GMR

The Original by FUNPEC-RP

Comparative analysis and expression of Dof genes in *Citrus sinensis* during fruit development

L.M. Guaberto, T.B. dos Santos, A.F. Ribas and L.G.E. Vieira

Programa de Pós-Graduação em Agronomia, Universidade do Oeste Paulista, Presidente Prudente, SP, Brasil

Corresponding author: T.B. dos Santos
E-mail: tiagobio02@yahoo.com.br

Genet. Mol. Res. 18 (2): gmr18210
Received November 14, 2018
Accepted April 08, 2019
Published April 30, 2019
DOI http://dx.doi.org/10.4238/gmr18210

ABSTRACT. The DNA-binding One Finger (Dof) proteins are a large group of plant-specific transcription factors (TFs) that participate in several biological processes in plants. Although the Dof gene family has been studied in many plant species, detailed information on these genes and their fruit-specific expression during fruit development are not yet available in sweet orange (*Citrus sinensis* L. Osbeck). Here, we identified and characterized 24 *CsDof* genes in *C. sinensis* genome. Phylogenetic analysis showed that orange Dof proteins clustered into four major groups (A, B, C and D) and nine subgroups (A, B1, B2, C1, C2.1, C2.2, C3, D1, D2) based on the established *Arabidopsis thaliana* classification. The predicted *CsDof* genes could be mapped on all chromosomes, except for chromosomes 4 and 9. Most of these genes lacked introns or possessed just one intron. Based on freely RNA-seq data, 12 *CsDofs* with higher transcript abundance in fruits were selected for further analysis by semi-quantitative RT-PCR. This analysis revealed that the *CsDof* genes exhibit a variety of expression patterns during the early stages of fruit development (up to 90 days after anthesis), making it possible to establish three groups regarding their transcriptional activity. Among them, the *CsDof17* showed the highest expression in all sampling stages investigated, indicating that this isoform may play an important role in regulating the early development of sweet orange fruits. Our results provide some useful information for the utilization of the *CsDof* genes for crop improvement of this important fruit species.

Key words: Dof; Sweet orange; Fruit development; Expression profiles

INTRODUCTION

Citrus is one of the world's most important fruit crops. In spite of intensive genetic selection for better fruits in sweet orange (*Citrus sinensis* L. Osbeck), information about how these fruits develop and which genes control this process are still not clear in many respects. The changes during the development of the citrus fruits is highly coordinated and involves a series of physiological, biochemical and molecular changes allowing the development of an edible mature fruit (Wang et al., 2017). Citrus fruit development and ripening follow a typical sigmoid growth curve, characterized by two stages of slow growth with an interstitial period of rapid growth (Bain, 1958). The initial stage (I) after anthesis is characterized by slow growth with an intense cell division and may last approximately two-month. Subsequently, the fruit goes through a rapid growth period due to rapid cell enlargement, mainly of albedo cells, and water accumulation (Stage II). During maturation (Stage III), the fruit growth may continue but is considerably lower than in Stage II.

Studies of genes expressed in non-climacteric fruits, such as citrus, are important to understand the regulation of the development and maturation of these fruits. With the availability of annotated genome sequences of *C. sinensis* (Xu et al., 2013) it has become possible to analyze the genomic structure and function of genes families that have critical impact on biological and evolutionary processes in this important economic species. For example, it is known that during the development and ripening processes of citrus fruits, the transcriptional patterns of a large number of genes are specifically regulated by several transcription factors (TFs) (Wu et al., 2014). As such, being a highly controlled and complex biological process, fruit development in *C. sinensis* involve the participation of several families of TFs, such as WRKY, MADS, MYB, MYC, bHLH, HD-Zip and Dof (Ye et al., 2010; Wu et al., 2016b).

Dof domain proteins are plant-specific transcription factors with bifunctional binding activities with both DNA and proteins to regulate transcription in plant cells in response to developmental programs and environmental changes (Yanagisawa, 2004). Dof transcription factor family belongs to the class of zinc finger domains and typically contain between 200–400 amino acids with a highly conserved 52 Dof domain at the N-terminal with a single Cys2/Cys2 zinc finger structure in a conserved $CX_2CX_{21}CX_2C$ motif, which recognizes the specific cis-element of 5'-(AT)/AAAG-3' sites in DNA (Noguero et al., 2013). The Dof family has been studied widely in plants, such as *Arabidopsis* (Lijavetzky et al., 2003), pepper (Wu et al., 2016a), eggplant (Wei et al., 2018) among others. In addition, the Dof proteins were shown to be involved in a variety of biological processes, such as abiotic stress responses (Zhang et al., 2017; Corrales et al., 2017), seed germination (Santopolo et al., 2015), carbon and nitrogen metabolism (Kurai et al., 2011 and floral development (Cheng et al., 2018).

Compared with the comprehensive researches of this gene family in other plant species, to date no specific study have examined the Dof genes in sweet orange. In this work, we identified 24 members of the *CsDof* gene family in the genome of *C. sinensis*. Phylogenetic analysis revealed that the *CsDof* genes could be grouped into four major groups according to their gene structure and protein sequences. Additionally, we characterize the expression patterns of selected *CsDof* genes in the early stages of fruit development with semi-quantitative real-time reverse transcription polymerase chain reaction (RT-PCR) assays. Overall, this study provides the first genome-wide analysis of

the *C. sinensis* Dof gene family, and the results may provide useful insights for further investigation into the putative functions this family in the early fruit developmental stages.

MATERIAL AND METHODS

Identification and characterization of *CsDof* family members

In order to identify members of the Dof gene family in sweet orange, database searches were performed using the "keyword" option in the "*Citrus sinensis* - Annotation Project" database (Xu et al., 2013; CITRUS SINENSIS, 2016). Additionally, the retrieved sequences were individually compared to the sequences deposited and in the Phytozome database v.12 (https://phytozome.jgi.doe.gov/pz/portal.html/)]. These sequences were then used as queries in a BLAST search against sequences deposited in the NCBI database (National Centre for Biotechnology Information; http://www.ncbi.nlm.nih.gov/) to confirm their similarity with Dof genes of other species. The functional motifs or domains of Dof protein sequences (PF02701) were confirmed using the Pfam database (http://pfam.xfam.org/) and the ScanProsite tool (https://prosite.expasy.org/scanprosite/). Physicochemical parameters of the *CsDof* proteins were described using the ProtParam tool (http://web.expasy.org/protparam/), including the theoretical isoelectric point (pI), and the molecular weight (kDa). In addition, the grand average of hydropathicity was determined using the GRAVY calculator (http://www.gravy-calculator.de/) and the subcellular localization was predicted with Plant-mPLoc (Chou and Shen, 2010) (Table 1).

Table 1. Dof transcription factor genes identified and characterized in *Citrus sinensis*.

Citrus Dof	Phytozome Identifier	Protein length (aa)	kDa	pI	# Introns	S/a	GRAVY	Plant-mPLoc
*CsDof*1	orange1.1g046549m	250	27.11	8.59	0	0	-0.882	Nucleus
*CsDof*2	orange1.1g040161m	366	39.17	9.15	0	0	-0.528	Nucleus
*CsDof*3	orange1.1g024786m	262	27.23	8.49	0	0	-0.373	Nucleus
*CsDof*4	orange1.1g041420m	309	33.49	9.38	1	0	-0.679	Nucleus
*CsDof*5	orange1.1g021502m	311	33.99	9.38	0	0	-0.630	Nucleus
*CsDof*6	orange1.1g042156m	336	37.02	9.52	0	0	-1.038	Nucleus
*CsDof*7	orange1.1g047290m	172	20,08	8.88	0	0	-0.883	Nucleus
*CsDof*8	orange1.1g041099m	361	39.57	8.15	1	0	-0.674	Nucleus
*CsDof*9	orange1.1g025998m	245	27.54	4.66	1	0	-0.746	Nucleus
*CsDof*10	orange1.1g043493m	254	26.71	7.64	0	0	-0.473	Nucleus
*CsDof*11	orange1.1g021633m	310	34.87	7.06	1	0	-0.610	Nucleus
*CsDof*12	orange1.1g023047m	288	31.79	8.37	1.2	2	-0.734	Nucleus
*CsDof*13	orange1.1g011169m	492	53.23	6.26	1	1	-0.621	Nucleus
*CsDof*14	orange1.1g011752m	478	52.17	6.83	1	0	-0.729	Nucleus
*CsDof*15	orange1.1g017919m	364	38.36	9.06	1	0	-0.562	Nucleus
*CsDof*16	orange1.1g043617m	291	32.59	5.58	0	0	-0.657	Nucleus
*CsDof*17	orange1.1g011016m	495	54.82	7.24	1	0	-0.879	Nucleus
*CsDof*18	orange1.1g019668m	337	36.62	9.25	1	0	-0.622	Nucleus
*CsDof*19	orange1.1g036178m	330	36.09	6.83	0	0	-0.910	Nucleus
*CsDof*20	orange1.1g038013m	494	54.79	7.96	1	0	-0.943	Nucleus
*CsDof*21	orange1.1g020288m	328	36.62	8.87	1	0	-0.580	Nucleus
*CsDof*22	orange1.1g021974m	304	34.52	7.61	1	0	-0.769	Nucleus
*CsDof*23	orange1.1g024229m	270	30.15	8.81	2	0	-0.805	Nucleus
*CsDof*24	orange1.1g021641m	309	33.75	8.78	1	1	-0.708	Nucleus

kDa: Kilodalton.PI: Isoelectric point.S/a: alternative splicing.GRAVY: Grand average of hydropathicity index.Plant-mPLoc Predicting subcellular localization. Un* ("unassembled scaffold").

Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignment was performed using the CLC Main Workbench 8.0 program (CLC Bio, Aarhus, Denmark) to identify the conserved regions containing the Dof domain. To classify which group the Dof genes belonged to, a phylogenetic tree was generated using the Neighbor-Joining method using the program MEGA 7.0 and evaluated with bootstrap test of 1000 replicates for internal branch reliability (Tamura et al., 2013). The sequences of Dof proteins from *Arabidopsis thaliana* (*AtDofs*), according to the established classification and nomenclature of (Lijavetzky et al., 2003; Kushwaha et al., 2011), were used for tree constructions.

Chromosomal location and structure of *CsDof* genes

The mapping of the genes in the *C. sinensis* chromosomes and the numbers and positions of exons and introns were determined by comparing full-length cDNA sequences and the corresponding genomic DNA sequences of each gene using the information available in the genome database "*Citrus sinensis* - Annotation Project". All the identified putative Dof proteins were named as prefix “Cs” for *Citrus sinensis*, followed by Arabic numbers serially starting from 1 according to their chromosomal location. The genomic and CDS sequences were used for generated the exon/intron structure of the *CsDof* genes obtained from the online website Gene Structure Display Server (GSDS: http://gsds.cbi.pku.edu.cn/).

Plant Materials

To analyze the expression patterns of selected *CsDof* genes during fruit development, fruits samples were collected from adult plants of the mid-season cultivar (cv. Pêra) grafted on Rangpur lime (*Citrus limonia* Osbeck) cultivated at Universidade do Oeste Paulista (UNOESTE), Presidente Prudente - SP, Brazil, located at 22º 07 'S, 51º 27' W. After flowering, young fruits were identified on nine trees distributed in different positions in the experimental area for the purpose of having fruits reaching the same development stage at each sampling time. The fruits were collected individually from August to November, for a total of six sampling stages with 15-day intervals, resulting in a total harvest period of 90 days from the anthesis (Table 2). On each sampling date, three fruits were randomly collected at different positions of each tree, resulting in a total of 27 fruits for sample. In all experiments, the fruits were immediately frozen in liquid nitrogen and stored at -80°C freezer until the extraction of the total RNA.

Table 2. Diameter and fresh weight of fruits of *Citrus sinensis* (cv. Pêra) at the six sampling stages (days after flowering - DAF) used in this experiment.

Fruit sampling stages	Days after flowering (DAF)	Fruit diameter (cm) ± SD*	Fruit fresh weight (g) ± SD*
1	15	0.57 ± 0.05	0.64 ± 0.03
2	30	0.57 ± 0.05	2.1 ± 0.05
3	45	1.7 ± 0.05	6.0 ± 0,10
4	60	2.9 ± 0.11	13.0 ± 0.11
5	75	3.1 ± 0.5	18.2 ± 0.05
6	90	4.3 ± 0.11	36.5 ± 0.05

*Values are represented by means ± standard deviation (SD). (n = 5).

CsDof expression analysis

Out of a total of 24, 12 *CsDof* genes with higher normalized RPKM values in fruits relative to other tissues were selected from the "*Citrus sinensis* - Annotation Project" (Jiao et al. 2013; CITRUS SINENSIS, 2016) dataset for analysis of their expression profiles during the six sampling stages of fruit development (15, 30, 45, 60, 75 and 90 days after anthesis) (Table 2). For extraction of RNA, 300 mg of flavedo and albedo were pulverized in liquid nitrogen and the RNA was isolated using the Trizol Reagent procedure (Invitrogen®). After performing DNase (RNase-free, Invitrogen®) treatment in all samples, the RNA integrity was examined by 1% agarose gel electrophoresis, and concentration and purity were determined from the A_{260} and A_{260}/A_{280} ratio, respectively. Aliquots of 1 µg total RNA were used for first-strand cDNA synthesis using SuperScript III (Invitrogen®) and poly-A: oligo (dT) following the manufacturers recommendations in a final volume of 20 µL reaction. The cDNAs were stored at -20°C until used in the RT-PCR experiments. Three biological replicates of each sample were used for the semi-quantitative real-time reverse transcription polymerase chain reaction (RT-PCR) analysis.

Primers were designed by the Primer Express program (version 3.0 (Supplementary 1). Additionally, the primers were subjected to the Primer-BLAST tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) to check for non-specific annealing with *C. sinensis* sequences deposited at the NCBI. Preliminary RT-PCR reactions were performed to ensure that the primers could not amplify sequences other than the authentic target region. The optimum number of PCR cycles (15, 20, 25, 30, 35 and 40 cycles) and annealing temperatures (56, 57 and 58°C) were determined for each primer pair to avoid the saturation of DNA amplification. Based on that, the RT-PCR conditions used for the *CsDof* primers were 94°C for 5 min, followed by 35 cycles at 94°C for 1 min, 58°C for 1 min and 72°C for 2 min and a final extension of 72°C for 1 min. In the case of the reference gene used for normalization (*CsEF1-α*), only 20 cycles were used for amplification.

The amplicons were separated on 1% agarose gels, stained with ethidium bromide and imaged under UV light. The expression of *CsDofs* genes were calculated by densitometry using a computerized image analysis system. The captured images were processed using ImageJ 1.43 U software, essentially as described by Freschi et al. (2009). Transcript abundance of each *CsDof* genes was first compared with the expression of that genes in the leaves (control) and then normalized to the expression of the reference gene, *CsEF1-α* (Endo et al., 2006). Hence, the fold expression (FE) of each *CsDof* gene was calculated as FE = *CsDof* expression in the fruit/*CsDof* expression in the leaves divided by the expression of the reference gene *CsEF1-α* in the fruits/*CsEF1-α* in the leaves.

RESULTS

Identification and sequence analysis of *CsDof* genes

Twenty-four non-redundant sequences, designated as *CsDof1-CsDof24*, with the conserved Dof domain $CX_2CX_{21}CX_2C$ (Noguero et al., 2013) were identified from genome sequence analysis in the publicly available citrus databases. The predicted *CsDof*s were analyzed for sequence characteristics, subcellular localization and multiple sequence alignments. The deduced full-length amino acid sequence of the *CsDof*s ranged from 172 (*CsDof7*) to 495 aa (*CsDof17*), with an average size of 320 aa. The theoretical isoelectric point (pI) varied from 4.66 (*CsDof9*) to 9.52 (*CsDof6*) and the molecular weight (kDa) of the protein ranged from 26.71 (*CsDof*10) to 54.82 kDa (*CsDof*17) (Table 1). The subcellular localization of the protein was predicted to be in the nucleus with the Plant-mPLoc tool and the presence of Nuclear Localization Signals (NLS) were detected in all 24 *CsDof* proteins (Table 1).

The Dof domain, according to the National Center for Biotechnology Information conserved domain database (NCBI, CDD; https://www.ncbi.nlm.nih.gov/cdd), consists of approximately 52 residues and is located in the N-terminal region. In this study, we found a very similar amino acid distribution in the homeodomain, with 29 out of 52 amino acid residues being conserved in all 24 *CsDof* proteins (Figure 1).

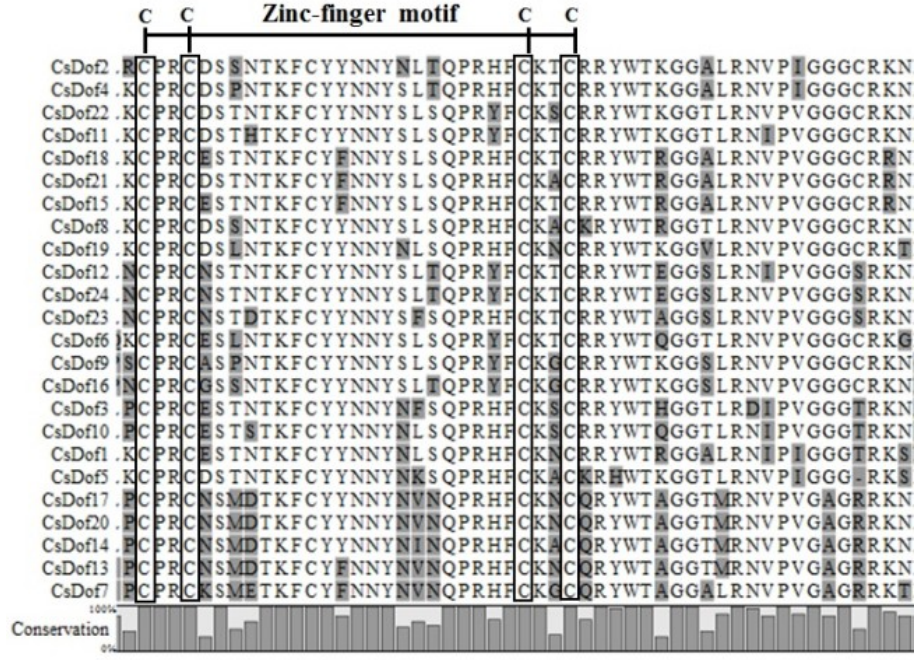

Figure 1. Multiple alignment of Dof domain of all *CsDof* proteins. Domains were aligned using CLC Main Workbench 8.0 program. Degree of sequence conservation is indicated below the alignment. The cysteine residues (C) putatively responsible for the zinc-finger structure are indicated above the sequence.

Chromosomal localization, structure and duplication of *CsDof* genes

The *CsDof* genes are located in almost all chromosomes, except for chromosomes 4 and 9 (Figure 2).

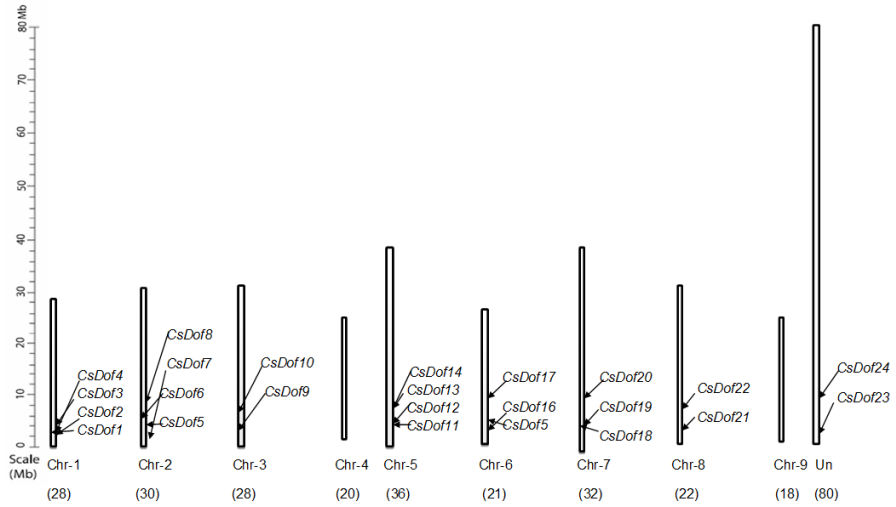

Figure 2. Chromosome localization of the 24 *CsDof* genes on 9 chromosomes of *Citrus sinensis*. The scale is in megabases (Mb). *CsDof*23 and *CsDof*24 lie within the unassembled scaffold.

All the genes were positioned towards the chromosome ends, with chromosomes 1, 2 and 5 containing the maximum occurrence of *CsDofs* (4 genes in each). Analysis of the exon-intron structure showed great variation among the *CsDof* genes (Figure 3). No introns were found in 37% of the *CsDofs* genes (*CsDofs 1, 2, 3, 5, 6, 7, 10, 16* and *19*) (Figure 3 and Table 1).

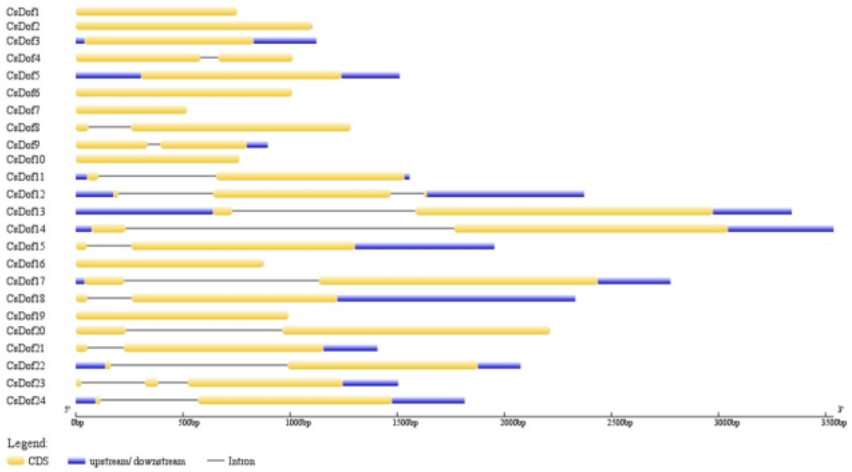

Figure 3. Gene structures of the 24 *CsDof* genes. Exons are represented by dark grey rectangles, whereas introns are shown as black lines. The scale is in pairs of bases (bp).

Thirteen genes contain a single intron (*CsDofs 4, 8, 9, 11, 13, 14, 15, 17, 18, 20, 21, 22* and *24*), mostly in the 5' region of the gene, while two have two introns (*CsDofs 18* and *23*). *CsDof12* presents one or two introns due to the occurrence of alternative splicing (Table 1).

Using only *CsDof* amino acid sequence data, trees were constructed using the Neighbor-Joining (NJ) method. Only the tree obtained with the NJ was used in this study (Figure 4), as the two algorithms resulted in almost identical topologies. From the analysis of the tree topology, we observed pairs of *CsDof*s with close homology (Figure 4), which may indicate the presence of paralogs.

However, by inspecting the pairs of *CsDof* genes located adjacently in the *C. sinensis* genome (Figure 2), we observed that only two neighboring homologous genes were separated by less than 200 kb (*CsDof3* and *CsDof4* genes, mapped 136 kb apart from each other on chromosome 1, and *CsDof18* and *CsDof19* located 158 kb apart on chromosome 7). Grouping of genes usually represent paralogues, but this is not always the case. Therefore, for the identification of possible paralogs, we adopted the following criteria: (a) the aligned nucleotide sequence covered >70% of the longer aligned gene; (b) the amino acid identity between the sequences was >70% (Yang et al., 2008). Based on that, only *CsDof14* and *CsDof17* (mapped on chromosomes 5 and 6, respectively) met these two criteria.

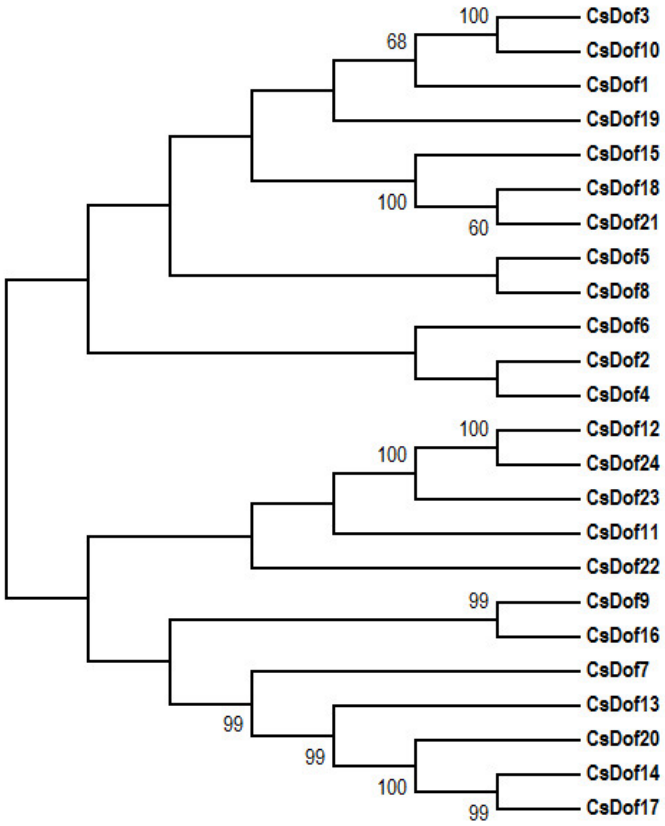

Figure 4. Phylogenetic tree constructed from the complete alignment of all amino acid residues of the predicted 24 *CsDof* proteins in C. sinensis. Bootstrap values above 50% are shown next to the relevant nodes.

Phylogenetic analysis of *CsDof*s

To assess phylogenetic relationships among the *C. sinensis* Dofs and to group them into the well-established groups and subfamilies (Lijavetzky et al., 2003; Yanagisawa, 2004), we constructed a phylogenetic tree based on multiple sequence alignments of *C. sinensis* and *A. thaliana* Dof protein sequences (Figure 5). The *CsDof* family were divided into four main groups (A, B, C and D) and 9 subgroups (A, B1, B2, C1, C2.1, C2.2, C3, D1 and D2) of proteins (Figure 5). These different groups and subfamilies were classified based on the phylogenetic tree using the *A. thaliana* genes as references according to Lijavetzky et al. (2003).

Groups C and D presented the largest number of genes (nine and seven, respectively), similarly to what was found in *Arabidopsis* (Figure 5 and Supplementary 2). The subfamily D1 is among those with the largest number of genes in both species used in the construction of the phylogenetic tree. On the other hand, only one of the 24 members of the *CsDof* family could be classified as C3. Groups A and B presented two and six genes, respectively, and group B could be divided into two subgroups, B1 and B2, with 3 members each (Figure 5 and Supplementary 2).

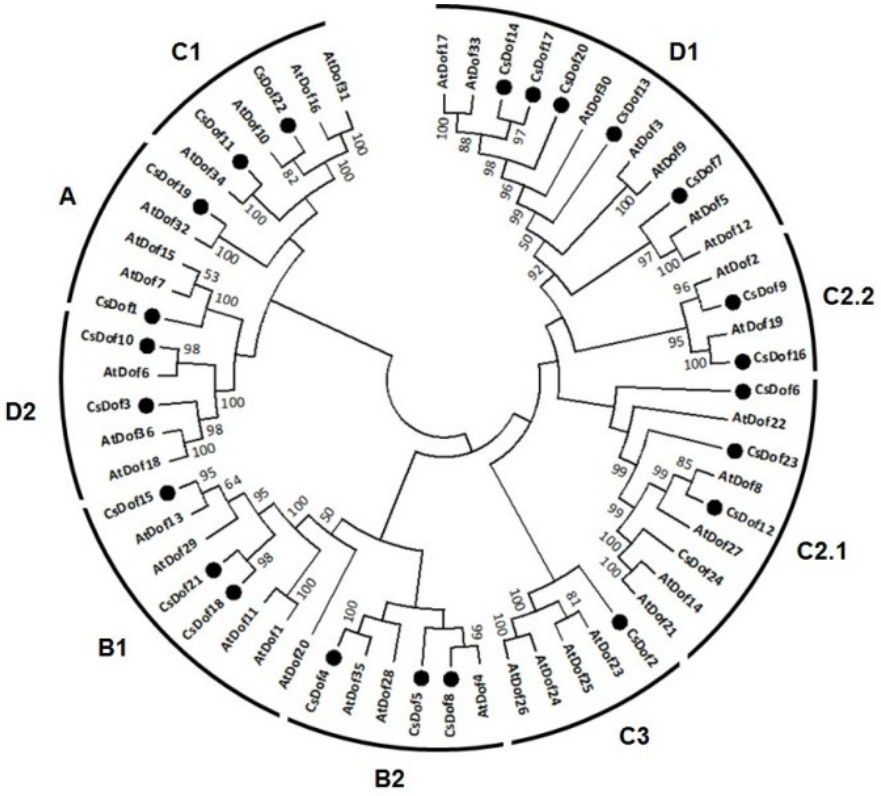

Figure 5. Joined unrooted phylogenetic tree of the *Citrus sinensis* and *Arabidopsis thaliana* Dof transcription factors. Bootstrapping values are indicated as percentages (when >50%) along the branches. The resulting gene groups and subgroups according to the report by Lijavetzky et al. (2003) are shown.

Expression analysis of *CsDof* genes in fruits

According to the RNA-seq expression data (Jiao et al., 2013; CITRUS SINENSIS, 2016), the *CsDof* genes are expressed in different parts of the plant, such as fruits, leaves, flowers and calluses (Supplementary 3). The highest number of *CsDof* genes was expressed in the fruits, with a total of 12 genes being up-regulated in this plant structure (*CsDofs 1, 5, 6, 9, 13, 14, 15, 17, 20, 22, 23* and *24*) (Supplementary 3).

To better understand the temporal expression patterns of the above selected *CsDofs* genes at early developmental stages of the sweet orange fruits, RT-PCR analyses were performed in fruits harvested every 15 days until reaching an average diameter of 4.3 cm and 36.5 g of fresh weight at day 90 after anthesis (Table 2). Considering the relative expression profiles in all stages, the transcript abundance levels the *CsDof* genes in fruits were divided into three groups. The first group is composed of *CsDofs* with high relative expression (*CsDof1, CsDof15, CsDof17, CsDof20, CsDof22*; > 8 times than in the leaves). The second group was made up of those genes with intermediate expression in at least in one stage (*CsDof13, CsDof14* and *CsDof24*; 3 to 8 times the mRNA level detected in the leaves). The remaining *CsDofs* showed low levels of activity in the fruits (*CsDof5, CsDof6, CsDof9* and *CsDof23*; below 3.5 times) (Figure 6).

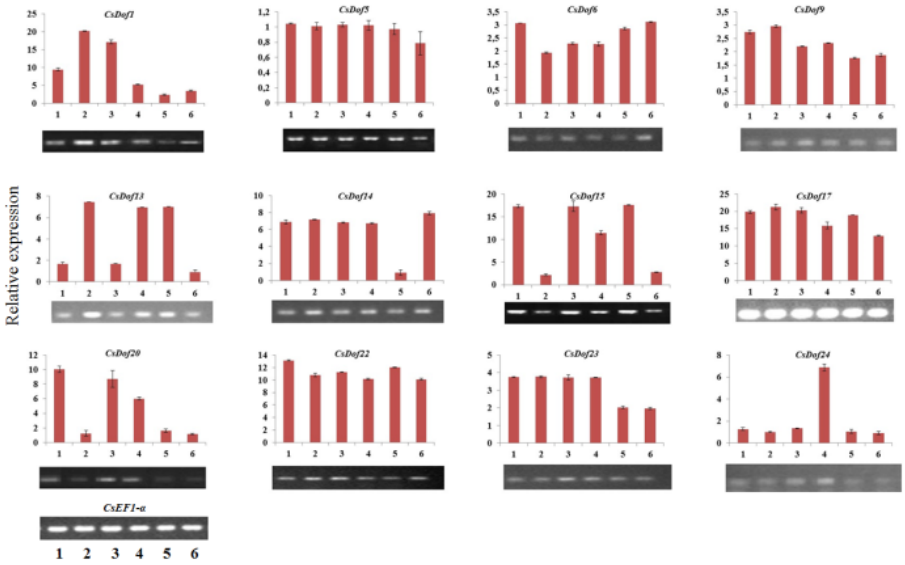

Figure 6. Expression profiles of 12 *CsDof*s genes at 15 (1), 30 (2), 45 (3), 60 (4), 75 (5) and 90 (6) days after flowering (DAF) in *Citrus sinensis* cv 'Pêra' fruits determined by semi-quantitative RT–PCR and normalized to that of CsEF1- α. Photographic images of gel electrophoresis are shown at the bottom of each graph. The data represent the means ± standard error (SE) for three biological replicates.

The *CsDof17* gene had a high transcriptional activity in all fruit development stages analyzed in this work. Although showing lower mRNA abundance than *CsDof17*, the *CsDof22* gene also had a comparative uniform relative expression in the fruits from the beginning to the end of the sampling period (Figure 6). Diversely, *CsDof1* showed an increasing level of expression in the early stages of development followed by a large

decrease as the fruit development progressed. Despite their high expression, *CsDof15* and *CsDof20* showed marked fluctuations in the number of transcripts, depending on the period of data collection (Figure 6).

Within the second group of genes, *CsDof24* presented low mRNA levels in all stages, except at 60 days after anthesis (Figure 6). On the other hand, an intermediate and stable expression level of the *CsDof14* gene was observed throughout the sampling period, with only a marked decline (circa 7 X) in stage 5 (75 DAF). The abundance of *CsDof13* gene transcripts oscillated significantly during the different stages of fruit development, with the time of peak expression in stages 2, 4 and 5 (Figure 6).

The semi-quantitative RT-PCR results showed that the lowest transcriptional level was for the *CsDof5* at all stages of development of the fruits. *CsDof23,* also included in the group with low expression levels, was relatively more expressed at the beginning of the development of the fruit (up to 60 days after anthesis), with a decline of about 50% in mRNA levels in the last two sampling periods. *CsDof6* and *CsDof9* were found to express comparatively similar low mRNA levels during the fruit development, showing only slight fluctuations (to a maximum of 2-fold higher) relative to their expression in leaf tissues (control).

DISCUSSION

The functions of Dof transcription factors in a vast range of developmental processes in plants make them targets for research in developmental biology and in plant biotechnology for crop improvement (Yanagisawa, 2004). Various studies have shown their great diversity both in terms of number and functions (Lijavetzky et al., 2003; Kushwaha et al., 2011; Cai et al., 2013; Ma et al., 2015, Zhang et al., 2017; Cheng et al., 2018; Garcia et al., 2018). However, there are few reports on these transcription factors in fruit trees of economic interest (Wu et al., 2016b). Thus, in this work, we performed a comprehensive study of the Dof gene family in *C. sinensis*, including phylogenetic, chromosomal location, gene structure and expression profiling analyses at the early stages of fruit development.

It was possible to identify 24 *CsDof* genes with a highly conserved Dof domain in all freely available citrus genome databases. This small number of genes is in sharp contrast with those reported in other species, where the number of Dof members ranged from 30 to 76 genes (Ma et al., 2015). Although with genomes largely differing in size [*C. sinensis* genome size = 367 Mbp (Arumuganathan and Earle, 1991); *Sorghum bicolor* genome size = 750 Mbp (Paterson et al., 2009) and *Brachypodium. distachyon* genome size = 272 Mb (Vogel et al., 2010)], the number of Dof genes found in *C. sinensis* is similar to these two latter-mentioned monocot species that carry 28 and 27 Dof genes, respectively (Kushwaha et al., 2011; Hernando-Amado et al., 2012).

It has been reported that Dof family genes in rice, tomato and *Arabidopsis* usually contain few (usually one or two) or no introns (Lijavetzky et al., 2003; Kushwaha et al., 2011; Cai et al., 2013). We found that a similar condition exists in *C. sinensis*: among the 24 *CsDof* members, 58% (14) contained at least one intron in their open reading frame (ORF) (Figure 3 and Table 1). Further analysis showed that one intron positions in *CsDof8, 15*, *18* and *21* were highly conserved (Figure 3), which suggests that splicing was conserved during the evolution of these four genes. Aside these intron positions, it was observed that non-conserved intron positions are more frequent in the *CsDof* members.

The 24 *CsDof* TFs are found to be coded by unique genes located throughout the chromosomes, except for chromosomes 4 and 9, with a maximum of four genes in chromosomes 1, 2 and 5 (Figure 2). Comparatively, in tomato the 34 Dof family genes are organized on 11 chromosomes out of the 12 chromosomes (Cai et al., 2013), whereas the 36 Dof genes of *Arabidopsis* are distributed among all five chromosomes (Lijavetzky et al., 2003). The location of the *CsDof* genes at the ends of the chromosome arms offers increased probability of inter-chromosomal exchanges during genome duplication events.

Paralogs are genes originated by a duplication event within the genome of the same species, usually evolving to distinct functions. Two or more genes located on the same chromosome may have originated from a tandem duplication event, whereas a segmental duplication event involves the transfer of sequences to one or more locations between different chromosomes. In tomato, 11 pairs of paralogous genes were detected, 8 from segmental events randomly scattered throughout the genome and 3 located on the same chromosome (Cai et al., 2013), indicating that both segmental and tandem duplications have contributed to the expansion of Dof gene family in this species. Here, only a pair of putative paralogs were identified (*CsDof14* and *CsDof17*, mapped on chromosomes 5 and 6), showing that *C. sinensis* Dof genes have undergone lower rate of duplication. The observed clustering of the *CsDofs* in the chromosomes seems to be the consequence of fusions and chromosomal translocations derived from the hybrid origin of *C. sinensis,* a cross between *C. grandis* and *C. reticulata* (Xu et al., 2013). Liu et al. (2014) in an analysis of the family of CCCH-zinc transcription factors in Clementine mandarin (*C. reticulata* × sweet orange), also observed the presence of only few paralogs, which suggests that the occurrence of duplications in citrus species is much less prevalent when compared to *Arabidopsis* and rice (Lijavetzky et al., 2003).

Phylogenetic analysis based on the grouping used for *Arabidopsis* (Lijavetzky et al., 2003; Kushwaha et al., 2011), showed that the *CsDof*s were clustered into 4 groups (A, B, C and D) and 9 subgroups (A, B1, B2, C1, C2.1, C2.2, C3, D1, D2) (Figure 5 and Supplementary 2). Among these, subgroups C2.1 and D1 constituted the largest clades (with 4 and 5 members, respectively), while the smallest clade was the C3 (1 member). The predominance of genes clustered in group D1 was similar to the observed in tomato (Cai et al., 2013), chinese cabbage (Ma et al., 2015) and also banana (Feng et al., 2016), a species where the D1 group showed 10 members.

Kushwaha et al. (2011) found great similarity between *AtDof17, AtDof30* and *AtDof33* genes in *Arabidopsis*, whose proteins, also denominated CDF (Cycling Dof Factor), are associated with the regulation of the photoperiod and control of flowering (Fornara et al., 2009). In this study, we found that these three above-mentioned *Arabidopsis* Dofs were clustered in the D1 subgroup together with *CsDof*s7, 13, 14, 17 and 20, which are more expressed in the sweet orange fruits according to the RNAseq data (Jiao et al., 2013). The *CsDof*3, classified in group D2, was clustered with *AtDof18* (OBP1) which, according to Skirycz et al. (2008), is involved in the regulation of the cell cycle. Interestingly, the C3 subgroup, which was originally identified in *Arabidopsis* (*AtDofs* 23, 24, 25 and 26), only contains one *C. sinensis* Dof gene (*CsDof*2). This subgroup is found only in a few species, especially in cruciferous plants, and no apparent homologues have been observed in rice, tomato, chinese cabbage and banana (Lijavetzky et al., 2003; Cai et al., 2013; Ma et al., 2015; Feng et al., 2016).

According to Araújo et al. (2005), depending on the genotype and growth conditions, the development of sweet orange fruits can be divided into 3 phases: an initial slow development, circa 20 days after fruit setting; a second phase of accelerated growth (approximately 120 days); followed by a slowdown in growth (until maturation), with no appreciable change in diameter. This highly complex process involves important changes in the color, flavor, aroma, texture and nutritional content of the pulp, being controlled by transcriptional regulation networks that involve several transcription factors (Gapper et al., 2013; Seymour et al., 2013; Cherian et al., 2014). Therefore, over the last few years, more and more studies have been reported on the function of transcription factors families in sweet orange and other species during fruit ripening and development (Karlova et al., 2014; Wu et al., 2016b). Since gene expression patterns can provide valuable information on their function, we analyzed the transcriptional profile of 12 *CsDof* genes, reported as having higher normalized RPKM values in fruits relative to other tissues in a publicly available developmental transcriptome RNA-seq normalized RPKM dataset (Jiao et al., 2013), at the early stages of fruit development.

In our case, the sigmoidal growth of the fruits of *C. sinensis* cv. Pêra during the sampling period (maximum size and weight of 43 mm diameter and 36.5 g, respectively) was similar to the data obtained by Araujo (2010) for fruits of the cultivar Valencia. As expected, based on the expression of Dof transcripts in sweet orange and other fruit species (Cai et al., 2013; Feng et al., 2016, Wu et al., 2016b), the *CsDof* genes analyzed showed different mRNA levels during the early stages of sweet orange fruit development analyzed here (Figure 6). Expression of *CsDof17* transcripts was the highest at all stages analyzed, indicating that this gene probably plays an important role in the development of sweet orange fruits. Specifically, at the beginning of fruit development (stage 1; up to 15 days after anthesis), the most expressed genes were *CsDof1*, *15, 17, 20* and *22*. In contrast, *CsDof5* remained mostly undetectable across the six sampled stages. *CsDof6* and *CsDof9* were also found to express low mRNA levels during all fruit development stages. In the case of *CsDof9*, we detected a further decrease in its mRNA expression as fruit ripening progressed. It is interesting to mention that *CsDof9* (orange1.1g025998m) was found to be significantly down-regulated in a bud late-ripening mutant in comparison with its wild-type counterpart (Fengjie 72–1') during a later sampling period (150 - 240 DAF) than the one used in this work (Wu et al., 2016b).

The phylogenetic analysis showed that genes highly expressed in the sweet orange fruits (*CsDof17* and *20*) were clustered in the subgroup D1, which contains the genes *SlDof3* and *SlDof22* and the genes *MaDof2*, *3, 10* and *20* that were reported to be largely transcribed in tomato and banana fruits, respectively (Cai et al., 2013; Feng et al., 2016). The expression data analysis of the identified paralogous pair revealed a high level of expression divergence in the fruit tissues. While the *CsDof17* gene showed high expression level in all the analyzed sampling dates, its putative paralogous *CsDof14* showed only intermediate expression at five of the fruit stages.

The comparison of the RNA-seq data (Supplementary 1) to those obtained here by semi-quantitative RT-PCR (Figure 6) revealed that the genes *CsDof5, 9, 13, 17, 20, 22* and *23* exhibit similar relative expression patterns between the two methods for transcript abundance estimation, with more transcripts detected in the fruits than in the leaves. On the other hand, the *CsDof1* and *CsDof14* genes exhibited similar transcriptional profiles in fruits and leaves by the RNA-seq data, while the results obtained here with RT-PCR

showed that this correspondence only occurred after stage 4 (60 to 90 DAF) and at stage 5 (75 DAF) for *CsDof1* and *CsDof14*, respectively. Unlike RNA-seq data obtained by Jiao et al. (2013), in this work the samplings were performed at different stages of the early development of the fruits, so we were able to particularize Dof expression levels, whereas RNA-seq data was generated from fruits sampled at a single stage or from a pool of developmental stages (Xu et al., 2013). In addition, it should be noted that these few divergences could be due to the different capacities for detecting transcripts between the two methods of analysis, the use of different genotypes and / or the biological complexity resulting from the environmental effects at the plant culture sites used for the collection of the fruits.

CONCLUSIONS

In this study, the analysis of sequence data existing in publicly available databases was performed to identify 24 genes encoding Dof transcription factors from *C. sinensis* genome. The phylogenetic relationship of these proteins among themselves and with their orthologs from Arabidopsis allow us to group *CsDofs* into four groups and 9 subgroups. To understand the contribution of *CsDofs* in fruit development, mRNA levels of 12 selected Dof genes were monitored at initial stages of fruit growth. Differential expression of these transcription factors evidences that they should play an important role in the regulation of sweet orange fruit development. In particular, the expression profile changes detected in the *CsDof* genes examined here may serve as biomarkers for evaluating key developmental transitions of sweet orange fruits and how they are affected by growth conditions under controlled environments or in the field.

ACKNOWLEDGMENTS

LGEV was supported by scholarship from National Counsel of Technological and Scientific Development (CNPq).

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Araújo EF and Roque N (2005). Taxonomia dos citros. In: Mattos Júnior D, de Negri JD, Pio RM. Pompeu Júnior J. (eds) Citros. Campinas: Instituto Agronômico e Fundag. 1: 125–145.

Bain JM (1958). Morphological, anatomical, and physiological changes in the developing fruit of the Valencia orange. (*Citrus sinensis* (L) Osbeck). *Aus. J. Bot*. 6: 1–23.

Cai X, Zhang Y, Zhang C, Zhang T, et al. (2013). Genome-wide analysis of plant-specific Dof transcription factor family in tomato. *J. Int. Plant Biol*. 55: 552–566.

Cheng Z, Hou D, Liu J, Li X, et al. (2018). Characterization of moso bamboo (*Phyllostachys edulis*) Dof transcription factors in floral development and abiotic stress responses. *Genome*. 61: 151–156.

Cherian S, Figueroa CR and Nair H (2014). 'Movers and shakers' in the regulation of fruit ripening: a cross-dissection of climacteric versus nonclimacteric fruit. *J. Exp. Bot*. 65: 4705–4722.

Chou KC and Shen HB (2010). Plant-mPLoc: A Top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One*. 5(6): e11335.

CITRUS SINENSIS (2016). Disponível em: <http://citrus.hzau.edu.cn/orange/>. Acesso em: 01 nov 2017.

Corrales AR, Carrillo L, Lasierra P, Nebauer SG, et al. (2017). Multifaceted role of cycling DOF factor 3 (CDF3) in the regulation of flowering time and abiotic stress responses in Arabidopsis. *Plant Cell. Environ*. 40: 748–764.

Endo T, Shimada T, Fujii H and Omura M (2006). Cloning and characterization of 5 MADS-box cDNAs isolated from citrus fruit tissue. *Sci. Hortic*. 109: 315–321.

Feng BH, Han YC, Xiao YY, Kuang JF, et al. (2016). The banana fruit Dof transcription factor MaDof23 acts as a repressor and interacts with MaERF9 in regulating ripening-related genes. *J. Exp. Bot*. 67: 2263–2275.

Fornara F, Panigrahi KC, Gissot L, Sauerbrunn N, et al. (2009). Arabidopsis DOF transcription factors act redundantly to reduce constans expression and are essential for a photoperiodic flowering response. *Develop. Cell*. 17: 75–86.

Freschi L, Nievola CC, Rodrigues MA, Domingues DS, et al. (2009). Thermoperiod affects the diurnal cycle of nitrate reductase expression and activity in pineapple plants by modulating the endogenous levels of cytokinins. *Physiol. Plant.* 137: 201–212.

Gapper NE, Mcquinn RP and Giovannoni JJ (2013). Molecular and genetic regulation of fruit ripening. *Plant Mol. Biol*. 82: 575–591.

Garcia V, Ribas AF, Vieira LGE and dos Santos TB (2018). *In silico* analysis of the Dof transcription factor family in *Coffea canephora*. *Coll. Agrariae*. 14: 99–111.

Hernando-Amado, S. González-Calle V, Carbonero P and Barrero-Sicilia C (2012). The family of DOF transcription factors in *Brachypodium distachyon*: phylogenetic comparison with rice and barley DOFs and expression profiling. *BMC Plant Biol*. 12: 1–12.

Jiao WB, Huang D, Xing F, Hu Y, et al. (2013). Genome-wide characterization and expression analysis of genetic variants in sweet orange. *Plant J*. 75: 954–964.

Karlova R, Chapman N, David K, Angenent GC, et al. (2014). Transcriptional control of fleshy fruit development and ripening. *J. Exp. Bot*. 65: 4527–4541.

Kurai T, Wakayama M, Abiko T, Yanagisawa S, et al. (2011). Introduction of the ZmDof1 gene into rice enhances carbon and nitrogen assimilation under low-nitrogen conditions. *Plant Biotechnol. J*. 9: 826–837.

Kushwaha H, Gupta S, Singh VK, Rastogi S, et al. (2011). Genome wide identification of Dof transcription factor gene family in sorghum and its comparative phylogenetic analysis with rice and Arabidopsis. *Mol. Biol. Rep*. 38: 5037–5053.

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

Liu S, Khan MR, Li Y, Zhang J, et al. (2014). Comprehensive analysis of CCCH‐type zinc finger gene family in citrus (Clementine mandarin) by genome‐wide characterization. *Mol. Gen. Genomics*. 289: 855–872.

Ma J, Li MY, Wang F, Tang J, et al. (2015). Genome-wide analysis of Dof family transcription factors and their responses to abiotic stresses in Chinese cabbage. *BMC Genomics*. 31: 16–33.

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

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, et al. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature*. 457: 551–556.

Santopolo S, Boccaccini A, Lorrai R, Ruta V, et al. (2015). DOF AFFECTING GERMINATION 2 is a positive regulator of light-mediated seed germination and is repressed by DOF AFFECTING GERMINATION 1. *BMC Plant Biol.* 15: 72 10.1186/s12870-015-0453-1.

Seymour GB, Ostergaard L, Chapman NH, Knapp S, et al. (2013). Fruit development and ripening. *Annu. Rev. Plant Biol*. 64: 219–241.

Skirycz A, Radziejwoski A, Busch W, Hannah MA, et al. (2008). The DOF transcription factor OBP1 is involved in cell cycle regulation in *Arabidopsis thaliana*. *Plant J*. 56: 779–792.

Tamura K, Stecher G, Peterson D, Filipski A, et al. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol*. 30: 2725–2729.

Vogel JP, Garvin DF, Mockler TC, Schmutz J, et al. (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*. 463: 763–768.

Wang JH, Liu JJ, Chen KL, Li HW, et al. (2017). Comparative transcriptome and proteome profiling of two *Citrus sinensis* cultivars during fruit development and ripening. *BMC Genomics*. 18: 984.

Wu J, Xu Z, Zhang Y, Chai L, et al. (2014). An integrative analysis of the transcriptome and proteome of the pulp of a spontaneous late-ripening sweet orange mutant and its wild type improves our understanding of fruit ripening in citrus. *J. Exp. Bot*. 65: 1651–1671.

Wu Z, Chen J, Cui J, Xu X, et al. (2016a). Genome-wide identification and expression profile of dof transcription factor gene family in pepper (*Capsicum annuum* L.). *Front. Plant Sci*.7: 574.

Wu J, Fu L and Yi H (2016b). Genome-wide identification of the transcription factors involved in citrus fruit ripening from the transcriptomes of a late ripening sweet orange mutant and its wild type. *Plos One*. 11: 4,e0154330.

Xu Q, Chen LL, Ruan X, Chen D, et al. (2013). The draft genome of sweet orange (*Citrus sinensis). Nature Genet*. 45: 59–66.

Yanagisawa S (2004). Dof domain proteins: Plant-specific transcription factors associated with diverse phenomena unique to plants. *Plant Cell. Physiol.* 45: 386-391.

Yang S, Zhang X, Yue JX, Tian D, et al. (2008). Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genomics.* 280: 187–198.

Ye JL, Zhu AD, Tao NG, Xu Q, et al. (2010). Comprehensive analysis of expressed sequence tags from the pulp of the red mutant 'Cara Cara' navel orange (*Citrus sinensis* Osbeck). *J. Int. Plant Biol*. 52: 856–867.

Zhang Z, Yuan L, Liu X and Wang XY (2017). Evolution analysis of Dof transcription factor family and their expression in response to multiple abiotic stresses in *Malus domestica*. *Gene*. 639: 137–148.