Journal of Applied Horticulture, 28(1): 17-28, 2026
https://doi.org/10.37855/jah.2026.v28i01.3
Journal of Applied Horticulture
ISSN: 0972-1045

Genome-wide analysis of Dof transcription factors in Mediterranean olive reveals structural features, domestication-driven evolution and contribution to fruit trait variation

Inchirah Bettaïeb*, Jihen Hamdi* and Dhia Bouktila#

*Laboratory of Genetics, Biodiversity and Bioresources Exploitation (LR11ES41), Higher Institute of Biotechnology of Monastir, University of Monastir, 5000 Monastir (Tunisia). *Authors Inchirah Bettaieb and Jihen Hamdi contributed equally to this work and share first authorship. #E-mail: dhia_bouktila2000@yahoo.fr*

Abstract

The DNA-binding one zinc finger (Dof) transcription factors play essential roles in plant growth and adaptation. This study presents a comprehensive characterization of the Dof gene family in the cultivated Mediterranean olive (*Olea europaea* subsp. *europaea* var. *europaea*), focusing on structural features, evolutionary history, and functional insights based on linking transcriptomic data with morpho-biochemical profiles across three olive cultivars. We identified 76 *O. europaea* var. *europaea* Dof (OeDof) genes encoding 81 putative proteins, a larger set than previously reported in wild olive (*O. europaea* var. *sylvestris*), suggesting family gene family expansion during domestication. OeDof proteins showed considerable variation in their physicochemical properties but maintained a conserved Dof-binding domain, characterized by nine variable regions that may modulate target site specificity. Fifteen conserved motifs were identified, including subgroup-specific motifs likely contributing to functional specialization. Several motifs were annotated, suggesting OeDof involvement in olive fruit metabolic processes. Phylogenetic analysis classified OeDof genes into four groups and nine subgroups, revealing orthologous relationships with Arabidopsis Dof genes. Duplication analysis identified 27 paralogous gene pairs, mostly under purifying selection (Ka/Ks < 1), with several gene pairs showing relaxed selection pressure (0.5 < Ka/Ks < 1), suggesting that, while selective pressure maintains protein function, a relatively high tolerance for nonsynonymous substitutions (*i.e* . functional divergence) exists. A unique gene pair (*OeDof19/OeDof49*) underwent positive selection, highlighting its potential role in key biological processes. RNA-Seq analysis classified OeDof genes into six distinct expression patterns across the fruit of three olive cultivars with contrasting morpho-biochemical traits, identifying candidate genes for further functional validation. This genome-wide analysis offers valuable insights into the domestication-linked evolutionary and functional dynamics of Dof genes in Mediterranean olive, laying the foundation for their application in genomics-assisted breeding for improved fruit traits and oil quality.

Key words: *Olea europaea*, Dof transcription factor family, In silico analysis, phylogenetic analysis, expression profiling, candidate genes

Introduction

The Mediterranean olive tree (*Olea europaea* subsp. *europaea* var. *europaea*), a perennial diploid tree of the *Oleaceae* family, was domesticated around 6000 years ago in Asia Minor and has a genome size of about 1.38 Gbp (Unver *et al*., 2017; Cruz *et al*., 2016). It is a major agricultural crop in the Mediterranean region because it is a source of olive oil (Hamed *et al*., 2025). Olive oil is rich in unsaturated fatty acids and other important secondary metabolites. It also contains more than 30 different phenolic compounds (Gouvinhas *et al*., 2017) that are strong antioxidants and free radical scavengers (Ghanbari *et al*., 2012). In fact, olive oil has great benefits for human health (El and Karakaya, 2009), especially for cardiovascular disease. Given the economic and nutritional importance of olives, it is critical to improve some features of this tree, such as disease resistance and biotic interaction (Valverde *et al*., 2023), agronomic traits (Barone *et al*., 2014), yield performance (Barone *et al*., 2014) and oil quality (García-Rodríguez *et al*., 2017). In particular, transcription factor (TF) genes represent an important group targeted for crop improvement because they are involved in the control of many traits of agronomic importance.

The DNA-binding one zinc finger (Dof) gene family is a family of plant-specific transcription factors that is widespread in higher plants. This family plays a key role in many biological processes, such as metabolism regulation, phytohormone signaling, seed development, cell differentiation and abiotic and biotic stress responses (Noguero *et al*., 2013; Gupta *et al*., 2015). Dof proteins generally contain two major functional domains (Gupta *et al*., 2015). One is the DNA-binding one zinc finger (Dof) domain (PF02701) in the N-terminal region, which consists of 50–52 amino acid residues and contains a highly conserved single zinc-finger structure formed by the $CX_2CX_{21}CX_2C$ motif. The second domain is a regulatory C-terminal domain.

Despite the availability, since 2016, of the whole genome assembly of the cultivated Mediterranean olive tree, *Olea europaea* subsp. *europaea* var. *europaea* (Cruz *et al*., 2016), Dof gene family remains largely uncharacterized in this botanical variety. Genome-wide investigations of important regulatory gene families, such as the Dof family, in the domesticated form olive tree will help researchers and breeders identify gene family members, gain a better understanding of the mechanisms by which these genes control metabolic pathways and make use of selected

functional candidates, especially those involved in fruit traits and oil content and quality. Genetic variations that have arisen during domestication bottlenecks can be identified by comparing the gene families involved in metabolism in wild and domestic olive trees, as domestication involves accelerated selection for traits that are beneficial to human cultivation, such as increased yield, fruit size and biochemical content. This anthropic selective pressure also often leads to changes in gene expression patterns and regulation pathways (Breton *et al*., 2009). In this study, we performed genome-wide identification of the Dof gene family members in cultivated olive and comprehensively predicted their physicochemical features, phylogenetic relationships, structural motifs, selection modes and expression patterns in three olive cultivars 'Arbequina', 'Frantoio Selection' and 'Nikitskii', which have previously been identified as having significant variations in fatty acid metabolism and flavonoid concentration in olive fruits (Niu *et al*., 2022). Our findings shed light on the molecular evolution of OeDofs and their functional properties, laying the groundwork for *O. europaea* breeding programs aimed at genetically increasing olive oil quality by Dof candidate gene manipulations.

Materials and methods

Database searches for Dof family members in the domesticated form *Olea europaea*: The olive tree gene information for the entire genome was retrieved from (GIGA)n DB (http://gigadb.org/dataset/100201). This information contains the genome assembly OE6 produced from the leaves of the Spanish millennial olive tree, Santander, belonging to *Olea europaea* subsp. *europaea* variety *europaea* (Cutivar 'Farga') (Bioproject submitted by the Centro Nacional de Análisis Genómico, Spain, and registered at Genbank under accession PRJEB4992) (Cruz *et al*., 2016). Proteins featuring the Dof domain were retrieved from InterProScan annotations of 79.910 *O. europaea* var. *europaea* protein products using the Pfam ID PF02701. The Dof protein sequences of *Arabidopsis thaliana* were downloaded from the iTAK - Plant Transcription Factor & Protein Kinase Identifier and Classifier (http://itak.feilab.net/cgi-bin/itak/index.cgi). All amino acid sequences obtained were further checked for the presence of a highly conserved Dof domain at the N-terminus using the Pfam database (https://pfam.xfam.org/) and the National Center for Biotechnology Information (NCBI) Conserved Domain Database (CDD) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Additionally, for *O. europaea*, Dof protein-coding sequences (CDSs) were also downloaded from GigaDB (http://gigadb.org/dataset/100201) for use in selection pressure analyses.

Analysis of protein features and subcellular localization: The physicochemical properties of the OeDof amino acid sequences were determined using the Expert Protein Analysis System (ExPASy) program ExPASy (http://web.expasy.org/protparam/) and EMBOSS Pepstats (https://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/). In addition, the subcellular localization of OeDof proteins was predicted using the WoLF PSORT tool (https://wolfpsort.hgc.jp/) and further validated with DeepLoc 2.0 tool (https://services.healthtech.dtu.dk/services/DeepLoc-2.0/).

Multiple sequence alignment, phylogenetic analysis and identification of conserved motifs: Multiple sequence alignments were performed on the Dof amino acid sequences using ClustalW, implemented in MEGA5.0 software (Tamura *et al*., 2011) with default settings. Subsequently, a maximum likelihood (mL) phylogenetic tree was generated based on the resulting alignment using 1,000 bootstrap replications and the Jones–Taylor–Thornton (JTT) model. Furthermore, conserved motifs of each deduced OeDof protein were identified using the MEME online program (http://meme-suite.org) with the following parameter settings: motif length = 20–50 residues; maximum number of motifs = 15; and the distribution of one single motif was "any number of repetitions".

The Ka/Ks ratios and duplication dates of duplicate Dof genes in *Olea europaea* var. *europaea*: The Ka and Ks values were calculated using the Nei–Gojobori approach implemented in MEGA 5.0. Next, the date of duplication (million years ago, MYA) was calculated using a synonymous mutation rate of one substitution per synonymous site per year as $T = Ks/2\lambda$, where $\lambda = 1.8\times10^{-9}$ (Barghini *et al*., 2015).

***In silico* expression profiling of *Olea europaea* var. *europaea* Dof genes**:

The RNA-Seq Sequence Read Archive (SRA) datasets for three olive cultivars analyzed by Niu *et al*. (2022) were retrieved from the European Nucleotide Archive (ENA; www.ebi.ac.uk/ena) in FASTQ format. HiSat2 Galaxy Version 2.2.1+galaxy0, available online through Galaxy Europe at https://usegalaxy.eu/, was utilized to align adapter-free RNA-seq reads to the domesticated olive genome. Employing Cufflinks, accessible through Galaxy Europe at https://usegalaxy.eu/, we quantified mapped reads per annotated transcript, using "Fragments per Kilobase of exon model per Million mapped fragments" (FPKM). Using the web tool heatmap2 implemented in Galaxy Version 3.1.1+galaxy1 (https://usegalaxy.eu/), the FPKM expression values were normalized with z-score (gene expression value in gene of interest - mean expression across all genes / standard deviation), and then applied to create a heat map with Euclidian distance and average linkage methods, for distance measurement and clustering, respectively.

Results and discussion

Dof transcription factor family identification and structure analysis: A total of 77 *O. europaea* and 36 *A. thaliana* Dof genes were initially identified. After validating the integrity of the Dof domain using the Pfam database, 76 *O. europaea* Dof genes encoding 81 proteins, including isoforms, and 36 *A. thaliana* Dof genes encoding 47 proteins were retained.

Multiple sequence alignment revealed that the zinc finger domain among OeDof proteins was characterized by four highly conserved cysteine residues and a general $CX_2CX_{21}CX_2C$ structure. In addition to the four highly conserved cysteine residues, 12 other residues were found to be perfectly (100%) conserved in the Dof domains of all 81 OeDofs proteins (Fig. 1). These highly conserved residues were nearly identical to those of the Dof domains of other plants, such as cassava, pepper, and sugar beet (Wu *et al*., 2016; Zou *et al*., 2019; Hamdi *et al*., 2021). However, the sharp amino acid structure revealed at least nine variable motifs or residues among the Dof domains of the Mediterranean olive (Fig. 1): (I) The first motif is N/K/H/A/G/D/E/Q residue, which follows the conserved CPRC sequence; (II) The second is the di-amino acid TN/ SN/ LN/ PN/ TK/ LD/

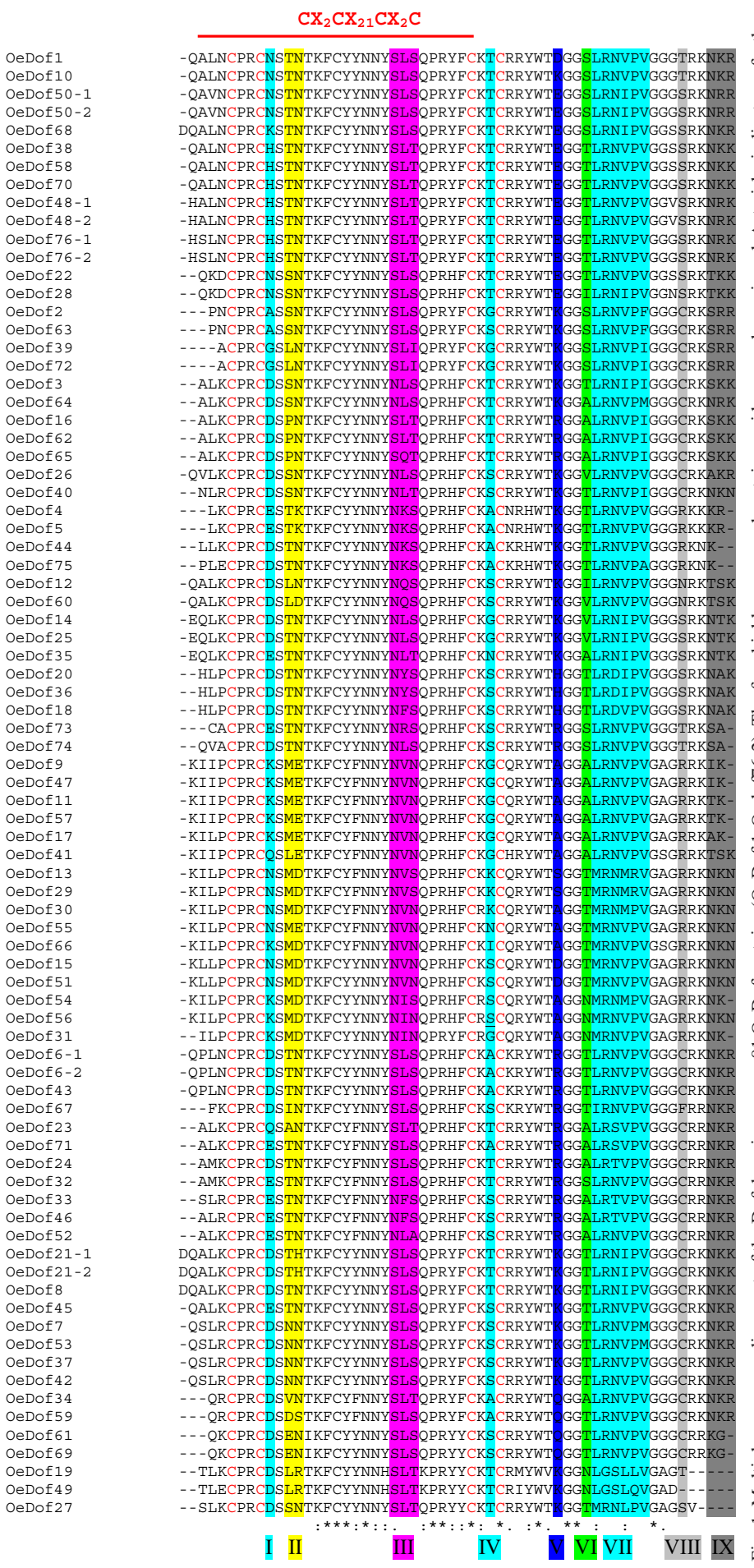

Fig. 1. Multiple sequence alignment of the Dof domains among 81 OeDof proteins (OeDof 1-Oedof76-2). The four highly conserved cysteine residues are shown in red. Asterisks indicate perfectly conserved residues. The dots represent variation in residues within strongly conserved groups, whereas colons represent variation in residues within weakly conserved groups. Colored columns (I–IX) correspond to variable regions identified in the present study, within the OeDof domain.

ME/ LE/ MD/ IN/ AN/ TH/ NN/ VN/ DS/ EN/ LR; (III) The third is the tri-amino acid SLS/ SLT/ SLI/ NLS/ SQT/ NLT/ NKS/ NQS/ NYS/ NFS/ NRS/ NVN/ NVS/ NIS/ NIN/ NLA, which comes after FCYY/FNNY/H, the middle conserved domain containing middle C residue; (IV) the fourth variable region is the residue S/N/ T/A/G/K/I, which is inside the second C-C bridge; ; (V) the fifth variable region is represented by the residue D/S/E/K/R/H/A/Q, which follows the second C-C bridge; (VI) the sixth variable region is represented by the residue S/T/I/A/ V/N, in the C-terminus of the Dof domain; (VII) the seventh variable region is represented by the hexa-amino acid LRNVPV/ LRNIPV/ LRNVPF/ LRNVPI/ LRNIPI/ LRNVPM/ LRNVPA/ LRDIPV/ LRDVPV/ MRNMRV/ MRNMPV/ MRNVPV/ IRNVPV/ LRSVPV/ LRTVPV/ LGSLLV/ LGSLQV/ MRNLP, in the C-terminus of Dof domain; (VIII) the eighth variable region is represented by the residue T/S/C/R/N/F, and (IX) the ninth variable region is represented by the tri-amino acid NKK/ NKR/ NRR/ NRK/ TKK/ SRR/ SKK/ AKR/ NKN/ KR-/ K--/ TSK/ NTK/ NAK/ SA-/ IK-/ TK-/ AK-/ NK-/ KG-, in the C-terminus of the Dof domain. These variable structural patterns within the Dof domain may aid in determining the appropriate target site for a specific Dof protein, since specific TF protein domains can be associated with a wide range of downstream gene functional characteristics.

Comparative Analysis of *Dof* gene family size in *Oleaceae* and non-*Oleaceae* plant genomes: The number of OeDof genes (76) was greater than in non-*Oleaceae* species, namely *Arabidopsis* (36 AtDofs), *Beta vulgaris* (21 BvDofs) (Hamdi *et al*., 2021), and *Brachypodium distachyon* (27 BdDofs) (Hernando-Amado *et al*., 2012) (Table 1). The olive tree's comparatively higher number of Dof genes compared to other plants is attributable to at least three whole-genome duplication events

that affected its evolutionary history, particularly considering the most recent whole-genome duplication (WGD) event specific to the *O. europaea* lineage, which would have occurred around 10 million years ago (Mya), as suggested by Julca *et al*. (2018). This WGD event likely provided the genomic context for the expansion of gene families, including the Dof gene family, in the olive tree lineage. On the other hand, the average number of Dof genes per Mb (density) was greater in Arabidopsis than all species including *O. europaea* var. *europaea* and *O. europaea* var. *sylvestris* (Table 1), which is explainable by the relatively small genome size of the of *A. thaliana*, probably due to both large and small-scale deletions in the Arabidopsis genome (Oyama *et al*., 2008).

Comparing the number of Dof genes between domesticated (*O. europaea* var. *europaea*) and wild (*O. europaea* var. *sylvestris*) olive trees revealed that the Dof family in *O. europaea* var. *europaea* (76 Dof genes reported in the current study based on OE6 genome assembly) is larger than the size previously reported in the wild olive tree *O. europaea* var. *sylvestris* (51 Dof genes; Mariyam Shafiq *et al*., 2021), despite the fact that the cultivated olive variety has a slightly smaller genome than the wild olive tree, while having more Dof gene family members (Table 1). This observation suggests gene expansion within specific gene families through selective pressures during the domestication and breeding processes of the olive tree. The process of plant domestication, even though relatively recent, can lead to the expansion of certain gene families. A number of recent pangenome analyses have revealed significant gene presence/absence variation within a species, with gene retention or loss being influenced by selection throughout domestication and breeding. For example, comparisons of rice cultivars and wild rice species showed that the *Submergence1* (*Sub1A*) haplotype involved in submergence tolerance appeared recently, after rice domestication (Fukao *et al*., 2009). Besides, Zhao *et al*. (2018) showed that gene content increased in rice with a median of 1,171 additional genes in domesticated Japanese accessions compared to wild accessions. Furthermore, in pepper (*Capsicum annuum*), which was domesticated from wild chiltepin pepper (*Capsicum annuum* var. *glabriusculum*) approximately 8,000 years ago (Lopez-Moreno *et al*., 2023), comparisons between wild and domestic species showed that tandemly-duplicated genes involved in capsaicin biosynthesis likely contributed to the diversification of pepper pungency (Qin *et al*., 2014), which may be the basis for variation in pungency among domesticated pepper varieties. As a result, it has been established that genes related to stress response and plant agronomic processes are positively selected in cultivars during domestication (Zhao *et al*., 2018), and the Dof gene family may be among these genes.

Physicochemical properties of OeDof transcription factors: The physical and chemical characteristics of OeDof proteins are presented in Table 2. The average molecular weight of the OeDof proteins was 96.89 kDa with polypeptide sizes ranging from 144 aa (OeDof41) to 857 aa (OeDof1), with an average of 313 aa. The predicted isoelectric point (pI) ranged from 4.73 (OeDof2) to 9.97 (OeDof27), with an average pI of 8.82 for all proteins. Twenty-one (21) of the 81 identified OeDof proteins had an acidic pI (< 7.0), while the remaining 60 proteins (74%) had basic pI values (> 7.0). Proteins with a basic pI are positively charged at physiological pH, facilitating strong DNA binding essential for transcriptional regulation, while about 26% of OeDof proteins may favor protein–protein interactions. The GRAVY was less than zero for all OeDof proteins, indicating the hydrophilic nature of these proteins. This hydrophilic feature of OeDof proteins is generally expected, because it facilitates their movement in the cytoplasm (before entering the nucleus) and the nucleoplasm. All OeDofs had a greater percentage of aliphatic amino acids (roughly twofold) than aromatic amino acids, which suggested that these proteins were high in aliphatic amino acids. The proportion of positively charged amino acids was slightly greater than that of negatively charged amino acids (12.71% and 10.61%, respectively). Finally, the subcellular localization analysis showed that most OeDof proteins (91%) are nuclear, confirming their role in transcriptional regulation, while a few are targeted to chloroplasts (six OeDof genes) and mitochondria (one OeDof gene), as observed in other species such as *Cleistogenes songorica* and *Stevia rebaudiana* (Wang *et al*., 2021), suggesting possible regulatory functions in organelle-specific processes such as photosynthesis and respiration. The broad physicochemical diversity within this gene family may be related to the functional diversity and evolutionary dynamics of these genes, providing a valuable window into the complex interplay between genotype, phenotype and environment, as well as insights into the mechanisms driving plant adaptation and diversity.

Table 1. Dof gene family sizes and their genomic proportions in different genomes.

	Dicots				Monocots
	Oleaceae		*Amaranthaceae*	*Brassicaceae*	*Poaceae*
	Mediterranean olive *(Olea europaea* Subsp. *europaea* var. *europaea)* (Cultivated form)	Mediterranean olive *(Olea europaea* Subsp. *europaea* var. *sylvestris)* (Wild form)	*Beta vulgaris* (Sugar beet)	*Arabidopsis thaliana* (Thale cress)	*Brachypodium distachyon*
Total Dof genes	76 (a)	51(a)	21 (a)	36 (a)	27 (a)
Total protein-coding genes in genome	56,349 (b)	50,684 (c)	27,421 (d)	27,416 (e)	34,310 (f)
Dof transcription factor genes (%)	0.134	0.100	0.076	0.131	0.078
Genome size (Mb)	1 380 (b)	1 460 (c)	758 (d)	135 (e)	355 (f)
Average number of Dof genes per Mb	0.055	0.034	0.027	0.266	0.076

(a) *Olea europaea* subsp. europaea var. europaea data from current study; *Olea europaea* subsp. europaea var. sylvestris data from Mariyam Shafiq *et al*. (2021); Beta vulgaris data from Hamdi *et al*. (2021); Arabidopsis thaliana data from current study; and Brachypodium distachyon data from Hernando-Amado *et al*. (2012). (b): Cruz *et al*. (2016). (c): Unver *et al*. (2017). (d) : Dohm *et al*. (2014). (e): Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). (f): Ensembl Plants (https://plants.ensembl.org/Brachypodium_distachyon/Info/Annotation/#assembly).

Table 2. Basic information of the 81 putative Dof family proteins studied in *Olea europaea* subsp. europaea var. europaea

Protein ID	Systematic protein name	Protein length (aa)	Aliphatic (%)	Aromatic (%)	Negativey charged residues (Asp+Glu) (%)	Positively charged residues (Arg+Lys) (%)	Theoretical isoelectic point (Pi)	Molecular weight (Da)	Grand average of hydropathicity (GRAVY)	Subcellular localization
OE6A003769P1	OeDof1	857	23.921	10.852	10.618	12.719	8.82	96893.64	-0.485	Nucleus
OE6A004082P1	OeDof2	277	21.661	11.552	13.357	9.025	4.73	30946.44	-0.555	Chloroplast
OE6A005199P1	OeDof3	323	18.266	9.288	7.120	8.979	8.78	35059.05	-0.622	Nucleus
OE6A008023P1	OeDof4	238	15.546	10.084	8.824	9.244	7.56	26125.91	-0.741	Nucleus
OE6A009098P1	OeDof5	238	16.387	9.664	9.244	9.664	7.56	26089.92	-0.74	Nucleus
OE6A009231P1	OeDof6_1	333	18.919	9.61	6.307	8.108	8.89	36818.96	-0.679	Nucleus
OE6A009231P2	OeDof6_2	335	19.104	9.552	6.269	8.060	8.89	37046.22	-0.673	Nucleus
OE6A009236P1	OeDof7	257	16.732	12.062	9.338	8.913	8.09	28770.95	-0.835	Nucleus
OE6A012425P1	OeDof8	308	15.909	11.364	10.390	8.117	5.69	34347.08	-0.721	Nucleus
OE6A013516P1	OeDof9	160	18.125	13.750	11.875	16.25	9.15	18266.63	-0.866	Nucleus
OE6A013813P1	OeDof10	275	16.727	12	9.091	10.182	8.46	30549.09	-0.748	Nucleus
OE6A016186P1	OeDof11	160	18.75	10.625	10.000	15.000	9.2	18067.48	-0.81	Chloroplast
OE6A017479P1	OeDof12	268	19.776	11.94	10.821	10.821	7.03	29524.72	-0.698	Nucleus
OE6A017875P1	OeDof13	518	21.663	8.511	11.412	11 218	6.89	56463.87	-0.734	Nucleus
OE6A020362P1	OeDof14	241	14.938	10.373	8.713	8.714	6.88	26375.9	-0.891	Nucleus
OE6A021342P1	OeDof15	496	19.556	7.258	6.452	11.492	6.48	54158.47	-0.801	Nucleus
OE6A028156P1	OeDof16	341	19.412	10	8.235	9.412	8.55	37054.19	-0.629	Nucleus
OE6A028175P1	OeDof17	162	19.753	11.728	10.494	13.580	8.8	18138.47	-0.764	Chloroplast
OE6A029093P1	OeDof18	230	20.87	12.609	7.392	7.826	7.65	24240.03	-0.386	Nucleus
OE6A029570P1	OeDof19	342	19.883	10.234	5.555	10.526	9.43	37093.69	-0.518	Nucleus
OE6A029845P1	OeDof20	233	18.884	13.305	6.867	7.726	8.13	24609.44	-0.389	Nucleus
OE6A029854P1	OeDof21-1	308	16.558	13.312	9.740	7.792	6.13	34483.2	-0.734	Nucleus
OE6A029854P2	OeDof21-2	303	16.832	13.201	9.901	7.920	6.14	33783.36	-0.761	Nucleus
OE6A030482P1	OeDof22	255	12.941	12.941	9.412	11.373	8.87	28542.55	-0.886	Nucleus
OE6A030707P1	OeDof23	350	18	9.429	5.142	7.714	9.26	37261.21	-0.607	Chloroplast
OE6A030884P1	OeDof24	341	19.062	9.677	4.985	7.331	9.3	36347.48	-0.515	Nucleus
OE6A031703P1	OeDof25	243	14.815	10.288	8.642	8.642	6.88	26604.11	-0.912	Nucleus
OE6A032821P1	OeDof26	327	17.737	11.315	8.563	8.257	6.75	35174.49	-0.725	Nucleus
OE6A032828P1	OeDof27	222	16.667	10.811	5.856	15.766	9.97	24771.54	-1.051	Nucleus
OE6A037315P1	OeDof28	221	15.385	10.407	7.240	10.860	9.26	24481.29	-0.787	Nucleus
OE6A038851P1	OeDof29	517	21.857	8.317	11.412	11.218	6.86	56408.71	-0.735	Nucleus
OE6A039076P1	OeDof30	526	20.342	8.935	11.787	10.646	6.18	57239.35	-0.771	Nucleus
OE6A039274P1	OeDof31	494	22.02	8.687	13.738	11.515	5.49	54527	-0.712	Nucleus
OE6A039684P1	OeDof32	341	19.941	8.798	4.986	7.331	9.34	36223.35	-0.481	Nucleus
OE6A040492P1	OeDof33	333	18.619	9.309	6.006	8.108	9.23	36064.04	-0.612	Nucleus
OE6A043222P1	OeDof34	274	17.883	9.124	4.380	8.394	9.77	29754.16	-0.756	Nucleus
OE6A044457P1	OeDof35	244	20.492	9.836	8.197	9.016	8.03	26600.31	-0.729	Nucleus
OE6A044794P1	OeDof36	233	18.884	13.305	6.867	7.726	8.13	24581.39	-0.399	Nucleus
OE6A047040P1	OeDof37	257	14.008	11.673	9.338	10.506	8.39	28934.05	-0.929	Nucleus
OE6A049521P1	OeDof38	257	19.066	11.284	8.560	9.727	8.44	28467.52	-0.705	Nucleus
OE6A052002P1	OeDof39	266	22.556	10.526	9.022	8.646	6.43	29114.77	-0.379	Nucleus
OE6A053383P1	OeDof40	306	20.588	9.15	6.209	7.517	8.65	32753.43	-0.468	Nucleus
OE6A057303P1	OeDof41	144	21.528	9.722	9.723	16.667	9.46	16330.65	-0.823	Mitochondrion
OE6A063670P1	OeDof42	257	14.008	11.673	9.338	10.506	8.39	28934.05	-0.929	Nucleus
OE6A067414P1	OeDof43	335	20	9.254	6.269	7.762	8.77	36780.94	-0.646	Nucleus
OE6A067535P1	OeDof44	228	18.421	10.526	8.333	9.211	8.19	24812.43	-0.668	Nucleus
OE6A068158P1	OeDof45	289	17.647	12.111	7.612	8.996	8.48	31326.85	-0.624	Nucleus
OE6A069998P1	OeDof46	333	18.619	9.61	6.006	7.808	9.1	36100.01	-0.602	Nucleus
OE6A074022P1	OeDof47	160	20	13.75	11.875	15.625	9.04	18342.77	-0.794	Nucleus
OE6A075838P1	OeDof48-1	255	17.647	12.549	8.235	10.980	9.12	28211.44	-0.715	Nucleus
OE6A075838P2	OeDof48-2	270	18.148	12.222	8.518	10.741	8.99	29858.27	-0.702	Nucleus
OE6A078176P1	OeDof49	343	18.95	10.204	6.122	11.079	9.43	37513.09	-0.617	Nucleus
OE6A080432P1	OeDof50-1	280	17.857	11.429	8.214	9.286	8.48	30926.26	-0.726	Nucleus
OE6A080432P2	OeDof50-2	287	17.77	11.498	8.711	9.408	8.17	31823.26	-0.751	Nucleus
OE6A082367P1	OeDof51	502	20.518	7.371	13.745	11.354	5.3	54748.14	-0.761	Nucleus
OE6A084000P1	OeDof52	357	20.728	8.683	6.162	8.963	9.41	38800.5	-0.572	Nucleus
OE6A084922P1	OeDof53	257	16.342	12.062	9.338	10.117	8.09	28754.91	-0.856	Nucleus
OE6A085395P1	OeDof54	544	21.691	10.11	11.765	12.500	8.2	60382.17	-0.714	Nucleus
OE6A085809P1	OeDof55	489	19.632	8.589	12.679	10.838	5.74	53247.05	-0.826	Nucleus
OE6A086431P1	OeDof56	494	22.47	8.704	13.360	10.931	5.4	54432.98	-0.7	Nucleus

Table 2 continued

Protein ID	Systematic protein name	Protein length (aa)	Aliphatic (%)	Aromatic (%)	Negativey charged residues (Asp+Glu) (%)	Positively charged residues (Arg+Lys) (%)	Theoretical isoelectic point (Pi)	Molecular weight (Da)	Grand average of hydropathicity (GRAVY)	Subcellular localization
OE6A087193P1	OeDof57	160	19.375	10.625	10.000	15.000	9.2	18049.44	-0.798	Chloroplast
OE6A087330P1	OeDof58	257	19.066	11.284	8.56	9.727	8.44	28467.52	-0.705	Nucleus
OE6A088921P1	OeDof59	267	17.978	8.989	4.869	9.738	9.93	28944.27	-0.805	Nucleus
OE6A089438P1	OeDof60	269	19.703	10.781	10.781	10.037	6.21	29260.16	-0.715	Nucleus
OE6A089500P1	OeDof61	266	16.917	8.647	6.391	8.271	8.94	29049.91	-0.921	Nucleus
OE6A092064P1	OeDof62	340	19.412	10	8.235	9.412	8.55	37054.19	-0.629	Nucleus
OE6A094590P1	OeDof63	431	27.378	11.137	11.136	8.584	5.14	47904.38	-0.192	Nucleus
OE6A094591P1	OeDof64	322	17.391	10.248	7.453	9.317	8.77	35256.31	-0.669	Nucleus
OE6A097466P1	OeDof65	341	19.941	10.557	8.211	10.851	9.12	37581.96	-0.663	Nucleus
OE6A104771P1	OeDof66	418	22.727	9.569	9.330	11.244	8.9	45539.07	-0.545	Nucleus
OE6A106374P1	OeDof67	269	21.561	10.781	5.204	8.550	9.28	29398.94	-0.465	Nucleus
OE6A107309P1	OeDof68	269	17.472	9.294	10.781	10.781	6.73	29649.01	-0.723	Nucleus
OE6A107432P1	OeDof69	266	17.293	8.647	6.391	7.895	8.76	29006.89	-0.89	Nucleus
OE6A107773P1	OeDof70	267	18.727	11.236	8.240	9.738	8.64	29625.91	-0.678	Nucleus
OE6A109933P1	OeDof71	334	19.461	9.581	4.790	7.785	9.31	35838.04	-0.555	Nucleus
OE6A111461P1	OeDof72	266	22.556	9.774	8.646	9.022	7.52	28940.54	-0.392	Chloroplast
OE6A112039P1	OeDof73	251	20.717	8.765	9.164	7.968	6.07	26861.82	-0.462	Nucleus
OE6A115677P1	OeDof74	246	19.919	8.537	8.537	8.131	6.42	25814.42	-0.511	Nucleus
OE6A116930P1	OeDof75	234	16.667	10.684	8.120	8.975	8.26	25695.18	-0.824	Nucleus
OE6A119867P1	OeDof76-1	251	17.131	12.351	8.367	11.156	9.06	27818.98	-0.722	Nucleus
OE6A119867P2	OeDof76-2	266	17.669	12.03	8.646	10.902	8.93	29467.85	-0.708	Nucleus
Mean	-	313.636	23.921	10.852	10.618	12.719	8.820	96893.640	-0.485	

Phylogenetic relationships of OeDof genes: An unrooted ML phylogenetic tree was built using 81 OeDof proteins from *O. europaea* and 47 from Arabidopsis. As shown in Fig. 2, the phylogenetic tree divided all Dof genes into four groups (A, B, C, and D) and nine subgroups (A, B1, B2, C1, C2.1, C2.2, C3, D1, and D2), in accordance with Lijavetzky *et al*. (2003)'s classification of the *A. thaliana* Dof family. Among all subgroups, subgroup D1 was the largest, including 16 members of OeDof proteins, followed by subgroups B2 and C2.1, which had 14 members each. Subgroups D2, C1 and A were composed of 13, 8 and 6 members, respectively. Subgroups B1 and C2.2 each had 4 members. Finally, subgroup C3 was the smallest, containing only two members (Fig. 2).

The phylogenetic classification of Dof genes in the domesticated olive, *O. europaea* var. *europaea*, exhibited minor differences with the previously reported wild olive (*O. europaea* var. *sylvestris*) (Mariyam Shafiq *et al*., 2021), in terms of the number of subgroups as well as the number of members within each subgroup. In the wild olive genome, there were 8 subgroups A, B1, B2, C1, C2.1, C2.2, D1 and D2 with 3, 6, 12, 4, 10, 4, 9 and 3 Dof members, respectively (Mariyam Shafiq *et al*., 2021). The wild olive genome lacked the smallest Dof subgroup discovered in Arabidopsis and in the current study on the domesticated olive, namely C3 (which has just two members). Furthermore, when compared to the wild olive, almost all Dof subgroups of the domesticated olive expanded in members' number, likely providing adaptive advantages and agronomic features to the farmed variety.

On the functional side, the current phylogenetic comparative analysis involving *O. europaea* with the model angiosperm *A. thaliana* may serve as a phylogeny-based functional prediction, as orthologs between Dof genes of *O. europaea* and *A. thaliana* may be helpful for inferring putative functions of OeDof genes because orthologs often retain the same function (Tatusov *et al*., 1997). For example, *OeDof21*, encoding the two isoforms OeDof21-1 and OeDof21-2, shares evolutionary history with *AtDof34* (At5g62940) (Fig. 2), which has been reported to play specific roles related to lipid synthesis (Wang *et al*., 2007). Additional experimental studies are required to confirm whether *OeDof21* contributes through regulatory functions to lipid metabolism in the olive tree, as observed in its Arabidopsis counterpart.

Motif composition of OeDof proteins: A total of 15 conserved motifs were identified from all the OeDof proteins using the MEME tool (Table 3). As illustrated in Fig. 3, members who were clustered into the same phylogenetic group or subgroup had similar motif compositions, providing further confirmation of the evolutionary relationships among the OeDof genes. Motif 1, which is widely present in all the OeDof TFs, was annotated as the Dof domain. Multiple additional conserved motifs were shared across different groups, such as motif 7, which was found in groups D and C, and motifs 10, 11, 13 and 14, which were found in groups B and D. However, some other conserved motifs appeared exclusively in specific groups or subgroups, implying that these group- or subgroup-specific motifs may determine the specific functions of the members in these groups or subgroups. For example, motifs 6 and 8 were only identified in group C (C1 and C2.1); motifs 2, 3, 4, 5 and 9 were only present in subgroup D1; motif 12 was found only in subgroup C2.1; and motif 15 was detected only in subgroup C1. Strikingly, group A and subgroup C3 lack any motif in addition to the Dof domain, which indicates that their OeDofs might have fewer functions than the other members of the Dof family in *O. europaea*. Among all OeDof proteins, those from subgroup D1 exhibited the most complex motif composition (5 specific motifs), which has also been observed in the same subgroup (D1) in the wild olive tree (Mariyam Shafiq *et al*., 2021).

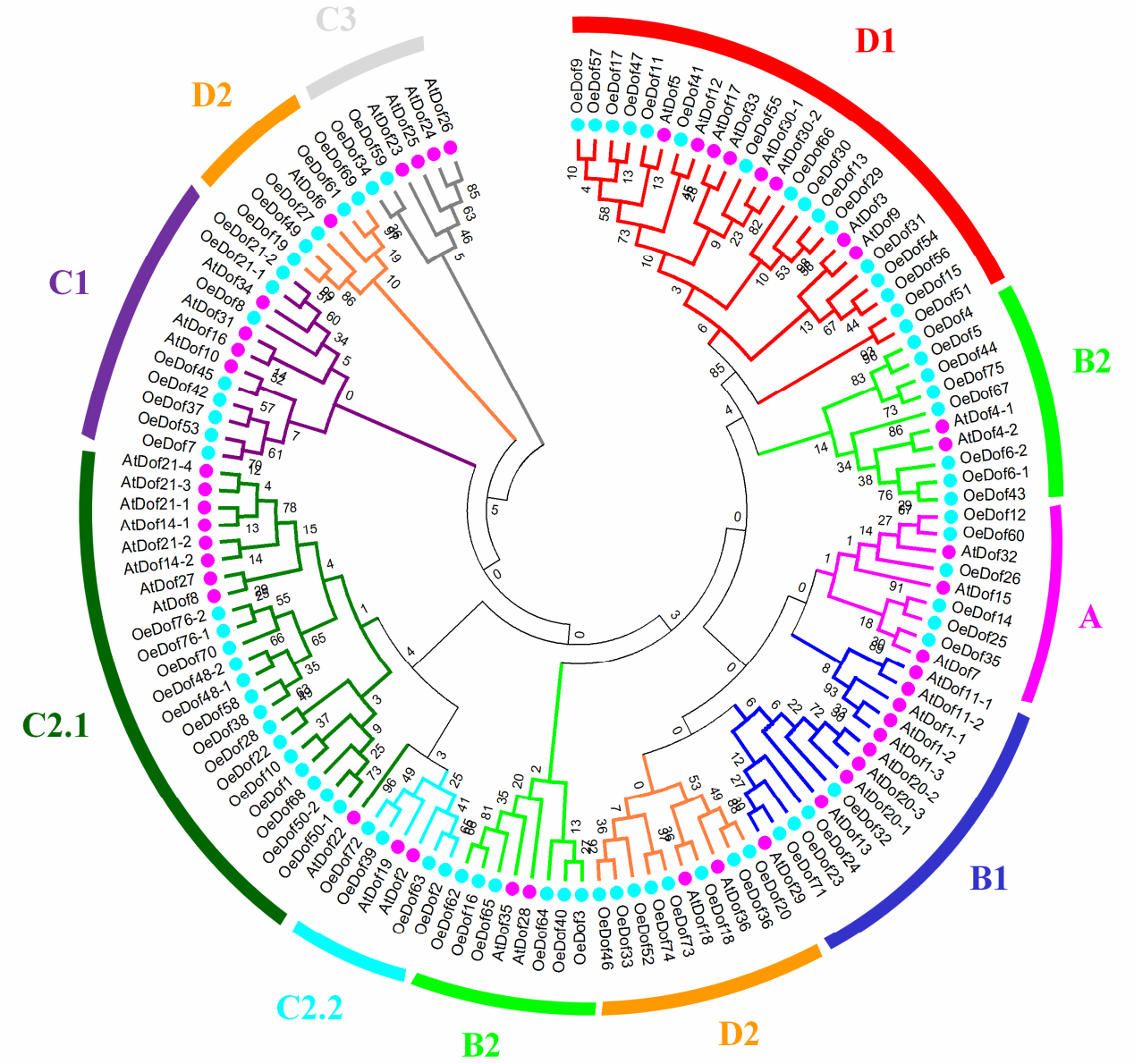

Fig. 2. Unrooted phylogenetic tree representing the relationships among Dof proteins of *Olea europaea* (blue circles) and Arabidopsis thaliana (pink circles). The differently-colored arcs represent different subgroups of Dof proteins designated based on reference phylogeny in A. thaliana (Lijavetzky *et al*., 2003).

Two motifs could be annotated in addition to motif 1 representing the Dof domain found in the N-terminal conserved region. Motif 2 was predicted to the GI-binding domain, while motif 9 represented the FKF1-binding domain, both positioned at the C-terminus (Jia *et al*., 2019). Both motifs were previously identified in Dof transcription factors from other plants, such as *Glycine soja*, *Medicago truncatula*, *Sorghum bicolor*, and *Theobroma cacao* (Jia *et al*., 2019). The GI-binding domain plays a role in controlling the circadian clock, particularly by regulating the stability of key clock proteins such as TIMING OF CAB EXPRESSION1 (TOC1) and PSEUDO RESPONSE REGULATOR5 (PRR5) (Ito *et al*., 2012). Kim *et al*. (2023) reported that certain genes involved in lipid metabolism in Arabidopsis are subject to circadian control, in particular the central clock regulators *LATE ELONGATED HYPOCOTYL* (*LHY*) and *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*). This suggests an involvement of motif 2 in controlling metabolic processes, further emphasizing the interconnectedness of circadian rhythms and metabolic pathways in plants, since the clock influences the expression of genes involved in lipid metabolism (Kim *et al*., 2023). On the other hand, reduced expression of *SlFKF1* in tomato leads to delayed flowering, altered leaf architecture, and lower lycopene accumulation, highlighting its crucial role in coordinating developmental processes, fruit quality, and metabolic regulation (Shibuya *et al*., 2021). Interestingly, motifs 2 and 9 were jointly present in the C-terminus of 10 OeDofs belonging only to subgroup D1 (OeDofs 13, 15, 29, 30, 31, 51, 55, 56, 57 and 66), which makes this group interesting for studying circadian regulation, reproductive function, as well as quality, maturation and metabolic processes in the olive fruits. According to Mariyam Shafiq *et al*. (2021), most Dof genes in Group D1 of wild olive were associated with reproduction and circadian rhythm-related gene ontology (GO) terms, including “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”, suggesting the retention of these function and their possible

Table 3. Multilevel consensus sequence identified by MEME among OeDof proteins in *Olea europaea* var. europaea.

Motifs	Number of amino acids	Number of occurrences	E-value	Consensus sequence	Annotation
1	50	81	8.8e-3801	CPRCDSTNTKFCYYNNYSLSQPRHFCKTCRRYWTKGGTLRNVPVGGGCRK	Dof domain
2	38	10	9.0e-210	EKNSESSIVIPKTLRIDDPDEAAKSSIWATLGIKYDSV	GI-binding domain
3	28	16	6.8e-139	KPDKSENDQSETKKSQQKTLKKPDKILP	-
4	50	9	7.6e-158	TAANSNNEGGKTGLLEPQMKNINGFPPPVPCLPGVPWPVPWNSAVPVPAI	-
5	50	8	8.5e-146	HYRHITISEALQAARIDLPNGFHHPTFKPNGTVLSFTPDSPLCESMASVL	-
6	39	19	3.0e-150	SPLTPDSSPPSSSSQNPKIHDGQDLNLAYPGLPQFLEFV	-
7	39	17	1.4e-142	SFRLLFPFEELSNTAESEVEQSKGQENSNGYWNGMLGGG	-
8	20	20	1.8e-116	MLTCSKPSIEKKPRPQKEQA	-
9	28	10	3.6e-112	DKKKHVASTSTVLQANPAALSRSLNFQE	FKF1-binding domain
10	35	7	1.5e-100	QQFPFLGGLDPSPAGLYQFQGGMEQSSFVGETSQA	
11	39	7	2.3e-122	PPPPPPQPHATGGTGSVRPGSMAERARLANIPMPEAALK	
12	50	9	5.2e-109	SSNNINSSSSSPPSALSLLDLLRTSIGLNSFIPTPAVERNAVFASGFPFQ	
13	28	7	1.3e-089	MDFSSIPAYLDPANWQQQQNHQIGGSST	
14	28	7	4.3e-081	LETSIESLSSINQDLHWKLQQQRLAMLY	
15	39	4	6.5e-080	GGTNTQIEELSVAQTRVKQEMHNAREDERRELWGFPWQF	

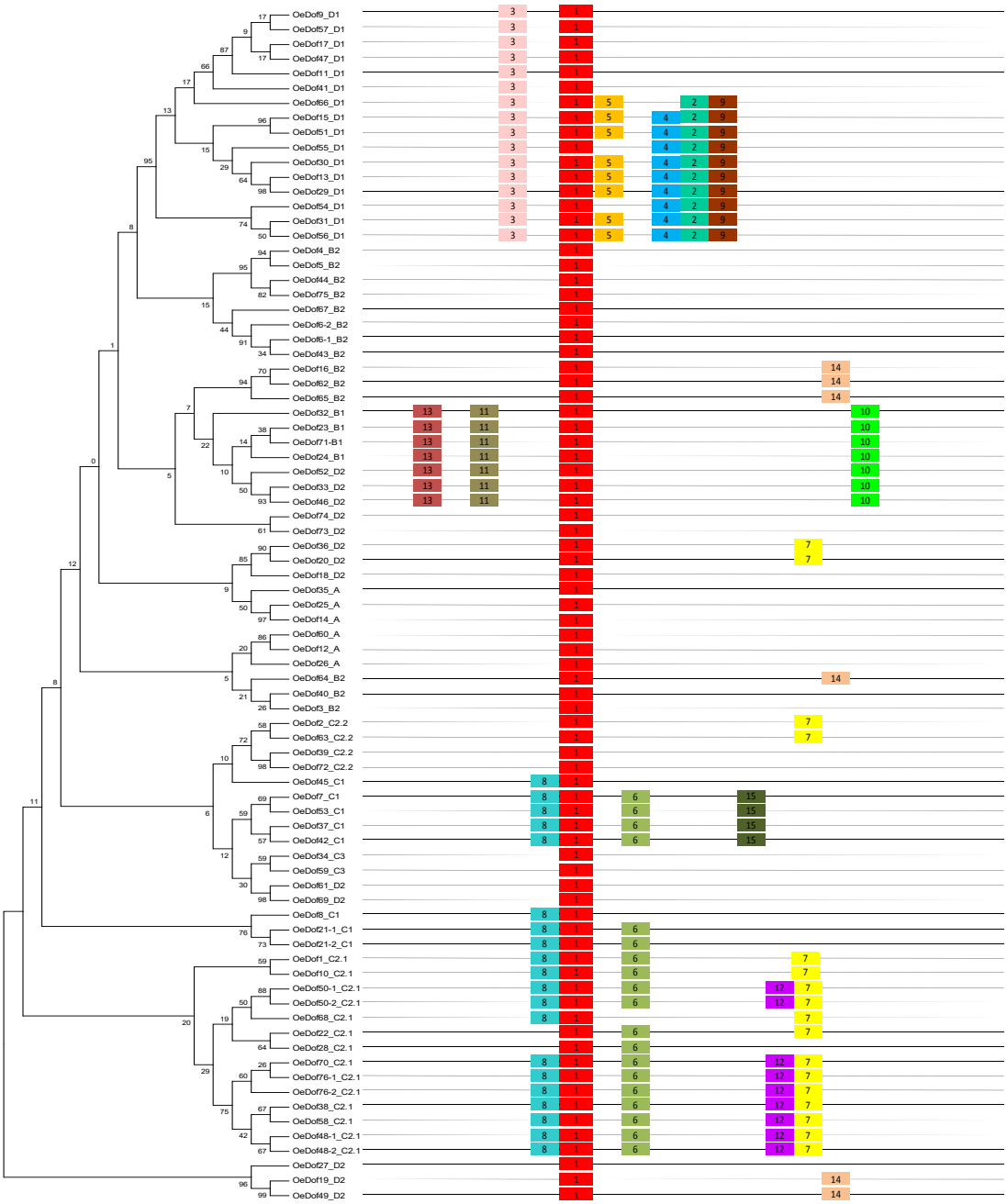

Fig. 3. Phylogenetic relationships and architecture of conserved protein motifs in Dof transcription factors from *Olea europaea* var. europaea. A phylogenetic tree was constructed based on the full-length sequences of OeDof proteins using MEGA 5.2 software. Fifteen conserved motifs (Dof and 2–15) are displayed in different colored boxes.

Table 4. Estimation of selection pressure and dates of the duplication events in *Olea europaea* subsp. europaea var. europaea.

Phylogenetic (sub)groups	Mean duplication date (Mya)	Gene 1	Gene 2	[a] Ks	[a] Ka	Ka/Ks	Duplication date (Mya)
Group A	54.16	*OeDof25*	*OeDof14*	0.012	0.004	0.333(-) [a]	3.33
		OeDof60	*OeDof12*	0.366	0.153	0.418(-)	101.66
Subgroup B1	98.33	*OeDof23*	*OeDof71*	0.354	0.081	0.228(-)	98.33
Subgroup B2	53.23	*OeDof4*	*OeDof5*	0.012	0.011	0.916(-)	3.33
		OeDof44	*OeDof75*	0.254	0.185	0.728(-)	70.55
		OeDof6-1	*OeDof43*	0.309	0.076	0.245(-)	85.83
		OeDof16	*OeDof62*	0.004	0	N/D [b]	N/D [b]
		OeDof40	*OeDof3*	2.218	0.963	N/D [b]	N/D [b]
Subgroup C1	6.66	*OeDof7*	*OeDof53*	0.024	0.002	0.083(-)	6.66
		OeDof37	*OeDof42*	0	0	N/D [b]	N/D [b]
Subgroup C2.1	126.19	*OeDof1*	*OeDof10*	0.226	0.057	0.252(-)	62.77
		OeDof22	*OeDof28*	0.249	0.100	0.401(-)	69.16
		OeDof70	*OeDof76-1*	0.888	0.193	0.217(-)	246.66
		OeDof38	*OeDof58*	0	0	N/D [b]	N/D [b]
Subgroup C2.2	45.27	*OeDof2*	*OeDof63*	0.292	0.081	0.277(-)	81.11
		OeDof39	*OeDof72*	0.034	0.021	0.617(-)	9.44
Subgroup C3	145.27	*OeDof34*	*OeDof59*	0.523	0.102	0.195(-)	145.27
Subgroup D1	112.38	*OeDof9*	*OeDof57*	0.865	0.133	0.153(-)	240.27
		OeDof17	*OeDof47*	0.833	0.164	0.196(-)	231.38
		OeDof15	*OeDof51*	0.262	0.081	0.309(-)	72.77
		OeDof13	*OeDof29*	0.017	0.009	0.529(-)	4.72
		OeDof31	*OeDof56*	0.046	0.030	0.652(-)	12.77
Subgroup D2	24.02	*OeDof33*	*OeDof46*	0.017	0.012	0.705(-)	4.72
		OeDof74	*OeDof73*	0.278	0.139	0.5(-)	77.22
		OeDof36	*OeDof20*	0	0.002	N/D [b]	N/D [b]
		OeDof61	*OeDof69*	0.021	0.002	0.095(-)	5.83
		OeDof19	*OeDof49*	0.030	0.039	1.3(+)	8.33

[(a)] Signs (+) and (−) refer to positive and negative selection, respectively. [(b)] Cells containing dashes refer to gene pairs excluded from the calculation of Ka/Ks and/or duplication dating, as very high or very low Ks values may result in biased estimates.

diversification through domestication.

Tracing the evolutionary history of *OeDof* genes: **duplication events, selection modes and functional dynamics**: To explore how the OeDof paralogous genes survived after the duplication events, a total of 27 paralogous gene pairs were identified from the phylogenetic tree (Fig. 3). We calculated the synonymous (Ks) and nonsynonymous (Ka) substitution rates as well as the Ka/Ks ratios of these gene pairs using gene and coding sequence (CDS) data (Table 4). As a result, we found that the Ka/Ks ratio of most OeDof paralogous gene pairs (22 out of 27) was lower than 1 (Table 4), indicating that these genes evolved through purifying selection, which tends to remove deleterious mutations at the protein level during the evolutionary history of the domesticated form olive genome. This result suggested that no significant functional divergence occurred after the duplication events giving birth to most OeDof paralogous gene pairs, consistent with the results of the study by Mariyam Shafiq *et al*. (2021), which also indicated that the Dof genes in the wild olive tree underwent purifying selection. Importantly, our findings demonstrated that a unique paralogous gene pair (*OeDof19*/*OeDof49*) underwent positive selection. As reported by Julca *et al*. (2020), it is possible that some genes *O. europaea*, involved in lipid, carbohydrate or amino acid metabolisms, as well as stress tolerance, solute transport and RNA processing may make part of selective scanning regions of domestication, and therefore could undergo neofunctionalization as part of the domestication evolutionary process. As a result, genes *OeDof19* and *OeDof49* may be candidates for key biological processes related to olive breeding; however this hypothesis requires further exploration.

The synonymous substitution rate (Ks) was then used to estimate the dates of duplication for the paralogous genes. The predicted date for gene duplication was estimated in the range from ~3.33 Mya for paralogous pairs *OeDof25*/*OeDof14* and *OeDof4*/*OeDof5*, to ~246 Mya for paralogous pair *OeDof70*/*OeDof76*. Notably, eight pairs of paralogous Dof genes (*OeDof25/14*, *OeDof4/5*, *OeDof7/53*, *OeDof39/72*, *OeDof13/29*, *OeDof33/46*, *OeDof61/69*, and *OeDof19/49*) underwent relatively recent duplications (<10 Mya). Among them, six exhibited Ka/Ks ratios exceeding 0.5, indicating that while negative selection influenced these duplicates, relaxed purifying selection may have facilitated significant functional divergence post-duplication. It is known that the most recent gene duplication wave comprised olive-specific gene duplications and followed the evolutionary divergence of *O. europaea* (subtribe: *Oleinae*) and *Populus angustifolia* occurring ~10 Mya (Julca *et al*., 2018). It turns out that a substantial amount of Dof gene family expansion occurred following the speciation of olive, which was characterized by gene duplication dynamics as well as relaxed selection pressure. Positive and/or relaxed selection acting on paralogs suggests that the most recent OeDof genes were part of an evolutionary process characterized by gene birth and functional flexibility. Functional diversification prior to and during the olive domestication would have conferred new adaptation specificities. Our findings provide evolutionary insights into the functional dynamics of the Dof gene family in *O. europaea*, shedding light on the timings and most probable fates of gene duplication events in this species.

***In silico* expression analysis of *OeDof* genes across cultivars with distinct fruit traits**: **linking transcriptomic and morpho-metabolomic profiles**: To understand the general expression pattern of common olive Dof genes and to clarify their putative functions

in controlling the metabolisms of fatty acids inside the fruit of different olive genotypes, we mapped whole transcriptome reads generated in Niu *et al*. (2022) from three common olive cultivars with distinct morphological and biochemical fruit traits, 'Arbequina', 'Frantoio Selection' and 'Nikitskii I' against the Mediterranean olive genome, then we retrieved the quantitative expression values of the OeDof genes in each cultivar. The results revealed variable expression profiles of OeDofs across the three cultivars. In total, 46 differentially-expressed genes (DEGs) were identified. Fig. 4 depicts a heatmap representation of the expression profiles of OeDof genes. As a whole, we grouped DEGs into 6 clusters based on their expression profiles (Fig. 4). Cluster I contains 9 OeDof genes (*OeDof1*, *-4*, *-12*, *-26*, *-30*, *-54*, *-56*, *-67* and *-73*) that are uniformly highly expressed in all three olive cultivars. Cluster II includes four OeDof genes (*OeDof5*, *-31*, *-50*, and *-22*) that are highly expressed in oil-rich cultivars, 'Arbequina' and 'Frantoio selection,' which are known for their moderate flesh-to-pit ratios and relatively high oil content (Niu *et al*., 2022). These cultivars may exhibit similar genetic control over fruit development, potentially influenced by the significantly high expression of cluster II genes. Cluster III includes two genes, *OeDof44* and *OeDof23*, which are abundantly expressed in 'Frantoio selection,' a cultivar known for its high linoleic acid content in fruits (Niu *et al*., 2022). The potential role of these genes in linoleic acid production warrants further investigation. Cluster IV consists of two genes, *OeDof55* and *OeDof14*, which exhibit the highest expression levels exclusively in the 'Arbequina' cultivar. Given that 'Arbequina' has the highest flavonoid content (801.87 µg/g FW) among the three cultivars (Niu *et al*., 2022), these genes should be further investigated for their role in flavonoid biosynthesis. Cluster V is the largest, containing 28 OeDof genes with low expression levels across all three olive cultivars. Finally, cluster VI only contains *OeDof2*, which is highly expressed in both 'Arbequina' and 'Nikitskii

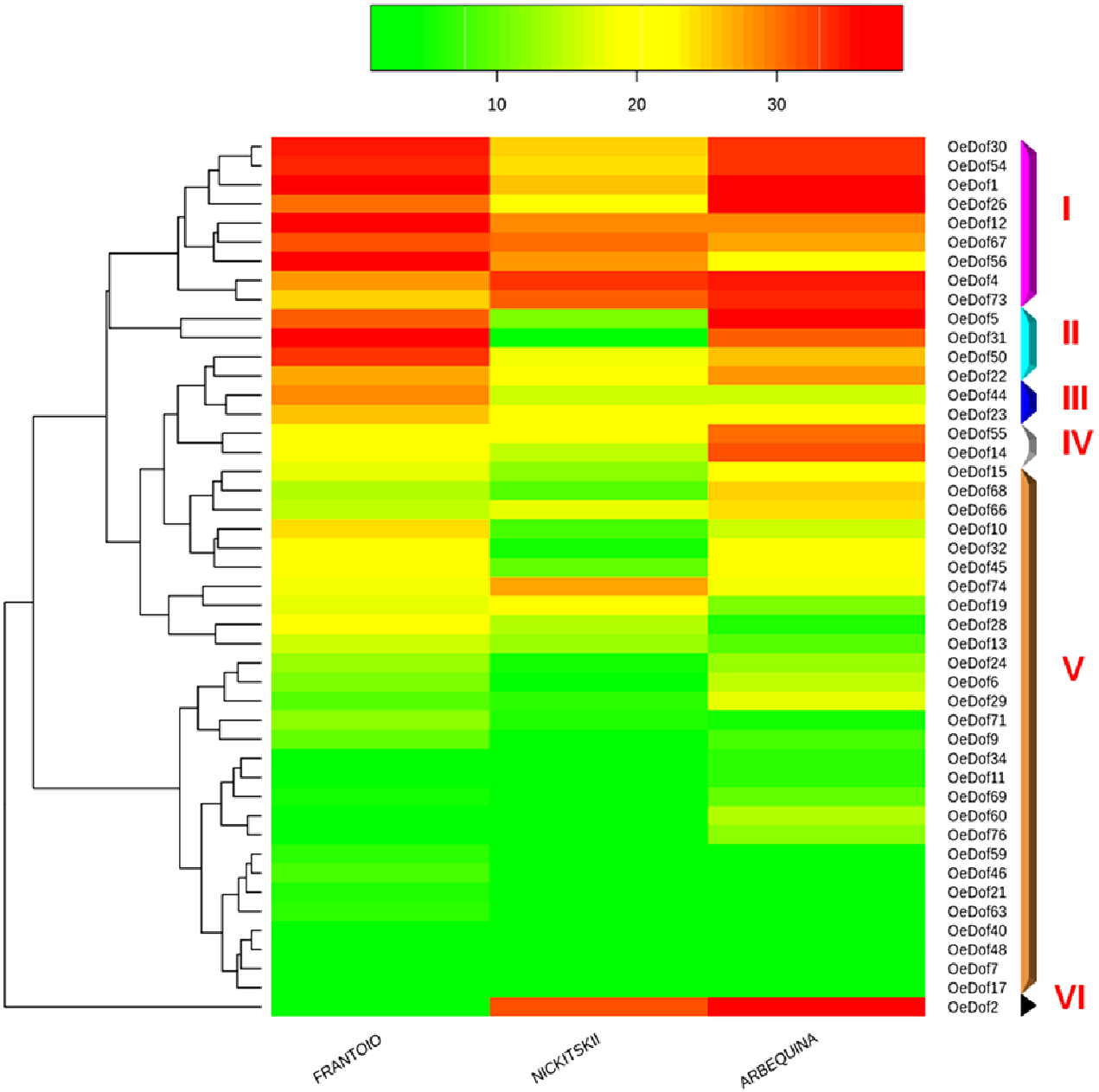

Fig. 4. The expression profiles of the 46 OeDof genes in fruit among three Mediterranean olive cultivars, 'Arbequina', 'Frantoio selection' and 'Nikitskii I', with different fruit morphological and biochemical traits, based on FPKM values. Gradient with 3 colors (green, yellow and red) indicate, low, average and high expression levels, respectively, for each gene.

I'. Compared to 'Frantoio selection,' both cultivars have higher levels of oleic (C18:1), palmitic (C16:0), and heptadecenoic (C17:1) acids (Niu *et al*., 2022), suggesting a potential role for OeDof2 in their biosynthesis.

Overall, this analysis highlights candidate OeDof genes and coordinated expression modules potentially involved in fatty acid and flavonoid biosynthesis, as well as fruit morphology, across three cultivars with distinct biochemical profiles. To confirm their specific roles, these candidate genes and expression modules require functional characterization.

This study provides a comprehensive genome-wide characterization of the domesticated olive Dof gene family, an important target for crop improvement due to its role in controlling key agronomic traits. Our findings reveal 81 Dof TFs encoded by 76 genes in the domesticated olive genome, with distinct physicochemical properties and motif compositions. The larger number of Dof genes detected in the cultivated olive genome compared with the reported wild olive genome suggests a possible expansion of this gene family in cultivated olive; however, this conclusion should be interpreted cautiously because genome assembly quality, annotation depth and gene-identification criteria may influence gene counts. Genes within the same phylogenetic groups/subgroups shared similar structural motifs, supporting functional predictions based on orthology with Arabidopsis and motif annotations. Most OeDof genes have undergone purifying selection, with one gene pair showing signs of positive selection, suggesting a role in olive domestication. RNA-Seq analysis identified candidate OeDof genes and clustered expression modules across three cultivars with distinct morpho-biochemical profiles, which could be related to fatty acid and flavonoid metabolism, as well as fruit morphology. In order to confirm their specific roles, these candidate genes and expression gene clusters require further functional validation. This study offers valuable insights into the molecular, evolutionary and preliminary functional features of Dof genes in *O. europaea* var. *europaea* and lays the groundwork for future functional assays and application of these genes in genetic improvement programs.

Acknowledgements

We thank all members of the Laboratory of Genetics, Biodiversity & Bioressources Exploitation, LR11ES41 (University of Monastir), for assistance. This study was financially supported by the Ministry of Higher Education and Scientific Research (Tunisia).

References

Barghini, E., L. Natali, T. Giordani, R.M. Cossu, S. Scalabrin, F. Cattonaro, H. Šimková, J. Vrána, J. Doležel, M. Morgante and A. Cavallini, 2015. LTR retrotransposon dynamics in the evolution of the olive (*Olea europaea*) genome. *DNA Res*., 22(1): 91-100.

Barone, E., M.L. Mantia, A. Marchese and F.P. Marra, 2014. Improvement in yield and fruit size and quality of the main Italian table olive cultivar 'Nocellara del Belice'. *Scientia Agricola,* 71(1): 52-57.

Breton, C., J.F. Terral, C. Pinatel, F. Medail, F. Bonhomme and A. Berville, 2009. The origins of the domestication of the olive tree. *C. R. Biol.,* 332: 1059-1064.

Cruz, F., I. Julca, J. Gómez-Garrido, D. Loska, M. Marcet-Houben, E. Cano, B. Galán, L. Frias, P. Ribeca, S. Derdak, M. Gut, M. Sánchez-Fernández, J.L. García, I.G. Gut, P. Vargas, T.S. Alioto and T. Gabaldón, 2016. Genome sequence of the olive tree, *Olea europaea*. *GigaScience,* 5: 29.

Dohm, J.C., A.E. Minoche, D. Holtgräwe, S. Capella-Gutiérrez, F. Zakrzewski, H. Tafer, O. Rupp, T.R. Sörensen, R. Stracke, R. Reinhardt, A. Goesmann, T. Kraft, B. Schulz, P.F. Stadler, T. Schmidt, T. Gabaldón, H. Lehrach, B. Weisshaar and H. Himmelbauer, 2014. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature, 505: 546-549.

El, S.N. and S. Karakaya, 2009. Olive tree (*Olea europaea*) leaves: Potential beneficial effects on human health. *Nutr. Rev*., 67: 632-638.

Fukao, T., T. Harris and J. Bailey-Serres, 2009. Evolutionary analysis of the Sub1 gene cluster that confers submergence tolerance to domesticated rice. *Ann. Bot.*, 103(2): 143-150.

García-Rodríguez, R., A. Belaj, C. Romero-Segura, C. Sanz and A.G. Pérez, 2017. Exploration of genetic resources to improve the functional quality of virgin olive oil. *J. Funct. Foods,* 38: 1-8.

Ghanbari, R., F. Anwar, K.M. Alkharfy, A.H. Gilani and N. Saari, 2012. Valuable nutrients and functional bioactives in different parts of olive (*Olea europaea* L.): A review. *Int. J. Mol. Sci*., 13: 3291-3340.

Gouvinhas, I., N. Machado, C. Sobreira, R. Dominguez-Perles, S. Gomes, E. Rosa and A. Barros, 2017. Critical review on the significance of olive phytochemicals in plant physiology and human health. *Molecules*, 22(11): 1986.

Gupta, S., N. Malviya, H. Kushwaha, J. Nasim, N.C. Bisht, V.K. Singh and D. Yadav, 2015. Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. *Planta,* 241(3): 549-562.

Hamdi, J., N. Kmeli and D. Bouktila, 2021. Genome-wide survey of sugar beet (*Beta vulgaris* subsp. *vulgaris*) Dof transcription factors reveals structural diversity, evolutionary expansion and involvement in taproot development and biotic stress response. *Biologia*, 76: 2421-2436.

Hamed, H.H., A.A. Abdelhafez, A.A. Rashedy and S. Shaddad, 2025. Toward sustainable oliviculture: Effects of replacing mineral with organic nitrogen on floral behaviour, fruit quality and yield. *J. Appl. Hortic.*, 27(3): 494-500.

Hernando-Amado, S., V. González-Calle, P. Carbonero and C. Barrero-Sicilia, 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.

Ito, S., Y.H. Song and T. Imaizumi, 2012. LOV domain-containing F-box proteins: light-dependent protein degradation modules in *Arabidopsis. Mol. Plant,* 5: 573-582.

Jia, B., X. Xie, M. Wu, Z. Lin, J. Yin, S. Lou, Y. Huang, Z. Hu, 2019. Understanding the functions of endogenous DOF transcript factor in *Chlamydomonas reinhardtii*. *Biotechnol. Biofuels*, 12: 67.

Julca, I., M. Marcet-Houben, F. Cruz, J. Gómez-Garrido, B.S. Gaut, C.M. Díez, I.G. Gut, T.S. Alioto, P. Vargas and T. Gabaldón, 2020. Genomic evidence for recurrent genetic admixture during the domestication of Mediterranean olive trees (*Olea europaea* L.). *BMC Biol*., 18(1): 148.

Julca, I., M. Marcet-Houben, P. Vargas and T. Gabaldón, 2018. Phylogenomics of the olive tree (*Olea europaea*) reveals the relative contribution of ancient allo- and autopolyploidization events. *BMC Biol.,* 16(1): 15.

Kim, S.C., K.N. Edgeworth, D.A. Nusinow and X. Wang, 2023. Circadian clock factors regulate the first condensation reaction of fatty acid synthesis in Arabidopsis. *Cell Rep.*, 42(12): 2211-2247.

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

Lopez-Moreno, H., A.C. Basurto-Garduño, M.A. Torres-Meraz, E. Diaz-Valenzuela, S. Arellano-Arciniega, J. Zalapa, R.J.H. Sawers, A. Cibrián-Jaramillo and L. Diaz-Garcia, 2023. Genetic analysis and QTL mapping of domestication-related traits in chili pepper (*Capsicum annuum* L.). *Front. Genet*., 14: 1101401.

Mariyam Shafiq, M., M. Haseeb, R. Muhammad Atif, S.A.A.A. Naqvi, N. Ali, M.A. Javed, F. Gillani and M.S. Haider, 2021. Genome-wide identification and characterization of a plant-specific Dof transcription factor gene family in olive (*Olea europaea*) and its comparison with Arabidopsis. *Hortic. Environ. Biotechnol.*, 62: 949-968.

Niu, E., W. Hu, J. Ding, W. Wang, A. Romero, G. Shen and S. Zhu, 2022. GC-MS/LC-MS and transcriptome analyses revealed the metabolisms of fatty acid and flavonoid in olive fruits (*Olea europaea* L.). *Sci. Hortic.*, 299: 111017.

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

Oyama, R.K., M.J. Clauss, N. Formanová, J. Kroymann, K.J. Schmid, H. Vogel, K. Weniger, A.J. Windsor and T. Mitchell-Olds, 2008. The shrunken genome of *Arabidopsis thaliana. Plant Syst. Evol.,* 273: 257-271.

Qin, C., C. Yu, Y. Shen, X. Fang, L. Chen, J. Min, J. Cheng, S. Zhao, M. Xu, Y. Luo, Y. Yang, Z. Wu, L. Mao, H. Wu, C. Ling-Hu, H. Zhou, H. Lin, S. González-Morales, D.L. Trejo-Saavedra, H. Tian, X. Tang, M. Zhao, Z. Huang, A. Zhou, X. Yao, J. Cui, W. Li, Z. Chen, Y. Feng, Y. Niu, S. Bi, X. Yang, W. Li, H. Cai, X. Luo, S. Montes-Hernández, M.A. Leyva-González, Z. Xiong, X. He, L. Bai, S. Tan, X. Tang, D. Liu, J. Liu, S. Zhang, M. Chen, L.K. Zhang, L. Zhang, Y. Zhang, W. Liao, Y. Zhang, M. Wang, X. Lv, B. Wen, H. Liu, H. Luan, Y. Zhang, S. Yang, X. Wang, J. Xu, X. Li, S. Li, J. Wang, A. Palloix, P.W. Bosland, Y. Li, A. Krogh, R.F. Rivera-Bustamante and L. Herrera-Estrella, 2014. Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. *Proc. Natl. Acad. Sci. U.S.A.*, 111(14): 5135-5140.

Shibuya, T., M. Nishiyama, K. Kato and Y. Kanayama, 2021. Characterization of the FLAVIN-BINDING, KELCH REPEAT, F-BOX 1 Homolog SlFKF1 in tomato as a model for plants with fleshy fruit. *Int. J. Mol. Sci*., 22(4): 1735.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar, 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol*., 28: 2731-2739.

Tatusov, R.L., E.V. Koonin and D.J. Lipman, 1997. A genomic perspective on protein families. *Science*, 278: 631-637.

Unver, T., Z. Wu, L. Sterck, M. Turktas, R. Lohaus, Z. Li, M. Yang, L. He, T. Deng, FJ. Escalante, C. Llorens, FJ. Roig, I. Parmaksiz, E. Dundar, F. Xie, B. Zhang, A. Ipek, S. Uranbey, M. Erayman, E. Ilhan, O. Badad, H. Ghazal, D.A. Lightfoot, P. Kasarla, V. Colantonio, H. Tombuloglu, P. Hernandez, N. Mete, O. Cetin, M.V. Montagu, H. Yang, Q. Gao, G. Dorado and YV de Peer, 2017. Genome of wild olive and the evolution of oil biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.*, 114: E9413-E9422.

Valverde, P., D. Barranco, F.J. López-Escudero, C. Díez and C. Trapero, 2023. Efficiency of breeding olives for resistance to *Verticillium* wilt. *Front. Plant Sci.*, 14: 1149570.

Wang, H.W., B. Zhang, Y.J. Hao, J. Huang, A.G. Tian, Y. Liao, J.S. Zhang and S.Y. Chen, 2007. The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic *Arabidopsis plants. Plant J*., 52: 716-729.

Wang, P., Z. Yan, X. Zong, Q. Yan and J. Zhang, 2021. Genome-wide analysis and expression profiles of the Dof family in *Cleistogenes songorica* under temperature, salt and ABA treatment. *Plants,* 10(5): 850.

Wu, Z., J. Cheng, J. Cui, X. Xu, G. Liang, X. Luo, X. Chen, X. Tang, K. Hu and C. Qin, 2016. Genome-Wide Identification and Expression Profile of Dof Transcription Factor Gene Family in Pepper (*Capsicum annuum* L.). *Front. Plant Sci*., 7: 574.

Zhao, Q., Q. Feng, H. Lu, Y. Li, A. Wang, Q. Tian, Q. Zhan, Y. Lu, L. Zhang, T. Huang, Y. Wang, D. Fan, Y. Zhao, Z. Wang, C. Zhou, J. Chen, C. Zhu, W. Li, Q. Weng, Q. Xu and Z.-X. Wang, 2018. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet*., 50(2): 278-284.

Zou, Z., J. Zhu and X. Zhang, 2019. Genome-wide identification and characterization of the Dof gene family in cassava (*Manihot esculenta*). *Gene*, 687: 298-307.

Received: October, 2025; Revised: November, 2025;
Accepted: December, 2025