

RESEARCH ARTICLE

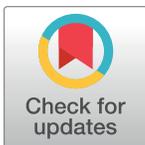
Genome-wide analysis and comparison of the DNA-binding one zinc finger gene family in diploid and tetraploid cotton (*Gossypium*)

YuZhe Li¹ , Zhen Liu¹ , KaiYue Zhang¹, SenYi Chen¹, MengJie Liu¹, QingDe Zhang^{2*}

1 School of Biotechnology and Food Science, Anyang Institute of Technology, Anyang, Henan, China, **2** School of Life Sciences, Henan University, Kaifeng, Henan, China

 These authors contributed equally to this work.

* qdzhang@henu.edu.cn



Abstract

The Dof (DNA-binding one zinc finger) transcription factor family is a representative of plant-specific classes of transcription factors. In this study, we performed a genome-wide screening and characterization of the Dof gene family within two tetraploid species *Gossypium barbadense*, *Gossypium hirsutum*, and two diploid species *Gossypium arboreum*, *Gossypium raimondii*. 115, 116, 55 and 56 Dof genes were identified respectively and all of the genes contain a sequence encoding the Dof DNA-binding domain. Those genes were unevenly distributed across 13/26 chromosomes of the cotton. Genome comparison revealed that segmental duplication may have played crucial roles in the expansion of the cotton Dof gene family, and tandem duplication also played a minor role. Analysis of RNA-Seq data indicated that cotton Dof gene expression levels varied across different tissues and in response to different abiotic stress. Overall, our results could provide valuable information for better understanding the evolution of cotton Dof genes, and lays a foundation for future investigation in cotton.

OPEN ACCESS

Citation: Li Y, Liu Z, Zhang K, Chen S, Liu M, Zhang Q (2020) Genome-wide analysis and comparison of the DNA-binding one zinc finger gene family in diploid and tetraploid cotton (*Gossypium*). PLoS ONE 15(6): e0235317. <https://doi.org/10.1371/journal.pone.0235317>

Editor: Eric Jellen, Brigham Young University, UNITED STATES

Received: March 4, 2020

Accepted: June 9, 2020

Published: June 29, 2020

Copyright: © 2020 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by China Postdoctoral Science Foundation grant number 2018M632761.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Transcription factors play a very vital role in gene regulation at transcriptional level. The Dof (DNA-binding one zinc finger) is a plant-specific transcription factor having multiple roles such as carbon assimilation, light-mediated regulation, seed maturation and germination [1]. Dof specifically bind AAAG sequences of plant gene promoters with the Dof DNA-binding domain [2–4]. In spite of high level homology in the Dof domain, the rest of the sequences are divergent, coinciding with their expected diverse functions [1, 3].

Cotton (*Gossypium*) is one of the most important agronomic genera in the world. Furthermore, cotton is also an excellent model system for studying polyploidization and cell elongation [5–8]. Current understanding recognizes more than 50 species within the cotton genus, with both diploid and polyploid members. Cotton is commonly grouped into eight diploid genomic groups, designated A-G and K, and one tetraploid genomic group, namely AD. All

tetraploid cotton species came from interspecific hybridization between the A-genome species and the D-genome species [9, 10].

In recent years, as more and more plant genome data have been published, genome-wide analysis has become an very effective way for gene function prediction from a large family of genes [11, 12] and there are an increasing number of reports about cotton gene families [13–17]. The Dof gene family has been extensively studied in many plant species, such as *Arabidopsis thaliana*, *Oryza sativa* [18], *Jatropha curcas* [19] and *Setaria italic* [20]. Furthermore, the Dof gene family of *Gossypium hirsutum* was also studied [21, 22]. Because of the importance of Dof gene in various physiological processes, it would be necessary to perform a genome-wide identification and comparative analysis of Dof family in different cotton species. Whole genome sequenes of two cultivated tetraploid species, upland cotton (*Gossypium hirsutum*) and island cotton (*Gossypium barbadense*) [23], and two diploid species *Gossypium arboreum* [24] and *Gossypium raimondii* [25] provided an opportunity to reveal the traits of cotton Dof gene family at genome-wide level. In the present study, we performed a comprehensive analysis of cotton Dof genes, including their gene structure, motif compositions, chromosome distribution, duplication patterns and expression profiles. This study will provide valuable clues for functional characterization of Dof gene family in cotton.

Materials and methods

Identification and characterization of the cotton Dof genes

The *G. hirsutum* [23] and *G. barbadense* [23] genome sequences were downloaded from CottonGen (<https://www.cottongen.org/>), The genome sequences of *G. arboreum* [24] were downloaded from NCBI (BioProject ID: PRJNA382310), and the *G. raimondii* [25] genome sequence was download from <https://cottonfgd.org/>. The candidate genes were searched by BLASTP [26] using a E value of 1e-10 and the known Dof proteins from Arabidopsis were taken as queries. Then the hidden Markova model file (PF02701) was downloaded from the Pfam website (<http://pfam.xfam.org/>) and was used as the query to identify all possible Dof sequences with HMMER software [27]. Furthermore, NCBI CD-Search (<https://www.ncbi.nlm.nih.gov/cdd/>) and Search Pfam tools (<http://pfam.xfam.org/search>) were used to confirm the candidate sequence. The biophysical properties of the Dof proteins were calculated using the ExPASy online server tool (<https://www.expasy.org/>).

Phylogenetic and gene structure analysis of Dof proteins

Previous studies have shown that there are 36 Dof proteins in *Arabidopsis thaliana* [18]. In this study, we included these *Arabidopsis thaliana* Dof proteins in the phylogenetic tree. The ClustalX [28] was used to align Dof protein sequences and MEGA-X [29] program was used to construct a neighbor-joining phylogenetic tree with 1000 bootstrap replicates. Dof gene sequences were loaded into TBtools (<http://www.tbtools.com/>) to obtain exon-intron structure. To identify protein-conserved motifs of cotton Dof, a MEME [30] search was performed, the maximum number of motif was set to 10.

Chromosomal localization, synteny analysis and gene duplication of cotton Dof genes

The chromosome locations of all Dof genes were obtained from the genome annotation files of four cotton species and Mapchart [31] was used to visually map the chromosomal location. Gene duplication events were analyzed using MCScanX [32] and the result data were plotted

by TBtools. Thereafter, the synonymous (Ks) and nonsynonymous (Ka) substitution rates of Dof genes were calculated by KaKs_Calculator 2.0 [33].

Expression profile analysis in various tissues of cotton Dof genes

The original expression data for *G. hirsutum* and *G. barbadense* Dof genes of multiple tissues and under salinity, PEG, cold, heat conditions and normal condition (CK) for 1h, 3h, 6h, 12h and 24h were retrieved from NCBI BioProject database (PRJNA490626). The software Trimmomatic [34] was used to remove the adapters and to perform quality control. The program hisat2 [35] was used to map the reads to the genomes, then the expression profile of Dof genes was obtained with FPKM value using Cufflinks [36], then the results were log transformed and a heatmap was generated by MeV [37].

Results

Genome-wide identification and characterization of Dof gene family in cotton

We used a whole-genome scan to identify genes that encode proteins containing the Dof domain by both BLASTP and HMMER. In the present study, we identified 115, 116, 55 and 56 Dof genes from *G. hirsutum*, *G. barbadense*, *G. arboreum* and *G. raimondii*. The gene number in tetraploid cotton is almost twice that of diploid cotton, and is more than in rice (30 Dof genes) and Arabidopsis (36 Dof genes) [18]. The length of these cotton Dof protein sequences mainly centered on the range of 164~543 amino acid residues. Correspondingly, the molecular weights were mainly distributed from 18318.89 Da to 59589.04 Da. The predicted isoelectric point of Dof proteins varied from 4.77 to 9.92 (S1 Table). The Dof gene family has a wide range of characteristics, this is similar both in cotton and other species [1, 18, 19].

The 342 Dof family members were classified into 3 groups: A, B, C (Fig 1), and the genome/subgenomes of each analyzed cotton have similar member number in the 3 groups (Fig 1, S1 Table).

Gene structure and conserved motifs of the cotton Dof gene family

Our results revealed that the number of exons varied from 1 to 4 in cotton Dof gene family. Most of genes have 1 (43.6%) or 2 (48.8%) exons, and only 1 gene in *G. hirsutum* and 1 gene in *G. raimondii* contains 4 exons (S1 Fig).

Dof protein usually has a DNA-binding domain of approximate 40~60 amino acid residues in the N-terminus. This domain contains a highly-conserved CX2CX21CX2C single zinc-finger structure, which is essential for the zinc finger configuration and loop stability [1, 3, 4]. In this study, all of the cotton Dof protein sequences were loaded into MEME to identify the conserved motifs. The results show that a total of ten conserved motifs were observed. Among them, motif-1 is a common motif in all cotton Dof proteins, corresponding to the CX2CX21CX2C single zinc-finger structure in the Dof domain (S1 and S2 Figs). Some of the Dof proteins only contain motif-1, while others have extra specific motifs, which may be relevant to different functions.

Chromosomal locations and gene duplication events of the cotton Dof gene family

The results show that the 342 cotton Dof genes were widely but unevenly distributed on 13/26 cotton chromosomes (S3 Fig), which is similar to millet [20], banana [34] and Physic Nut [19]. As gene replication plays an important role in the occurrence of novel functions and gene

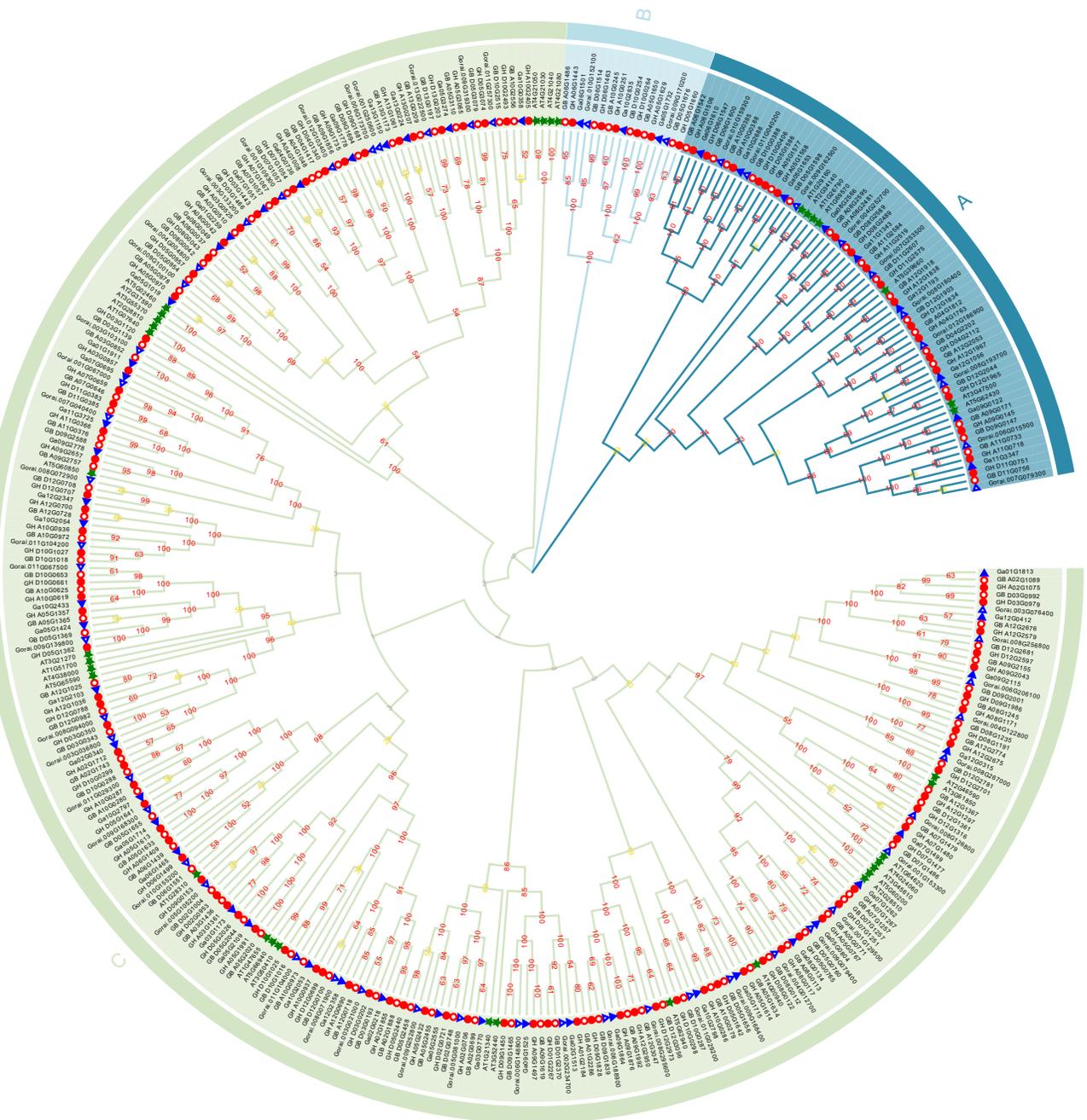


Fig 1. Phylogenetic tree of the Dof gene family. Bootstrapping values are indicated as percentages along the branches. The different background colors indicate different groups. Tetraploid *G. hirsutum* is indicated by red solid circle and tetraploid *G. barbadense* is indicated by red hollow circle; diploid *G. arboreum* is indicated by blue solid triangle and *G. raimondii* is indicated by blue hollow triangle; the *Arabidopsis thaliana* is indicated by green star.

<https://doi.org/10.1371/journal.pone.0235317.g001>

expansion, in this study, we analyzed the duplication events of cotton Dof genes. According to our MCSan analysis, 128, 125 duplication gene-pairs were found between diploid *G. arboreum* A-genome and tetraploid *G. hirsutum*, *G. barbadense* A-subgenome respectively. 137 144

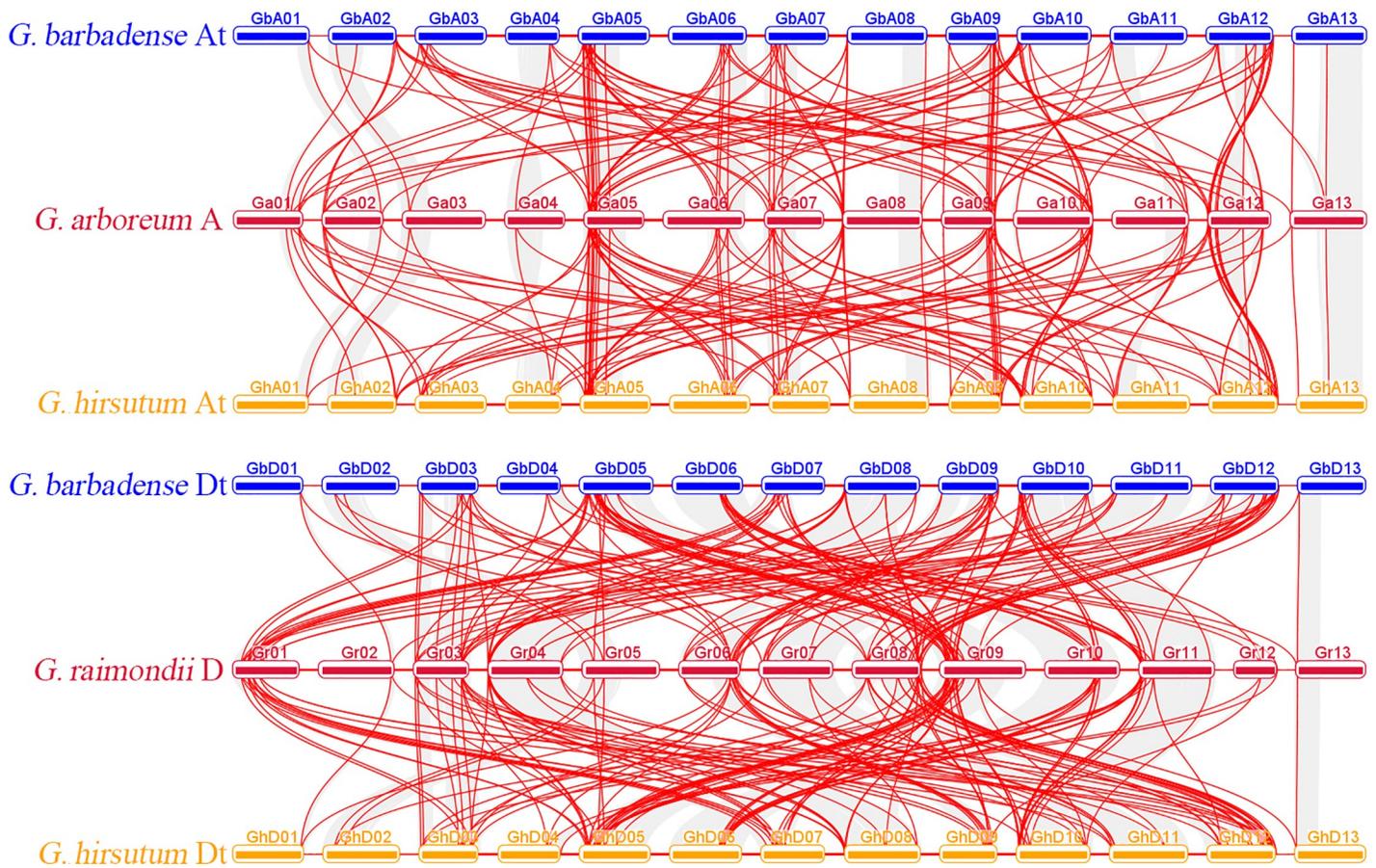


Fig 2. The sub-genome distribution and synteny analysis of cotton Dof genes. The red lines indicate duplicated Dof pairs, the gray lines indicate collinear blocks.

<https://doi.org/10.1371/journal.pone.0235317.g002>

duplication gene-pairs were found between diploid *G. raimondii* D-genome and tetraploid *G. hirsutum*, *G. barbadense* D-subgenome respectively (Fig 2). In addition, we also identified the tandem duplication events. According to Holub, a chromosomal region within 200 kb containing two or more genes is defined as a tandem duplication event [19]. Fourteen Dof genes were clustered into six tandem repeat event regions in both *G. hirsutum* and *G. barbadense*, six Dof genes were clustered into three tandem repeat event regions in *G. arboreum*, and seven Dof genes were clustered into three tandem repeat event regions in *G. raimondii*.

To further infer the phylogenetic mechanisms of cotton Dof family, we constructed a collinear maps associated with all of the four cotton species analyzed (S4 Fig). Some collinear pairs were identified between all of the four cotton species, such as GB_A05G1633/GB_D05G1655/Ga05G1714/GH_A05G1613/GH_D05G1641/Gorai.009G168300, indicating that these orthologous pairs may already exist before the ancestral divergence. In contrast, some collinear gene pairs were not found in one or more of the four cotton species, such as Gorai.003G036800/Ga02G0340/GB_A02G1743/GB_D03G0343/GH_A02G1712/GH_D03G0350, which may indicate that these orthologous pairs formed after the divergence of the four cotton species (S4 Fig).

We further calculated the Ka/Ks ratios for genes pairs between A/At and D/Dt genomes/subgenomes, and the majority of orthologous Dof gene pairs Ka/Ks ratio were between 0.2 and 0.3 (Fig 3), suggesting that the cotton Dof gene family might have experienced purifying selective pressure during evolution.

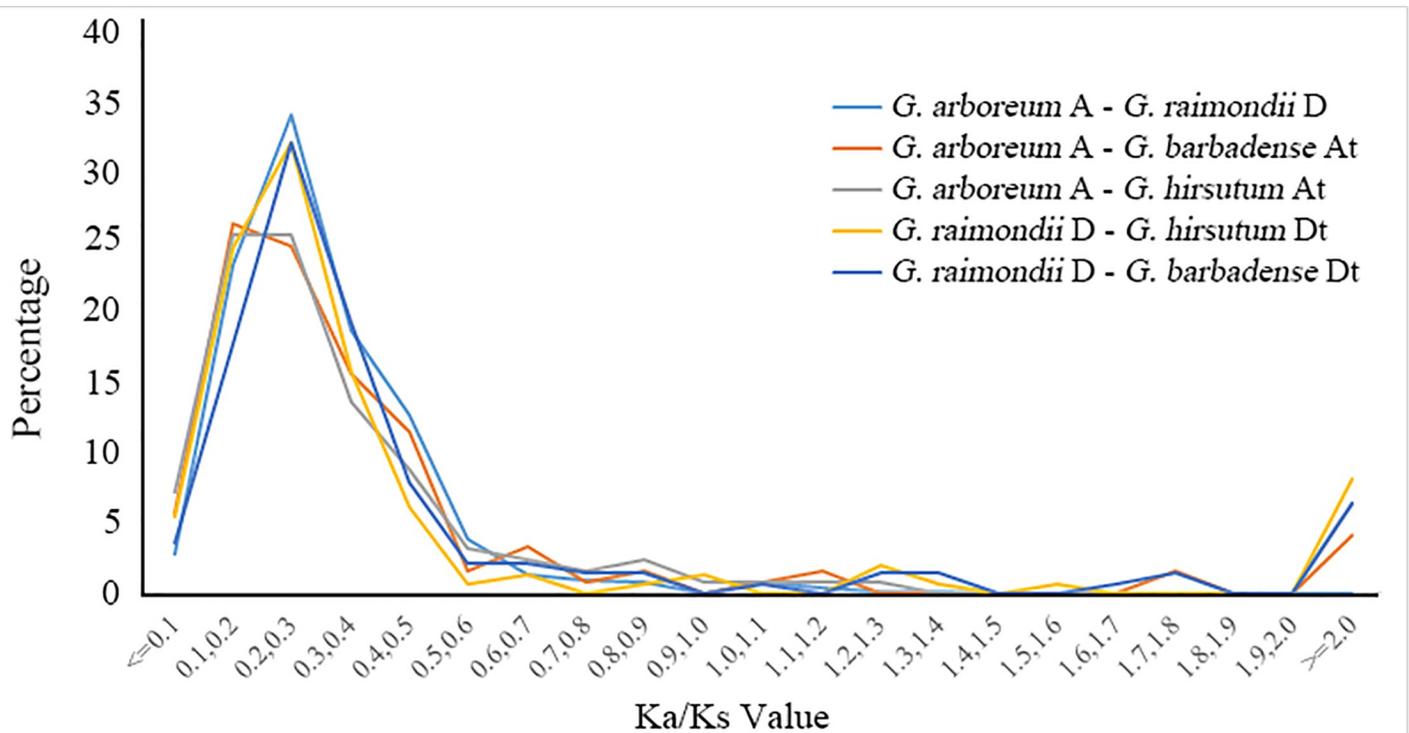


Fig 3. The distribution of Ka/Ks.

<https://doi.org/10.1371/journal.pone.0235317.g003>

Expression pattern of the Dof genes

Transcriptome data was used to explore the variation of cotton Dof genes expression across tissues. RNAseq data for leaf, sepal, root and stem of *G. hirsutum* and torus, sepal, pistil, stem and root of *G. barbadense* were downloaded and analyzed. Our results show that the expression level of most Dof genes vary greatly in different tissues (Fig 4), which is similar to the expression pattern of millet Dof genes [20]. At the same time, few Dof genes have similar expression in different tissues, for example GH_A12G1967, GH_D12G1965 and GH_D05G0765 expressed highly across all tested tissues, while others, such as GH_D12G0788 and GH_A12G1036, expressed far lower in all tested tissues (Fig 4A). In general, these expression patterns indicate that paralogous Dof genes differ considerably in their biological regulatory functions.

To further confirm the potential functions of Dof genes in abiotic stress responses, the expression of *G. hirsutum* and *G. barbadense* Dof genes under salinity, PEG, cold, heat conditions and normal condition for 1h, 3h, 6h, 12h and 24h was analyzed (Fig 5). GH_D03G0202, GH_A02G1855 and GH_A09G1497 of the 115 *G. hirsutum* Dof genes, and GB_D09G1465, GB_A09G1619 and GB_D03G0193 of the 116 *G. barbadense* Dof genes were not expressed in all detected samples. Most Dof genes were significantly induced/repressed by multiple treatments. For instance, GH_A11G0718, GH_D11G0751 of *G. hirsutum* and GB_A11G0733, GB_D11G0756 of *G. barbadense* responded to salinity, PEG, cold and heat treatments significantly. Interestingly, all of these genes were up-regulated by salinity, PEG, cold stress but were down-regulated by heat stress treatment. In addition, GH_D06G1463, GH_A09G2657 of *G. hirsutum* and GB_D06G1514 of *G. barbadense* were repressed by all tested treatments. In contrast, other Dof genes exhibited preferential expression under different conditions. For

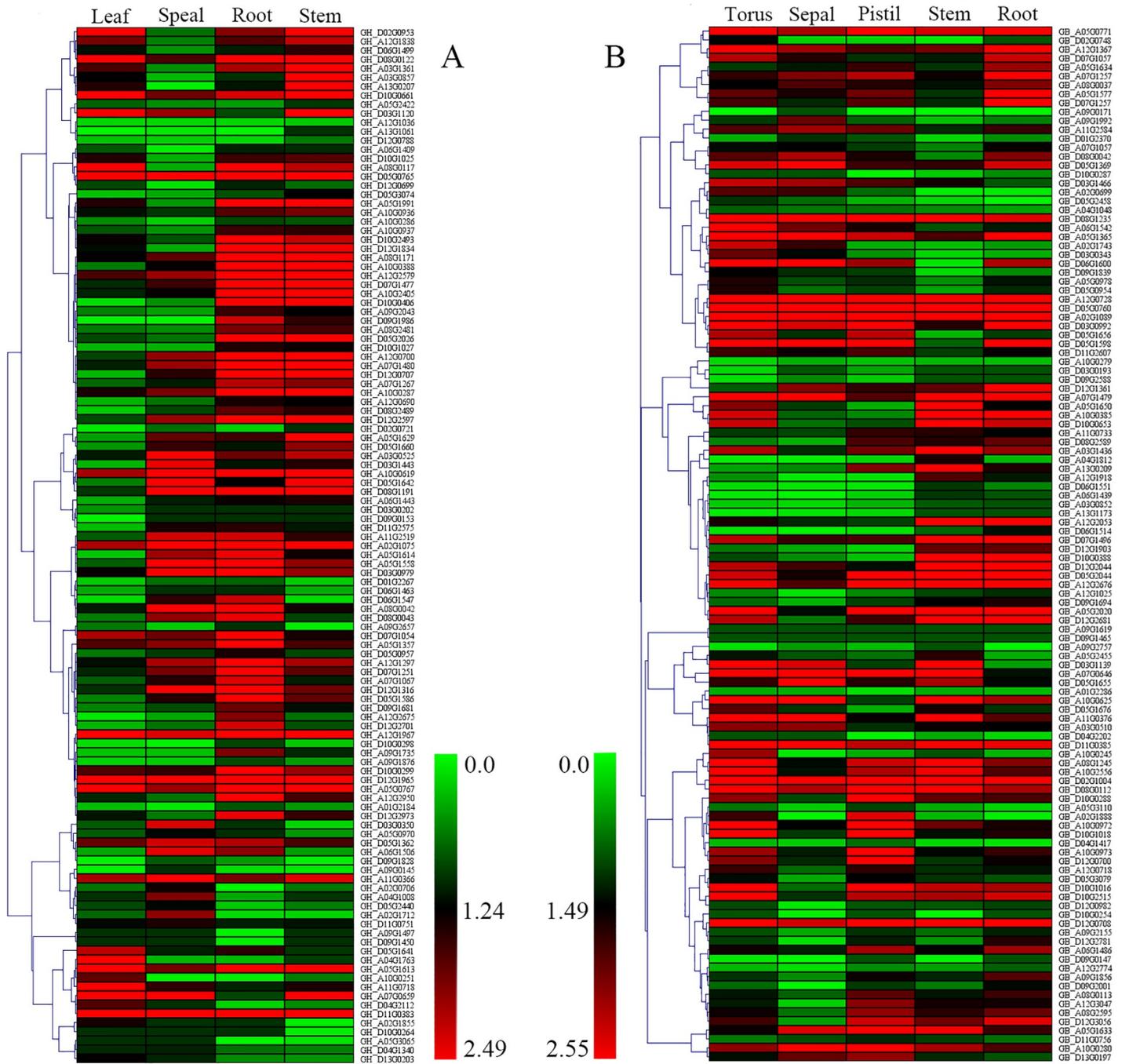


Fig 4. Expression patterns of Dof genes under different tissue. A: *G. hirsutum*, B: *G. barbadense*.

<https://doi.org/10.1371/journal.pone.0235317.g004>

instance, GH_D12G1965 of *G. hirsutum* and GB_A09G2757 of *G. barbadense* were induced significantly by salinity and cold stress but not obviously by heat and PEG stress. Overall, these results demonstrated that the cotton Dof gene family displayed different expression patterns under diverse environmental stress conditions, suggesting that these genes were responsive to stress treatments.

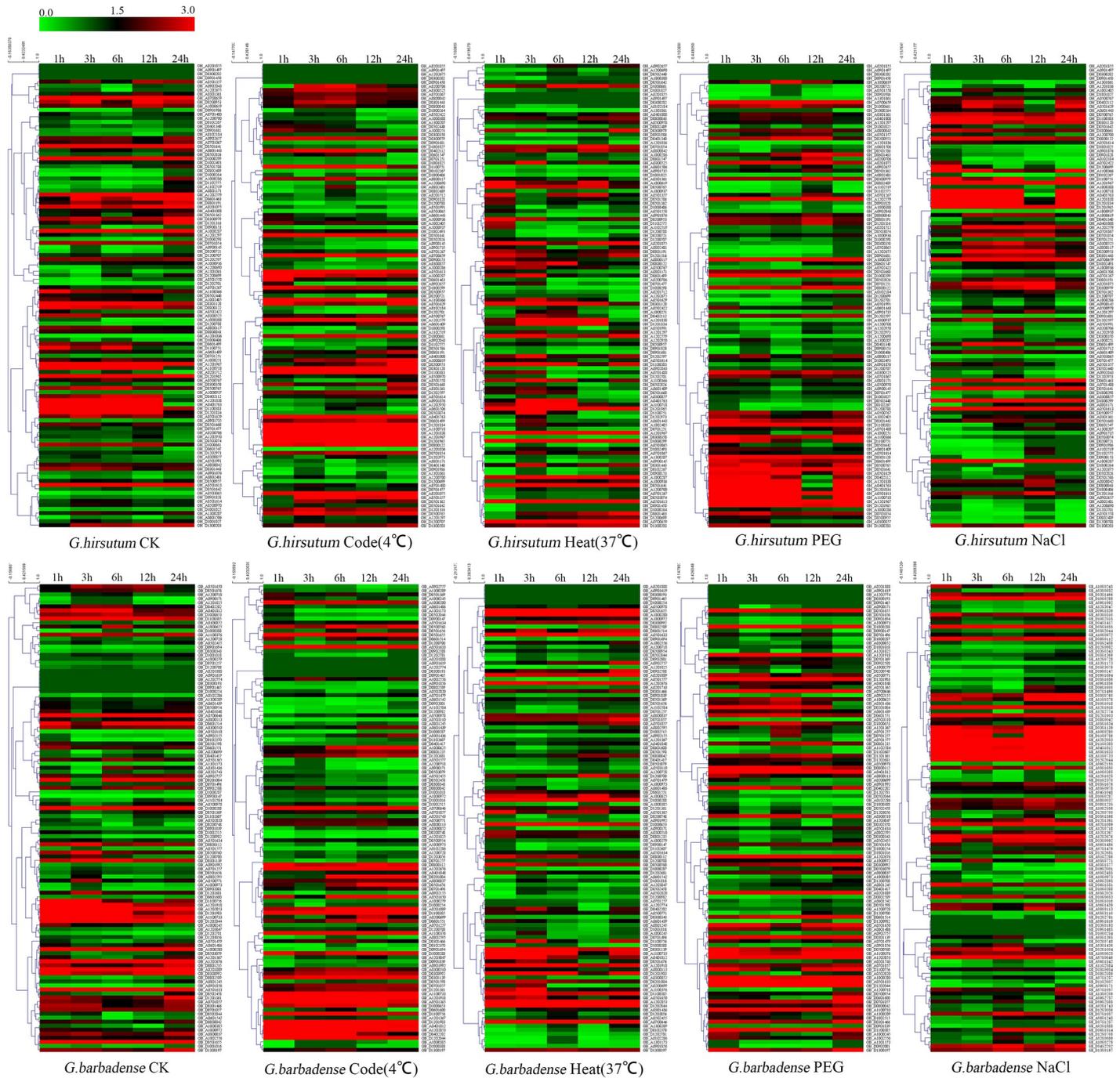


Fig 5. Expression patterns of Dof genes under normal condition (CK), cold, heat, PEG and salinity conditions for 1h, 3h, 6h, 12h and 24h.

<https://doi.org/10.1371/journal.pone.0235317.g005>

Discussion

Transcription factors are an important group targeted for crop improvement and a lot of efforts have been made to reveal the whole set of transcription factors [35–37]. In the present study, we completed genome-wide analysis of cotton Dof genes by bioinformatic analysis and 115 and 116 Dof genes were identified from tetraploid *G. barbadense* and *G. hirsutum*, 55 and

56 Dof genes were identified from diploid *G. arboreum* and *G. raimondii*. The number of genes in tetraploid cotton is almost twice that of diploid cotton, but the duplication gene-pairs in tetraploid cotton is significantly more than that in diploid cotton. The Dof gene density decreased from 0.3/Mb to 0.05/Mb in Arabidopsis and cottons. The reason for this discrepancy might be the variable status of genome duplications in Arabidopsis and cottons [18, 34, 38]. The cotton Dof family members were classified into 3 groups, the same group shared more similar gene structures which suggest evolutionary conservation in cotton Dof gene evolution. The gene expansion of the Dof family in cotton mainly resulted from segmental duplication, and tandem duplication also played a minor role. The Dof duplicated gene pairs tended to be subjected to positive selection, which may play important roles in the adaptive phenotypes of cotton. In this study, the gene expression of Dof gene family were identified in salinity, PEG, cold, heat conditions and normal condition stresses. The expression profile demonstrated the broad involvement of cotton Dof genes in different abiotic stressed treatments. In addition, cotton Dof gene expression has tissue-specific characteristics.

Because *G. hirsutum* is the main source of textile fiber, the study of cotton Dof gene family was focused on *G. hirsutum* in previous studies [21, 22]. Studies have shown that *G. hirsutum* Dof gene family constitutively expressed in leaves, roots and stems, accumulated highest in leaves. The salinity and cold treatments induced *G. hirsutum* Dof transcript accumulation, and the overexpressed of Dof showed significantly higher salt and cold tolerance over the wild-type plants [21]. Moreover, genome-wide study shown that there were 114 Dof genes in *G. hirsutum*, the phylogeny, duplication, and chromosomal locations of *G. hirsutum* Dof gene family in previous studies are similar to ours [22]. In this study, we performed a genome-wide analysis and comparison of the Dof gene family within two tetraploid cotton species and two diploid cotton species. Gene structure, conserved motifs and Ka/Ks distribution of Dof gene family in the four cotton species were analyzed for the first time. In addition, Dof gene expression was analyzed by RNA-Seq data in our study which is different with RT-PCR in previous studies. Therefore, our study will further broaden our insights into the evolution and functional elucidation of Dof gene family in cotton.

Conclusions

A genome-wide bioinformatics analysis of cotton Dof genes was performed in this study. Protein lengths, molecular weights, and theoretical isoelectric points of cotton Dofs vary greatly. Gene structure analysis demonstrated that 92.4% cotton Dof genes have 1 or 2 exons. Conserved motif, phylogenetic tree and expression pattern were also analyzed in our study. On the whole, this study provides an extensive resource for understanding the Dof genes in cotton.

Supporting information

S1 Table. Classification and characterization of cotton Dof genes.

(XLSX)

S1 Fig. Phylogenetic relationships, gene structure and architecture of conserved protein motifs in Dof genes. The motifs, numbers 1~10, are displayed in different colored boxes. The sequence length can be estimated using the scale at the bottom.

(EPS)

S2 Fig. Sequence logo of the cotton Dof protein conserved motif1. The font size represents the frequency of the respective amino acid.

(TIF)

S3 Fig. Chromosomal locations of cotton Dof genes.
(EPS)

S4 Fig. The chromosomal distribution and interchromosomal relationships of cotton Dof genes. The lines indicate duplicated Dof pairs.
(TIF)

Author Contributions

Conceptualization: YuZhe Li, QingDe Zhang.

Data curation: YuZhe Li, SenYi Chen, MengJie Liu.

Formal analysis: YuZhe Li, Zhen Liu, SenYi Chen, MengJie Liu.

Funding acquisition: QingDe Zhang.

Investigation: YuZhe Li, Zhen Liu.

Methodology: YuZhe Li, Zhen Liu, MengJie Liu, QingDe Zhang.

Project administration: Zhen Liu, QingDe Zhang.

Resources: Zhen Liu, KaiYue Zhang, MengJie Liu.

Software: YuZhe Li, KaiYue Zhang, SenYi Chen, MengJie Liu.

Supervision: Zhen Liu, KaiYue Zhang.

Validation: Zhen Liu, KaiYue Zhang, MengJie Liu.

Visualization: Zhen Liu.

Writing – original draft: Zhen Liu, QingDe Zhang.

Writing – review & editing: Zhen Liu, QingDe Zhang.

References

1. Yanagisawa S. The Dof family of plant transcription factors. *Trends Plant Sci.* 2002; 7: 555–560. [https://doi.org/10.1016/s1360-1385\(02\)02362-2](https://doi.org/10.1016/s1360-1385(02)02362-2) PMID: 12475498
2. Shigyo M., Tabei N., Yoneyama T. and Yanagisawa S. Evolutionary processes during the formation of the plant-specific Dof transcription factor family. *Plant Cell Physiol.* 2007; 48: 179–185. <https://doi.org/10.1093/pcp/pci044> PMID: 17132629
3. Gupta S., Malviya N., Kushwaha H., Nasim J., Bisht N.C. and Singh V.K., et al. Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. *Planta.* 2015; 241: 549–562. <https://doi.org/10.1007/s00425-014-2239-3> PMID: 25564353
4. Noguero M., Atif R.M., Ochatt S. and Thompson R.D. The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. *Plant Sci.* 2013; 209: 32–45. <https://doi.org/10.1016/j.plantsci.2013.03.016> PMID: 23759101
5. Sun Q., Du X., Cai C., Long L., Zhang S. and Qiao P., et al. To Be a Flower or Fruiting Branch: Insights Revealed by mRNA and Small RNA Transcriptomes from Different Cotton Developmental Stages. *Sci Rep.* 2016; 6: 23212. <https://doi.org/10.1038/srep23212> PMID: 26983497
6. Su J., Ma Q., Li M., Hao F. and Wang C. Multi-Locus Genome-Wide Association Studies of Fiber-Quality Related Traits in Chinese Early-Maturity Upland Cotton. *Front Plant Sci.* 2018; 9: 1169. <https://doi.org/10.3389/fpls.2018.01169> PMID: 30166989
7. Shang B., Zang Y., Zhao X., Zhu J., Fan C. and Guo X., et al. Functional characterization of GhPHOT2 in chloroplast avoidance of *Gossypium hirsutum*. *Plant Physiol Biochem.* 2019; 135: 51–60. <https://doi.org/10.1016/j.plaphy.2018.11.027> PMID: 30500518
8. Wang P., Zhang S., Qiao J., Sun Q., Shi Q. and Cai C., et al. Functional analysis of the GbDWARF14 gene associated with branching development in cotton. *PeerJ.* 2019; 7: e6901. <https://doi.org/10.7717/peerj.6901> PMID: 31143538

9. Wu Y., Liu F., Yang D.G., Li W., Zhou X.J. and Pei X.Y., et al. Comparative Chloroplast Genomics of *Gossypium* Species: Insights into Repeat Sequence Variations and Phylogeny. *Front Plant Sci.* 2018; 9: 376. <https://doi.org/10.3389/fpls.2018.00376> PMID: 29619041
10. Renny-Byfield S., Gong L., Gallagher J.P. and Wendel J.F. Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. *Mol Biol Evol.* 2015; 32: 1063–1071. <https://doi.org/10.1093/molbev/msv001> PMID: 25573906
11. Cai Y., Cai X., Wang Q., Wang P., Zhang Y. and Cai C., et al. Genome sequencing of the Australian wild diploid species *Gossypium austral* highlights disease resistance and delayed gland morphogenesis. *Plant Biotechnol J.* 2019.
12. Guo Z.H., Ma P.F., Yang G.Q., Hu J.Y., Liu Y.L. and Xia E.H., et al. Genome Sequences Provide Insights into the Reticulate Origin and Unique Traits of Woody Bamboos. *Mol Plant.* 2019; 12: 1353–1365. <https://doi.org/10.1016/j.molp.2019.05.009> PMID: 31145999
13. Lu T., Zhang G., Wang Y., He S., Sun L. and Hao F. Genome-wide characterization and expression analysis of PP2CA family members in response to ABA and osmotic stress in *Gossypium*. *PeerJ.* 2019; 7: e7105. <https://doi.org/10.7717/peerj.7105> PMID: 31231596
14. Zhang G., Lu T., Miao W., Sun L., Tian M. and Wang J., et al. Genome-wide identification of ABA receptor PYL family and expression analysis of PYLs in response to ABA and osmotic stress in *Gossypium*. *PeerJ.* 2017; 5: e4126. <https://doi.org/10.7717/peerj.4126> PMID: 29230363
15. Lu T., Zhang G., Sun L., Wang J. and Hao F. Genome-wide identification of CBL family and expression analysis of CBLs in response to potassium deficiency in cotton. *PeerJ.* 2017; 5: e3653. <https://doi.org/10.7717/peerj.3653> PMID: 28828254
16. Sun Q., Wang G., Zhang X., Zhang X., Qiao P. and Long L., et al. Genome-wide identification of the TIFY gene family in three cultivated *Gossypium* species and the expression of JAZ genes. *Sci Rep.* 2017; 7: 42418. <https://doi.org/10.1038/srep42418> PMID: 28186193
17. Chen M., Li K., Li H., Song C.P. and Miao Y. The Glutathione Peroxidase Gene Family in *Gossypium hirsutum*: Genome-Wide Identification, Classification, Gene Expression and Functional Analysis. *Sci Rep.* 2017; 7: 44743. <https://doi.org/10.1038/srep44743> PMID: 28300195
18. Lijavetzky D., Carbonero P. and Vicente-Carbajosa J. Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC Evol Biol.* 2003; 3: 17. <https://doi.org/10.1186/1471-2148-3-17> PMID: 12877745
19. Wang P., Li J., Gao X., Zhang D., Li A. and Liu C. Genome-Wide Screening and Characterization of the Dof Gene Family in Physic Nut (*Jatropha curcas* L.). *Int J Mol Sci.* 2018; 19.
20. Zhang L., Liu B., Zheng G., Zhang A. and Li R. Genome-wide characterization of the SiDof gene family in foxtail millet (*Setaria italica*). *Biosystems.* 2017; 151: 27–33. <https://doi.org/10.1016/j.biosystems.2016.11.007> PMID: 27894970
21. Su Y., Liang W., Liu Z., Wang Y., Zhao Y. and Ijaz B., et al. Overexpression of GhDof1 improved salt and cold tolerance and seed oil content in *Gossypium hirsutum*. *J Plant Physiol.* 2017; 218: 222–234. <https://doi.org/10.1016/j.jplph.2017.07.017> PMID: 28888163
22. Li H., Dou L., Li W., Wang P., Zhao Q. and Xi R., et al. Genome-Wide Identification and Expression Analysis of the Dof Transcription Factor Gene Family in *Gossypium hirsutum* L. *Agronomy.* 2018; 8: 186.
23. Hu Y., Chen J., Fang L., Zhang Z., Ma W. and Niu Y., et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet.* 2019; 51: 739–748. <https://doi.org/10.1038/s41588-019-0371-5> PMID: 30886425
24. Du X., Huang G., He S., Yang Z., Sun G. and Ma X., et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat Genet.* 2018; 50: 796–802. <https://doi.org/10.1038/s41588-018-0116-x> PMID: 29736014
25. Paterson A.H., Wendel J.F., Gundlach H., Guo H., Jenkins J. and Jin D., et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature.* 2012; 492: 423–427. <https://doi.org/10.1038/nature11798> PMID: 23257886
26. Matsuda F., Tsugawa H. and Fukusaki E. Method for assessing the statistical significance of mass spectral similarities using basic local alignment search tool statistics. *Anal Chem.* 2013; 85: 8291–8297. <https://doi.org/10.1021/ac401564v> PMID: 23944154
27. Finn R.D., Clements J. and Eddy S.R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011; 39: W29–W37. <https://doi.org/10.1093/nar/gkr367> PMID: 21593126
28. Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A. and McWilliam H., et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23: 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404> PMID: 17846036

29. Hall B.G. Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol.* 2013; 30: 1229–1235. <https://doi.org/10.1093/molbev/mst012> PMID: 23486614
30. Bailey T.L., Boden M., Buske F.A., Frith M., Grant C.E. and Clementi L., et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009; 37: W202–W208. <https://doi.org/10.1093/nar/gkp335> PMID: 19458158
31. Voorrips R.E. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered.* 2002; 93: 77–78. <https://doi.org/10.1093/jhered/93.1.77> PMID: 12011185
32. Wang Y., Tang H., Debarry J.D., Tan X., Li J. and Wang X., et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012; 40: e49. <https://doi.org/10.1093/nar/gkr1293> PMID: 22217600
33. Wang D., Zhang Y., Zhang Z., Zhu J. and Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics.* 2010; 8: 77–80. [https://doi.org/10.1016/S1672-0229\(10\)60008-3](https://doi.org/10.1016/S1672-0229(10)60008-3) PMID: 20451164
34. Dong C., Hu H. and Xie J. Genome-wide analysis of the DNA-binding with one zinc finger (Dof) transcription factor family in bananas. *Genome.* 2016; 59: 1085–1100. <https://doi.org/10.1139/gen-2016-0081> PMID: 27831816
35. Xu F., Liu H., Xu Y., Zhao J., Guo Y. and Long L., et al. Heterogeneous expression of the cotton R2R3-MYB transcription factor GbMYB60 increases salt sensitivity in transgenic Arabidopsis. *Plant Cell, Tissue and Organ Culture (PCTOC).* 2018; 133: 15–25.
36. Jiang Y., Xie Q., Wang W., Yang J., Zhang X. and Yu N., et al. Medicago AP2-Domain Transcription Factor WRI5a Is a Master Regulator of Lipid Biosynthesis and Transfer during Mycorrhizal Symbiosis. *Mol Plant.* 2018; 11: 1344–1359. <https://doi.org/10.1016/j.molp.2018.09.006> PMID: 30292683
37. Zhang P., Wang R., Ju Q., Li W., Tran L.P. and Xu J. The R2R3-MYB Transcription Factor MYB49 Regulates Cadmium Accumulation. *Plant Physiol.* 2019; 180: 529–542. <https://doi.org/10.1104/pp.18.01380> PMID: 30782964
38. Mun J.H., Kwon S.J., Yang T.J., Seol Y.J., Jin M. and Kim J.A., et al. Genome-wide comparative analysis of the Brassica rapa gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol.* 2009; 10: R111. <https://doi.org/10.1186/gb-2009-10-10-r111> PMID: 19821981