

ARTICLE

Transcriptome analyses of the *Dof*-like gene family in grapevine reveal its involvement in berry, flower and seed development

Danielle Costenaro da Silva^{1,6,7}, Vítor da Silveira Falavigna^{1,6}, Marianna Fasoli^{2,8}, Vanessa Buffon³, Diogo Denardi Porto⁴, Georgios Joannis Pappas Jr⁵, Mario Pezzotti², Giancarlo Pasquali¹ and Luís Fernando Revers³

The Dof (DNA-binding with one finger) protein family spans a group of plant transcription factors involved in the regulation of several functions, such as plant responses to stress, hormones and light, phytochrome signaling and seed germination. Here we describe the *Dof*-like gene family in grapevine (*Vitis vinifera* L.), which consists of 25 genes coding for Dof. An extensive *in silico* characterization of the *VviDofL* gene family was performed. Additionally, the expression of the entire gene family was assessed in 54 grapevine tissues and organs using an integrated approach with microarray (cv Corvina) and real-time PCR (cv Pinot Noir) analyses. The phylogenetic analysis comparing grapevine sequences with those of Arabidopsis, tomato, poplar and already described *Dof* genes in other species allowed us to identify several duplicated genes. The diversification of grapevine *DofL* genes during evolution likely resulted in a broader range of biological roles. Furthermore, distinct expression patterns were identified between samples analyzed, corroborating such hypothesis. Our expression results indicate that several *VviDofL* genes perform their functional roles mainly during flower, berry and seed development, highlighting their importance for grapevine growth and production. The identification of similar expression profiles between both approaches strongly suggests that these genes have important regulatory roles that are evolutionally conserved between grapevine cvs Corvina and Pinot Noir.

Horticulture Research (2016) 3, 16042; doi:10.1038/hortres.2016.42; Published online 31 August 2016

INTRODUCTION

Dof (DNA-binding with one finger) proteins are plant exclusive transcription factors characterized by a conserved 52 amino acid segment encompassing a C₂C₂-type zinc finger.¹ The Dof domain is usually found in the N-terminal region of these proteins, specifically binding AAAG sequences in gene promoters. The C-terminal region is responsible to interact with other regulatory proteins in order to activate gene expression (reviewed in Yanagisawa¹ and Hir and Bellini²). In contrast to the highly conserved DNA-binding domain, sequences outside the Dof domain diverge widely, suggesting that Dof proteins are regulating several biological functions. Indeed, functional studies conducted so far confirm this assumption, demonstrating Dof involvement in seed germination,^{3–9} seed development,^{10–14} flowering control,^{15–18} flower abscission,¹⁹ pollen development,²⁰ among many other functions (recently reviewed in Gupta *et al.*,²¹ Hir and Bellini² and Noguero *et al.*²²).

After the first isolation of a complementary DNA from maize encoding a Dof domain-containing protein, these transcription factors have been identified in a variety of angiosperms including Arabidopsis, barley, pea, potato, pumpkin, rice, tobacco and wheat.¹ Thereafter, *Dof*-like genes were identified in several other species such as grapevine,²³ *Chlamydomonas reinhardtii*, *Jatropha curcas*, maritime pine, pea, the moss *Physcomitrella patens*, sorghum,

soybean and sweet potato.^{21,22} In terms of family composition, the first genome-wide report for *Dof* genes was for Arabidopsis, which presented 36 Dof-encoding genes and one pseudogene.¹ Thereafter, many studies identifying *Dof* genes have been reported, with 30 members in rice,²⁴ 41 in poplar,²⁵ 26 in barley,²⁶ 19 in moss,²⁷ 12 in the fern *Selaginella moellendorffii*,²⁷ 10 in loblolly pine,²⁷ 31 in wheat,²⁸ 27 in *Brachypodium distachyon*,²⁹ 10 in pine,³⁰ 34 in tomato,³¹ 78 in soybean,³² 25 in sugarcane,³³ 76 in Chinese cabbage,³⁴ 46 in maize,³⁵ 35 in potato³⁶ and 46 in carrot.³⁷

We have previously conducted a representational difference analysis aiming to identify genes related to seedlessness during specific developmental stages in grapevine.²³ Among the expressed sequence tags identified, a putative *Dof* gene was isolated in the developmental stage of four-weeks after fruit set. We therefore decided to further investigate the presence of other genes in the grapevine genome potentially encoding Dof proteins. Here we demonstrate that 25 *Dof*-like genes are present in grapevine (*VviDofL*). *In silico* analyses, including a phylogenetic approach that revealed clusters of paralogous and orthologous genes, allowed the prediction of putative functions for these genes. Finally, the expression profiles of all 25 *VviDofL* genes were established by whole-transcriptome data and strongly suggested

¹Graduate Program in Cell and Molecular Biology, Center for Biotechnology, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 91501-970, Brazil; ²Dipartimento di Biotecnologie, Università degli Studi di Verona, Verona 37134, Italy; ³Embrapa Uva e Vinho, Bento Gonçalves, RS 95701-008, Brazil; ⁴Embrapa Semiárido, Petrolina, PE 56302-970, Brazil and ⁵Department of Cell Biology, Universidade de Brasília, Brasília, DF 70910-900, Brazil.

Correspondence: LF Revers (luis.revers@embrapa.br)

⁶These authors contributed equally to this work.

⁷Present address: Syngenta Seeds Ltda, Uberlândia, MG, 38407-049, Brazil.

⁸Present address: E. & J. Gallo Winery, Modesto, CA 95354, USA.

Received: 30 June 2016; Revised: 8 August 2016; Accepted: 9 August 2016

the involvement of some of these genes in the regulation of berry, flower and seed development.

MATERIAL AND METHODS

Bioinformatic analysis of nucleotide and amino acid sequences
A *VviDofL* sequence previously identified by the group²³ was used as query in a genome-wide search for putative Dofs using the *BLASTP* tool³⁸ implemented in the 12X Grape Genome Browser V1 (<http://genomes.cribi.unipd.it/>),^{39,40} which was verified in the V2 version with the same 25 *VviDofL*. In order to confirm the results, the *Dof*-seed sequence (accession number PF02701) available in the Pfam database⁴¹ was used to perform a HMM search using Hmmer v3.0 (<http://hmmer.org/>) to find predicted grapevine proteins in the 12X Grape Genome Browser V1. All predicted gene models identified by both strategies were compared with the Pfam database⁴¹ in order to confirm the presence of the *Dof* domain (PF02701), and the ones without the characteristic domain of the gene family were excluded from further analyses. The chromosomal locations of *VviDofL* genes were obtained at the Grape Genome Browser, whereas marker locations were obtained at the National Center for Biotechnology Information database (<http://www.ncbi.nlm.nih.gov/>). All data were compiled into a figure using the *MapChart* v.2.2 (Wageningen, Netherlands) software.⁴²

Deduced amino acid sequences of all grapevine *DofL* proteins were searched for the presence of conserved domains using MEME Suite v.4.9.0.⁴³ Default parameters were used with the following exceptions: the motif distribution among sequences was set to any number of repetitions; the maximum number of motifs was set to 15; the maximum motif width was defined between 10 and 52 amino acids; and an *e*-value cutoff of 1^{-10} was adopted. All identified motifs were annotated by comparison with conserved motifs in Pfam⁴¹ and SMART⁴⁴ databases. Protein subcellular localization was predicted using the Plant-mPLOC database (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/#>).⁴⁵ Full-length protein sequences of grapevine, Arabidopsis, poplar, tomato and other already described *Dof*-encoding genes (searched in March, 2016) were aligned using the ClustalW software.⁴⁶ Accession numbers are shown in Supplementary Table S1. Two phylogenetic trees were inferred using MrBayes v.3.2.6 software⁴⁷ employing the mixed amino acid substitution model in the default settings. The first tree performed a 14.81 million generation run, sampled every 100 generations and the first 25% trees were discarded as burn-in. The remaining ones were summarized in a consensus tree. The alignment of the second tree was cured to eliminate poor alignment positions and divergent regions using Gblocks 0.91b (Barcelona, Spain) software in default settings, except gap positions that were set to half.⁴⁸ The cured alignment was composed of 50 amino acids spanning the *Dof* domain region. The tree was analyzed as previously described but using 9.2 million generation run. Phylogenetic trees were visualized and edited using FigTree v.1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Analysis of whole *VviDofL* transcriptome

In order to fully characterize the expression of *VviDofL* genes during grapevine development, 54 tissues and organs of *Vitis vinifera* cv Corvina (clone 48, rootstock 41B) were analyzed using a comprehensive genome microarray. Plant materials used for the construction of the database encompassed all major grapevine organs in different developmental stages, and included data from berries during post-harvest withering, green and woody tissues, as well as specialized tissues such as pollen. Genome-wide transcriptome analysis was performed by microarray as described.⁴⁹ The expression data was analyzed by hierarchical clustering on the whole 54-sample data set. Pearson's correlation distance was used as the metric, and T-MeV v4.81 (Boston, MA, USA) software was used to create the transcriptional profiles dendrograms for both

genes and samples.⁵⁰ Expression data are shown as normalized based on the mean center genes/rows adjustment.

Real-time PCR

Grapevine (*Vitis vinifera* L. cv Pinot Noir grafted on Paulsen 1103) samples from at least seven independent plants located in experimental plots at Embrapa Uva e Vinho (Bento Gonçalves/RS, Brazil) were sampled during the 2008/2009 growing season. Three biological replicates consisted of tissues from three independent plants kept separately throughout the analysis. The developmental stages were defined as described by Coombe⁵¹ (modified E-L system) except when not indicated: roots of *in vitro* cultivated plants, 5 cm diameter leaves, tips of 10 cm stems (E-L 12), inflorescences and tendrils from pre-anthesis (E-L 17), summer buds (E-L 31), 7 mm-large berries (E-L 31), pre-véraison and véraison berries (E-L 34–35). Pre-véraison and véraison berries were sampled from the same bunch by separating half-colored berries from green colored berries. All samples were frozen in liquid nitrogen in the field and stored at -80°C until RNA extraction.

Primer pairs were designed using the Primer3 program (<http://frodo.wi.mit.edu/primer3>, Supplementary Table S2). Total RNA was extracted from frozen tissues using the Purelink RNA Reagent (Invitrogen, Carlsbad, CA, USA) and protocols recommended by the manufacturer. All RNA preparations were treated with DNase (Fermentas, Waltham, MA, USA) in order to eliminate residual DNA contamination. Quantity and quality of total RNAs were evaluated with the Qubit Quantitation Platform (Invitrogen) and standard 1.5% agarose gel electrophoresis in $1\times$ MOPS buffer. Complementary DNA was prepared using M-MLV Reverse Transcriptase (Promega, Madison, WI, USA). Transcript levels were determined by real-time PCR using a StepOne Plus PCR system (Applied Biosystems, Foster City, CA, USA) and SYBR Green (Applied Biosystems). The cycling protocol consisted of one step at 95°C for 5 min, followed by 40 cycles at 95°C for 15 s and 60°C for 35 s, and finished by a dissociation curve between 60°C and 95°C . The specificity of PCR amplifications was assessed by the presence of a single peak in the melting curves. Biological samples ($n=3$) were analyzed in three technical replicates. Glyceraldehyde 3-phosphate dehydrogenase and tubulin were used as reference genes⁵² using the $2^{-\Delta\Delta\text{Ct}}$ method.⁵³

RESULTS

Identification and annotation of grapevine *DofL* genes

In order to identify putative *Dof* proteins in grapevine, a *VviDofL* sequence previously identified by the group²³ was used as query in *BLAST* searches, resulting in the identification of 25 genes possibly coding for *Dof* proteins. This result was further confirmed in a HMM search using the *Dof*-seed sequence available at the Pfam database (PF02701) against the grapevine genome database. Putative *Dof* genes were named according to Grimplet *et al.*⁵⁴ using phylogenetic trees provided by Gramene⁵⁵ (Figure 1). Grapevine *Dof* domains were aligned and all highly conserved residues described to the DNA-binding *Dof* domain were mapped in the 25 sequences (Supplementary Figure S1). All predicted gene models identified along with their chromosomal location and deduced protein length are presented in Figure 1. The putative functions of the grapevine *Dof* proteins were further investigated by predicting their subcellular localization using the Plant-mPLOC database.⁴⁵ Only three sequences (*VviDofL23*, *VviDofL24* and *VviDofL25*) not had their predicted subcellular location to the nucleus (Supplementary Table S3). Finally, intron and exon structures, as well as *Dof* domains, were determined and are displayed in Supplementary Figure S2 using FancyGene v1.4.⁵⁶

The 25 putative *Dof* genes were found to be distributed in 14 out of the 19 grapevine chromosomes, as well as in the unplaced contigs chromosome (ChrUn), which contains sequences whose

Gene name	Genome locus	Chromosomal location	Size (aa)	Motif distribution
<i>VviDofL1.4</i>	VIT_00s0652g00010	chrUn:33581346..33582988	295	
<i>VviDofL1.6</i>	VIT_03s0063g01350	chr3:4802389..4802910	173	
<i>VviDofL2.1</i>	VIT_06s0004g03420	chr6:4265287..4267184	296	
<i>VviDofL3.5</i>	VIT_08s0105g00170	chr8:7109736..7110548	270	
<i>VviDofL5.6</i>	VIT_17s0000g08290	chr17:9361008..9362865	308	
<i>VviDofL6</i>	VIT_17s0000g06310	chr17:6845904..6848869	473	
<i>VviDofL7</i>	VIT_01s0026g02580	chr1:12212183..12215486	477	
<i>VviDofL8</i>	VIT_14s0108g00980	chr14:29617368..29620681	511	
<i>VviDofL9</i>	VIT_08s0056g01230	chr8:1937113..1940026	494	
<i>VviDofL10</i>	VIT_18s0001g15730	chr18:9622093..9623242	281	
<i>VviDofL11</i>	VIT_02s0025g02250	chr2:2012768..2014765	288	
<i>VviDofL12</i>	VIT_16s0098g01420	chr16:21635846..21637912	287	
<i>VviDofL13</i>	VIT_15s0046g00150	chr15:17188767..17190484	316	
<i>VviDofL14</i>	VIT_08s0007g00180	chr8:14550709..14554703	345	
<i>VviDofL15</i>	VIT_13s0019g01410	chr13:3000612..3002046	337	
<i>VviDofL16</i>	VIT_06s0004g04520	chr6:5492077..5493425	320	
<i>VviDofL17</i>	VIT_10s0003g00030	chr10:1318306..1319562	299	
<i>VviDofL18</i>	VIT_17s0000g04850	chr17:5240463..5241938	318	
<i>VviDofL19</i>	VIT_00s0218g00040	chrUn:13957686..13958702	338	
<i>VviDofL20</i>	VIT_00s0253g00060	chrUn:17750019..17752319	352	
<i>VviDofL21</i>	VIT_10s0003g00040	chr10:1335706..1337565	308	
<i>VviDofL22</i>	VIT_09s0002g02490	chr9:2271830..2273304	216	
<i>VviDofL23</i>	VIT_07s0255g00020	chr7:14346778..14347676	220	
<i>VviDofL24</i>	VIT_18s0001g11310	chr18:13865318..13866466	226	
<i>VviDofL25</i>	VIT_10s0003g01260	chr10:2577215..2577834	159	

Figure 1. Identification of grapevine *DofL* genes. Genome locus is provided by the 'Grape Genome' (<http://genomes.cripi.unipd.it/>). Chromosomal localization and deduced peptide lengths are shown. Schematic view of the conserved motifs between *VviDofL*-deduced protein sequences was performed by the *MEME* suite.⁴³ Colored boxes represent conserved motifs (Supplementary Figure S3). The height of the motif box is proportional to $-\log(P \text{ value})$, with the maximum height denoting a P value of e^{-10} . Gray lines represent non-conserved sequences. *VviDofL18* was previously named *VvDof1* by Costenaro-da-Silva *et al.*²³

physical position on specific chromosomes have not yet been defined. As shown in Figure 2, chromosomes 8, 10 and 17 have three *DofL* genes each, chromosomes 6 and 18 both have two *DofL* genes and chromosomes 1, 2, 3, 7, 9, 13, 14, 15 and 16 have one *DofL* gene each. ChrUn also contains three *DofL* genes. Genome annotated scaffolds and molecular markers mapped near to *DofL* genes in each grapevine chromosome are also represented in Figure 2.

Phylogenetic studies of predicted *DofL* proteins

The analysis of the deduced grapevine *DofL* protein sequences using the *MEME* software revealed the existence of several different motifs apart from the conserved *Dof* domain. Eleven motifs were identified with an e -value support lower than 1^{-10} and their annotation was performed using Pfam and SMART databases. However, only the *Dof* domain could be annotated (cyan motif in Figure 1 and in Supplementary Figure S3). The

overall motif distribution of grapevine *DofL* sequences is shown in Figure 1. The *Dof* motif was the only one shared between all sequences. The other ten motifs were distributed unequally between *DofL* proteins (Figure 1). The consensus sequences of all motifs are shown in Supplementary Figure S3.

A phylogenetic study was carried out to evaluate evolutionary relationships within *DofL* genes. To this purpose, 167 sequences from 19 different plant species were used, including all identified sequences from *Arabidopsis*, grapevine, poplar and tomato, in addition to *Dof* proteins already described in barley, *C. reinhardtii*, *J. curcas*, maize, maritime pine, moss, pea, potato, pumpkin, rice, sorghum, soybean, sweet potato, tobacco and wheat (Supplementary Table S1). On the basis of previous phylogenetic analysis,^{18,22,26} the *Dof* sequence from *C. reinhardtii* was considered as a common *Dof* ancestor and was employed to root the tree. The phylogenetic tree generated by using full alignments produced high nodal support values (Figure 3). The tree topology

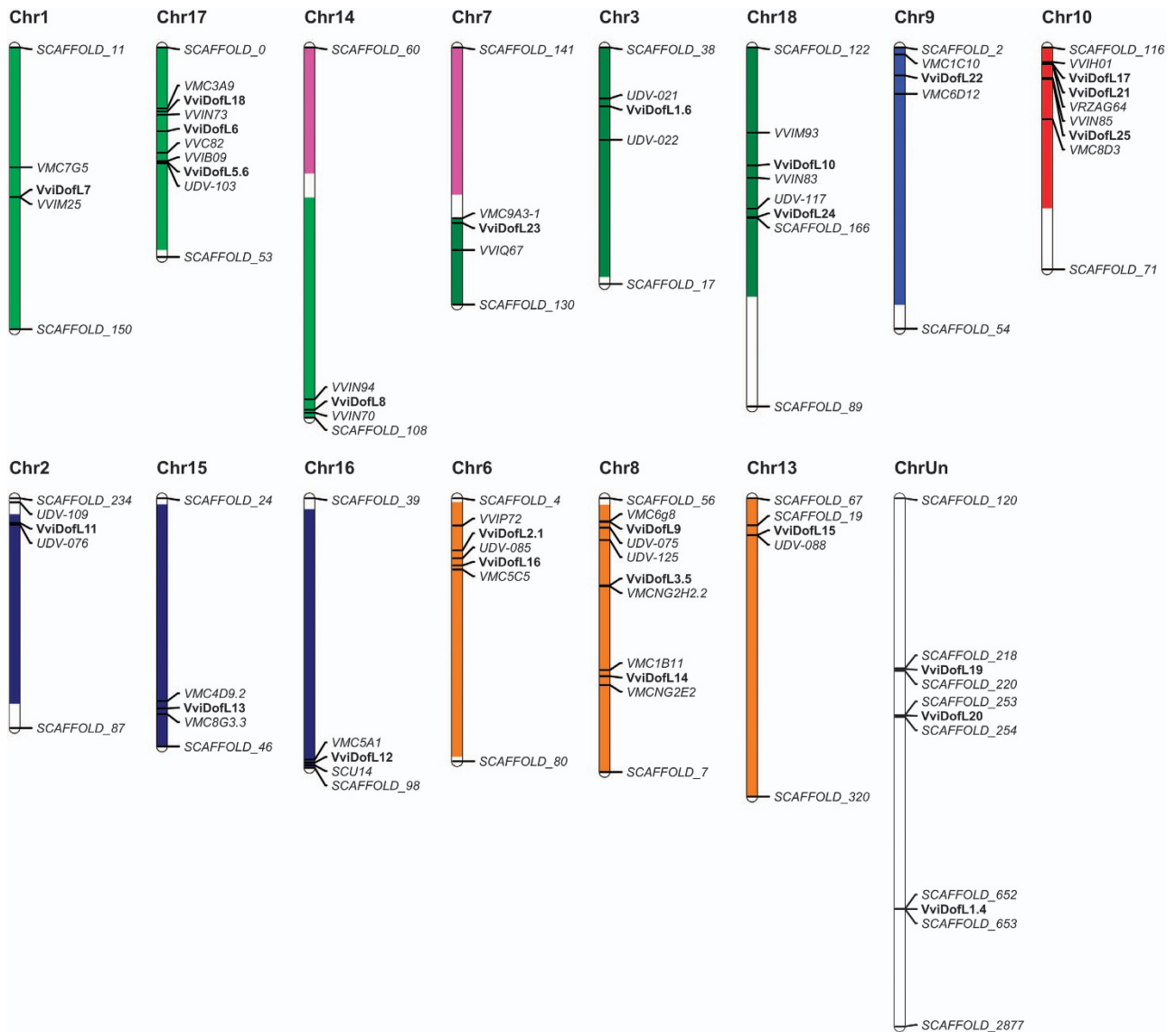


Figure 2. Chromosomal location of 25 *VviDofL* genes. Paralogous regions in the putative ancestral constituents of the grapevine genome are depicted in the color scheme following Jaillon *et al.*⁴⁰ Molecular markers and scaffolds that help positioning genes are in italics.

obtained by the Bayesian Inference led to the identification of six major clusters of orthologous groups (MCOGs), numbered from I to VI. Interestingly, all MCOGs identified displayed at least one member from grapevine, poplar and tomato. Surprisingly, when curing the alignment, which converged to the Dof domain spanning region as the sole input sequence for the phylogenetic analysis, no trees with acceptable support values were produced (Supplementary Figure S4).

From the phylogenetic tree (Figure 3), 40 putative paralogs and 14 putative orthologs were identified among all 167 sequences analyzed. The majority of the paralogous pairs belonged to poplar (19) followed by Arabidopsis (10), tomato (9), barley and sorghum (one pair each). No paralogs were identified among grapevine sequences. Interestingly, eleven grapevine sequences formed clusters with paralogs from poplar (10) or tomato (1), namely *VviDofL1.4*, *5.6*, *7*, *9*, *10*, *11*, *12*, *16*, *17*, *20* and *23* (clades colored in black in Figure 3). Five putative orthologs were identified among grapevine: *VviDofL1.6*/*PpiDof5*, *VviDofL22*/*PtDof19*, *VviDofL21*/*PtDof20*, *VviDofL18*/*SiDof21* and *VviDofL19*/*SiDof31* (these clades are collapsed in Figure 3). Finally, *VviDofL2.1* formed a cluster with the orthologous pair *GmDof11*/*PsDof1*.

Expression analysis of *DofL* genes

The expression profiles of putative grapevine *VviDofL* genes were retrieved from a global transcriptomic atlas comprising 54 tissues, organs or developmental stages.⁴⁹ Overall, higher transcript accumulation was found in rachis, tendrils, buds and whole inflorescences, while low levels were identified in petals, stamens, pollen and berry skins. Hierarchical clustering analysis was carried out in order to find groups of *VviDofL* genes with similar transcript level profiles across samples. Groups were mostly composed of samples originated from the same tissue/organ (Figure 4). However, *VviDofL* expression patterns in buds, stems and berries were not totally clustered. A clear distinction could be observed between green/vegetative and mature/woody tissues and organs. In addition, senescing leaves presented marked differences in the patterns of *VviDofL* gene expression relatively to other leaf samples.

Seven major patterns of *VviDofL* mRNA accumulation were identified through samples (A–G in Figure 4). Genes from group A presented higher transcript levels in several berry stages (mainly ripening to mid-ripening). In addition, *VviDofL25* showed high expression in pollen, root, seedling and senescing leaves. Genes

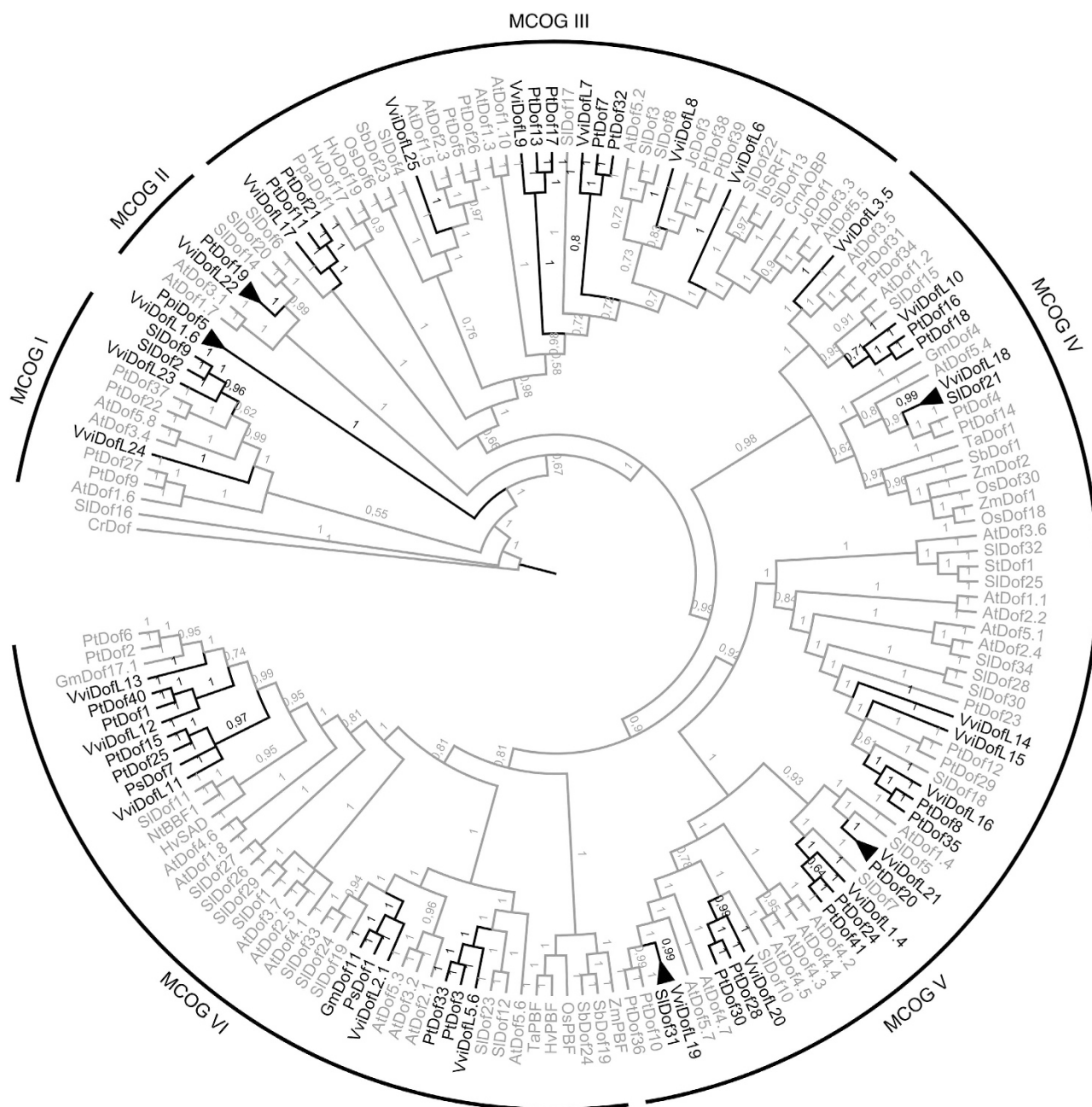


Figure 3. Phylogenetic tree of the full-length alignment of 167 *Dof* protein sequences. The resulting major clusters of orthologous genes (MCOGs) were numbered from I to VI. The tree was inferred using MrBayes v.3.2.6.⁴⁷ Branch support is given by posteriori probability values shown next to the corresponding nodes (when > 0.50). All accession codes used in the phylogenetic analysis are depicted in Supplementary Table S1. Clades comprising grapevine orthologs are collapsed; clades of grapevine proteins clusterized with paralogs or orthologs of other species are colored in black. At, *Arabidopsis thaliana*; Cm, *Cucurbita maxima*; Cr, *C. reinhardtii*; Gm, *Glycine max*; Hv, *Hordeum vulgare*; Ib, *Ipomoea batatas*; Jc, *J. curcas*; Nt, *Nicotiana tabacum*; Os, *Oriza sativa*; Ppa, *P. patens*; Ppi, *Pinus pinaster*; Ps, *Pisum sativum*; Pt, *Populus trichocarpa*; Sb, *Sorghum bicolor*; Sl, *Solanum lycopersicum*; St, *Solanum tuberosum*; Ta, *Triticum aestivum*; Vv, *V. vitifera*; Zm, *Zea mays*.

from group B were mainly induced in inflorescences, buds and seeds at the stages of post-fruit set to ripening. Interestingly, genes from group C showed low transcript amounts in all berry stages, except at the post-fruit set stage. Moreover, these genes were induced in tendrils, rachis and bud stages. Transcripts from this group were presented in low levels in all four seed stages analyzed, with the exception of *VvDofL20*, *VvDofL22* and *VvDofL24* that were induced in seed at the fruit set and post-fruit set stages. Considering the three genes that integrate group D, only *VvDofL5.6* and *VvDofL16* were induced in rachis and berries from véraison to ripening stages, with *VvDofL5.6* showing

the highest expression levels. Genes from group E and F showed minor expression variation in the samples analyzed. However, *VvDofL10* was highly induced in petals, stamens and flowers. Finally, genes from group G were mainly induced in berry stages (véraison to post-harvest withering III), with *VvDofL13* and *VvDofL14* showing the highest expression levels. In summary, a broad expression pattern of *VvDofL* genes was found across the 54 grapevine developmental stages analyzed.

In order to complement the whole-transcriptome data, steady-state mRNA levels of all *VvDofL* genes were investigated by real-time PCR in nine grapevine organs and tissues of the Pinot Noir

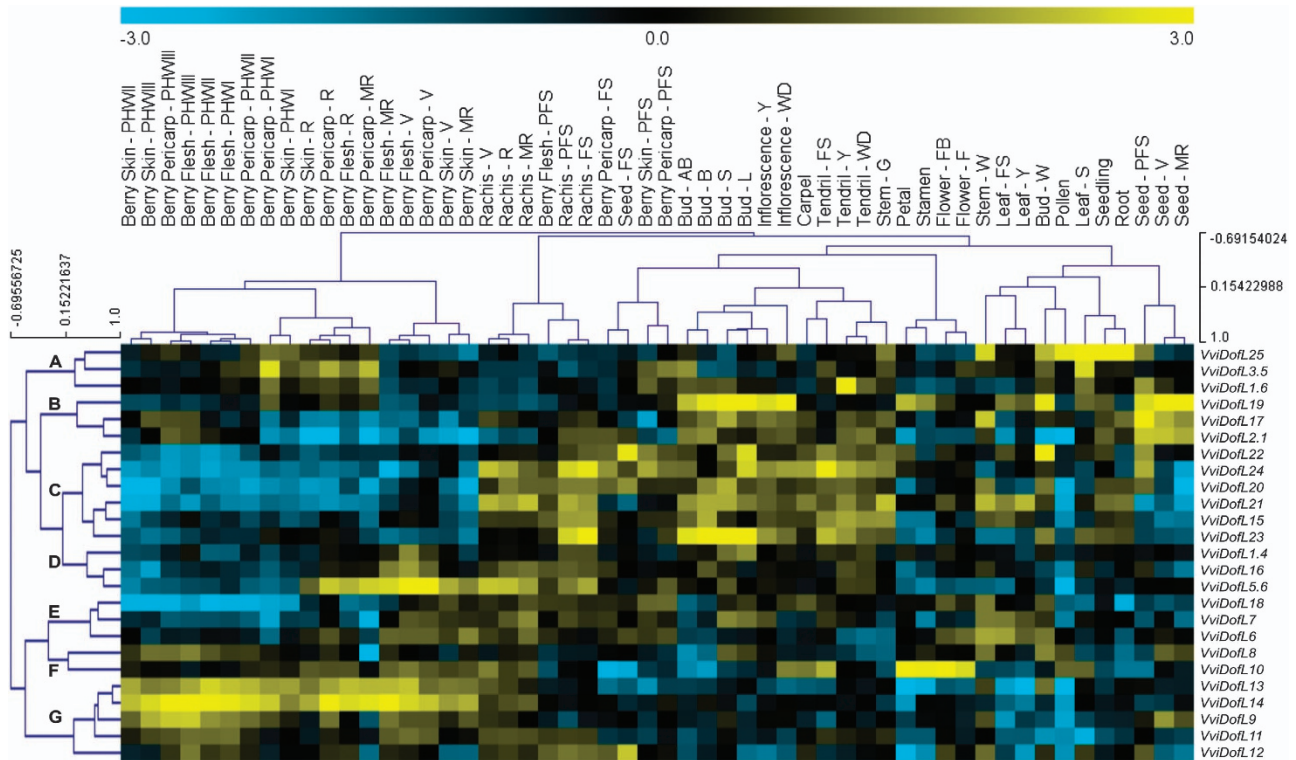


Figure 4. Transcript levels of *VviDofL* genes across 54 several grapevine cv Corvina tissues and organs. The color scheme used to represent expression level is yellow/blue: black boxes indicate low variation in gene expression, blue boxes indicate a fold decrease, and yellow boxes indicate a fold increase in relation to the mean value. Samples and genes were hierarchically clustered based on the average Pearson's distance, resulting in seven major patterns of *VviDofL* mRNA levels (A–G). Abbreviations in the sample headline: for buds: AB, after-burst; B, bud burst; L, latent bud; S, bud swell; W, winter bud; for inflorescences: WD, well developed; Y, young; for flowers: F, flowering; FB, flowering begins; for tendrils: FS, mature; WD, well developed; Y, young; for leaves: FS, mature; S, senescing; Y, young; for berry pericarp, berry skin, berry flesh, seed and rachis: FS, fruit set; MR, mid-ripening; PFS, post-fruit set; PHWI, post-harvest withering (first month); PHWII, post-harvest withering (second month); PHWIII, post-harvest withering (third month); R, ripening; V, véraison; for stems: G, green; W, woody. Data obtained as described in Fasoli *et al.*⁴⁹

grapevine variety. The overall grapevine *Dof* gene expression was quite diverse in these samples (Figure 5). Eleven *VviDofL* genes exhibited their highest expressions levels in inflorescences (*VviDofL3.5*, *VviDofL8*, *VviDofL10*, *VviDofL14*, *VviDofL16*, *VviDofL17*, *VviDofL18*, *VviDofL19*, *VviDofL20*, *VviDofL21*, and *VviDofL24*), three in stems (*VviDofL6*, *VviDofL7* and *VviDofL23*), two in stems and inflorescences (*VviDofL1.4* and *VviDofL2.1*), one gene in summer buds and inflorescences (*VviDofL22*), one in inflorescences, pré-véraison and véraison berries (*VviDofL5.6*), one in roots (*VviDofL25*), one in pré-véraison berries (*VviDofL1.6*), and one in véraison berries (*VviDofL13*). The genes with the highest expression specificity in the grapevine organs tested were *VviDofL6*, *VviDofL8*, *VviDofL10*, *VviDofL19* and *VviDofL25*. None of the *VviDofL* genes seemed to be specific to tendrils or leaves, or to have any pronounced level of mRNA accumulation in these organs. After several trials of reamplification, no measurable amplicons for *VviDofL9*, *VviDofL11*, *VviDofL12* and *VviDofL15* was obtained (Figure 5).

DISCUSSION

Velasco *et al.*³⁹ identified 62 families of transcription factors in the consensus genome sequence of the heterozygous grapevine 'Pinot Noir' and, among them, *Dof* transcription factors. Corroborating those findings, we were able to identify, phylogenetically characterize, and describe the patterns of steady-state mRNA levels of 25 grapevine *Dof* genes. Among the seeded plant genomes annotated for *Dof* genes so far, grapevine, sugarcane and pine showed the smallest numbers.^{21,22} Although 25 putative *Dof* genes were found in grapevine and sugarcane,³³ only 10 were identified

in pine.^{27,30} Other species showed more members for this gene family, ranging from 26 in barley to 78 in soybean.^{21,22,26,32}

In the present work, the four characteristic cysteine residues for zinc docking, as well as other well conserved sequences, were identified in all 25 grapevine *DofL* proteins, suggesting that their *Dof* domains are functional (Supplementary Figure S1). Additionally, the predicted physical localization of *DofL* genes in grapevine chromosomes was presented (Figure 2). In addition to the *Dof* domain, a plethora of other motifs is shared among *DofL* proteins (Figure 1). This finding was already observed in several other reports.^{22,24–26,29–31} The presence, absence or position of these motifs may affect the functional role of each *Dof* transcription factor. However, no similarity with known domains was found when comparing these sequences with the Pfam or SMART databases. Surprisingly, the comparison of the 10 grapevine non-*Dof* motifs identified (Supplementary Figure S3) with non-*Dof* motifs of other studies^{22,24–26,29–31} revealed similarities between the studies. This finding reinforces the need to conduct further investigations to unveil the functional roles of these still uncharacterized domains.

Comparative genomic studies are able to track characteristic features in multiple genomes. The identification of orthologous and paralogous sequences, along with functional information, can be useful tools to predict gene function.^{57,58} Moreover, gene duplication events are a common feature during genome and gene family evolution, and this seems to be the case in the *Dof* expansion. Moreno-Risueno *et al.*²⁶ suggested that duplication events from an ancestral *Dof* gene would have triggered the expansion of this family, resulting in neo-, sub- and

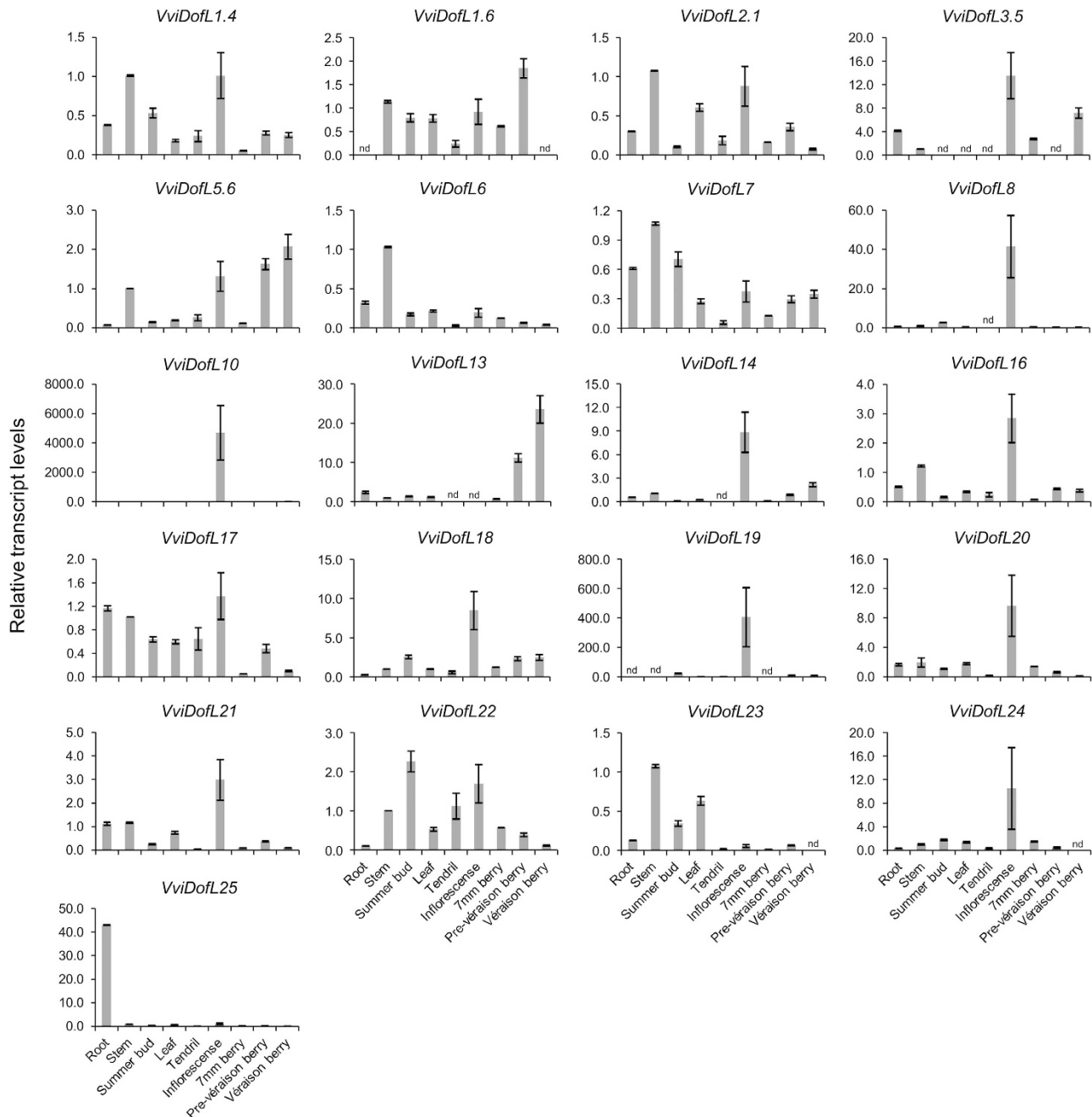


Figure 5. Expression profiles of *VviDofL* genes in nine grapevine cv Pinot Noir vegetative and reproductive organs by real-time PCR. Relative transcript levels in stems were set to 1, except *VviDofL19* whose leaf samples were set to 1. ND, non-detected. Standard error bars are shown.

pseudogenization processes that likely lead to their high range of functions. Indeed, *Dof* genes are exclusive of the green lineage, increasing from a single-copy gene in *C. reinhardtii* to a multicopy family with numerous functions in seeded plants.¹⁸ In this scenario, the identification of paralogous and orthologous genes along with functional information about *Dof* family members can be used to infer gene function, although further investigations are required to confirm these predictions.

Dof sequences from grapevine, *Arabidopsis*, poplar and tomato, in addition to *Dof* proteins already described in other species, were compared using a phylogenetic approach with the aim of improving the knowledge about the function of the *VviDofL* genes (Figure 3). The results fit well with previous reports, given that our analysis rendered the same 22 *Dof* paralogous and orthologous

pairs identified by Yang *et al.*,²⁵ 11 identified by Yanagisawa¹ and 7 identified by Cai *et al.*³¹ It's worth to mention that several other pairs identified by these groups formed new paralogs or orthologs in our set of sequences. This finding can be explained by the large set of sequences used in our analysis. As one could expect in a comparative genome analysis, the number of orthologous sequences found between grapevine and poplar (12 pairs; Figure 3) was higher than the one found between poplar and rice (four pairs in Yang *et al.*²⁵). Additionally, the strategy of comparing *Dof* genes from species with established functions yielded four pairs of orthologs besides the ones with poplar (Figure 3). These *Dof* genes were already characterized in maritime pine and tomato and may be an important resource to better understand the *VviDofL* functions in grapevine. Interestingly,

grapevine sequences did not form paralogs suggesting that neofunctionalization events could be held accountable in the expansion of this gene family in grapevine. As several authors have previously observed,^{24–26} the *Dof* domain is extremely well conserved among proteins and, for phylogenetic comparisons, the only strategy that generate consistent trees was based on sequences outside the *Dof* domain. In accordance, curing of the protein alignment was not effective in producing better-structured trees given that the Bayesian algorithm was not able to branch the tree (Supplementary Figure S4). Lijavetzky *et al.*²⁴ recognized that several clusters remained with poor supporting values in their phylogenetic analysis as the consequence of using only the conserved *Dof* domain (50 amino acids in length) in alignments.

Genome-wide studies describing the transcript accumulation patterns of *Dof* genes are found in the literature, reporting differential expression of *Dof* genes among tissues, organs or in response to abiotic stresses,^{25,26,28–31,34–37,59} but little information is available for grapevine *Dof* genes.²³ In this work, by exploring genome-wide analysis by microarray in combination with real-time PCR, specific *VviDofL* transcriptional patterns were found. Transcriptome data obtained from various grapevine samples showed different *VviDofL* expression signatures according to the tissue/organ tested (Figure 4). Interestingly, some tissues/organs and mature/woody samples tended to display transcript accumulation profiles considerably distant from their vegetative/green counterparts. This trend was also observed for the whole set of organ-specific grapevine genes when screened for mRNA levels in the same samples.⁴⁹ Furthermore, several expression profiles were obtained by real-time PCR (Figure 5) and support the microarray data. This finding is very interesting given that the plant material used in the two approaches is highly divergent, suggesting their involvement in similar processes regardless the cultivar. Finally, these results indicate that functional roles are being played by *VviDofL* genes during specific developmental stages, especially in reproductive tissues.

A recent review on the subject highlighted the importance of *Dof* genes during three main processes: metabolism regulation, seed development and tissue differentiation.²² The whole transcriptome data strongly suggest a clear relationship between *VviDofL* genes and the development of berries, seeds and flowers. Although most *VviDofL* genes presented low levels of transcripts during berry development (Figure 4), genes from group G, mainly *VviDofL13* and *VviDofL14*, were highly induced in berries from véraison to post-harvest withering III stages. In addition, *VviDofL3.5* (group A) displayed higher expression levels in berries at mid-ripening to ripening stages, and *VviDofL5.6* (group D) showed high transcript levels in berries from véraison to ripening stages. Similar results were obtained by real-time PCR for *VviDofL5.6* and *VviDofL13* (Figure 5). Costenaro-da-Silva *et al.*²³ reported a *DofL* gene (*VviDofL18* in this work) being induced during berry developmental stages 2–6 weeks after fruit set; however, this relationship still lacks an in-depth characterization. Berry development in grapevine is a finely tuned complex process, and the results described in this work suggest that *VviDofL3.5*, *5.6*, *13* and *14* are important players in the regulation of berry development in grapevine. Complementarily, there is a widely accepted positive correlation between berry weight and seed content in grapevine segregating populations, possibly due to the action of growth regulators produced by seeds.⁶⁰ In this respect, seeds can influence berry development, also having important roles in enological traits, given that seeds are one of the major providers of tannins during berry growth.⁶¹ Indeed, several *VviDofL* genes were highly induced in seed samples and they may be playing roles during seed development. This seems to be the case of genes from group B and genes *VviDofL20*, *VviDofL22* and *VviDofL24* from group C (Figure 4). These genes presented a remarkable accumulation of transcripts in all seed stages analyzed

(fruit set to mid-ripening) and might be acting together in seed development. Interestingly, *VviDofL2.1* from group B, which was almost exclusively expressed in seeds from post-fruit set to mid-ripening stages (Figure 4), exhibited orthology with *GmDof11* and *PsDof1* (Figure 3). *GmDof11*, along with *GmDof4*, is involved in the regulation of genes associated with the synthesis of fatty acids, controlling the accumulation of lipids in soybean seeds.⁵⁹ It is tempting to speculate that *GmDof11* and *VviDofL2.1* share conserved functions, although functional studies are needed to confirm or refute these statements. Furthermore, several functional studies reported the involvement of *Dof* proteins in the regulation of seed content, germination and/or development. *Dof* proteins from barley (*HvSAD* and *HvPBF*), maize (*ZmPBF*) and rice (*OsPBF*) were shown to activate several genes during seed development that are responsible for 15% of the total endosperm protein content.^{10–14} These *Dof* transcription factors also participate on the mobilization of storage compounds in order to provide nutrients for seed germination and seedling growth.⁸ In addition, *DAG1* (*Dof Affecting Germination 1*) and *DAG2* act on a maternal switch that is responsible for controlling seed germination in *Arabidopsis*.^{3–6} In this context, the transcriptional data gathered in this work suggest that several *VviDofL* genes are playing important roles during the course of berry and seed development.

Eleven *VviDofL* genes were mainly expressed in inflorescences when screened for mRNA levels by real-time PCR (Figure 5), and *VviDofL10*, *VviDofL19* and *VviDofL24* also presented flowering-specific transcription profiles in the whole transcriptome data (Figure 4). The identification of similar expression profiles between microarray and real-time PCR approaches, comparing samples from different cultivars, orchard management practices and climate growth conditions, strongly suggest that these genes have important regulatory roles during flowering that may be evolutionally conserved between grapevine varieties. In the literature, few reports about *Dof* functions during flowering are found,^{19,20,62} although some *Dofs* were described by their involvement in daylength regulation of flowering through interactions with *CONSTANS*.^{15–18} However, inside a subclade of MCOG V, *VviDofL19* is very close to *AtDof4.7* (Figure 3), which is involved in the control of floral abscission as part of a transcription complex that directly regulates the expression of cell wall hydrolysis enzymes.¹⁹ Moreover, *ZmDof1* from maize was described as a transcriptional repressor of pollen development.²⁰ Finally, flower-specific expression of *Dof* genes was already identified for *AtDof4.2*, which negatively and positively regulates the transcription of flavonoid and hydroxycinnamic acid production, respectively.⁶² Taken together, our findings strongly suggest that several *VviDofLs* might be acting in specific functions during flower development. Although promising, the genes herein identified still need further investigations in order to unveil their functional roles.

In conclusion, we identified the probable full set of *DofL* genes in grapevine. The phylogenetic analysis resulted in the identification of seven MCOGs that contains members from 18 plant species. Our results confirm that recurrent duplications and diversification from an original *Dof* ancestor led to the formation of this complex family of transcription factors specific to Viridiplantae. In view of important genome duplication events leading to gene redundancy in the history of plant diversification (see Jaillon *et al.*⁴⁰ for example), a combination of phylogenetically inferred relationships with functional data will be essential to effectively establish conserved and diverged roles of the *DofL* gene family in grapevine and in other plant species. Finally, our expression data revealed the involvement of *VviDofL* genes mainly in berry, seed and flower development in grapevine. In this work, we gathered consistent results that turn the *VviDofL* genes into good candidates to better understand the regulatory network associated with flowering and berry development.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Dr Roberto C Togawa for providing us with the access to the Bioinformatics Laboratory at Embrapa Recursos Genéticos e Biotecnologia. This work was supported by 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior' (CAPES, Brazil), 'Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil; Proc. 475440/2004-8; GP research fellowship 311361/2009-9), 'Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul' (FAPERGS, Brazil; Proc. 04/1184.6) and Embrapa (Brazil).

REFERENCES

- Yanagisawa S. The Dof family of plant transcription factors. *Trends Plant Sci* 2002; **7**: 555–560.
- Le Hir R, Bellini C. The plant-specific Dof transcription factors family: new players involved in vascular system development and functioning in Arabidopsis. *Front Plant Sci* 2013; **4**: 164.
- Gabriele S, Rizza A, Martone J, Circelli P, Costantino P, Vittorioso P. The Dof protein DAG1 mediates PIL5 activity on seed germination by negatively regulating GA biosynthetic gene AtGA3ox1. *Plant J* 2010; **61**: 312–323.
- Gualberti G, Papi M, Bellucci L, Ricci I, Bouchez D, Camilleri C *et al.* Mutations in the Dof zinc finger genes DAG2 and DAG1 influence with opposite effects the germination of Arabidopsis seeds. *Plant Cell* 2002; **14**: 1253–1263.
- Papi M, Sabatini S, Bouchez D, Camilleri C, Costantino P, Vittorioso P. Identification and disruption of an Arabidopsis zinc finger gene controlling seed germination. *Genes Dev* 2000; **14**: 28–33.
- Santopolo S, Boccaccini A, Lorrain R, Ruta V, Caputo D, Minutello E *et al.* DOF AFFECTING GERMINATION 2 is a positive regulator of light-mediated seed germination and is repressed by DOF AFFECTING GERMINATION 1. *BMC Plant Biol* 2015; **15**: 72.
- Isabel-LaMoneda I, Diaz I, Martinez M, Mena M, Carbonero P. SAD: a new DOF protein from barley that activates transcription of a cathepsin B-like thiol protease gene in the aleurone of germinating seeds. *Plant J* 2003; **33**: 329–340.
- Martinez M, Rubio-Somoza I, Fuentes R, Lara P, Carbonero P, Diaz I. The barley cystatin gene (*lcy*) is regulated by DOF transcription factors in aleurone cells upon germination. *J Exp Bot* 2005; **56**: 547–556.
- Rueda-Romero P, Barrero-Sicilia C, Gómez-Cadenas A, Carbonero P, Oñate-Sánchez L. Arabidopsis thaliana DOF6 negatively affects germination in non-after-ripened seeds and interacts with TCP14. *J Exp Bot* 2012; **63**: 1937–1949.
- Diaz I, Martinez M, Isabel-LaMoneda I, Rubio-Somoza I, Carbonero P. The DOF protein, SAD, interacts with GAMYB in plant nuclei and activates transcription of endosperm-specific genes during barley seed development. *Plant J* 2005; **42**: 652–662.
- Dong G, Ni Z, Yao Y, Nie X, Sun Q. Wheat Dof transcription factor WPBF interacts with TaQM and activates transcription of an alpha-gliadin gene during wheat seed development. *Plant Mol Biol* 2007; **63**: 73–84.
- Hwang Y-S, Ciceri P, Parsons RL, Moose SP, Schmidt RJ, Huang N. The maize O2 and PBF proteins act additively to promote transcription from storage protein gene promoters in rice endosperm cells. *Plant Cell Physiol* 2004; **45**: 1509–1518.
- Marzábal P, Gas E, Fontanet P, Vicente-Carbajosa J, Torrent M, Ludevid MD. The maize Dof protein PBF activates transcription of gamma-zein during maize seed development. *Plant Mol Biol* 2008; **67**: 441–454.
- Yamamoto MP, Onodera Y, Touno SM, Takaiwa F. Synergism between RPBF Dof and RISBZ1 bZIP activators in the regulation of rice seed expression genes. *Plant Physiol* 2006; **141**: 1694–1707.
- Fornara F, Panigrahi KCS, Gissot L, Sauerbrunn N, Ruhl M, Jarillo JA *et al.* Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Dev Cell* 2009; **17**: 75–86.
- Imaizumi T, Schultz TF, Harmon FG, Ho LA, Kay SA. FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. *Science* 2005; **309**: 293–297.
- Li D, Yang C, Li X, Gan Q, Zhao X, Zhu L. Functional characterization of rice OsDof12. *Planta* 2009; **229**: 1159–1169.
- Lucas-Reina E, Romero-Campero FJ, Romero JM, Valverde F. An evolutionarily conserved DOF-CONSTANS module controls plant photoperiodic signaling. *Plant Physiol* 2015; **168**: 561–574.
- Wei P-C, Tan F, Gao X-Q, Zhang X-Q, Wang G-Q, Xu H *et al.* Overexpression of AtDOF4.7, an Arabidopsis DOF family transcription factor, induces floral organ abscission deficiency in Arabidopsis. *Plant Physiol* 2010; **153**: 1031–1045.
- Chen X, Wang D, Liu C, Wang M, Wang T, Zhao Q *et al.* Maize transcription factor Zmdof1 involves in the regulation of Zm401 gene. *Plant Growth Regul* 2012; **66**: 271–284.
- Gupta S, Malviya N, Kushwaha H, Nasim J, Bisht NC, Singh VK *et al.* Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. *Planta* 2015; **241**: 549–562.
- Noguero M, Atif RM, Ochat S, Thompson RD. The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. *Plant Sci* 2013; **209**: 32–45.
- Costenaro-da-Silva D, Passaia G, Henriques JAP, Margis R, Pasquali G, Revers LF. Identification and expression analysis of genes associated with the early berry development in the seedless grapevine (*Vitis vinifera* L.) cultivar Sultanine. *Plant Sci* 2010; **179**: 510–519.
- Lijavetzky D, Carbonero P, Vicente-Carbajosa J. Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC Evol Biol* 2003; **3**: 17.
- Yang X, Tuskan GA, Cheng Z-M. Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. *Plant Physiol* 2006; **142**: 820–830.
- Moreno-Risueno MA, Martínez M, Vicente-Carbajosa J, Carbonero P. The family of DOF transcription factors: from green unicellular algae to vascular plants. *Mol Genet Genomics* 2007; **277**: 379–390.
- Shigyo M, Tabei N, Yoneyama T, Yanagisawa S. Evolutionary processes during the formation of the plant-specific Dof transcription factor family. *Plant Cell Physiol* 2007; **48**: 179–185.
- Shaw LM, McIntyre CL, Gresshoff PM, Xue G-P. Members of the Dof transcription factor family in *Triticum aestivum* are associated with light-mediated gene regulation. *Funct Integr Genomics* 2009; **9**: 485–498.
- Hernando-Amado S, González-Calle V, Carbonero P, Barrero-Sicilia C. The family of DOF transcription factors in Brachypodium distachyon: phylogenetic comparison with rice and barley DOFs and expression profiling. *BMC Plant Biol* 2012; **12**: 202.
- Rueda-López M, García-Gutiérrez A, Cánovas FM, Ávila C. The family of Dof transcription factors in pine. *Trees* 2013; **27**: 1547–1557.
- Cai X, Zhang Y, Zhang C, Zhang T, Hu T, Ye J *et al.* Genome-wide analysis of plant-specific Dof transcription factor family in tomato. *J Integr Plant Biol* 2013; **55**: 552–566.
- Guo Y, Qiu L-J. Genome-wide analysis of the Dof transcription factor gene family reveals soybean-specific duplicable and functional characteristics. *PLoS One* 2013; **8**: e76809.
- Gupta S, Kushwaha H, Singh VK, Bisht NC, Sarangi BK, Yadav D. Genome wide in silico characterization of Dof transcription factor gene family of sugarcane and its comparative phylogenetic analysis with Arabidopsis, rice and sorghum. *Sugar Tech* 2014; **16**: 372–384.
- Ma J, Li M-Y, Wang F, Tang J, Xiong A-S. Genome-wide analysis of Dof family transcription factors and their responses to abiotic stresses in Chinese cabbage. *BMC Genomics* 2015; **16**: 33.
- Chen Y, Cao J. Comparative analysis of Dof transcription factor family in maize. *Plant Mol Biol Report* 2015; **33**: 1245–1258.
- Venkatesh J, Park SW. Genome-wide analysis and expression profiling of DNA-binding with one zinc finger (Dof) transcription factor family in potato. *Plant Physiol Biochem* 2015; **94**: 73–85.
- Huang W, Huang Y, Li M, Wang F, Xu Z, Xiong A. Dof transcription factors in carrot: genome-wide analysis and their response to abiotic stress. *Biotechnol Lett* 2016; **38**: 145–155.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–410.
- Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D *et al.* A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One* 2007; **2**: e1326.
- Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A *et al.* The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 2007; **449**: 463–467.
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C *et al.* The Pfam protein families database. *Nucleic Acids Res* 2012; **40**: D290–D301.
- Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 2002; **93**: 77–78.
- Bailey TL, Boden M, Buske F a, Frith M, Grant CE, Clementi L *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 2009; **37**: W202–W208.
- Leticia I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* 2012; **40**: D302–D305.
- Chou KC, Shen HB. Plant-mPLoc: A top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One* 2010; **5**: e11335.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**: 4673–4680.

- 47 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S *et al.* MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 2012; **61**: 539–542.
- 48 Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**: 540–552.
- 49 Fasoli M, Dal Santo S, Zenoni S, Tornielli GB, Farina L, Zamboni A *et al.* The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. *Plant Cell* 2012; **24**: 3489–3505.
- 50 Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA *et al.* TM4 microarray software suite. *Methods Enzymol* 2006; **411**: 134–193.
- 51 Coombe BG. Growth stages of the grapevine: Adoption of a system for identifying grapevine growth stages. *Aust J Grape Wine Res* 1995; **1**: 104–110.
- 52 Reid KE, Olsson N, Schlosser J, Peng F, Lund ST. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biol* 2006; **6**: 27.
- 53 Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 2001; **25**: 402–408.
- 54 Grimplet J, Adam-Blondon A-F, Bert P-F, Bitz O, Cantu D, Davies C *et al.* The grapevine gene nomenclature system. *BMC Genomics* 2014; **15**: 1077.
- 55 Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* 2009; **19**: 327–335.
- 56 Rambaldi D, Ciccarelli FD. FancyGene: Dynamic visualization of gene structures and protein domain architectures on genomic loci. *Bioinformatics* 2009; **25**: 2281–2282.
- 57 Thornton JW, Desalle R. Gene family evolution and homology: genomics meets phylogenetics. *Annu Rev Genomics Hum Genet* 2000; **1**: 41–73.
- 58 Zhang J. Evolution by gene duplication: An update. *Trends Ecol Evol* 2003; **18**: 292–298.
- 59 Wang H-W, Zhang B, Hao Y-J, Huang J, Tian A-G, Liao Y *et al.* The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic Arabidopsis plants. *Plant J* 2007; **52**: 716–729.
- 60 Doligez A, Bertrand Y, Farnos M, Grolier M, Romieu C, Esnault F *et al.* New stable QTLs for berry weight do not colocalize with QTLs for seed traits in cultivated grapevine (*Vitis vinifera* L.). *BMC Plant Biol* 2013; **13**: 217.
- 61 Kennedy JA, Troup GJ, Pilbrow JR, Hutton DR, Hewitt D, Hunter CR *et al.* Development of seed polyphenols in berries from *Vitis vinifera* L. cv. Shiraz. *Aust J Grape Wine Res* 2000; **6**: 244–254.
- 62 Skirydz A, Jozefczuk S, Stobiecki M, Muth D, Zanor MI, Witt I *et al.* Transcription factor AtDOF4.2 affects phenylpropanoid metabolism in Arabidopsis thaliana. *New Phytol* 2007; **175**: 425–438.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016

Supplementary Information for this article can be found on the *Horticulture Research* website (<http://www.nature.com/hortres>)