



Genome-wide identification and characterization of a plant-specific *Dof* transcription factor gene family in olive (*Olea europaea*) and its comparison with *Arabidopsis*

Mariyam¹ · Muhammad Shafiq¹ · Muhammad Haseeb¹ · Rana Muhammad Atif² · Syed Agha Armaghan Asad Abbas Naqvi¹ · Numan Ali¹ · Muhammad Arshad Javed¹ · Fizza Gillani¹ · Muhammad Saleem Haider¹

Received: 10 December 2020 / Revised: 9 May 2021 / Accepted: 11 May 2021 / Published online: 14 June 2021
© Korean Society for Horticultural Science 2021

Abstract

DNA binding with one finger (*Dof*) proteins are encoded by a ubiquitous plant-specific transcription factor gene family that plays a critical role in various biological processes including fruit ripening and organogenesis. The wild olive (*Olea europaea* var. *sylvestris* v1.0) genome was used to identify *Dof* gene family members using a set of bioinformatics tools. Gene structure, chromosome locations, phylogeny, protein motifs, miRNA targets and tissue-specific expression patterns were analyzed. Here, we identified 51 potential *Dof* genes unevenly distributed on all chromosomes and a few scaffolds. *Dof* proteins in olive clustered into eight subgroups (D1, B2, C3, C2.2, C1, C2.1, B1, and A) based on the established *Arabidopsis* classification. The prevalence of segmental duplication was observed as compared to tandem duplication, and this was the main factor underlying the expansion of the *Dof* gene family in olive. Tissue-specific expression profiling of *OleaDof* genes revealed that the majority of *OleaDof* genes were highly expressed in flowers, stem and meristem tissues. In seed and meristem tissues, cis-regulatory element (CRE) analysis revealed the presence of elements that are specifically responsive to light, circadian, auxin, and ABA. In addition, a comparative analysis between *Dof* genes in olive and *Arabidopsis* revealed eight groups or sub-families, although the C3 group of *Arabidopsis* was not represented in olive. This extensive genome evaluation of the *Dof* gene family in olive presents a reference for cloning and functional analysis of the members of this gene family.

Keywords Olive · *Dof* · Plant specific transcription factor · Genome-wide · Bioinformatics

1 Introduction

DNA binding with One Finger (*Dof*) transcription factors is crucial in plant growth and development (Gupta et al. 2015; Malviya et al. 2015). This plant-specific transcription factors gene family encodes proteins with a highly conserved domain of 50–52 amino acids and a C2C2-type typical zinc finger motif, a DNA binding motif, at the N-terminus (Song

et al. 2016; Zou et al. 2013). *Dof* transcription factors play various roles in processes that are unique in plants including nitrogen assimilation (Wang et al. 2013a; Yanagisawa et al. 2004), accumulation of seed storage proteins (Dong et al. 2007), carbon metabolism (Gupta et al. 2015), association with the intracellular trafficking of proteins (Chen et al. 2013), endosperm-specific responses (Diaz et al. 2005), defence responses (Takano et al. 2013), seed germination (Noguero et al. 2013), tolerance of drought and salt (Ma et al. 2015), photoperiodic control of flowering (Fornara et al. 2009), regulation of the formation of branches, shoots and seed coats (Zou et al. 2013), and regulation of genes associated with stomatal functioning and morphogenesis (Negi et al. 2013), as well as being related to the circadian cycle (Yang et al. 2011).

The existence of huge variation of *Dof* genes in terms of number in various crops indicates a great likelihood of diversification in functions. Genome-wide analysis on the relative

Communicated by Seon-In Yeom, Ph.D.

✉ Muhammad Shafiq
shafiq.iags@pu.edu.pk

¹ Faculty of Agricultural Sciences, University of the Punjab, P.O BOX. 54590, Lahore, Pakistan

² Department of Plant Breeding and Genetics, Faculty of Agriculture, University of Agriculture, Faisalabad, Pakistan

phylogeny of *Arabidopsis* and rice *Dof* gene families showed 36 and 30 *Dof* genes respectively (Lijavetzky et al. 2003). Likewise, attempts have been made to study the evolutionary characteristics of *Dof* gene families by differentiating 34, 36 and 41 *Dof* genes of tomato, *Arabidopsis* and poplar, respectively (Cai et al. 2013; Yang and Tuskan 2006). The number of potential *Dof* genes in *Solanum tuberosum* (Venkatesh and Park 2015), *Hordeum vulgare*, *Capsicum annuum* (Kang et al. 2016; Wu et al. 2016b), *Chrysanthemum morifolium* (Song et al. 2016), and *Cucurbita* sp. is 35, 24, 34, 33, 20, and 36, respectively (Hernando-Amado et al. 2012; Mena et al. 2002; Moreno-Risueno et al. 2007).

Olea europaea var *sylvestris* is a wild form of the olive tree and is commonly named Oleaster. Mainly found in the Mediterranean Basin, it is considered to be one of the oldest trees worldwide. Previous studies have provided evidence about cultivated olive trees, i.e., *Olea europaea* L. var. *europaea*, being more similar to oleaster species and, hence, support the idea that oleasters have an ancestral relationship with cultivated olive trees (Kassa et al. 2019; Kyriakopoulou and Kalogianni 2020). It is a small, evergreen tree and a diploid species ($2n = 2x = 46$) with an estimated genome size of 3.19 ± 0.047 pg/2C DNA (https://plants.ensembl.org/Olea_europaea_sylvestris/Info/Index). Wild olive is resistant to certain diseases, environmental and climatic conditions (Beghe et al. 2017). The beneficial properties of both wild and cultivated olive oil for human health have led both to have high economic and nutritional value, but these properties have made olive oil one of the agricultural products most susceptible to counterfeit and fraud. Wild olive has a higher level of antioxidant activity and phenolic content, as well as tocopherolic and orthodiphenolic contents that are either equal to or higher to those in extra virgin cultivated olive oil (Bouarroudj et al. 2016). Due to high antioxidant activity, phenolic extracts from leaves of wild olive have been investigated for use in foodstuffs, food additives and functional food materials (Lafka et al. 2013; Mohamed et al. 2007). Cosmetics and pharmaceutical industries have been using wild olive for its valuable characteristics to manufacture products. Research has shown that wild olive has antimicrobial activity against certain bacterial pathogens that target humans (Paudel et al. 2011). Wild olive and its cultivated form have been used to produce several food supplements (Colombo 2016). The complete genome of *Olea europaea* var. *sylvestris* has been sequenced (Unver et al. 2017).

The Mediterranean olive tree (*O. europaea* subsp. *europaea*) is one of the first domesticated trees and a major agricultural crop of high importance in the Mediterranean region because it is the source of olive oil (Baldoni et al. 2006; Breton et al. 2009; Diez et al. 2015; Lumaret and Ouazzani 2001). Despite the rising importance of olive as an economic and nutritious oil fruit crop, no significant research about its *Dof* transcription factors has been reported. The main

objective of this study was to identify and characterize the genes belonging to the *Dof* transcription factor family in the wild olive genome (Cruz et al. 2016) using various bioinformatics tools. Briefly, a systematic approach was followed to identify *Dof* genes from the wild olive genome. Their chromosomal distribution, intron/exon distribution pattern, presence of conserved domains and cis-regulatory elements were also investigated. The comparative phylogenetic analysis of *Dof* genes from wild olive and *A. thaliana* was also carried out to determine the orthologous relationships and to discern their probable functions. Our extensive genome-wide evaluation of *Dof* gene family members in wild olive provides a reference and an opportunity for functional analysis and cloning of the members of this gene family in other olive subspecies.

2 Materials and methods

2.1 Database search and retrieval of sequence

The amino acid sequence of the *Dof* DNA-binding domain was retrieved from Pfam i.e., PF02701. The 59 aa sequence *Dof* domain from *A. thaliana* (Accession no NP_175581) (<http://pfam.xfam.org/>) (Finn et al. 2014). was used for the identification of *Dof* protein-encoding genes in the wild olive proteome database at Phytozome v12 (<https://phytozome.jgi.doe.gov/pz/portal.html/>; https://phytozome-next.jgi.doe.gov/info/Oeuropaea_v1_0/) using BLAST-P (Protein-basic local alignment search tool) (Goodstein et al. 2014). The retrieved amino acid sequences were analyzed using the simple modular architecture research tool (SMART available at <http://smart.embl-heidelberg.de/>) (Letunic and Bork 2018), and the NCBI CDD (Conserved Domain Database) (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi/>) (Lu et al. 2020) with the default parameters. Any predicted proteins lacking the *Dof* conserved domain (*PF02701*) (<https://pfam.xfam.org/family/PF02701/>) were excluded.

2.2 Determination of physio-chemical properties of olive *Dof* proteins

The protein length (amino acid residues), molecular weight, and theoretical pI of *OeuDof* proteins were predicted using the ProtParam tool (<http://web.expasy.org/protparam/>) (Gasteiger et al. 2005). The information for gene IDs, chromosomal positions, and sequences of genes and proteins were retrieved from Phytozome. These *OeuDof* genes were renamed according to the order of their physical positions. The nuclear localization signals in olive *Dof* proteins were predicted through an online server NLSdb (<https://roslab.org/services/nlsdb/>) (Cokol et al. 2000). Subcellular

localization of *OeuDofs* was predicted using the online tool WoLF PSORT (<https://wolfsort.hgc.jp/>) (Horton et al. 2006).

2.3 Gene structure analysis

To investigate the intron/exon arrangement of *OeuDofs*, the genomic and coding sequences of identified genes were retrieved from the Phytozome database. Moreover, the gff3 file of the olive genome was also retrieved from Phytozome v12. These sequences were further used to draw the gene structure using Gene Structure Display Server (GSDS v2.0) (Hu et al. 2015) (available at <http://gsds.cbi.pku.edu.cn/>).

2.4 Multiple sequence alignment and phylogenetic analysis

The amino acid sequences of Dof proteins were aligned using Clustal W version 2.1 (Thompson et al. 2003, 1994), and the phylogeny was created through MEGA X v2.0 (Kumar et al. 2018) with neighbour-joining (NJ) and bootstrapping set at 1000 replications with partial deletion. In all, 51 olive Dof and 35 *Arabidopsis* Dof protein sequences were used for phylogenetic analysis.

2.5 Cis-regulatory elements and conserved motif recognition

For the analysis of promoter regions, a sequence of 1000-bp upstream was retrieved from the initiation codon for each putative *OeuDof* gene. PlantCare database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Rombaux et al. 1999) was then used to predict *cis*-regulatory elements in these sequences and validated in the PLACE databases (<http://www.dna.affrc.go.jp/PLACE/>) (Higo et al. 1998, 1999).

Multiple EM for Motif Elicitation (MEME) (<http://meme.nbcrl.net/meme/>) (Bailey et al. 2015) was used to analyze motifs using the predicted protein sequences of the *OeuDofs* with the maximum number of motifs set as 20. The minimum and maximum width of motifs were set to 6 and 50, respectively, as default values along with other factors.

2.6 Gene duplication and synteny analysis

The time of divergence of the olive *Dof* gene family was estimated using Ks and Ka values. Protein sequence alignments were made using Clustal W and then the Ka and Ks substitution rates were determined using the Nei-Gojobori model through MegaX Software. The rate variation among sites was modelled with a gamma distribution (shape parameter = 1). The Ka/Ks ratios were calculated. The parameters were configured as described in the software package

manuals. The Ka/Ks ratios were estimated to predict the rates of molecular evolution for each paralogous gene pair. The time of divergence (T) was estimated by $T = Ks/2\lambda$, where λ represents the value of 1.8×10^{-9} (Bettaieb and Bouktila 2020).

The Multiple Collinearity Scan toolkit (MCScanX) was adopted to analyze the gene duplication events, with the default parameters (Wang et al. 2013b). To exhibit the synteny relationship of the paralogous *Dof* genes obtained from the olive, the syntenic analysis map was constructed using the Micro Synteny view software in TBtools (Chen et al. 2020).

2.7 Transcriptome analysis

To analyze the organ-specific expression profile of *OeuDof* at various development stages, we obtained previously generated RNA-seq data for olive plant tissues including fruit, flower, leaf, meristem, root and stem of mature olive trees under field conditions (Ramirez-Tejero et al. 2020). For expression profiling, Reads Per Kilobases per Million mapped reads (RPKM) values from RNA-seq data were \log_2 transformed. Expression patterns with hierarchical clustering are displayed in Heatmap illustrator in TBtools (Chen et al. 2020).

2.8 Putative microRNA target site analysis

The micro-RNA (miRNA) datasets of the olive tree were retrieved from Gene Expression Omnibus (GEO) NCBI (Gardiner 2010; Iwamoto 2016; Konishi 2007; Miyashima 2019) in an experiment related to drupe ripening. To identify miRNAs that could target the wild olive *OeuDof* genes, the CDS sequences of all wild *OeuDof* genes were searched for sequences complementary to miRNAs, using psRNATarget (<https://plantgrn.noble.org/psRNATarget/analysis?function=3/>) (Samad 2017) with default parameters.

3 Results

3.1 Identification of *Dof* genes in olive

To identify the *Dof* genes, the sequence of the Dof domain was BLAST searched against the whole genome sequence of a wild olive that was retrieved from the Phytozome database. An initial analysis led to the identification of 53 proteins. The proteins encoded by the same gene isoforms, as well as proteins containing a truncated Dof DNA-binding domain were excluded from the analysis. A total of 51 non-redundant *OeuDof* genes were identified and used for further analysis. These non-redundant Dof protein sequences from wild olive included the highly conserved four cysteine

residues that coordinate with zinc ion and is a typical feature of Dof proteins. Within the highly conserved sequences of olive Dof domain, 24 out of 50 amino acids were found to be 100% conserved in all the Dof domain sequences (Fig. 1). Other conserved residues observed were Cys¹, Pro², Arg³, Cys⁴, Ser⁶, Lys¹⁰, Phe¹¹, Cys¹², Tyr¹³, Asn¹⁵, Asn¹⁶, Tyr¹⁷, Gln²¹, Pro²², Arg²³, Cys²⁶, Cys²⁹, Arg³¹, Trp³³, Thr³⁴, Gly³⁶, Gly³⁷, Arg⁴⁰, Gly⁴⁵ (Fig. 1) while the other 24 or 26 amino acids were discovered to be variable in all the OeuDof proteins. The OeuDof genes encode proteins ranging from 162 to 574 amino acids in length and with a molecular weight that ranges from 16.2 to 63.3 kDa, with OeuDof 41 being the smallest and OeuDof 23 being the longest protein (Table 1). The isoelectric points of identified proteins ranged from 4.73 to 10.00. A total of 9 OeuDofs namely OeuDof 7, OeuDof 41, OeuDof 43, OeuDof 44, OeuDof 45, OeuDof 46, OeuDof 48, OeuDof 49 and OeuDof 51 showed nuclear localization signals (NLSs) as predicted using NLSdb (<https://roslab.org/services/nlsdb/>) (Cokol et al. 2000). The NLS signal for OeuDof 41, OeuDof 43, OeuDof 44, OeuDof 45, OeuDof 46, OeuDof 48, and OeuDof 49 were found to be GAGRRK while OeuDof 7 and OeuDof 51 showed PKK-GRK and TLASMR, respectively (Table S2).

3.2 Gene structures and recognition of conserved motifs and domains

The organization of exons and introns provides the backbone of genes and helps in assisting verification for the study of evolutionary relationships between genes or organisms (Koralewski and Krutovsky 2011). Their numbers and distribution patterns are an evolutionary mark for a gene family. A comprehensive demonstration of the exon–intron structures of olive Dof genes along with phylogeny revealed that the gene structure pattern was consistent with the phylogenetic analysis. The number of introns varied from one to seven in olive (Fig. 2; Table S5). Twenty-six OeuDof genes are without an intron (50.9%), twenty OeuDof genes have one intron (39.2%), four OeuDof genes have two introns (7.8%), one

(OeuDof33) gene contains 3 introns and one (OeuDof 23) gene contains 7 introns (Table S5; Fig. 2). All of the OeuDof genes in subfamily D2 possessed no introns, while the number of introns in the OeuDof gene subfamily C2.1 varied from zero to three (Table S5). Similar to the Dof genes studied in various species, some Dof genes in olive possess no intron while other Dof genes possess multiple introns, up to seven (Table S5; Fig. 2).

The identification and distributions of 20 motifs within all the wild olive Dof proteins were studied using the MEME program (Fig. 3). The presence of the Dof domain was consistent among all the OeuDof proteins. It was observed that the Dof genes present in the same group encode motifs that are alike, which suggests that these conserved motifs take an essential part in activities that are specific in a group or subgroup. All 51 olive Dof genes encode the same Dof domain. In OeuDof23 an additional domain Fibronectin type-III was found. This FN3 like the domain is also present at the C-terminus of cucumisin proteins, a serine protease from melon fruits (Wen et al. 2016). The distribution of similar motifs among various Dof genes suggests that such genes might have come into existence as a result of gene expansion.

3.3 Comparative phylogenetic relatedness of olive Dof gene family with Arabidopsis

To investigate the evolutionary relationships between the Dof transcription factors of wild olive and *Arabidopsis thaliana*, a Neighbor-Joining (NJ) phylogenetic tree was constructed through MEGA X by aligning their full-length protein sequences. The results depicted that 51 OeuDof proteins were distributed among 8 subgroups named D1, B2, C3, C2.2, C1, C2.1, B1, A, and D2 (Table S1; Fig. 4). Group D1 consisted of 16 Dof proteins in which 7 are *Arabidopsis* Dof-like proteins, ATG69570, ATG26790, AT5G39660, AT3G47500, AT5G62430, AT1G29160 and AT2G34140, while the remaining ones are olive OeuDof 47, OeuDof 41, OeuDof 42, OeuDof 48, OeuDof 49, OeuDof 43, OeuDof 46, OeuDof 44, and OeuDof 45.

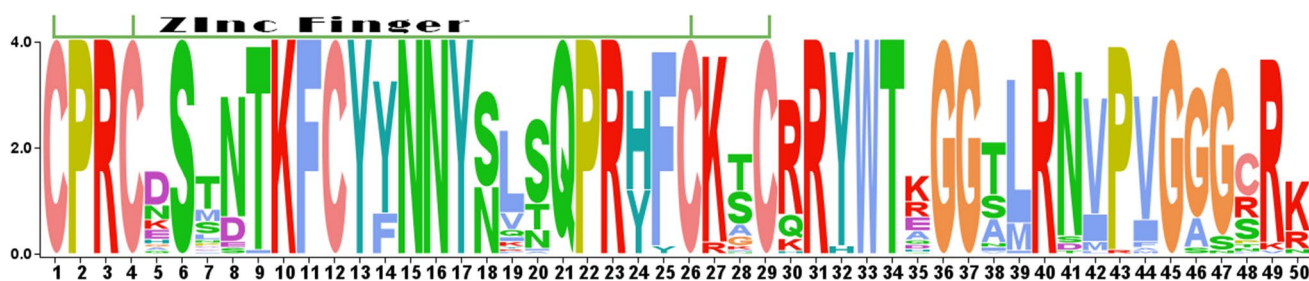


Fig. 1 Sequence logos based on alignments of all olive Dof domains. Dof domains are highly conserved across all 51 Dof proteins in olive. Multiple alignment analysis of 51 typical olive Dof domains was performed with ClustalW. The bit score indicates the information con-

tent for each position in the sequence. (Cys) in the Dof domain are conserved and are present at positions 1, 4, 12, 26, and 29. The zinc finger motif is also indicated as the green line

Table 1 Information about 51 non-redundant Dof genes discovered from the genome of olive

DOF gene name	Source accession no	Chromosome no	Chromosome location (bp)	Direction	Size (AA)		pI	Mw (KD)	Dof Sequence	
					Genome	Peptide			Start	End
<i>OeuDof 1</i>	Oeu052805	18	5,561,837. 5,562,660	F	824	244	8.04	26.5	29	83
<i>OeuDof 2</i>	Oeu032329	20	8,254,985. 8,255,953	F	969	261	9.21	28.6	34	88
<i>OeuDof 3</i>	Oeu017116	18	12,845,937. 12,846,959	R	1023	340	8.83	36.9	36	90
<i>OeuDof 4</i>	Oeu031941	10	33,057,160. 33,057,966	F	807	268	7.03	29.5	23	79
<i>OeuDof 5</i>	Oeu004885	15	18,733,277. 18,734,209	R	933	310	8.65	33.1	51	106
<i>OeuDof 6</i>	Oeu012259	2	583,011. 583,982	F	972	323	8.78	35.0	33	88
<i>OeuDof 7</i>	Oeu048982	Scaffold 532	58,395. 59,204	F	810	269	5.96	29.1	22	78
<i>OeuDof 8</i>	Oeu003808	Scaffold 1135	153,966. 154,991	F	1026	341	9.12	37.5	35	90
<i>OeuDof 9</i>	Oeu036157	Scaffold 309	307,559. 308,736	R	1178	333	9.28	35.9	77	133
<i>OeuDof 10</i>	Oeu004890	15	18,793,875. 18,794,567	R	693	230	7.65	24.2	26	81
<i>OeuDof 11</i>	Oeu057255	18	31,085,205. 31,085,960	F	756	251	5.97	26.8	47	100
<i>OeuDof 12</i>	Oeu032935	18	25,731,540. 25,732,241	F	702	233	8.13	24.6	26	81
<i>OeuDof 13</i>	Oeu007713	Scaffold 12,813	3. 1624	R	1622	298	5.69	33.3	45	102
<i>OeuDof 14</i>	Oeu047445	Scaffold 4979	7714. 8400	F	687	228	8.19	24.7	34	87
<i>OeuDof 15</i>	Oeu056097	6	18,590,646. 18,591,455	R	810	269	9.28	29.5	20	74
<i>OeuDof 16</i>	Oeu038944	20	11,109,220. 11,110,029	F	810	269	8.77	29.1	18	74
<i>OeuDof 17</i>	Oeu035511	Scaffold 3018	53,416. 54,626	F	1211	341	9.30	36.4	78	133
<i>OeuDof 18</i>	Oeu031135	7	15,785,883. 15,786,422	R	540	179	10.0	20.2	27	78
<i>OeuDof 19</i>	Oeu049234	Scaffold 537	592,158. 593,386	F	1229	341	9.34	36.2	75	130
<i>OeuDof 20</i>	Oeu064790	Scaffold 994	132,995. 134,153	R	1159	335	8.77	36.7	52	108
<i>OeuDof 21</i>	Oeu060032	Scaffold 83	504,289. 505,500	R	1212	334	9.31	35.8	74	130
<i>OeuDof 22</i>	Oeu005921	4	12,645,151. 12,645,894	R	744	247	8.75	27.5	26	82
<i>OeuDof 23</i>	Oeu022270	Scaffold 1985	60,199. 64,538	R	4340	574	8.70	63.3	19	75
<i>OeuDof 24</i>	Oeu043709	9	4,681,108. 4,681,812	R	705	234	8.26	25.6	34	87
<i>OeuDof 25</i>	Oeu024355	15	31,061,969. 31,063,222	R	1254	350	9.03	37.2	74	130
<i>OeuDof 26</i>	Oeu037045	11	1,884,207. 1,885,633	F	1427	278	8.15	30.8	34	90

Table 1 (continued)

DOF gene name	Source accession no	Chromosome no	Chromosome location (bp)	Direction	Size (AA)		pI	Mw (KD)	Dof Sequence	
					Genome	Peptide			Start	End
<i>OeuDof 27</i>	Oeu010751	Scaffold 1406	252,504. 253,505	R	1002	194	9.60	21.9	38	94
<i>OeuDof 28</i>	Oeu038532	11	779,351. 780,160	F	810	269	8.47	29.9	24	80
<i>OeuDof 29</i>	Oeu041074	12	11,924,420. 11,925,544	F	1125	221	9.02	24.5	18	73
<i>OeuDof 30</i>	Oeu004964	15	19,737,831. 19,739,455	F	1625	365	8.07	40.2	39	95
<i>OeuDof 31</i>	Oeu023671	1	10,831,636. 10,832,831	F	1196	254	8.99	28.4	19	73
<i>OeuDof 32</i>	Oeu064789	Scaffold 994	122,612. 123,412	F	801	266	8.94	29.0	44	97
<i>OeuDof 33</i>	Oeu022049	22	12,944,083. 12,945,831	F	1749	372	8.98	41.4	36	93
<i>OeuDof 34</i>	Oeu022050	22	12,934,614. 12,935,839	F	1226	269	6.18	29.7	36	93
<i>OeuDof 35</i>	Oeu046008	14	1,728,041. 1,728,889	R	849	282	4.91	31.5	25	79
<i>OeuDof 36</i>	Oeu042946	3	10,077,888. 10,078,616	F	729	242	8.83	26.7	24	80
<i>OeuDof 37</i>	Oeu004984	16	7,213,102. 7,214,019	R	918	305	9.20	33.8	75	131
<i>OeuDof 38</i>	Oeu017719	17	10,509,590. 10,511,334	F	1745	276	8.65	31.0	13	67
<i>OeuDof 39</i>	Oeu028649	2	26,380,398. 26,381,397	R	1000	266	6.43	29.1	14	67
<i>OeuDof 40</i>	Oeu032971	Scaffold 2757	1205. 2038	R	834	277	4.73	31.0	25	79
<i>OeuDof 41</i>	Oeu039921	16	13,290,133. 13,290,621	F	489	162	8.80	18.2	48	103
<i>OeuDof 42</i>	Oeu003911	3	11,409,236. 11,409,670	R	435	144	9.26	16.2	46	102
<i>OeuDof 43</i>	Oeu002092	18	27,052,674. 27,055,067	R	2394	502	5.30	54.7	151	207
<i>OeuDof 44</i>	Oeu041729	Scaffold 3906	12,738. 14,651	F	1914	483	6.45	53.4	147	202
<i>OeuDof 45</i>	Oeu047006	Scaffold 489	247,841. 250,172	R	2332	494	5.48	54.4	151	207
<i>OeuDof 46</i>	Oeu047289	10	24,920,724. 24,923,268	R	2545	496	6.53	54.2	142	198
<i>OeuDof 47</i>	Oeu054867	6	23,318,463. 23,320,482	R	2020	493	8.97	53.8	195	251
<i>OeuDof 48</i>	Oeu019467	7	13,822,655. 13,828,037	R	5383	539	6.18	58.6	149	205
<i>OeuDof 49</i>	Oeu015423	Scaffold 1611	155,895. 160,222	F	4328	525	6.68	57.4	156	212
<i>OeuDof 50</i>	Oeu043810	Scaffold 4261	23,052. 23,876	F	825	274	9.77	29.7	34	89
<i>OeuDof 51</i>	Oeu056096	6	18,549,774. 18,550,577	F	804	267	9.75	28.9	35	89

AA, amino acid sequence length; MW, molecular weight; pI, isoelectric point

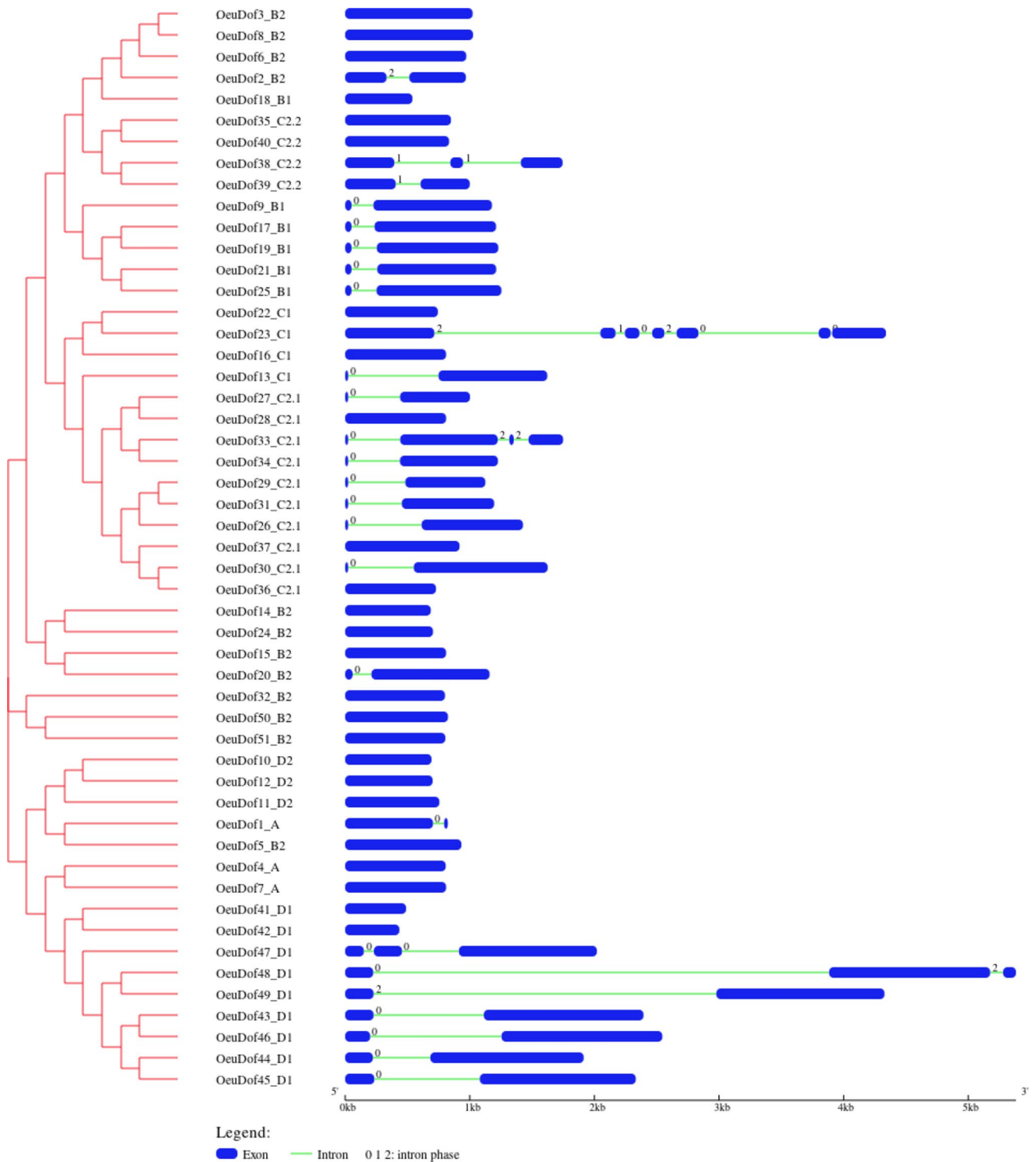


Fig. 2 Phylogenetic relationships and gene structures of *Dof* genes from the olive. **a** The phylogenetic tree was constructed based on the full-length sequences of olive *Dof*. **b** Exon–intron structures of the olive *Dof* genes. Blue boxes indicate exons; green lines indicate introns

Group C3 consisted of 4 *Dof*-like proteins of *Arabidopsis* AT4G21030, AT4G21040, AT4G21050 and AT4G21080, and none of the *Dof*-like proteins present in this clade belongs to olive. Group B2 contained 15 *Dof*-like

proteins in which only 3 are of *Arabidopsis*, AT4G38000, AT5G65590, AT1G28310, while, the rest of the positions belong to Olive, OeuDof 5, OueDOF2, OeuDof 6, OeuDof 3, OeuDof 8, OeuDof 24, OeuDof 14, OeuDof

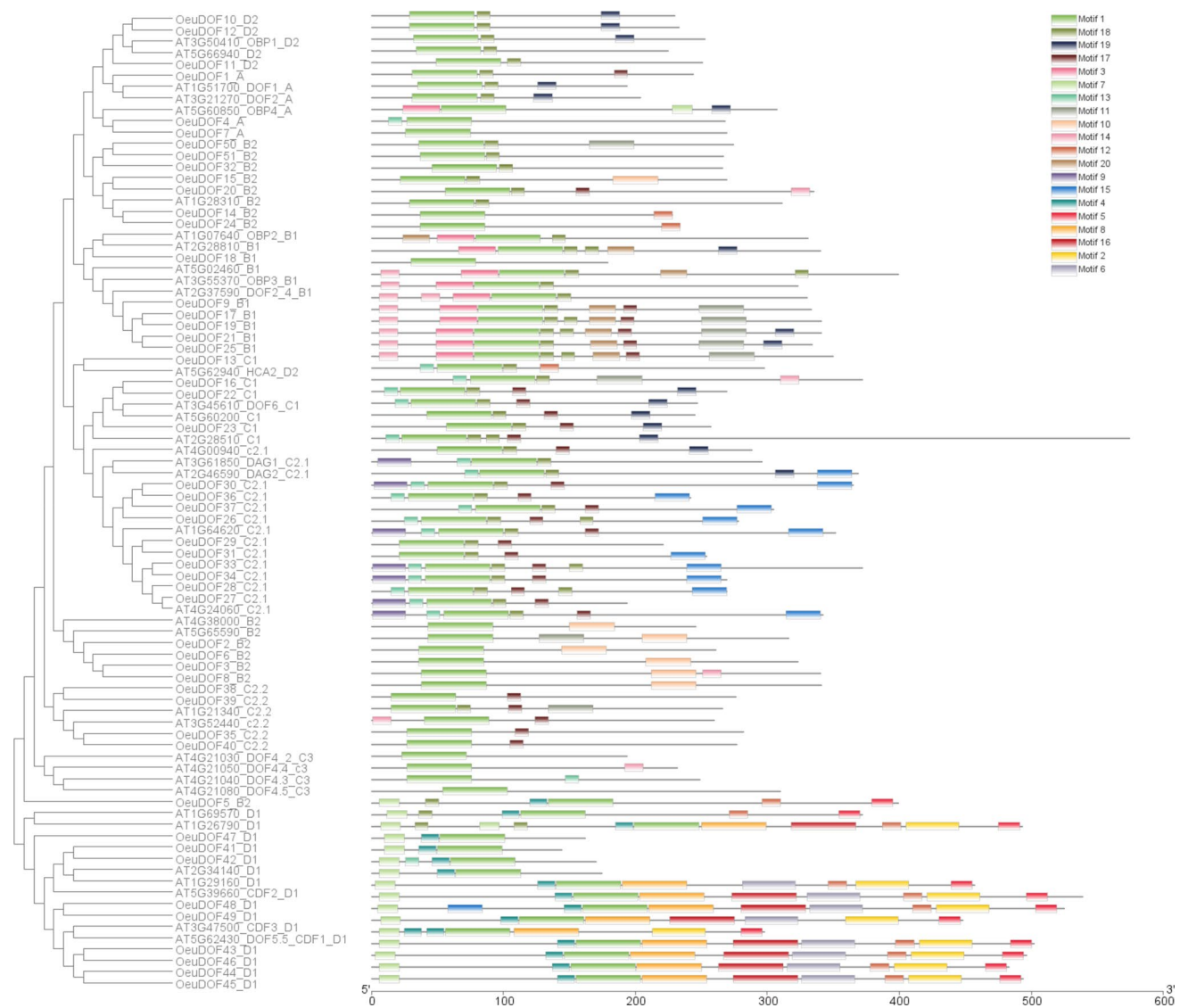


Fig. 3 Distribution of 20 motifs on 51 Dof proteins of olive. Analysis was carried out using MEME version 4.9.0 and interlinking it with the phylogenetic tree to develop a good understanding of their asso-

ciation. The bars represent motifs with different colour codes for different types of motifs

20, OeuDof 15, OeuDof 51, OeuDof 50 and OeuDof 32. Group C1 contained 8 Dof-like proteins in which 4 are of *Arabidopsis* AT5G62940, AT2G28510, AT3G45610 and AT5G60200 and 4 are of olive, OeuDof 13, OeuDof 23, OeuDof 16 and OeuDof 22. Group C2.1 had 15 Dof-like proteins, 5 of them are of *Arabidopsis* AT4G00940, AT3G61850, AT1G64620, AT2G46590 and AT4G24060, while the remaining 10 belong to olive OeuDof 26, OeuDof 37, OeuDof 30, OeuDof 36, OeuDof 29, OeuDof 31, OeuDof 33, OeuDof 34, OeuDof 28, and OeuDof27. Group B1 consisted of 11 Dof-like proteins, 5 proteins of *Arabidopsis*, AT1G07640, AT2G28810, AT5G02460, AT3G55370, AT2G37590 and 6 proteins of olive, OeuDof

18, OeuDof 9, OeuDof 17, OeuDof 19, OeuDof 21 and OeuDof 25. Group A consisted of 6 Dof-like proteins, AT5G60850, AT3G21270, AT1G51700, OeuDof 4, OeuDof 7 and OeuDof 1, 3 each of *Arabidopsis* and olive respectively. Group D2 consisted of 5 Dof-like proteins, 2 of *Arabidopsis*, AT3G50410, AT5G66940, and 3 of olive, OeuDof 10, OeuDof 11 and OeuDof 12. Proteins in common clades usually seem to show similarity in structure and function (Fig. 4). So, all the Dof-like proteins in similar clades may have a similar structure as well as functions. 18 amino acids present in the Dof domain sequence of olive and *Arabidopsis* were found to exist in the same location (Fig. S1).

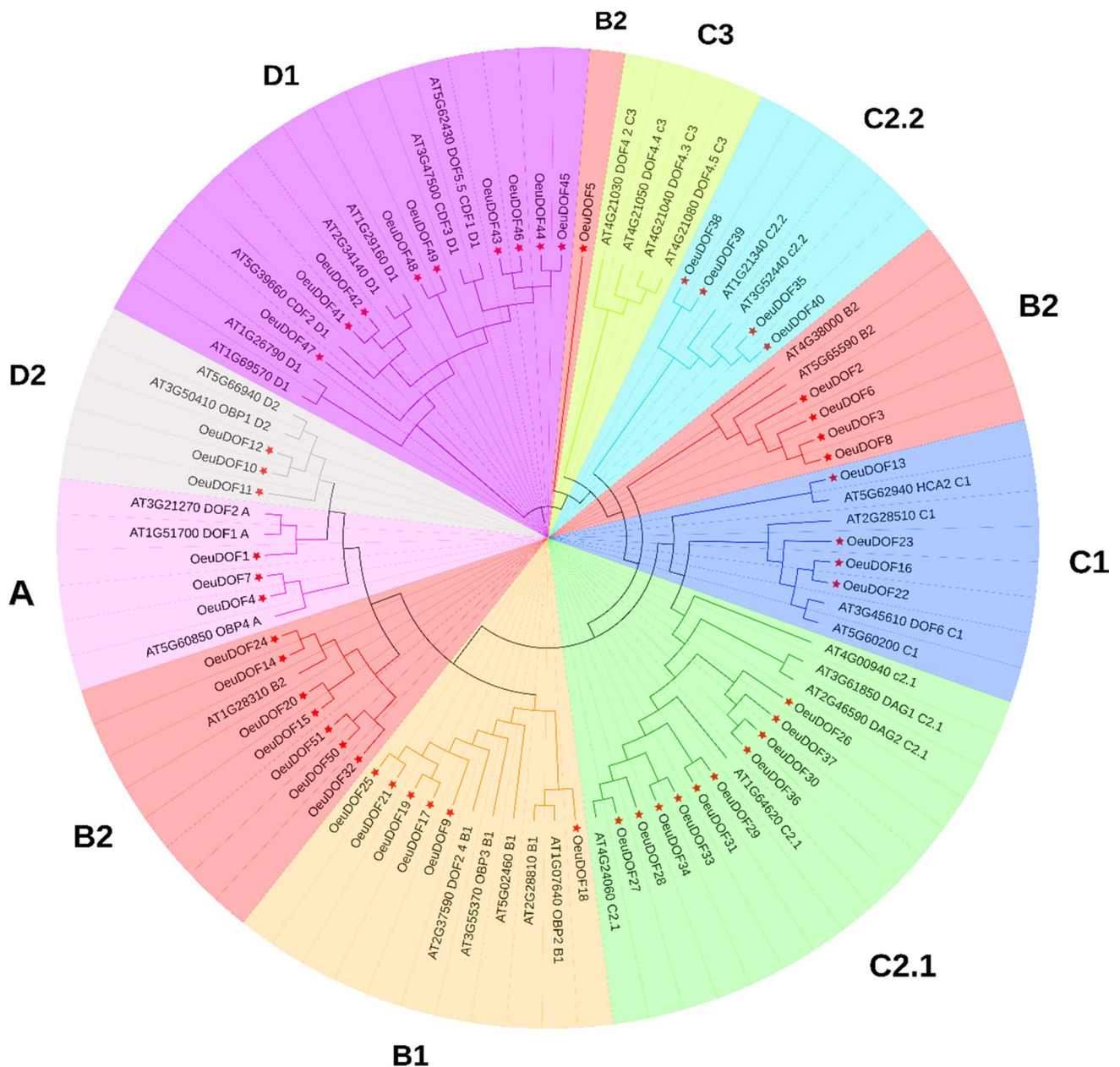


Fig. 4 Phylogenetic relationships between OeuDof and AtDof Proteins. OeuDof proteins are marked with red stars. The evolutionary history was inferred using the UPGMA method with 1000 Bootstrap. This analysis involved 86 amino acid sequences. All ambiguous posi-

tions were removed for each sequence pair (pairwise deletion option). There was a total of 852 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (P.H.A. and R.R., 1973; J., 1985; E. and L., 1965; S. et al., 2018)

3.4 Location of chromosomes and assessment of gene duplication of olive *Dof* genes

Distribution on chromosomes of the analyzed *Olea europaea* var. *Sylvestris* *Dof* genes demonstrated that OeuDof genes were present on various chromosomes. The maximum number, five, of *Dof* genes were located on chromosome 18. Chromosome 15 contained four *Dof* genes, three *Dof* genes were located on chromosome 6, two *Dof* genes were

present on each of the chromosomes 2, 3, 7, 10, 11, 16, 20, and 22, and one *Dof* gene was reported on chromosomes 1, 4, 9, 12, 14, and 17. A total of 17 OeuDofs were identified to be present on the scaffolds that have not been assigned to any chromosome in the wild olive genome assembly by Phytozome Database v12, yet (Fig. 5a).

Furthermore, synteny analysis was performed for OeuDof genes to assess segmental and tandem duplication of the OeuDof gene family in chromosomal location (Fig. 5a, b).

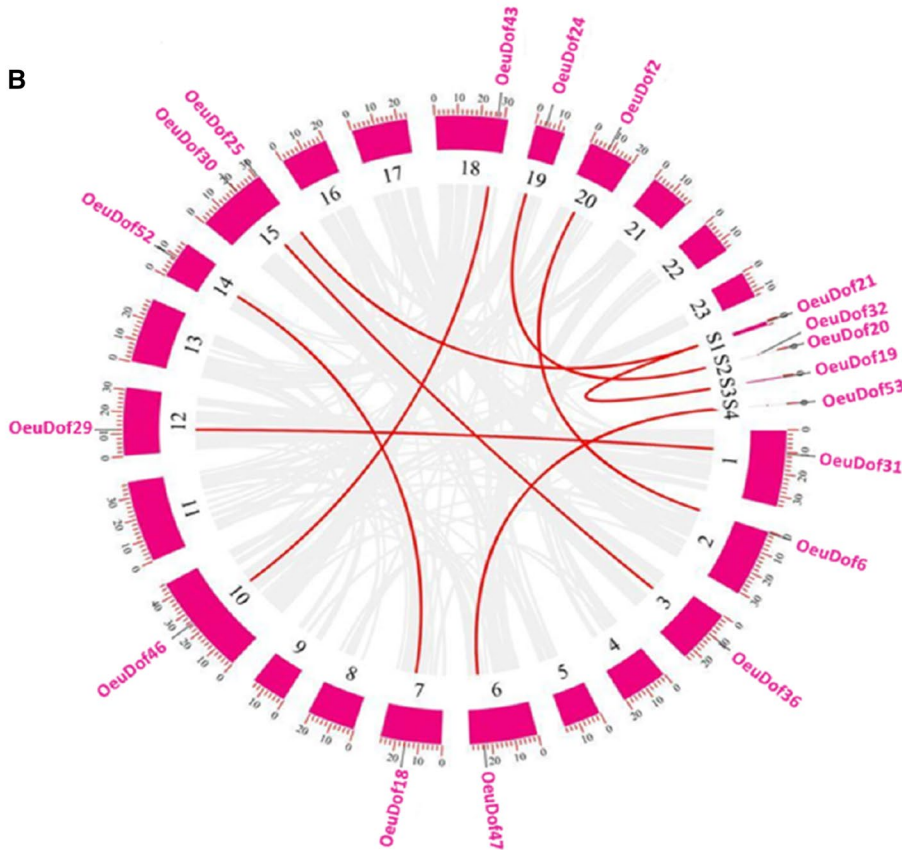
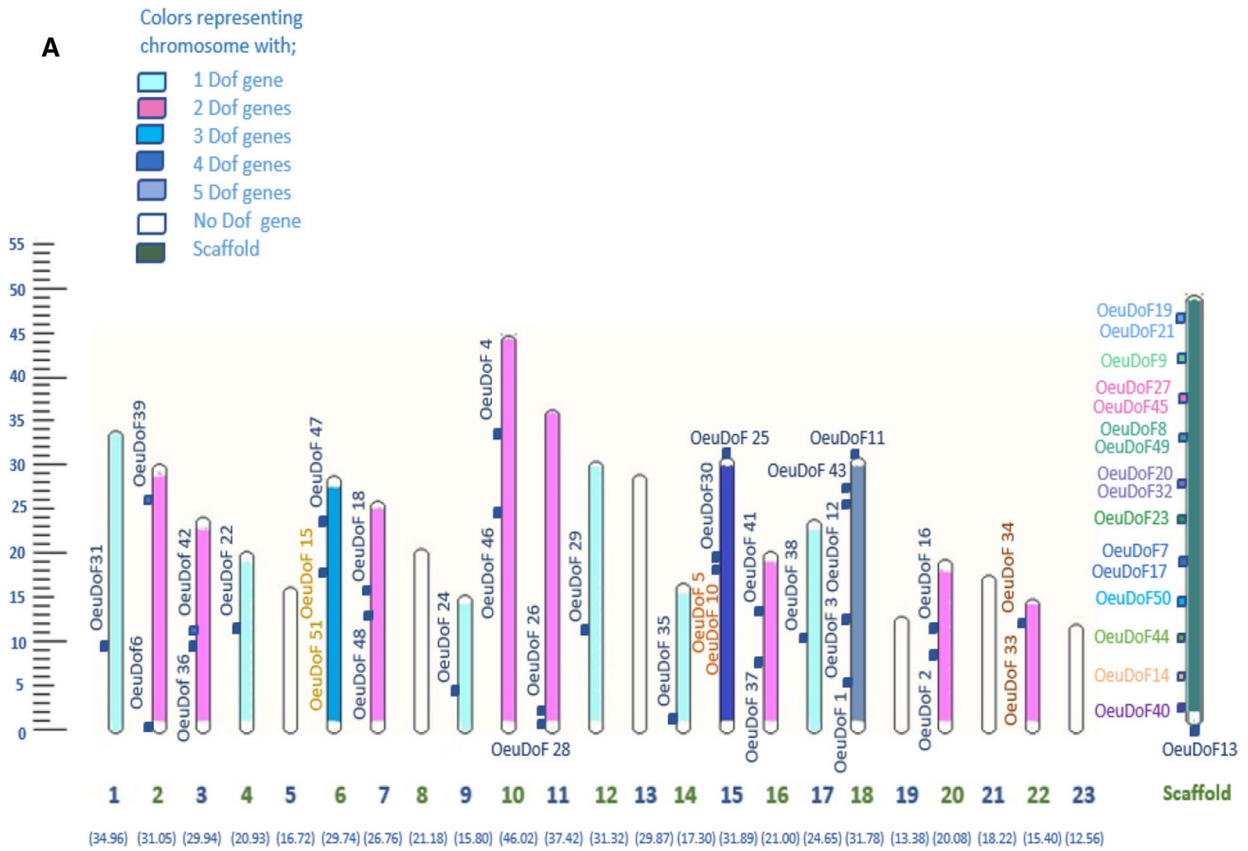


Fig. 5 a Distribution of *OeuDof* genes on olive chromosomes. *Dof* genes that are present on the same location within the same chromosome are coloured differently than the rest of the present genes. Arctic blue colour represents chromosome having 1 *Dof* gene, pink 2, azure blue 3, dark blue 4, royal blue 5, and white colour represents chromosome with no *Dof* genes in them, respectively. The scale represents a 10 Mb chromosomal distance. Genes on the scaffold are mapped imaginary due to lack of full scaffold length data. **b** Genome-wide synteny analysis of olive *Dof* genes showing the dominance of segmental duplication and rare occurrence of tandem duplication

In olive *Dof* genes, 8 paralogous gene pairs were distributed non uniformly in the whole olive genome, which suggested that these genes might have emerged from segmental duplication, whereas 1 paralogous gene pair located together on the same scaffold might result from tandem duplication (Fig. 5a, b).

The date of duplication of genes was also estimated through MEGA-X using pairwise alignment that provided Ks and Ka values and then Ka/Ks was calculated manually (Fig. 6). Ks depicts the number of synonymous substitutions per synonymous site, whereas Ka shows the number of nonsynonymous substitutions per nonsynonymous site, and the ratio of nonsynonymous (Ka) versus synonymous (Ks) mutation was represented by Ka/Ks. This ratio ranged from 0.36 in the *OeuDof 20/OeuDof 32* pair, to 0.86 in the *OeuDof 30/OeuDof 36* pair. The predicted date for tandem gene duplication of the paralogous group *OeuDof 20/OeuDof 32* was calculated to be 526.60 Mya while for the remaining 8 paralogous pairs the segmental duplication date was estimated in the range from 36.50 Mya for paralogous pair *OeuDof 2/OeuDof 6*, to 196.70 Mya for paralogous pair *OeuDof 47/OeuDof 53*. All 9 paralogous group pairs in wild olive had Ka/Ks ratios greater than 0.3 but less than 1, which suggests a probability of considerable functional divergence after the occurrence of duplication due to purifying selection.

3.5 Analysis of cis-regulatory elements

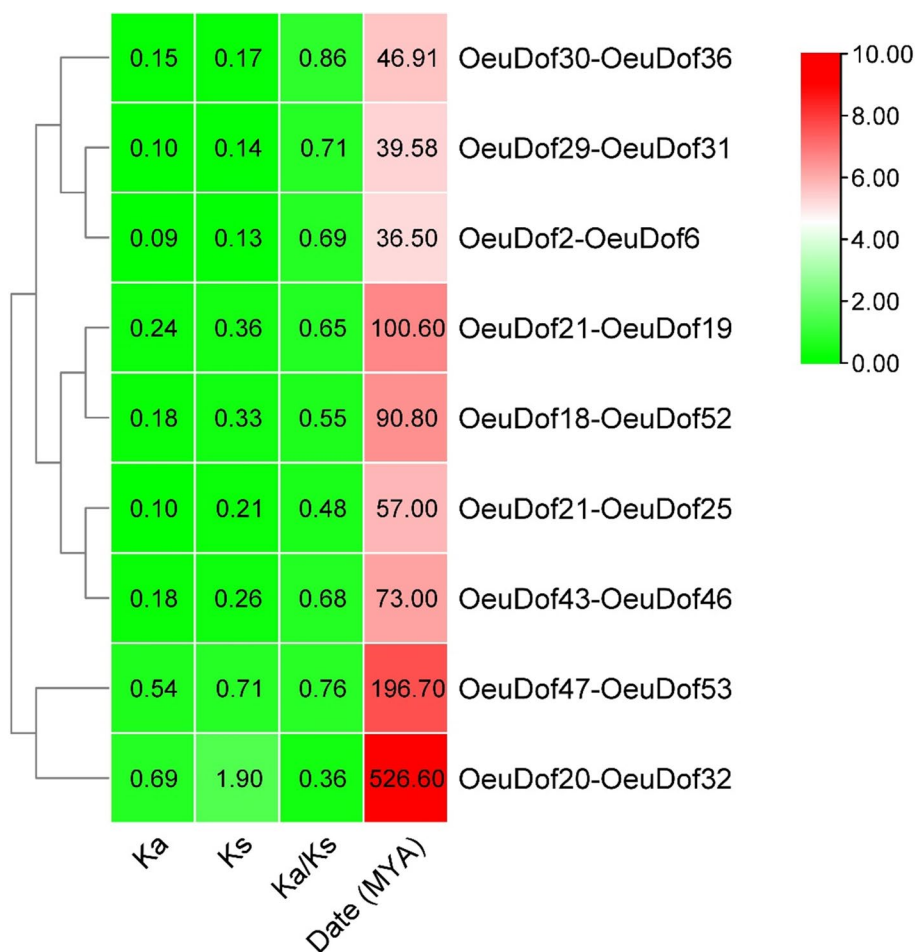
The spatio-temporal transcriptomic expression of genes is affected by the presence and organization of various cis-regulatory elements at the binding site of transcription factors on the promoter region. In-silico analysis of various cis-regulatory elements can be employed to evaluate the putative functions of genes (Bulow and Hehl 2016; Jones and Vandepoele 2020). Cis-regulatory elements with annotated functions such as response to light, seed-specific, endosperm specific, hormone specific, meristem specific, and stress were observed (Fig. 7; Fig. S2). Notably, 35 out of 51 *OeuDof* genes contain the ARE element that is essential for anaerobic induction, 35 *OeuDof* genes possess Box 4 elements, fragment of a conserved DNA module that takes part in light responsiveness, 33 *OeuDof* genes have the ABRE

element that is involved in the abscisic acid response, 25 *OeuDof* genes possess the TGACG element which is responsive to methyl Jasmonic acid, 17 *OeuDof* genes possess the TCA element which is involved in salicylic acid responsiveness, 13 *OeuDof* genes showed the wound-responsive WUN motif, 12 *OeuDof* genes showed TC-rich repeats that show responses in defense and stress, 11 *OeuDof* genes possess the CAT-box that is related to meristem expression, and the MBS element that is related to drought-inducibility, 9 *OeuDof* genes possess the LTR element that is involved in low-temperature responsiveness, 8 *OeuDof* genes possess the auxin responsive TGA element, 6 *OeuDof* genes showed the GC-motif that is involved in anoxic inducibility, 5 *OeuDof* genes possess RY-elements that are specific to seed regulation, 4 *OeuDof* genes possess the GCN4_motif that is involved in endosperm expression, 3 *OeuDof* genes possess the Circadian element, which is involved in circadian control, and the gibberellin responsive GARE-motif. 2 *OeuDof* genes showed the MSA-like element involved in cell cycle regulation. However, *OeuDof 40* and *OeuDof 49* did not show any of the above-listed elements. The cis-regulatory elements identified among 51 *Dof* genes of olive along with their functional annotations are shown in Fig. 7 and Fig. S2.

3.6 Expression analysis of olive *Dof* gene in different organs

Differential expression patterns of all wild olive *Dof* genes in various developmental stages were also analyzed using the available RNA seq data (Ramirez-Tejero et al. 2020). In the following experiment, the data was collected from different anatomical tissues which included fruit, flower, leaf, meristem, root and stem of a mature cultivated olive tree under field conditions. So, these six plant organs were studied in two biological replicates. The expression profiles of wild *OeuDof* are represented in the form of a heat map (Fig. 8). The expression of only 5 *OeuDof* genes was observed in this experiment, which was *OeuDof 1*, *OeuDof 3*, *OeuDof 4*, *OeuDof 11* and *OeuDof 42* because only these 5 hits were found from the available RNA seq data. A simple hypothesis is that since the RNA seq data was extracted from a mature cultivated olive tree, the remaining *OeuDof* genes that were not expressed might have roles and functions in the initial and developmental stages rather than in the mature stage or maybe these *OeuDof* genes have evolved and hence lost their original structure in cultivated olive. Because of the high similarity between these wild *OeuDof* genes to their relative hit RNA seq data, the expression of these *OeuDof* genes can be deduced. Based on gene expression in various tissues, *OeuDof 1* can be grouped with *OeuDof 11*, and similarly, *OeuDof 3* can be grouped with *OeuDof 42* as the level of expression and pattern of these genes is similar; however, the extent of expression of these genes in different organs

Fig. 6 Time of gene duplication estimated for different paralogous pairs of olive *Dof* genes based on K_s and K_a values. Analyses were conducted using the Nei-Gojobori model. K_a represents the number of nonsynonymous substitutions per nonsynonymous site and K_s is the number of synonymous substitutions per synonymous site. While K_a/K_s represents the ratio of nonsynonymous (K_a) versus synonymous (K_s) mutations



was different from each other. *OeuDof 1* showed maximum expression in the stem, followed by a flower which shows their involvement in stem and flower development. Moreover, it showed negligible expression in the roots. On the other hand, *OeuDof 3* showed significant expression only in flower, and very slight expression in leaf and meristem. It was not expressed in fruits, stem and roots. The *OeuDof 4* gene showed expression in all of the organs under study, with maximum expression in leaf, stem and meristem. *OeuDof 11* was highly expressed in meristem and stem regions, and moderately expressed in roots, leaves, and flowers, and it also showed low expression in the fruit. Lastly, *OeuDof 42* showed expression in all regions/organs of olive except leaf, showing its highest expression in flower and meristem regions.

3.7 miRNA targeting of *OeuDof* genes during drupe ripening in *O. europaea* var. *sylvestris*

miRNA analysis of two cultivars of cultivated olive (Cassanese and Leucocarpa) was carried out at 100 DAF (days after flowering) and 130 DAF to analyze the miRNAs that were involved in drupe ripening. For each cultivar, samples

of 30 drupes were taken for both 100 DAF and 130 DAF (Carbone et al. 2019). In the overall experiment, 19 miRNA sequences (ranging from 28 to 32 nucleotides) that were related to 10 wild olive (*Olea europaea* var. *sylvestris*) *Dof* genes were observed. In cultivar Cassanese, at 100 DAF, 5 miRNAs were identified targeting 4 wild olive *Dof* genes (*OeuDof 40*, *OeuDof 44*, and *OeuDof 47*) whereas, at 130 DAF, 8 miRNAs were observed targeting 7 wild olive *Dof* genes (*OeuDof 11*, *OeuDof 15*, *OeuDof 28*, *OeuDof 40*, *OeuDof 46*, *OeuDof 47*, and *OeuDof 48*). In cultivar Cassanese, at 100 DAF, wild olive *Dof* genes *OeuDof 40*, *OeuDof 44*, and *OeuDof 47* each had 1 matched miRNA sequence, while wild olive *Dof* genes *OeuDof 44* had 2 miRNA sequences that matched its sequence.

miRNA sequences were also observed from a second olive cultivar under study (i.e., Leucocarpa). Initially, the results of the 100 DAF sample were analyzed. In 100 DAF sample, 3 miRNA sequences were found corresponding to 3 wild olive *Dof* genes (one sequence per gene). The length of Leucocarpa miRNA strands consisted of 30–31 nucleotide sequences. These miRNAs showed alignment with wild olive genes *OeuDof 1*, *OeuDof 44*, and *OeuDof 48*

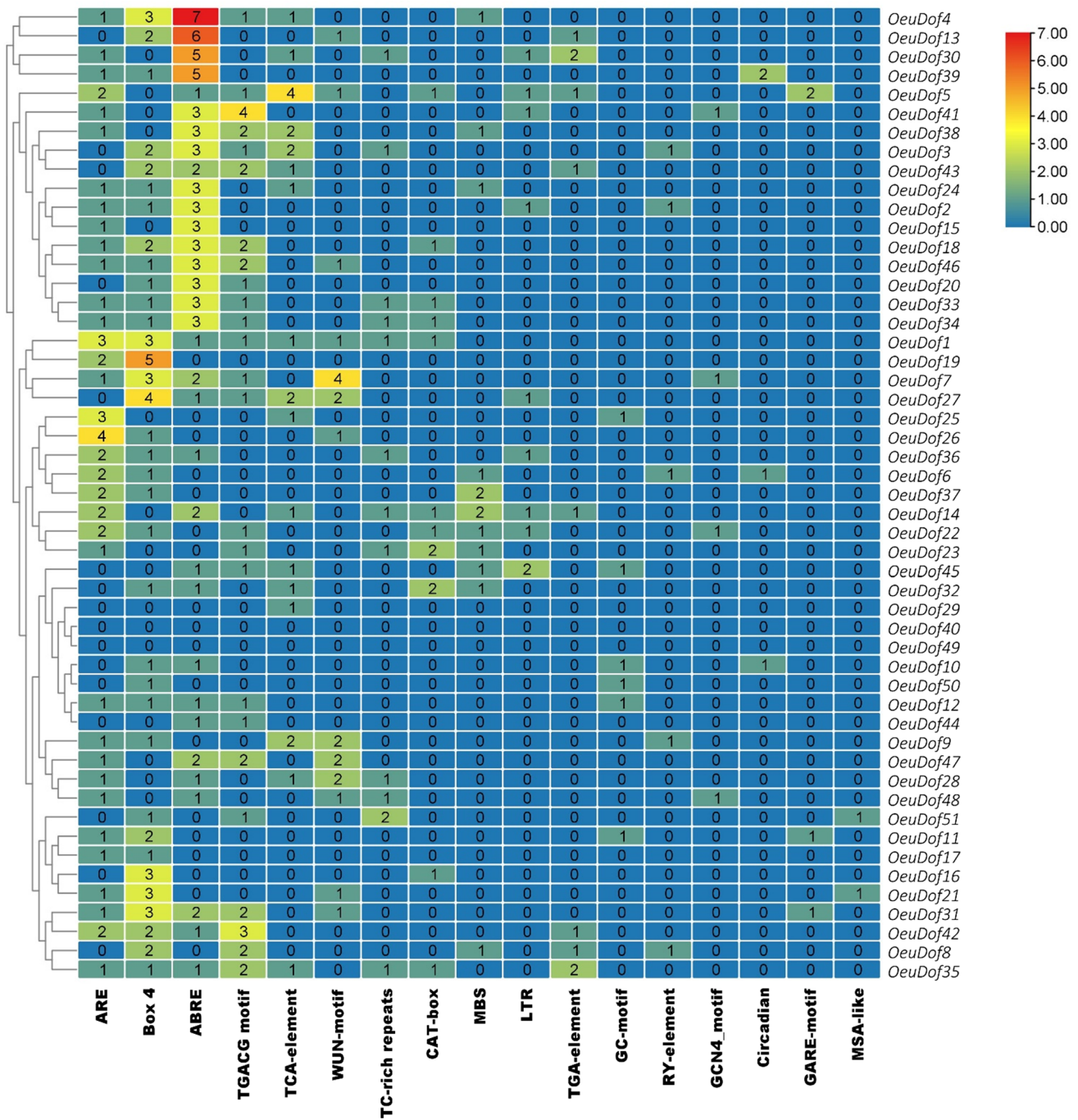


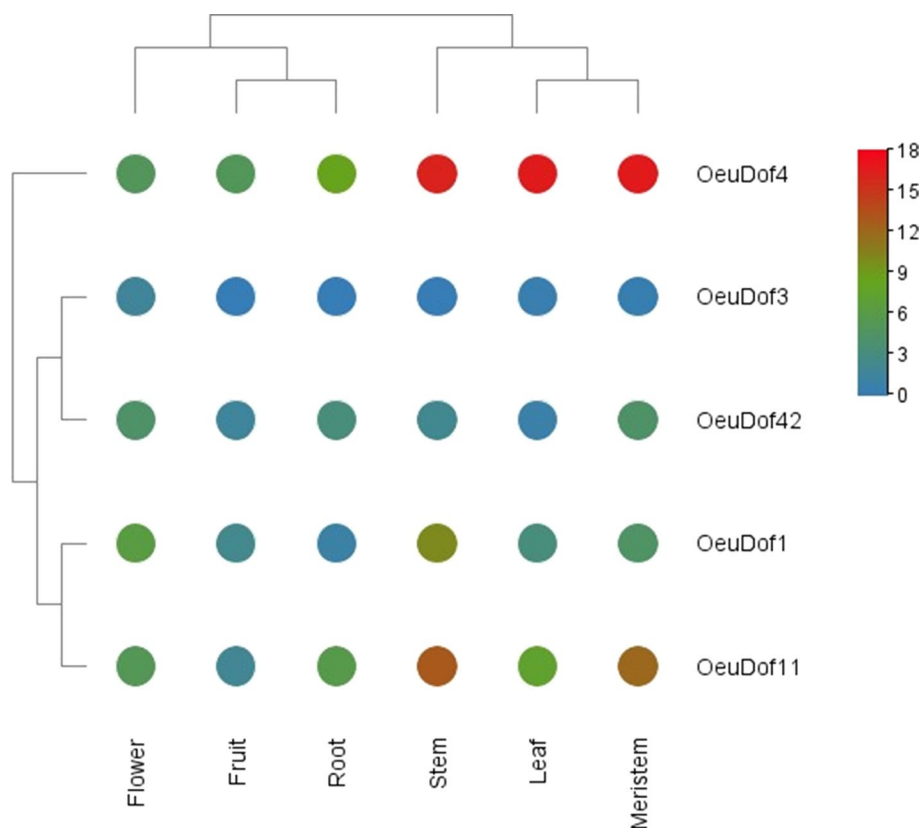
Fig. 7 Different *cis*-acting elements in putative *OeuDof* promoters. Elements are associated with abiotic stresses, hormone responses, growth and development. Colour legends indicating the number of *cis*-elements found in each *OeuDof* gene

(Table S3), and lastly, from the fourth sample (which was of cultivar Leucocarpa at 130 DAF), 3 miRNA sequences were found that aligned with 3 wild olive *Dof* genes, *OeuDof 11*, *OeuDof 40*, and *OeuDof 44*. The strand length of these miRNA sequences consisted of 29–30 nucleotides.

3.8 Putative miRNA targets in *Olea europaea*

The miRNAs sequences were retrieved from the Plant MicroRNA Encyclopedia database. Those miRNAs that could potentially target wild olive (*Olea europaea* var. *sylvestris*) *Dof* genes were identified with the help of

Fig. 8 Heat map of the expression profile for the olive *Dof* genes in different organs in a mature olive tree. The x-axis represents the names of the six organs in a mature olive tree, and the y-axis represents different *OeuDof* genes. The expression levels of *OeuDof* genes are revealed by different colours, which increase from blue to red



psRNATarget online tool (<https://plantgrn.noble.org/psRNATarget/analysis>). A total of 88 miRNAs were found that targeted 28 out of the total 51 wild *OeuDof* genes. The remaining 23 *OeuDof* genes were not targeted by any of these miRNAs (Table S4). The length of these miRNAs ranged from 20 to 22 amino acids. The number of miRNAs targeting these genes ranged from 1 to 19 miRNAs per the wild *OeuDof* gene. *OeuDof* 2, 4, 7, 15, 19, 22, 27, 35, 37, 39, 43, and 46 are the genes that were targeted by only 1 mature miRNA. On the other hand, *OeuDof* 16, 34 and 47 were targeted by 2 mature miRNA each. *OeuDof* 9, 26, 32, 33, 41, and 49 all were targeted by 3 miRNAs. Furthermore, *OeuDof* 20 was targeted by 4 miRNAs, *OeuDof* 23 by 5 miRNAs, *OeuDof* 44 by 7 miRNAs, *OeuDof* 45, and *Dof* 48 were each targeted by 8 miRNAs. Lastly, *OeuDof* 40 was targeted by 19 miRNAs (Table S4). Thus, *OeuDof* 40 was the only gene that was targeted by the maximum number of miRNAs. In terms of the groups, Group D1 was targeted the most, as it was targeted by 33 mature miRNAs. On the other hand, Group A was targeted by only 3 miRNAs, which is the least among all groups. Group B1, B2, C1, C2.1 and, C2.2 were targeted by 4, 8, 8, 10, and 21 miRNAs, respectively (Table S4).

4 Discussion

Transcription factors (TFs) are important regulatory molecules and have the main role in the regulation of gene transcription and networking. Characterization and identification of transcription factors provide a better understanding of plant growth and development under environmental stimuli (Jones and Vandepoele 2020; Wen et al. 2016; Yanagisawa and Schmidt 1999).

According to the phylogenetic and domain analysis of *A. thaliana* (Lijavetzky et al. 2003), citrus (Wu et al. 2016a), and eggplant (Wei et al. 2018), *Dof* transcription factors were divided into nine subfamilies (Group A, B1, B2, C1, C2.1, C2.2, C3, D1 and D2). In this study, we used a recently released wild olive (*Olea europaea* var. *sylvestris*) genome database (https://phytozome-next.jgi.doe.gov/info/Oeuropaea_v1_0/) to identify 51 wild olive *Dof* genes at the genome level (Table 1). 51 *Dof* genes of wild olive were classified into eight subfamilies (Group A, B1, B2, C1, C2.1, C2.2, C3, D1 and D2) using the phylogenetic analysis (Fig. 4; Table S1), and this subfamily grouping was based on the original classification of *Arabidopsis* *Dof* genes; however, *Arabidopsis* *Dof* C3 subfamily

(Zou et al. 2013), AtDOF4.2 (AT4G21030), was missing in the wild olive genome. AtDOF4.2 (AT4G21030) helps in seed coat formation and regulates shoot branching in *Arabidopsis* (Zou et al. 2013). The number of *Dof* genes in wild olive was lower than in banana (74 *MaDof*) (Dong et al. 2016) and Chinese cabbage (76 *BrATDof*) (Ma et al. 2015), but greater than in rice (30 *OsDof*) (Yang and Tuskan 2006), *Arabidopsis* (36 *AtDof*) (Yang and Tuskan 2006), and tomato (34 *SiDof*) (Cai et al. 2013).

The exon–intron structure can also be used as evidence for understanding the evolutionary relationships among genes or organisms (Bondarenko and Gelfand 2016; Koralewski and Krutovsky 2011). The predicted exon–intron association revealed that a total of 26 wild olive *Dof* genes out of 51 were intron-less (Table S5); on the contrary, 10 wild olive *Dof* genes possessed only one intron that was present on the upstream end of the *Dof* domain, and also, another 10 wild olive *Dof* genes possessed only one intron that was present on the downstream end of the *Dof* domain. It was observed that some of the intron and non-intron containing genes were classified in the same group. In general, the wild olive *OeuDof* genes present in the same subfamily shared similar exon–intron structures (Fig. 2), but differences were present in different subfamilies. Similarity of exon–intron structures has also been noticed in *Arabidopsis*, rice and, soybean (Lijavetzky et al. 2003; Gu et al. 2013), which suggested that these structures are evolutionary preserved.

Classification of *OeuDof* genes was also verified by the conserved motif analysis. All of the *OeuDof* protein sequences were imported into the MEME analysis tool to identify the conserved motifs. As a result, a total of twenty conserved motifs were observed, which were statistically significant with *E*-values less than 1×10^{-40} (Fig. 3, Fig. S3, Fig. S4 and Table S6). The motifs of *OeuDof* proteins identified by MEME were between 15 and 50 amino acids in length. Among them, Motif-1 is a common motif in all wild olive *Dof* proteins, corresponding to the CX2CX21CX2C single zinc-finger structure in the *Dof* domain, which is the highly homologous core region of the *Dof* family (Fig. 1). While all of the Group D2 proteins and many of the Group B2, C2.2, and A proteins only contain Motif-1, some *Dof* proteins have extra specific motifs, which may be relevant to different functions. The *Dof* proteins from Group D1 had the most complex motif pattern, and Motif2, Motif 3, Motif 5, Motif 10, and Motif 14 specific for this group. While B1 group members have a relatively simple motif pattern as compared to Group D1, they also had group-specific motifs, such as Motif-7, and Motif-16, but not all the group members have these specific motifs. To understand the potential roles of the Group D1-specific motifs, GO annotations of the Group D1 genes in *Arabidopsis* were checked. Interestingly, we found that in comparison with the *Arabidopsis Dof* genes (Table S7) in other groups, most of the *AtDof* genes

in Group D1 have flower development-related annotations, such as “flower development”, “flowering”, “negative regulation of short-day photoperiodism”, “negative regulation of long-day photoperiodism”, “regulation of timing of the transition from vegetative to reproductive phase”, and “vegetative to the reproductive phase transition of meristem”, which implied the possible functional divergence of the *Dof* genes in group D1 (Table S7).

The distribution of motifs among the wild olive *Dof* proteins (Fig. 3) is indicative of its evolutionary relationship as deduced by the phylogenetic tree (Gupta et al. 2015; Malviya et al. 2015). The motif data analysis by MEME (Fig. 3), and domain analysis using NCBI CDD (Fig. 3) and the alignment of the wild olive *OeuDof* protein sequences (Fig. 1) revealed a highly conserved *Dof* domain, which was observed at the N-terminal region of 42 *OeuDof* genes, and 9 *OeuDof* genes have the *Dof* domain in the central region (Figs. 1, 3). *Dof* transcription factors have been evolutionarily conserved among plants. Apart from the *Dof* domain, nineteen distinct motifs were identified that were differentially distributed among the wild olive *OeuDof* genes (Fig. 3). Meanwhile, at least one or two conservative motif types and spatial distributions in the wild olive *OeuDof* genes are present in the same subfamily while some differences were present between the member of different subfamilies, implying certain functional similarities of the wild olive *Dof* members within the same subfamily. In addition, the wild *OeuDof* genes showed structural conservation in subfamilies and were consistent with other plants such as *Arabidopsis*, banana, rice, and, chickpea (Dong et al. 2016; Lijavetzky et al. 2003; Nasim et al. 2016; Yang and Tuskan 2006). In addition, as predicted by *in silico* analyses, 9 deduced wild olive *OeuDof* genes harboured NLSs to guide their localization in the nucleus (Table S2), but subcellular localization analysis using online tool WoLF PSORT (<https://wolffpsort.hgc.jp/>), predicted nuclear localization in all *OeuDof* proteins, except, *OeuDof 18* (Table S2).

The duplication of genes can be predicted from their locations on the chromosome. This means that two or more genes that are present on the same chromosome can be the result of tandem duplication, whereas duplicated genes that are located on different chromosomes might be the result of segmental duplication (Panchy et al. 2016). The highly dense presence of olive *Dof* genes on chromosomes 18 and 15 (Fig. 5a) is an indication of tandem duplication, but instead of tandem duplication, the predominance of segmental duplication was observed using the synteny analysis (Fig. 5b). The predominance of segmental duplication was also observed in the chickpea (Nasim et al. 2016), and pigeon pea (Malviya et al. 2015) *Dof* gene families. The phenomenon of gene duplication is the main driver of expansion in the gene family, and perhaps, the increase in the number of *Dof* genes in higher plants can be due to duplication of

the domain during the evolution of eukaryotic plants (Taylor and Raes 2004; Moore and Purugganan 2005).

The ratio of Ka/Ks provides an understanding of the selection pressure on the substitution of amino acids (Fig. 6). A ratio of Ka/Ks < 1 suggests the possibility of purifying selection whereas Ka/Ks ratios > 1 suggest the likelihood of positive selection (Yang and Bielawski 2000; Hurst 2002). Generally, evaluation of selective pressure provides selective leads for amino acid sequences altered in a protein and are also necessary for interpreting functional residues and functional protein shifts (Morgan et al. 2010). Ka/Ks ratios of the sequences from the different olive *Dof* groups vary remarkably. Despite the differences, the estimated values of Ka/Ks ranged from 0.36 to 0.86, which being less than 1 suggested that the *Dof* sequences present in each group underwent strong purifying selection pressure and positive selection might have acted on only a few sites during the process of evolution (Fig. 6).

Cluster analysis provides important clues about the function of olive *Dof* genes. *OeuDof* genes showed specific spatial and temporal expression patterns in different organs and developmental stages. As mentioned above, data for 5 out of 51 *OeuDof* genes were found in the data that was downloaded from the NCBI GEO dataset experiments (Ramirez-Tejero et al. 2020). From the expression comparison graph, it can be observed that different *OeuDof* genes are expressed in different organs (Fig. 8). *OeuDof 4* and *OeuDof 1* were expressed more in flower meristem and stem as compared to the underground part which suggested their role in stem meristem and leaf development. *OeuDof 4* and *OeuDof 1* were found in group A that includes *AtDof 1* (AT1G51700), *AtDof 2* (AT3G21270) and *AtOBP 4* (AT5G60850) (Peng and Weselake 2011; Ramirez-Parra et al. 2017; Rymen et al. 2017; Xu and Cai 2019; Xu et al. 2016). *AtDof 1* (AT1G51700) was expressed during the early globular embryo stage, whereas *AtDof 2* (AT4G38000) exhibited a similar expression profile during seed development. These genes are generally expressed all over the plant, more specifically in root, stem, leaf, flower, seed, guard cell, plant embryo, and pollen (Peng and Weselake 2011).

OeuDof 3 (a group B2 member) had shown expression in flowers and no expression in stem, root and fruit. This pattern revealed that this gene might be associated with the reproductive function of the *europaea* var. *sylvestris*. In *Arabidopsis*, the members of group B2 are AT5G65590, AT4G38000 and AT1G28310 (Moreno-Risueno et al. 2007; Yanagisawa 2002). These results are consistent with their orthologue partner in Group B2 that also showed expression during early flower development in *Arabidopsis* (Wellmer et al. 2006), which is a key process in the life cycle of a plant during which floral patterning and the specification of floral organs is established (Wellmer et al. 2006). The cis-regulatory analysis also predicted that *OeuDof3* has roles relating

to light and during abiotic stress (Fig. 7). *OeuDof 11* also showed a strong expression that was very similar to *OeuDof 4* but had a little lower expression as compared to *OeuDof 4*. *OeuDof 11* (Group D2) is closely related to AT5G62940, AT5G66940 and AT3G50410, which are involved in the regulation of cambium formation and vascular tissue development, particularly at a very early stage during inflorescence stem development, and promotes both cambium activity and phloem specification, but prevents xylem specification. These are also expressed in carpel, cauline leaf, collective leaf structure, flower, flower pedicel, hypocotyl, inflorescence meristem, petal, plant embryo, root, seed, sepal, shoot apex, shoot system, stamen, stem, and vascular tissue of leaf; collectively, in the whole plant (Guo 2009; Miyashima 2019; Yanagisawa 2002). In olive, the orthologs of these three *Arabidopsis* proteins are *OeuDof 10*, *OeuDof 11*, and *OeuDof 12*. So, it can be inferred that these *OeuDof* proteins may also have similar roles and functions in the wild olive plant as of their orthologs in *Arabidopsis*.

In the end, *OeuDof 42* (Group D1) appeared to have slight to moderate expression in various organs and parts of the plant. The members of group D1 in *Arabidopsis* are AT1G29160, AT2G34140, AT3G47500 and AT5G39660, which are transcriptional repressors of *CONSTANS* expression and thus regulate the photoperiodic flowering response (Fornara 2009; Fornara et al. 2009; Imaizumi 2005). These proteins are orthologs of the CYCLING DOF FACTOR 1 (CDF1), which interacts with FKF1 and regulates *CO* expression (Imaizumi 2005).

MicroRNAs are very important regulators in plants that regulate almost every biological process ranging from growth and development to combating pathogens and maintaining proper internal conditions (Carbone et al. 2019; Samad 2017; Spanudakis 2014; Terzi 2008). miRNAs are highly conserved among different species, meaning that each microRNA performs a specific function, regardless of the type of species in which they were observed. Cyclic DOF (CDFs) play an important role in blue light signalling. *AtCDF2* acts as a transcriptional activator or repressor of a group of microRNA (miRNA) genes and binds to the primary miRNA transcripts (Sun et al. 2015). *CDF2* is a suppressor of miRNA biosynthesis and acts by targeting Dicer-like 1 (DCL1) complex and suppresses the processing of primary miRNAs. *CDF2* works in the same pathway as miR159 or miR172 to control flowering (Sun et al. 2015). *OeuDof 42* (a putative CDF in olive and Group D1) appeared to have slight to moderate expression in various organs (flowers) and parts of the plant. *SICDF4* gene expression was detected during tomato fruit ripening, whereas *SICDF5* transcripts were abundant only in green fruit, and *SICDF2* showed similar expression in green and red fruit (Corrales et al. 2014). Five wild olive *Dof* genes, including *OeuDof40*, *OeuDof43*, *OeuDof 44*, *OeuDof 47*, and *OeuDof 48* belong to group D1 and

are CDF partners in olive. In this study of possible miRNA targets, olive CDFs (*OeuDof 40*, *OeuDof 43*, *OeuDof 44*, *OeuDof 47*, and *OeuDof 48*) were detected by 19 miRNAs during drupe development in olive and also from the miRNA data downloaded from different miRNA databases.

OeuDof 47 was expressed only in cultivar Cassanese in both 100 DAF and 130 DAF samples. Whereas *OeuDof 48* was observed in “Cassanese 130 DAF” and “Leucocarpa 100 DAF” samples. It can also be inferred from the miRNA analysis that *OeuDof 40* and *OeuDof 44* showed expression in the highest number of samples. *OeuDof 40* was present in all samples except for Leucocarpa at 100 DAF, and *OeuDof 44* was found in all samples except for the “Cassanese 130 DAF” sample. The results clearly showed the variation of expression of the miRNAs. As both cultivars differ in the colour of epidermis and mesocarp of the fruit, and the expression of *OeuDof* genes also differ among the cultivars; these results imply that *OeuDof* genes might also play some role in the pigmentation of the olive fruit. *OeuDof 48* and *OeuDof 49* are also CDF partners and belong to group D1 (Fig. 4; Table 2; Table S1). From the data gathered from plant microRNA encyclopedia (<http://pmiren.com/>), both of these *OeuDof* genes were found to be targeted by miR164 at only a single site. Moreover, it can also be inferred that inhibition of expression of *OeuDof 48* and *OeuDof 49* can induce flowering, and additionally, can also enhance the drought resistance capability of the plant. miR166 regulates floral development by affecting the morphogenesis of flowers (Jung and Park 2007). In olive, this miR166 was observed to target two genes, *OeuDof40*, and *OeuDof35*, and it can be inferred that by inhibiting these *OeuDof* genes floral development can be regulated (Table 2). miR172 has been observed to be involved in multiple processes (Jung et al., 2014; Kim et al., 2014; Li et al., 2016; Sun et al., 2015; Yamashino et al., 2013; Yang et al., 2015). This miRNA has roles at the transitional stages like the transition from juvenile to the adult stage; it is also important in regulating proper shifting of the plant from vegetative to the reproductive stage during maturity. Additionally, it also regulates the proper development of the flower of the plant. In olive, this miR172 was seen to target three genes, *OeuDof44*, *OeuDof45*, and *OeuDof47*, targeting a specific site in each of

these genes (Table 2). The miR172 acting upon *OeuDof44* and *OeuDof45* had two different but nearly identical nucleotide sequences. The orthologues of *OeuDof44*, *OeuDof45*, and *OeuDof47* in *Arabidopsis* are CDFs and are involved in the suppression of *CONSTANS* which are required for floral initiation and development (Fornara 2009; Imaizumi 2005; Sun et al. 2015). So, these three *OeuDof* genes might have the opposite function of any or all of the abovementioned miR172 roles.

The miR159 family is a very abundant miRNA family and represents one of the most ancient miRNAs in the plant kingdom (Wu et al. 2009), which has three members, miR159a, miR159b, and miR159c in *Arabidopsis*. miR159 is involved in the transition from the vegetative to reproductive phase in the plant, as well as in the regulation of flower development (Wu et al. 2009). This miRNA group was found to target 3 *OeuDof* genes which are *OeuDof9*, *OeuDof20*, and *OeuDof32* (Table 2). In all three of these genes, miR159 was targeted at three different regions as three types of miR159 were observed targeting these *Dof* genes. This might suggest that *OeuDof9*, *OeuDof20*, and *OeuDof32* might be involved in promoting vegetative growth within the plant and inhibiting floral and reproductive development.

5 Conclusion

In this study, we reported a comprehensive analysis of *OeuDof* transcription factor genes in the wild olive genome. The 51 *OeuDof* genes were categorized into Eight subgroups and some of the structural and functional properties of each *OeuDof* member were characterized. Most of the *OeuDof* genes were involved in flower and stem development. miRNA data on possibly targeted *OeuDof* genes during drupe development in olive suggested their role in fruit growth and development. The detailed computational inspection of olive *Dof* proteins revealed in the current study might be used for selection and cloning at the molecular level, portraying gene expression and studying their interactions with different transcription factors. The presence of similar numbers of *Dof* genes in some plants such as 33 in tomato, 34 in pepper, and 35 in potato and relatively more *Dof* genes in

Table 2 Information about miRNAs, their functions and target IDs

miRNA ID	Function	Target ID
miR159	Transition from vegetative to reproductive phase in the plant; regulation of flower development	<i>OeuDof9</i> , <i>OeuDof20</i> , and <i>OeuDof32</i>
miR164	Inhibits flowering and drought resistance capability of the plant	<i>OeuDof 48</i> and <i>OeuDof 49</i>
miR166	Regulates floral development by affecting flower morphogenesis	<i>OeuDof40</i> , and <i>OeuDof35</i>
miR172	Role in transitional stages like from juvenile to adult stage transition; regulates proper plant shifting from vegetative to reproductive stage during maturity; regulate proper flower development	<i>OeuDof44</i> , <i>OeuDof45</i> , and <i>OeuDof47</i>

other plants like 78 in soybean, and 51 in olive suggests that duplications might have led to the expansion of *Dof* gene family in some species.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13580-021-00366-7>.

Acknowledgements This work was supported by the University of the Punjab, Lahore, Pakistan.

Author contributions MS, M, MH, SAAAA, RMA, NA, MAJ and FG were involved in data analysis. MS and RMA provided overall direction and experimental design. SAAAA, MH, NA and M wrote the manuscript. All authors read and approved the final manuscript.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

References

- Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME suite. *Nucleic Acids Res* 43:W39–W49
- Baldoni L, Tosti N, Ricciolini C, Belaj A, Arcioni S, Pannelli G, Germana MA, Mulas M, Porceddu A (2006) Genetic structure of wild and cultivated olives in the central Mediterranean basin. *Ann Bot* 98:935–942
- Beghe D, Piotti A, Satovic Z, de la Rosa R, Belaj A (2017) Pollen-mediated gene flow and fine-scale spatial genetic structure in *Olea europaea* subsp. *europaea* var. *sylvestris*. *Ann Bot* 119:671–679
- Bettaieb I, Bouktila D (2020) Genome-wide analysis of NBS-encoding resistance genes in the Mediterranean olive tree (*Olea europaea* subsp. *europaea* var. *europaea*): insights into their molecular diversity, evolution and function. *Tree Genet Genomes* 16:1–21
- Bondarenko VS, Gelfand MS (2016) Evolution of the exon–intron structure in ciliate genomes. *PLoS ONE* 11:e0161476
- Bouarroudj K, Tamendjari A, Larbat R (2016) Quality, composition and antioxidant activity of Algerian wild olive (*Olea europaea* L. subsp. *Oleaster*) oil. *Ind Crops Prod* 83:484–491
- Breton C, Terral JF, Pinatel C, Medail F, Bonhomme F, Berville A (2009) The origins of the domestication of the olive tree. *C R Biol* 332:1059–1064
- Bulow L, Hehl R (2016) Bioinformatic identification of conserved *cis*-sequences in coregulated genes. *Methods Mol Biol* 1482:233–245
- 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
- Carbone F, Bruno L, Perrotta G, Bitonti MB, Muzzalupo I, Chiappetta A (2019) Identification of miRNAs involved in fruit ripening by deep sequencing of *Olea europaea* L. transcriptome. *PLoS ONE* 14:e0221460
- Chen H, Ahmad M, Rim Y, Lucas WJ, Kim JY (2013) Evolutionary and molecular analysis of D of transcription factors identified a conserved motif for intercellular protein trafficking. *New Phytol* 198:1250–1260
- Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* 13:1194–1202
- Cokol M, Nair R, Rost B (2000) Finding nuclear localization signals. *EMBO Rep* 1:411–415
- Colombo M (2016) Botanicals authentication in food, food supplements and herbal medicinal products. *Plant Biosyst Int J Deal All Aspects Plant Biol* 150:22–26
- Corrales AR, Nebauer SG, Carrillo L, Fernandez-Nohales P, Marques J, Renau-Morata B, Granell A, Pollmann S, Vicente-Carbajosa J, Molina RV, 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
- Cruz F, Julca I, Gómez-Garrido J, Loska D, Marcet-Houben M, Cano E, Galán B, Frias L, Ribeca P, Derdak S (2016) Genome sequence of the olive tree, *Olea europaea*. *Gigascience* 5:29
- Diaz I, Martinez M, Isabel-LaMoneda I, Rubio-Somoza I, Carbonero P (2005) The DOF protein, SAD, interacts with GAMYB in plant nuclei and activates transcription of endosperm-specific genes during barley seed development. *Plant J* 42:652–662
- Diez CM, Trujillo I, Martinez-Urdiroz N, Barranco D, Rallo L, Marfil P, Gaut BS (2015) Olive domestication and diversification in the Mediterranean Basin. *New Phytol* 206:436–447
- Dong G, Ni Z, Yao Y, Nie X, 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
- Dong C, Hu H, Xie J (2016) Genome-wide analysis of the DNA-binding with one zinc finger (Dof) transcription factor family in bananas. *Genome* 59:1085–1100
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J (2014) Pfam: the protein families database. *Nucleic Acids Res* 42:D222–D230
- Fornara F, Panigrahi KC, Gissot L, Sauerbrunn N, Rühl M, Jarillo JA, 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
- Gardiner J, Sherr I, Scarpella E (2010) Expression of DOF genes identifies early stages of vascular development in Arabidopsis leaves. *Int J Dev Biol* 54(8–9):1389–1396
- Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExpASY server. The proteomics protocols handbook. Springer, Berlin, pp 571–607
- Goodstein D, Batra S, Carlson J, Hayes R, Phillips J, Shu S, Schmutz J, Rokhsar D (2014) Phytozome comparative plant genomics portal
- Gu X, Zou Y, Su Z, Huang W, Zhou Z, Arendsee Z, Zeng Y (2013) An update of DIVERGE software for functional divergence analysis of protein family. *Mol Biol Evol* 30:1713–1719
- Guo Y (2009) Dof5.6/HCA2, a Dof transcription factor gene, regulates interfascicular cambium formation and vascular tissue development in Arabidopsis. *Plant Cell* 21:3518–3534
- Gupta S, Malviya N, Kushwaha H, Nasim J, Bisht NC, Singh V, Yadav D (2015) Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. *Planta* 241:549–562
- Hernando-Amado S, Gonzalez-Calle V, Carbonero P, 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
- Higo K, Ugawa Y, Iwamoto M, Higo H (1998) PLACE: a database of plant cis-acting regulatory DNA elements. *Nucleic Acids Res* 26:358–359
- Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27:297–300
- Horton P, Park K-J, Obayashi T, Nakai K (2006) Protein subcellular localization prediction with WoLF PSORT. In: Proceedings of the 4th Asia-Pacific bioinformatics conference. World Scientific, Singapore, pp 39–48

- Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31:1296–1297
- Hurst LD (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet* 9:486–487
- Imaizumi T (2005) FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. *Science* 309:293–297
- Iwamoto M, Tagiri A (2016) Micro RNA-targeted transcription factor gene RDD 1 promotes nutrient ion uptake and accumulation in rice. *Plant J* 85(4):466–477
- Jones DM, Vandepoele K (2020) Identification and evolution of gene regulatory networks: insights from comparative studies in plants. *Curr Opin Plant Biol* 54:42–48
- Jung JH, Park CM (2007) MIR166/165 genes exhibit dynamic expression patterns in regulating shoot apical meristem and floral development in Arabidopsis. *Planta* 225:1327–1338
- Jung JH, Lee S, Yun J, Lee M, Park CM (2014) The miR172 target TOE3 represses AGAMOUS expression during Arabidopsis floral patterning. *Plant Sci* 215–216:29–38
- Kang W-H, Kim S, Lee H-A, Choi D, Yeom S-I (2016) Genome-wide analysis of Dof transcription factors reveals functional characteristics during development and response to biotic stresses in pepper. *Sci Rep* 6:33332
- Kassa A, Konrad H, Geburek T (2019) Molecular diversity and gene flow within and among different subspecies of the wild olive (*Olea europaea* L.): a review. *Flora* 250:18–26
- Kim BH, Kwon Y, Lee BH, Nam KH (2014) Overexpression of miR172 suppresses the brassinosteroid signaling defects of bak1 in Arabidopsis. *Biochem Biophys Res Commun* 447:479–484
- Konishi M, Yanagisawa S (2007) Sequential activation of two Dof transcription factor gene promoters during vascular development in *Arabidopsis thaliana*. *Plant Physiol Biochem* 45(8):623–629
- Koralewski TE, Krutovsky KV (2011) Evolution of exon–intron structure and alternative splicing. *PLoS ONE* 6:e18055
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549
- Kyriakopoulou CI, Kalogianni DP (2020) Genetic identification of the wild form of olive (*Olea europaea* var. *sylvestris*) using allele-specific real-time PCR. *Foods* 9:467
- Lafka T-I, Lazou AE, Sinanoglou VJ, Lazos ES (2013) Phenolic extracts from wild olive leaves and their potential as edible oils antioxidants. *Foods* 2:18–31
- Li W, Wang T, Zhang Y, Li Y (2016) Overexpression of soybean miR172c confers tolerance to water deficit and salt stress, but increases ABA sensitivity in transgenic *Arabidopsis thaliana*. *J Exp Bot* 67:175–194
- Lijavetzky D, Carbonero P, Vicente-Carbajosa J (2003a) Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC Evol Biol* 3:17
- Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A (2020) CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* 48:D265–D268
- Lumaret R, Ouazzani N (2001) Plant genetics. Ancient wild olives in Mediterranean forests. *Nature* 413:700
- 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
- Malviya N, Gupta S, Singh V, Yadav M, Bisht N, Sarangi B, Yadav D (2015) Genome wide in silico characterization of Dof gene families of pigeonpea (*Cajanus cajan* (L.) Millsp.). *Mol Biol Rep* 42:535–552
- Mena M, Cejudo FJ, Isabel-Lamonedá I, Carbonero P (2002) A role for the DOF transcription factor BPBF in the regulation of gibberellin-responsive genes in barley aleurone. *Plant Physiol* 130:111–119
- Miyashima S, Roszak P, Sevilem I, Toyokura K, Blob B, Heo JO, Mellor N, Help-Rinta-Rahko H, Otero S, Smet W, Boekschoten M (2019) Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Nature* 565(7740):490–494
- Mohamed R, Pineda M, Aguilar M (2007) Antioxidant capacity of extracts from wild and crop plants of the Mediterranean region. *J Food Sci* 72:S059–S063
- Moore RC, Purugganan MD (2005) The evolutionary dynamics of plant duplicate genes. *Curr Opin Plant Biol* 8:122–128
- Moreno-Risueno MA, 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
- Morgan CC, Loughran NB, Walsh TA, Harrison AJ, O’Connell MJ (2010) Positive selection neighboring functionally essential sites and disease-implicated regions of mammalian reproductive proteins. *BMC Evol Biol* 10:39
- Nasim J, Malviya N, Kumar R, Yadav D (2016) Genome-wide bioinformatics analysis of Dof transcription factor gene family of chickpea and its comparative phylogenetic assessment with Arabidopsis and rice. *Plant Syst Evol* 302:1009–1026
- Negi J, Moriwaki K, Konishi M, Yokoyama R, Nakano T, Kusumi K, Hashimoto-Sugimoto M, Schroeder JI, Nishitani K, Yanagisawa S (2013) A Dof transcription factor, SCAP1, is essential for the development of functional stomata in Arabidopsis. *Curr Biol* 23:479–484
- Noguero M, Atif RM, Ochatt S, Thompson RD (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, Shiu SH (2016) Evolution of gene duplication in plants. *Plant Physiol* 171:2294–2316
- Paudel S, Magrati T, Lamichhane JR (2011) Antimicrobial activity of wild olive crude extracts in vitro. *Int J Pharma Sci Res* 2:110–113
- Peng FY, Weslake RJ (2011) Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis. *BMC Genomics* 12:286
- Ramírez-Parra E, Perianez-Rodríguez J, Navarro-Neila S, Gude I, Moreno-Risueno MA, Del Pozo JC (2017) The transcription factor OBP4 controls root growth and promotes callus formation. *New Phytol* 213:1787–1801
- Ramírez-Tejero JA, Jiménez-Ruiz J, Leyva-Pérez MD, Barroso JB, Luque F (2020) Gene expression pattern in olive tree organs (*Olea europaea* L.). *Genes* 11(5):544
- Rombauts S, Déhais P, Van Montagu M, Rouzé P (1999) PlantCARE, a plant cis-acting regulatory element database. *Nucleic Acids Res* 27:295–296
- Rymen B, Kawamura A, Schafer S, Breuer C, Iwase A, Shibata M, Ikeda M, Mitsuda N, Koncz C, Ohme-Takagi M, Matsui M, Sugimoto K (2017) ABA suppresses root hair growth via the OBP4 transcriptional regulator. *Plant Physiol* 173:1750–1762
- Samad AF, Sajad M, Nazaruddin N, Fauzi IA, Murad A, Zainal Z, Ismail I (2017) MicroRNA and transcription factor: key players in plant regulatory network. *Front Plant Sci* 8:565
- Song A, Gao T, Li P, Chen S, Guan Z, Wu D, Xin J, Fan Q, Zhao K, Chen F (2016) Transcriptome-wide identification and expression profiling of the DOF transcription factor gene family in *Chrysanthemum morifolium*. *Front Plant Sci* 7:199
- Spanudakis E, Jackson S (2014) The role of microRNAs in the control of flowering time. *J Exp Bot* 65(2):365–380
- Sun Z, Guo T, Liu Y, Liu Q, Fang Y (2015) The roles of Arabidopsis CDF2 in transcriptional and posttranscriptional regulation of primary microRNAs. *PLoS Genet* 11:e1005598
- Takano M, Haque MA, Odaira S, Nakata K, Sasaki N, 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.* <https://doi.org/10.5511/plantbiotechnology.12.1207a>
- Taylor JS, Raes J (2004) Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet* 38:615–643
- Terzi LC (2008) Regulation of flowering time by RNA processing. Springer, Berlin
- Thompson JD, Higgins DG, Gibson TJ (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
- Thompson JD, Gibson TJ, Higgins DG (2003) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protocols Bioinform* 1:2–3
- Unver T, Wu Z, Sterck L, Turktas M, Lohaus R, Li Z, Yang M, He L, Deng T, Escalante FJ (2017) Genome of wild olive and the evolution of oil biosynthesis. *Proc Natl Acad Sci* 114:E9413–E9422
- Venkatesh J, Park SW (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, Li K (2013a) Overexpression of *Arabidopsis* Dof1, GS1 and GS2 enhanced nitrogen assimilation in transgenic tobacco grown under low-nitrogen conditions. *Plant Mol Biol Report* 31:886–900
- Wang Y, Li J, Paterson AH (2013b) MCScanX-transposed: detecting transposed gene duplications based on multiple colinearity scans. *Bioinformatics* 29:1458–1460
- Wei Q, Wang W, Hu T, Hu H, Mao W, Zhu Q, Bao C (2018) Genome-wide identification and characterization of Dof transcription factors in eggplant (*Solanum melongena* L.). *PeerJ* 6:e4481
- Wellmer F, Alves-Ferreira M, Dubois A, Riechmann JL, Meyerowitz EM (2006) Genome-wide analysis of gene expression during early *Arabidopsis* flower development. *PLoS Genet* 2:e117
- Wen CL, Cheng Q, Zhao L, Mao A, Yang J, Yu S, Weng Y, Xu Y (2016) Identification and characterisation of Dof transcription factors in the cucumber genome. *Sci Rep* 6:23072
- Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS (2009) The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell* 138:750–759
- Wu J, Fu L, Yi H (2016a) Genome-wide identification of the transcription factors involved in citrus fruit ripening from the transcriptomes of a late-ripening sweet orange mutant and its wild Type. *PLoS ONE* 11:e0154330
- Wu Z, Cheng J, Cui J, Xu X, Liang G, Luo X, Chen X, Tang X, Hu K, Qin C (2016b) Genome-wide identification and expression profile of Dof transcription factor gene family in pepper (*Capsicum annuum* L.). *Front Plant Sci* 7:574
- Xu P, Cai W (2019) Nitrate-responsive OBP4-XTH9 regulatory module controls lateral root development in *Arabidopsis thaliana*. *PLoS Genet* 15:e1008465
- Xu P, Chen H, Ying L, Cai W (2016) AtDOF5.4/OBP4, a DOF transcription factor gene that negatively regulates cell cycle progression and cell expansion in *Arabidopsis thaliana*. *Sci Rep* 6:27705
- Yamashino T, Yamawaki S, Hagui E, Ishida K, Ueoka-Nakanishi H, Nakamichi N, Mizuno T (2013) Clock-controlled and FLOWERING LOCUS T (FT)-dependent photoperiodic pathway in *Lotus japonicus* II: characterization of a microRNA implicated in the control of flowering time. *Biosci Biotechnol Biochem* 77:1179–1185
- Yanagisawa S (2002) The Dof family of plant transcription factors. *Trends Plant Sci* 7:555–560
- Yanagisawa S, Schmidt RJ (1999) Diversity and similarity among recognition sequences of Dof transcription factors. *Plant J* 17:209–214
- Yanagisawa S, Akiyama A, Kisaka H, Uchimiya H, Miwa T (2004) Metabolic engineering with Dof1 transcription factor in plants: improved nitrogen assimilation and growth under low-nitrogen conditions. *Proc Natl Acad Sci* 101:7833–7838
- Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 15:496–503
- Yang X, Tuskan GA (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 J, Yang M-F, Zhang W-P, Chen F, 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 FX, Zhu GF, Wang Z, Liu HL, Huang D (2015) A putative miR172-targeted CeAPETALA2-like gene is involved in floral patterning regulation of the orchid *Cymbidium ensifolium*. *Genet Mol Res* 14:12049–12061
- Zou HF, Zhang YQ, Wei W, Chen HW, Song QX, Liu YF, Zhao MY, Wang F, Zhang BC, Lin Q (2013) The transcription factor AtDOF4.2 regulates shoot branching and seed coat formation in *Arabidopsis*. *Biochem J* 449:373–388

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.