Identification, Characterization and Expression Profiles of Dof Transcription Factors in Common Walnut (Juglans Regia L.)

Hanif Khan¹#, Irfan Ullah¹#, Umar Zeb², Sharif Ullah¹, Peng Zhao¹*

¹Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Xi’an, Shaanxi 710069, China
²Department of Biology, The University of Haripur, KPK, Pakistan

Received: 27 February 2020
Accepted: 18 July 2020

Abstract

Juglans regia is a hardwood tree of economic importance and produced high-quality wood and edible nuts. Dof (DNA binding with one finger) are plant-specific transcription factors (TF) family and has an important role in the development of different tissues. However, no study of the Dof gene family in Persian walnut (J. regia) has been reported to our knowledge. We identified 39 JrDof genes in the genome of J. regia. Dof family genes clustered into 06 groups based on the phylogenetic tree and contain a highly conserved motif. The motif composition and conserved domain analysis showed that the common walnut Dof gene family contains one conserved motif and one conserved domain Zf-Dof. Most of the Dof genes contain 1 and 2 exons, as revealed by structural analysis. All JrDof genes distributed unevenly on 16 chromosomes with the maximum member (4 genes) found on chromosome 3, 4, 6, 7, and 12. High expression of JrDof12-1, JrDof15-3, JrDof6-4, JrDof5-2, and JrDof13-2, (female flowers), JrDof10-1, JrDof3-3, and JrDof14-1, (male flowers), JrDof15-3 and JrDof13-2, (embryo), JrDof4-2, JrDof19-1, JrDof15-2, and JrDof12-4, (leaves, fruit and roots) were identified. RNA-Seq data also confirm that Dof genes play a role in the development of J. regia male and female flowers.

Keywords: DNA binding with one finger, Juglans regia, expression profile, transcription factor, Phylogenetic analysis

Introduction

The Dof domain includes a C2C2 zinc finger motif with 50-52 amino acids, which is characterized by the AAAG- element in the promoter [1, 2]. In plants, Dof transcription factor share a highly conserved domain (zf-Dof) [3]. The Dof domain has bifunctional connections of DNA-protein and protein-protein interactions [4]. Dof proteins contain a conserved Zf-Dof domain located at the N-terminal region and a transcriptional regulation domain at the C-terminus [5]. The primary Dof gene (ZmDof1) was first recognized in maize, which is known for its response to light and transcriptional regulation of genes involved in carbon absorption [6, 7]. Dof TFs are not found in human beings, yeast, drosophila, and other eukaryotes.
Transcription factors (TFs) are sequence-specific DNA binding proteins that can hinder or promote gene expression [8]. Thirty-six putative Dof genes have been recognized in the genome of Arabidopsis thaliana [9]. In Arabidopsis thaliana, a total of 9/36 genes were identified that regulated flowering, phytocrome signaling, and seed germination [10, 11]. In Arabidopsis, Dof genes contain DAG2 and DAG1 genes related to seed germination [12]. The genes CDF1, CDF2, and CDF3 controlled the photoperiod of flowering plants [13], while phytocrome signaling and cell cycle regulated by OBPI, and OBP3 [10, 14]. In rice, the gibberellin-regulated expression characterized by OsDof3 [15, 16]. The phosphoenolpyruvate carboxylase gene together with Dof1 in corn has been revealed as an activator of gene expression related to carbohydrate metabolism [17, 18], and maize Dof2 seems like a repressor capable of blocking Dof1 transactivation [19]. The carbon metabolism in wheat is associated with TaDof1 [20]. The JrGRAS2 is a beneficial factor for plant high-temperature tolerance associated with Dof TF in Persian walnut [21]. Dof genes are involved in multiple biological processes such as the regulation of flower development [22], pigmentation and seed proteins [23], nitrogen and carbon assimilation [18], seed germination [4] hormone response [24] and light-mediated regulation [25]. Herbaceous plant species were the main focused of these studies. Contrasting the Dof genes of economically important woody and fruit tree species have received less consideration. This study provided valuable clues for the functional characterization of Dof gene family members in J. regia.

Persian walnut is a diploid (2n = 32), large, wind-pollinated, monoeocious, dichogamous, enduring, perennial tree and vital nut tree on earth belonging to the family Juglandaceae [26, 27]. Juglans are the most important tree because of its wood and nut since ancient times [28-30]. We used public transcriptome data to examine Dof gene expression profiles in different tissues of J. regia. In this study, to understand the potential role and characteristics of Dof gene family, the phylogenetic analysis, chromosomal locations, conserved domain, gene structure, protein structure, and RNA-Seq data for different tissues were investigated. To know the importance of the Dof genes in flowers, we sampled and analyzed the expression level of female and male flowers. This study provides the first genome-wide analysis of the Persian walnut Dof gene family, and these findings will be useful for understanding the putative functions of Persian walnut Dof genes.

Material and Methods

Dof Transcription Factors Identification in J. regia

The Persian walnut whole protein sequence was downloaded from National Center for Biotechnology Information (NCBI). Members of the Dof gene family of Arabidopsis were downloaded from The Arabidopsis Information Resource (TAIR) website [31]. To search against Persian walnut protein sequences, we used Arabidopsis Dof protein sequences as a query using a basic local alignment search tool (BLAST), considering those with an E-value less than 1×10⁻10. We implemented a profile hidden Markov model (HMM) in HMMER v.3.2.1 for window (http://hmmer.org/download.html/) [32] with default parameters to search for Dof proteins and Dof domains in the protein family (Pfam) database (http://pfam.xfam.org/) [33].

Dof Protein Alignment, Phylogenetic Analysis, Pfam Domain Detection, Chromosome Location, and Protein Domains Analysis of Common Walnut

An unrooted phylogenetic tree was constructed by using MEGA software (Molecular evolutionary genetics analysis, Pennsylvania State University: State College, PA, USA) v.7.0 based on neighbor-joining (NJ) method [29]. The phylogenetic NJ tree was constructed with the pairwise deletion of 1000 bootstraps and a Poisson model [34]. To search for the presence of potential domains, the Pfam web server (http://pfam.xfam.org/) [33], was used, and 39 Dof protein sequences were detected. The simple modular architecture research tool (SMART) program (http://smart.embl-heidelberg.de/) [35], also detected the same domains obtained from Pfam (The large collection of protein family's database) with an E-value cutoff of 1.0 to validate the results. We categorized these sequences into various subfamilies, and their distribution on the 16 chromosomes was visualized using MapChart (http://mapchart.software.informer.com/2.2/) with default parameters [36]. A Conserved Domain Database (CDD) search was conducted in NCBI (http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) [37] to confirm the Dof domain of each sequence.

Motif, Gene Structure and Protein Structure Analysis of J. Regia

To analyze the number of exons and introns, we used the coding sequences (CDS) of Dof transcription factors to BLAST against the J. regia genome sequences with a top E-value less than 1×10⁻30. The whole CDS database were downloaded from (https://treegenesdb.org/FTP/Genomes/Jure/v1.1/annotation/). The online gene structure display server was used to illustrate exon and intron structures (http://gsds.cbi.pku.edu.cn/) [38]. The genome browser was used to search for related Persian walnut gene sequences (https://www.ncbi.nlm.nih.gov/ genome/). The motif identification used the MEME program with default parameters, the maximum number of motifs (10), and the optimum motif width (30-50) [37].
Plant Materials, Treatments, and Collections

We collected three male and nine female flowers from common walnut individual trees to evaluate the expression of Dof transcription factors at various developmental stages growing in the Qinling Mountain. The female flowers were collected on 10 April, 15 April, 22 April, and 1 May respectively as 3, 3, 2, and 1 replicate. The male flowers were collected on 10 April, 11 April, and 2 May, respectively. The first opening of female flowers in this study occurred on April 10, April 15, and April 22, and there was a complete opening of female flowers (specifically, the stigma was not fully developed on April 15 and April 22) and 1 May was the end date. The three respective stages for male flowers appear on April 10, April 11, and May 2. After harvesting, the pericarp was immediately dissected, and the flesh was frozen in liquid nitrogen and stored at −80ºC [39]. To isolate total RNA, we used RNA-prep Pre-Plant Kit (Tiangen, Beijing, China) [40]. Using NEBNext Ultra RNA Library Prep Kit (NEB, Beverly, MA, USA), RNA-seq libraries were produced. The pair end sequence was performed with Novogene (Bioinformatics Technology Co., Ltd., Beijing, China) (www.novogene.cn) on the Illumina HiSeq2500 platform to produce 100 bp reads with default parameters. In Trinity, by using default settings, the de novo transcriptome was assembled [41] based on the J. regia genome reference [42]. We investigated RNA-seq data to identify the spatial and temporal expression patterns of the Dof gene family. The transcriptome sequencing datasets were deposited to Bio-Project identifier (ID) PRJNA358784, which was used to conduct RNA-seq of different male and female flowers in J. regia. Also, at the initial flowering stage of germination, we analyzed the total RNA seq data from the male and female flowers. We quantified these gene expression patterns using Cufflinks with default parameters based on their fragments per kilobase of exon per million reads mapped (FPKM) values [43] and using Heml 1.0 software with default parameters to represent these results [44]. Additionally, the ggplot2 R package was used for the Dof transcription factors expression with error bars [45]. Analysis of differential gene expression (DESeq) was carried out using the package DESeq R v.1.1.1. Genes found by DESeq with adjusted P-value/>0.05 were allocated as expressed differentially. We also normalized the number of reads from the RNA-seq information for differential gene expression [44].

Microarray Expression Profiles of Dof Transcription Factors

We downloaded public transcriptional raw data to study the expression pattern in different tissues of common walnut (https://treegenesdb.org/FTP/Genomes/Jure.v1.0/transcriptome/rawreads/) [46]. We used the bowtie software (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) with default parameters to align sequences with long reference sequences [47]. The reference sequences were downloaded from (https://treegenesdb.org/FTP/Genomes/Jure.v1.4/genome/), using Cufflinks software (http://cufflinks.cbch.umd.edu/) [48] with default parameters to assemble transcripts, estimate their abundances, and test for differential expression and regulation in RNA sequencing (RNA-Seq), and represented these results using Heml software (Huazhong University of Science and Technology, Wuhan, China, v.1.0) [41]. A hierarchical map based on normalization data was built and viewed with the Mev software (Multiple Experiment Viewer) (George Washington University, Washington, DC, USA, v.4.9.0) [49]. A heat map is a graphical illustration of data that depicts the individual values of a matrix as colors that enable readers to easily comprehend the information [49].

Results and Discussion

Characteristics of JrDof Gene Family

Gene family research in recent years is important for understanding gene structure, function, and evolution [50]. The Dof genes are plant-specific transcription factors and are widespread in several plant species involved in various biological processes. To date, most work has focused on the functions of Dof genes, and previously 36 AtDof genes identified in Arabidopsis [2, 6], 38 CcDof genes in pigeon pea [51], 36 in cucumber [52], 34 SlDof genes in tomato [53], 78 GmDof genes in soybean [54], 76 BraDof genes in Chinese cabbage [55], 35 StDof genes in potato [56], 42 in Barrel clover [57], pepper with 33 [58], 30 OsDof genes in rice [59], and 20 Dof genes in Chrysanthemum morifolium [1].

Dof Gene Family Identification and Phylogenetic Relationship

We identified a total of 39 Dof genes based on the J. regia whole reference genome (Fig. 1). In this study, the number of JrDof genes identified was slightly higher than that of Arabidopsis, rice, sorghum, and poplar [2, 5]. Furthermore, conserved domain analysis also confirmed that all JrDofs at their N-terminus contain Zf-Dof domain that is the basic characteristics of Dof gene family (Fig. 2) [60]. The neighbor-joining (NJ) phylogenetic tree showed that the JrDof genes are divided into six subfamilies (A to F) in Persian walnut (Fig. 1), and according to previous studies, this classification was performed [1, 34, 55]. The Dof members from Arabidopsis and rice were classified into four groups (a, b, c and d), which is slightly distinct from the present phylogenetic analysis with Dofs of J. regia [61]. In the phylogenetic tree analysis, subfamilies “A” with 12 Dof family members were
the largest clades, and “D” with three Dof family members, were the smallest clades (Fig. 1).

Dof Genes Motifs Identification and Conserved Motifs Analysis

Putative motifs were predicted by the MEME program to reveal the diversification of Dof genes in Persian walnut, and 10 conserved motifs were identified. Graphical view of the motifs identified was presented in Fig. 3(a, b). Motifs have different structural variations that provide more information about its involvement in many biological processes (Fig. 3a, b). Among 10 motifs identified, the widely distributed Motif 1 belongs to the Dof domain, which is involved in protein interaction as well as DNA binding [62, 63]. TFs sometimes contain multiple DNA-binding domains. For example, plant-specific WRKY TFs possess different numbers of WRKY DNA-binding domains, which allows the proteins to be classified into subgroups [64]. However, in the case of Dof proteins, only a single copy of the Dof domain can consistently be found in their N-terminal regions (Fig. 3a, b). In the J. regia Dof gene family, Motif 1 has been identified in all genes, while Motif 10 belongs only to group B members, which indicates that Motif 10 in members of group B is specific to the evolution (Fig. 3a). Motifs 6, 8, 4, 9, 3, 5, and 2 are specific to group F. To understand the Dof genes in Persian walnut, and the structural information provides further insight into these motifs (Fig. 3b). However, high divergence
in their structures was found between the different groups, reflecting the complex nature of the function of Dof proteins in J. regia. In the same group, the majority of JrDof genes shared similar motifs indicating that these conserved motifs are closely related and implying functional similarities between the proteins [65]. In total, 41 conserved motifs were identified in poplar, Arabidopsis, and rice Dof protein sequences [1]. The motif distribution indicated that the genes containing the same motifs were likely produced via gene expansion within the same groups. Conserved motifs were further evaluated to analyze the structural characteristics of Dof genes. The results of our analysis showed that the Dof gene family contained at least one conserved domain (Zf-Dof) and two specific domains (PRP8 and PLN03237) (Fig. 2) [5, 66]. All the JrDofs contained the Zf-Dof conserved domain, the JrDof 2-1 contains the PRP8, and JrDof 10-1 contains PLN03237 specific domain, respectively (Fig. 2). Gene structure analysis and position of conserved motif provide further details about this family evolutionary relationships in J. regia [54].

**Dof Genes Chromosomal Locations and Exon/Intron Structure**

Locations of Dof genes and genetic linkage on the chromosome were identified using MapChart [36]. The JrDof genes were then renamed according to their location on the chromosome (Fig. 4) [53]. Our results
showed that the Dof genes distributed unevenly on 16 chromosomes of Persian walnut, and clustered only on few chromosomes in many plant species (Fig. 4) [51, 55]. The chromosomal location and length of Dof genes are shown in Fig. 4. The highest number of Dof genes (four genes) was found on chromosome 3, 4, 6, 7, and 12, followed by three genes on chromosome 15. In contrast, chromosomes 2, 8, 9, and 10 contain only 1 Dof gene, respectively (Fig. 4). As four Juglans Dof was located on individual chromosomes (Fig. 4), the same phenomenon was observed in barrel clover [57].

To know about the gene structure of the Dof family in the genome of J. regia, we compared diverse exon-intron organizations of the Dof family. To determine the structural heterogeneity of J. regia Dof genes, we analyzed the characterization of the exon-intron structure in the genomic DNA sequences. The number of exons predicted among the Dof genes were comparatively lower, varying from one to three, 19 genes having one exon and no introns, other 19 genes have 2 exons [67] while one member (JrDof 9-1) having three exons (Fig. 5) [56, 68]. The diverse status of exon and intron splicing might be meaningful for the JrDof gene evolution [54]. The Dof genes of the same group also have similar gene structures, such as intron number and exon length. The similar structural features in the Persian walnut genome may be related to their similar functions [68]. The intron number and intron-exon organization of Dof genes in the J. regia genome were quite like Arabidopsis and rice [17], soybean [69].
expression analysis in many plant species has been performed at various developmental stages for different tissues through sequencing that is publicly available. RNA-Seq data for the analysis of gene expression profiles is a convenient tool. We carried out RNA-Seq analysis for various J. regia tissues, including root, somatic embryo, embryo, young leaf, leaves, hull dehiscing, hull cortex, hull peel, immature hull, vegetative buds, and fruit (Fig. 6).

Analysis of the expression profile showed a higher expression of some Dof genes in multiple tissues. For instance, JrDof10-1, JrDof7-1, JrDof12-1, JrDof15-3, JrDof13-2, JrDof4-3, JrDof6-3, JrDof6-4, JrDof14-2, JrDof5-2, and JrDof16-1 were highly expressed in female flowers, while JrDof10-1, JrDof6-4, JrDof12-1, JrDof15-3, JrDof3-3, JrDof5-1, JrDof9-1, JrDof14-1, JrDof3-2, JrDof14-2, JrDof13-2, and JrDof16-1 expressed highly in male flowers (Fig. 6) [1, 62, 72]. Previous studies showed that HbDof12a is the most expressed gene in both female and male flowers, while HbDof08b is the most expressed in roots [68]. In male flowers, on May 2, all Dof genes have a high level of expression than in the other two periods (Fig. 6). In total, 23 Dof genes expressed highly in the female and male flowers (reproductive tissues) (Fig. 6). Importantly, we also noticed 14 Dof genes (JrDof13-1, JrDof4-2, JrDof15-2, JrDof6-3, JrDof5-3, JrDof7-1, JrDof12-1, JrDof7-2, JrDof6-3, JrDof15-2, JrDof12-3, JrDof12-1, JrDof7-3, JrDof6-2, JrDof16-2, JrDof2-1, JrDof13-1, JrDof10-1, JrDof7-1, JrDof4-1, JrDof16-1, JrDof14-2, JrDof5-2, and JrDof14-1) highly expressed in somatic embryo, embryo, vegetative buds, young leaf, leaves, hull cortex, immature hull, hull dehiscing, hull peel, root, and fruit (vegetative tissues) (Fig. 6) [73]. Dof genes that differ in the pattern of expression have also been reported in other species, e.g. Arabidopsis, rice, poplar, and Chinese cabbage [55, 74]. These results are consistent with previous results that Dof genes were ubiquitously expressed in all tissues with possible redundant functions in higher plant tissues [75, 76]. High expression in root tissues predicting its key role in the development of root, these findings are consistent with previous studies in which Dof genes high expression were observed in roots of soybean and other plants [2, 7, 22, 50, 54, 77]. Expression profiles showed that there is no high and specific expression of Dof genes in leaves, suggesting that Dof gene family had little or no role in the development of leaves, which is in contrast with the previous study in which HbDof14b was highly expressed in leaf (Fig. 6) [78, 79]. Dof gene expression in flower tissues was also observed in barrel clover [80]. Previous studies also showed that Dof genes play an important role in the reproductive and vegetative tissues in different plants [19, 27, 52, 58, 77, 78]. Analysis of transcript expression of various Persian walnut tissues (female and male flowers, embryo, leaves, hull, fruit, and leaves) showed a high level of expression in females and male flowers (reproductive tissues), while other vegetative tissues showed a comparatively low degree of expression as compared to reproductive tissues (Fig. 7). The expression of CmDOF20 and CmDOF21

Expression Profiles of Dof Genes in Different J. regia Tissues

It was important to answer the question, whether the JrDof protein structure had any correlation with their functions in different tissues/organs. Based on the sequence similarity of conserved motifs, a total of 10 conserved motifs including the zf-Dof domain were identified (Fig. 2 and 3a) [70, 71]. Gene

and tomato [53]. The number of introns in J. regia Dof genes ranging from zero to two (Fig. 5). The number of introns in soybean and Barrel clover, with only 0 to 1, was reported to be very small [57]. The Dof genes in Cucumber also have 0 to 2 introns that predict structural similarity with other species [52]. Dof proteins found to be highly conserved and comparable to previous Dof gene studies in other species.

![Fig. 5. Intron-Exon structure of 39 Dof genes in J. regia genome. Colored boxes indicate Exons, and gray lines represent Introns; 0, 1, and 2 represent different types of phases. Phase 0: located between two consecutive codons; Phase 1: splitting codon between the first and second nucleotides; Phase 2: between the second and third nucleotides of a codon.](image-url)
was significantly higher in reproductive organs than that in vegetative organs based on the previous report which supports our results [1]. The highly expressed or differentially expressed JrDof genes reported in this study play a regulatory role in J. regia development (Fig. 6 and 7). However, additional research is needed to determine the functions of the JrDof genes.

Conclusions

In Persian walnut (J. regia), we identified a total of 39 Dof genes and clustered into six groups (A-F). Dof transcription factors were further characterized according to the phylogenetic analysis, gene organization, conserved motifs, and Dof domain.

Fig. 6. Expression pattern of multiple tissues of common walnut based on Dof transcription factor. Analysis of expression patterns of all tissues in common walnut using RNA sequencing (RNA-Seq). The heatmap was drawn in log-10 transformed expression values. Red represents relatively high, and the green represents relatively low expression.
Based on the conserved domains, Dof genes contain a conserved domain (Zf-Dof) in all genes. Moreover, the analysis of the expression profile based on RNA-Sequencing showed that the Dof transcription factors reveal diverse patterns of expression in different Persian walnut tissues. Most of the common walnut Dof transcription factors expressed highly in male and female flowers. A total of 16 Dof genes were highly expressed in female and male flowers, while others in reproductive tissues were low in expression or no expression. Also, a total of 14 Dof genes highly expressed in vegetative tissues. In the development of female and male flowers, the important role of Dof genes was observed, that can be further studied and used for the improvement of J. regia development in term of nuts. In conclusion, these results provide a base for studying the potential function of Persian walnut Dof transcription factors.

**Acknowledgements**

This study was funded by the Natural Science Foundation of Shaanxi Province of China (2019JM-008),
Shaanxi Academy of Science Research Funding Project (2019K-06), the National Natural Science Foundation of China (No. 41471038; No. 31200500; No. 31860215), the Program for Excellent Young Academic Backbones funded by Northwest University, Shaanxi Academy of Science Research Funding Project (Y19Z604F12), and Opening Foundation of Key Laboratory of Resource Biology and Biotechnology in Western China (Northwest University), Ministry of Education (ZSK2018009), and the public health specialty in the Department of Traditional Chinese Medicine (2017-66, 2018-43, 2019-68).

Conflicts of Interest
The authors declare no conflict of interest.

References


23. GABRIELE S., RIZZA A., MARTONE J., CIRCELLI P., COSTANTINO P., VITTORIOSO P. The Dof protein...
DAG1 mediates PIL5 activity on seed germination by negatively regulating GA biosynthetic gene ArGA3ox1. Plant J, 61 (2), 312, 2010.


44. WICKHAM H. ggplot2: elegant graphics for data analysis: Springer; 2016.
duplicable and functional characteristics. PLoS ONE, 8 (9), e76809, 2013.