BioSystems 151 (2017) 27–33

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Genome-wide characterization of the SiDof gene family in foxtail millet (*Setaria italica*)

Li Zhang[a], Baoling Liu[a], Gewen Zheng[c], Aiying Zhang[b], Runzhi Li[a,*]

[a] Institute of Molecular Agriculture & Bioenergy, Shanxi Agricultural University, Taigu, 030801, China
[b] Institute of Millet, Shanxi Academy of Agricultural Sciences, Changzhi, 046011, China
[c] Seed Management Station, Agriculture Department of Shanxi Province, Taiyuan, 030006, China

ARTICLE INFO

Article history:
Received 20 February 2016
Received in revised form
21 November 2016
Accepted 24 November 2016
Available online 25 November 2016

Keywords:
Setaria italica
Dof transcription factors
Phylogenetic analysis
Expression profiling

ABSTRACT

Dof (DNA binding with one finger) proteins, which constitute a class of transcription factors found exclusively in plants, are involved in numerous physiological and biochemical reactions affecting growth and development. A genome-wide analysis of *SiDof* genes was performed in this study. Thirty five *SiDof* genes were identified and those genes were unevenly distributed across nine chromosomes in the *Seteria italica* genome. Protein lengths, molecular weights, and theoretical isoelectric points of SiDofs all vary greatly. Gene structure analysis demonstrated that most *SiDof* genes lack introns. Phylogenetic analysis of SiDof proteins and Dof proteins from *Arabidopsis thaliana*, rice, sorghum, and *Setaria viridis* revealed six major groups. Analysis of RNA-Seq data indicated that *SiDof gene* expression levels varied across roots, stems, leaves, and spike. In addition, expression profiling of *SiDof* genes in response to stress suggested that *SiDof* 7 and *SiDof* 15 are involved in drought stress signalling. Overall, this study could provide novel information on SiDofs for further investigation in foxtail millet.

1. Introduction

The Dof (DNA binding with one finger) protein family belongs to a class of transcription factors exclusively found in plants. These proteins have been identified in different plant species and are involved in the regulation of numerous biochemical reactions during growth and development. Such reactions include responses to phytohormones (Boccaccini et al., 2014; Moreno-Risueno et al., 2007), fatty acid biosynthesis (Wang et al., 2007), carbon and nitrogen metabolism (Kurai et al., 2011; Santos et al., 2012), responses to light (Santopolo et al., 2015), glucosinolate biosynthesis (Skirycz et al., 2006), and abiotic stress responses (Ma et al., 2015). These functions have been specifically demonstrated in several major crops. For example, transgene studies found that Zm*Dof1* (MNB1a) from maize can regulate gene expression in response to light, as well as metabolic pathways in other C4 plants (Yanagisawa and Sheen, 1998). In rice, *OsDof 13* expression alters significantly in response to stress under nitrogen-poor conditions (Washio, 2001), and *OsDof 3* appears to be involved in the gibberellic acid response pathway during seed germination (Washio, 2003).

Dof proteins are typically 200–400 amino acids in size and have two regions: a conserved, specific DNA-binding domain in the N-terminal and a variable C-terminal regulatory domain (Imaizumi et al., 2005). Typically, a Ser residue exists as a skeletal structure between the N- and C-terminal domains, possibly acting as a molecular hinge connecting them (Umemura et al., 2004). The N-terminal consists of approximately 52 amino acids with a single Cys2/Cys2 zinc finger structure in a conserved $CX_2CX_{21}CX_2C$ motif, which recognizes (AT)/AAAG sites in DNA (Yanagisawa, 2002). The variable C-terminal allows Dof proteins to be regulated by signals from multiple metabolic pathways, resulting in numerous transcriptional outputs and a diverse range of functions (Noguero et al., 2013).

The Dof gene family has been characterized in *Arabidopsis* (Yanagisawa, 2002), rice (Lijavetzky et al., 2003), sorghum (Kushwaha et al., 2011), soybean (Guo and Qiu, 2013), tomato (Cai et al., 2013) potato (Venkatesh and Park, 2015), Chinese cabbage (Ma et al., 2015), pigeon pea (Malviya et al., 2015), cucumber (Wen et al., 2016), and pepper (Wu et al., 2016). However, no full-scale studies have been undertaken in foxtail millet (*Setaria italica*), a stress-tolerant crop with a short life cycle, high inbreeding levels, and a small genome characterized by relatively low counts of repetitive DNA. These traits make *S. italica* a perfect model for biological investigations in the related Panicoid family, including questions on

* Corresponding author.
E-mail addresses:
zhangli7912@163.com (L. Zhang), 839856567@qq.com (B. Liu), liqy0126@126.com (G. Zheng), zay1012@126.com (A. Zhang), rli2001@126.com (R. Li).

http://dx.doi.org/10.1016/j.biosystems.2016.11.007
0303-2647/

genome architecture, phylogeny, and physiology (Lata et al., 2013). Moreover, the *S. italica* genome recently became available, along with RNA sequencing data from different tissues under drought stress (Zhang et al., 2012; Qi et al., 2013), increasing the research applicability of this crop.

In this study, we used available *S. italica* resources to perform a broad, genome-wide characterization of the *S. italica* Dof (SiDof) gene family. Based on genomic sequences, we identified 35 SiDof proteins in *S. italica* that we further divided into six groups. Their gene structure, chromosomal distribution, and motifs were also investigated. Additionally, *SiDof* gene functions were predicted through evaluations of expression profiles across different tissues and under drought stress. These results will provide the foundation for further research on *SiDofs* functions in drought resistance.

2. Materials and methods

2.1. Database queries

The *S. italica* genome was searched in Phytozome version 11.0 (https://phytozome.jgi.doe.gov/pz/portal.html) using 'DOF' and 'Setaria italica' as keywords. In addition, 130 Dof transcription-factor protein sequences of four species (*Arabidopsis thaliana*, rice, sorghum, and *Setaria viridis*) were downloaded from PTFD 3.0 (http://planttfdb.cbi.pku.edu.cn/family.php?fam=DOF)(Table S1). Each non-redundant sequence was checked in the SMART (Simple Modular Architecture Research Tool) database (http://smart.emblheidelberg.de) for the presence of the conserved Dof domain.

2.2. Chromosomal localization and gene structure analysis of SiDofs

For each S*iDof* gene, the chromosome number, plus the gene start and end positions, were identified in Phytozome version 11.0 (Table 1). The *SiDof* genes were then plotted onto the respective chromosomes according to the ascending order of their physical position (in base pairs, bp). The resultant physical map was created in MapChart version 2.2. The gene structure of *SiDofs* was constructed via aligning cDNA sequences with corresponding genomic DNA sequences in the Gene Structure Display Server (GSDS 2.0, http://gsds.cbi.pku.edu.cn/index.php).

2.3. Motifs analysis of the SiDof proteins

Protein motifs were identified using MEME (Multiple Expectation Maximization for Motif Elicitation, http://meme.nbcr.net/meme/). The analysis was performed with the following settings: number of repetitions, any; maximum number of motifs, 15; and optimum width of the motif, 10–60.

2.4. Dof proteins sequences alignment and phylogenetic tree construction

Mutiple sequence alignment was performed on the full protein sequences of *S. italica* and four other species (*A. thaliana*, rice, sorghum, and *S. viridis*). The alignment was then used to construct a maximum likelihood phylogenetic tree in MEGA 6.0. Bootstrap values were calculated with 1000 iterations.

2.5. In silico expression profiling of SiDofs

RNA sequencing data from four tissue types in *S. italica* (spica, stem, leaf, and root) were retrieved from the European Nucleotide Archive (http://www.ebi.ac.uk/ena SRX128226 [spica], SRX128225 [stem], SRX128224 [leaf], SRX128223 [root)]). In addition, RNA sequencing data from the *S. italica* dehydration-stress library (SRR629694) and control library (SRR629695) were retrieved from the European Nucleotide Archive (http://www.ebi.ac.uk/ena) (Qi et al., 2013). All paired-end Illumina reads from the four different tissue samples were mapped onto the *S. italica* gene sequences using Bowtie2, and the number of mapped reads was normalized with the RPKM (reads per kilobase per million) method. Based on the RPKM value of each *SiDof* gene in the respective tissues, a heat map showing tissue-specific expression was generated using TIGR Multi Experiment Viewer (MeV4.9.0) software.

3. Results and discussion

3.1. Validation of SiDof proteins and analysis of physicochemical properties

The keyword search in Phytozome yielded 36 *SiDof* genes. Verification of Dof domain presence using SMART resulted in the exclusion of one gene lacking the conserved Dof domain (Seita.1G336300). Thirty-five transcripts were named (*SiDof* 1 to *SiDof* 35) according to their order on the chromosomes (Table 1).

We used the ProtParam tool from ExPASy (http://web.expasy.org/protparam/) to analyse the protein size, molecular weight, and theoretical isoelectric point (pI) of these SiDof proteins. The longest (623 amino acids, aa) and shortest (168 aa) protein sequences belonged to SiDof 15 and SiDof 17, respectively. The protein lengths of other SiDof proteins were approximately 328 aa. Molecular weights of the identified SiDof proteins ranged from 17.19 kDa (SiDof 17) to 67.89 kDa (SiDof 15), and theoretical pI varied from 4.81 (SiDof 19) to 10.06 (SiDof 17).

3.2. Chromosomal localization of SiDof genes

A chromosome localization analysis of *SiDof* genes on *S. italica* chromosomes was performed, revealing 35 *SiDof* genes distributed unevenly across nine chromosomes (Fig. 1). *SiDof* genes clustered densely on the ends of chromosomes (note Chr8 and Chr9 in particular; Fig. 1). The largest number of genes (nine) was located on Chr9. In contrast, only one gene was found on Chr4 and 6. Chromosomes 1 and 8 each contained four *SiDof* genes, while chromosomes 2, 3, 5, and 7 had 3, 6, 5, and 2 *Dof* genes, respectively.

A similar uneven distribution of *Dof* genes was observed in *Arabidopsis*, with 36 *AtDof* genes in total across all five chromosomes (Yanagisawa, 2002). In rice, *Dof* genes are unevenly distributed among 11 chromosomes, with none on Chr11 (Lijavetzky et al., 2003). In sorghum, *Dof* genes are distributed among nine chromosomes, with none on Chr10 (Kushwaha et al., 2011).

3.3. Gene structure analysis of SiDof genes

Gene structure analysis using GSDS revealed 20 *SiDof* genes lacking introns and 15 with only one intron (Fig. 2). Most introns were located near the 5′ end of the *SiDof* genes.

Similar structural features can be found in other plants. For instance, among the 36 *AtDof* genes from *Arabidopsis*, 18 have no intron and 15 have one (Yanagisawa, 2002). In rice, approximately half of 30 *OsDof* genes have no intron (Lijavetzky et al., 2003). In sorghum, 21 of 28 *SbDof* genes have no intron (Kushwaha et al., 2011). Overall, these data suggest strong similarities in *Dof* gene structures, and possibly in protein regulatory function, between foxtail millet, rice, *Arabidopsis*, and sorghum.

3.4. Conserved domain analysis of SiDof proteins

Fifteen motifs were identified after conserved-domain analysis of 35 SiDof proteins using MEME (Fig. 3 and Table 2). All SiDofs contained motif 1, which represents the conserved, 50-aa Dof domain