,600-9,185,770	1171	252	28381.57	Nucleus	9.28	-0.894
<i>SiDof15</i>	SIN_1009420	LOC105167652	LG1: 10,741,936-10,742,844	909	302	31986.02	Nucleus	6.17	-0.678
<i>SiDof16</i>	SIN_1009530	LOC105167857	LG1: 11,671,599-11,672,309	711	236	24637.58	Nucleus	5.78	-0.701
<i>SiDof17</i>	SIN_1010375	LOC105167892	LG3: 8,646,525-8,647,241	717	238	24921.17	Nucleus	8.2	-0.728
<i>SiDof18</i>	SIN_1011214	LOC105167893	LG11: 13,978,880-13,979,722	843	280	29980.41	Nucleus	9.57	-0.536
<i>SiDof19</i>	SIN_1011668	LOC105168584	LG7: 4,440,768-4,441,825	756	251	26842.1	Nucleus	8.63	-0.518
<i>SiDof20</i>	SIN_1013094	LOC105168787	LG11: 11,136,758-11,137,603	846	281	23450.09	Nucleus	4.89	-0.28
<i>SiDof21</i>	SIN_1013324	LOC105170336	LG2: 4,460,598-4,461,491	894	297	32901.17	Nucleus	4.47	-0.589
<i>SiDof22</i>	SIN_1017363	LOC105170685	LG3: 11,375,833-11,376,732	900	299	32125.22	Nucleus	5.98	-0.697
<i>SiDof23</i>	SIN_1018780	LOC105174207	LG4: 3,971,861-3,974,149	2289	430	47595.85	Nucleus	6.06	-0.779
<i>SiDof24</i>	SIN_1019847	LOC105174708	LG8: 2,620,144-2,620,506	363	120	13173.59	Nucleus	10.4	-1.03
<i>SiDof25</i>	SIN_1020428	LOC105175506	LG3: 4,087,700-4,088,614	915	304	33502.16	Nucleus	7.04	-0.692
<i>SiDof26</i>	SIN_1021128	LOC105175553	LG2: 13,201,838-13,202,701	864	287	31376.75	Nucleus	8.48	-0.638
<i>SiDof27</i>	SIN_1024161	LOC105175554	LG9: 5,205,549-5,206,379	831	276	30097.56	Nucleus	8.6	-0.651
<i>SiDof28</i>	SIN_1024519	LOC105176376	LG9: 2,530,186-2,531,085	900	299	32104.59	Nucleus	8.39	-0.659
<i>SiDof29</i>	SIN_1025235	LOC105177260	LG15: 8,227,690-8,228,706	1017	338	36430.69	Nucleus	9	-0.496
<i>SiDof30</i>	SIN_1027039	LOC105177686	LG8: 8,191,599-8,192,516	918	305	33262.06	Nucleus	6.49	-0.557
<i>SiDof31</i>	SIN_1027065	LOC105177758	LG8: 8,013,200-8,014,375	1,176	327	35413.22	Nucleus	6.98	-0.553
<i>SiDof32</i>	SIN_1027102	LOC105178230	LG8: 7,821,914-7,822,393	480	159	17966.22	Nucleus	9.12	-0.914

644
645
646

647

Table 2 Key domains found in the 32 *Dof* proteins of *S. indicum*

Sl. No.	Domain Name	Description	Functions	Reference
1	'zf-Dof'	'DNA-binding with one finger' (zf-Dof) domain proteins (52 aa residues) are TFs with a highly conserved 'DNA-binding domain' with a single Cys2/Cys2 zinc finger (pfam02701)	Act as a transcriptional activator or repressor in various biological processes in plants	(Shuichi Yanagisawa, 2004; Umemura et al., 2004)
2	'Lipase_GDSL'	GDSL-like Lipase/Acylhydrolase (pfam00657)	Play a vital role in growth and development, and stress tolerance in plants	(Ding et al., 2019; Shen et al., 2022)

Table 3 The conserved motifs present in 32 *Dof* proteins of *S. indicum*

Motif name	E-value	Sites	Width	Motif sequence
Motif 1	3.4e-1402	31	50	CPRCDSTNTKFCYYNNYNLSQPRHFCKSCRRYWTKGGLTRNVPVGGGCRK
Motif 2	7.50E-71	5	41	EEHREEKNSEKSVWVPKTLRIDDPGEAAKSSIWATLGIKND
Motif 3	2.50E-56	4	50	QFRHJTIPEALQAARIDLPGNFHHPALKPNGTILSFGPDSPLCESMASVL
Motif 4	3.10E-38	5	30	KGDEKKQAVNTSTVLQANPAALSRJSFHE
Motif 5	2.10E-32	5	28	YAAPYWGCAVPGAWNVPWVSPPSPTVNQ
Motif 6	1.30E-31	4	40	PPLMPNINPFPPVPCFPGVPWPYPWNSVQPSPAVSPPAY
Motif 7	2.60E-29	5	25	KTEEDQSETTNSQEKTLLKPKDKILP
Motif 8	1.10E-20	5	18	EVKDPEIKLFGKTIELPE
Motif 9	2.90E-08	5	21	IJVCSPKQEQKKPRPAKEQAL
Motif 10	5.90E-08	2	50	DNYGGNJDLSLEESIITLTDIPSTASVPWPGPDSMSPMDLPNCWNWNEIDA

Table 4 The list of 11 pair duplication events in *Dof* genes of *S. indicum* and their Ka/Ks ratio

Sl. No.	Gene 1	Gene 2	Non-synonymous substitution (Ka)	Synonymous substitution (Ks)	Ka/Ks
1	SIN_1001405	SIN_1006007	0.1564	1.0201	0.1533
2	SIN_1003692	SIN_1008018	0.7178	NaN	NaN
3	SIN_1003703	SIN_1019847	0.2203	0.6696	0.3290
4	SIN_1004081	SIN_1025235	0.2181	0.9622	0.2267
5	SIN_1005243	SIN_1027039	0.1984	0.6368	0.3115
6	SIN_1005292	SIN_1027065	0.2541	0.7646	0.3323
7	SIN_1005551	SIN_1011214	0.1276	0.7681	0.1661
8	SIN_1005860	SIN_1021128	0.3836	1.9718	0.1946
9	SIN_1009420	SIN_1017363	0.1388	0.9297	0.1492
10	SIN_1009530	SIN_1010375	0.1196	0.7141	0.1675
11	SIN_1024161	SIN_1024519	0.1925	0.9242	0.2083

655

656 **Table 5** *Cis*-regulatory elements in the promoter regions of *SiDof* genes

Types	Gene name	<i>SiDof1</i>	<i>SiDof12</i>	<i>SiDof3</i>	<i>SiDof4</i>	<i>SiDof5</i>	<i>SiDof6</i>	<i>SiDof7</i>	<i>SiDof8</i>	<i>SiDof9</i>	<i>SiDof10</i>	<i>SiDof11</i>	<i>SiDof12</i>	<i>SiDof13</i>	<i>SiDof14</i>	<i>SiDof15</i>	<i>SiDof16</i>	<i>SiDof17</i>	<i>SiDof18</i>	<i>SiDof19</i>	<i>SiDof20</i>	<i>SiDof21</i>	<i>SiDof22</i>	<i>SiDof23</i>	<i>SiDof24</i>	<i>SiDof25</i>	<i>SiDof26</i>	<i>SiDof27</i>	<i>SiDof28</i>	<i>SiDof29</i>	<i>SiDof30</i>	<i>SiDof31</i>	<i>SiDof32</i>	Total
Light response	AE-BOX	0	1	1	0	1	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	1	562
	BOX-4	9	15	3	3	3	6	8	6	8	5	4	3	5	8	4	6	7	5	6	6	12	9	10	4	11	10	9	5	1	7	16	8	
	G-box	3	7	1	2	0	5	3	5	1	1	4	8	5	4	10	3	0	2	3	6	6	13	7	0	9	4	4	5	4	1	6	4	
	GA-motif	0	1	0	0	0	1	0	1	0	2	1	0	0	2	1	0	1	0	0	1	0	0	1	1	0	0	0	0	2	0	0	0	
	GATA-motif	0	2	1	0	1	3	1	1	2	2	1	1	2	2	2	0	1	1	0	1	2	3	2	2	0	1	0	0	1	3	0	2	
	I-box	2	2	0	0	3	1	0	1	0	1	2	0	0	1	0	0	0	0	1	0	0	2	1	0	0	1	0	0	0	0	0	3	
	MRE	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	3	1	0	1	0	1	2	1	1	0	0	2	0	0	0	0	0	
	GT1-motif	1	1	0	2	0	1	5	0	3	0	4	1	2	0	3	3	1	1	1	3	7	3	2	6	0	0	1	4	1	1	2	0	
	Sp1	1	0	4	0	1	0	0	0	2	0	0	2	0	0	0	0	0	0	2	0	0	1	1	1	0	1	0	0	0	0	0	0	
ATCT-motif	0	0	0	1	0	1	1	1	0	1	0	0	0	0	1	3	0	0	0	0	1	1	0	0	0	2	1	0	1	1	1	1	0	
ACE	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	1	0	2	0	
Phytohormone	ABRE	2	6	1	2	0	4	3	4	0	1	3	6	4	3	9	3	0	3	5	6	7	13	6	0	8	3	3	4	5	2	7	3	373
	GARE-motif	1	0	0	0	0	1	3	0	0	1	1	0	2	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	1	0	
	ERE	4	2	1	7	0	0	0	2	2	5	3	5	3	4	7	0	3	3	5	3	2	5	3	2	5	2	4	0	5	1	1	4	
	TATC-box	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	1	3	0	0	0	0	0	1	0	0	1	
	P box	0	0	0	1	0	0	1	0	0	0	1	0	2	0	2	1	3	1	0	1	0	1	0	0	0	1	1	0	0	0	0	0	
	TCA-element	1	1	2	2	0	0	0	1	0	0	0	0	0	0	0	1	4	1	0	0	2	0	1	1	0	0	1	1	4	2	0	0	
	TGA-element	2	1	1	0	0	0	0	0	1	2	2	0	1	0	0	0	0	0	1	2	0	2	1	0	0	1	0	1	0	0	1	1	
Stress response	TGACG-motif	1	3	0	3	1	3	2	5	1	0	1	1	2	2	2	2	6	0	1	1	1	2	4	1	3	3	5	0	4	2	3	2	733
	ARE	1	3	2	1	2	0	3	2	2	3	3	2	6	1	2	2	5	6	2	4	2	0	2	2	0	5	2	0	2	0	3	1	
	Box S	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	
	DRE CORE	0	0	2	0	1	2	0	1	0	0	0	0	0	0	2	0	0	0	1	0	0	1	0	2	0	0	0	0	0	0	0	0	
	MBS	1	4	0	1	2	2	1	2	1	2	3	0	1	1	3	2	0	2	0	0	1	1	1	0	0	1	1	0	1	0	3	1	
	STRE	5	2	4	4	7	3	4	2	1	3	3	4	4	4	1	4	4	3	2	4	6	8	9	1	5	0	3	1	8	3	5	6	
	TC-rich repeats	1	1	3	0	1	0	0	1	0	0	0	1	0	1	0	0	0	2	2	1	0	0	0	1	1	0	0	0	1	0	0	0	
	W-box	1	0	1	0	1	0	0	0	1	0	0	2	0	1	0	0	0	1	0	2	1	0	4	1	0	1	0	0	1	0	1	0	
	WRE3	3	2	1	1	0	2	2	0	1	1	0	2	2	1	1	1	1	1	0	1	1	0	0	0	1	1	5	3	2	1	3	1	
	WUN-motif	2	2	1	1	0	1	2	0	1	4	0	0	2	3	4	1	3	0	2	1	0	1	0	2	1	4	1	1	0	0	1	1	
	CCAAT-box	0	0	0	0	2	1	1	0	0	0	1	1	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	1	1	0	
	MYB	5	2	1	2	5	5	11	1	3	3	5	3	6	0	3	4	4	4	2	2	7	3	2	2	8	3	2	5	2	5	7	5	
	MYB recognition site	0	0	0	0	2	1	1	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	1	1	0	
	Myb-binding site	1	1	0	0	0	3	4	0	2	1	0	0	2	0	1	0	2	2	0	0	1	1	0	0	4	1	0	1	0	1	4	0	
	MYC	2	1	6	6	10	7	11	2	7	7	3	6	9	8	4	7	5	4	7	4	8	5	2	4	2	2	3	8	3	2	4	5	
LTR	0	2	1	0	0	0	0	1	0	0	3	0	1	0	2	1	0	0	1	0	1	1	1	1	1	0	1	0	1	0	2	0		
Total no. of motifs		49	62	38	39	44	55	70	39	43	46	51	49	61	48	64	47	53	48	44	51	74	79	67	34	65	46	51	41	50	36	76	48	1668

657

658 **Table 6** The SSR found in the *SiDof* genes

Gene name	Gene ID	SSR nr.	SSR type	SSR	Size	Start	End
<i>SiDof8</i>	SIN_1005292	1	p2	(TC)13	26	89	114
<i>SiDof11</i>	SIN_1006007	1	p3	(TTA)7	21	1278	1298
<i>SiDof11</i>	SIN_1006007	2	p2	(TG)6	12	3906	3917
<i>SiDof11</i>	SIN_1006007	3	p1	(A)13	13	4422	4434
<i>SiDof11</i>	SIN_1006007	4	p1	(G)13	13	4645	4657
<i>SiDof11</i>	SIN_1006007	5	p2	(CT)11	22	5210	5231
<i>SiDof11</i>	SIN_1006007	6	p2	(TA)13	26	5313	5338
<i>SiDof11</i>	SIN_1006007	7	p2	(TA)6	12	8240	8251
<i>SiDof11</i>	SIN_1006007	8	p2	(GA)7	14	10042	10055
<i>SiDof11</i>	SIN_1006007	9	p1	(A)10	10	10820	10829
<i>SiDof11</i>	SIN_1006007	10	c	(AT)6(A)12	24	11225	11248
<i>SiDof19</i>	SIN_1011668	1	p3	(CCA)5	15	365	379
<i>SiDof23</i>	SIN_1018780	1	p2	(GT)12	24	175	198
<i>SiDof23</i>	SIN_1018780	2	p2	(GT)6	12	983	994
<i>SiDof26</i>	SIN_1021128	1	p3	(ATC)5	15	276	290
<i>SiDof31</i>	SIN_1027065	1	p2	(TC)9	18	75	92
<i>SiDof31</i>	SIN_1027065	2	p3	(TCC)6	18	574	591

659

660

Figures

Fig. 1 The phylogenetic tree showing evolutionary relationship of sesame *Dof* genes with other species. The circular phylogenetic tree was built using the *Dof* protein sequences of *S. indicum*, *A. thaliana*, *G. max*, *Z. mays*, *S. lycopersicum*, *O. sativa* and *M. truncatula*. The tree was built using MEGA11 considering ‘Maximum Likelihood’ method

Fig. 2 The gene structure (exon-intron) of member of the *Dof* gene family in sesame

Fig. 3 The conserved domains found in SiDof proteins. The domains were anticipated by CD-search of NCBI (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>)

Fig. 4 The conserved motifs present in 32 SiDof proteins. Motifs are labelled with different colored rectangles from 1 to 10

Fig. 5 Logos of different motifs (1-10) found in the ‘zf-Dof’ and ‘Lipase-GDSL’ domains of SiDof proteins

Fig. 6 Multiple protein alignment of *Dof* proteins in sesame (a). The ‘zf-Dof’ domain is labeled below the alignment (b). The alignment was made with web tool (<http://www.phylogeny.fr/>)

Fig. 7 The ‘three-dimensional’ (3D) structures of SiDof proteins. The secondary structure elements α -helices, β -sheets, and coils (disordered structures) are indicated for the predicted 3D structures of SiDof proteins. Spiral parts indicate α -helix and Ribbon like structures indicate β -sheet

Fig. 8 ‘Protein-protein interaction’ analysis of SiDof proteins. The red points in the figure represent the SiDof proteins, and the blue points illustrate the proteins that interacted with the SiDof proteins. The green lines indicate the interactions between the SiDof proteins.

Fig. 9 Genomic location of 32 *Dof* genes in sesame genome

Fig. 10 Syntenic association of *Dof* genes among *S. indicum*, *A. thaliana*, *G. max* and *O. sativa* representing the level of conservation. The coloured ribbons indicate the level and intensity of evolutionary conservation among different *SiDof* genes. The maximum strength is from red to blue

Fig. 11 Heatmap representation of the FMKM/RPKM expression pattern of *SiDof* genes in *S. indicum*. The FMKM value was retrieved from GEO dataset of the NCBI database. a) Expression levels in young leaf (GSE164511), b) Drought (PEG-treated root) (GSE148340) and c) RPKM expression in salt-tolerant and salt-sensitive shoots under salt-stress condition (Zhang et al., 2020). The maximum strength is from red to green

Figures

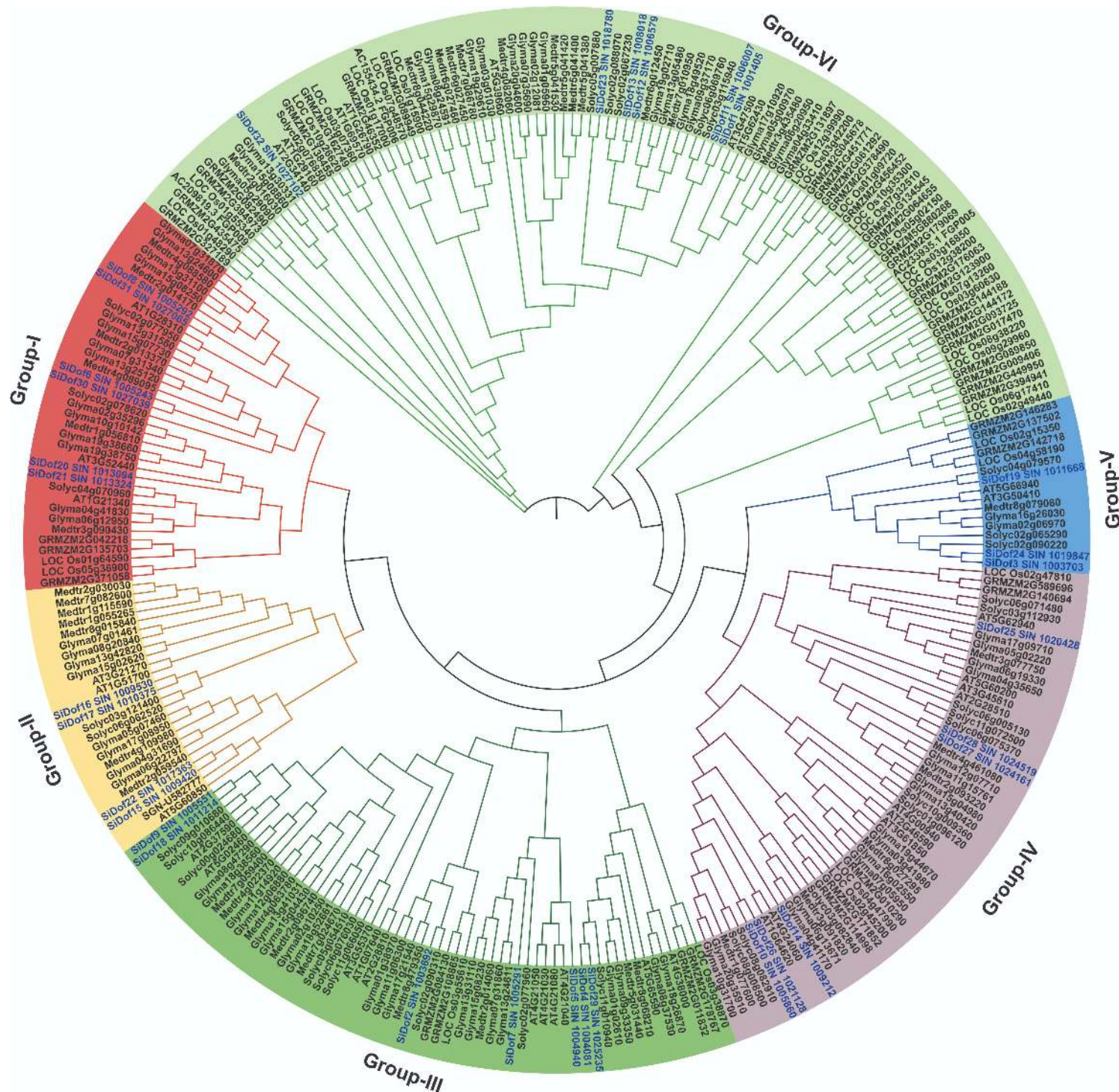


Figure 1

The phylogenetic tree showing evolutionary relationship of sesame *Dof* genes with other species. The circular phylogenetic tree was built using the *Dof* protein sequences of *S. indicum*, *A. thaliana*, *G. max*, *Z. mays*, *S. lycopersicum*, *O. sativa* and *M. truncatula*. The tree was built using MEGA11 considering 'Maximum Likelihood' method

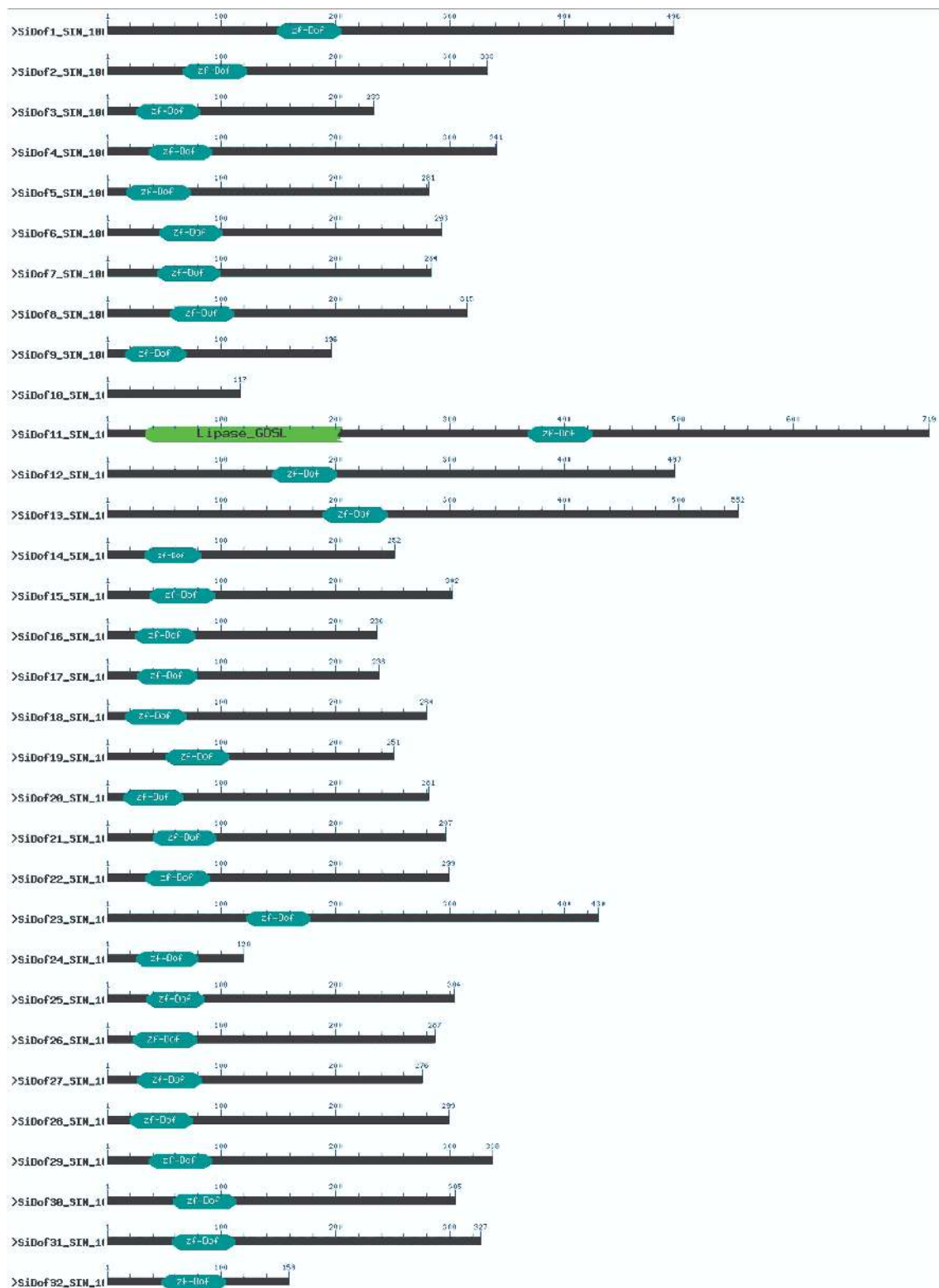


Figure 3

The conserved domains found in SiDof proteins. The domains were anticipated by CD-search of NCBI (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>)

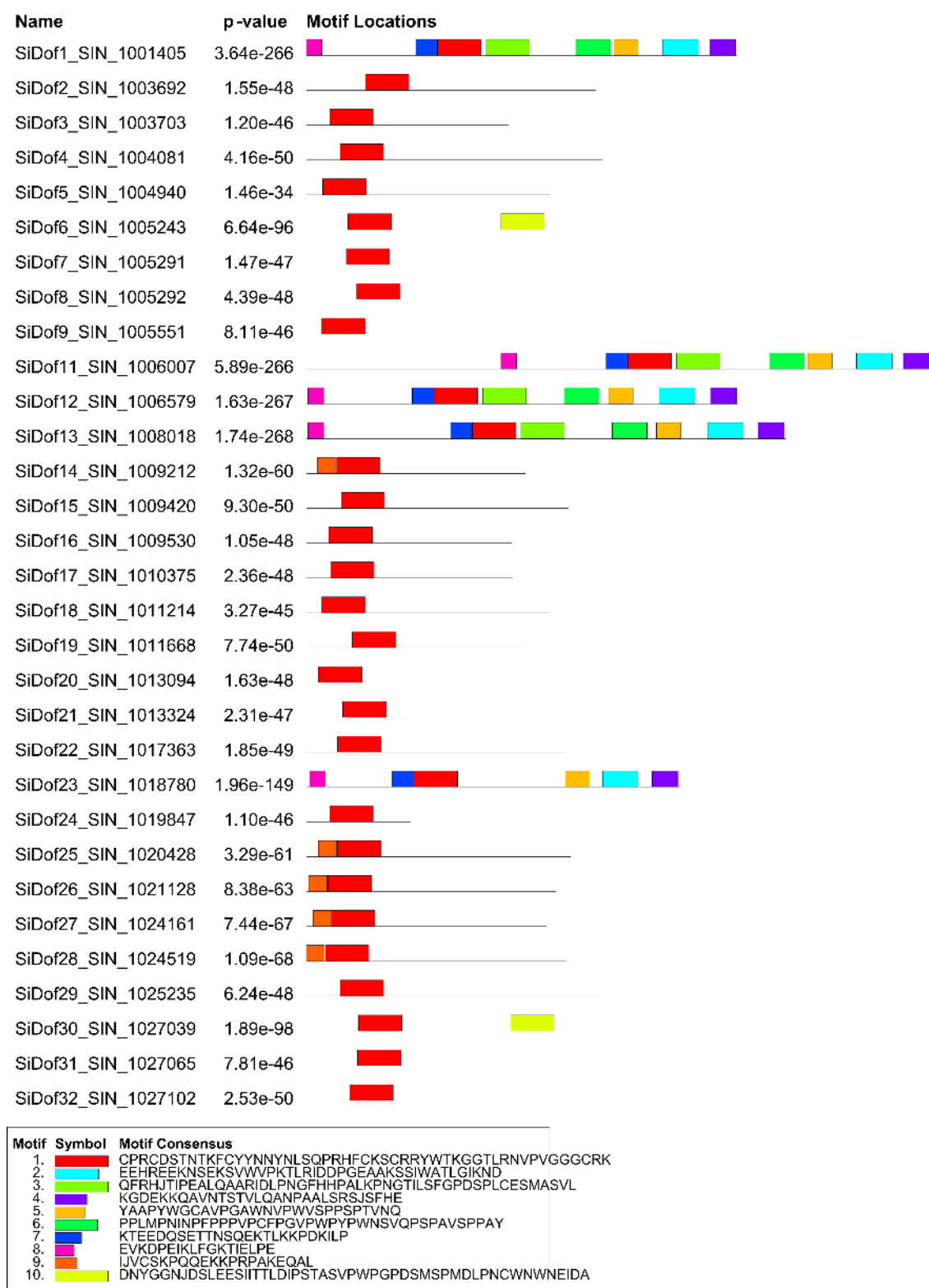


Figure 4

The conserved motifs present in 32 SiDof proteins. Motifs are labelled with different colored rectangles from 1 to 10



Figure 5

Logos of different motifs (1-10) found in the 'zf-Dof' and 'Lipase-GDSL' domains of SiDof proteins

a)

```

SiDof10_SIN_100 -----MEFLKSSTTATAGNMSSRGLMS-----SFMSIPVSDCSNSVYS-----SSAFVH-----
SiDof5_SIN_1004 -----MEQPTGAPQQQQ-----QAS-TPKCPKCAALDTKFRYFNNHSLAQPRHYFCKTCKRQWTVGGNLRNIPVGGKTRN
SiDof31_SIN_102 NLYTTATSSPTHQCPQKQIVADVQNSSFPSSSPTNNSVMQRSNQ-----DPHQPPKCPKCDSTNTKFCYNNNYSLSQPRHYFCKACKRYWTRGGTLRNVPVGGGCRK
SiDof8_SIN_1005 NCEKMMMIAPISSTTTNQWPQNQIQLDHKSFMASSSANKAMEKTTQ-----DPHHQPKCPKCDSTNTKFCYNNNYSLSQPRHYFCKACKRYWTRGGTLRNVPVGGGCRK
SiDof21_SIN_101 -----MFTSTATTSEYSILECPPTIRLVMDRITTSKWK-YN--NIE-IAPNCPKCAALDTKFRYFNNHSLAQPRHYFCKGCRRYWTKGGSRLNVPVGGGCRK
SiDof4_SIN_1004 -----MMSPDNKPAATKPGGKDESQSSNAARKAASS--RPQDPPKCPKCDSTNTKFCYNNNYSLSQPRHYFCKTCKRRYWTKGGALRNVPVGGGCRK
SiDof29_SIN_102 -----MMSPQDNIIIAAKPAGKGESQSSTGRKTSSA--RPQEPKCPKCDSTNTKFCYNNNYSLSQPRHYFCKTCKRRYWTKGGALRNVPVGGGCRK
SiDof7_SIN_1005 -----MEQRRNEEKLLKQQDQQQPPSRRLKPAEMQMAAAHNQH--QPP-PPQKCPKCDSTNTKFCYNNNYSLSQPRHYFCKGCRRYWTKGGTLRNVPVGGGCRK
SiDof15_SIN_100 -----MQDIHAIGGGGGGGGGGRFFGGGDRRLRPHS--HPSHQAKCPKCDSTNTKFCYNNNYSLSQPRHYFCKSCRRYWTKGGVLRNVPVGGGCRK
SiDof22_SIN_101 -----MQDMHAVGGGGGGGRFLGGGDRRLRPHN--HPTHQPKCPKCDSTNTKFCYNNNYSLSQPRHYFCKSCRRYWTKGGVLRNVPVGGGCRK
SiDof6_SIN_1005 -----MGLSSKQVSGDQGLDWGRTLQSGTAEKSKSPSIRRQQLH--QSS-EPKCPKCAALDTKFRYFNNHSLAQPRHYFCKSCRRYWTKGGTLRNVPVGGG-RK
SiDof30_SIN_102 DLDMGLGTELQVSGDHQGLDWGQPLQLQSGKAEKAKPPSGRRQQLK--QSS-EPKCPKCAALDTKFRYFNNHSLAQPRHYFCKSCRRYWTKGGTLRNVPVGGG-RK
SiDof14_SIN_100 -----MATAWPQGIGVVRPVENPRTSTEKKPR--P--NKE-QAKCPKCAALDTKFRYFNNHSLAQPRHYFCKTCKRRYWTKGGTLRNVPVGGGSRK
SiDof2_SIN_1003 GGERSKLSTPVCGLDGSPLSPSSSPSSSSTSVGTAAAAA--AAS-ENKCPKCDSTNTKFCYNNNYSLSQPRHYFCKTCKRRYWTKGGALRNVPVGGGCRK
SiDof23_SIN_101 TLTEAATSEITDDPGTADLEKEPLSPETPKKEEPSNAPQEKTLK--KPD-KIIPCPKCDSTNTKFCYNNNYSLSQPRHYFCKKCRRYWTKGGTLRNVPVGGGCRK
SiDof32_SIN_102 ---MAEVQENNSSGKILFGATITLKERSSSQANDRQESKKAEEK--KPD-KIIPCPKCDSTNTKFCYNNNYSLSQPRHYFCKGCRRYWTKGGALRNVPVGGGCRK
SiDof12_SIN_100 LKE-DATPAADEDKALSADSDARAVQTSKTEDDQIETGNSQEKTLK--KPD-KIIPCPKCDSTNTKFCYNNNYSLSQPRHYFCKKCRRYWTKGGTLRNVPVGGGCRK
SiDof13_SIN_100 LKDSNANPPIVEDSNAP-ADSDAGSMKMSKNEEDQAETSNQSEKSLK--KPD-KIIPCPKCDSTNTKFCYNNNYSLSQPRHYFCKKCRRYWTKGGTLRNVPVGGGCRK
SiDof1_SIN_1001 SHNANALSESDDNKTASIDDDVGSNAPKADNDQSDATNAQQLK--KPD-KIIPCPKCDSTNTKFCYNNNYSLSQPRHYFCKKCRRYWTKGGTLRNVPVGGGCRK
SiDof11_SIN_100 SHNPDAVVEDENPKTPSIDEGIVSEEHPKTENGQSDATNGQQLK--KPD-KIIPCPKCDSTNTKFCYNNNYSLSQPRHYFCKKCRRYWTKGGTLRNVPVGGGCRK
SiDof20_SIN_101 -----MERGWK--P--NVE-VSPACPCGDSNTKFCYNNNYSLSQPRHYFCKGCRRYWTKGGSRLNVPVGGGCRK
SiDof26_SIN_102 -----MDEIITCKSPSVERKVR--P--QKE-QAKCPKCDSTNTKFCYNNNYSLSQPRHYFCKTCKRRYWTEGGSLRNIPVGGGSRK
SiDof19_SIN_101 ---MDSLNTSLSPELITKHFTPIQPNTPEVSDSTHPRLSRYPSRAPEQKCPKCDSTNTKFCYNNNYSLSQPRHYFCKSCRRYWTRGGALRNVPVGGGTRK
SiDof25_SIN_102 -----MDSSPITTPSGDDSLACSRPLIERRLR--P--PHE-QSKCPKCDSTNTKFCYNNNYSLSQPRHYFCKTCKRRYWTKGGTLRNIPVGGGCRK
SiDof27_SIN_102 -----MGSQTLESMFVCTKPPQEKKP--RPAEQAKCPKCDSTNTKFCYNNNYSLSQPRHYFCKSCRRYWTKGGTLRNVPVGGGCRK
SiDof28_SIN_102 -----MLVCTKTQEKKPR--P--ADQ-QAKCPKCDSTNTKFCYNNNYSLSQPRHYFCKSCRRYWTKGGTLRNVPVGGGCRK
SiDof9_SIN_1005 -----MSERARLANIP--MPE-AAKCPKCDSTNTKFCYNNNYSLSQPRHYFCKTCKRRYWTRGGALRNVPVGGGCRK
SiDof18_SIN_101 -----MAERARLANIP--MPE-TAKCPKCDSTNTKFCYNNNYSLSQPRHYFCKTCKRRYWTRGGALRNVPVGGGCRK
SiDof3_SIN_1003 -----MATDPSEKSTRPGQVGAP--PSEPEKCPKCDSTNTKFCYNNNYSLSQPRHYFCKSCRRYWTHSGTLRDVPVGGGSRK
SiDof24_SIN_101 -----MATDTSLRKATKSAQVGAP--PPEPEKCPKCDSTNTKFCYNNNYSLSQPRHYFCKSCRRYWTHSGTLRDVPVGGGSRK
SiDof17_SIN_101 -----MQDPSLYSIIKAQVQQLP--Q--FPEQEKCPKCDSTNTKFCYNNNYSLSQPRHYFCKSCRRYWTKGGALRNIPVGGGSRK
SiDof16_SIN_100 -----MQNSSFSYQINAQAQQPQ--FPEQEKCPKCDSTNTKFCYNNNYSLSQPRHYFCKSCRRYWTKGGTLRNIPVGGGSRK

```

b)

CPRCSTNTKFCYNNNYSLSQPRHYFCKGCRRYWTKGGSRLNVPVGGGCRK

Figure 6

Multiple protein alignment of Dof proteins in sesame (a). The 'zf-Dof' domain is labeled below the alignment (b). The alignment was made with web tool (<http://www.phylogeny.fr/>)

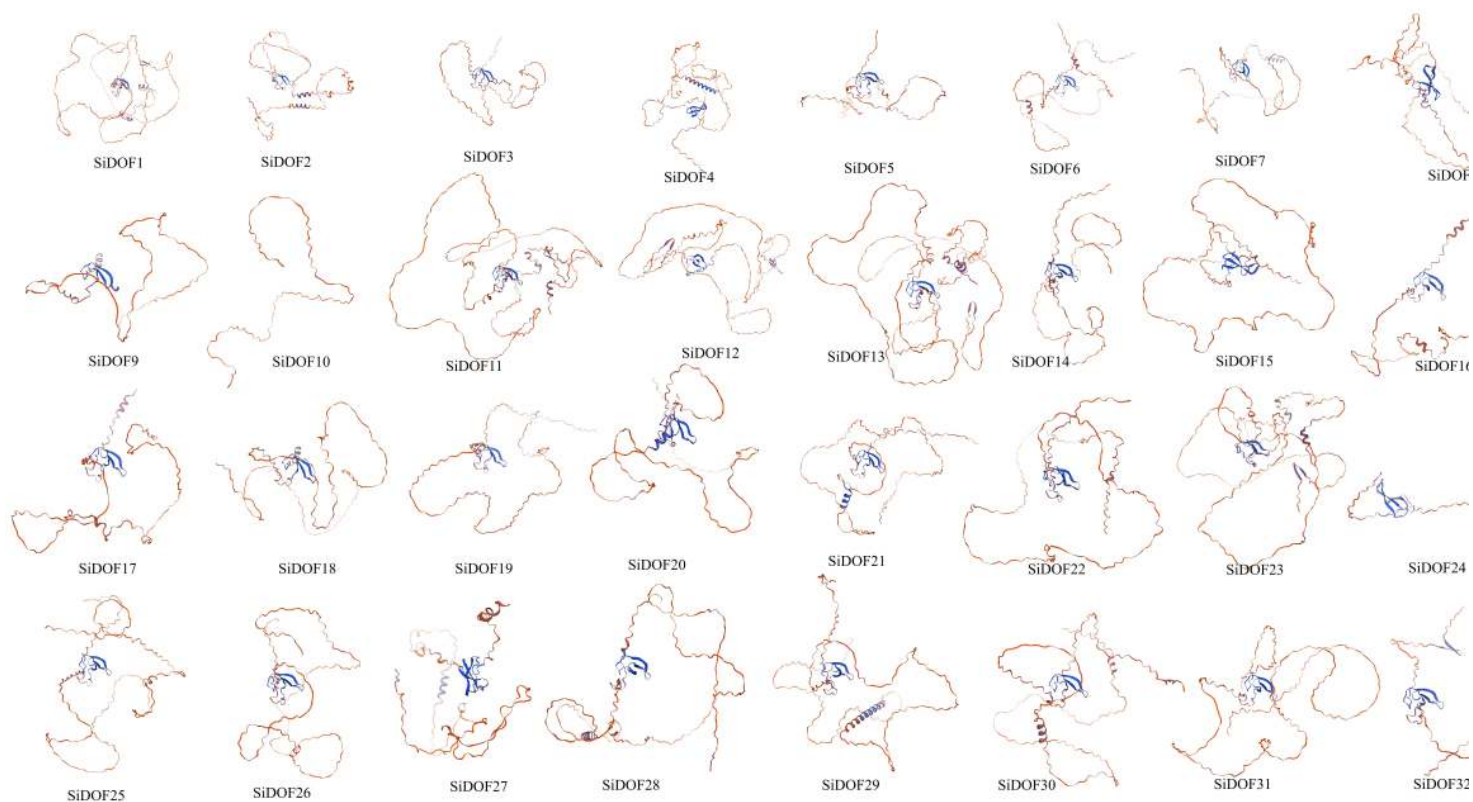


Figure 7

The 'three-dimensional' (3D) structures of SiDof proteins. The secondary structure elements α -helices, β -sheets, and coils (disordered structures) are indicated for the predicted 3D structures of SiDof proteins. Spiral parts indicate α -helix and Ribbon like structures indicate β -sheet

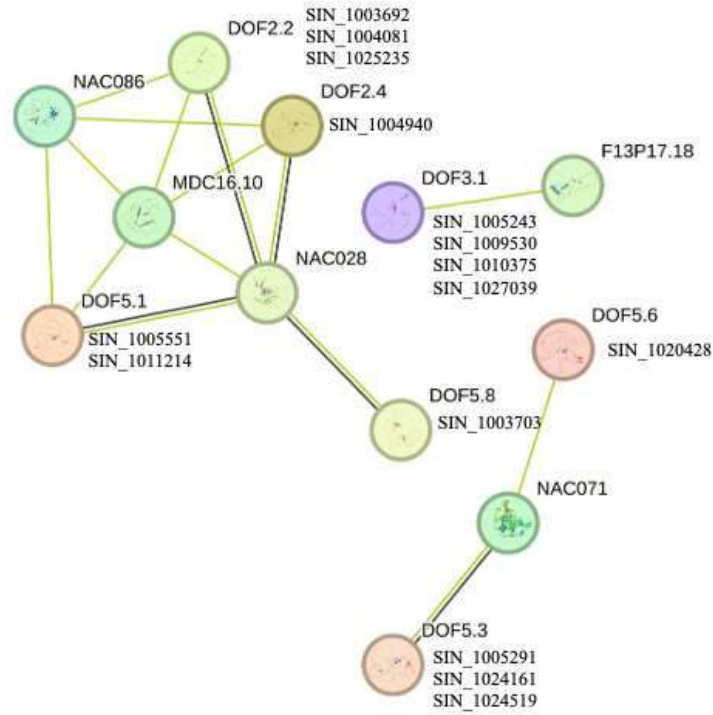


Figure 8

‘Protein-protein interaction’ analysis of SiDof proteins. The red points in the figure represent the SiDof proteins, and the blue points illustrate the proteins that interacted with the SiDof proteins. The green lines indicate the interactions between the SiDof proteins.

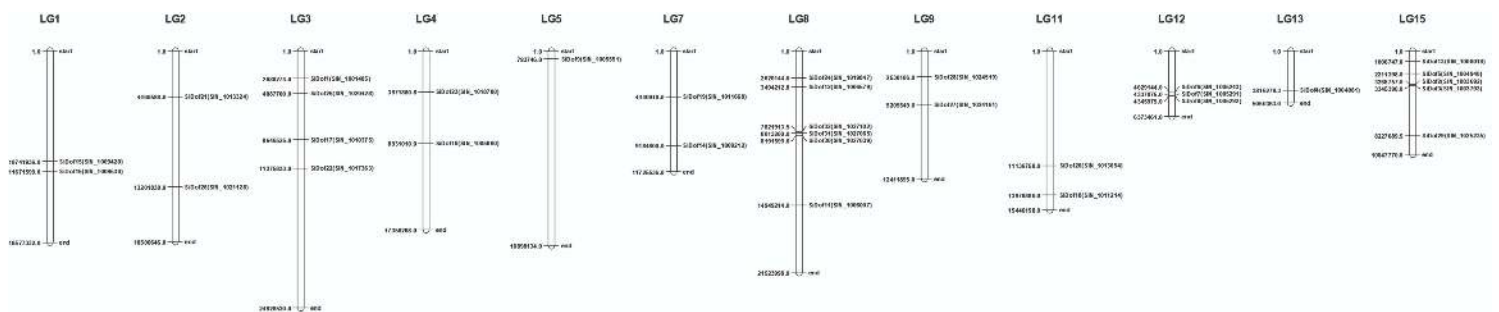


Figure 9

Genomic location of 32 *Dof* genes in sesame genome

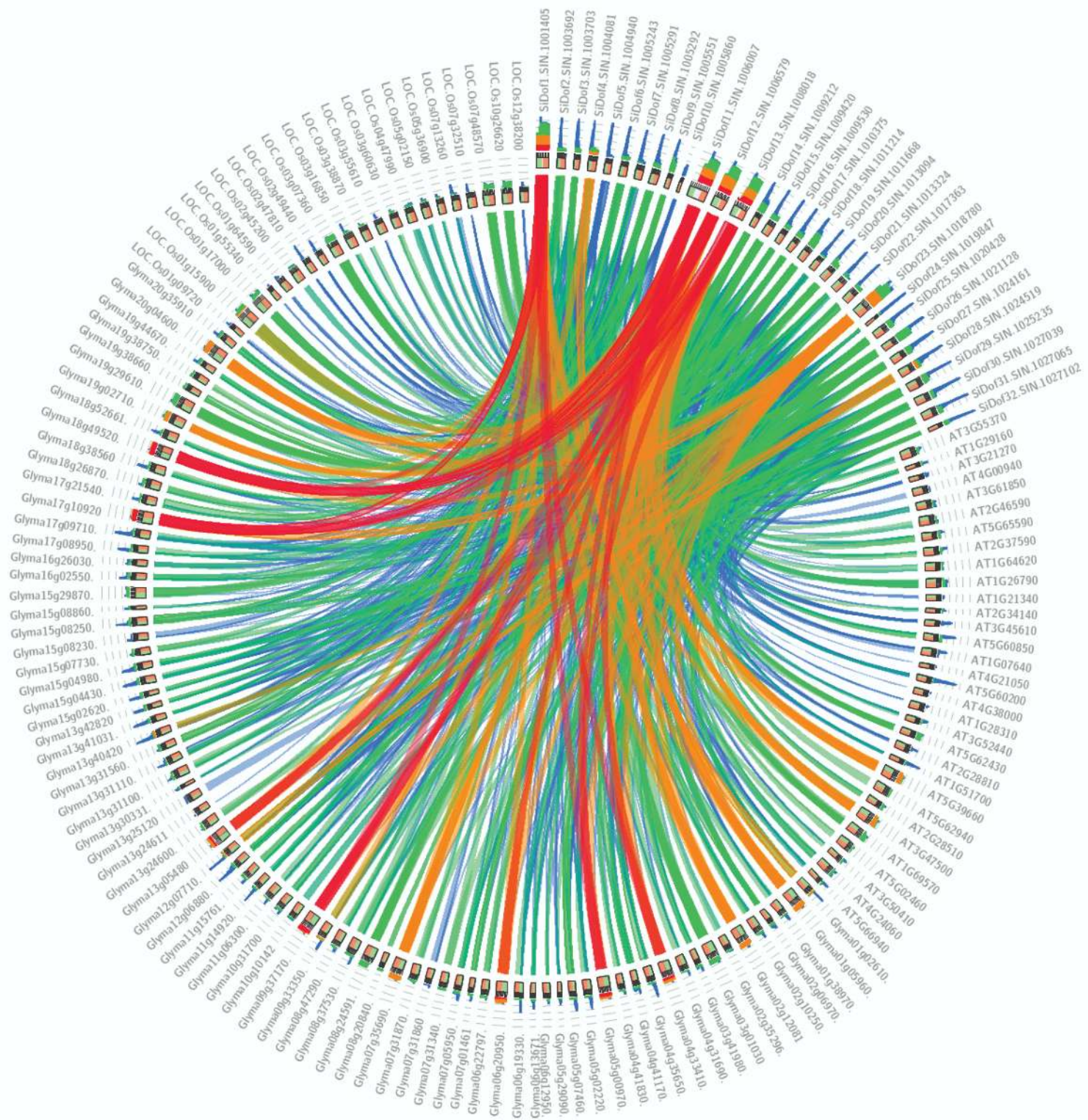


Figure 10

Syntenic association of *Dof* genes among *S. indicum*, *A. thaliana*, *G. max* and *O. sativa* representing the level of conservation. The coloured ribbons indicate the level and intensity of evolutionary conservation among different *SiDof* genes. The maximum strength is from red to blue

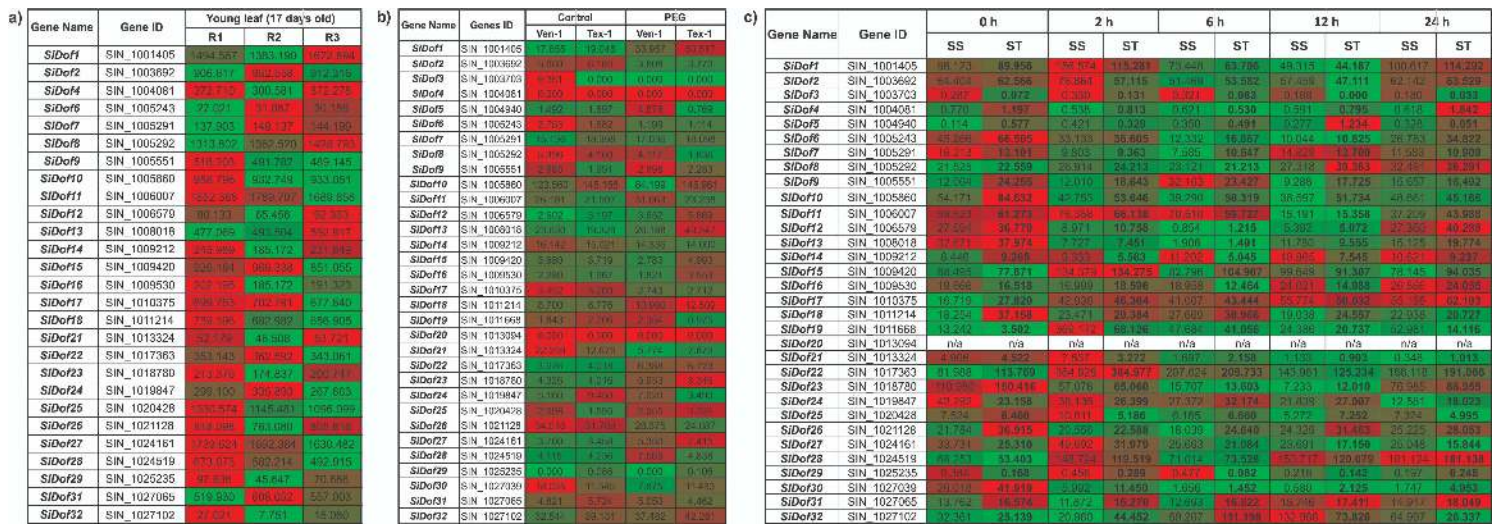


Figure 11

Heatmap representation of the FMKM/RPKM expression pattern of *SiDof* genes in *S. indicum*. The FMKM value was retrieved from GEO dataset of the NCBI database. a) Expression levels in young leaf (GSE164511), b) Drought (PEG-treated root) (GSE148340) and c) RPKM expression in salt-tolerant and salt-sensitive shoots under salt-stress condition (Zhang et al., 2020). The maximum strength is from red to green