Genetics and Molecular Research

Genome-wide identification and characterization of the Dof gene family in *Medicago truncatula*

Y.J. Shu, L.L. Song, J. Zhang, Y. Liu and C.H. Guo

Key Laboratory of Molecular Cytogenetics and Genetic Breeding of Heilongjiang Province, College of Life Science and Technology, Harbin Normal University, Harbin, Heilongjiang, China

Corresponding author: C.H. Guo
E-mail: kaku_2008@163.com

Genet. Mol. Res. 14 (3): 10645-10657 (2015)
Received February 5, 2015
Accepted June 8, 2015
Published September 9, 2015
DOI http://dx.doi.org/10.4238/2015.September.9.5

ABSTRACT. The DNA-binding one zinc finger (Dof) family is a classic plant-specific zinc-finger transcription factor family, which is involved in many important processes, including seed maturation and germination, plant growth and development, and light responses. Investigation of the *Medicago truncatula* genome revealed 42 putative Dof genes, each of which holds one Dof domain. These genes were classified into four groups based on phylogenetic analysis, which are similar to the groups reported for *Arabidopsis* and rice. Based on genome duplication analysis, it was found that the MtDof genes were distributed on all chromosomes and had expanded through tandem gene duplication and segmental duplication events. Two main duplication regions were identified, one from tandem duplication and another from segmental duplication. By analyzing high-throughput sequencing data from *M. truncatula*, we found that most of the MtDof genes showed specific expression patterns in different tissues. According to cis-regulatory element analysis, these MtDof genes are

regulated by different cis-acting motifs, which are important for the functional divergence of the MtDof genes in different processes. Thus, using genome-wide identification, evolution, and expression pattern analysis of the Dof genes in *M. truncatula*, our study provides valuable information for understanding the potential function of the Dof genes in regulating the growth and development of *M. truncatula*.

Key words: Dof transcription factor; *Medicago truncatula*; Phylogenetic analysis; Genomic duplication; Expression patterns; Cis-acting element

INTRODUCTION

The DNA-binding one zinc finger (Dof) transcription factor (TF) family is a group of plant-specific TFs, which belong to the class of zinc finger domains. The Dof DNA-binding domain is usually located close to the N-terminal region of the Dof protein and is characterized by a binding domain of 52 amino acid residues that is structured as a Cys2Cys2 zinc finger, which binds specifically to DNA sequences with a core recognition site 5'-T/AAAAG-3' (Noguero et al., 2013). Since the first Dof gene, *ZmDof1*, was isolated from maize (Yanagisawa and Izui, 1993), a large number of Dof genes have been identified from various plant species, but no Dof genes have been isolated from other eukaryotes, such as yeast, *Drosophila*, or humans (Noguero et al., 2013). In plant species, Dof genes appear to be more recent in origin than other TFs in plants, there are few members in the plant genome, and the number of Dof genes shows little diversity among different plant species. For example, there are 36 members in *Arabidopsis* (Lijavetzky et al., 2003), 30 in rice (Lijavetzky et al., 2003), 28 in sorghum (Kushwaha et al., 2011), 27 in *Brachypodium* (Hernando-Amado et al., 2012), 34 in tomato (Cai et al., 2013), 21 in castor bean (Jin et al., 2014), and 78 in soybean (Guo and Qiu, 2013).

Dof TFs have been reported to have a great diversity of functions and participate in many plant-specific metabolism or regulation processes, such as light regulation, seed germination, plant growth and development, and response to stress. The first Dof gene identified was from maize, *ZmDof1*, and it acts as a transcriptional activator of light-regulated physiological process (Yanagisawa and Izui, 1993). In *Arabidopsis*, a number of Dof genes have been found to participate as transcriptional regulators in light regulation processes. For example, CDF1, CDF2, and CDF3 are all found to be involved in the photoperiodic control of flowering by repressing the *CONSTANS* gene (Imaizumi and Kay, 2006), while the three Dof genes *HPPBF3, COG1*, and *OBP3* also participate in light regulation processes mediated by phytochromes A and B and cryptochrome 4. Other Dof genes from *Arabidopsis* are associated with seed germination, including *DAG1*, *DAG2*, and *AtDof6* (Gabriele et al., 2010), while *OsDof3* plays a similar role in the seed germination process in rice (Washio, 2003). In cereal seeds, most Dof genes are expressed during both seed germination and maturation, recognizing the 5'-TGTAAAG-3' motif, and are termed Prolamin-box binding factors (PBF). These Dof genes have a high degree of sequence similarity to other Dof genes, including *ZmPBF* from maize (Schneidereit et al., 2008), *OsPBF* from rice (Yamamoto et al., 2006), *WPBF* from wheat (Mena et al., 1998; Dong et al., 2007), *BPBF* from barley (Mena et al., 1998), and *FMPBF*

from finger millet (Gupta et al., 2011). These genes have similar expression profiles with similar functions in the development of the endosperm during the seed-filling phase. Furthermore, several Dof genes are induced by environmental stimuli and they are important modulators of plant responses to biotic and abiotic stress. *OBP2* is induced by infection with *Spodoptera littoralis* and it regulates accumulation of indole glucosinolate to counteract stress (Skirycz et al., 2006). Two Dof genes from wheat are significantly upregulated under drought stress, indicating that they may be involved in drought adaptation (Shaw et al., 2009). In addition, Dof genes are involved in physical interactions with other TFs such as bZIP, MYB, and WKRY; therefore, they have been implicated in the regulation of plant physiological processes (Diaz et al., 2005).

Medicago truncatula is an excellent legume model plant because of its small, diploid genome, short life cycle, self-fertility, and high genetic transformation efficiency (Young et al., 2011). Although Dof genes have been characterized in other species of plants, their functions are poorly understood in *M. truncatula*. In this study, we identified the Dof gene family in *M. truncatula*, as well as the characteristics of gene structures, phylogenetic relationships, chromosomal locations, expression patterns of conserved motifs, and promoter analysis of Dof genes in *M. truncatula*.

MATERIAL AND METHODS

Identification and classification of Dof genes in *Medicago truncatula*

M. truncatula genome and protein sequences were downloaded from the JCVI website (*M. truncatula* Genome Project v4.0, http://www.jcvi.org/medicago/) (Young et al., 2011). The Dof sequences of *Arabidopsis* and rice were downloaded from the *Arabidopsis* Information Resource website (http://www.arabidopsis.org/, v9; Lamesch et al., 2012) and the Rice Genome Annotation Project website (http://rice.plantbiology.msu.edu/, v5; Kawahara et al., 2013), respectively, as described by Guo and Qiu (2013), and then these sequences were blasted (Altschul et al., 1990) against the *M. truncatula* genome with expected values $\leq 1E^{-5}$. All hits were retrieved and searched using the Hidden Markov Model profile of the Dof domain (PF002701), which was downloaded from the Pfam website (pfam.sanger.ac.uk) (Finn et al., 2011, 2014). The Dof genes were confirmed by the presence of the Dof domain, and all the putative Dof proteins were aligned to *Arabidopsis* and rice Dof proteins for classification into different groups. Furthermore, all the annotation information of putative Dof genes was retrieved from the *M. truncatula* genome website, and the numbers and distributions of introns in Dof genes were investigated using *M. truncatula* genome annotation information.

Phylogenetic analysis of the Dof genes

All the candidate Dof protein sequences were aligned using ClustalW with default parameters (Thompson et al., 2002), and phylogenetic trees of all MtDof proteins were generated using MEGA (v4.0) with the neighbor-joining method using the following parameters: Poisson correction, pairwise deletion, and bootstrap (1000 replicates) (Tamura et al., 2007).

Analysis of conserved motifs in Dof proteins

All MtDof protein sequences were analyzed using the Multiple EM for Motif Elicitation (MEME) v4.8.1 software package (Bailey et al., 2006). An MEME search was performed using the following parameters: 1) optimum motif width was set to ≥6 and ≤200, 2) the maximum number of motifs was set to identify 30 motifs, 3) occurrences of a single motif are distributed among the sequences with model: zero or one per sequence (-mod zoops). The MEME motifs were annotated using the Pfam database.

Chromosomal locations and gene duplication of the Dof genes

The relative sequences of the Dof genes (genomic sequences, CDS sequences) were collated from the *M. truncatula* genome database. The Dof genes were blasted against each other to identify gene duplication, of which the similarity of the aligned regions was more than 85%. Meanwhile, the position information of all these Dof genes was investigated to draw their chromosome location images in *M. truncatula*, and the duplicated genes between different chromosomes were linked with colored lines. This plot was created using the Circos software (http://circos.ca/; Krzywinski et al., 2009).

In silico expression analysis of the Dof genes from *M. truncatula*

Genome-wide transcriptome data from different development tissues of *M. truncatula* generated using high-throughput sequencing were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov, Accession No.: SRX099057-SRX099062). The transcriptome data included six tissue types: root, nodule, blade, bud, seedpod, and flower, and the details were shown as introduction of experiment data. All the transcriptome data were analyzed and clustered in Matlab (R2012a) using the Bioinformatics Toolbox.

Cis-regulatory element analysis

For promoter analysis, 1000-bp sequences upstream from translational start sites (TSS) of the putative Dof genes were retrieved from the *M. truncatula* genome. These sequences were then subjected to a search in the PLACE database (http://www.dna.affrc.go.jp/PLACE/signalscan.html; Higo et al., 1999) to identify cis-regulatory elements, and those motifs with clearly more than 200 copies were collected for annotation analysis.

RESULTS

Identification and classification of the Dof genes in *M. truncatula*

To identify the full complement of the Dof genes in *M. truncatula*, the Dof genes from *Arabidopsis* and rice were used to perform a BLASTp search against the *M. truncatula* genome, and 43 proteins were identified as predicted Dof genes. To confirm our results, the Dof domain (PF002701) was employed to search against these predicted Dof genes, and all predicted Dof genes containing a typical Dof domain in the N-terminal region, except Medtr5g011660, were

verified as Dof TF genes. According to their locations on the chromosomes, these deduced Dof genes were named *MtDof1* through *MtDof42*, as shown in Table 1. The amino acid sequence lengths of MtDofs varied from 112 to 495 amino acids, and more than half of the MtDof genes contained no or only one intron by investigation of intron numbers.

Table 1. Summary information of the MtDof genes in *Medicago truncatula*.

Gene name	Gene locus	Gene location	MCOG group	No. of amino acids	No. of introns
MtDof1	Medtr1g055265	MtChr1:24425414-24425752	A	112	0
MtDof2	Medtr1g056810	MtChr1:24878803-24879723	C	288	0
MtDof3	Medtr1g077600	MtChr1:34645345-34648290	C	272	4
MtDof4	Medtr1g115590	MtChr1:52275374-52275787	A	137	0
MtDof5	Medtr2g013370	MtChr2:3591319-3593054	B	274	0
MtDof6	Medtr2g014060	MtChr2:3908409-3909419	A	336	0
MtDof7	Medtr2g014170	MtChr2:3944814-3946421	B	309	1
MtDof8	Medtr2g016030	MtChr2:4821709-4822571	D	161	0
MtDof9	Medtr2g030030	MtChr2:11258026-11258442	A	138	0
MtDof10	Medtr2g059540	MtChr2:24563204-24563824	A	206	0
MtDof11	Medtr2g093220	MtChr2:39739900-39742251	C	293	1
MtDof12	Medtr2g096740	MtChr2:41334660-41336526	B	288	3
MtDof13	Medtr3g077750	MtChr3:34975012-34977217	C	336	1
MtDof14	Medtr3g090430	MtChr3:41085453-41086565	C	332	0
MtDof15	Medtr3g091820	MtChr3:41895073-41896886	C	306	1
MtDof16	Medtr3g435480	MtChr3:11623421-11626578	D	465	1
MtDof17	Medtr4g022370	MtChr4:7434526-7436301	B	364	1
MtDof18	Medtr4g063780	MtChr4:23662915-23664638	B	334	1
MtDof19	Medtr4g082060	MtChr4:31770129-31773332	D	465	1
MtDof20	Medtr4g088580	MtChr4:35201795-35204051	B	384	1
MtDof21	Medtr4g089095	MtChr4:35721012-35723110	B	298	2
MtDof22	Medtr4g109980	MtChr4:45762554-45764077	A	320	0
MtDof23	Medtr4g461080	MtChr4:22482796-22483629	C	277	0
MtDof24	Medtr5g031440	MtChr5:13480884-13481894	B	336	0
MtDof25	Medtr5g041380	MtChr5:18187491-18188871	D	371	1
MtDof26	Medtr5g041400	MtChr5:18192425-18194195	D	363	1
MtDof27	Medtr5g041420	MtChr5:18197045-18198772	D	322	1
MtDof28	Medtr5g041530	MtChr5:18229014-18231781	D	381	1
MtDof29	Medtr6g012450	MtChr6:3773706-3777176	D	495	1
MtDof30	Medtr6g027450	MtChr6:9423194-9426212	D	329	1
MtDof31	Medtr6g027460	MtChr6:9430394-9432744	D	368	1
MtDof32	Medtr7g010950	MtChr7:2821578-2824992	D	486	1
MtDof33	Medtr7g024670	MtChr7:8130987-8132842	B	373	1
MtDof34	Medtr7g059400	MtChr7:21548976-21550742	B	348	1
MtDof35	Medtr7g082600	MtChr7:31660203-31660607	A	134	0
MtDof36	Medtr7g086780	MtChr7:33750642-33752953	D	422	1
MtDof37	Medtr8g015840	MtChr8:5209981-5211545	A	218	2
MtDof38	Medtr8g027295	MtChr8:9592389-9594622	C	269	2
MtDof39	Medtr8g044220	MtChr8:16949084-16952235	D	439	1
MtDof40	Medtr8g068210	MtChr8:28437211-28438227	B	338	0
MtDof41	Medtr8g079060	MtChr8:33739368-33740718	D	229	0
MtDof42	Medtr8g479350	MtChr8:33823499-33824721	B	343	0

MCOG, major clusters of orthologous groups.

Phylogenetic analysis of the Dof genes in *M. truncatula*

Using homology searches against *Arabidopsis* and rice Dof genes, MtDof TF were divided into four groups, as shown in Table 1. To determine the phylogenetic relationships of MtDof genes in detail, a phylogenetic tree was constructed based on the alignment of full length sequences of MtDof proteins. The phylogenetic tree confirmed that most MtDof TFs were classified into four groups, termed major clusters of orthologous groups (MCOG) A, B,

C, and D, shown in Figure 1, and just three MtDof TFs (*MtDof6*, *MtDof13*, and *MtDof41*) were not correctly classified. There were eight members in MCOG A, 13 members in MCOG B, 7 members in MCOG C, 13 members in MCOG D, and only MtDof13 was not classified into any group.

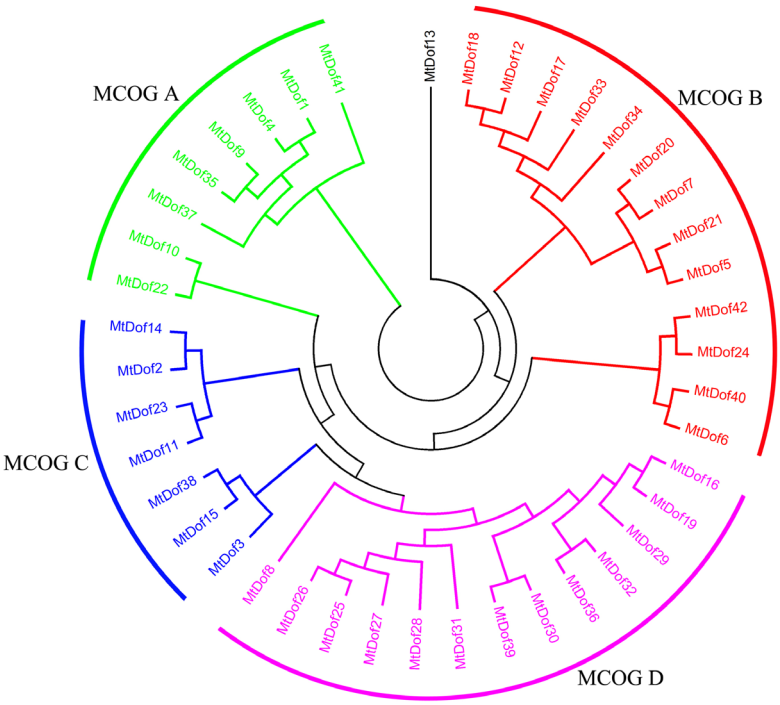

Figure 1. Phylogenetic tree of the MtDof genes in *Medicago truncatula*. The four major clusters of orthologous groups (MCOG) are highlighted.

Analysis of conserved motifs in Dof domain proteins

A total of 42 Dof genes from *M. truncatula* were further analyzed to identify conserved motifs shared among related proteins, and a total of 30 conserved motifs, named motif 1 to 30, were identified (Figure 2). Among these, the conserved motif encoding the Dof domain (Motif 1) was found in all of MtDof genes and was the most conserved motif in all the MtDof proteins. From the MEME results, we found that most of the closely related members in the phylogenetic tree had common motif compositions. MCOG A had two conserved motifs (motif 8 and 12), MCOG B and C had four similar motifs (motifs 14, 15, 29, and 30), while MCOG D had nine motifs (3, 4, 5, 6, 7, 11, 19, 21, and 22), and most of these were characteristic of MCOG D. The results of this motif analysis confirm that the Dof domain was conserved during the evolution of MtDof genes. Nevertheless, diversions of other motifs promoted differentiation of the Dof genes and the differences in motif distribution in different groups of the MtDof genes are sources of functional divergence in the MtDof genes.

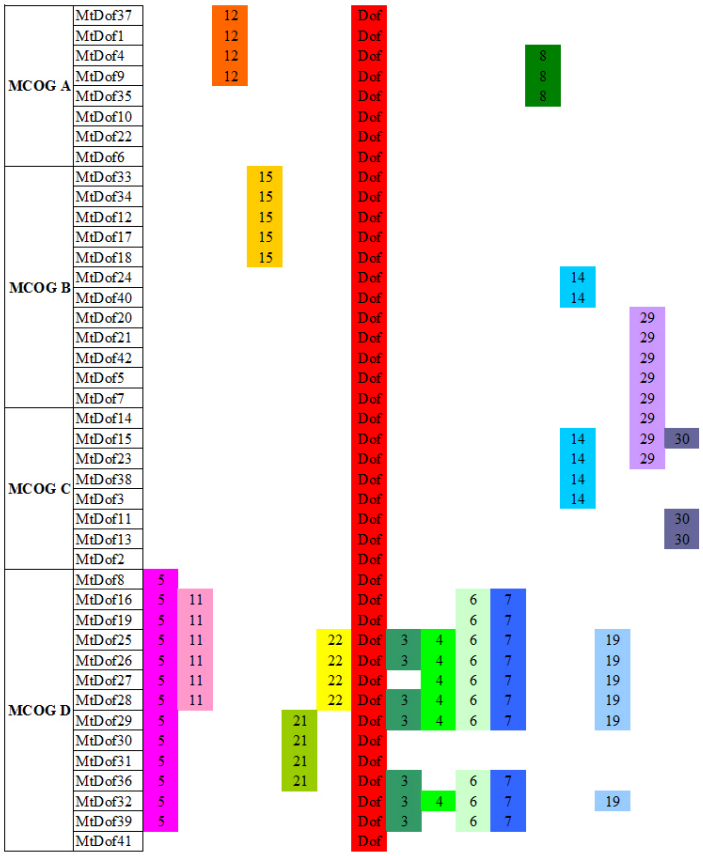

Figure 2. Distribution of conserved motifs in the MtDof proteins of *Medicago truncatula* identified using the Multiple EM for Motif Elicitation search tool. The proteins are divided into the four major clusters of orthologous groups (MCOG).

Chromosomal locations and duplication of the Dof genes in *M. truncatula*

The physical locations of the MtDof genes on *M. truncatula* chromosomes are displayed in Figure 3. In *M. truncatula*, the MtDof genes are distributed across the chromosomes and each chromosome holds some MtDof genes, ranging in number from three to eight. However, the MtDof genes are not randomly distributed on each chromosome; there are a number of gene clusters or gene hot regions on the chromosomes. For example, chromosome 5 has four MtDof genes in a short chromosome region (~45 kb), and chromosome 2 shows a similar gene cluster. In addition, using gene duplication analysis, we found 11 pairs of gene duplications, which arose from tandem duplications and segment duplications. Tandem duplications have resulted in MtDof gene clusters or hot regions, e.g., the MtDof cluster on chromosome 5, while segment duplication has resulted in many homologies of the MtDof genes between chromosomes, which have expanded the MtDof gene groups. For example, *MtDof1*, *4*, *9*, and *35* in the MCOG A are a product of genome segment duplication.

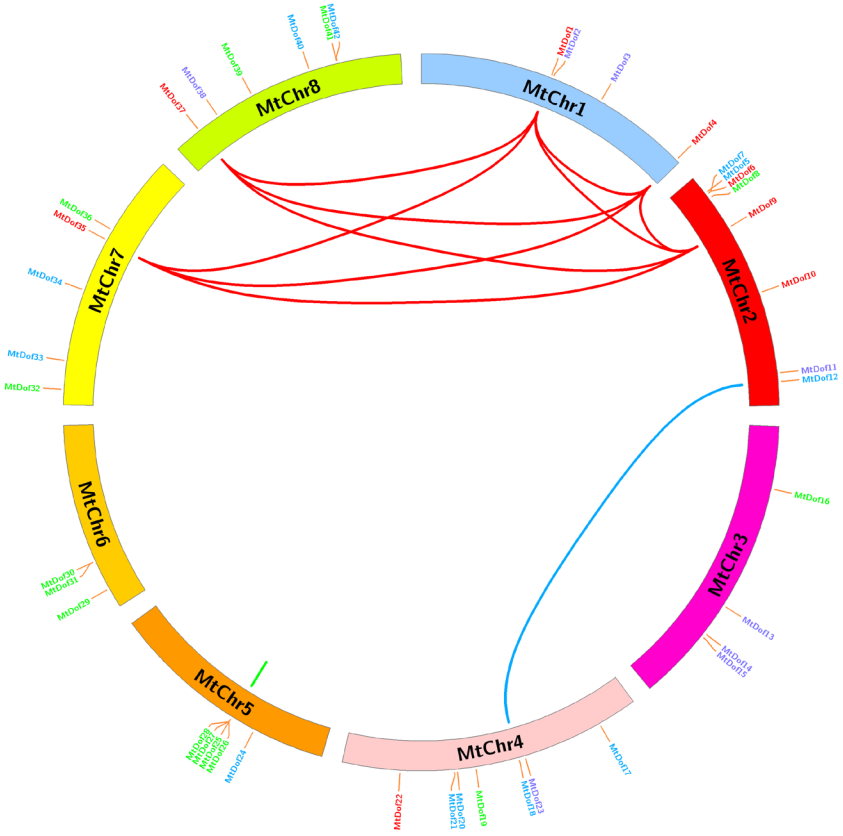

Figure 3. Chromosomal (chr) locations of the MtDof genes in *Medicago truncatula*. Red lines showed duplication between members of MCOG A, pure blue lines showed duplication between members of MCOG B, and green lines showed duplication between members of MCOG D.

Expression patterns analysis of the MtDof genes

Since the employment of high-throughput sequencing technology to determine gene sequences and conduct expression analysis, large quantities of sequence data have been deposited in NCBI. We downloaded *M. truncatula* transcriptome sequencing data for different development tissues, and from this, we collected all the MtDof gene expression data. Based on their expression patterns, the MtDof genes clustered into six groups, shown in Figure 4. Cluster A included 13 MtDof genes, most of them members of MCOG D, and they were expressed in nodules, blades, and buds. Clusters B, D, and F included 15 members in total, which were mainly from MCOG B and these genes were highly expressed in buds, seedpods, and flowers. Cluster C consisted of seven MtDof genes from MCOG A, three of them were highly expressed in roots, and the other four genes (*MtDof1*, *4*, *9*, and *35*) were not expressed in any tissue. The final cluster, E, consisted of seven Dof genes, which were mainly members of MCOG C and had broad expression patterns, being expressed in roots, buds, seedpods, and flowers, although their expression levels were not high.

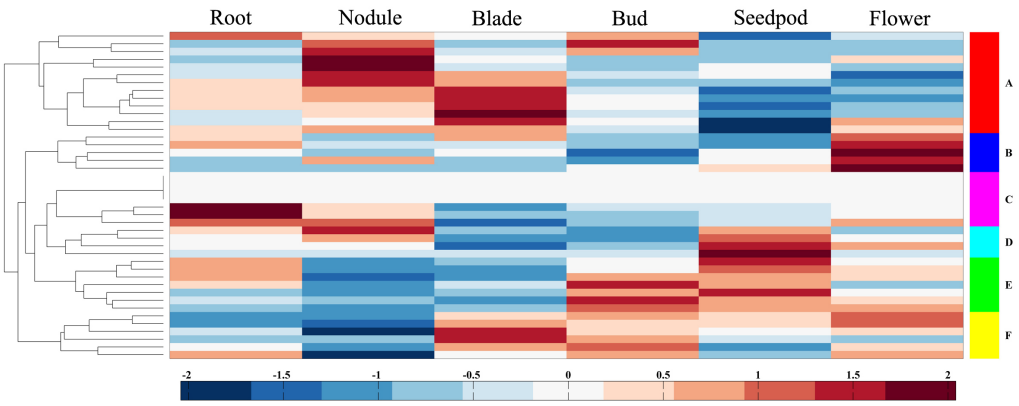

Figure 4. Heat map showing expression of MtDof in different tissues of *Medicago truncatula*, based on high-throughput sequencing data. The legend showed expressional level of MtDof, blue represented low expression, while red represented high expression. The letters A-F and relevant color paint indicated cluster A-F mentioned in manuscript.

Cis-regulatory element analysis of the MtDof genes

The 1000 bp sequences located upstream from the TSS of the MtDof genes were submitted to the PLACE website to identify putative cis-elements, and in total 220 motifs were identified from the TSS sequences, and those motifs with more than 200 copies were selected for further analysis (Table 2). The most abundant motif was the Dof binding site (DOFCOREZM, 754 copies), ranging from 9 to 35 copies in each MtDof gene. Other motifs, including some related to tissue development (ROOTMOTIFTAPOX1, POLLEN1LELAT52, CACTFTPPCA1, GTGANTG10, and CAATBOX1), some responsive to light (TATABOX5, GT1CONSENSUS, and GATABOX), and other TF-binding motifs (MYCCONSENSUSAT, WRKY71OS, ARR1AT, and EBOXBNNAPA), were also overrepresented in the promoter regions of the MtDof genes (details are shown in **Table S1**).

Table 2. Summary information of cis-acting regulatory DNA elements from the MtDof genes in *Medicago truncatula.*

Site ID	Site name	Motif sequence	No. of copies
S000028	CAATBOX1	CAAT	532
S000039	GATABOX	GATA	356
S000098	ROOTMOTIFTAPOX1	ATATT	383
S000144	EBOXBNNAPA	CANNTG	320
S000198	GT1CONSENSUS	GRWAAW	475
S000203	TATABOX5	TTATTT	225
S000245	POLLEN1LELAT52	AGAAA	313
S000265	DOFCOREZM	AAAG	754
S000378	GTGANTG10	GTGA	317
S000407	MYCCONSENSUSAT	CANNTG	320
S000447	WRKY71OS	TGAC	261
S000449	CACTFTPPCA1	YACT	683
S000454	ARR1AT	NGATT	424

DISCUSSION

In the present study, using comparative genomic and phylogenetic analysis, 42 MtDof genes were identified from *M. truncatula*, and these were classified into four groups, which is consistent with reports for other species, such as *Arabidopsis* and rice (Lijavetzky et al., 2003), and *Brachypodium* (Hernando-Amado et al., 2012). However, the number of MtDof genes belonging to the MCOG B group (13 MtDof genes) was slightly more than previously reported. In addition, all of the MtDof genes identified in the current study contained very few introns (most of them included no or only one intron; 88%, 37/42), which is similar to the discoveries in *Arabidopsis* and rice. Possessing fewer introns is thought to make MtDof genes more sensitive to transcriptional regulation, which facilitates a plant strong ability to adapt to diverse development processes and environmental stimuli (Jin et al., 2014).

The phylogenetic tree and conservation domain analysis showed that MtDof genes in the same group share similar motifs, and these conserved motifs play important roles in group specific functions. MtDof genes with similar motif compositions are likely to have emerged by gene duplication, including tandem duplications and segment duplications. For example, *MtDof25*, *26*, *27*, and *28* form a cluster of MtDof genes resulting from tandem duplications of a common ancestor MtDof gene and four genes containing similar motif compositions. These four genes were all classified into MCOG D, and furthermore, they had similar expression profiles in different tissues. Similarly, *MtDof1*, *4*, *9*, and *35* are the result of segment duplication events, but they are distributed on different chromosomes.

In addition to conservation motifs in proteins, promoter sequences have also played crucial roles in determining the divergence of MtDof gene functions. The results of the cis-elements analysis confirmed that the functional diversities of the MtDof genes are mainly involved in tissue development, response to environmental stress, and interactions with other TFs. For example, motifs related to tissue development, including CAATBOX1 related to tissue-specific promoter activity (Shirsat et al., 1989), ROOTMOTIFTAPOX1 related to root development (Elmayan and Tepfer, 1995), POLLEN1LELAT52 and GTGANTG10 related to flower development (Rogers et al., 2001; Filichkin et al., 2004), and CACTFTPPCA1 related to leaf development (Gowik et al., 2004), were extensively present in the MtDof genes, and thus, most of these were highly expressed in roots, blades, buds, flowers, and seedpods. In addition, light responsive elements were also identified, such as GT1CONSENSUS (Terzaghi and Cashmore, 1995), and these were widely represented in the MtDof genes.

Meanwhile, transcriptome analysis showed that most of the MtDof genes were widely expressed in different tissues, indicating that they may be involved in diverse physiological functions, confirming the functional divergence of the MtDof genes. It was notable that four MtDof genes, *MtDof1*, *4*, *9*, and *35*, were not expressed in any tissues. These MtDof genes are very similar to each other and they were duplicated from one locus and translocated onto different chromosomes by segment duplication, shown in Figure 3. These genes appeared to expand the MCOG A group; however, these genes are actually pseudogenes and they were not expressed in any of the six tissues analyzed, but they may be induced by other conditions not assessed in this study.

Considering the complexity of transcriptional regulation, TFs also control each other to perform more exact regulation. In previous reports, the Dof domain was known as a bi-functional domain, which was mediated not only by DNA-bindings but also by protein-protein

physical interactions (Zhang et al., 1995). From promoter analysis, we identified a number of TF binding sites in promoters of the MtDof genes, including DOFCOREZM, MYCCONSENSUSAT, WRKY71OS, ARR1AT, and EBOXBNNAPA. All the MtDof genes identified contained more than nine copies of DOFCOREZM elements, indicating that regulation by themselves was crucial for the execution of their functions (Yanagisawa and Schmidt, 1999). In addition, discovery of other TF-binding sites suggests that MtDof genes may be induced by those TFs, which mediate the response of the MtDof genes to other processes. For example, the ARR1AT element has been shown to precipitate the response of MtDof genes to auxin (Kim et al., 2010), while binding sites of the TF WRKY allow the MtDof genes to participate in responses to biotic stress or salicylic acid treatment, as described by Jin et al. (2014).

In summary, we identified 42 MtDof genes in *M. truncatula*, which were classified into four groups, consistent with previous studies. The classification, evolution, expression profiles, and promoters of these MtDof genes were investigated, and the results showed that the MtDof genes participate in regulation of plant tissue development processes. The information from this investigation will be useful for MtDof gene identification and characterization. However, further functional analyses of these genes will be needed to explore their biological roles in *M. truncatula*.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the MOST ‘‘863’’ project (#2013AA102607-5), grants from the Natural and Science Foundation of China (#31302019 and #31470571), and the Heilongjiang Province Postdoctoral Science Foundation (#LBH-Z14126).

Supplementary material

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, et al. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.

Bailey TL, Williams N, Misleh C and Li WW (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34: W369-W373.

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

Diaz I, Martinez M, Isabel-LaMoneda I, Rubio-Somoza I, et al. (2005). The DOF protein, SAD, interacts with GAMYB in plant nuclei and activates transcription of endosperm-specific genes during barley seed development. *Plant J.* 42: 652-662.

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

Elmayan T and Tepfer M (1995). Evaluation in tobacco of the organ specificity and strength of the rolD promoter, domain A of the 35S promoter and the 35S2 promoter. *Transgenic Res.* 4: 388-396.

Filichkin SA, Leonard JM, Monteros A, Liu PP, et al. (2004). A novel endo-beta-mannanase gene in tomato *LeMAN5* is associated with anther and pollen development. *Plant Physiol.* 134: 1080-1087.

Finn RD, Clements J and Eddy SR (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39: W29-W37.

Finn RD, Bateman A, Clements J, Coggill P, et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42: D222-D230.

Gabriele S, Rizza A, Martone J, Circelli P, et al. (2010). The Dof protein DAG1 mediates PIL5 activity on seed germination by negatively regulating GA biosynthetic gene *AtGA3ox1*. *Plant J.* 61: 312-323.

Gowik U, Burscheidt J, Akyildiz M, Schlue U, et al. (2004). cis-regulatory elements for mesophyll-specific gene expression in the C_4 plant *Flaveria trinervia*, the promoter of the C_4 phosphoenolpyruvate carboxylase gene. *Plant Cell* 16: 1077-1090.

Guo Y and Qiu LJ (2013). Genome-wide analysis of the Dof transcription factor gene family reveals soybean-specific duplicable and functional characteristics. *PLoS One* 8: e76809

Gupta N, Gupta AK, Singh NK and Kumar A (2011). Differential expression of *PBF Dof* transcription factor in different tissues of three finger millet genotypes differing in seed protein content and color. *Plant Mol. Biol. Rep.* 29: 69-76.

Hernando-Amado S, Gonzalez-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: 202.

Higo K, Ugawa Y, Iwamoto M and Korenaga T (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* 27: 297-300.

Imaizumi T and Kay SA (2006). Photoperiodic control of flowering: not only by coincidence. *Trends Plant Sci.* 11: 550-558.

Jin Z, Chandrasekaran U and Liu A (2014). Genome-wide analysis of the Dof transcription factors in castor bean (*Ricinus communis* L.). *Genes Genomics* 36: 527-537.

Kawahara Y, de la Bastide M, Hamilton J, Kanamori H, et al. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6: 4.

Kim HS, Kim SJ, Abbasi N, Bressan RA, et al. (2010). The DOF transcription factor Dof5.1 influences leaf axial patterning by promoting revoluta transcription in *Arabidopsis*. *Plant J.* 64: 524-535.

Krzywinski M, Schein J, Birol I, Connors J, et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19: 1639-1645.

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.

Lamesch P, Berardini TZ, Li D, Swarbreck D, et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40: D1202-1210.

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.

Mena M, Vicente-Carbajosa J, Schmidt RJ and Carbonero P (1998). An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin-box of a native B-hordein promoter in barley endosperm. *Plant J.* 16: 53-62.

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.

Rogers HJ, Bate N, Combe J, Sullivan J, et al. (2001). Functional analysis of cis-regulatory elements within the promoter of the tobacco late pollen gene *g10*. *Plant Mol. Biol.* 45: 577-585.

Schneidereit A, Imlau A and Sauer N (2008). Conserved cis-regulatory elements for DNA-binding-with-one-finger and homeo-domain-leucine-zipper transcription factors regulate companion cell-specific expression of the *Arabidopsis thaliana SUCROSE TRANSPORTER 2* gene. *Planta* 228: 651-662.

Shaw LM, McIntyre CL, Gresshoff PM and Xue GP (2009). Members of the Dof transcription factor family in *Triticum aestivum* are associated with light-mediated gene regulation. *Funct. Integr. Genomics* 9: 485-498.

Shirsat A, Wilford N, Croy R and Boulter D (1989). Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. *Mol. Gen. Genet.* 215: 326-331.

Skirycz A, Reichelt M, Burow M, Birkemeyer C, et al. (2006). DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in *Arabidopsis*. *Plant J.* 47: 10-24.

Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.

Terzaghi WB and Cashmore AR (1995). Light-regulated transcription. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 46: 445-474.

Thompson JD, Gibson TJ and Higgins DG (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics* 2: 2.3.

Washio K (2003). Functional dissections between GAMYB and Dof transcription factors suggest a role for protein-protein associations in the gibberellin-mediated expression of the *RAmy1A* gene in the rice aleurone. *Plant Physiol.* 133: 850-863.

Yamamoto MP, Onodera Y, Touno SM and Takaiwa F (2006). Synergism between RPBF Dof and RISBZ1 bZIP activators in the regulation of rice seed expression genes. *Plant Physiol.* 141: 1694-1707.

Yanagisawa S and Izui K (1993). Molecular cloning of two DNA-binding proteins of maize that are structurally different but interact with the same sequence motif. *J. Biol. Chem.* 268: 16028-16036.

Yanagisawa S and Schmidt RJ (1999). Diversity and similarity among recognition sequences of Dof transcription factors. *Plant J.* 17: 209-214.

Young ND, Debelle F, Oldroyd GE, Geurts R, et al. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480: 520-524.

Zhang B, Chen W, Foley RC, Buttner M, et al. (1995). Interactions between distinct types of DNA binding proteins enhance binding to ocs element promoter sequences. *Plant Cell* 7: 2241-2252.

