Agricultural Sciences Journal
Available online at http://asj.mnsuam.edu.pk/index.php
ISSN 2707-9716 Print
ISSN 2707-9724 Online
https://doi.org/10.56520/asj.v5i1.237

Research Article

GENOME-WIDE IDENTIFICATION AND CHARACTERIZATION OF PLANT-SPECIFIC *DOF* TRANSCRIPTION FACTOR GENE FAMILY IN CASHEW (*ANACARDIUM OCCIDENTALE*)

Maira Shakeel[1], Muhammad Shafiq[1]*, Syed Agha Armaghan Asad Abbas[2], Muhammad Haseeb[3], Numan Ali[3], Aqdas Batool[4], Muhammad Rizwan Tariq[4], Muhammad Saleem Haider[5]
[1]Department of Horticulture, University of the Punjab, Lahore, Pakistan
[2]Department of Agronomy, University of the Punjab, Lahore, Pakistan
[3]Department of Plant Pathology, University of the Punjab, Lahore, Pakistan
[4]Department of Food Sciences, University of the Punjab, Lahore, Pakistan
[5]Faculty of Agricultural Sciences, University of the Punjab, Lahore, Pakistan
**Corresponding Author; shafiq.iags@pu.edu.pk*

Abstract

DNA binding with one-finger (Dof) protein belongs to the plant-specific transcription factors (PSTFs) gene family. These transcription factors have a variety of roles in many biological processes in plants. However, there is limited research on their role in Cashew. A total of 67 *Dof* genes were found in the cashew genome and were classified into 11 subgroups (A, B1, B2, C1, C2, D, E1, E2, F1, F2, and F3) by comparing them with *Dof* genes from Arabidopsis and lettuce. Cashew *Dof* genes were present in 9 of these subgroups, except for A and F1. This article provides a detailed discussion of the gene structures, chromosome positions, phylogeny, subcellular localization, cis-regulatory analysis, protein motifs, and evolutionary patterns of *Dof* genes in cashew. The only type of duplication found in cashew was segmental duplication, which mainly contributes to the large *Dof* gene family. The analysis of cis-regulatory elements (CREs) revealed the presence of light, ethylene, seed, circadian, meristem, and auxin-sensitive elements, which are particularly sensitive to these factors. The article also includes a comparative analysis of the evolutionary or phylogenetic relationships between *Dof* genes from lettuce, cashew, and Arabidopsis. This study provides a comprehensive understanding of the *Dof* gene family in cashew and can serve as a guide for functional analysis and cloning of its gene family members.

Keywords: Cashew, Dof, Plant specific transcription factor, Genome-wide, Bioinformatics

(Received: 19, January 2023, Accepted: 20, April 2023) Cite as: Shakeel. M., Shafiq. M., Abbas. S. A. A. A., Haseeb. M., Ali. N., Batool. A., Tariq. M. R., Haider. M. S., 2023 Genome-wide identification and characterization of plant-specific *Dof* transcription factor gene family in Cashew (*Anacardium occidentale*). Agric. Sci. J. 5(1): 83-107.

1. INTRODUCTION

The DNA binds with one finger Dof factor and performs an efficient role in the growth of plant and its development. (Gupta et al., 2015; Malviya et al., 2015). A conserved domain of 50-52 amino acids with a C2C2-type standard zinc finger motif that is a DNA binding motif, is present at the N-terminus of Dof gene family members. (Song et al., 2016b; Zou et al., 2013). The Dof transcription factors are involved in various major roles in plants, such as assimilation of nitrogen (Wang et al., 2013; Yanagisawa et al., 2004), accrual of proteins that assemble in the seed (Dong et al., 2007), carbon metabolism (Gupta et al., 2015), association with the intracellular trafficking of protein (Chen et al., 2013), endosperm specific response (Diaz et al., 2005), defence response (Takano et al., 2013), sprouting of seed (Noguero et al., 2013), drought and salt tolerance (Ma et al., 2015; Ayoub et al., 2021), balancing of photoperiodic flowering (Fornara et al., 2009), regulation of branch and shoot alongwith seed coat formation (Zou et al., 2013), regulation of genes linked to stomata and morphogenesis (Negi et al., 2013), and are concerned with the circadian cycle as well (Yang et al., 2011).The existence of a wide variety of *Dof* genes numerically in different crops suggests a high probability of feature diversification. The relative phylogeny of the *Arabidopsis thaliana* and *Oryza sativa Dof* gene families revealed 36 and 30 *Dof* genes respectively, in a genome-wide study (Lijavetzky et al., 2003a). Similarly, 34, 36, and 41 *Dof* genes from *Solanum lycopersicum*, *Arabidopsis thaliana*, and *Populus ciliata*, correspondingly, are used for studying the evolutionary features of *Dof* gene families (Cai et al., 2013; Yang and Tuskan, 2006). The total potential *Dof* genes found in *Solanum tuberosum* (Venkatesh and Park, 2015), *Hordeum vulgare*, *Capsicum annuum* (Wu et al., 2016), *Chrysanthemum morifolium (Song et al., 2016a),* and *Cucurbita sp.* are 35, 24, 34, 33, 20, and 36, respectively (Hernando-Amado et al., 2012; Mena et al., 2002; Moreno-Risueno et al., 2007a). (*Anacardium occidentale* L.) belong to the family Anacardiaceous and is a tropical nut tree that is thought to have originated in Central Brazil (Danella Figo et al., 2019; Fernandes and Mesquita, 1995; Hasnain et al., 2018; Mah et al., 2017). The cashew crop expanded rapidly after European conquerors, mainly Portuguese, introduced it to Asia and Africa. It has since been a significant agricultural commodity exported by many countries. Approximately 4.7 million tonnes of raw cashew nuts were grown globally in 2011, which was equally divided between Africa and Asia, while in South America, roundabout 1.8 million tonnes of cashew apples were produced, with Brazil being the most prominent producer. Exponentially rising production volumes and prices of cashew kernel over the past couple of decades indicates the cashew crop's growing popularity and its importance. (Nair et al., 2009; Neto et al., 2001; Tyman and Kiong, 1978). Even though cashew is becoming a more important economic, balanced and nutrient rich fruit crop which is grown majorly for oil, no substantial study on its *Dof* TFs has been released yet. Using a variety of bioinformatics tools, the main goal of this research was to find and classify *Dof* TF family genes in the cashew genome. To summarize, *Dof* genes in the cashew genome were discovered using a systematic approach. The presence of conserved domains and cis-regulatory elements, as well as their chromosomal distribution, intron/exon distribution pattern, and presence of conserved domains, were all investigated. For the purpose of determining the orthologous relationship and their probable role, a comparative phylogenetic study of *Dof* from cashew, lettuce, and Arabidopsis was also performed. Our data and findings create a foundation to further study the evolutionary and functional characteristics of the *Dof* gene family in Cashew.

2. MATERIALS AND METHODS

2.1. Database search and retrieval of sequence

To identify *Dof* genes in the cashew genome, the amino acid sequence of a common *Dof* domain (Pfam i.e. PF02701) based on hidden Markov models (HMMs) (http://pfam.xfam.org/) (Finn et al., 2014) and the 59 AA *Dof* domain sequence of *A. thaliana* (Accession no.NP_175581) were used to search against the predicted cashew gene database collection at Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). The retrieved amino acid sequences were subjected to NCBI CDD (Conserved Domain Database) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Lu et al., 2020) with the default parameters.

2.2. Determination of physio-chemical properties of cashew Dof proteins

The ProtParam tool (http://web.expasy.org/protparam/) was used to figure out the length of amino acids, molecular weight, and pI values of *AoDof* proteins (http://web.expasy.org/protparam/) (Gasteiger et al., 2005). The Phytozome database was used to obtain gene IDs, chromosomal locations, gene and protein sequences. These *AoDof* genes were renamed in the order in which they were discovered.

2.3. Gene structure analysis

The genomic, peptide and coding sequences of known genes were extracted from the database for the investigation of the intron/exon structure of *AoDof*s. Then gene structure was drawn by using these sequences, with the help of Gene Structure Display Server (GSDS v2.0) (http://gsds.gao-lab.org/) (Hu et al., 2015).

2.4. Multiple sequence alignment and phylogenetic analysis

To perform a phylogenetic analysis, the amino acid sequences of *Dof* proteins were aligned using ClustalW version 2.1 (Thompson et al., 2003; Thompson et al., 1994). The phylogeny was then constructed using the neighbour-joining (NJ) method with bootstrapping set at 1000 replications and partial deletion, through the MEGA vX.0 program (Kumar et al., 2018). In total, 67 cashew Dof, 35 Arabidopsis, and 48 lettuce *Dof* protein sequences were included in the analysis.

2.5. Cis-regulatory elements and conserved motifs recognition

A 1000-bp upstream sequence from the initiation codon of the putative AoDof genes was used for examining the promoter region. Cis-regulatory elements were predicted from these sequences by Plant Care database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Rombauts et al., 1999).

The concluded protein sequences of the *AoDof* genes were analysed by the use of a Multiple EM for Motif Elicitation (MEME) programme (https://meme-suite.org/meme/) (Bailey et al., 2015), limiting the maximum number of motifs to 10. The default values for the motif were set to a minimum of 6 widths and a maximum of 50 widths, as well as other variables.

2.6. Gene ontology and subcellular localization analysis

Gene ontology (GO) enrichment information of each *AoDof* gene was retrieved from the PhytoMine available in the Phytozome database (https://phytozome.jgi.doe.gov/phytomine/begin.do) to investigate the particular involvement of the *Dof* genes of cashew in terms of molecular functions (MF), biological functions (BP), and certain cellular components (CC). Furthermore, a web-based tool called WoLF PSORT (https://wolfpsort.hgc.jp/) (Horton et al., 2006) was used for predicting the subcellular location of the 67 *Dof* proteins of cashew. Nuclear Localization Signal Database (NLSdb) (https://rostlab.org/services/nlsdb/) was used to predict the signals of nuclear localization in Dof proteins of cashew (Cokol et al., 2000).

2.7. Gene duplication and calculation of synonymous (Ks) and non-synonymous (Ka) substitution rates:

Phylogenetic, motif, and domain analysis data were used to produce gene pairs for the *AoDof* gene, which were then used to compute the Ka and Ks substitution rates using TBtools (Chen et al., 2020b). Gene pairs along with CDS sequence and protein sequence of *Dof*-like genes of cashew were used. The ratio of Ka and Ks was used for determining the molecular evolutionary rates of a single gene pair. The ratio Ka/Ks < 1 typically applies to purifying selection, Ka/Ks = 1 to neutral selection while Ka/Ks > 1 to positive selection (Yang and Bielawski, 2000). TBtools software (https://github.com/CJ-Chen/TBtools) was utilized to map the *AoDof* genes on scaffolds to display their distribution. The genes were either present on the same chromosome or they were a part of unrelated chromosomes, this location suggested the tandem duplication and segmental duplication respectively. The duplicated genes were marked with a line joining on the map.

3. RESULTS

3.1. Identification of the *Dof* genes in Cashew

A total of 67 Dof genes were identified in cashew using the BLAST search tool on the Phytozome database. These genes were grouped into different gene families based on their structure and function. Gene family analysis (GFA) of the *Dof* genes revealed that all non-redundant *Dof* protein sequences in cashew had four highly conserved cysteine residues that bind to zinc ions, which is a general characteristic of *Dof* proteins.

3.2. Conservation analysis of Dof genes in cashew:

Analysis of the conservation of *Dof* genes in cashew showed that within the strongly conserved sequences of the *Dof* domain, 32 out of 50 amino acids (i.e., 64% of the total AoDof amino acids) were 100% conserved in all *Dof* domain sequences. The conserved amino acids included Cys1, Pro2, Arg3, Cys4, Ser6, Thr9, Lys10, Phe11, Cys12, Tyr13, Asn15, Asn16, Tyr17, Gln21, Pro22, Arg23, Phe25, Cys26, Lys27, Cys29, Arg31, Tyr32, Tryp33, Thr34, Gly36, Gly37, Arg40, Asn41, Pro43, Gly45, Gly47, and Arg49.

3.3. Characterization of *Dof* proteins in cashew:

The *Dof* genes in cashew encode proteins with molecular weights ranging from 17.37 to 57.45 kDa and lengths ranging from 155 to 524 amino acids. AoDof58 was found to be the smallest protein, while *AoDof59* was the longest. The isoelectric points of the identified proteins ranged from 4.58 to 9.87, with *AoDof26* and *AoDof27* having the respective extreme values.

3.4. Nuclear localization signal analysis of Dof proteins in cashew:

Of the 67 identified Dof proteins, 20 (*AoDof25, AoDof28, AoDof41, AoDof47, AoDof48, AoDof49, AoDof54, AoDof55, AoDof56, AoDof57, AoDof58, AoDof59, AoDof60, AoDof61, AoDof62, AoDof63, AoDof64, AoDof65*, and *AoDof66*) were found to have a nuclear localization signal (NLS) when analysed using the NLSDB software. The NLS signal for 17 out of these 20 proteins was found to be GAGRRK, while for *AoDof25* and *AoDof28*, it was RNKRN, and for *AoDof48*, it was KKPDR.

3.5. Gene structures and recognition of conserved motifs and domain

The exon-intron structures of cashew *Dof* genes were examined in detail, along with phylogeny, and the pattern of gene structure was found to be consistent with the results of the phylogenetic study. The introns in cashew *Dof genes* ranged from zero to three (Fig.2, Table 1). Out of a total of 67 cashew *Dof* genes, 25 (37.3%) were intron-less, 34 (50.7%) had one intron, 7 (10.4%) had two introns, and one gene (*AoDof27*) had three introns (Table 1, and Fig. 2). In family D, all of the *AoDof* genes were intron-less, while the *AoDof* genes in subfamilies B1 and B2 had introns ranging from zero to two. Except for *AoDof27*, which has three introns, all *AoDof* genes in subfamily C1 had one intron. In *AoDof* genes in subfamily C2, E1, E2, F2, and F3, the number of introns ranged from zero to one. Ten conserved motifs were identified in all cashew *Dof* proteins using the MEME software (Fig 3). The *Dof* domain was found in all 67 cashew *Dof* genes. It was observed that *Dof* genes from the

Figure. 1 *Dof* domains are highly conserved across all 67 Dof proteins in cashew. All of the cashew *Dof* domains were aligned to create the sequence logos. The ClustalW programme (Thompson et al., 2003) in Mega-X was used to perform multiple alignment analysis of 67 typical cashew *Dof* domains. The SeqLogo programme in TBtools was used to produce the signature of the aligned sequences (Chen et al., 2020a). The bit score indicates the information content for each position in the sequence (Cys) are conserved in the *Dof* domain and can be found at positions 1, 4, 12, 26, and 29. The red line represents the zinc finger motif.

same clade had similar motifs, indicating that these conserved motifs are crucial for highly specific group or subgroup activities. The presence of related motifs in various *Dof* genes suggests that they may have arisen due to gene expansion (Fig 3).

3.6. Comparative phylogenetic relatedness of cashew *Dof* gene family with Arabidopsis

To investigate the evolutionary relationships of the cashew *Dof* gene family with Arabidopsis, a Neighbour-Joining (NJ) phylogenetic tree was constructed from full-length aligned peptide sequences of *AoDof* TFs, *Arabidopsis thaliana*, and *Lactuca sativa* in MEGA-X. The analysis classified 67 *AoDof* proteins into 11 subgroups: A, B1, B2, C1, C2, D, E1, E2, F1, F2, and F3. Subgroup A comprised 14 Dof proteins, out of which 10 were Arabidopsis (AT5G66940 D2, AT3G50410 OBP1 D2, AT4G21030 DOF4 2 C3, AT4G21050 DOF4.4 c3, AT2G46590 DAG2 C2.1, AT3G61850 DAG1 C2.1, AT1G28310 B2, AT4G38000 B2, AT4G21080 DOF4.5 C3 and AT4G21040 DOF4.3 C3) and 4 were from lettuce (*LsDOF15*, *LsDOF9*, *LsDOF4*, and *LsDOF3*). No *AoDof* protein belonged to this subgroup. B1 subgroup comprised a total of 34 Dof proteins, out of which 8 were Arabidopsis (AT1G64620 C2.1, AT5G62430 DOF5.5 CDF1 D1, AT1G69570 D1, AT1G26790 D1, AT2G34140 D1, AT1G29160 D1, AT3G47500 CDF3 D1 and AT5G39660 CDF2 D1), 9 were lettuce (*LsDOF43*, *LsDOF45*, *LsDOF42*, *LsDOF44*, *LsDOF46*, *LsDOF35*, *LsDOF36*, *LsDOF39* and *LsDOF40*), and 17 were from cashew (*AoDof41*, *AoDof47*, *AoDof48*, *AoDf49*, *AoDof54*, *AoDof55*, *AoDof56*, *AoDof57*, *AoDof58*, *AoDof59*, *AoDof60*, *AoDof61*, *AoDof62*, *AoDof63*, *AoDof64*, *AoDof65*, and *AoDof66*). The B2 subgroup contained only 7 *Dof* proteins, out of which 3 were from cashew (*AoDof38*, *AoDof42*, and *AoDof67*) and 4 were from lettuce (*LsDOF25*, *LsDOF26*, *LsDOF41*, and *LsDOF30*). This subgroup of *Dof* proteins did not contain any from Arabidopsis. Subgroup C1 contains 19 *Dof* proteins, out of which 5 are from Arabidopsis (AT5G65590 B2, AT3G55370 OBP3 B1, AT5G02460 B1, AT2G28810 B1, and AT2G37590 DOF24B1), 6 are from lettuce (*LsDOF14*, *LsDOF19*, *LsDOF20*, *LsDOF21*, *LsDOF22*, and *LsDOF23*), and 8 are from cashew (*AoDof29*, *AoDof24*, *AoDof23*, *AoDof19*, *AoDof21*, *AoDof25*, *AoDof27*, and *AoDof28*). Subgroup C2 has 7 Dof proteins, out of which only one is from Arabidopsis (AT4G00940 c2.1), 2 are from lettuce (*LsDOF31* and *LsDOF32*), and the remaining 4 are from cashew (*AoDof31*, *AoDof30*, *AoDof20*, and *AoDof22*). Subgroup D consists of 6 *Dof* proteins, out of which only one belongs to Arabidopsis (AT1G07640 OBP2 B1), one to lettuce (*LsDOF37*), and 4 belong to cashew (*AoDof7*, *AoDof43*, *AoDof11*, and *AoDof45*).

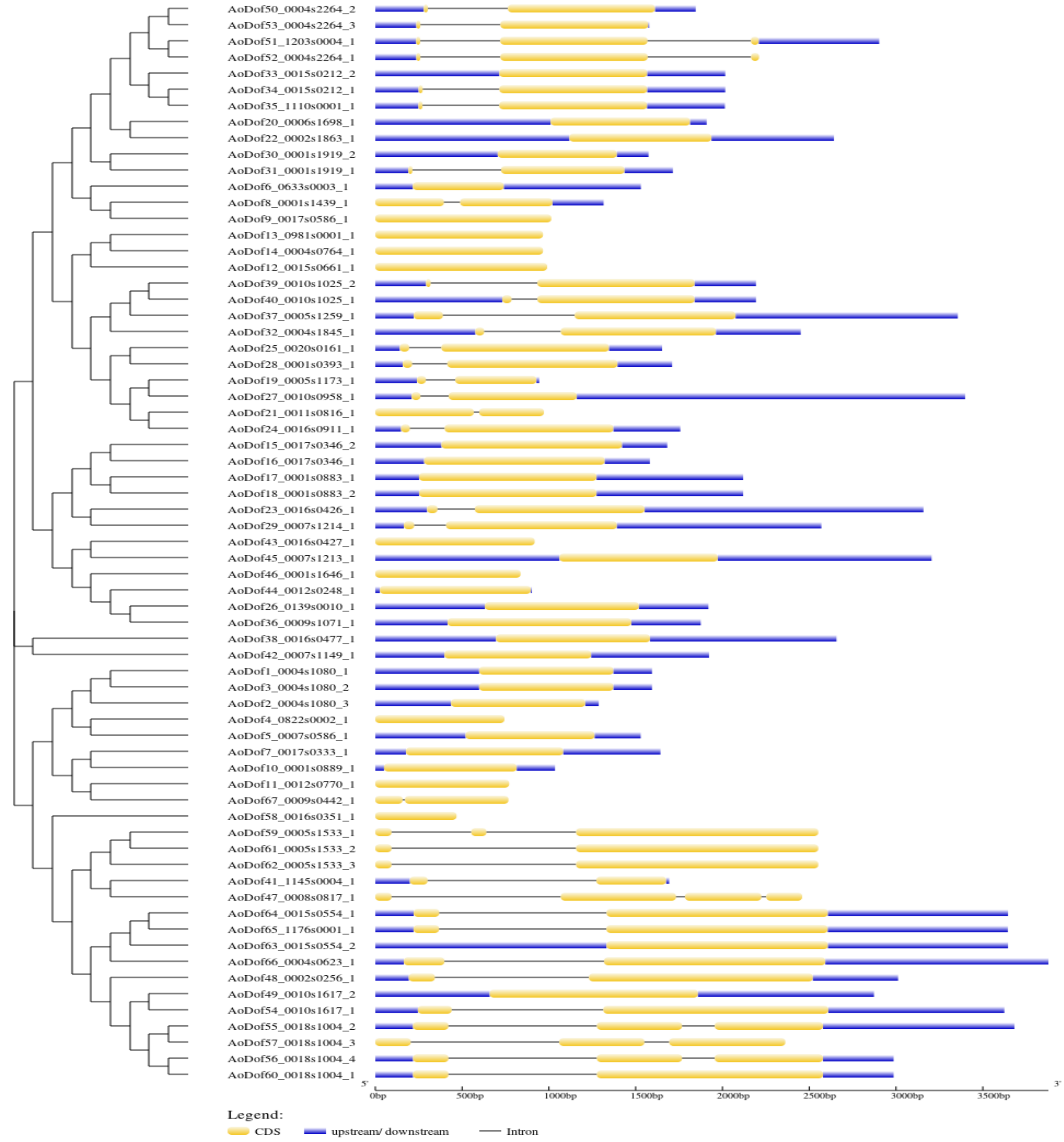


Figure. 2 Phylogenetic relationships and gene structures of the *Dof* genes of cashew. The full-length sequences of cashew *Dof* were used to establish the phylogenetic tree. Arrangements of exons and introns in the cashew *Dof* genes were also found out. Exons are marked by yellow bars, introns are indicated by black lines, and upstream and downstream regions are indicated by blue bars

Subgroup E1 consists of 11 *Dof* proteins, out of which one belongs to Arabidopsis (AT4G24060 C2.1), one to lettuce (*LsDOF28*), and the remaining 9 are from cashew (*AoDof53*, *AoDof52*, *AoDof51*, *AoDof50*, *AoDof35*, *AoDof34*, *AoDof33*, *AoDof4*, and *AoDof5*).

Subgroup E2 consists of 19 *Dof* proteins, out of which 5 are from Arabidopsis (AT1G51700 DOF1 A, AT1G21340 C2.2, AT5G62940 HCA2 D2, AT5G60200 C1, and AT3G45610 DOF6 C1), 9 are from lettuce (*LsDOF38*, *LsDOF24*, *LsDOF29*, *LsDOF11*, *LsDOF7*, *LsDOF1*, *LsDOF2*, *LsDOF6*, and *LsDOF8*), and 5 are from cashew (*AoDof3*, *AoDof2*, *AoDof1*, *AoDof10*, and *AoDof32*).

Subgroup F1 consists of only 3 *Dof* proteins, out of which one belongs to Arabidopsis (AT3G21270 DOF2 A) and

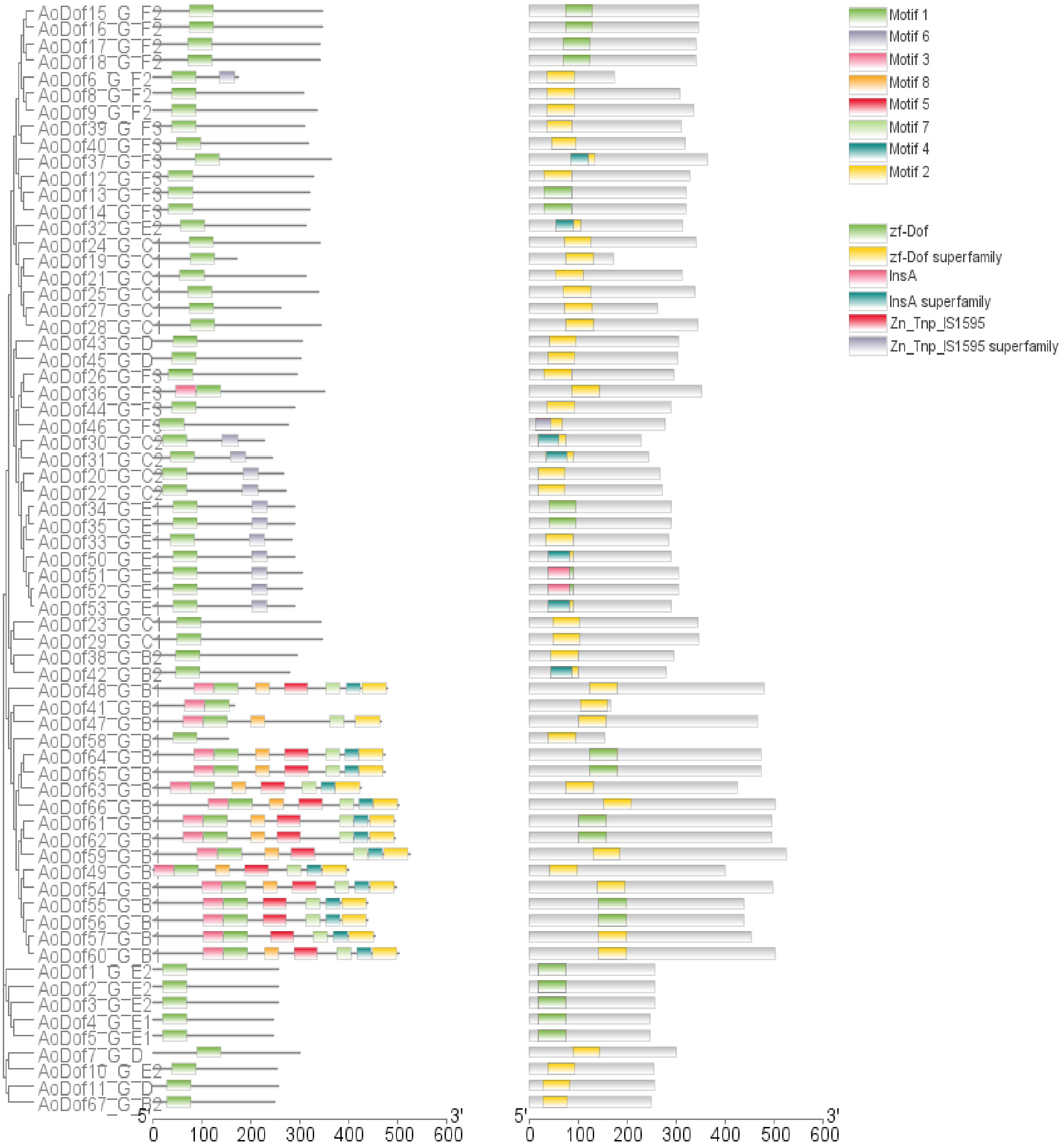


Figure. 3 MEME software was used to identify motifs using the deduced amino-acid sequences of the 67 *AoDof*s and interlinking them with phylogenetic tree and *Dof* domain using NCBI CDD (Bailey et al., 2015).

two belong to lettuce (*LsDOF33* and *LsDOF4*). Subgroup F2 consists of a total of 10 Dof proteins, out of which 3 belong to lettuce (*LsDOF13*, *LsDOF10*, and *LsDOF5*), and the remaining 7 are from cashew (*AoDof18*, *AoDof17*, *AoDof16*, *AoDof15*, *AoDof9*, *AoDof8*, and *AoDof6*).A total of 20 *Dof* proteins were present in Subgroup F3, including 3 from Arabidopsis (AT5G60850 OBP4A, AT3G52440 c2.2, and AT2G28510 C1), 7 from lettuce (*LsDOF12*, *LsDOF16*, *LsDOF17*, *LsDOF18*, *LsDOF27*, *LsDOF48*, and *LsDOF49*), and 10 from cashew (*AoDof12*, *AoDof13*, *AoDof14*, *AoDof26*, *AoDof36*, *AoDof37*, *AoDof39*, *AoDof40*, *AoDof44*, and *AoDof46*). This information is presented in (Figure 4) and also visually represented in (Figure 5).

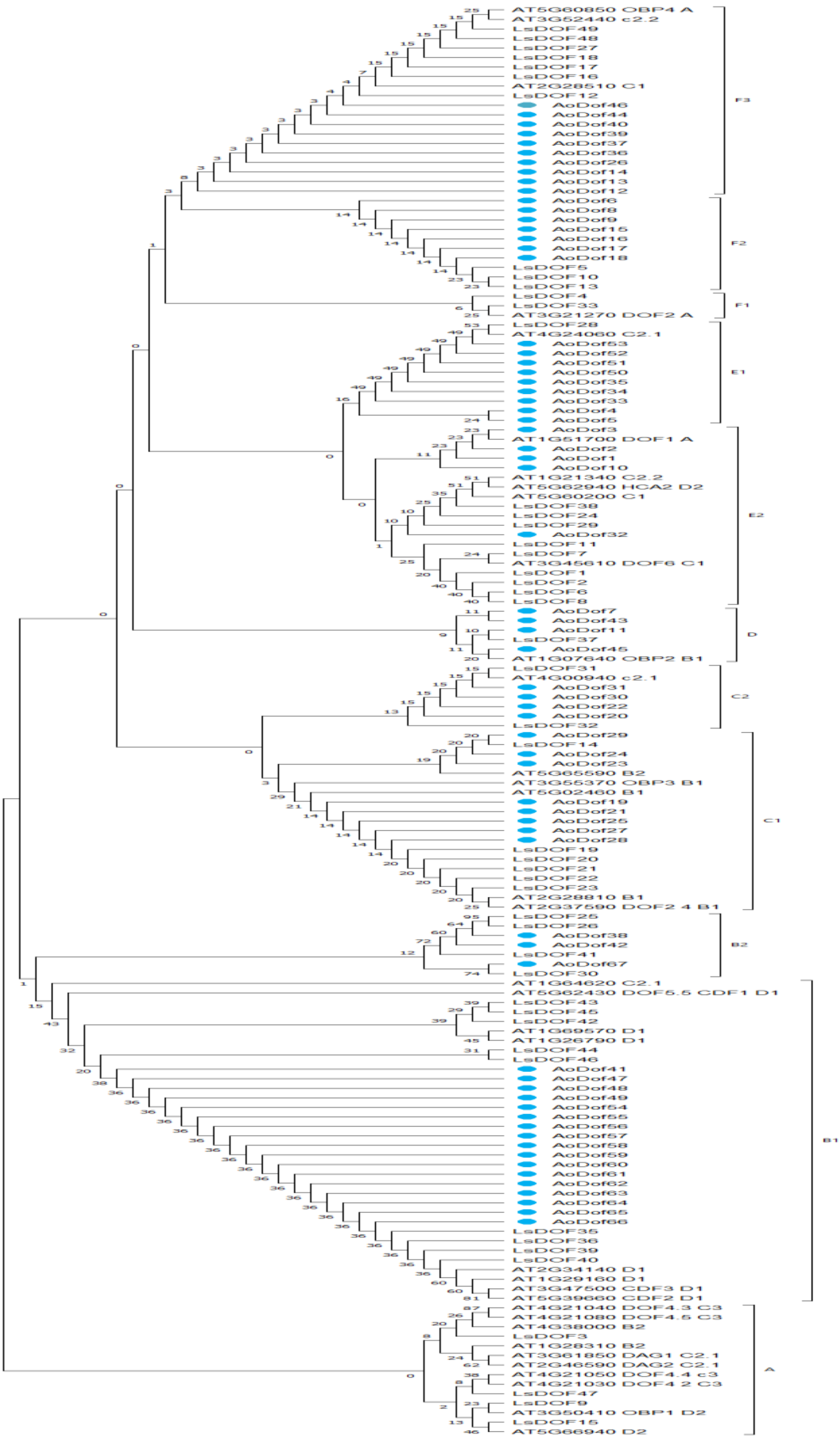


Figure. 4 Phylogenetic and evolutionary relationship between *Dof* gene family member of cashew, Arabidopsis and lettuce. Blue circles indicate proteins that are *AoDof* proteins. Using the UPGMA method and 1000 Bootstrap, the evolutionary past was inferred. MEGA X was used to perform evolutionary studies (Kumar et al., 2018; Kumar et al., 1994; Mello, 2018; S. et al., 2018).

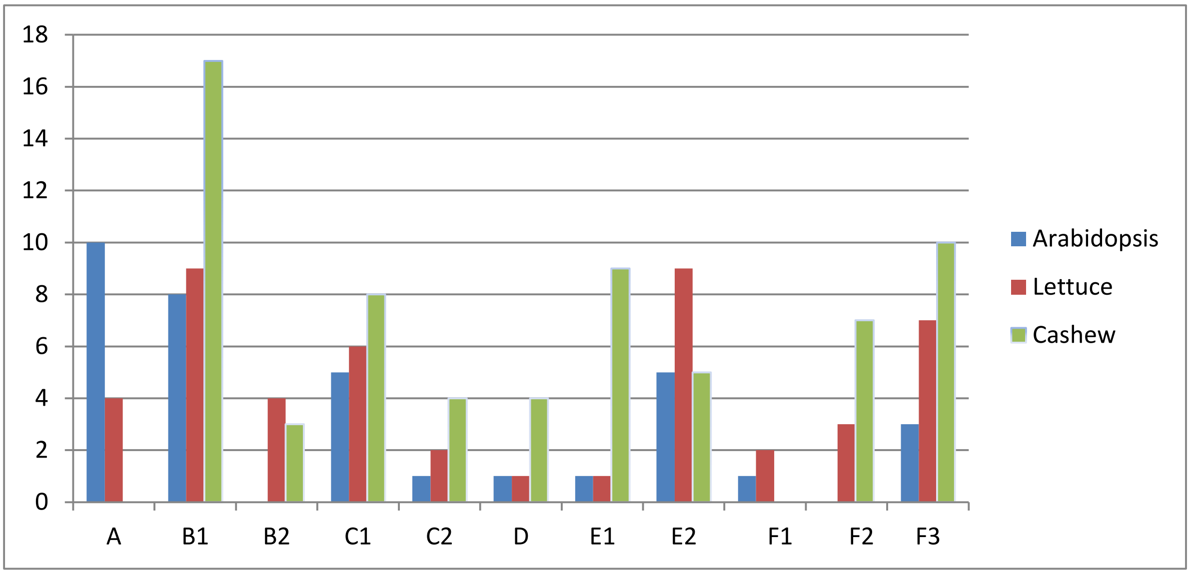


Figure. 5 Detailed information on the *Dof* members in each subgroup present in cashew, Arabidopsis and lettuce. X- axis represents groups and Y-axis number of species in group.

3.7. Location of chromosomes and assessment of gene duplication of cashew *Dof* genes

The study examined the chromosome distribution of Cashew *Dof* genes and identified the presence of *AoDof* genes on various scaffolds. The highest number of *Dof* genes were found on scaffold 4 with nine genes, while scaffold 1 had eight, and scaffolds 5, 10, 15, and 16 each had five *Dof* genes. Four *Dof* genes were found on scaffolds 7, 17, and 18, while scaffolds 2, 9, and 12 had two *Dof* genes. Only one *Dof* gene was located on scaffolds 6, 8, 11, 20, 13, 633, 822, 981, 1110, 1145, 1176, and 1203, which were also the scaffolds with the lowest number of *Dof* genes per scaffold ratio in the Cashew genome. The study also investigated the replication of the *Dof* gene family in the chromosomal role in Cashew plants. The results showed that 40 *Dof* genes originated from segmental replication, while the remaining 27 genes did not reveal any duplication. This finding indicates that approximately 60% of the total *AoDof* genes are segmentally duplicated, while the remaining 40% showed no duplication.

3.8. Assessment of Ka/Ks ratio and natural selection

MEGA-X was used to figure out the number of nonsynonymous substitutions per nonsynonymous site (Ka) and the number of synonymous substitutions per synonymous site (Ks) using pairwise alignment. The Ka/Ks ratio, which represents the ratio of nonsynonymous to synonymous mutations, was then manually determined. The ratio ranged from 0.11 in AoDof41/AoDof47 pair to up to 2.07 in AoDof33/AoDof34 pair. The Ka/Ks ratio denotes the likelihood of natural selection on the evolutionary path. If Ka/Ks is less than one, the selection is purifying; if it is equal to one, it is neutral; if it is greater than one, it is positive (Yangand Bielawski, 2000). Ka/Ks ratio of only the *AoDof33/AoDof34* pair was found greater than 1, symbolizing positive selection which is being favoured throughout evolution, while the ratio of all other pairs was less than 1, symbolizing purifying selection (Fig. 7).

Table1: Information about 67 putative *Dof* genes discovered from the genome of *A. occidentale*

Dof gene	**Accession No Phytozome**	**Scaffold no**	**Orientation**	**Location**	**Intron No**	**mRNA (bp)**	**AA length**	**pI**	**Mol. Weight**
AoDof1	0004s1080	4	forward	13259294..13260887	1	744	257	9.25	27857.65
AoDof2	0004s1080	4	forward	13259456..13260742	1	744	257	9.25	27857.65
AoDof3	0004s1080	4	forward	13259294..13260887	1	744	257	9.25	27857.65
AoDof4	0822s0002	822	reverse	57629..58372	0	717	247	8.13	26029.43
AoDof5	0007s0586	7	forward	10429846..10431374	0	744	247	8.13	26029.43
AoDof6	0633s0003	633	forward	56558..58087	1	804	174	9.80	19218.26
AoDof7	0017s0333	17	forward	2127402..2129044	0	906	301	8.43	31804.58
AoDof8	0001s1439	1	reverse	17787383..17788697	1	804	308	9.04	33344.86
AoDof9	0017s0586	17	forward	4581205..4582218	0	1014	337	8.97	36508.59
AoDof10	0001s0889	1	reverse	6091272..6092306	0	765	254	8.96	26226.32
AoDof11	0012s0770	12	reverse	13797390..13798160	0	771	256	5.94	27384.37
AoDof12	0015s0661	15	reverse	12515928..12516917	0	990	329	6.59	36082.45
AoDof13	0981s0001	981	reverse	17..3389	0	966	321	6.56	35162.27
AoDof14	0004s0764	4	reverse	11099111..11100076	0	966	321	6.56	35162.27
AoDof15	0017s0346	17	forward	2182635..2184316	0	1041	346	8.80	36956.90
AoDof16	0017s0346	17	forward	2182735..2184316	0	1041	346	8.80	36956.90
AoDof17	0001s0883	1	reverse	6018444..6020562	0	1041	340	8.92	36489.49
AoDof18	0001s0883	1	reverse	6018444..6020562	0	1041	340	8.92	36489.49
AoDof19	0005s1173	5	reverse	13187097..13188040	1	519	172	9.85	18730.18
AoDof20	0006s1698	6	reverse	17468461..17470369	0	804	267	9.01	29362.67
AoDof21	0011s0816	11	reverse	4913647..4914617	1	798	312	9.57	32929.76
AoDof22	0002s1863	2	reverse	18865824..18868465	0	819	272	9.07	29982.13
AoDof23	0016s0426	16	forward	2846945..2850102	1	1038	345	9.13	37507.68
AoDof24	0016s0911	16	forward	5943672..5945428	1	1023	340	9.30	36118.36
AoDof25	0020s0161	20	forward	4760397..4762048	1	1020	339	9.34	35851.66
AoDof26	0139s0010	139	reverse	51182..53100	0	831	295	4.58	33168.33
AoDof27	0010s0958	10	reverse	11917448..11920845	3	900	262	9.87	29105.49
AoDof28	0010s0958	1	reverse	2266641..2268350	1	1035	344	9.14	36346.38
AoDof29	0007s1214	7	reverse	16141145..16143714	1	1044	347	8.67	38011.45
AoDof30	0001s1919	1	reverse	23347720..23349293	1	735	228	9.53	25339.27
AoDof31	0001s1919	1	reverse	23347624..23349337	1	735	244	9.51	27062.25
AoDof32	0004s1845	4	reverse	18266846..18269296	1	945	314	6.44	34395.53
AoDof33	0015s0212	15	forward	2016264..2018279	1	927	285	9.35	30705.35
AoDof34	0015s0212	15	forward	2016264..2018279	1	927	291	9.25	31401.02
AoDof35	1110s0001	1110	forward	20616..22629	1	927	291	9.25	31401.02
AoDof36	0009s1071	9	reverse	6800192..6802066	0	831	352	5.19	39546.72
AoDof37	0005s1259	5	forward	13749792..13753147	1	954	364	6.34	39895.93
AoDof38	0016s0477	16	reverse	3174391..3177047	2	1083	295	9.06	32875.53
AoDof39	0010s1025	10	forward	12373041..12375233	1	978	310	7.56	34276.82
AoDof40	0010s1025	10	forward	12373041..12375233	1	978	319	8.61	35621.52
AoDof41	1145s0004	1145	forward	17661..19353	1	498	167	8.51	18812.38
AoDof42	0007s1149	7	reverse	15773735..15775657	0	846	281	9.11	31001.21
AoDof43	0016s0427	16	reverse	2858874..2859791	0	918	305	9.20	32979.21
AoDof44	0012s0248	12	forward	9436962..9437863	0	870	289	4.94	32345.68
AoDof45	0007s1213	7	forward	16132408..16135612	0	912	303	9.37	32769.82
AoDof46	0001s1646	1	forward	21073784..21074620	1	783	278	5.90	31275.09
AoDof47	0008s0817	8	reverse	5120751..5123209	2	1170	466	5.98	50618.01
AoDof48	0002s0256	2	reverse	3164207..3167218	1	1440	479	6.12	53400.45
AoDof49	0010s1617	10	forward	16315448..16318320	1	1428	399	8.80	43894.15
AoDof50	0004s2264	4	forward	20553292..20555137	1	876	290	8.43	31538.07
AoDof51	1203s0004	1203	forward	26707..29609	1	876	305	8.81	33376.17
AoDof52	0004s2264	4	forward	20553335..20555545	1	876	305	8.81	33376.17
AoDof53	0004s2264	4	forward	20553335..20554912	1	870	291	8.43	31724.28
AoDof54	0010s1617	10	forward	16314697..16318320	2	1623	496	6.48	54267.22
AoDof55	0018s1004	18	reverse	11958257..11961938	2	1596	438	5.77	48135.22
AoDof56	0018s1004	18	reverse	11958953..11961938	2	1530	438	5.77	48135.22
AoDof57	0018s1004	18	reverse	11958916..11961938	2	1539	453	5.65	49519.56
AoDof58	0016s0351	16	forward	2436374..2436841	0	468	155	9.28	17375.74

AoDof59	0005s1533	5	reverse	15577250..15580679	0	1377	524	5.79	57455.79
AoDof60	0018s1004	18	reverse	11958953..11961938	2	1530	501	5.82	54928.75
AoDof61	0005s1533	5	reverse	15577249..15580735	0	1377	495	5.89	54068.86
AoDof62	0005s1533	5	reverse	15577250..15580679	0	1377	495	5.89	54068.86
AoDof63	0015s0554	15	forward	11554872..11558516	1	1422	425	7.56	46476.15
AoDof64	0015s0554	15	forward	11554872..11558516	1	1422	473	6.37	51506.63
AoDof65	1176s0001	1176	forward	1611..5254	1	1422	473	6.37	51506.63
AoDof66	0004s0623	4	reverse	9331163..9335040	1	1422	502	8.32	55010.26
AoDof67	0009s0442	9	forward	2533219..2533985	1	702	250	6.89	26607.83

Table 2: Cashew *Dof* genes involved in biological process (BP) based on the known functions of orthologous Arabidopsis genes

Group	**Gene ID**	**Gene duplication group**	**GO Number (MF and BF)**	***Arabidopsis* ortholog genes**	**Putative Function of *Arabidopsis* orthologs**
E2	*AoDof1*	Segmental duplication	GO:0003677 GO:0006355	AT4G38000	Floral organ abscission
E2	*AoDof2*	No duplication	NF	AT1G21340.1	Regulate transcription of genes
E2	*AoDof3*	Segmental duplication	NF	AT1G21340.1	Regulate transcription of genes
E1	*AoDof4*	Segmental duplication	GO:0003677 GO:0006355	AT5G62940.1	Play an important role in xylem and phloem histogenesis
E1	*AoDof5*	Segmental duplication	GO:0003677 GO:0006355	AT5G62940.1	Responsive for xylem and phloem histogenesis
F2	*AoDof6*	No duplication	GO:0003677 GO:0006355	AT3G21270.1	Involved in regulation of transcription
D	*AoDof7*	Segmental duplication	GO:0003677 GO:0006355 GO:0015078 GO:0015986 GO:0045263	AT5G60850.1	Involved in transcriptional control
F2	*AoDof8*	No duplication	GO:0003677 GO:0006355	AT3G21270.1	Involved in transcriptional control
F2	*AoDof9*	No duplication	GO:0003677 GO:0006355	AT3G21270.1	Involved in transcriptional control
E2	*AoDof10*	Segmental duplication	GO:0003677 GO:0006355	AT3G21270.1	Involved in transcriptional control
D	*AoDof11*	Segmental duplication	GO:0003677 GO:0006355	AT3G21270.1	Involved in transcriptional control
F3	*AoDof12*	No duplication	GO:0003677 GO:0006355	AT3G21270.1	Involved in transcriptional control
F3	*AoDof13*	No duplication	GO:0003677 GO:0006355	AT3G21270.1	Involved in transcriptional control
F3	*AoDof14*	No duplication	GO:0003677 GO:0006355	AT3G21270.1	Involved in transcriptional control
F2	*AoDof15*	No duplication	NF	AT1G07640.3	Glucosinolate biosynthetic process regulation transcription, insect response, Jasmonic acid response, wounding response
F2	*AoDof16*	Segmental duplication	GO:0003677 GO:0006355	AT1G07640.3	Glucosinolate biosynthetic process regulation transcription, insect response, Jasmonic acid response, wounding response
F2	*AoDof17*	No duplication	GO:0003677 GO:0006355	AT1G07640.3	Glucosinolate biosynthetic process regulation transcription, insect response, Jasmonic acid response, wounding response
F2	*AoDof18*	No duplication	NF	AT1G07640.3	Glucosinolate biosynthetic process regulation

					transcription, insect response, Jasmonic acid response, wounding response
C1	*AoDof19*	Segmental duplication	GO:0003677 GO:0006355	AT3G21270.1	Involved in transcriptional control
C2	*AoDof20*	Segmental duplication	GO:0003677 GO:0006355	AT1G64620	Involved in transcriptional control
C1	*AoDof21*	Segmental duplication	GO:0003677 GO:0006355	AT1G21340.1	Involved in transcriptional control
C2	*AoDof22*	No duplication	GO:0003677 GO:0006355	AT1G64620	Involved in transcriptional control
C1	*AoDof23*	Segmental duplication	GO:0003677 GO:0006355	AT1G21340.1	Involved in transcriptional control
C1	*AoDof24*	Segmental duplication	GO:0003677 GO:0006355	AT3G21270.1	Involved in transcriptional control
C1	*AoDof25*	No duplication	GO:0003677 GO:0006355	AT1G64620	Involved in transcriptional control
F3	*AoDof26*	No duplication	GO:0003677 GO:0006355	AT1G21340.1	Involved in transcriptional control
C1	*AoDof27*	Segmental duplication	GO:0003677 GO:0006355	AT2G28510.1	Involved in transcriptional control
C1	*AoDof28*	Segmental duplication	GO:0003677 GO:0006355	AT5G62940.1	Phloem or xylem histogenesis
C1	*AoDof29*	Segmental duplication	GO:0003677 GO:0006355	AT2G28510.1	Involved in transcriptional control
C2	*AoDof30*	Segmental duplication	NF	AT1G64620	Involved in transcriptional control
C2	*AoDof31*	Segmental duplication	GO:0003677 GO:0006355	AT1G64620	Regulate transcription of genes
E2	*AoDof32*	Segmental duplication	GO:0003677 GO:0006355	AT5G60850.1	Regulation of transcription
E1	*AoDof33*	Segmental duplication	NF	AT1G64620	Regulate transcription of genes
E1	*AoDof34*	Segmental duplication	GO:0003677 GO:0006355	AT1G64620	Regulate transcription of genes
E1	*AoDof35*	No duplication	GO:0003677 GO:0006355	AT1G64620	Regulate transcription of genes
F3	*AoDof36*	No duplication	GO:0003677 GO:0006355	AT1G21340.1	Regulate transcription of genes
F3	*AoDof37*	No duplication	GO:0003677 GO:0006355	AT4G00940	Regulate transcription of genes
B2	*AoDof38*	Segmental duplication	GO:0003677 GO:0006355	AT2G28810.1	Regulate transcription of genes
F3	*AoDof39*	No duplication	NF	AT4G00940	Regulate transcription of genes
F3	*AoDof40*	No duplication	GO:0003677 GO:0006355	AT4G00940	Regulate transcription of genes
B1	*AoDof41*	Segmental duplication	GO:0003677 GO:0006355	AT1G29160.1	Regulate transcription of genes, seed coat development
B2	*AoDof42*	Segmental duplication	GO:0003677 GO:0006355	AT1G28310.2	Regulation of transcription
D	*AoDof43*	Segmental duplication	GO:0003677 GO:0006355	AT5G60850.1	Regulation of transcription
F3	*AoDof44*	No duplication	GO:0003677 GO:0006355	AT1G21340.1	Regulation of transcription,
D	*AoDof45*	Segmental duplication	GO:0003677 GO:0006355 GO:0043401	AT5G62940.1	phloem or xylem histogenesis
F3	*AoDof46*	No duplication	GO:0003677 GO:0006355	AT1G21340.1	Regulation of transcription,
B1	*AoDof47*	Segmental duplication	GO:0003677 GO:0006355	AT1G29160.1	Regulate transcription of genes, seed coat development
B1	*AoDof48*	Segmental duplication	GO:0003677 GO:0006355	AT3G21270.1	Regulation of transcription
B1	*AoDof49*	Segmental duplication	NF	AT1G29160.1	Regulate transcription of genes, seed coat development

E1	*AoDof50*	No duplication	NF	AT1G64620	Regulation of transcription
E1	*AoDof51*	No duplication	GO:0003677 GO:0006355	AT1G64620	Regulation of transcription
E1	*AoDof52*	No duplication	GO:0003677 GO:0006355	AT1G64620	Regulation of transcription
E1	*AoDof53*	Segmental duplication	NF	AT1G64620	Regulate transcription of genes
B1	*AoDof54*	Segmental duplication	GO:0003677 GO:0006355	AT1G29160.1	Regulate transcription of genes, seed coat development
B1	*AoDof55*	Segmental duplication	NF	AT1G29160.1	Regulate transcription of genes, seed coat development
B1	*AoDof56*	No duplication	NF	AT1G29160.1	Regulate transcription of genes, seed coat development
B1	*AoDof57*	No duplication	NF	AT1G29160.1	Regulate transcription of genes, seed coat development
B1	*AoDof58*	Segmental duplication	GO:0003677 GO:0006355 GO:0005506 GO:0016702 GO:0055114	AT1G64620	Regulate transcription of genes
B1	*AoDof59*	Segmental duplication	GO:0003677 GO:0006355	AT1G29160.1	Regulate transcription of genes, seed coat development
B1	*AoDof60*	Segmental duplication	GO:0003677 GO:0006355	AT1G29160.1	Regulate transcription of genes, seed coat development
B1	*AoDof61*	Segmental duplication	NF	AT1G29160.1	Regulate transcription of genes, seed coat development
B1	*AoDof62*	Segmental duplication	NF	AT1G29160.1	Regulate transcription of genes, seed coat development
B1	*AoDof63*	Segmental duplication	NF	AT1G29160.1	Regulate transcription of genes, seed coat development
B1	*AoDof64*	No duplication	GO:0003677 GO:0006355	AT1G29160.1	Regulate transcription of genes, seed coat development
B1	*AoDof65*	Segmental duplication	GO:0003677 GO:0006355	AT1G29160.1	Regulate transcription of genes, seed coat development
B1	*AoDof66*	Segmental duplication	GO:0003677 GO:0006355	AT1G29160.1	Regulate transcription of genes, seed coat development
B2	*AoDof67*	No duplication	GO:0003677 GO:0006355	AT1G64620	Regulate transcription of genes

Analysis of Cis-regulatory elements in the promoter region of cashew *Dof* genomes revealed the presence of various elements responsible for essential physiological processes. In total, 67 *AoDof* genes were analysed, and all of them contained cis-acting elements in the promoter and enhancer regions. The cis-regulatory elements related to abscisic acid responsiveness were present in 53 of the 67 (79%) *AoDof* genes, with the highest levels of presence in *AoDof17* and *AoDof18*. In addition, 65 (97%) *AoDof* genes had different types of light-responsive elements, and 64 (95%) *AoDof* genes were part of a conserved DNA module regulating light responsiveness, with maximum presence in *AoDof1, AoDof2, AoDof3, AoDof5, AoDof27,* and *AoDof54*.

The analysis also revealed that 48 (71%) *AoDof* genes were required for anaerobic induction, with *AoDof47* having the highest number. Among the 67 genes, 49 (73%) were involved in MeJA-responsiveness, with the highest presence in *AoDof63*, *AoDof64*, and *AoDof65*. Similarly, 54 (80%) *AoDof* genes contained ethylene-responsive elements, with maximum presence in *AoDof50*, *AoDof51*, *AoDof52*, and *AoDof53*, while 30 (45%) genes were related to drought-inducibility, with *AoDof48* having the most among others.

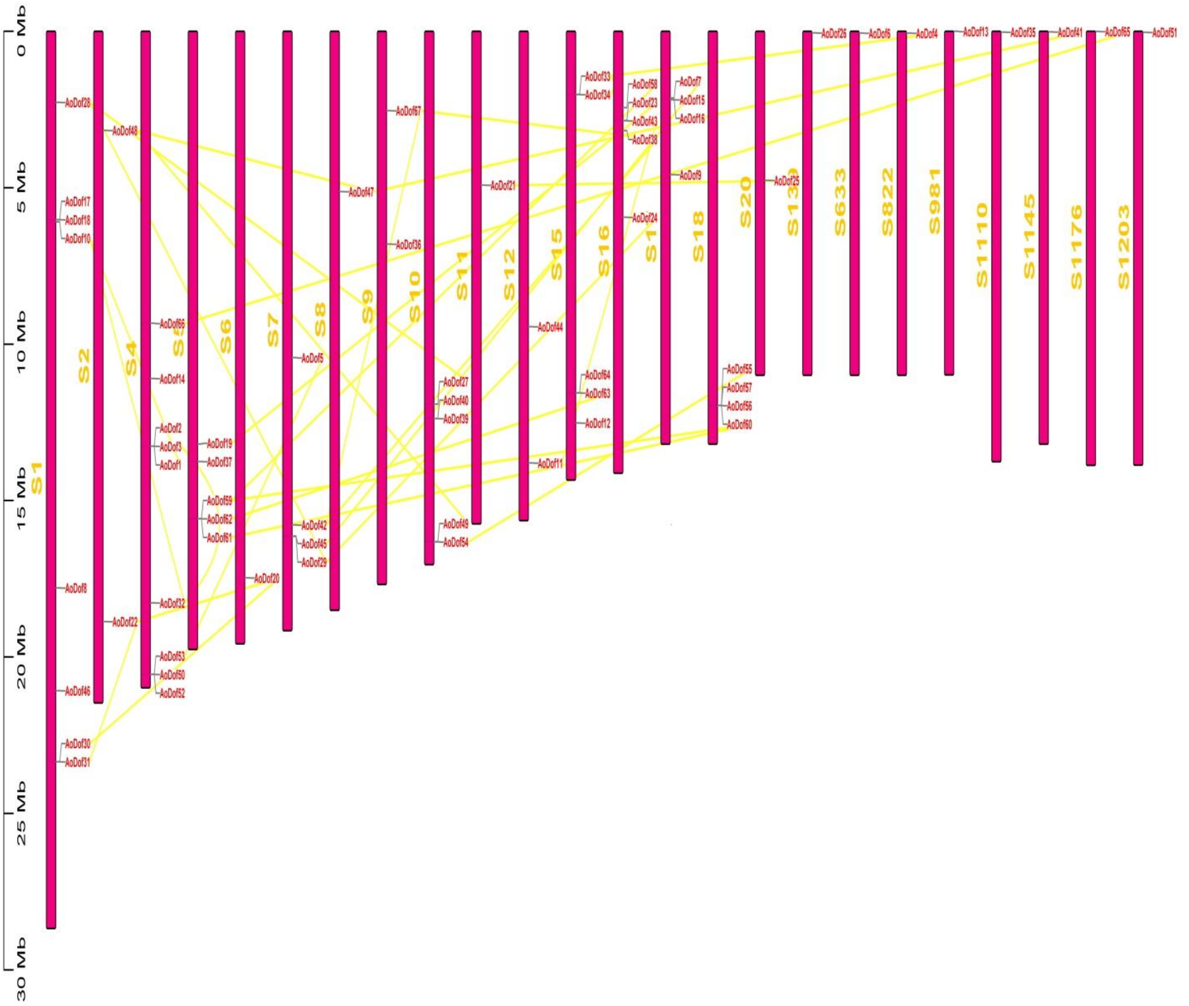


Figure. 6 Distribution of *Dof* genes on cashew scaffolds. The scale represents a bp (base pair) chromosomal distance. Chromosome map was generated using TBtools (Chen et al., 2020a). The yellow lines link the segmented duplicate gene. *Dof* genes in cashew showing the dominance of segmental duplication.

The analysis also revealed that 58 (87%) *AoDof* genes had elements co-related with transcriptional activation under stress, with *AoDof42* showing maximum presence, whereas 48 (72%) genes had elements involved in stress resistance and growth, with maximum abundance in *AoDof17* and *AoDof18*. The AAGAA motif was found in 39 (58%) *AoDof* genes, the AT~TATA box was present in 55 genes (82%), and the TATA box was found in 35 (52%) *AoDof* genes. In addition, 12 (18%) *AoDof* genes had elements involved in endosperm expression in small amounts, 3 (4%) had elements involved in palisade mesophyll cell differentiation in small amounts, and 21 (31%) had elements associated with low-temperature responsiveness. Furthermore, 15 (22%) *AoDof* genes were slightly involved in circadian control, and 16 (24%) contained gibberellin-responsive cis-regulatory elements.

3.9. Gene Ontology (GO) annotation and subcellular localization of cashew *Dof* genes:

To classify the 67 cashew *Dof* genes, the Phytozome database was used to identify their biological processes, molecular functions, and cellular components using Gene Ontology (GO). The results of the GO functional classification are presented in (Table 2). The GO enrichment study revealed that two GO terms (GO:0003677, GO:0006355) were significantly enriched for almost all *AoDof* genes. Additionally, one GO term (GO:0043401) was found only in *AoDof45* of subgroup D. In subgroup D of *AoDof7*, three GO terms (GO:0015078, GO:0015986, and GO:0045263) were discovered. Similarly, in subgroup B1, *AoDof58* was discovered to be a member of three GO terms (GO:0005506, GO:0016702, and GO:0055114).

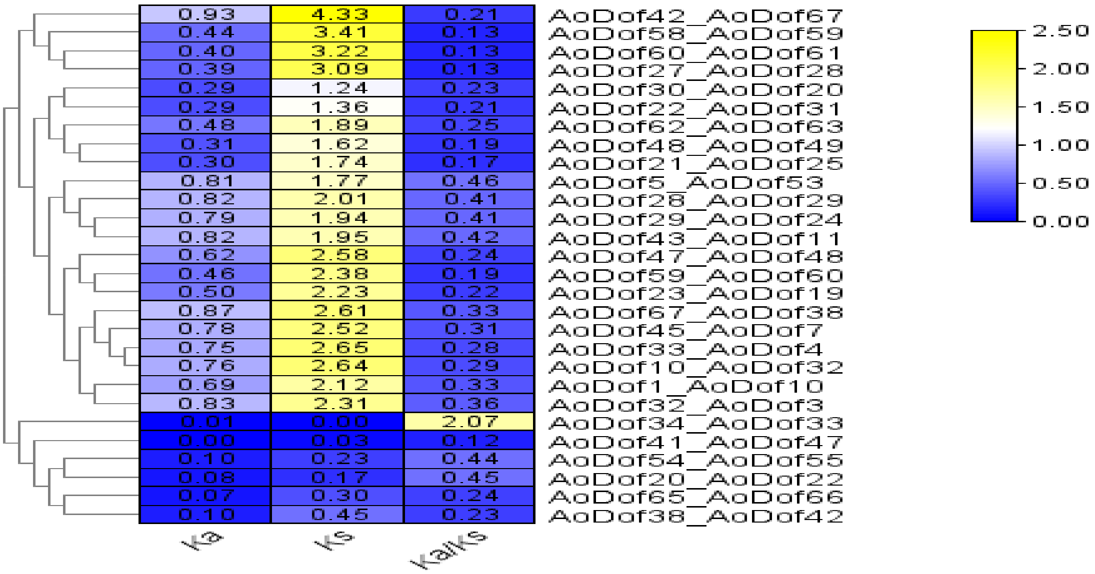
3.10. Subcellular localization:

The subcellular localization of the cashew *Dof* genes was also investigated. Most of the genes expressed themselves in the nucleus. However, some *AoDof* genes were found to be expressed in different locations. *AoDof7, AoDof36*, and *AoDof10* were mostly expressed in the chloroplast. *AoDof33* was found in the cytonuclear, and *AoDof64* and *AoDof65* were expressed in the cytoplasm. It is important to note that *AoDof7* and *AoDof10* seem to be truncated, as they do not possess the Nuclear localization signal (NLS), possibly due to a point mutation that led to the loss of this signal, although they have the conserved *Dof* domain and motif (Fig 9).

4. Discussion

From available cashew nucleotide sequences (https://phytozome-next.jgi.doe.gov/info/Aoccidentale v0 9), this analysis identified 67 *AoDof* genes, which were characterized with the existence of a strongly conserved *Dof* domain. In cashew, the *Dof* gene family was found to be scattered over 9 clades. The number of *Dof* genes in cashew are less than in wheat (96 *TaDof*) (Liu et al., 2020) and more than in cucumber (36 *CsDof*) (Wen et al., 2016), *Arabidopsis (36AtDof)*, tomato *(SlDof34) (Cai et al., 2013)* and rice (*OsDof30*) (Lijavetzky et al., 2003b). Despite this, the cashew genome is 488 Mb (Megabase pair), almost 4.2 times more than Arabidopsis (115 Mb) and 1.2 times more than rice (420 Mb). To date, 377.95 Mb (77.4%) of the 488 Mb cashew tree genome has been compiled (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org Aoccidentale er), perhaps all AoDof genes have probably not been detected owing to the limitations of the presently available database. Gene family analysis has become popular method for studying gene structure, complexity, role, development or evolution. Comparative analysis of the *Dof* family members between two species is used in this study to analyse the different roles of the members of the cashew *Dof* family and to aid further analysis of gene functioning. The arrangement of exons and introns may also be used to deduce evolutionary interactions between genes or species (Bondarenko and Gelfand, 2016; Koralewski and Krutovsky, 2011). In cashew, some *Dof* genes were intron-less while others have up to three introns. The majority of intron-containing and intron-less genes expressed similar trends and were hence, grouped into the same respective clades (Table 1 and Fig.2). Analogous cases have also been found to occur in rice and arabidopsis, implying that evolutionary conservation has occurred (Lijavetzky et al., 2003c). Only *AoDof*27 showed a considerable variance in introns and exons numbers between the members of same group in the respective groups. The maximum number of *Dof* genes were found on scaffold no 4 which totalled 9. While on the other extreme, scaffolds 6, 8, 11, 20, 13, 633, 822, 981, 1110, 1145, 1176, and 1203, each had only one *Dof* gene per scaffold in them. All these 67 *AoDof* genes were scattered across 24 scaffolds in entire cashew genome.

The conserved motifs in the cashew *Dof* family were looked at by using the MEME software. Most of the genes of cashew *Dof* of the same group or similar subgroup had similar motifs, implying that these motifs are conserved and are essential for group or subgroup functions. In all 67 *AoDof* genes, Motif 1 was shown to be present. However, there was a lot of variation in the arrangements of the individual groups or subgroups. Subgroup B1, for example, contains motifs 3, 7, 2,5,4 and 8, indicating the complexity of *Dof* protein activity in cashew. Some motifs were also found to be conserved in certain groups; for example, motif 1 is found in all *AoDof* genes and Motif 6 is found in subgroup B2. The distribution of motifs revealed that genes with similar motifs are most likely the product of gene expansion which was thought to occur within the same groups or subgroups. To put it another way, ancestral genes with different motif frameworks emerge early in evolution and have remained the same during evolution. Examination of the locations of conserved motifs (Fig. 3) provides additional information about the cashew *Dof* family's ancestral relationships; study of conserved motifs also complements and confirms the findings of phylogenetic research (Fig. 3) Gene expression and cis-regulatory analysis also showed some interesting results. Just like the typical *Dof* genes of various other species, most of the *AoDof* genes are also expressed in the nucleus, while others were expressed in various locations, according to the sub-cellular localization of *AoDof* genes (Fig 9). The expression of *AoDof*10 and *AoDof*36 in chloroplasts suggests that they can play a part in photosynthesis regulation and development of chlorophyll and hence, chloroplast. Spatio-temporal transcriptomic expression of genes was also analysed by observing cis-regulatory elements of the *AoDof* genes. Cis-regulatory elements, thus associated with these *Dof* genes were responsible for certain vital physiological processes like light response; stress-related; seed, endosperm, meristem and hormone specific processes. The occurrence of cis-regulatory factors of light response was the most common and significant of all of them. This suggests that *Dof* genes often play some crucial part in the direct and indirect regulation of processes that are commonly (but not specifically) associated with light (Fig 8). Duplication is the fundamental driving force which evolves the *Dof* genes over time. Segmental replication occurs when two or more genes are duplicated on separate chromosomes, while tandem duplication occurs when two or more genes are duplicated on the same chromosome (Panchy et al., 2016). For example, the cucumber genome contains two tandemly duplicated gene pairs and six segmentally duplicated gene pairs (Wen et al., 2016). *PtrDof* genes were present in both segmental and tandem duplicated regions of up to 49 percent (20 out of 41) of poplar genes (Yang and Tuskan, 2006). In apple, a total of 57 *MdDof* genes were



Ka	Ks	Ka/Ks	
0.93	4.33	0.21	AoDof42_AoDof67
0.44	3.41	0.13	AoDof58_AoDof59
0.40	3.22	0.13	AoDof60_AoDof61
0.39	3.09	0.13	AoDof27_AoDof28
0.29	1.24	0.23	AoDof30_AoDof20
0.29	1.36	0.21	AoDof22_AoDof31
0.48	1.89	0.25	AoDof62_AoDof63
0.31	1.62	0.19	AoDof48_AoDof49
0.30	1.74	0.17	AoDof21_AoDof25
0.81	1.77	0.46	AoDof5_AoDof53
0.82	2.01	0.41	AoDof28_AoDof29
0.79	1.94	0.41	AoDof29_AoDof24
0.82	1.95	0.42	AoDof43_AoDof11
0.62	2.58	0.24	AoDof47_AoDof48
0.46	2.38	0.19	AoDof59_AoDof60
0.50	2.23	0.22	AoDof23_AoDof19
0.87	2.61	0.33	AoDof67_AoDof38
0.78	2.52	0.31	AoDof45_AoDof7
0.75	2.65	0.28	AoDof33_AoDof4
0.76	2.64	0.29	AoDof10_AoDof32
0.69	2.12	0.33	AoDof1_AoDof10
0.83	2.31	0.36	AoDof32_AoDof3
0.01	0.00	2.07	AoDof34_AoDof33
0.00	0.03	0.12	AoDof41_AoDof47
0.10	0.23	0.44	AoDof54_AoDof55
0.08	0.17	0.45	AoDof20_AoDof22
0.07	0.30	0.24	AoDof65_AoDof66
0.10	0.45	0.23	AoDof38_AoDof42

Figure. 7 Ks and Ka values were calculated. Analyses were conducted using the ka_ks calculator present in TBtools (Chen et al., 2020a).

located in tandem duplicated regions, while 13 *MdDof* genes were duplicated both segmentally and tandemly (Chen et al., 2020c; Kang and Wang, 2013; Yang et al., 2018; Zhang et al., 2018). In this research, we discovered that more than half of the *AoDof* genes (40 out of 67) in the cashew genome had segmental duplications, but unlike the *Dof* family of majority of the other plants, cashew genome lacked tandem duplication, implying that the only source of *AoDof* gene expansion in cashew is segmental duplication. Other plants, such as cotton, have shown similar extraordinary effects (Li et al., 2018). While *AoDof* genes only showed segmental duplication, which was dispersed randomly in the whole cashew genome. Out of the total 67 *AoDof* genes, 40 *Dof* genes were the probable result of segmental duplication in the cashew genome, while the remaining 27 *Dof* genes showed no signs of duplication (Fig. 6), making the percentage of segmental and no-duplication out of total *AoDof* genes as 60% and 40%. Furthermore, the number of segmental duplications in each clade was also figured out. Clades A and F1 did not contain any *AoDof* gene and 14 out of 17 *AoDof* genes of clade B1 showed segmental duplication event. 2 out of 3 members of clade B2 showed segmental duplication. Furthermore, 7 out of 8 in clade C1; 3 out of 4 in clade C2; all 4 in clade D; 5 out of 9 in clade E1; 4 out of 5 in clade E2; 1 out of 7 in clade F2 and none of the clade F3 members showed segmental duplication. This implies that whatever the expansion, divergence and evolution have been in *AoDof*, is a result of segmental duplication only. There is no role of tandem duplication in this regard, as it was not found to occur in case of *AoDof.*

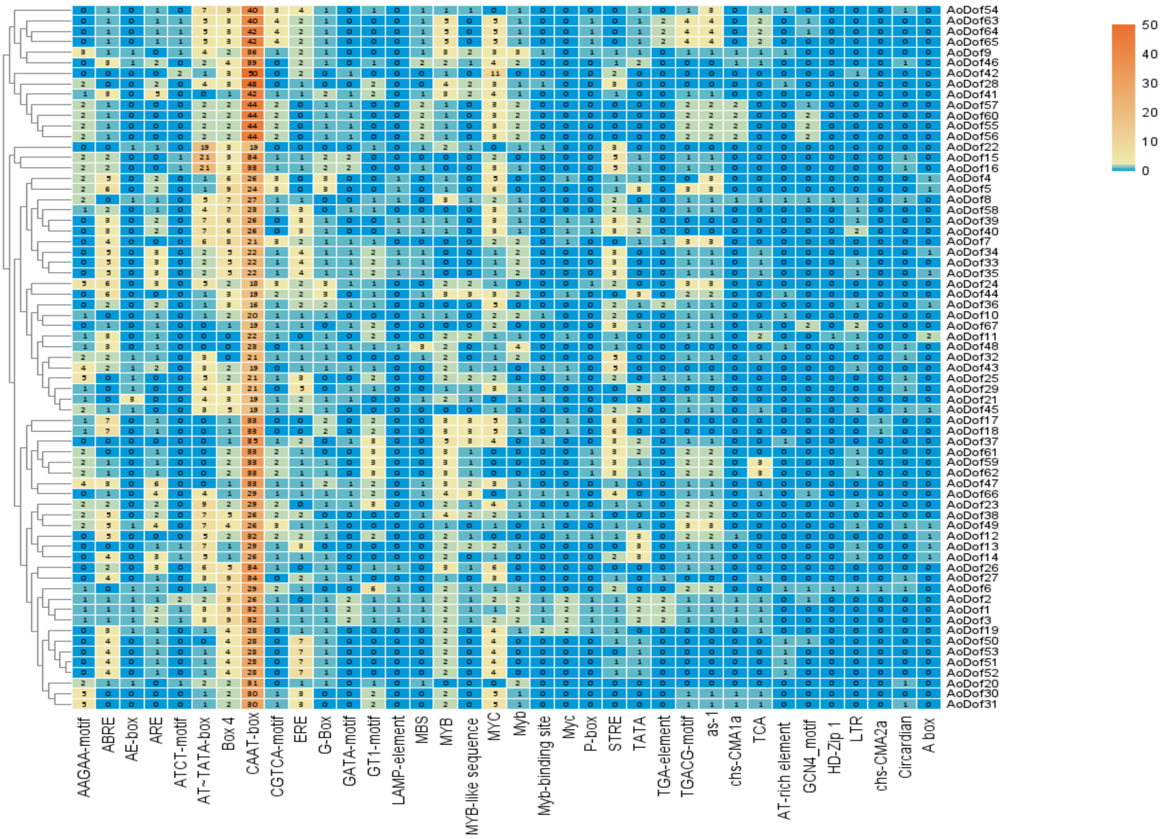
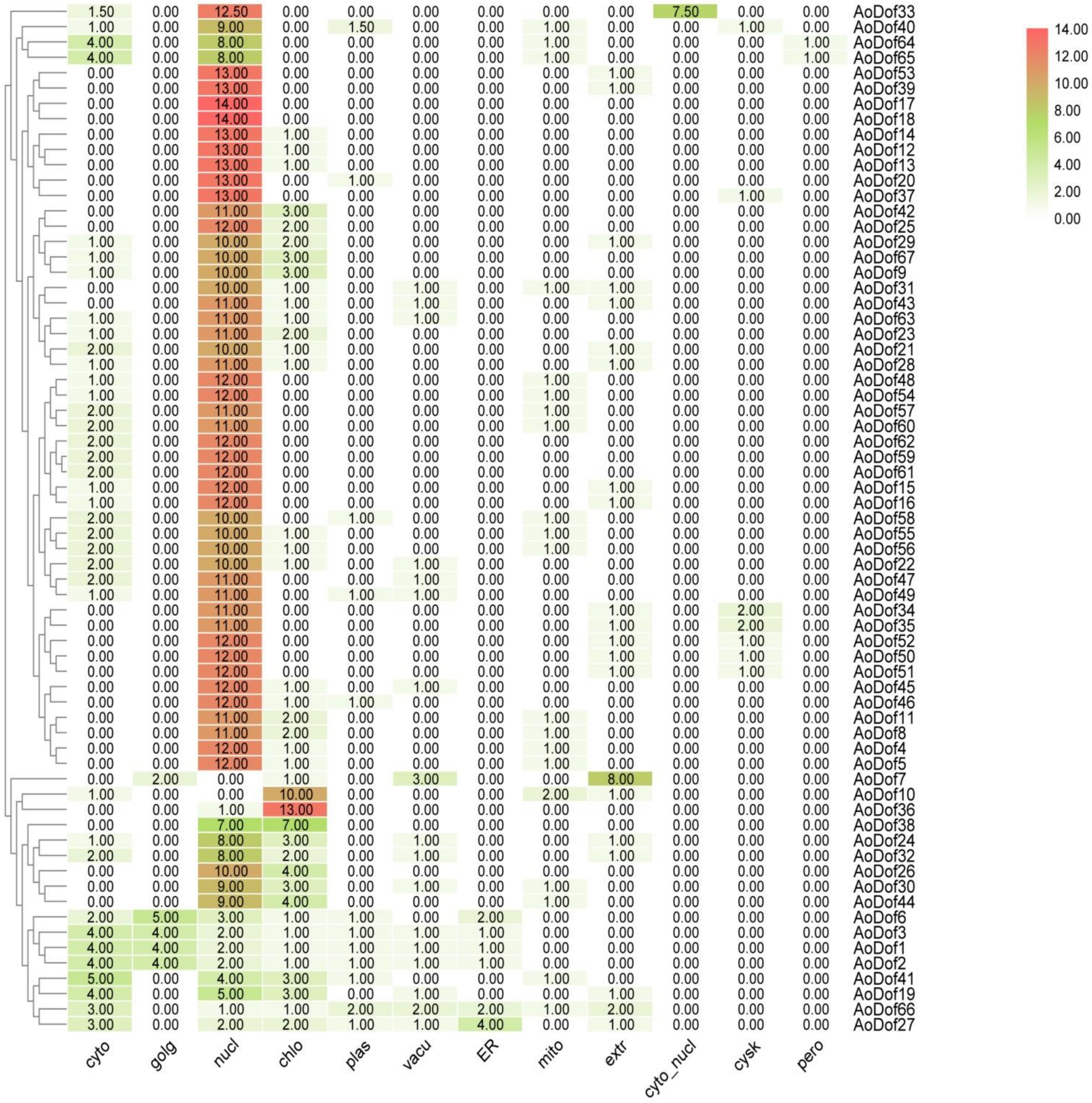


Figure. 8 Cis-regulatory elements (CREs) analysis in putative *AoDof* gene interlinked with phylogenetic tree.

The number of nonsynonymous substitutions per nonsynonymous site is denoted by Ka, whereas the number of synonymous substitutions per synonymous site is denoted by Ks, and the ratio of nonsynonymous to synonymous substitutions is expressed by Ka/KS. The Ka/Ks ratio gives a degree of selective pressure. Purifying selection on a gene pair reveals that it must have been purged by natural selection, most likely due to deleterious effects. Positively selected gene pairs, on the other hand, suggest that they may have been beneficial during the evolution of the two duplicates.



cyto	golg	nucl	chlo	plas	vacu	ER	mito	extr	cyto_nucl	cysk	pero	
1.50	0.00	12.50	0.00	0.00	0.00	0.00	0.00	0.00	7.50	0.00	0.00	AoDof33
1.00	0.00	9.00	0.00	1.50	0.00	0.00	1.00	0.00	0.00	1.00	0.00	AoDof40
4.00	0.00	8.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	AoDof64
4.00	0.00	8.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	AoDof65
0.00	0.00	13.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	AoDof53
0.00	0.00	13.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	AoDof39
0.00	0.00	14.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof17
0.00	0.00	14.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof18
0.00	0.00	13.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof14
0.00	0.00	13.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof12
0.00	0.00	13.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof13
0.00	0.00	13.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof20
0.00	0.00	13.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	AoDof37
0.00	0.00	11.00	3.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof42
0.00	0.00	12.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof25
1.00	0.00	10.00	2.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	AoDof29
1.00	0.00	10.00	3.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof67
1.00	0.00	10.00	3.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof9
0.00	0.00	10.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	0.00	0.00	AoDof31
0.00	0.00	11.00	1.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	AoDof43
1.00	0.00	11.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof63
1.00	0.00	11.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof23
2.00	0.00	10.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	AoDof21
1.00	0.00	11.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	AoDof28
1.00	0.00	12.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof48
1.00	0.00	12.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof54
2.00	0.00	11.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof57
2.00	0.00	11.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof60
2.00	0.00	12.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof62
2.00	0.00	12.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof59
2.00	0.00	12.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof61
1.00	0.00	12.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	AoDof15
1.00	0.00	12.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	AoDof16
2.00	0.00	10.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof58
2.00	0.00	10.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof55
2.00	0.00	10.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof56
2.00	0.00	10.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof22
2.00	0.00	11.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof47
1.00	0.00	11.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof49
0.00	0.00	11.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	2.00	0.00	AoDof34
0.00	0.00	11.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	2.00	0.00	AoDof35
0.00	0.00	12.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	AoDof52
0.00	0.00	12.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	AoDof50
0.00	0.00	12.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	AoDof51
0.00	0.00	12.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof45
0.00	0.00	12.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof46
0.00	0.00	11.00	2.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof11
0.00	0.00	11.00	2.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof8
0.00	0.00	12.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof4
0.00	0.00	12.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof5
0.00	2.00	0.00	1.00	0.00	3.00	0.00	0.00	8.00	0.00	0.00	0.00	AoDof7
1.00	0.00	0.00	10.00	0.00	0.00	0.00	2.00	1.00	0.00	0.00	0.00	AoDof10
0.00	0.00	1.00	13.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof36
0.00	0.00	7.00	7.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof38
1.00	0.00	8.00	3.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	AoDof24
2.00	0.00	8.00	2.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	AoDof32
0.00	0.00	10.00	4.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	AoDof26
0.00	0.00	9.00	3.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof30
0.00	0.00	9.00	4.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof44
2.00	5.00	3.00	1.00	1.00	0.00	2.00	0.00	0.00	0.00	0.00	0.00	AoDof6
4.00	4.00	2.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	AoDof3
4.00	4.00	2.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	AoDof1
4.00	4.00	2.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	AoDof2
5.00	0.00	4.00	3.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	AoDof41
4.00	0.00	5.00	3.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	AoDof19
3.00	0.00	1.00	1.00	2.00	2.00	2.00	1.00	2.00	0.00	0.00	0.00	AoDof66
3.00	0.00	2.00	2.00	1.00	1.00	4.00	0.00	1.00	0.00	0.00	0.00	AoDof27

Figure. 9: Sub cellular localization in putative *AoDof* genes interlinked with phylogenetics showed the presence of nuclear localization.

If Ka/Ks is less than one, the selection is purifying if it is equal to one, it is neutral; if it is greater than one, it is positive (Yang and Bielawski, 2000). The ratios of Ka to Ks in cashew *Dof* genes were hence, calculated. The ratio of *AoDof*41/*AoDof*47 pair showed the minimum was equivalent to 0.11, the least among all. While the *AoDof*33/*AoDof*34 pair had a ratio of 2.07, which was the highest in *AoDof* genes. Only the *AoDof*33/*AoDof*34 pair showed positive selection, due to having Ka/Ks ratio more than 1. This indicates that this is the only gene pair whose duplication has been favoured throughout evolution. All the remaining gene pairs had the Ka/Ks ratio lesser than 1, implying purifying selection (Fig 7).

A merged phylogenetic tree of aligned *AoDof* sequences, together with the reference genes of *AtDof* and *LsDof* was constructed by neighbour-joining (NJ) method, to figure out the evolutionary relationship between *AoDof* genes. The tree showed 11 clades (A, B1, B2, C1, C2, D, E1, E2, F1, F2 and F3) of *Dof* genes. And out of these clades, 9 contained *AoDof* genes while other two (A and F1) clades had no *Dof* genes of cashew. *AoDof* proteins were majorly detected in subgroup B1 (having 17 *AoDof* proteins) while subgroup A and F1 contained none of them. Proteins of same clade are structurally comparable and exhibit similar functions and as well (Fig. 4, 5). Resultantly, it can be inferred that all *Dof* proteins of same clades have similar structures and functions. Genes which are usually duplicated inside the same genome are called paralogs, but genes which are duplicated in different genomes are called orthologs, likely because of the taxonomic lineage separation (Thornton and DeSalle, 2000). Orthologs perform the similar task, but paralogs go through a variety of dissimilar functions (Tatusov et al., 1997). So identifying a gene family's paralogs and orthologs is critically important to comprehend its functional diversity and its dissimilarity in distribution throughout the various organisms. In total, two paralogous pairs of *AoDof* genes were found while five orthologs of *AoDof* were identified in Lettuce and Arabidopsis. Clade E1 contained the first paralogous among *AoDof4* and *AoDof5* and clade D had the second paralogous pair among *AoDof7* and *AoDof43* while no other clade contained any other paralogous pair. In addition, five pairs of orthologs of *AoDof* genes were also discovered. Clade B2, C1 and E2 were the clades having one pair of orthologs which were *AoDof67/LsDof30, AoDof29/LsDof14* and *AoDof3/AT1G51700* respectively. Only clade D had two orthologous pairs, which were *AoDof11/LsDof37* and *AoDof45/AT1G07640*. While all the remaining clades were left with no orthologue. A paralogous pair of genes refers to the growth of a gene family following the split between dicots and monocots. While the presence of orthologous pairs in monocots and dicots shows that some ancestral *Dof* genes existed in a common ancestor before the split of monocots and dicots, which occurred around 170-235 Mya (Blanc and Wolfe, 2004). According to the detailed relative evolutionary study of Cashew alongwith Lettuce and Arabidopsis, it was revealed that *AoDof* proteins in such clades are more intently connected to Lettuce as compared to Arabidopsis, as in a taxonomic tree Cashew is more connected to Lettuce than Arabidopsis. This also demonstrates the development of *Dof* proteins in parallel with the expansion of the Plantae Kingdom as a whole.

5. REFERENCES

Ayoub. N., Saddique. M. A. B., Ali. Z., Iqbal. S., Sher. M. A., Rehman. S. U., Farid. B. 2021. Morphological characterization of the rice genotypes at early growth stages and determinationof better parent heterosis for salinity tolerance traits under NaCl stress. Agric. Sci. J. 3(1): 96-106

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. Nucleic Acids Res. 43, W39-W49.

Blanc, G., and Wolfe, K. H. (2004). Widespread Paleopolyploidy in Model Plant Species Inferred from Age Distributions of Duplicate Genes. The Plant Cell. 16, 1667-1678.

Bondarenko, V. S., and Gelfand, M. S. (2016). Evolution of the Exon-Intron Structure in Ciliate Genomes. *PLoS One.* 11, e0161476.

Bulow, L., and Hehl, R. (2016). Bioinformatic Identification of Conserved Cis-Sequences in Coregulated Genes. *Methods Mol Biol* 1482, 233-45.

Cai, X., Zhang, Y., Zhang, C., Zhang, T., Hu, T., Ye, J., Zhang, J., Wang, T., Li, H., and Ye, Z. (2013). Genome-wide analysis of plant-specific Dof transcription factor family in tomato. J. Integr. Plant Biol. 55, 552-566.

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y. and Xia, R., 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol. plant. 13(8), pp.1194-1202.

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., and Xia, R. (2020b). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant. 13, 1194-1202.

Chen, H., Ahmad, M., Rim, Y., Lucas, W. J., and Kim, J. Y. (2013). Evolutionary and molecular analysis of D of transcription factors identified a conserved motif for intercellular protein trafficking. New Phytol. 198, 1250-1260.

Chen, P., Yan, M., Li, L., He, J., Zhou, S., Li, Z., Niu, C., Bao, C., Zhi, F., Ma, F., and Guan, Q. (2020c). The apple DNA-binding one zinc-finger protein MdDof54 promotes drought resistance. Hortic Res. 7, 195.

Cokol, M., Nair, R., and Rost, B. (2000). Finding nuclear localization signals. *EMBO* reports. 1, 411-415.

Corrales, A. R., Nebauer, S. G., Carrillo, L., Fernandez-Nohales, P., Marques, J., Renau-Morata, B., Granell, A., Pollmann, S., Vicente-Carbajosa, J., Molina, R. V., and Medina, J. (2014). Characterization of tomato Cycling Dof Factors reveals conserved and new functions in the control of flowering time and abiotic stress responses. J Exp Bot .65, 995-1012.

Danella Figo, D., De Amicis, K., Neiva Santos de Aquino, D., Pomiecinski, F., Gadermaier, G., Briza, P., Eduardo Santos Galvao, C., Bussador do Amaral, J., de Oliveira Martins, C., Fernandes Morato Castro, F., Kalil, J., and Souza Santos, K. (2019). Cashew Tree Pollen: An Unknown Source of IgE-Reactive Molecules. Int J Mol Sci. 20.

Diaz, I., Martinez, M., Isabel-LaMoneda, I., Rubio-Somoza, I., and & Carbonero, P. (2005). The DOF protein, SAD, interacts with GAMYB in plant nuclei and activates transcription of endosperm-specific genes during barley seed development. The Plant Journal. 42, 652-662.

Dominguez-Figueroa, J., Carrillo, L., Renau-Morata, B., Yang, L., Molina, R. V., Marino, D., Canales, J., Weih, M., Vicente-Carbajosa, J., Nebauer, S. G., and Medina, J. (2020). The Arabidopsis Transcription Factor CDF3 Is Involved in Nitrogen Responses and Improves Nitrogen Use Efficiency in Tomato. Front Plant Sci. 11, 601558.

Dong, G., Ni, Z., Yao, Y., Nie, X., and Sun, Q. (2007). Wheat Dof transcription factor WPBF interacts with TaQM and activates transcription of an alpha-gliadin gene during wheat seed development. Plant Mol. Biol. 63, 73-84.

Fernandes, L., and Mesquita, A. M. (1995). Anacardium occidentale (cashew) pollen allergy in patients with allergic bronchial asthma. J Allergy Clin Immunol. 95, 501-4.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., and Mistry, J. (2014). Pfam: the protein families database. Nucleic acids res. 42, D222-D230.

Fornara, F., Panigrahi, K. C., Gissot, L., Sauerbrunn, N., Rühl, M., Jarillo, J. A., and Coupland, G. (2009). Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. Dev cell. 17, 75-86.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In "The proteomics protocols handbook", pp. 571-607. Springer.

Gupta, S., Arya, G. C., Malviya, N., Bisht, N. C., and Yadav, D. (2016). Molecular cloning and expression profiling of multiple Dof genes of Sorghum bicolor (L) Moench. Mol Biol Rep. 43, 767-74.

Gupta, S., Malviya, N., Kushwaha, H., Nasim, J., Bisht, N. C., Singh, V., and Yadav, D. (2015). Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. Planta. 241, 549-562.

Hasnain, M. S., Rishishwar, P., Rishishwar, S., Ali, S., and Nayak, A. K. (2018). Extraction and characterization of cashew tree (Anacardium occidentale) gum; use in aceclofenac dental pastes. Int J Biol Macromol. 116, 1074-1081.

Hernando-Amado, S., Gonzalez-Calle, V., Carbonero, P., and Barrero-Sicilia, C. (2012). The family of DOF transcription factors in Brachypodium distachyon: phylogenetic comparison with rice and barley DOFs and expression profiling. BMC Plant Biol. 12, 202.

Horton, P., Park, K.-J., Obayashi, T., and Nakai, K. (2006). Protein subcellular localization prediction with WoLF PSORT. *In* "Proceedings of the 4th Asia-Pacific bioinformatics conference", pp. 39-48. World Scientific.

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. Bioinform.31, 1296-7.

Humera Amin, S. D., Muhammad Shafiq, Muhammad Arshad Javed, Muhammad Saleem Haider (2020). Genome-Wide Bioinformatics Analysis of DOF Transcription Factor Gene Family of Asparagus and Its Comparative Phylogenetic Assessment with Arabidopsis. Int. J. Biosci. 17, 103-123.

Jones, D. M., and Vandepoele, K. (2020). Identification and evolution of gene regulatory networks: insights from comparative studies in plants. Curr Opin Plant Biol. 54, 42-48.

Kang, H. B., and Wang, J. H. (2013). Adaptive control of 5 DOF upper-limb exoskeleton robot with improved safety. ISA Trans. 52, 844-52.

Koralewski, T. E., and Krutovsky, K. V. (2011). Evolution of exon-intron structure and alternative splicing. PloS one. 6, e18055.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35, 1547-1549.

Kumar, S., Tamura, K., and Nei, M. (1994). MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Comput Appl Biosci. 10, 189-91.

Kushwaha, H., Gupta, S., Singh, V. K., Bisht, N. C., Sarangi, B. K., and Yadav, D. (2013). Cloning, in silico characterization and prediction of three dimensional structure of *SbDof1, SbDof19, SbDof23 and SbDof24* proteins from Sorghum [*Sorghum bicolor* (L.) Moench]. Mol Biotechnol. 54, 1-12.

Kushwaha, H., Gupta, S., Singh, V. K., Rastogi, S., and Yadav, D. (2011). Genome wide identification of Dof transcription factor gene family in sorghum and its comparative phylogenetic analysis with rice and Arabidopsis. Mol Biol Rep. 38, 5037-53.

Li, H., Dou, L., Li, W., Wang, P., Zhao, Q., Xi, R., Pei, X., Liu, Y., and Ren, Z. (2018). Genome-Wide Identification and Expression Analysis of the Dof Transcription Factor Gene Family in *Gossypium hirsutum L.* Agronomy. 8, 186.

Lijavetzky, D., Carbonero, P., and Vicente-Carbajosa, J. (2003a). Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. BMC Evol. Biol. 3, 17.

Lijavetzky, D., Carbonero, P., and Vicente-Carbajosa, J. (2003b). Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. BMC Evol. Biol. 3, 1-11.

Lijavetzky, D., Carbonero, P., and Vicente-Carbajosa, J. (2003c). Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. BMC Evol Biol.3, 17.

Liu, Y., Liu, N., Deng, X., Liu, D., Li, M., Cui, D., Hu, Y., and Yan, Y. (2020). Genome-wide analysis of wheat DNA-binding with one finger (Dof) transcription factor genes: evolutionary characteristics and diverse abiotic stress responses. BMC Genom. 21, 1-18.

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., and Marchler-Bauer, A. (2020). CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 48, D265-D268.

Ma, J., Li, M.-Y., Wang, F., Tang, J., and Xiong, A.-S. (2015). Genome-wide analysis of Dof family transcription factors and their responses to abiotic stresses in Chinese cabbage. BMC genom. 16, 33.

Mah, E., Schulz, J. A., Kaden, V. N., Lawless, A. L., Rotor, J., Mantilla, L. B., and Liska, D. J. (2017). Cashew consumption reduces total and LDL cholesterol: a randomized, crossover, controlled-feeding trial. Am J Clin Nutr. 105, 1070-1078.

Malviya, N., Gupta, S., Singh, V., Yadav, M., Bisht, N., Sarangi, B., and Yadav, D. (2015). Genome wide in silico characterization of Dof gene families of pigeonpea (*Cajanus cajan (L)* Millsp.). Mol. Biol. Rep. 42, 535-552.

Mariyam, Shafiq, M., Haseeb, M., Atif, R. M., Naqvi, S. A. A. A. A., Ali, N., Javed, M. A., Gillani, F., and Haider, M. S. (2021). Genome-wide identification and characterization of a plant-specific Dof transcription factor gene family in olive (Olea europaea) and its comparison with

Arabidopsis. HEB. 62(6), pp.949-968.

Mello, B. (2018). Estimating timetrees with MEGA and the TimeTree resource. Mol. Biol. Evol. 35, 2334-2342.

Mena, M., Cejudo, F. J., Isabel-Lamoneda, I., and Carbonero, P. (2002). A role for the DOF transcription factor BPBF in the regulation of gibberellin-responsive genes in barley aleurone. Plant Physiol. 130, 111-9.

Moreno-Risueno, M. A., Martinez, M., Vicente-Carbajosa, J., and & Carbonero, P. (2007a). The family of DOF transcription factors: from green unicellular algae to vascular plants. Mol Genet Genomics. 277, 379-90.

Moreno-Risueno, M. Á., Martínez, M., Vicente-Carbajosa, J., and Carbonero, P. (2007b). The family of DOF transcription factors: from green unicellular algae to vascular plants. MGG 277, 379-390.

Nair, P. K., Melnick, S. J., Wnuk, S. F., Rapp, M., Escalon, E., and Ramachandran, C. (2009). Isolation and characterization of an anticancer catechol compound from *Semecarpus anacardium*. J Ethnopharmacol. 122, 450-6.

Negi, J., Moriwaki, K., Konishi, M., Yokoyama, R., Nakano, T., Kusumi, K., Hashimoto-Sugimoto, M., Schroeder, J. I., Nishitani, K., and Yanagisawa, S. (2013). A Dof transcription factor, SCAP1, is essential for the development of functional stomata in Arabidopsis. Curr Bio. 23, 479-484.

Neto, V. Q., Narain, N., Silva, J. B., and Bora, P. S. (2001). Functional properties of raw and heat processed cashew nut (*Anacardium occidentale*, L.) kernel protein isolates. Nahrung. 45, 258-62.

Nilsen, K. T., Walkowiak, S., Xiang, D., Gao, P., Quilichini, T. D., Willick, I. R., Byrns, B., N'Diaye, A., Ens, J., Wiebe, K., Ruan, Y., Cuthbert, R. D., Craze, M., Wallington, E. J., Simmonds, J., Uauy, C., Datla, R., and Pozniak, C. J. (2020). Copy number variation of TdDof controls solid-stemmed architecture in wheat. Proc Natl Acad Sci U S A. 117, 28708-28718.

Noguero, M., Atif, R. M., Ochatt, S., and Thompson, R. D. (2013). The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. Plant Sci. 209, 32-45.

Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016). Evolution of Gene Duplication in Plants. Plant Physiol. 171, 2294-316.

Rombauts, S., Déhais, P., Van Montagu, M., and Rouzé, P. (1999). PlantCARE, a plant cis-acting regulatory element database. Nucleic acids res. 27, 295-296.

S., K., G., S., M., L., C., K., and K., T. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol. Biol. Evol.1547-1549.

Shaw, L. M., McIntyre, C. L., Gresshoff, P. M., and Xue, G. P. (2009). Members of the Dof transcription factor family in Triticum aestivum are associated with light-mediated gene regulation. Funct Integr Genomics. 9, 485-98.

Song, A., Gao, T., Li, P., Chen, S., Guan, Z., Wu, D., Xin, J., Fan, Q., Zhao, K., and Chen, F. (2016a). Transcriptome-Wide Identification and Expression Profiling of the DOF Transcription Factor Gene Family in Chrysanthemum morifolium. Front Plant Sci. 7, 199.

Song, A., Gao, T., Li, P., Chen, S., Guan, Z., Wu, D., Xin, J., Fan, Q., Zhao, K., and Chen, F. (2016b). Transcriptome-wide identification and expression profiling of the DOF transcription factor gene family in

Chrysanthemum morifolium. Front. Plant Sci. 7, 199.

Takano, M., Haque, M. A., Odaira, S., Nakata, K., Sasaki, N., and Nyunoya, H. (2013). Overexpression of a tobacco Dof transcription factor BBF1 stimulates the transcription of the tobacco mosaic virus resistance gene N and defense-related responses including ROS production. Plant Biotechnol J.12.1207 a.

Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997). A Genomic Perspective on Protein Families. Science 278, 631-637.

Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. Current protocols in bioinform. 2.3. 1-2.3. 22.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids res. 22, 4673-4680.

Thornton, J. W., and DeSalle, R. (2000). Gene Family Evolution and Homology: Genomics Meets Phylogenetics. Annu Rev Genomics Hum Genet ANNU REV GENOM HUM G. 1, 41-73.

Tyman, J. H., and Kiong, L. S. (1978). Long chain phenols: Part XI. Composition of natural cashew nutshell liquid (*Anacardium occidentale*) from various sources. Lipids. 13, 525-32.

Venkatesh, J., and Park, S. W. (2015). Genome-wide analysis and expression profiling of DNA-binding with one zinc finger (Dof) transcription factor family in potato. Plant Physiol Biochem. 94, 73-85.

Wang, Y., Fu, B., Pan, L., Chen, L., Fu, X., and Li, K. (2013). Overexpression of Arabidopsis Dof1, GS1 and GS2 enhanced nitrogen assimilation in transgenic tobacco grown under low-nitrogen conditions. Plant Mol. Biol. Rep. 31, 886-900.

Wei, Q., Wang, W., Hu, T., Hu, H., Mao, W., Zhu, Q., and Bao, C. (2018). Genome-wide identification and characterization of Dof transcription factors in eggplant (Solanum melongena L.). PeerJ. 6, e4481.

Wen, C. L., Cheng, Q., Zhao, L., Mao, A., Yang, J., Yu, S., Weng, Y., and Xu, Y. (2016). Identification and characterisation of Dof transcription factors in the cucumber genome. Sci Rep 6, 23072.

Wu, Z., Cheng, J., Cui, J., Xu, X., Liang, G., Luo, X., Chen, X., Tang, X., Hu, K., and Qin, C. (2016). Genome-wide identification and expression profile of Dof transcription factor gene family in pepper (Capsicum annuum L.). Front. Plant Sci. 7, 574.

Yanagisawa, S., Akiyama, A., Kisaka, H., Uchimiya, H., and Miwa, T. (2004). Metabolic engineering with Dof1 transcription factor in plants: improved nitrogen assimilation and growth under low-nitrogen conditions. Proceedings of the National Academy of Sciences. 101, 7833-7838.

Yang, J., Yang, M.-F., Zhang, W.-P., Chen, F., and Shen, S.-H. (2011). A putative flowering-time-related Dof transcription factor gene, JcDof3, is controlled by the circadian clock in Jatropha curcas. Plant Sci. 181, 667-674.

Yang, Q., Chen, Q., Zhu, Y., and Li, T. (2018). Identification of MdDof genes in apple and analysis of their response to biotic or abiotic stress. Funct Plant Biol. 45, 528-541.

Yang, X., and Tuskan, G. A. (2006). Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. Plant physiol. 142, 820-830.

Yang, Z., and Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 15, 496-503.

Zhang, Z., Yuan, L., Liu, X., Chen, X., and 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.

Zou, H.-F., Zhang, Y.-Q., Wei, W., Chen, H.-W., Song, Q.-X., Liu, Y.-F.,

Zhao, M.-Y., Wang, F., Zhang, B.-C., and Lin, Q. (2013). The transcription factor AtDOF4. 2 regulates shoot branching and seed coat formation in Arabidopsis. Biochem. 449, 373-388.