

RESEARCH ARTICLE

Open Access



Genome-wide analysis of wheat DNA-binding with one finger (Dof) transcription factor genes: evolutionary characteristics and diverse abiotic stress responses

Yue Liu^{1†}, Nannan Liu^{1†}, Xiong Deng¹, Dongmiao Liu¹, Mengfei Li¹, Dada Cui¹, Yingkao Hu^{1*} and Yueming Yan^{1,2*} 

Abstract

Background: DNA binding with one finger (Dof) transcription factors play important roles in plant growth and abiotic stress responses. Although genome-wide identification and analysis of the DOF transcription factor family has been reported in other species, no relevant studies have emerged in wheat. The aim of this study was to investigate the evolutionary and functional characteristics associated with plant growth and abiotic stress responses by genome-wide analysis of the wheat Dof transcription factor gene family.

Results: Using the recently released wheat genome database (IWGSC RefSeq v1.0), we identified 96 wheat Dof gene family members, which were phylogenetically clustered into five distinct subfamilies. Gene duplication analysis revealed a broad and heterogeneous distribution of *TaDofs* on the chromosome groups 1 to 7, and obvious tandem duplication genes were present on chromosomes 2 and 3. Members of the same gene subfamily had similar exon-intron structures, while members of different subfamilies had obvious differences. Functional divergence analysis indicated that type-II functional divergence played a major role in the differentiation of the *TaDof* gene family. Positive selection analysis revealed that the Dof gene family experienced different degrees of positive selection pressure during the process of evolution, and five significant positive selection sites (30A, 31 T, 33A, 102G and 104S) were identified. Additionally, nine groups of coevolving amino acid sites, which may play a key role in maintaining the structural and functional stability of Dof proteins, were identified. The results from the RNA-seq data and qRT-PCR analysis revealed that *TaDof* genes exhibited obvious expression preference or specificity in different organs and developmental stages, as well as in diverse abiotic stress responses. Most *TaDof* genes were significantly upregulated by heat, PEG and heavy metal stresses.

Conclusions: The genome-wide analysis and identification of wheat DOF transcription factor family and the discovery of important amino acid sites are expected to provide new insights into the structure, evolution and function of the plant *Dof* gene family.

Keywords: Wheat, Dof transcription factors, Phylogenetics, Evolution, Transcript expression, Abiotic stress

* Correspondence: yingkaohu@cnu.edu.cn; yanym@cnu.edu.cn

[†]Yue Liu and Nannan Liu contributed equally to this work.

¹College of Life Science, Capital Normal University, Xisanhuan Beilu No. 105, 100048 Beijing, People's Republic of China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Transcription factors (TFs) involve in activating or inhibiting the activity of RNA polymerase to regulate the spatiotemporal expression of the target genes by recognizing specific DNA sequence elements present in the promoter region of the gene [1]. DNA binding with one finger (Dof) transcription factors are plant-specific. Dof proteins are generally 200–400 amino acids long with a highly conserved Dof domain of 50–52 amino acids, which is structured as a C2C2-type zinc finger that recognizes a cis-regulatory element with the common core sequence of 5'-AAAG-3' [2–4]. Unlike the conserved N-terminal domain, a transcriptional regulatory domain at the C-terminal of Dof proteins varies greatly, which can react with different regulatory proteins or substances to activate or inhibit gene transcription [3].

The Dof domain is a bifunctional domain that mediates both DNA-protein and protein-protein interactions [5, 6]. The first protein-protein interaction was observed in the *Arabidopsis thaliana* Dof domain protein OBP1, which interacted with bZIP proteins associated with stress responses. OBP1 specifically increased the binding of the OBF proteins to octopine synthase ('ocs') element sequences [7]. The Dof transcription factor prolamins-box binding factor (PBF) in maize can activate the gene expression of cereal storage protein by binding to the P-box present in the prolamins gene (zein) promoter. Meanwhile, PBF can interact with bZIP protein Opaque2 (O2) to activate gamma-zein expression and regulate the protein content of the endosperm in maize [5]. OsDof3 in rice regulates the gibberellin response through interaction with GAMYB [8]. The Dof protein SAD from barley can activate transcription of endosperm-specific genes through interaction with R2R3MYB protein [9]. AtDof3.2 from *A. thaliana* acts as a negative regulator of seed germination and interacts with a positive regulator of seed germination TCP14 [10].

Dof TFs have important functions in plant growth and development, as well as various environmental stress responses. The function of Dof genes in *A. thaliana* has been extensively studied, and several AtDof genes have been shown to function in plant growth and C/N metabolism [11, 12], shoot branching and seed coat formation [13], vascular tissue development and interfascicular cambium formation [14], photoperiodic control of flowering [15–17], morphogenesis and stomata functioning [18], and abiotic stress tolerance [17]. PBF (RPBF) Dof, an activator of seed storage protein genes in rice, participates in the regulation of endosperm-specific gene expression [19]. OsDof12 regulates flowering in long-day condition, and is inhibited by dark treatment [20]. The PBF Dof in maize is involved in seed protein and starch biosynthesis [21, 22]. ZmDof3 plays an important role in maize endosperm development [23]. Maize Dof1 can

activate the *PEPC* gene expression and enhance transcription from the promoters of pyruvate kinase and orthophosphate dikinase [24]. Several tomato Dof genes were found to participate in the control of flowering time and abiotic stress responses [25]. In wheat, PBF homologous genes *TaDof2* (WPBF-A), *TaDof3* (WPBF-D) and *TaDof6* (WPBF-B) were found to locate on the A, B and D genomes of common wheat (*Triticum aestivum* L.), respectively [26, 27]. Wheat PBF could transactivate the transcription of the native alpha-gliadin promoter by binding to the intact prolamins-box [28]. *TaDof1* was found to participate in the process of nitrogen assimilation and control the expression of genes involved in nitrogen assimilation, specifically *GS* and *GOGAT* [29]. A recent report has shown that the *TaDof2*, *TaDof3* and *TaDof6* genes are involved in water-deficit response [30].

As a DNA-binding protein, the first Dof protein (ZmDof1) was identified and characterized in maize [31]. Subsequently, a large number of *Dof* genes have been found in different plant genomes. The number of *Dof* genes identified from genome-based surveys varies depending on the plant species, such as 36 in *A. thaliana* [3], 30 in rice [32], 26 in barley [33], 28 in sorghum [34], 27 in *Brachypodium distachyon* [35], 34 in tomato [36], 46 in maize [37], 36 in cucumber [38], 33 in pepper [39] and 29 in eggplants [40]. Common wheat, as an allohexaploid species, has a huge genome (up to 17 GB) and more than 85% repeat sequences, leading to the slow progress of wheat genome sequencing. Earlier work only identified 31 *Dof* genes in bread wheat [27]. Since 2018, the wheat genome project has made great progress, and the wheat genome data have been updated to IWGSC Annotation v1.0 by the International Wheat Genome Sequencing Consortium (IWGSC) through an improvement of the current wheat chromosome level assembly [41]. The completion of the sequencing of the wheat genome will accelerate the studies on the structure, evolution and function of the wheat *Dof* gene family.

In this study, using common wheat genome database (IWGSC RefSeq v1.0), we conducted a comprehensive genome-wide analysis on the structural characterization, molecular evolution and expression profiling of the wheat *Dof* gene family, which can provide new information for further understanding the evolution characteristics and function of plant *Dof* genes.

Results

Genome-wide identification of wheat *Dof* genes

Firstly, 36 and 30 *Dof* protein sequences from *A. thaliana* and rice were downloaded from the PlantTFDB v4.0 database (Table S1). These sequences were used as queries for searches in the recently released *Triticum aestivum* (common or bread wheat) genome database (IWGSC RefSeq

v1.0). Then SMART and Pfam websites were used to further identify whether candidate sequences have conserved Dof domain. Ultimately, a total of 96 members of the Dof transcription factor gene family were identified from wheat. For convenience, these *TaDof* genes were assigned names from *TaDof1* to *TaDof96* as listed in Table S2.

The number of amino acids of the *TaDof* encoding proteins varied from 152 to 539 amino acids, their pI values ranged from 4.66 to 10.46 with an average of 8.05 and weakly alkaline, and their molecular weights were from 15.77 to 58.13 kDa, with an average of 33.38 kDa. Their detailed information is shown in Table S2. These results indicated that variations in the amino acid sequence length of Dofs may be associated with adaptation to different functional requirements and physical/chemical properties.

Chromosome location and genes duplication of 96 *TaDof* genes

Based on the IWGSC database, the physical locations of the *TaDof* genes on the corresponding chromosomes are depicted in Fig. 1. All *TaDof* genes identified could be mapped on the chromosomes from 1A to 7D. Evidently, the *TaDof* genes were unevenly distributed on different chromosomes, including 34 *TaDof* genes in chromosome A, 32 in chromosome B, and 30 in chromosome D. Most *TaDofs* genes had corresponding homologous genes on the A, B, and D chromosomes. In particular, chromosome 3 with 27 *TaDof* members from *TaDof38* to *TaDof64* had the highest density, and they were closely arranged at the lower part of the chromosomes, but chromosome 7 only contained three *TaDof* genes (*TaDof94*, *TaDof95* and *TaDof96*). Interestingly, we found that the genes located on chromosome 4A were opposite to the position of the homologous genes located on chromosome 4B and chromosome 4D. In addition, segmental duplication and tandem duplication analysis revealed that *TaDofs* transcription factor family was not generated by segmental duplication, but obvious tandem duplication genes were present at the ends on the chromosomes 2 and 3.

Subcellular localization of *TaDof* proteins

The predicted cellular localizations by the five different software programs showed that all 96 *TaDof* proteins were located in the nucleus (Table S2). Then three *TaDof* proteins (*TaDof2*, *TaDof3* and *TaDof6*) were chosen to further perform transient expression to verify the subcellular localization predictions. The results showed that strong green fluorescent signals of the three GFP fusion proteins were observed in the nucleus (Fig. 2), confirming that these *TaDof* proteins were located in the nucleus. These results are consistent with

the transcription factor characteristics and the software predictions.

Phylogenetic relationships and molecular characterization of *TaDof* transcription factors

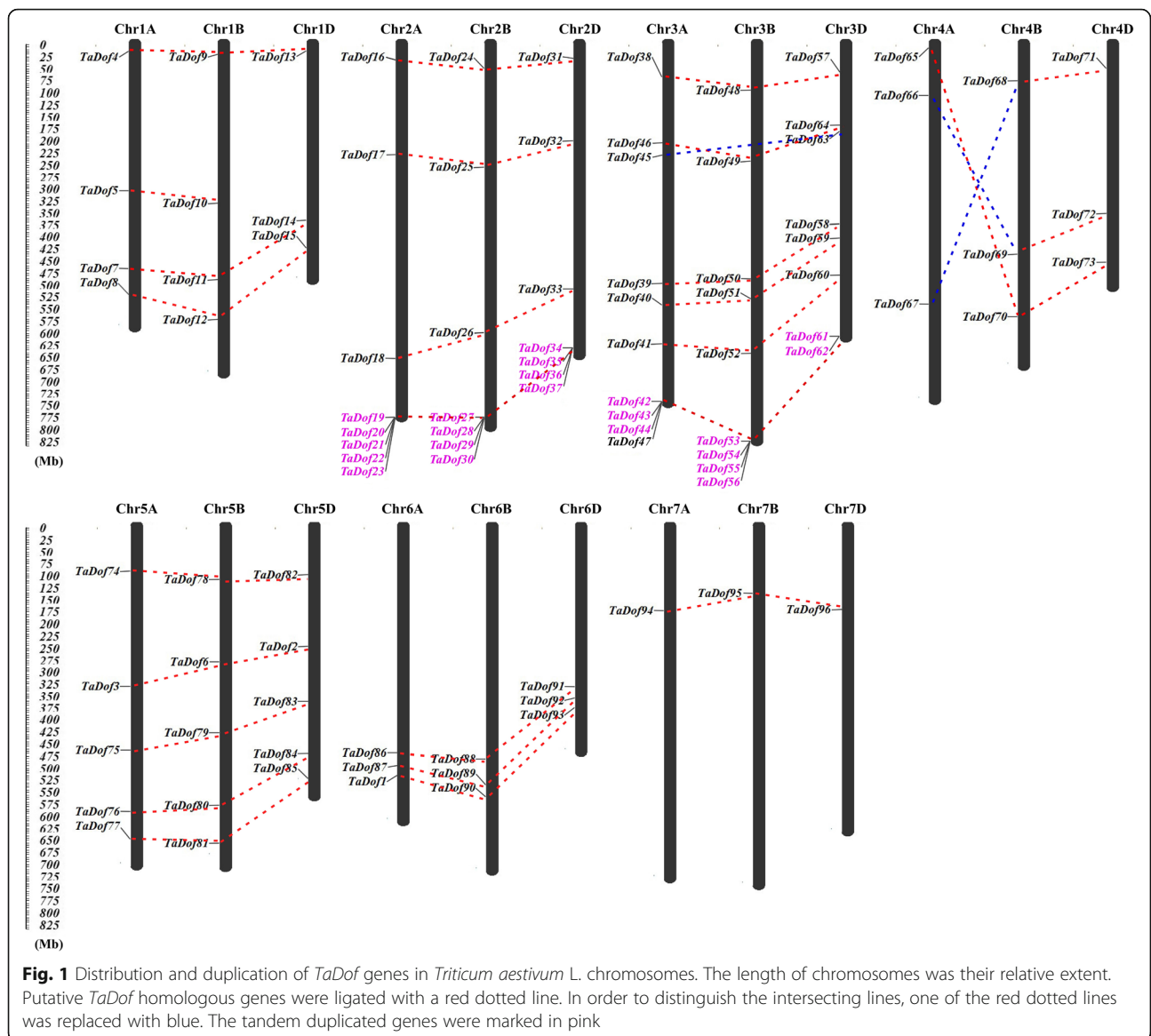
Multiple sequence alignments of the 162 Dof proteins were performed to construct a Bayesian phylogenetic tree (Fig. 3) and a Neighbor-joining (NJ) phylogenetic tree (Fig. S1). The trees revealed that the 96 *TaDof* genes in wheat were classified into five subfamilies (Group A–E) based on later topology and structural similarity analysis, among which Group D was the largest branch with 28 *TaDof* members. Both Group B and C had 20 members, followed by Group A with 15 members and Group E with 13 members. As anticipated, the wheat phylogenetic trees constructed by the NJ method (Fig. 4a), maximum likelihood method and minimal evolution method (Fig. S2) showed a similar topological structure for the five subfamilies.

The exon-intron structures of the 96 *TaDof* gene members were analyzed by comparing the CDSs and the complete gene sequences using the GSDSv2.0, and the results are shown in Fig. 4b, including CDS, intron and UTR structures. The number of introns in the *TaDof* genes was extremely small, with 0–2 introns in each gene. Except for *TaDof43* with two introns, 51 *TaDof* genes (53.13%) had only one intron, and the remaining 46 *TaDof* genes (45.83%) had no intron (Table S2). In addition, the members of the same subfamily generally had similar number of introns. The intron length also varied greatly among different subfamilies, likely resulting from the absence or gain of introns during long-term evolutionary processes.

To further investigate the diversity of *Dof* genes in wheat, the MEME program was used to analyze the potential motif composition in the *Dof* gene family. In total, 15 different motifs were identified (Fig. 4c and Fig. S3). Motif 1, a conserved Dof domain, was uniformly observed in all *TaDof* proteins. Except for individual members, the same subfamily of *Dof* proteins generally shared similar motif number, type and spatial arrangement, implying similar functions of *Dof* proteins in the same subfamily. Group A and B contained fewer motifs, almost exclusively had Motif 1. The specific motif 15 was present in Group C, the conserved motif 2 was present in Group D, and the conserved motifs 4, 6, and 9 occurred in Group E. The remaining motifs 7, 8, 11, 12, 13 and 14 were variable.

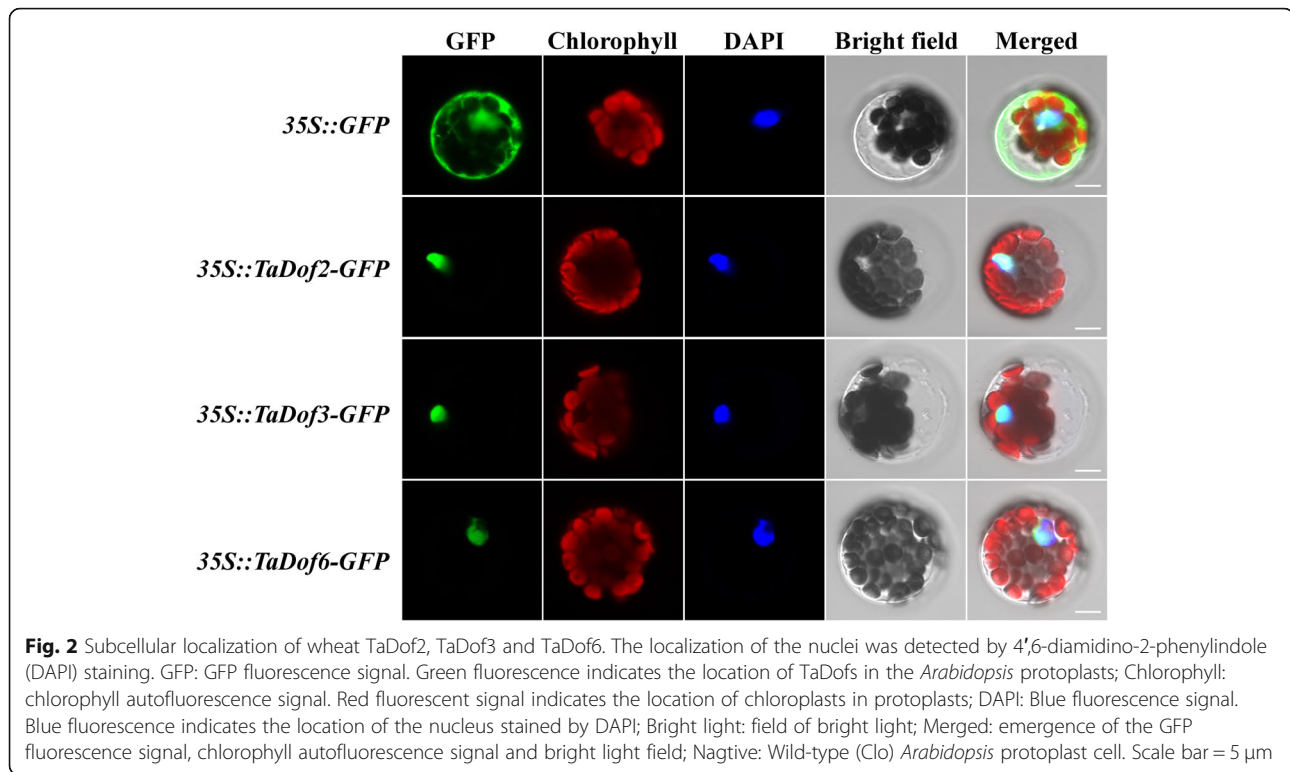
Functional divergence analysis of *TaDof* transcription factors

The DIVERGE v3.0 software combined with the posterior probability analysis method [42–44] was used to estimate the type-I and type-II functional divergences of the



gene group in the Dof family. The results showed that, except for subfamily pairs Group A/Group B, Group A/Group C, Group D/Group C and Group B/Group C, the type-I function divergence coefficient (θ_I) among other groups ranged from 0.177 to 0.418, which is significantly larger than 0. Among them, likelihood ratio test (LRT) values of the subfamily pairs Group A/Group E, Group C/Group E and Group D/Group E were significantly different ($p < 0.05$), indicating that the possible presence of type-I divergence sites during the evolution between groups of wheat Dof proteins. No significant type-I functional divergence was found among other groups. Similarly, the type-II functional divergence coefficient (θ_{II}) ranged from -0.157 to 0.164, indicating that type-II functional divergence sites may also be present (Table 1).

Critical amino acid sites were identified in five groups of *TaDof* subfamilies in the analysis of type-I and type-II functional divergence. In this study, $Q_k > 0.8$ was used as a threshold to screen important amino acid sites, which can reduce the occurrence of false positives. As shown in Table 1, only one type-I functional divergence amino acid site (30A) was detected between Group D and Group E, indicating that the evolutionary rate of this amino acid site might change significantly. Eleven type-II functional divergence sites were found, including 30A, 32A, 33A, 45 K, 47E, 52 K, 55 N, 66 M, 71Y, 75A and 94G. These may be the key amino acid sites affecting physical and chemical properties of *TaDof* proteins. Apparently, the type-II functional divergence sites were significantly more abundant than type-I functional divergence site, indicating that the type-II functional



divergence played a major role in the differentiation of the TaDof gene family. In particular, the amino acid site 30A belonged to both type-I and type-II functional divergence sites, suggesting that the evolutionary rate and physicochemical properties of this site have changed concurrently (Fig. 5a and b).

Positive selection, coevolution and three-dimensional (3D) structure analysis of TaDofs

The CODEML program in the PAML v4.4 software was used for positive selection analysis and positive selection site identification for the TaDof gene family. The site model used in this study included M0 (one scale) and M3 (discrete) as well as M7 (beta) and M8 (beta and ω) based on the previous method [45]. By comparing M0 and M3 models, we found that the twice log-likelihood difference of the models ($2\Delta\ln L$) was 883.03, indicating that certain amino acid sites might be undergone strong positive selection pressure. Comparison between M7 and M8 models can determine whether TaDof gene family members were subjected to positive selection pressure during the evolutionary process. The results revealed that the value of $2\Delta\ln L$ between the two models was 2036.983 with an extremely significant statistical difference. The estimated ω value of the M8 model was 2.55223, which is much higher than 1, indicating that some TaDof amino acid sites were strongly affected by positive selection. In total, 11 positive selectivity amino acid sites were detected in the M8 model, including one

significant (102G, $p < 0.05$) and four extremely significant (30A, 31 T, 33A, 104S, $p < 0.01$) positive selection sites (Table 2).

CAPS, a distance-sensitive coevolutionary analysis software for amino acids, was used to detect the TaDof gene family [46], and nine coevolution sites were detected (Table S3). Among them, 8 groups were adjacent in the primary structure, and most of them were distributed in different locations outside the functional structure domain (Fig. 5a).

The 3D structures of TaDof6 proteins constructed by the online software PHYRE2 showed that five significant and extremely significant sites were located on the 3D structure of TaDof6 protein (Fig. 5b-c), which were mainly located at the N-terminal of the Dof protein. These suggest that the N-terminal of TaDof proteins might encounter more positive selection pressures during the evolutionary process.

Analysis of cis-acting elements in wheat Dof transcription factors

The potential cis-acting elements in the promoter region among 96 TaDof transcription factor members were analyzed by the online tool PlantCARE, which can benefit the understanding of TaDof gene expression and function [47]. In total, seven types of cis-acting elements were found in the promoter region of the TaDof genes, as shown in Table S4.

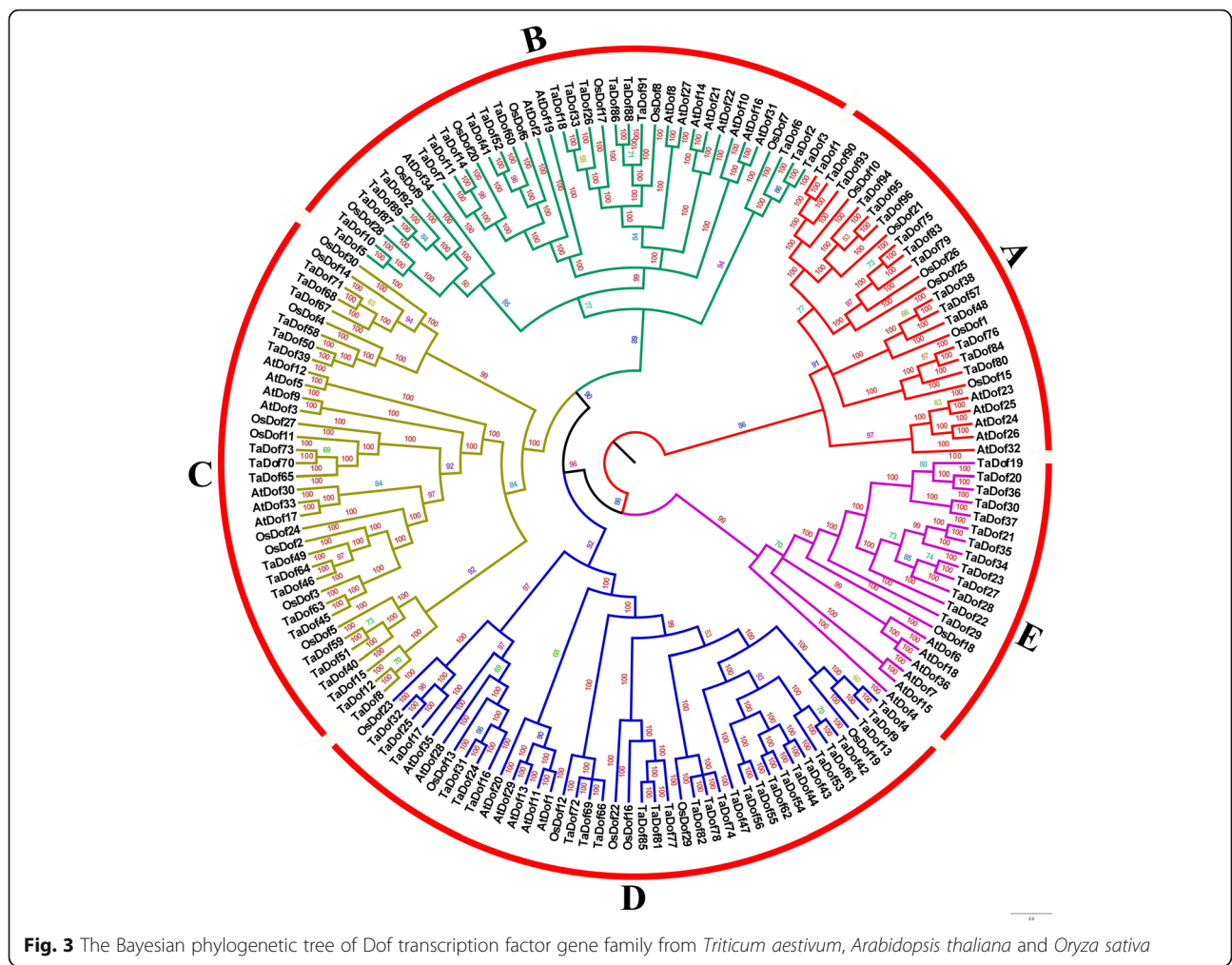


Fig. 3 The Bayesian phylogenetic tree of Dof transcription factor gene family from *Triticum aestivum*, *Arabidopsis thaliana* and *Oryza sativa*

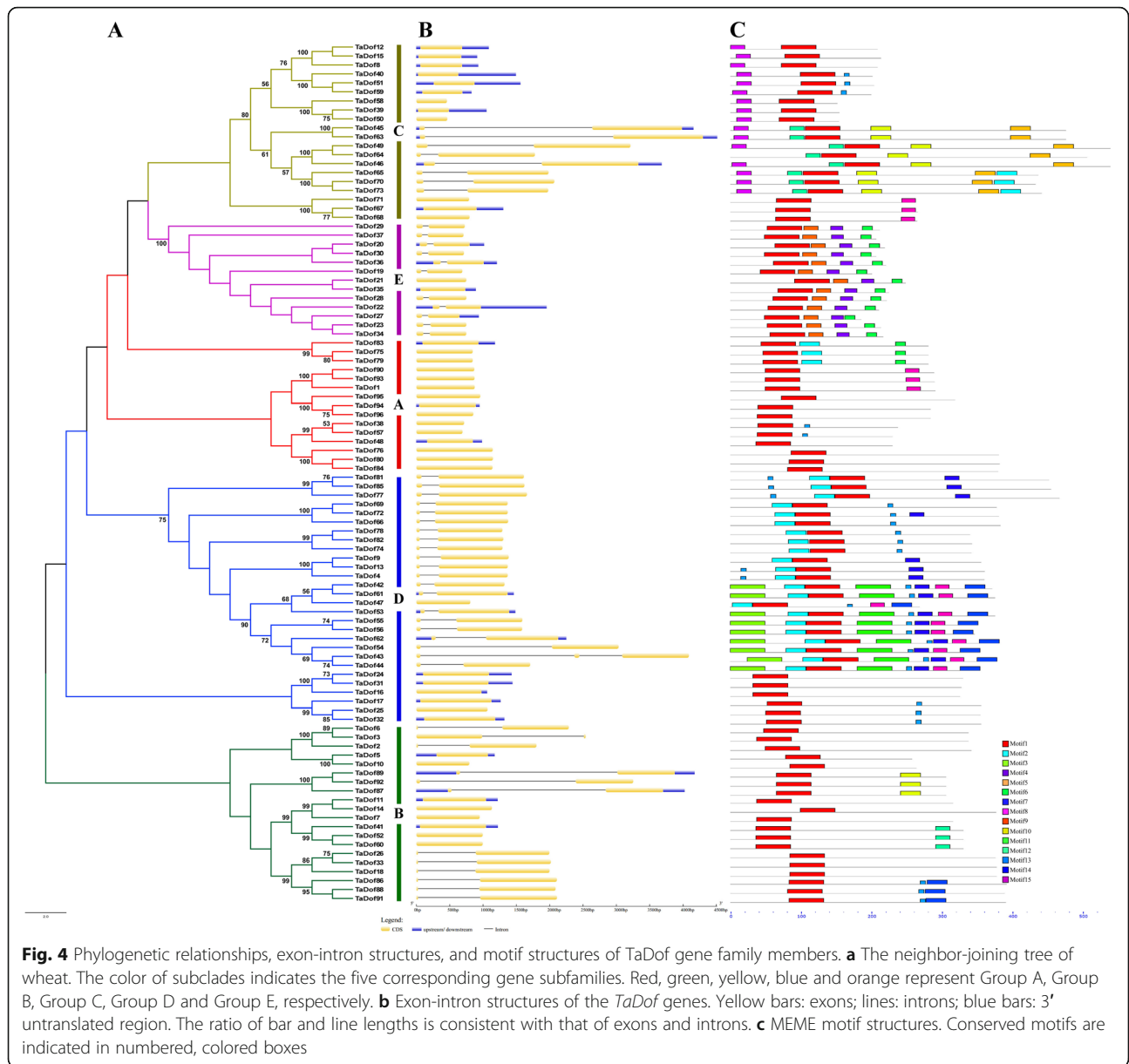
Light responsive elements are a very abundant class of *cis*-acting elements among the *TaDof* gene family members, including G-box, Sp1, and Box 4. G-box seems to be the most abundant type of light responsive elements in the *TaDof* gene family, with a cumulative number of 263. Only 17 members of the *TaDof* gene family did not contain G-box while the remaining members had at least one G-box copy. Hormone responsive elements, mainly including TATC-box, GARE-motif, TCA-element, TGA-element, ABRE, TGACG-motif and CGTCA-motif, participate in response to gibberellin, salicylic acid (SA), auxin, abscisic acid (ABA) and methyl jasmonic acid (MEJA). Among them, ABRE (87.5%), TGACG-motif (78.12%) and CGTCA-motif (79.17%) were present in a large number of members, with an average number of copies of 3.12, 2.67 and 2.66, respectively.

Environmental stress-related elements are also noteworthy. For instance, the GC-motif (68 copies) and ARE (108 copies), which are involved in the regulation of gene expression in the absence of oxygen stress, were found to be relatively abundant. Some *TaDof* genes

harbored MBS (MYB binding site, which is involved in drought-inducibility), indicating that the expression of these *TaDof* genes can be influenced by drought. And *TaDof* gene family members may be involved in defensive damage recovery response and temperature change response due to the presence of WUN-motif, TC-rich repeats, and LTR elements.

Additionally, CCGTCC-box, CAT-box and O₂-site accounted for 33.6, 25.1 and 21.7% of the total number of development related elements, respectively. These elements are involved in the expression and activation of meristematic tissues and the regulation of gliadin metabolism. The promoter-related elements TATA-box and CAAT-box had the largest number of copies per gene, with an average of 13.96 and 15.47, respectively. Except for the promoter of *TaDof86* that had no TATA-box, the other members of the wheat *TaDof* protein family all contained these two types of *cis*-acting elements related to transcriptional regulation.

In addition, we also counted the number of each *cis*-acting element present in the subfamily. Interestingly,



the results showed that the subfamily has a clear preference for the *cis*-acting elements contained in the seven major classes of *cis*-acting elements. For example, for light responsive elements, subfamily A had a relatively large number of A-box element, subfamily D had many Box-4 and TCCC-motif elements. Among the development related elements, subfamily A had the largest number of CCGTCC-box, while subfamily D had 59.3% of the total number of O2-site. For hormone response elements, subfamily A had the most TGACG-motif and CGTCA-motif elements. The number of AREs in the environmental stress-related elements in subfamily D accounted for 51.5% of the total. Particularly, TATA-box and CAAT-box of promoter binding elements were extremely abundant in subfamily D (Table. S5).

Expression of *TaDof* genes in different organs and developmental stages

Analysis of RNA-seq data of different organs at different developmental stages found that *TaDof* genes showed different expression patterns in different organs and developmental stages (Fig. 6a and Table S6). In general, the 96 *TaDof* genes could be divided into five groups with distinct expression patterns (Cluster I-V). The seven genes in Cluster I with exhibited a high expression level in leaf, stem and spike, especially at the early developmental stages. However, the expression of some *TaDof* genes, such as *TaDof43*, *TaDof38*, *TaDof6* and *TaDof54*, was either very low or undetectable in certain developmental stages. Cluster II contained 17 *TaDof*

Table 1 Functional divergence between subfamilies of the *TaDof* gene family

Group1	Group2	Type I			Type II	
		$\theta_I \pm s.e.$	LRT	Sites with $Q_k > 0.8$	$\theta_{II} \pm s.e.$	Sites with $Q_k > 0.8$
A	B	0.007 ± 0.022	0	None	-0.044 ± 0.128	30A
A	C	0.034 ± 0.022	0	None	-0.157 ± 0.157	None
A	D	0.224 ± 0.090	0.596	None	-0.012 ± 0.129	None
A	E	0.418 ± 0.218	0.786	None	-0.007 ± 0.116	52 K, 66 M, 94G, 105D
D	B	0.177 ± 0.022	0	None	-0.048 ± 0.129	71Y
D	C	0.022 ± 0.022	0	None	-0.129 ± 0.163	None
D	E	0.272 ± 0.095	3.989	30A	0.053 ± 0.106	30A , 32A, 33E, 47E 52 K, 55 N, 66 M, 94G
B	C	-0.053 ± 0.022	0	None	-0.094 ± 0.164	71Y
B	E	0.416 ± 0.125	1.895	None	0.037 ± 0.118	32A, 45 K, 52 K, 55 N, 66 M, 71Y, 75A, 94G
C	E	0.310 ± 0.197	1.115	None	-0.082 ± 0.151	47E

Note: θ_I and θ_{II} respectively refer to the coefficients of Type-I and Type- II functional divergence between two groups; LRT, Likelihood Ratio Test; Q_k , posterior probability. All sites were located on the sequence of *TaDof6* according to the results of multiple sequence alignment. The bolted amino acid sites are both Type-I and Type- II functional divergence site

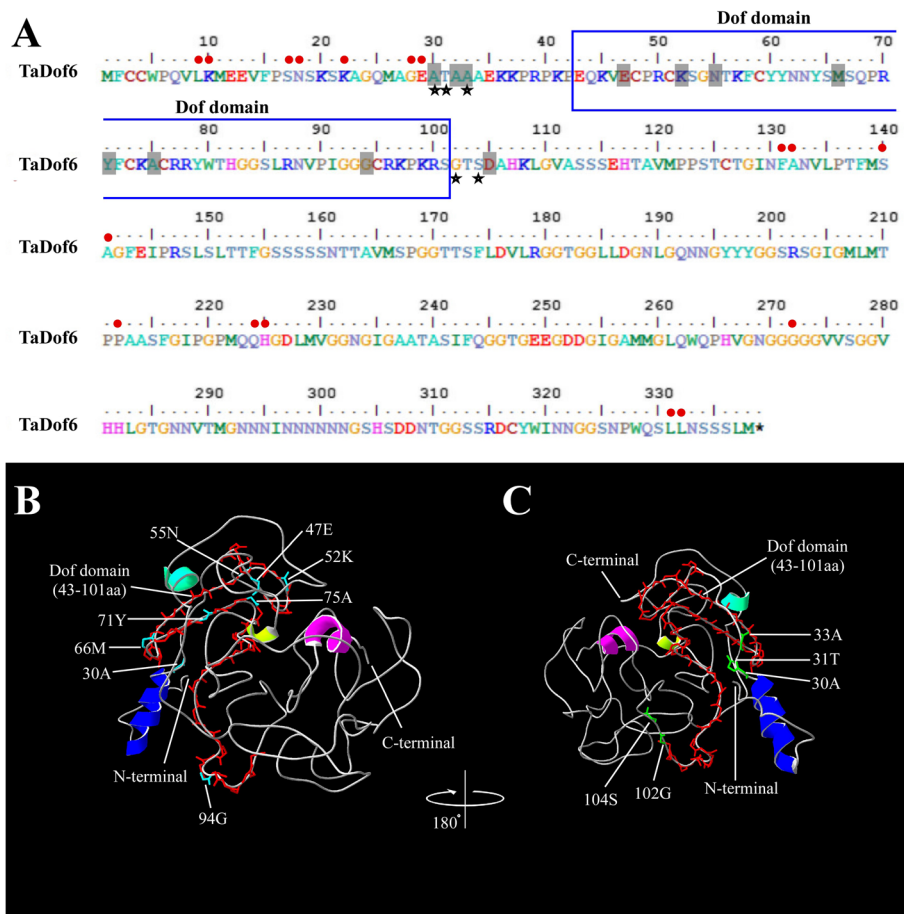


Fig. 5 Protein sequences and model building of the 3D structure of *TaDof6*. **a** Protein sequence of *TaDof6*. The critical amino acid sites of functional divergence, positive selection and coevolution are labeled respectively with gray shadows, black pentagams and red circles, respectively. **b** Schematic diagram of 3D structure of *TaDof6*. The precise positions of seven critical functional divergence sites were colored in light blue. **c** Schematic diagram of 3D structure of *TaDof6* obtained by **(b)** rotating 180 degrees anti-clockwise. Five significant positive selection sites are colored in green

Table 2 Tests for positive selection among *TaDofs* gene family using site-specific models

Models	np	Estimates of parameters ^a	lnL	2ΔlnL	Positively selected sites ^b
M0 (one-ratio)	191	$\omega = 0.05384$	– 4539.929	883.03 (M3 vs M0)	Not allowed
M3 (discrete)	195	$p_0 = 0.43174, p_1 = 0.26150, p_2 = 0.30676$ $\omega_1 = 0.00039,$ $\omega_2 = 0.03024,$ $\omega_3 = 0.23081$	– 4098.415		None
M7 (beta)	192	$p = 0.19976, q = 2.34994$	– 4083.224	2036.983 (M8 vs M7)	Not allowed
M8 (beta & ω)	194	$p_0 = 0.99999, p = 0.82856, q = 1.22843$ ($p_1 = 0.00001$), $\omega = 2.55223$	– 5101.716		30A**, 31 T**, 33A**, 45 K, 67S, 82H, 85S, 102G*, 103 T, 104S**, 105D

Note: *, $p < 0.05$ and **, $p < 0.01$ (χ^2 test). lnL, log likelihood. 2ΔlnL, twice the log-likelihood difference of the models. a, ω was estimated under model M0, M3, M7, M8. b, the number of amino acid sites estimated to have undergone positive selection. np, number of free parameters

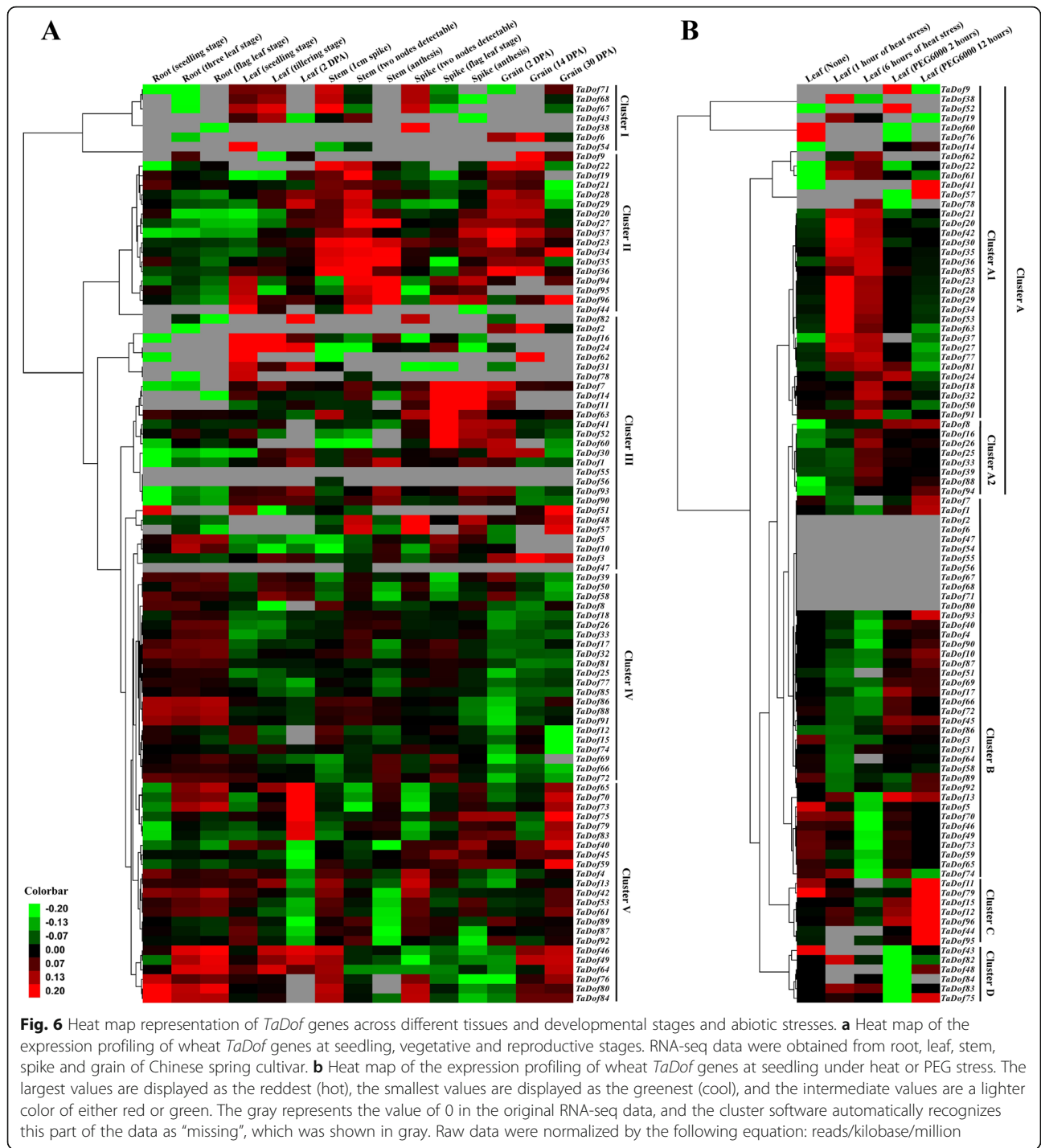
genes that were significantly expressed in all stages of the stem growth, and generally with a low expression in root and high expression in seeds at two and 14 days post anthesis (DPA). In Cluster III, 27 *TaDof* genes were significantly expressed in spikes or grains, but some genes, such as *TaDof55*, were not expressed at different growth stages in five organs. In Cluster IV, 22 genes were preferentially expressed in root and their expression levels were relatively low in both leaves and seeds. Cluster V included 23 *TaDof* genes that were significantly expressed at certain periods in all organs, such as *TaDof40*, *TaDof45* and *TaDof59* which were expressed at the late stages of grain development.

It is worth mentioning that most homologous genes had very similar expression patterns during growth and development (Fig. 6a and Fig. 8). Further analysis found that some *TaDof* genes which clustered in the same subfamily of the phylogenetic tree (Fig. 3) generally also had similar expression patterns. For example, except for *TaDof30*, all other members of Group E which clustered in Cluster II were significantly expressed in both stems and seeds. However, some other *TaDof* members, even the homologous genes with highly conserved amino acid sequences in the same subfamily showed distinct expression patterns. For example, 20 *TaDof* genes in Group C displayed four expression patterns (Cluster I/III/VI/V), of which the expression of two pairs of homologous genes *TaDof45/TaDof63* and *TaDof51/TaDof59* was distinct. Both *TaDof45* and *TaDof59* were clustered in Cluster V and showed higher expression in the spike and grain, while *TaDof63* and *TaDof51* were clustered in Cluster III and had a peak expression in both spike and root. In particular, the homologous genes *TaDof2*, *TaDof3*, and *TaDof6* clustered in a small branch of the phylogenetic tree showed a high expression in the endosperm, but the corresponding expression values detected in the qRT-PCR experiment were particularly low in other tissues (Fig. S4).

Expression profiling of *TaDof* genes in response to various abiotic stresses

The publicly available RNA-seq data of wheat leaves under polyethylene glycol (PEG)-simulated drought and heat stresses were used to show the expression profile of the *TaDof* genes (Table S6 and Fig. 6b). Since 10 *TaDof* genes lacked RNA-Seq atlas data, only 86 genes were analyzed, which were divided into four distinct expression patterns (Cluster A-D in Fig. 6b). Cluster A could be further divided into cluster A1 and cluster A2, and contained 43 genes (50%). Cluster A1 containing 35 genes was significantly upregulated under heat treatment, and cluster A2 containing 8 genes showed obvious upregulation after heat and PEG treatments. Cluster B had 30 genes (34.9%), and more than half of them were upregulated under PEG stress, but all of them were downregulated under heat stress. Cluster C included seven *TaDof* genes whose expression was downregulated under heat stress, but upregulated under PEG treatment, particularly at 12 h. Six *TaDof* genes in Cluster D were generally downregulated at the early stages of seedling growth when exposed to PEG stress, but were upregulated at 12 h after PEG treatment.

To further validate the expression profile of *TaDof* genes in different organs under various abiotic stresses, we selected 17 representative *TaDof* genes from the five groups for qRT-PCR analysis, and their primer sequences are listed in Table S7. In leaves, 11 genes (*TaDof79*, *TaDof94*, *TaDof95*, *TaDof96*, *TaDof45*, *TaDof49*, *TaDof64*, *TaDof16*, *TaDof31*, *TaDof89* and *TaDof29*) responded to almost all stress treatments evaluated. In particular, three *TaDof* genes displayed a significantly upregulated expression in response to multiple abiotic stresses, including *TaDof96* under all stress except for heat stress, *TaDof26* under ABA, PEG, and cold treatments, and *TaDof35* under Cr^{3+} , Cd^{2+} , and heat treatments. All *TaDof* genes except *TaDof96* and *TaDof26* were significantly downregulated under cold



treatment, whereas all genes, except *TaDof16*, *TaDof26* and *TaDof49*, were significantly upregulated under heat treatment. However, both *TaDof16* and *TaDof26* were significantly upregulated under PEG treatment. These results are generally consistent with the RNA-seq data described above (Table S8). According to the results of RNA-seq datas and qRT-PCR, the expression levels of the most homologous genes in response to abiotic stress

were significantly different. For example, *TaDof96* was significantly upregulated under all stress treatments except for heat stress, but the homologous genes (*TaDof94* and *TaDof95*) were downregulated under most stress treatments (Figs. 6b and 7).

In addition, seven *TaDof* genes (*TaDof79*, *TaDof95*, *TaDof96*, *TaDof49*, *TaDof64*, *TaDof16* and *TaDof29*) showed a sensitive response to hormone stress, in which

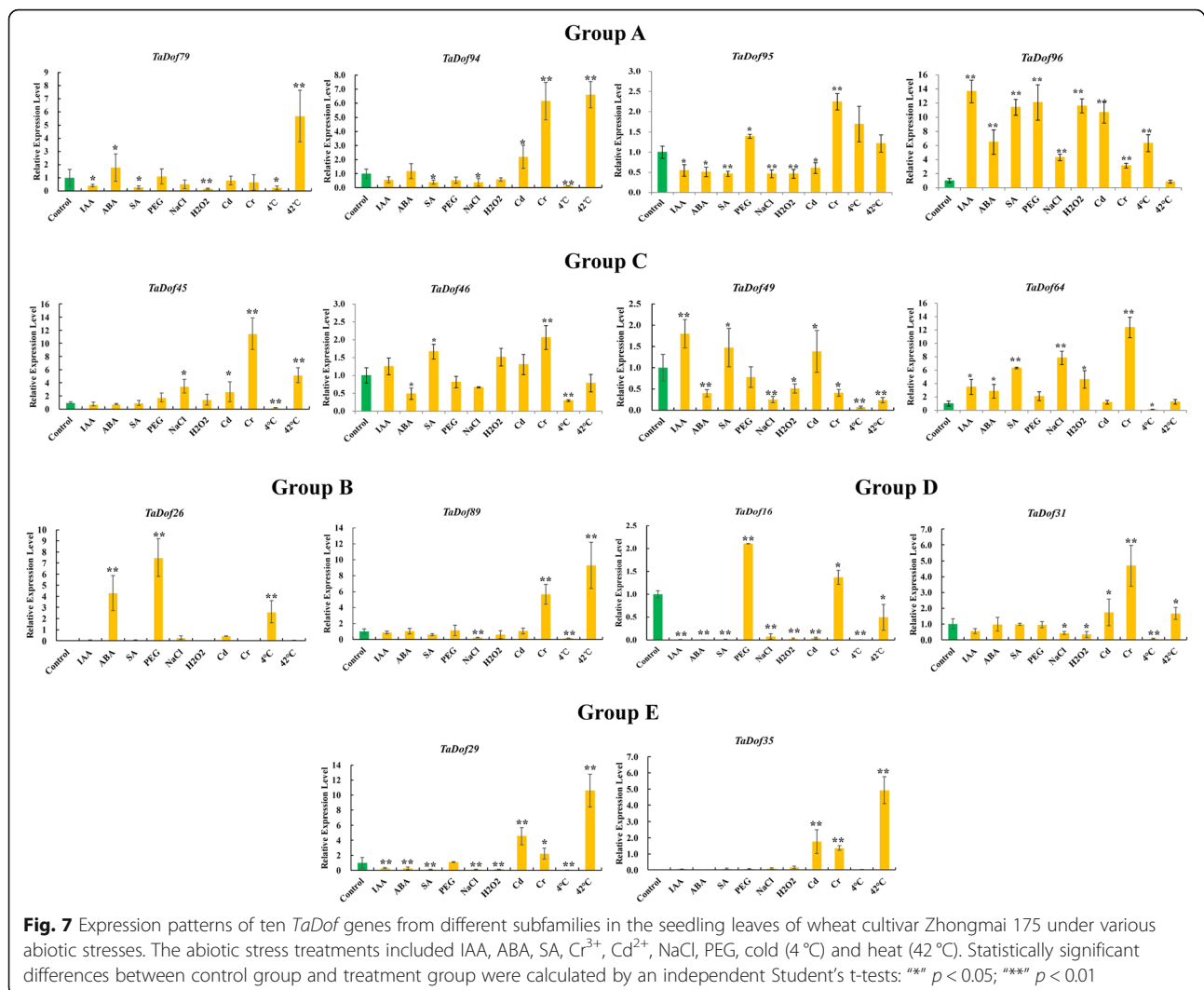


Fig. 7 Expression patterns of ten *TaDof* genes from different subfamilies in the seedling leaves of wheat cultivar Zhongmai 175 under various abiotic stresses. The abiotic stress treatments included IAA, ABA, SA, Cr^{3+} , Cd^{2+} , NaCl, PEG, cold (4 °C) and heat (42 °C). Statistically significant differences between control group and treatment group were calculated by an independent Student's t-tests: "*" $p < 0.05$; "**" $p < 0.01$

TaDof96 and *TaDof64* were significantly upregulated under three hormone stresses, *TaDof49* was significantly upregulated under IAA and SA treatment and the other genes were generally downregulated under hormonal stress. When subjected to heavy metal stress, the expression of most *TaDof* genes was significantly upregulated, especially under Cr^{3+} stress. Additionally, six *TaDof* genes (*TaDof79*, *TaDof95*, *TaDof49*, *TaDof16*, *TaDof31*, and *TaDof29*) were significantly downregulated after oxidative stress. Under salt stress, seven *TaDof* genes (*TaDof94*, *TaDof95*, *TaDof89*, *TaDof16*, *TaDof49*, *TaDof31* and *TaDof29*) were downregulated, whereas three *TaDof* genes (*TaDof96*, *TaDof45*, and *TaDof64*) were significantly upregulated (Fig. 7).

In grains, three highly expressed *TaDof* genes (*TaDof2*, *TaDof3* and *TaDof6*) showed up-and-down expression patterns along with grain development, except for *TaDof6* under nitrogen treatment. Their expression was highly induced and inhibited by high-nitrogen and low-

nitrogen treatment, respectively. The effects of nitrogen stress on the expression of the *TaDof3* and *TaDof6* genes were more obvious during grain development (Fig. 8a). Under drought stress, three *TaDof* genes were significantly upregulated in the early developmental stages of the wheat cultivars Zhongmai 175 and Jimai 22 (Fig. 8b).

Discussion

Molecular characterization and evolution of wheat Dof transcription factors

According to the cluster analysis results of *A. thaliana* and rice [3, 32], the sorghum Dof gene family was divided into six subfamilies (Group A-F), of which Group B had only four *AtDof* members, namely *AtDof4.2*, *AtDof4.3*, *AtDof4.4* and *AtDof4.5* [34]. In this study, we used the recently released wheat genome database to identify 96 *Dof* genes at the genome-wide level (Table S2). The 96 *Dof* genes were classified into five subfamilies (Group

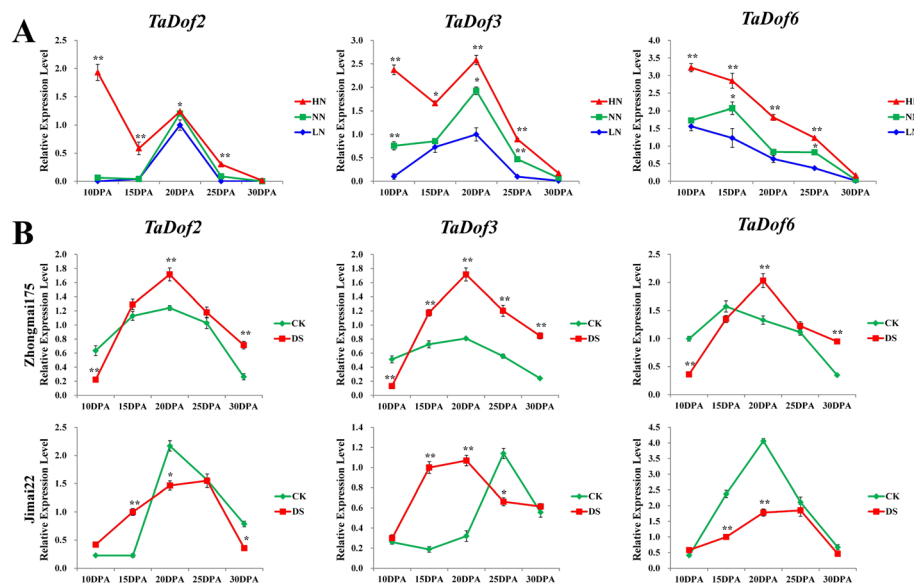


Fig. 8 Expression pattern of wheat *TaDof2/3/6* genes. **a** Expression patterns of *TaDof2/3/6* genes during grain development of wheat cultivar Zhongmai 175 and in response to high and low nitrogen treatments. LN (low-nitrogen): without fertilization application after sowing; NN (normal nitrogen): fertilization with 180 kg/hm²; HN (high-nitrogen): fertilization with 240 kg/hm². **b** Expression patterns of *TaDof2/3/6* genes during grain development of wheat cultivars Zhongmai 175 and Jimai 22 and in response to drought stress. CK: the control group with two-irrigation at jointing and anthesis stages; DS: drought stress group without irrigation after sowing. Expression of the wheat *Ubiquitin* gene was used as a reference gene. Statistically significant differences between control group and treatment group were calculated by an independent Student's t-tests: ***p* < 0.05; ****p* < 0.01

A-E) using the Bayesian method, in which the four *AtDof* members were categorized into Group A (Fig. 3), slightly different from sorghum *Dof* gene classification.

The exon-intron organization can be used as supporting evidence to determine the evolutionary relationships among genes or organisms [48, 49]. In general, the wheat *TaDof* genes in the same subfamily share similar exon-intron structures, but differences are present in different subfamilies. The distribution of motifs among *Dof* proteins is indicative of evolutionary relationship as deduced by phylogenetic tree [50]. The results shown in Fig. 4 demonstrated that the sequence structure of motif 1 identified by MEME was consistent with the *Dof* domain, which may be involved in binding to a particular promoter sequence. Meanwhile, at least one or two conservative motif types and spatial arrangements in *TaDofs* are present in the same subfamily, but obvious differences occur between different subfamilies, implying certain functional similarities of *Dof* members within the same subfamily. In addition, the structural conservation of the *TaDof* genes in subfamilies was consistent with other plants such as *Arabidopsis* and rice [32], barley [33], sorghum [34] and *B. distachyon* [35].

Wheat genomes (BBAADD) consist of three related sub-genomes, and they were derived from three different diploid species, respectively [51]. A total of 96 *TaDofs* transcription factor family members were identified in this study. They were unequally distributed in sub-

genomes A, B, and D, with 34, 32, and 30 members, respectively (Fig. 1). This suggests that there may be a loss of homologous genes during evolution. On the other hand, the retained genes and their distribution on chromosomes provided valuable reference for the polyploidization of wheat chromosomes. It is worth noting that the *TaDof* gene on chromosome 4A was opposite to the distribution of homologous genes on chromosomes 4B and 4D, basically consistent with the recent research [52]. This indicates that the chromosome 4A arm had been reversed during the evolution of hexaploid wheat [53]. We also found that the genes located at the lower part of chromosomes 2 and 3 were closely arranged, and their sequences were highly consistent. They meet the definition of tandem duplication, so they are considered as tandem duplication genes formed during long-term evolution. Interestingly, although the sequences of *TaDof47* and *TaDof42–44* are highly similar, there are too many genes between them to consider as a tandem duplication gene. Moreover, it didn't meet the requirements of segmental duplication gene, so we speculate that it may be caused by the insertion of a short chromosome segmental between *TaDof44* and *TaDof47* during the evolution. The functional diversity of proteins is generally resulted from gene duplication events or the formation of new species [54]. According to our results, type-II functional divergence sites were significantly more abundant than type-I functional divergence sites

(Table 1), demonstrating the important roles of the physical and chemical property changes of amino acids in the diversification among the subfamilies. All seven type-II functional divergence sites (47E, 52 K, 55 N, 66 M, 71Y, 75A and 94G) are located in the Dof domain (Fig. 5), indicating that they may be the main driving force promoting the variation of the Dof domain and the functional divergence of the wheat Dof family. At the molecular level, the amino acid mutations that increase adaptation to the environment were retained through positive selection [55]. Five sites (30A, 31 T, 33A, 102G and 104S) were found to have been subjected to strong positive selection, and may play important roles in functional divergence. Meanwhile, complex coevolutionary networks play important roles in the stability of protein structure and function during evolutionary process [46]. Nine coevolutionary amino acid sites detected in TaDof family members (Table S3) may play roles in maintaining the spatial structure of Dof proteins. These coevolutionary sites may not only perform an important function, but also play a key role in the evolution of the wheat Dof gene family. Therefore, these special amino acid sites might provide useful information for further deciphering the functional attributes of *TaDof* genes.

Potential roles of *TaDof* genes in tissue differentiation and organ development

Phylogenetic and expression analyses can provide important clues about the potential functions of wheat *Dof* genes. As reported in other plants [36, 56, 57], the *TaDof* genes showed a specific and preferential expression in different organs and developmental stages (Fig. 6a), implying their involvement in plant growth and development. Phylogenetic analysis revealed that eight *TaDof* genes (*TaDof45/46/49/63/64/65/70/73*) showed a high sequence similarity to the five *A. thaliana* *Cycling DOF Factors (CDF1–5)* in the Group C subfamily, and they were expressed in the developing grains and stem/spike at the flowering stage (Fig. 3). *CDF1* represses the transcription of *Constans (CO)* and thereby represses flowering in *A. thaliana* [58]. Thus, it is likely that these *TaDof* genes also play roles in spike and seed development.

The homologous *TaDof2*, *TaDof3* and *TaDof6* genes were highly expressed in the early and middle stages of grain development (Fig. 6a). They were closely clustered with *RPBF (OsDof7)* in the phylogenetic tree (Fig. 3), suggesting their functional similarity. The previous study showed that *RPBF* gene, predominantly expressed in the maturing endosperm and coordinately expressed with seed storage protein genes, also participated in the quantitative regulation of genes expressed in the endosperm through cooperating with *RISBZ1* [19]. Wheat WPBF can trans-activate the native alpha-gliadin gene promoter via interaction with the 5'-TGTAAG-3' motif

[28]. As PBF homologues, the proteins encoded by the *TaDof2*, *TaDof3* and *TaDof6* genes activate wheat prolamin gene expression during seed development [26, 27]. Furthermore, the promoter activities of WPBF were observed in the vascular system and the seeds of transgenic *A. thaliana*, suggesting that WPBF functions not only in seed development, but also during other plant growth and developmental stages [28]. We found that some development-related elements present in *TaDof2/3/6* are involved in seed-specific regulation (O2-site, RY-element) and meristem expression (CAT-box) or specific activation (CCGTCC-box) (Table S4), which could play important roles in regulating plant growth and seed development.

It is well established that stems contained abundant vascular tissue. The *TaDof* genes in Group E showed predominantly higher transcript levels in stem (Fig. 6), suggesting they may play a role in vascular tissue development. *A. thaliana* Dof factor OBP1 (OBF-binding factor-1) is implicated in a more general control of cell division, and plays an important role in cell cycle re-entry, acting as a transcriptional regulator of key cell cycle genes [59]. Promoter activity analysis suggested that *AtDof5.8* may function in the primary processes of leaf vasculature formation. *AtDof5.8* promoter sequences also contain *cis*-elements for the stage-specific expression during vascular development [60]. In this study, we detected a large number of the meristem-specific activation-related element CCGTCC-box in the *TaDof* gene promoter regions of Group E, and each gene contained more than 2.3 copies (Table S5), indicating that these *TaDof* genes may also function in various vascular development processes.

Expression and functions of *TaDof* genes in response to abiotic stresses

Dof transcription factors are involved in various abiotic stress responses through the regulation of multiple metabolic pathways. Dof proteins are implicated in responses to plant hormones, such as auxins [61] and gibberellins [8]. When a stressor causes an increase in auxin, the plant activates ABA and other pathways, ultimately promoting the expression of stress-defense genes [62, 63]. The *StDof* genes in potato showed either ABA-dependent or ABA-independent expression pattern [64]. We detected a numerous phytohormone regulation-related elements in the promoter region of the 96 *TaDof* family members, such as the TCA-element, GARE-motif, TGA-element, ABRE and CGTCA-motif (Table S4). qRT-PCR analysis also revealed that the expression of *TaDof* genes can be induced by IAA, ABA and SA (Fig. 7), indicating that these *cis*-elements related to the phytohormone response play important roles in plant hormone pathways.

Reactive oxygen species (ROS) in plants remain at a relatively stable level under normal physiological conditions. However, heavy metal stress will cause an imbalance in the production and clearance of H₂O₂, leading to oxidative damage to plant cells due to excessive accumulation of ROS [65, 66]. We found that eight *TaDof* genes (*TaDof95*, *TaDof96*, *TaDof46*, *TaDof49*, *TaDof64*, *TaDof16*, *TaDof31* and *TaDof29*) were significantly up-regulated under heavy metal stress (Fig. 7). We speculate that *TaDof* genes can act as important regulators in the dynamic regulation of ROS clearance pathways.

The *AtDof* gene *CDF3* is highly induced by drought, extreme environment as well as ABA. Studies showed that *cdf3-1*, a *CDF3* T-DNA insertion mutant, is much more sensitive to drought and low temperature stresses, whereas *CDF3* overexpression promotes the tolerance to drought, cold and osmotic stress [12, 17]. Five homologues of *A. thaliana* CDFs (*SlCDF1-5*) were reported to be differentially induced in response to osmotic, salt, heat, and low-temperature stresses [25]. Most of the *Dof* genes in Chinese cabbage were also quickly upregulated by salt, drought, heat, and cold stress treatments [56]. Furthermore, the *Dof14-15* and *Dof1* genes in wheat were significantly upregulated under drought and salt stresses [27]. In this study, most *TaDof* genes were upregulated under heat stress, and the RNA-Seq data also revealed that the majority of the *TaDof* genes were responsive to heat and PEG stresses (Fig. 6). In particular, the *TaDof16*, *TaDof26* and *TaDof96* genes were upregulated two-, seven- and twelve fold after PEG stress, respectively (Fig. 7). A large number of elements related to environmental stress were also present in the promoter region of *TaDof* genes such as LTR and MBS, and each member had more than 1.2 copies of ARE, GC-motif and WUN-motif (Table S4), which may play an important role in responding to various abiotic stresses.

Introduction of the maize *ZmDof1* gene into rice can increase nitrogen assimilation and enhance plant growth under low-nitrogen conditions [67]. *Dof1* and *GS* work together to regulate the nitrogen metabolism pathway in plants, and to enhance nitrogen assimilation in transgenic tobacco plants grown under low-nitrogen conditions [68]. The *TaDof2*, *TaDof3* and *TaDof6* genes were highly expressed during grain development under drought stress, which improved the synthesis of storage proteins and gluten quality [30]. Consistent with previous studies, in this study, the expression of the *TaDof2*, *TaDof3* and *TaDof6* genes was upregulated under drought and nitrogen stresses (Fig. 8). These results suggest the potential roles of *TaDof* genes in the response to abiotic stresses.

Conclusions

A total of 96 *TaDof* transcription factor genes were identified from the wheat genome database, which were

classified into five subfamilies (Group A-E). The members of the *TaDof* family almost had no introns and all contained a conserved Dof domain. Type-II function divergence was identified as the main reason for the functional diversification of *TaDof* transcription factors. The *Dof* gene family underwent different degrees of positive selection pressure in the evolutionary process. The nine coevolutionary sites were identified, which may play important roles in maintaining the structural and functional stability of *TaDof* proteins. Depending on the type and number of *cis*-acting elements, the *TaDof* genes may be regulated by a variety of hormones and environmental factors. RNA-seq data analysis found that *TaDof* genes showed multiple expression patterns, with obvious expression specificity and preference in different organs and developmental stages, suggesting their potential roles in tissue differentiation and organ development. qRT-PCR analysis further showed that some *TaDof* genes were significantly upregulated in response to single and multiple abiotic stressors, suggesting that they are involved in stress resistance. Our study provides valuable information for further understanding the molecular evolutionary mechanism and functional traits of the plant *Dof* gene family.

Methods

Identification of Dof transcription factors in wheat

The amino acid sequences of 36 and 30 *Dof* transcription factors of *Arabidopsis thaliana* and *Oryza sativa* were firstly downloaded from the plant transcription factor database (PlantTFDB v4.0, <http://planttfdb.cbi.pku.edu.cn/>) [32, 69]. Then the obtained Dofs protein sequences of rice and *Arabidopsis* were used for BLASTP operation (E-value $\leq 1e-5$) to retrieve possible *TaDofs* in wheat from the published common wheat genome database (IWGSC RefSeq v1.0) in Ensemble Plants (<http://plants.ensembl.org/>). The SMART (<http://smart.embl-heidelberg.de/>) [70] and Pfam (<http://pfam.xfam.org/>) [71] websites were used to further verify if the identified proteins had a conserved Dof domain.

Chromosome location and gene duplication analysis of *TaDof* genes

The chromosome location of *TaDof* genes was obtained from the IWGSC database and visualized by the MapInspect software. The duplication gene pairs in the *Dof* gene family were identified based on the criteria of the previous studies [52, 72]. For segmental duplication analysis, we used a pair of non-homologous but similar genes as the center to take 10 genes from the upstream and downstream respectively to compare the similarity of these genes. Based on the similarity and arrangement of the genes on the chromosomes near the two genes, it was determined whether the two genes were segment

duplicate genes. The tandem duplicate analysis of *TaDof* genes was performed according to the method described by Yin et al. [73].

Subcellular localization of TaDof proteins

The prediction and experimental verification of subcellular localization of TaDofs were performed according to the website and experimental method described by Liu et al. [74].

Phylogenetic analysis

Multiple sequence alignment of the amino acid sequences of Dof proteins were performed based on MUSCLE program (<http://www.ebi.ac.uk/Tools/msa/muscle/>) [75, 76]. Subsequently, the Bayesian and N-J phylogenetic trees were constructed based on the method by Han et al. [77]. Furthermore, maximum likelihood and minimal evolution methods were applied for the tree construction to validate the results of the N-J method in MEGA5.0.

Sequence characteristics analysis

The PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) was used to analyze cis-acting elements in the 1500 bp promoter region upstream of the *TaDofs* [47]. The exon-intron structures map of *TaDofs* were detected by comparing the coding sequences (CDS) and genomic sequences in Gene Structure Display Server v2.0 website (GSDS, <http://gsds.cbi.pku.edu.cn/>) [78]. Conserved motifs of the TaDof protein sequences were detected by MEME website (<http://meme-suite.org/tools/meme>), the maximum number of motifs was set to 15 and the remaining operating parameters were unchanged [79, 80]. Theoretical pI/MW of TaDofs was calculated by the Compute pI/MW tool (http://web.expasy.org/compute_pi/).

Molecular evolution analysis of the TaDofs gene family

DIVERGE 3.0 software package was used to detect the presence of type I and type II functional divergence sites in different subfamilies, and the critical value of posterior probability (Qk) was set to 0.8 [43]. The positive selection analysis of the Dof protein family was performed using site model in the PAML4.4 software package [45, 81, 82]. Coevolution sites between amino acids were identified using Coevolution Analysis using Protein Sequences (CAPS) in PERL software [46, 83].

Three-dimensional (3D) structure analysis of TaDof proteins

The 3D structures of TaDof proteins were constructed using PHYRE2 website (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) [84], and visualized by Pymol software (<http://pymol.org/>).

TaDof gene expression analysis by RNA-seq data

The RNA-seq data of the *TaDof* genes were downloaded from the expVIP website (<http://www.wheat-expression.com/>) [85], and cluster analysis was performed by Cluster 3.0 and TreeView software.

Plant material and abiotic stress treatments

The Chinese wheat cultivar Zhongmai 175 seedlings were cultivated into two leaves and one heart stage according to the culture conditions of Han et al. [77]. Then seedlings were treated with the following conditions: heavy metal stress with 300 μ M CrCl₃ and CdCl₂, drought stress with 20% (w/v) PEG 6000, salinity stress with 200 mM NaCl, oxidative stress with 15 mM H₂O₂, hormone stress with 100 μ M ABA (abscisic acid) and SA (salicylic acid), and 10 μ M IAA (indole-3-acetic acid), cold stress under 4 °C and heat stress under 42 °C. The samples from heat stress were collected at 2 h, and other treated seedlings were harvested at 12 h. Three biological replicates were set up for each sample, which were quickly frozen in liquid nitrogen and stored at -80 °C refrigerator.

Meanwhile, two elite Chinese wheat cultivars Zhongmai 175 and Jimai 22 were planted at the experimental station of China Agricultural University, Wuqiao, Hebei Province (116°37'23"E and 37°16'02"N) during the 2016–2018 wheat growing season. Field experiments included two-irrigation at jointing and anthesis stages for control group (CK), no-irrigation after sowing for drought stress treatment (DS), low-nitrogen (LN) without fertilization application after sowing, normal nitrogen (NN) fertilization with 180 kg/hm² and high-nitrogen (HN) fertilization with 240 kg/hm². Each treatment contained three biological replicates and each plot had 20 m². Plants were marked after flowering, and grain samples from five developmental stages of 10, 15, 20, 25 and 30 days post-anthesis (DPA) were harvested. All samples were quickly collected and immediately frozen in liquid nitrogen, and then stored at -80 °C prior to analysis.

qRT-PCR analysis

Total RNA extraction, primer design, qRT-PCR procedures and data analysis were performed according to Cao et al. [86], and three biological replicates were used for each sample. Ubiquitin gene was used as the reference gene. The expression levels of the target genes were determined according to the expression changes relative to the reference gene. At the beginning of the experiment, the relative stability of the reference gene was tested by using various organs and wheat materials treated with different abiotic stress treatments.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-6691-0>.

Additional file 1: Figure S1. The Neighbour-Joining tree of Dof transcription factor gene family from *Triticum aestivum* L., *Arabidopsis thaliana* and *Oryza sativa* L.

Additional file 2: Figure S2. The phylogenetic tree of Dof transcription factor gene family from *Triticum aestivum* L. (A) Minimum Evolution Tree. (B) Maximum Likelihood tree.

Additional file 3: Figure S3. Motifs of TaDof proteins.

Additional file 4: Figure S4. Expression profiling of *TaDof2*, *TaDof3* and *TaDof6* genes in seven wheat tissues and organs.

Additional file 5: Table S1. The nomenclature, characteristics of Dof genes and their deduced proteins in *Arabidopsis thaliana* and *Oryza sativa* subsp. *jaпонica*. The data (accession numbers and amino acid sequences) of Dof transcription factor for *Arabidopsis* and rice are obtained from the Plant Transcription Factor Database (<http://plantfdb.cbi.pku.edu.cn/family.php?sp=Ath&fam=Dof>; <http://plantfdb.cbi.pku.edu.cn/family.php?sp=Os&fam=Dof>). **Table S2.** The nomenclature, characteristics of Dof genes and their deduced proteins in *Triticum aestivum* L. The accession numbers and sequences (genomic, amino acid, and CDS sequences) of the 96 wheat Dof genes were obtained from *Triticum aestivum* genome data (IWGSC) in Ensemble Plants (http://plants.ensembl.org/Triticum_aestivum/Info/Index). **Table S3.** Coevolution sites in TaDofs. **Table S4.** Cis-element analysis of 1500 bp nucleotide sequences data upstream of the translation initiation codon of *TaDof* genes. **Table S5.** The number of each cis-acting element contained in the subfamily. **Table S6.** The RNA-Seq atlas data of the *TaDof* genes.

Table S7. Primers used for qRT-PCR of Dof genes in wheat. **Table S8.** The relative expression values in qRT-PCR and RNA-Seq atlas data of *TaDofs* under abiotic stresses. (A) The relative expression values of *TaDofs* in qRT-PCR under abiotic stresses. (B) The RNA-Seq atlas data of *TaDofs* under abiotic stresses.

Abbreviations

ABA: Auxin, abscisic acid; CDF: Cycling DOF Factors; CDS: Coding sequence; CK: Control group; Dof: DNA binding with one finger; DPA: Days post anthesis; DS: Drought stress treatment; GFP: Green fluorescent protein; GSDS: Gene structure display server; HN: High-nitrogen; IAA: Indole-3-acetic acid; IWGSC: International Wheat Genome Sequencing Consortium; LN: Low-nitrogen; LRT: Likelihood ratio test; MBS: MYB binding site; MCMC: Markov Chain Monte Carlo; MEJA: Methyl jasmonic acid; NJ: Neighbor-joining; NN: Normal nitrogen; OBP: OBF-binding factor; PBF: Prolamin-box binding factor; PEG: Polyethylene glycol; PlantTFDB: Plant Transcription Factor Database; RPPB: Rice PBF; SA: Salicylic acid; 3D: Three-dimensional; TFs: Transcription factors

Acknowledgements

The English in this document has been checked by professional editor, native speakers of English. For a certificate, please see: www.editorbar.com/order/cert/ LE201906130135.

Authors' contributions

YL and NN L carried out the bioinformatic analysis, experimental treatments and qRT-PCR. XD, DM L, MF L and DD C participated in the study and helped to draft the manuscript. YM Y and YK H designed the study and provide guidance on the whole study. All authors read and approved the final manuscript.

Funding

This research was financially supported by grants from National Key R & D Program of China (2016YFD0100502) and the National Natural Science Foundation of China (31771773, 31471485). The funders had no role in the study process including preparation of the manuscript.

Availability of data and materials

The 96 wheat Dof genes data and their accession numbers listed in Table S2 were derived from *Triticum aestivum* genome data (IWGSC) in Ensemble

Plants (http://plants.ensembl.org/Triticum_aestivum/Info/Index), and all accession numbers allow access to public data. The wheat genome sequencing data access URL in Ensemble Plants is ftp://ftp.ensemblgenomes.org/pub/release-46/plants/fasta/triticum_aestivum/dna/. The data (accession numbers and amino acid sequences) of Dof transcription factor for *Arabidopsis* and rice are obtained from the Plant Transcription Factor Database (<http://plantfdb.cbi.pku.edu.cn/family.php?sp=Ath&fam=Dof>; <http://plantfdb.cbi.pku.edu.cn/family.php?sp=Os&fam=Dof>). The RNA-seq data of the 96 wheat Dof genes with specific accession numbers from Ensemble Plants (http://plants.ensembl.org/Triticum_aestivum/Info/Index) are available on the expVIP website (<http://www.wheat-expression.com/>), and the transcriptome data of hexaploid wheat tissues, cultivars and stress conditions access URL in expVIP website is https://opendata.earlham.ac.uk/wheat/under_license/toronto/Ramirez-Gonzalez_etal_2018-06025-Transcriptome-Landscape/expvip/RefSeq_1.1/ByTranscript/. All data generated or analyzed during this study are included in this published article and its supplementary information files.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Life Science, Capital Normal University, Xisanhuan Beilu No. 105, 100048 Beijing, People's Republic of China. ²Hubei Collaborative Innovation Center for Grain Industry (HCICGI), Yangtze University, Jingzhou 434025, China.

Received: 26 July 2019 Accepted: 24 March 2020

Published online: 03 April 2020

References

- Gupta S, Malviya N, Kushwaha H, Nasim J, Bisht NC, Singh VK, Yadav D. Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. *Planta*. 2015;241:549–62.
- Yanagisawa S, Schmidt RJ. Diversity and similarity among recognition sequences of Dof transcription factors. *Plant J*. 1999;17:209–14.
- Yanagisawa S. The Dof family of plant transcription factors. *Trends Plant Sci*. 2002;7:555–60.
- Umemura Y, Ishiduka T, Yamamoto R, Esaka M. The Dof domain, a zinc finger DNA-binding domain conserved only in higher plants, truly functions as a Cys2/Cys2 Zn finger domain. *Plant J*. 2004;37:741–9.
- Vicente-Carbajosa J, Moose SP, Parsons RL, Schmidt RJ. A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator opaque2. *Proc Natl Acad Sci U S A*. 1997;94:7685–90.
- Kang HG, Singh K. Characterization of salicylic acid-responsive *Arabidopsis* Dof domain proteins: overexpression of OBP3 leads to growth defects. *Plant J*. 2000;21:329–39.
- Zhang B, Chen W, Foley RC, Büttner M, Singh KB. Interactions between distinct types of DNA binding proteins enhance binding to ocs element promoter sequences. *Plant Cell*. 1996;7:2241–52.
- Washio K. 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*. 2003;133:850–63.
- Diaz I, Vicente-Carbajosa J, Abraham Z, Martinez M, Moneda II, Carbonero P. The GAMYB protein from barley interacts with the DOF transcription factor BPDF and activates endosperm-specific genes during seed development. *Plant J*. 2002;29:453–64.
- Rueda-Romero P, Barrero-Sicilia C, Gomez-Cadenas A, Carbonero P, Onate-Sanchez L. *Arabidopsis thaliana* DOF6 negatively affects germination in non-after-ripened seeds and interacts with TCP14. *J Exp Bot*. 2012;63:1937–49.
- Kang HG, Foley RC, Onate-Sanchez L, Lin CGT, Singh KB. Target genes for OBP3, a Dof transcription factor, include novel basic helix-loop-helix domain proteins inducible by salicylic acid. *Plant J*. 2003;35:362–72.

12. Renau-Morata B, Molina RV, Carrillo L, Cebolla-Cornejo J, Sánchez-Perales M, Pollmann S, Domínguez-Figueroa J, Corrales AR, Flexas J, Vicente-Carbajosa J, Medina J, Nebauer SG. Ectopic expression of *CDF3* genes in tomato enhances biomass production and yield under salinity stress conditions. *Front Plant Sci.* 2017;8:660.
13. Zou HF, Zhang YQ, Wei W, Chen HW, Song QX, Liu YF, Zhao MY, Wang F, Zhang BC, Lin Q, Zhang WK, Ma B, Zhou YH, Zhang JS, Chen SY. The transcription factor AtDOF4.2 regulates shoot branching and seed coat formation in *Arabidopsis*. *J Biochem.* 2013;449:373–88.
14. Guo Y, Qin G, Gu H, Qu LJ. *Dof5.6/HCA2*, a Dof transcription factor gene, regulates interfascicular cambium formation and vascular tissue development in *Arabidopsis*. *Plant Cell.* 2009;21:3518–34.
15. Imaizumi T, Schultz TF, Harmon FG, Ho LA, Kay SA. FK1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in *Arabidopsis*. *Science.* 2005;309:293–7.
16. Fornara F, Panigrahi KC, Gissot L, Sauerbrunn N, Ruhl M, Jarillo JA, Coupland G. Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Dev Cell.* 2009;17:75–86.
17. Corrales AR, Carrillo L, Lasiera P, Nebauer SG, Domínguez-Figueroa J, Renau-Morata B, Pollmann S, Granell A, Molina RV, Vicente-Carbajosa J, Medina J. Multifaceted role of cycling DOF factor 3 (*CDF3*) in the regulation of flowering time and abiotic stress responses in *Arabidopsis*. *Plant Cell Environ.* 2017;40:748–64.
18. Negi J, Moriwaki K, Konishi M, Yokoyama R, Nakano T, Kusumi K, Hashimoto-Sugimoto M, Schroeder JP, Nishitani K, Yanagisawa S, Iba K. A Dof transcription factor SCAP1 is essential for the development of functional stomata in *Arabidopsis*. *Curr Biol.* 2013;23:479–84.
19. Yamamoto MP, Onodera Y, Touno SM, Takaiwa F. Synergism between RPF Dof and RISBZ1 bZIP activators in the regulation of rice seed expression genes. *Plant Physiol.* 2006;141:1694–707.
20. Li D, Yang C, Li X, Gan Q, Zhao X, Zhu L. Functional characterization of rice OsDof12. *Planta.* 2009;229:1159–69.
21. Wu Y, Messing J. Rapid divergence of prolamin gene promoters of maize after gene amplification and dispersal. *Genetics.* 2012;192:507–19.
22. Zhang Z, Zheng X, Yang J, Messing J, Wu Y. Maize endosperm-specific transcription factors O2 and PBF network the regulation of protein and starch synthesis. *P Natl Acad Sci USA.* 2016;113:10842.
23. Qi X, Li S, Zhu Y, Zhao Q, Zhu D, Yu J. ZmDof3, a maize endosperm-specific Dof protein gene, regulates starch accumulation and aleurone development in maize endosperm. *Plant Mol Biol.* 2017;93:1–14.
24. Yanagisawa S. Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. *Plant J.* 2000;21:281–8.
25. Corrales AR, Nebauer SG, Carrillo L, Fernández-Nohales P, Marqués J, Renau-Morata B, Granell A, Pollmann S, Vicente-Carbajosa J, Molina RV, Medina J. Characterization of tomato Cycling Dof Factors reveals conserved and new functions in the control of flowering time and abiotic stress responses. *J Exp Bot.* 2014;65:995–1012.
26. Ravel C, Nagy IJ, Martre P, Sourdille P, Dardevet M, Balfourier F, Pont C, Giancola S, Prud S, Charmet G. Single nucleotide polymorphism, genetic mapping, and expression of genes coding for the DOF wheat prolamin-box binding factor. *Funct Integr Genomic.* 2006;6:310–21.
27. Shaw LM, McIntyre CL, Gresshoff PM, Xue GP. Members of the Dof transcription factor family in *Triticum aestivum* are associated with light-mediated gene regulation. *Funct Integr Genomic.* 2009;9:485–98.
28. Dong GQ, Ni ZF, Yao YY, Nie XL, Sun QX. Wheat Dof transcription factor WPBF interacts with TaQM and activates transcription of an alpha-gliadin gene during wheat seed development. *Plant Mol Biol.* 2007;63:73–84.
29. Kumar R, Taware R, Gaur VS, Guru SK, Kumar A. Influence of nitrogen on the expression of TaDof1 transcription factor in wheat and its relationship with photosynthetic and ammonium assimilating efficiency. *Mol Biol Rep.* 2009;36:2209–20.
30. Zhou J, Liu D, Deng X, Zhen S, Wang Z, Yan Y. Effects of water deficit on breadmaking quality and storage protein compositions in bread wheat (*Triticum aestivum* L.). *J Sci Food Agric.* 2018;98:4357–68.
31. Yanagisawa S, Izui K. Molecular cloning of two DNA binding proteins of maize that are structurally different but interact with the same sequence motif. *J Biol Chem.* 1993;268:16028–36.
32. Lijavetzky D, Carbonero P, Vicente-Carbajosa J. Genomewide comparative phylogenetic analysis of the rice and *Arabidopsis* Dof gene families. *BMC Evol Biol.* 2003;3:1–11.
33. Moreno-Risueno MA, Martínez M, Vicente-Carbajosa J, Carbonero P. The family of DOF transcription factors: from green unicellular algae to vascular plants. *Mol Gen Genomics.* 2007;277:379–90.
34. Kushwaha H, Gupta S, Singh VK, Rastogi S, Yadav D. Genome wide identification of Dof transcription factor gene family in sorghum and its comparative phylogenetic analysis with rice and *Arabidopsis*. *Mol Biol Rep.* 2011;38:5037–53.
35. Hernando-Amado S, González-Calle V, Carbonero P, Barrero-Sicilia C. The family of DOF transcription factors in *Brachypodium distachyon*: phylogenetic comparison with rice and barley DOFs and expression profiling. *BMC Plant Biol.* 2012;12:202.
36. Cai X, Zhang Y, Zhang C, Zhang T, Hu T, Ye J, Zhang J, Wang T, Li H, Ye Z. Genome-wide analysis of plant-specific Dof transcription factor family in tomato. *J Integr Plant Biol.* 2013;55:552–66.
37. Chen Y, Cao J. Comparative analysis of Dof transcription factor family in maize. *Plant Mol Biol Report.* 2015;33:1245–58.
38. Wen CL, Cheng Q, Zhao L, Mao A, Yang J, Yu S, Weng Y, Xu Y. Identification and characterization of Dof transcription factors in the cucumber genome. *Sci Rep.* 2016;6:23072.
39. Wu Z, Cheng J, Cui J, Xu X, Liang G, Luo X, Chen X, Tang X, Hu K, Qin C. Genome-wide identification and expression profile of Dof transcription factor gene family in pepper (*Capsicum annuum* L.). *front. Plant Sci.* 2016;7:574.
40. Wei Q, Wang W, Hu T, Hu H, Mao W, Zhu Q, Bao C. Genome-wide identification and characterization of Dof transcription factors in eggplant (*Solanum melongena* L.). *PeerJ.* 2018;6:e4481.
41. International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science.* 2018;361:661.
42. Gu X. Statistical methods for testing functional divergence after gene duplication. *Mol Biol Evol.* 1999;16:1664–74.
43. Gaucher EA, Gu X, Miyamoto MM, Benner SA. Predicting functional divergence in protein evolution by site-specific rate shifts. *Trends Biochem Sci.* 2002;27:315–21.
44. Gu X. A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. *Mol Biol Evol.* 2006;23:1937–45.
45. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586–91.
46. Fares MA, McNally D. CAPS: coevolution analysis using protein sequences. *Bioinformatics.* 2006;22: 2821–2.
47. Lescot M. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 2002;30:325–7.
48. Koralewski TE, Krutovsky KV. Evolution of exon-intron structure and alternative splicing. *PLoS One.* 2011;6:e18055.
49. Malviya N, Gupta S, Singh VK, Yadav MK, Bisht NC, Sarangi BK, Yadav D. Genome wide *in silico* characterization of Dof gene families of pigeonpea (*Cajanus cajan* (L) Millsp.). *Mol Biol Rep.* 2015;42:535–52.
50. Gupta S, Kushwaha H, Singh VK, Bisht NC, Sarangi BK, Yadav D. Genome wide in silico characterization of Dof transcription factor gene family of sugarcane and its comparative phylogenetic analysis with *Arabidopsis*, rice and *sorghum*. *Sugar Tech.* 2014;16:372–84.
51. Feldman M, Levy AA. Genome evolution due to allopolyploidization in wheat. *Genetics.* 2012;192:763–74.
52. Zhao Y, Ma R, Xu D, Bi H, Xia Z, Peng H. Genome-wide identification and analysis of the AP2 transcription factor gene family in wheat (*Triticum aestivum* L.). *Frontiers in Plant Sci.* 2019;10:1286.
53. Ma J, Gao S, Stiller J, Jiang QT, Lan XJ, Liu YX, Zheng YL. Identification of genes bordering breakpoints of the pericentric inversions on 2B, 4B, and 5A in bread wheat (*Triticum aestivum* L.). *Genome.* 2015;58:385–90.
54. Gu X. Functional divergence in protein (family) sequence evolution. *Genetica.* 2003;118:133–41.
55. Fetterman CD, Rannala B, Walter MA. Identification and analysis of evolutionary selection pressures acting at the molecular level in five forkhead subfamilies. *BMC Evol Biol.* 2008;8:261.
56. Ma J, Li MY, Wang F, Tang J, Xiong AS. Genome-wide analysis of Dof family transcription factors and their responses to abiotic stresses in Chinese cabbage. *BMC Genomics.* 2015;16:33.
57. Song A, Gao T, Li P, Chen S, Guan Z, Wu D, Xin J, Fan Q, Zhao K, Chen F. Transcriptome-wide identification and expression profiling of the DOF transcription factor gene family in *Chrysanthemum morifolium*. *Front Plant Sci.* 2016;7:199.

58. Lucas-Reina E, Romero-Campero FJ, Romero JM, Valverde F. An evolutionarily conserved DOF-CONSTANS module controls plant photoperiodic signaling. *Plant Physiol.* 2015;168:561–74.
59. Skirycz A, Radziejowski A, Busch W, Hannah MA, Czeszejko J, Kwaśniewski M, Zanol MI, Lohmann JU, De Veylder L, Witt I, Mueller-Roeber B. The DOF transcription factor OBP1 is involved in cell cycle regulation in *Arabidopsis thaliana*. *Plant J.* 2008;56:779–92.
60. Konishi M, Yanagisawa S. Sequential activation of two Dof transcription factor gene promoters during vascular development in *Arabidopsis thaliana*. *Plant Physiol Biochem.* 2007;45:623–9.
61. Baumann K, De Paolis A, Costantino P, Gualberti G. The DNA binding site of the Dof protein NtBBF1 is essential for tissuespecific and auxin-regulated expression of the rolB oncogene in plants. *Plant Cell.* 1999;11:323–34.
62. Rock CD, Sun X. Crosstalk between ABA and auxin signaling pathways in roots of *Arabidopsis thaliana* (L.) Heynh. *Planta.* 2005;222:98–106.
63. Park JE, Park JY, Kim YS, Staswick PE, Jeon J, Yun J, Kim SY, Kim J, Lee YH, Park CM. GH3-mediated auxin homeostasis links growth regulation with stress adaptation response in *Arabidopsis*. *J Biol Chem.* 2007;282:10036–46.
64. Venkatesh J, Park SW. Genome-wide analysis and expression profiling of DNA-binding with one zinc finger (Dof) transcription factor family in potato. *Plant Physiol Biochem.* 2015;94:73–85.
65. Yuan HM, Liu WC, Jin Y, Lu YT. Role of ROS and auxin in plant response to metal-mediated stress. *Plant Signal Behav.* 2013;8:e24671.
66. Shahid M, Pourrut B, Dumat C, Nadeem M, Aslam M, Pinelli E. Heavy-metal-induced reactive oxygen species: phytotoxicity and physicochemical changes in plants. *Rev Environ Contam Toxicol.* 2014;232:1–44.
67. Kurai T, Wakayama M, Abiko T, Yanagisawa S, Aoki N, Ohsugi R. Introduction of the *ZmDof1* gene into rice enhances carbon and nitrogen assimilation under low-nitrogen conditions. *Plant Biotechnol J.* 2011;9:826–37.
68. Wang Y, Fu B, Pan L, Chen L, Fu X, Li K. Overexpression of *Arabidopsis Dof1*, *GS1* and *GS2* enhanced nitrogen assimilation in transgenic tobacco grown under low-nitrogen conditions. *Plant Mol Biol Report.* 2013;31:886–900.
69. Jin JP, Tian F, Yang DC, Meng YQ, Kong L, Luo JC, Gao G. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 2017;45:D1040–5.
70. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 2018;46:D493–6.
71. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2016;44:D279–85.
72. Wang M, Yue H, Feng K, Deng P, Song W, Nie X. Genome-wide identification, phylogeny and expression profiles of mitogen activated protein kinase kinase kinase (MAPKKK) gene family in bread wheat (*Triticum aestivum* L.). *BMC Genomics.* 2016;17:668.
73. Yin G, Xu H, Xiao S, Qin Y, Li Y, Yan Y, Hu Y. The large soybean (*Glycine max*) WRKY TF family expanded by segmental duplication events and subsequent divergent selection among subgroups. *BMC Plant Biol.* 2013;13:148.
74. Liu N, Dong L, Deng X, Liu Y, Liu D, Li M, Hu Y, Yan Y. Genome-wide identification, molecular evolution, and expression analysis of auxin response factor (ARF) gene family in *Brachypodium distachyon* L. *BMC Plant Biol.* 2018;18:336.
75. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 2004;5:113.
76. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.
77. Han Z, Liu Y, Deng X, Liu D, Liu Y, Hu Y, Yan Y. Genome-wide identification and expression analysis of expansin gene family in common wheat (*Triticum aestivum*, L.). *BMC Genomics.* 2019;20:101.
78. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 2015;31:1296–7.
79. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol.* 1994;2:28–36.
80. Moore RC, Purugganan MD. The early stages of duplicate gene evolution. *P Natl Acad Sci USA.* 2003;100:15682–7.
81. Anisimova M, Bielawski JP, Yang Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol.* 2001;18:1585–92.
82. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 2005;22:1107–18.
83. Song W, Qin Y, Zhu Y, Yin G, Wu N, Li Y, Hu Y. Delineation of plant caleosin residues critical for functional divergence, positive selection and coevolution. *BMC Evol Biol.* 2014;14:124.
84. Kelley LA, Sternberg MJ. Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc.* 2009;4:363–471.
85. Kaur S, Dhugga KS, Beech R, Singh J. Genome-wide analysis of the cellulose synthase-like (Csl) gene family in bread wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 2017;17:193.
86. Cao H, He M, Zhu C, Yuan L, Dong L, Bian Y, Zhang W, Yan Y. Distinct metabolic changes between wheat embryo and endosperm during grain development revealed by 2D-DIGE-based integrative proteome analysis. *Proteomics.* 2016;16:1515–36.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

