Pak. J. Bot., 54(4): 1375-1383, 2022. DOI: http://dx.doi.org/10.30848/PJB2022-4(33)

GENOME-WIDE IDENTIFICATION AND CHARACTERIZATION OF THE *DOF* GENE FAMILY IN *PRUNUS SIBIRICA*

HONGGUO LI[2], NINGNING GOU[1,3], CHEN CHEN[1,3], KAIJUN LIU[4] , LIN WANG[1,3*] AND TANA WUYUN[1,3*]

[1]*State Key Laboratory of Tree Genetics and Breeding, Non-timber Forest Research and Development Center, Chinese Academy of Forestry, Zhengzhou, 450003, China*
[2]*Experimental Center of Tropical Forestry, Chinese Academy of Forestry, Pingxiang 532600, China*
[3]*Kernel-Apricot Engineering and Technology Research Center of State Forestry and Grassland Administraion, Chinese Academy of Forestry, Zhengzhou, 450003, China*
[4]*Wulatehou City Nursery of National forestry and Grassland Administration, Chinese Academy of Forestry, Wulatehou 015543, China*
[*]*Corresponding author's email: tanatanan@163.com, wanglin1815@163.com*

Abstract

Prunus sibirica (Siberian apricot) is a species of the family Rosaceae, section Armeniaca (Lam.) Koc, which has enormous potential for oil and food raw materials. DNA binding with one finger (*Dof*) transcription factors play major roles in plant biological processes. However, the evolutionary and functional information of the *Dofs* in *P. sibirica* remain unclear. Here, we conducted a genome-wide screening and characterization of the *Dofs* in *P. sibirica*, and 24 putative *PsDofs* were identified, which were distributed across all eight chromosomes. Phylogenetic analysis showed that *PsDofs* were divided into four major groups (A, B, C, and D). The gene structure and conserved motifs of *PsDofs* were also predicted. The expression profiles of *PsDofs* exhibited different expression patterns in flower buds, flowers, leaves, fruits, and kernels. This study provides an important foundation for better understand the evolution and function of *PsDofs*.

Key words: *Dof* gene family, Phylogenetic analysis, Gene expression, *Prunus sibirica*.

Introduction

Prunus sibirica is a member of the family Rosaceae. It is broadly distributed in mountainous areas of northern and northeastern China, Eastern and southeastern Mongolia, eastern Siberia and coastal areas of Russia (Wang, 2011; Wang, 2012). *P. sibirica* can grow in poor soil and can tolerate drought and salinization. It is regarded as a crucial wood oil and also food species. The kernel of *P. sibirica* contained 45.65-51.47% crude fat, 20.93-30.55% crude protein, and 4.75-5.96% laetrile (Yin *et al*., 2017).

The *Dof* is a typical plant transcription factor, which contains 200-400 aa. It has a highly conserved *Dof* domain at the N-terminal, which is a zinc-finger structure formed by a $CX_2CX_{21}CX_2C$ motif. (Yanagisawa, 2002; Chen *et al*., 2012). According to phylogenetic analysis, *Dof* were divided into four groups which composed several subgroups (Lijavetzky *et al*., 2003). *Dofs* play diverse role in various biological processes of the plant growth and development, such as, *PBF* regulates the expression of endosperm specific storage protein genes in cereal (Vicente *et al*., 1997). The tobacco *NT-BBF1* regulated the expression of the proto-oncogene *rolB* in microtubules and apical meristers in response to auxin signals (Baumann, 1999). The expression *GmDof4* and *GmDof11* was related to the lipid content of seeds in *Soybean* (Wang *et al*., 2007). The *AtOBP1* was related to the plant defense mechanisms in *A.thaliana* (Chen *et al*., 2010). *OsDof12* could regulate the expression of *Hd3a* and *OsMADS1*4 and promote flowering in *Oryza sativa* (Li *et al*., 2009). A total of five candidate *PpeDofs* were highly expressed at the dormancy release stage, which may be involved in dormancy in peach (Chen *et al*., 2017). The expression levels of most *Dofs* in vegetative organs were higher than those in reproductive organs in apple (Wang *et al*., 2021). By contrast, little information about *Dofs* in *P. sibirica* is available.

In this study, a genome-wide analysis of *Dofs* in *P. sibirica* genome was performed. The gene structures, sequence characteristics, evolutionary relationships and gene expression profiles were comprehensively investigated. This research would facilitate further studies of *Dofs* in *P. sibirica* and other plants in Rosaceae.

Material and Methods

Identification of *Dofs* in the *P. sibirica*: The *Dof* sequences of *A. thaliana* were downloaded from PlantTFDB (http://planttfdb.cbi.pku.edu.cn/) and were used as queries in a BLAST (e-value =$1e^{-10}$ and a minimum amino acid identity of 50%) search against *P. sibirica* genome. Then, the *Dof* domain (PF02701) were obtained from Pfam (http://pfam.xfam.org) (Finn *et al*., 2016). Hidden Markov Model (HMM) (Finn *et al*., 2011) was used to search against reference genome. Finally, *Dof* conservative domains of candidate sequences were examined using SMART (http://smart.embl.de) (Ivica *et al*., 2015) and NCBI CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The putative isoelectric points (PIs) and molecular weights (MWs) of the candidate *Dofs* were predicted using the ExPASy Proteomics Server (http://expasy.org//protparam/).

Phylogenetic tree analysis: The full-length protein sequences from *P. sibirica*, *O. sativa* (Lijavetzky *et al*., 2003) and *A. thaliana* (http://planttfdb.cbi.pku.edu.cn/) (Yanagisawa, 2002) were aligned using clustalx2.0 (Larkin *et al*., 2007). The phylogenetic tree was constructed in MEGAX (Sudhir *et al*., 2018) using Neighbor-joining method, bootstrap values were calculated for 1000 replicates.

Gene structure and conserved motifs in *PsDofs*: The conserved motif domain of *PsDofs* were performed by MEME (http://meme-suite.org/tools/meme), the parameters were set to 20 different motifs with a width of 6-200 aa, and were visualized using TBtools (0.668636) (Chen *et al*., 2018). Exon and intron components of *PsDofs* were analyzed using GSDS2.0 (http://gsds.cbi.pku.edu.cn/index.php).

Gene duplication and evolutionary analysis of *PsDofs*: The chromosomal location of *PsDofs* were from genome annotation files. BLASTP was used to search collinearity for each of *PsDofs* (e-value < $1e^{-5}$, the top 5 matches). Then, the replication events were examined by MCScanX (Wang *et al*., 2012). Ks (synonymous substitution rate) and Ka (nonsynonymous substitution rate) were calculated and the results were visualized by TBtools (0.668636) (Chen *et al*., 2018).

Expression analysis using RNA-seq data: RNA-seq data, including leaf, flower bud, flower, and the fruit (F1-F6) and kernel (K1-K6) at six development stages. Two biological replicates for each sample were selected. FPKM (fragments per kilobase of exon per million fragments mapped) of each gene was calculated to present the expression level of *PsDofs*. The expression patterns were based on the transformed data of log2 (FPKM+1) values and min-max normalization by Heat map in TBtools (0.668636) (Chen *et al*., 2018).

Results

Identification and characterization of *PsDofs*: A BLASTP search was performed using the known *Dof* protein sequences of *A. thaliana* as queries to determine the *Dofs* of *P. sibirica* genome. Pfam, NCBI, CDD, and SMART searches were further used to ensure that the predicted sequences contained conserved domains. A total of 24 *PsDofs*, which harbor a $CX_2CX_{21}CX_2C$ zinc finger pattern, were identified in *P. sibirica* genome (Fig. 1). The length, putative molecular weights, and theoretical PIs ranging from 223 to 515 aa were 24.0 to 55.3 kDa, and 4.7 to 9.4, respectively (Table 1).

Phylogenetic analysis and classification of *PsDofs*: In order to reveal the evolutionary relationship of *PsDofs*, we constructed the phylogenetic tree based on the aligned protein sequences of 24 *PsDofs*, 36 *AtDofs* and 30 *OsDofs*. The *Dofs* were classified into four major groups (group A, group B, group C, and group D) (Fig. 2), which were consistent with previously described (Explain: the results were from the previously described by Lijavetzky *et al*., 2003) (Lijavetzky *et al*., 2003). Two and six *PsDofs* were in group A and group B, respectively. The most *PsDofs* (9) from group C, which were further clustered into four subgroups (subgroup C1, subgroup C2.1, subgroup C2.2, and subgroup C3). Group D contained seven members, which were divided into three subgroups (subgroup D1, subgroup D2, and subgroup D3). All *PsDofs* belonged to other 10 groups/subgroups. Except subgroups C3 and D3, which were specific to *A. thaliana* and *O. sativa*.

Chromosomal location and duplication of *PsDofs*: All 24 *PsDofs* were unevenly distributed on 8 chromosomes of *P. sibirica* (Fig. 3A), chromosomes 4, 5, and 6 contained four *PsDofs*, three *PsDofs* on chromosomes 2 and 7, respectively, and two *PsDofs* on chromosome 1, 3, and 8, respectively. We further analyzed duplication events of the *Dofs* of *P. sibirica* and *Prunus persica*. Ten segmental duplication existed in *PsDofs* (Table S1). A total of 42 pairs of orthologous *Dofs* were found between *P. sibirica* and *P. persica* (Fig. 3B, Table S2), they were divided into 8 categories, including A (3pairs), B1 (8pairs), B2 (3pairs), C1 (3pairs), C2.1 (9pairs), C 2.2(3pairs), D1 (11pairs), D2 (2pairs). Among these segmental duplication events, only *PsDof24* and *PpeDof2* belonged the different groups in Phylogenetic analysis. The triplicated gene pairs, including *PsDof3*, *PsDof4*, and *PsDof5*, *PsDof12*, *PsDof13*, and *PsDof14*, and *PsDof19*, *PsDof20*, and *PsDof21*, were observed in *P. sibirica* (Fig. 3A), as previously shown in *P. persica* (Chen *et al*., 2017). We further calculated the synonymous/ nonsynonymous substitution (Ka/Ks) ratios of segmental duplicated gene pairs in *P. sibirica* and between *P. sibirica* and *P. persica* (Table S1, Table S2). The Ka/Ks < 1 in *PsDofs,* suggesting the *PsDofs* had undergone purification selection during evolution, Ka/Ks analysis of *P. sibirica* and *P. persica* ratios were also below consistent <1, suggesting an overall signature of purifying selection or constraint on *PsDofs* and *PpeDof*.

Gene structure and conserved motif analysis of *PsDofs*: The exon-intron structure for the *PsDofs* were aligned and compared. Most of the different groups/subgroups of *PsDofs* have different exon-intron structures. Totally, *PsDofs* in groups/subgroups B1, C2.1, D1, and D2 had two exons, expect *PsDof22* had three exons, while *PsDofs* in groups/subgroups A and C2.2 had only one exon (Fig. 4A). The conserved motifs were analyzed by the MEME. A total of 15 motifs were identified (Fig. 4B), Almost all the *PsDofs* in the same groups/subgroups had a similar motifs. All of the *PsDofs* contained motif 1, which represented the conserved *Dof* domain, Motif 2 was widely present in subgroups B1 and C2.2, while, motif 3, 4, and 6 were unique to the subgroups B1, C2.1, and C2.2, respectively, and motif 9, 10, 11,14, and 15 were widely present in subgroups D1 (Table S3).

Expression patterns of *PsDofs* in various tissues: The expression patterns of *PsDofs* in leaves, flower buds, flowers, kernels, and fruits were analyzed. The expression patterns of *PsDofs* showed significant differences in various tissues and at different stages of fruit and kernel development. In general, the most genes in same group showed similar expression patterns, suggesting that there may be similar functions or functional redundancy, such as, group A (*PsDof1* and *PsDof2*) was highly expressed during the fruit development stages, whereas subgroup C2.2 (*PsDof15*, *PsDof16*, and *PsDof17*) presented lower expression during the fruit and kernel development process. In addition, we found the *PsDof2*, *PsDof5*, and *PsDof9* were highly expressed in leaves, flower buds, and flowers, respectively (Fig. 5). The expressions of *PsDof8*, *PsDof9*, *PsDof12*, *PsDof14*, *PsDof20*, and *PsDof23* were high at the early development stage of fruit and kernel. *PsDof20* were also exhibited higher expressed in mature fruit and kernel. The expression of *PsDof22* was downregulated during the kernel development. *PsDofs* showed different expression patterns, suggesting they may play different roles in growth and development of *P. sibirica*.

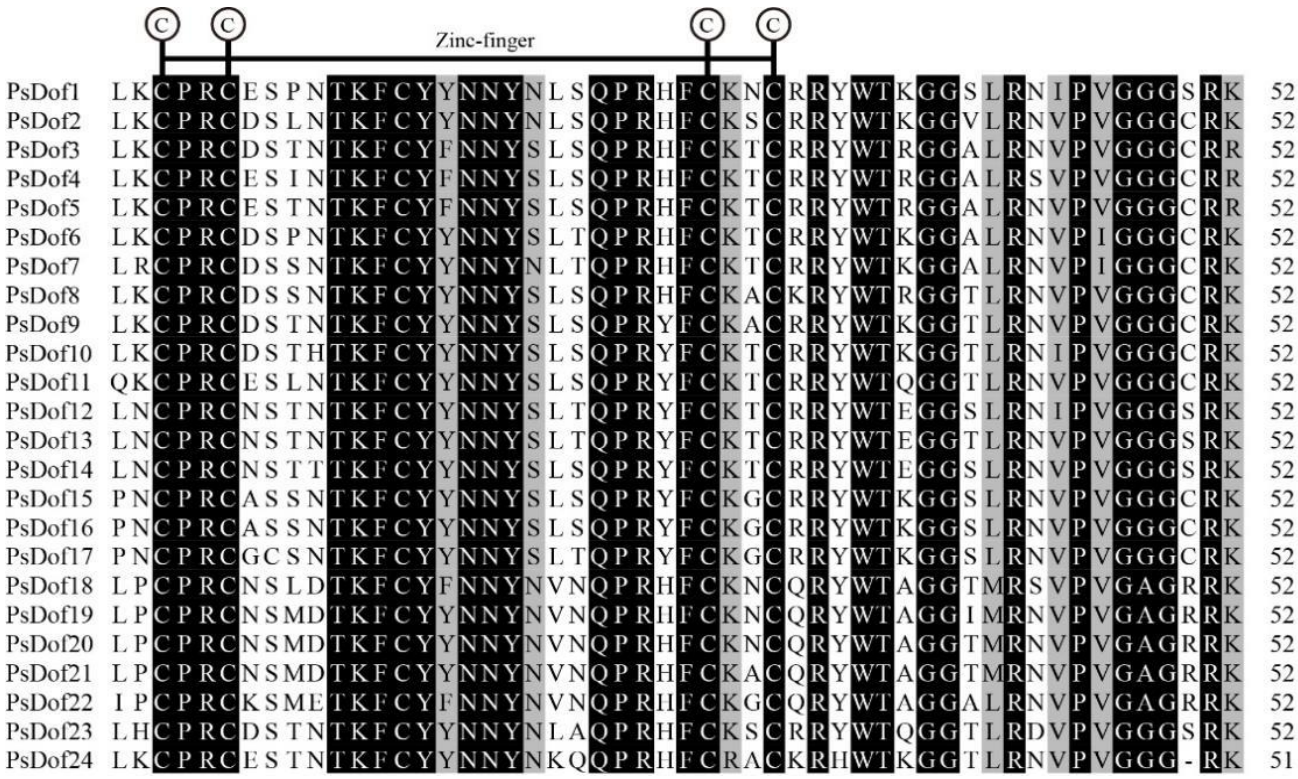

Fig. 1. Sequence alignment of *PsDofs*. The four cysteine residues putatively responsible of the zinc-finger structure were indicated. Identical amino acids were highlighted in black or gray, respectively.

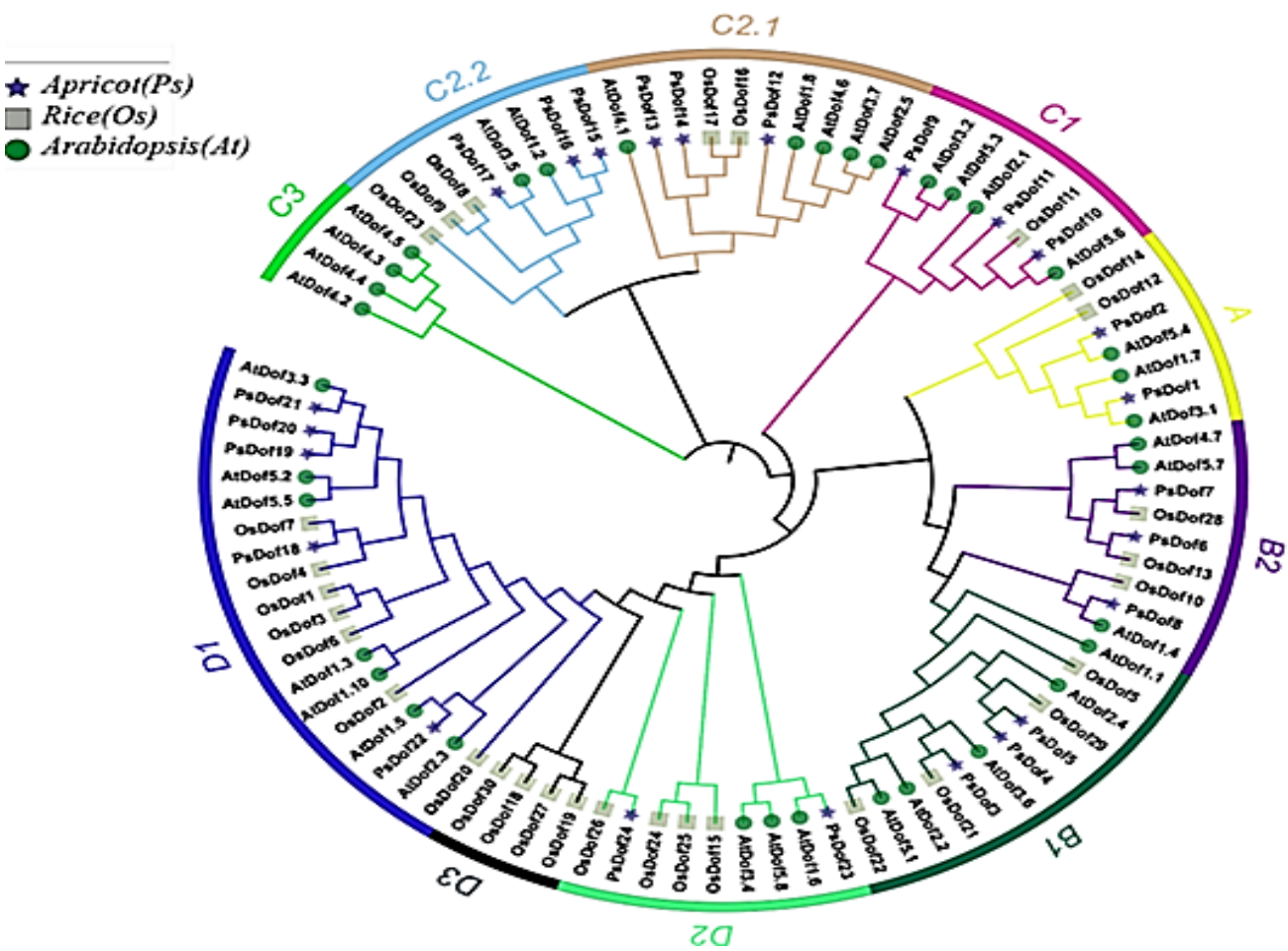

Fig. 2. The phylogenetic tree of *Dofs* in *P. sibirica.*

Table 1. Information of *PsDofs*.

Subfamily	Gene name	Gene ID	Length (aa)	Molecular weight (kDa)	PIs
A	*PsDof1*	PaF106G0302690000.01.T02	223	24.0	8.74
A	*PsDof2*	PaF106G0506763600.01.T01	329	34.9	7.72
B1	*PsDof3*	PaF106G0604694300.01.T01	378	39.8	9.36
B1	*PsDof4*	PaF106G0201899100.01.T01	359	38.8	8.80
B1	*PsDof5*	PaF106G0604280000.01.T01	362	39.2	8.94
B2	*PsDof6*	PaF106G0607803000.01.T01	442	46.6	8.85
B2	*PsDof7*	PaF106G0708772700.01.T01	313	34.3	9.59
B2	*PsDof8*	PaF106G0403372100.01.T01	466	51.2	8.87
C1	*PsDof9*	PaF106G0604306300.01.T01	338	37.4	7.53
C1	*PsDof10*	PaF106G0503960200.01.T01	326	36.3	6.55
C1	*PsDof11*	PaF106G0403372400.01.T01	292	31.6	8.70
C2.1	*PsDof12*	PaF106G0503855900.01.T01	316	34.5	9.00
C2.1	*PsDof13*	PaF106G0203082700.01.T01	278	30.5	8.15
C2.1	*PsDof14*	PaF106G0202067300.01.T02	314	34.7	8.20
C2.2	*PsDof15*	PaF106G0810489100.01.T01	322	35.7	4.87
C2.2	*PsDof16*	PaF106G0810493600.01.T01	322	35.6	4.86
C2.2	*PsDof17*	PaF106G0705344600.01.T01	313	34.0	6.06
D1	*PsDof18*	PaF106G0705597800.01.T01	474	51.5	8.10
D1	*PsDof19*	PaF106G0100647100.01.T01	515	55.0	5.98
D1	*PsDof20*	PaF106G0302204300.01.T01	509	55.3	5.86
D1	*PsDof21*	PaF106G0504001500.01.T01	466	50.8	5.28
D1	*PsDof22*	PaF106G0405639500.01.T01	247	27.8	9.42
D2	*PsDof23*	PaF106G0100868700.01.T02	265	28.2	5.87
D2	*PsDof24*	PaF106G0403391500.01.T01	272	29.8	9.94

Supplementary Table 1. *PsDofs* duplicates in *Prunus sibirica.*

Seq_1		Seq_2		Duplicated model	Method	Ka	Ks	Ka/Ks
B1	PsDof4	B1	PsDof3	Segmental	NG	0.395273	2.145324	0.184249
B1	PsDof4	B1	PsDof5	Segmental	NG	0.451398	1.491249	0.302698
B1	PsDof5	B1	PsDof3	Segmental	NG	0.447382	2.720978	0.164419
C2.1	PsDof13	C2.1	PsDof14	Segmental	NG	0.332825	1.59872	0.208182
C2.1	PsDof13	C2.1	PsDof12	Segmental	NG	0.254364	1.12857	0.225386
C2.1	PsDof14	C2.1	PsDof12	Segmental	NG	0.320575	2.830863	0.113243
C2.2	PsDof15	C2.2	PsDof16	Segmental	NG	0.001318	0	NA
D1	PsDof19	D1	PsDof20	Segmental	NG	0.286441	1.538793	0.186146
D1	PsDof19	D1	PsDof21	Segmental	NG	0.332244	1.755818	0.189225
D1	PsDof20	D1	PsDof21	Segmental	NG	0.268606	1.317079	0.20394

Discussion

Dof transcription factors originated before the differentiation of green algae and terrestrial plants (Miguel *et al*., 2007; Moreno *et al*., 2010), Genome-wide surveys showed that multiple copies of *Dof* in higher plants and only one or two copies in lower plants (Ma *et al*., 2015). To date, no comprehensive analysis of the *PsDof* has been reported in *P. sibirica*, and the functions of *PsDof* are unclear. In this study, the *PsDof* family were analyzed, including sequence features, phylogeny, chromosomal locations, gene structures, duplication events, and expression pattern. A total of 24 *PsDofs* with a single *Dof* domain of CX_2 CX_{21} CX_2C zinc finger pattern were identified. The number of *PsDof* present in *P. sibirica* was similar as *P. persica* (25) and *grapevine* (25) (Da *et al*., 2016; Chen *et al*., 2017), and less than those in *A. thaliana* (36), *O. sativa* (30), and poplar (41) (Yanagisawa, 2002; Lijavetzky *et al*., 2003; Yang *et al*., 2006).

Gene replication includes tandem replication, fragment replication and whole-genome replication, which played an important role in the expansion of gene family in the process of evolution (Cannon *et al*., 2004). In *P. sibirica*, a total of ten segmental duplication and three sets of triplicate genes in 24 *PsDofs*, are reported with no tandem duplication.

The phylogenetic analysis showed the relationships of *Dofs* among *P. sibirica*, *A. thaliana,* and *O. sativa*, which were classified into four groups and ten subgroups. Most of the groups/subgroups contained different numbers of *Dofs* among the three species, such as, all subgroups contained *Dofs* of the *P. sibirica*, *A. thaliana*, and *O.sativa*, except subgroups C3 and D3, suggesting that the two subgroups may have existed before the species diverges.

The expression profiles are important clues for researching the putative functions of genes. *PsDofs* show the different expression patterns, suggesting that they may act a different part in growth and development. *PsDof1*

and *PsDof2* were high expressed in fruit, which suggested that played key role in the development of fruit, these findings were consistent with previous studies in which the *Dofs* were involved in fruit development and ripening of *Malus pumila* (Wang *et al*., 2021) and *Musa acuminata* (Feng *et al*., 2016). More than 50% of the *Dof* genes in *Oryza sativa* were expressed during the seed development process (Gaur *et al*., 2011), and in our study, *PsDof20* were highly expressed in the early development stage of kernel, which was homologous with *OsDof1* and played involved in the accumulation of grain protein and yield traits at grain filling stage (Nidhi *et al*., 2012). In addition, the high expression of *PsDof2*, *PsDof5*, and *PsDof9* were observed in leaves, flower buds, and flowers, respectively, suggesting the *PsDofs* displayed differential expression patterns among various tissues, which was consistent with the report in *Malus pumila* (Wang *et al*., 2021), *Solanum lycopersicum* (Cai *et al*., 2013), *Manihot esculenta* (Zou *et al*., 2018), and *Jatropha curcas* (Zou *et al*., 2019).

Supplementary Table 2. *PsDofs* duplicates in *P. sibirica* and *P. persica*.

Seq_1		Seq_2		Duplicated model	Method	Ka	Ks	Ka/Ks
A	PsDof1	A	PpeDof1	Segmental	NG	0.003934435	0.032056	0.1227357
A	PsDof2	A	PpeDof3	Segmental	NG	0.014782113	0.05971	0.24756441
B1	PsDof3	B1	PpeDof6	Segmental	NG	0.012702382	0.055729	0.22793314
B1	PsDof3	B1	PpeDof5	Segmental	NG	0.448308008	2.680058	0.16727549
B1	PsDof4	B1	PpeDof4	Segmental	NG	0.01835868	0.058074	0.31612669
B1	PsDof4	B1	PpeDof6	Segmental	NG	0.412142002	2.347838	0.17554106
B1	PsDof4	B1	PpeDof5	Segmental	NG	0.449245381	1.433281	0.3134384
B1	PsDof5	B1	PpeDof4	Segmental	NG	0.457510659	1.599041	0.28611562
B1	PsDof5	B1	PpeDof5	Segmental	NG	0.004804821	0.016119	0.29808764
B1	PsDof5	B1	PpeDof6	Segmental	NG	0.433582194	2.759777	0.15710768
B2	PsDof6	B2	PpeDof8	Segmental	NG	0.005915246	0.060913	0.09711002
B2	PsDof7	B2	PpeDof9	Segmental	NG	0.07503478	0.161501	0.46460979
B2	PsDof8	B2	PpeDof7	Segmental	NG	0.008026066	0.024277	0.33060044
C1	PsDof9	C1	PpeDof12	Segmental	NG	0.041091979	0.07638	0.53799722
C1	PsDof10	C1	PpeDof11	Segmental	NG	0.001293661	0.050609	0.02556167
C1	PsDof11	C1	PpeDof10	Segmental	NG	0.010560899	0.034348	0.30746957
C2.1	PsDof12	C2.1	PpeDof13	Segmental	NG	0.273522106	1.184133	0.23098935
C2.1	PsDof12	C2.1	PpeDof14	Segmental	NG	0.343832185	2.973967	0.11561398
C2.1	PsDof12	C2.1	PpeDof15	Segmental	NG	0.004140797	0.00921	0.44961592
C2.1	PsDof13	C2.1	PpeDof13	Segmental	NG	0.004982028	0.033745	0.14763686
C2.1	PsDof13	C2.1	PpeDof14	Segmental	NG	0.364326025	1.306909	0.27876921
C2.1	PsDof13	C2.1	PpeDof15	Segmental	NG	0.253294854	1.116708	0.22682275
C2.1	PsDof14	C2.1	PpeDof14	Segmental	NG	0.006054523	0.015818	0.38276623
C2.1	PsDof14	C2.1	PpeDof13	Segmental	NG	0.357183968	1.532305	0.23310233
C2.1	PsDof14	C2.1	PpeDof15	Segmental	NG	0.299892951	4.320895	0.06940528
C2.2	PsDof15	C2.2	PpeDof17	Segmental	NG	0.012166539	0.045183	0.26927222
C2.2	PsDof16	C2.2	PpeDof17	Segmental	NG	0.013530641	0.045183	0.29946279
C2.2	PsDof17	C2.2	PpeDof16	Segmental	NG	0.023822647	0.034948	0.68165814
D1	PsDof18	D1	PpeDof23	Segmental	NG	0.021479796	0.040498	0.53038801
D1	PsDof19	D1	PpeDof18	Segmental	NG	0.011911765	0.039895	0.29857456
D1	PsDof19	D1	PpeDof20	Segmental	NG	0.285475121	1.524316	0.18728085
D1	PsDof19	D1	PpeDof22	Segmental	NG	0.332096552	1.594718	0.20824781
D1	PsDof20	D1	PpeDof18	Segmental	NG	0.283736554	1.461601	0.19412719
D1	PsDof20	D1	PpeDof20	Segmental	NG	0.005115836	0.044109	0.11598237
D1	PsDof20	D1	PpeDof22	Segmental	NG	0.267252489	1.270039	0.21042857
D1	PsDof21	D1	PpeDof18	Segmental	NG	0.356295368	1.609586	0.22135837
D1	PsDof21	D1	PpeDof20	Segmental	NG	0.264561793	1.301785	0.20323005
D1	PsDof21	D1	PpeDof22	Segmental	NG	0.009289482	0.045841	0.20264669
D1	PsDof22	D1	PpeDof21	Segmental	NG	0.01620225	0.027359	0.59221659
D2	PsDof23	D2	PpeDof24	Segmental	NG	0	0.053883	0
D2	PsDof23	D2	PpeDof25	Segmental	NG	0.358263555	0.858885	0.41712655
D2	PsDof24	A	PpeDof2	Segmental	NG	0.130738468	0.180443	0.72454225

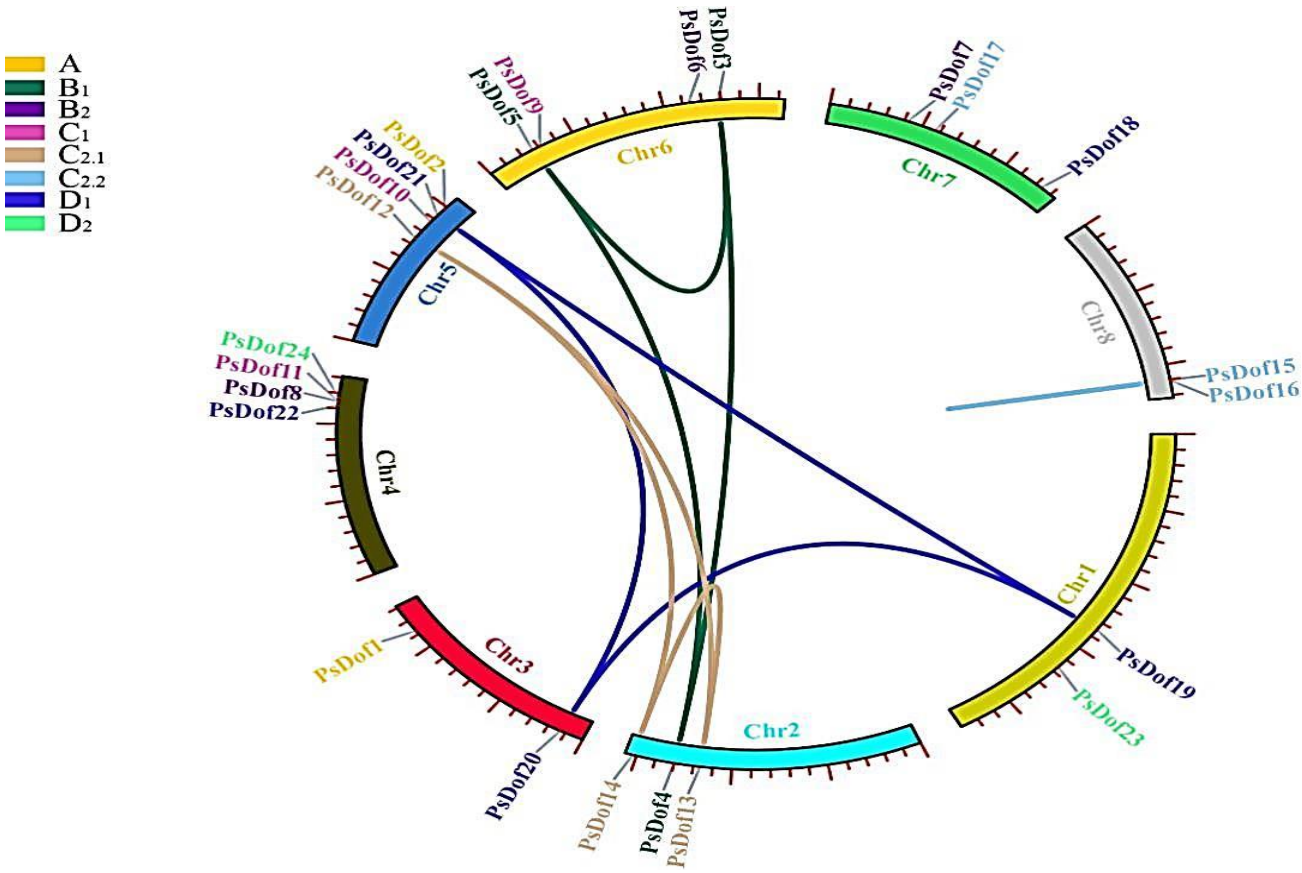

Fig. 3A. Chromosomal locations and collinear of *PsDofs*.

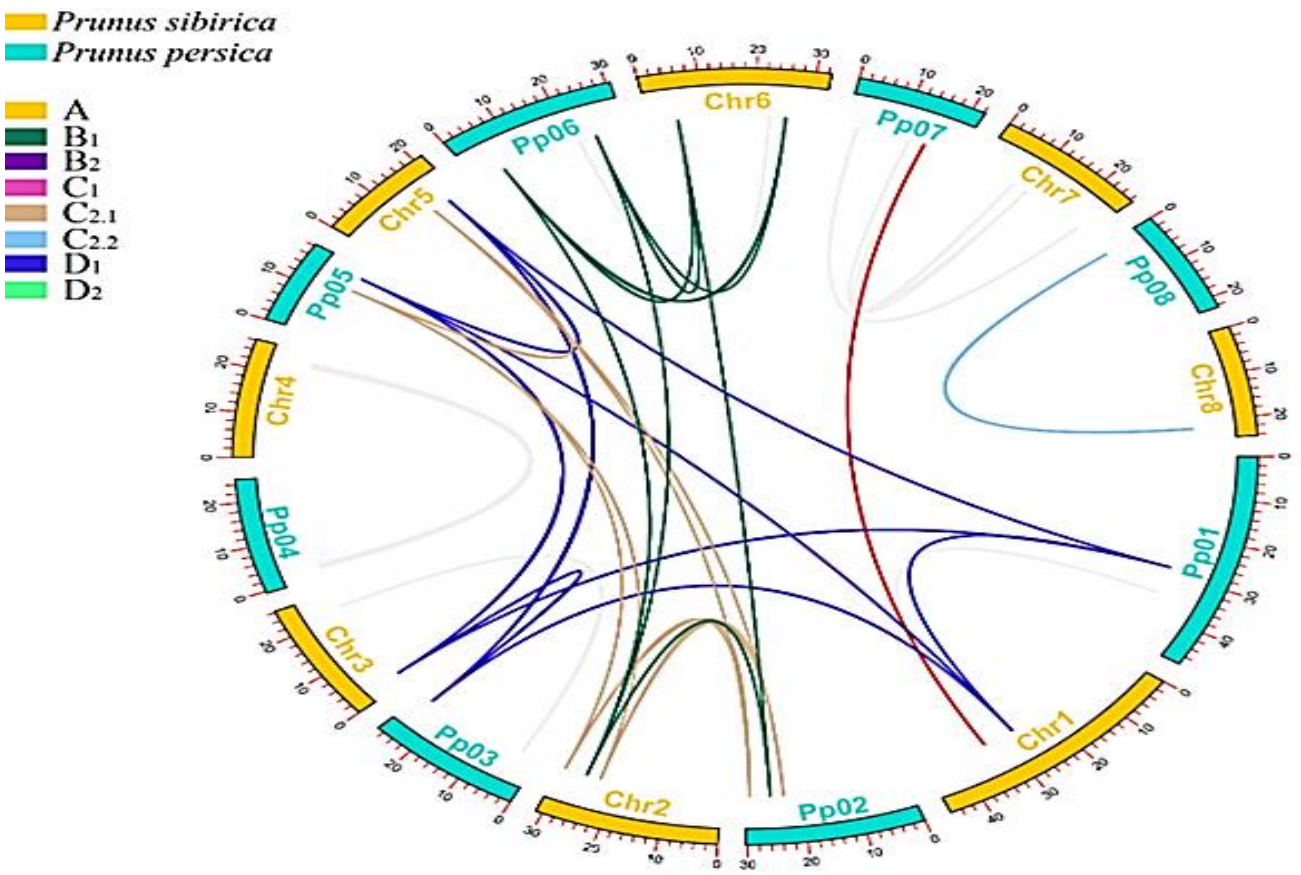

Fig. 3B. Collinear of *P. persica* and *P. sibirica.*

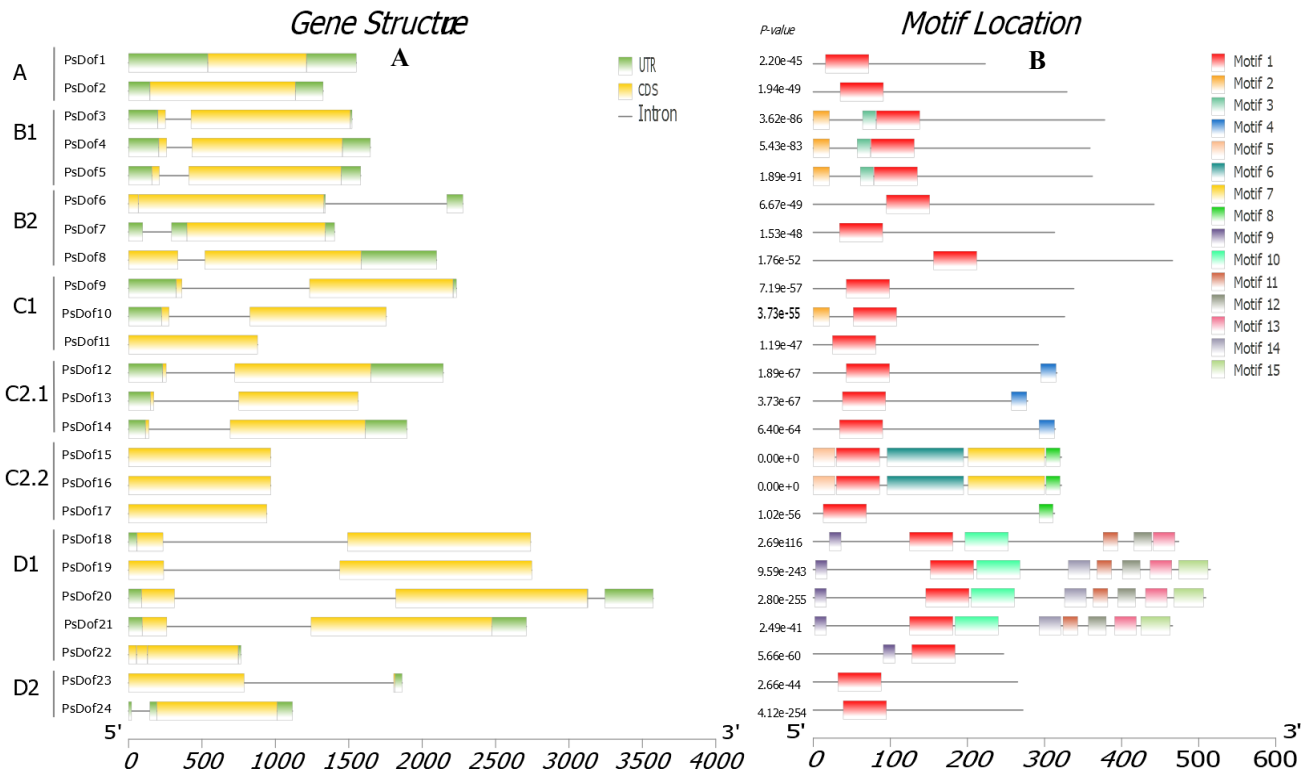

Fig. 4. Gene structure and conserved motifs of *PsDofs*. (A) The exon-intron structures displayed by using GSDS. (B) The distribution of conserved motifs of *Dof* proteins. The different motifs are represented by different color block.

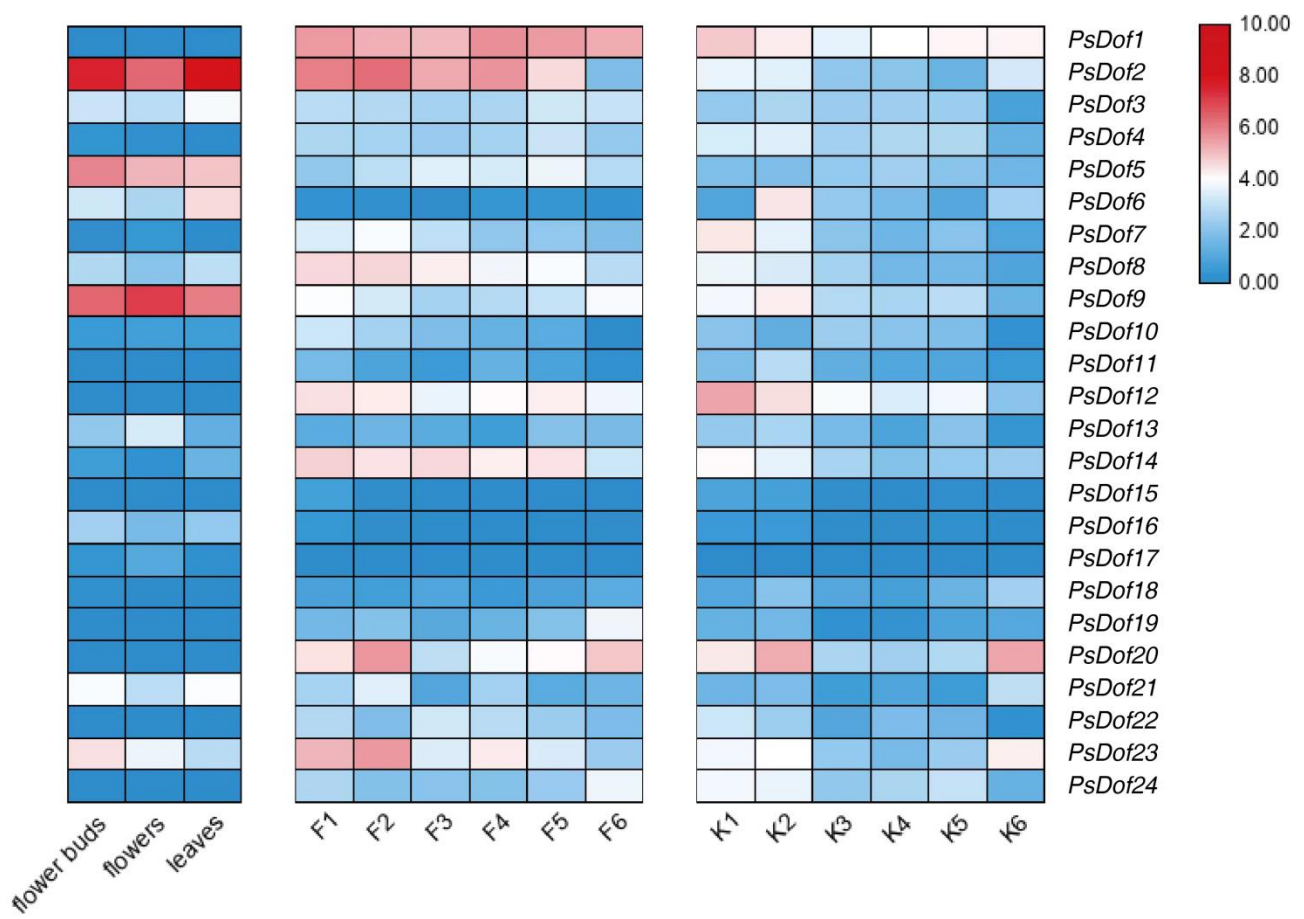

Fig. 5. The temporal and spatial expression patterns of *PsDofs*. Fruit (F) and Kernel (K) during different development stages 25(F/K1), 35(F/K2), 45(F/K3), 55(F/K4), 65(F/K5), and 75(F/K6) days after bloom (DAB).

Supplementary Table 3. Analysis of conserved motifs in *PsDof* proteins.

Motif[a]	Best match[b]	E_value	Width
Motif_1	EQALKCPRCDSTNTKFCYYNNYSLSQPRHFCKTCRRYWTKGGTLRNVPVGGGCRKNK	4.3E-1124	57
Motif_2	MVFSSIPVYLDPPNWQQQPNHH	1.80E-14	22
Motif_3	IRPGSMSDRARMAKIPQP	6.70E-08	18
Motif_4	QNKGQGDSTGYWNGMLGGGSW	7.50E-05	21
Motif_5	MFSAPVEQMLQCPSGPFITMDKRSWNKPH	1.50E-12	29
Motif_6	ADRASMSCFNHNSSSSDDTSGQYSSGTDNQPGGGNGSDIDLAAVFAKFLNNNSSNPAD EHDHLDQDHEPNLVISSSELNDVDGSQNSSKADQDLVEAVDH	4.10E-31	100
Motif_7	VAPDHHDHQHQQIQEENVQSFMGINHDQQQQDDMNIHQFGLQGLLGNDDQVVQDVF WSDDAATTSSLTSSTASFSWQPMVHLQELDYSLPSDDDHMKIPT	1.30E-48	100
Motif_8	LCSDNWSSFDFSGFEVFSR	2.40E-05	19
Motif_9	ESKDPAIKLFGKTIPL	2.60E-13	16
Motif_10	ASHYRHITISEALQTAQADAPNGAHHPSLKSNGSVLTFGGDAPLCESMASVLNLADK	1.00E-26	57
Motif_11	YPPAPYWGCAVPGPWNIPWL	6.20E-08	20
Motif_12	SGPNSPTLGKHSRDGDILKEESSE	1.70E-08	24
Motif_13	RVWIPKTLRIDDPSEAAKSSIWATLGIKN	2.90E-42	29
Motif_14	QIPCFPGAPWPYPWNSAQWPSPFPPPPFC	1.60E-15	29
Motif_15	GGLFKAFQSKGDQKNHVTEASPVLQANPAALSRSLNFQE	1.40E-21	39

Note: a Numbers correspond to the motifs described in Figure 4b.

Conclusion

In conclusion, we identified 24 *PsDofs* in *P. sibirica*. These genes were distributed on all eight chromosomes and were classified into four clusters. Gene structure and motifs were highly conserved of each group. We observed three sets of triplicate *Dofs* in both *P. persica* and *P. sibirica*, and analyzed the expression patterns of *PsDofs*, suggesting that *PsDofs* may play different roles in growth and development of *P. sibirica*. Our comprehensive analysis will provide a foundation for further studies the function of *Dofs* in *P. sibirica* and other species in Rosaceae.

Acknowledgements

This work was supported by the Fundamental Research Funds for the National Key R&D Program of China (2018YFD1000606, 2019YFD1000601) and the Central Non-profit Research Institution of Chinese Academy of Forestry (CAFYBB2020ZY003, CAFYBB 2017ZA004-4)

References

Baumann. 1999. The DNA binding site of the Dof protein NtBBF1 is essential for tissue-specific and auxin-regulated expression of the rolB oncogene in plants. *Plant Cell*, 11(3) : 323-333.

Cai, X.F., Y.Y. Zhang, C.J. Zhang, T.Y. Zhang, T.X. Hu, J. Ye, J.H. Zhang, T.T. Wang, H.Y. Li and Z. B. Ye. 2013. Genome-wide Analysis of Plant-specific Dof Transcription Factor Family in Tomato. *J. Integr. Plant Biol.,* 55(6)(-): 552-566.

Cannon, S.B., A. Mitra, A. Baumgarten, N.D. Young and G. May. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC. Plant. Biol.,* 4(1): 796-815.

Chen, C., X. Rui, C. Hao and Y. He. 2018. TBtools, a Toolkit for biologists integrating various HTS-data handling tools with a user-friendly interface. *BioRxiv*: 289660. doi. org/10.1101/289660.

Chen, M., X. Liu, L. Huan, M. Y. Sun, L. Liu, X. D. Chen, D. S. Gao and L. Li. 2017. Genome-wide analysis of Dof family genes and their expression during bud dormancy in peach (*Prunus persica*). *Sci. Hort.*, 214: 18-26.

Chen, W., G. Chao and K.B. Singh. 2010. The promoter of a H_2O_2-inducible, Arabidopsis glutathione S-transferase gene contains closely linked OBF- and OBP1-binding sites. *Plant. J.*, 10(6): 955-966.

Chen, X.Y., D.X. Wang, C. Liu, M.Z. Wang, T. Wang, Q. Zhao and J.J. Yu. 2012. Maize transcription factor ZmDof1 involves in the regulation of Zm401 gene. *Plant. Growth. Regul.*, 66(3): 271-284.

Da, D.C., V. Falavigna, M. Fasoli, V. Buffon, D.D. Porto, G.J. Pappas, M. Pezzotti, G. Pasquali and L.F. Revers. 2016. Transcriptome analyses of the Dof-like gene family in grapevine reveal its involvement in berry, flower and seed development. *Hort. Res.,* 3: 16042.

Dong, G., Z. Ni, Y. Yao, X. Nie and Q. Sun. 2007. Wheat Dof transcription factor WPBF interacts with TaQM and activates transcription of an alpha-gliadin gene during wheat seed development. *Plant. Mol. Biol.,* 63(1): 73-84.

Feng, B.H., Y.C. Han, Y.Y. Xiao, J.F. Kuang, Z.Q. Fan, J.Y. Chen and W.J. Lu. 2016. The banana fruit Dof transcription factor MaDof23 acts as a repressor and interacts with MaERF9 in regulating ripening-related genes. *J. Exp. Bot.,* (8): 2263-2275.

Finn, R.D., J. Clements and S.R. Eddy. 2011. HMMER web server: interactive sequence similarity searching. *Nucl. Acids. Res.*, 39 (Web Server issue): 29-37.

Finn, R.D., C. Penelope, R.Y. Eberhardt, S.R. Eddy, M. Jaina, A.L. Mitchell, S.C. Potter, P. Marco, Q. Matloob and S.V. Amaia. 2016. The Pfam protein families database: towards a more sustainable future. *Nucl. Acids. Res.*, 44(D1): D279-85.

Gaur, V.S., U.S. Singh and A. Kumar. 2011. Transcriptional profiling and in silico analysis of Dof transcription factor gene family for understanding their regulation during seed development of rice Oryza sativa L. *Eur. Mol. Biol. Organ. Rep.,* 38(4): 2827-2848.

Ivica, L., D. Tobias and B. Peer. 2015. SMART: recent updates, new developments and status in 2015. *Nucl. Acids. Res.*, 43 (Database issue): D257-60.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. Mcgettigan, H. Mcwilliam, F. Valentin, I.M. Wallace, A. Wilm and R. Lopez. 2007. Clustal W and Clustal X version

2.0. *Bioinformatics*, 23(21): 2947-2948.

Li, D., C. Yang, X. Li, Q. Gan, X. Zhao and L. Zhu. 2009. Functional characterization of *rice* OsDof12. *Planta*, 229(6): 1159-1169.

Lijavetzky, D., P. Carbonero and J.V. Carbajosa. 2003. Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC. Evol. Biol.*, 3(1): 631-7.

Ma, J., M.Y. Li, F. Wang, J. Tang and A.S. Xiong. 2015. Genome-wide analysis of Dof family transcription factors and their responses to abiotic stresses in Chinese cabbage. *BMC. Genom.*, 16(1): 33.

Miguel, Á.M., M. Martínez, J.V. Carbajos and P. Carbonero. 2007. The family of Dof transcription factors: from green unicellular algae to vascular plants. *Mol. Gen. Genet.*, 277(4): 379-390.

Moreno, M.Á., I. Díaz, L. Carrillo, R. Fuentes and P. Carbonero. 2010. The HvDof19 transcription factor mediates the abscisic acid-dependent repression of hydrolase genes in germinating barley aleurone. *Plant. J.*, 51(3): 352-365.

Nidhi, G., A.K. Gupta and C.T. Kumar. 2012. Spatial distribution pattern analysis of Dof1 transcription factor in different tissues of three Eleusine coracana genotypes differing in their grain colour, yield and photosynthetic efficiency. *Eur. Mol. Biol. Organ. Rep.,* 39(3): 2089-2095.

Sudhir, K., S. Glen, M. Li, K. Christina and T. Koichiro. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.,* 35(6): 1547-1549.

Vicente, J., S.P. Moose and P.R.J. Schmidt. 1997. A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator opaque2. *Procd. Natl. Acad. Sci. U. S. A.,* 94(14): 7685-7690.

Wang, H.W., B. Zhang, Y.J. Hao, J. Huang, A.G. Tian, J.S. Zhang and S.Y. Chen. 2007. The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic Arabidopsis plants. *The. Plant. J.*, 52(4): 716-729.

Wang, L. 2011. Resource investigation and distribution pattern of three Armeniaca species. *Forest. Res. Manag.*, (05): 65-70.

Wang, L. 2012. Evaluation of siberian apricot (*Prunus sibirica* L.) germplasm variability for biodiesel properties. *J. Amer. Oil. Chem. Soc.,* 89(9): 1743-1747.

Wang, X.L., L. Peng, J. Wang, J.J. Jia and L.P. Tang. 2021. Bioinformatics and expression analysis of Apple Dof transcription factor. *Jiangsu Agri. J.,* 37(02): 480-492.

Wang, Y., H. Tang, D.J. Du, T. Xu, J. Li, X. Wang, L.H. Jin, M. Barry and G. Hui. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucl. Acids. Res.*, 40(7): e49.

Yang, X.H., A. Tuskan and C.Z. Ming. 2006. Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. *Plant Physiol.*, 142(3): 820-830.

Yanagisawa, S. 2002. The Dof family of plant transcription factors. *Trends in Plant Sci.*, 7(12): 555-560.

Yin M.Y., H.M. Liu, W.Q. Bao, H. Zhao and T.N. Wuyun. 2017. Nucleolar phenotypic variation and superior plant selection of Inner Mongolia Siberian apricot. *Forest. Res.*, 030(006): 961-968.

Zou, Z. and X. Zhang. 2019. Genome-wide identification and comparative evolutionary analysis of the Dof transcription factor family in physic nut and castor bean. *Peer J.*, 7: e6354.

Zou, Z., J.L. Zhu and X.C. Zhang. 2018. Genome-wide identification and characterization of the Dof gene family in cassava (*Manihot esculenta*). *Gene,* 687.

(Received for publication 14 December 2020)