RESEARCH ARTICLE

Genome-wide identification and expression analysis of the NCED family in cotton (Gossypium hirsutum L.)

QingHua Li*, XianTao Yu*, Long Chen, Gang Zhao, ShiZhou Li, Hao Zhou, Yu Dai, Na Sun, YongFei Xie, JunShan Gao, DaHui Li, Xu Sun, Ning Guo*

College of Life Sciences, Anhui Agricultural University, Hefei, Anhui, China

These authors contributed equally to this work.
* guoing020@163.com

Abstract

Abscisic acid (ABA) is an important plant hormone that plays multiple roles in regulating growth and development as well as in stress responses in plants. The NCED gene family includes key genes involved in the process of ABA synthesis. This gene family has been found in many species; however, the function of the NCED gene family in cotton is unclear. Here, a total of 23 NCED genes (designated as GhNCED1 to GhNCED23) were identified in cotton. Phylogenetic analysis indicated that the identified NCED proteins from cotton and Arabidopsis could be classified into 4 subgroups. Conserved motif analysis revealed that the gene structure and motif distribution of proteins within each subgroup were highly conserved. qRT-PCR and ABA content analyses indicated that NCED genes exhibited stage-specific expression patterns at tissue development stages. GhNCED5, GhNCED6 and GhNCED13 expression was similar to the change in ABA content, suggesting that this gene family plays a role in ABA synthesis. These results provide a better understanding of the potential functions of GhNCED genes.

Introduction

Abscisic acid (ABA) plays an important role in the process of plant growth, development and stress tolerance [1]. In plants, ABA can be synthesized directly or indirectly, and the indirect C40 pathway for ABA synthesis is utilized in most higher plants [2]. In the C40 pathway, zeaxanthin is converted to violaxanthin by zeaxanthin epoxidase (ZEP) and then converted to neoxanthin. Violaxanthin and neoxanthin are cleaved into xanthoxin by catalysis of 9-cis-epoxycarotenoid dioxygenase (NCED) [3], and xanthoxin is converted to abscisic aldehyde by short-chain alcohol dehydrogenase (ABA2). Finally, abscisic aldehyde is oxidized to ABA by aba-aldehyde oxidase (AAO).

9-Cis-epoxycarotenoid dioxygenase (NCED) has been identified in many plants and is a rate-limiting enzyme for ABA biosynthesis. Early research indicated that a slight divergence occurred in the NCED subfamily and the exon position is highly conserved across diverse lineages. A slow evolutionary rate occurred in NCED genes, which might facilitate tissue-specific functional divergence in NCED subfamilies [4].
The NCED gene was first cloned in the maize ABA-deletion mutant Vp14 [3, 5]. Furthermore, nine NCED genes have been identified in Arabidopsis, and five of these genes (AtNCED2, AtNCED3, AtNCED5, AtNCED6, and AtNCED9) are related to ABA synthesis [6]. The study demonstrated that improved drought tolerance by manipulation of the AtNCED3 gene results in the accumulation of endogenous ABA [7]. Subsequently, NCED genes have been identified in Tamatim [8], cowpea [9] and rice [10].

Cotton (Gossypium hirsutum l.) is one of the most widely cultivated and important crops in the world [11]. Upon completion of whole-genome sequencing of cotton, the whole-genome sequence of the upland cotton genetic standard line TM-1 was recently published [12, 13]. However, investigations of NCED genes from cotton are limited. The role of the NCED family in cotton development has not been clearly identified. We studied the whole genome of upland cotton and identified all the potential GhNCED family genes using bioinformatics analysis. We identified three genes related to the change in ABA content during cotton seedling growth and development. The findings of this study provide new insights into functional differentiation and evolution of the NCED gene family in plants.

Materials and methods

Plant materials and genomic database

Upland cotton TM-1 was provided by the Beijing Academy of Agricultural Sciences. Cotton materials were cultured in light incubator at 21˚C with 16 h of light and 8 h of darkness per d. The samples were collected at 15, 18, 21, 24, and 27 d and stored at -80˚C for further analysis. The genome information of cotton were derived from the database (http://mascotton.njau.edu.cn), and the genome data of Arabidopsis were derived from the database (http://www.arabidopsis.org)/.

Identification of the NCED gene family in cotton

BioEdit software was used to establish a local database using the genome-wide amino acid (AA) sequence of upland cotton as a data source. The AA sequences of A. thaliana NCED proteins were used as query sequences, and the BLASTP algorithm (E-value = 0.001) in the established local database of upland cotton was used to search for NCED genes in cotton. To confirm members of the upland cotton NCED family, a hidden Markov model of the NCED domain was used in the Pfam (http://pfam.wustl.edu/hmmsearch.shtml) (E-value = 0.001) and SMART (http://smart.embl-heidelberg.de) databases. The ClustalW tool in MEGA5.0 software [14] was used to conduct multiple sequence alignment of the obtained NCED gene sequences to remove repeat sequences. The obtained AA sequences were analyzed online using ExPAsy (http://www.expasy.org/) to determine the physical parameters of the GhNCED proteins, including their isoelectric points (PIs) and molecular weights (MWs).

Conserved motifs and gene structure analysis of the GhNCED gene family

Multiple Expectation Maximization for Motif Elicitation (MEME) (http://meme.sdsc.edu/meme4_3_0/intro.html) was used to analyze the conserved motifs of NCED gene family proteins in upland cotton [15]. The maximum value of the motif was set to 30, and the length of the motif was set between 6 and 200. Then, the TBtools tool was used to visualize the conserved motifs. Gene Structure Display Server (GSDS, http://gsds.cbi.pku.edu.cn/) analytical tools were used for visualization of the exon substructure of NCED family genes.
Phylogenetic analysis of the NCED gene family in cotton and Arabidopsis
The ClustalW algorithm provided in MEGA 6.0 software was used to conduct multisequence alignment of the AA sequence of the NCED genes in upland cotton and Arabidopsis. All the NCED protein sequences were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/). The phylogenetic tree was constructed by using the neighbor-joining (NJ) method with 1000 bootstrap test.

Chromosome localization of the GhNCED gene family
The starting position of each NCED gene on the chromosome was obtained from the upland cotton genome database. The chromosomal locations of these genes were mapped using MapInspect software (http://mapinspect.software.informer.com).

ABA extraction and HPLC analysis
Cotton seedlings were grown in a light incubator in the laboratory. The first true leaves from seedlings germinated at 15 d (0 d). ABA was extracted from cotton leaves every 3 d (15, 18, 21, 24 and 27 d) and detected by high-performance liquid chromatography (HPLC). For ABA extraction, approximately 1 g of cotton leaves was weighed and extracted with methanol at 4˚C for 12 h. Each mix was centrifuged for 15 min at 5000 rpm, and the supernatants were then stored at 4˚C. Each precipitate was also extracted with methanol twice. Polyvinylpolypyrrolidone (PVPP) was added to adsorb phenolics and pigments from the resulting supernatants. Each sample was oscillated (200 rpm for 1 h, at 4˚C) and frozen in a cryogenic refrigerator (-80˚C) to form a frozen body. The frozen bodies were thawed (4˚C) and then centrifuged (10000 rpm for 15 min). Each precipitate was washed with methanol three times and then rotary-evaporated to dryness. The resulting precipitates were resuspended in methanol and were filtered to produce filtrates for HPLC detection (UV 210 nm, 1 mL/min).

Expression analysis of the GhNCED gene
The qRT-PCR primers are listed in S1 Table. Total RNA was extracted from upland cotton leaves at different developmental stages using TRizol Reagent. cDNA synthesis was performed using an RT reagent kit (Tiangen, China). qRT-PCR was analyzed in a 20-μL reaction system (including 10 μL SYBR ™; Premix Ex Taq™ II (2 x), 2 μL cDNA, and 0.8 μL upstream and downstream primers) and a simple procedure (50˚C for 2 min; 40 cycles at 95˚C for 30 s, 95˚C for 5 s, and 60˚C for 20 s; and a final extension at 72˚C for 10 min). The GhUBQ gene was used as a control, and each sample was repeated 3 times. The relative expression levels were calculated using the 2^−Ct method. All experiments included three biological replicates.

Results
Identification of the GhNCED gene family
To identify the NCED gene family of G. hirsutum, we performed BlastP searches against the cotton genomics databases using the NCED domain in Arabidopsis thaliana. A total of twenty-three NCED family members were identified from G. hirsutum, and these genes were provisionally named GhNCED1 to GhNCED23 (Table 1). In addition, the basic information for all the GhNCED genes was obtained from the ExPAsy online website, and the lengths, MWs, PIs and chromosomal locations of the GhNCED genes are listed in Table 1. The results showed that MW exceeded 50 for all the GhNCED genes except for GhNCED14 and GhNCED19. In addition, most of the GhNCED genes had PI values less than 7, indicating that the proteins
encoded by the GhNCED genes are weak acids and that these proteins may play a role in cells in an acidic environment.

**Genetic structure and phylogenetic evolution of the GhNCED family**

According to the structural features of GhNCED proteins, the GhNCED genes can be divided into four subgroups (Fig 1). We analyzed the motif distribution using the MEME web server. Five genes (GhNCED21, GhNCED23, GhNCED18, GhNCED20 and GhNCED22) contained 5–12 introns, whereas two genes (GhNCED17 and GhNCED19) contained only one intron each. The remaining GhNCED genes had no introns; each only had a single coding sequence. Clustering analysis revealed that these genes were in the same branch, indicating that their functions are highly consistent.

**Analysis of the phylogenetic relationships of the GhNCED genes**

The phylogenetic tree of NCED proteins from *G. hirsutum* and *Arabidopsis thaliana* was constructed with Mega6.0 software (Fig 2). There were 23 candidate GhNCED proteins from *G. hirsutum* and 9 AtNCED proteins from *Arabidopsis thaliana*. Through comparison of the sequence features of their NCED proteins (Fig 3), we found that the NCED genes in upland cotton and *Arabidopsis* could be divided into four subgroups: Group I, Group II, Group III and Group IV. We found 5 AtNCED genes (AtNCED2, AtNCED3, AtNCED5, AtNCED6 and AtNCED9) and 14 GhNCED genes (GhNCED1 to GhNCED14) that were directly related to ABA synthesis, and these genes clustered together in Group I. AtNCED1 and AtNCED4 as well as GhNCED15, GhNCED16, GhNCED17 and GhNCED19 were clustered together in Group II.

---

Table 1. Information for the GhNCED gene family in cotton.

| Gene   | Protein ID | Chromosome Localization | ORF | Number of AAs | MW  | PI   |
|--------|------------|-------------------------|-----|---------------|-----|------|
| GhNCED1| CotAD_25649| 9                       | 1821| 607           | 67.64| 6.06 |
| GhNCED2| CotAD_75526| scaffold                | 1818| 606           | 67.71| 6.41 |
| GhNCED3| CotAD_45615| scaffold                | 1806| 602           | 67.29| 6.27 |
| GhNCED4| CotAD_27475| 7                       | 1518| 506           | 56.61| 5.59 |
| GhNCED5| CotAD_41298| 10                      | 1806| 602           | 67.26| 7.35 |
| GhNCED6| CotAD_04169| 5                       | 1716| 572           | 63.26| 6.06 |
| GhNCED7| CotAD_45595| scaffold                | 1791| 597           | 65.47| 6.53 |
| GhNCED8| CotAD_14690| scaffold                | 1716| 572           | 63.12| 6.00 |
| GhNCED9| CotAD_13950| 13                      | 1767| 589           | 65.19| 5.96 |
| GhNCED10| CotAD_09370| 7                       | 1767| 589           | 65.19| 6.10 |
| GhNCED11| CotAD_35755| scaffold                | 1791| 597           | 65.64| 6.60 |
| GhNCED12| CotAD_11450| 6                       | 1806| 602           | 67.45| 8.78 |
| GhNCED13| CotAD_67889| scaffold                | 1806| 602           | 67.52| 8.94 |
| GhNCED14| CotAD_26725| 11                      | 1059| 353           | 38.99| 8.81 |
| GhNCED15| CotAD_42837| 8                       | 1815| 605           | 66.41| 6.71 |
| GhNCED16| CotAD_17632| 8                       | 1806| 602           | 66.32| 6.95 |
| GhNCED17| CotAD_44591| 6                       | 1509| 503           | 57.12| 8.32 |
| GhNCED18| CotAD_34311| 11                      | 1899| 633           | 71.03| 6.24 |
| GhNCED19| CotAD_27691| 9                       | 1221| 407           | 46.09| 7.01 |
| GhNCED20| CotAD_00217| 9                       | 1812| 604           | 67.55| 6.01 |
| GhNCED21| CotAD_34301| 11                      | 1680| 560           | 62.87| 6.53 |
| GhNCED22| CotAD_00216| 9                       | 1710| 570           | 64.23| 5.97 |
| GhNCED23| CotAD_22927| 10                      | 1701| 567           | 63.18| 6.25 |

https://doi.org/10.1371/journal.pone.0246021.t001
AtNCED7, GhNCED18, GhNCED20 and GhNCED22 were clustered together in Group III. The remaining three genes (AtNCED8 with GhNCED21 and GhNCED23) were clustered together in Group IV.

Analysis of the conserved motifs of the GhNCED gene family

Thirty conserved motifs of NCED proteins in upland cotton were identified using the MEME web server (Fig 4). The types and numbers of motifs of the genes within each group were more similar within groups than among groups. In Group I, four GhNCED genes (GhNCED1, GhNCED3, GhNCED5 and GhNCED13) exhibited conserved motifs. Among them, the conserved motif of GhNCED4 is different from other genes, and it may play a different role in ABA synthesis. In Group II, Group III and Group IV, the motifs of the GhNCED genes were more conservative, suggesting that these genes may have the same function and may have been highly conserved in the course of evolution.

Chromosome localization of the GhNCED gene family

Mapinspect software was used to locate 23 GhNCED genes on 26 chromosomes of G. hirsutum (Fig 5). In total, 17 NCED genes were mainly distributed in eight sections of Chr 5, 6, 7, 8, 9,
10, 11 and 13 in *G. hirsutum*. Among these chromosomes, Chr 9 contained the largest number (four) of *GhNCED* genes; Chr 11 contained three *GhNCED* genes; Chr 6, 7, 8 and 10 contained two *GhNCED* genes each; and Chr 1 and 13 contained only one *GhNCED* gene each. According to the definition of tandem replication [12, 13], six replication events occurred between *GhNCED* genes.

Expression analysis of ABA in upland cotton at different stages

When the first true leaf grows, the plant is in the seedling stage. Cotton appears in the rapid growth stage. The ABA content was assessed to understand the dynamic changes in ABA in the growth and development stages of cotton. An ABA standard curve was drawn using an ABA standard from SIGMA (Fig 6). ABA solutions at five concentrations (0.05, 0.10, 0.15, 0.20 and 0.30 mg/mL) were prepared for HPLC determination, and an ABA standard curve was obtained. ABA was extracted from cotton leaves every 3 d (Fig 7). The results showed that the ABA content varied among the different stages of upland cotton. The ABA content was high in the early stage, decreased at 18 d and then increased to different extents at 21, 24 and 27 d. However, the ABA content was not significantly different among stages of upland cotton.
Expression analysis of the GhNCED gene family

To analyze the expression of GhNCED genes at five developmental stages, primers for 23 NCED genes were designed (S1 Table), and qRT-PCR was used to study the expression patterns of the five developmental stages at 15, 18, 21, 24 and 27 d after germination (Fig 8). Most of the Group I genes exhibited higher expression in the early stage than in the late stage. The change trend of GhNCED gene expression was similar to that of ABA content. Group I GhNCED genes exhibit functions similar to those of Arabidopsis NCED genes and directly affect ABA synthesis. Other genes, such as GhNCED12, GhNCED16, GhNCED17 and GhNCED18, exhibited low expression in the early stage and high expression in the later stage. These results indicate that these genes may function in ABA synthesis inhibition. The expression trends of other genes were more variable, indicating that they play indirect roles in ABA synthesis.

Discussion

In recent years, a large number of NCEDs have been identified and analyzed in various species using bioinformatics technology. Research progress has been slower for the GhNCED family in cotton than in other species. To date, insufficient information is available about NCEDs in cotton. In the present study, 23 GhNCED genes were identified from the cotton genome database.

Gene loss and duplication are the main mechanisms driving species evolution, generating novel gene functions and promoting the formation of gene families. NCED family members...
Fig 4. Distribution of the conserved motifs of NCED proteins in *G. hirsutum* L.

https://doi.org/10.1371/journal.pone.0246021.g004

Fig 5. Chromosomal distribution map of the NCED genes in *Gossypium hirsutum* L. The chromosome numbers are indicated at the top of each bar, while the size of a chromosome is indicated by its relative length. The unit of the left scale is Mb, and the short line indicates the approximate position of the *GhNCED* gene on the corresponding chromosome.

https://doi.org/10.1371/journal.pone.0246021.g005
have been identified in many plants. Gene duplication and differentiation are the principal pathways for forming new genes and extending gene functions [16]. Members that cluster in similar subgroups commonly exhibit similar functions. Therefore, phylogenetic analysis could facilitate the development of functional genomics. In the present study, we identified 32 NCED proteins with full-length domain sequences from cotton and Arabidopsis and classified them into 4 subfamilies on the basis of their sequence structures and phylogenetic relationships.

Phylogenetic trees revealed fourteen GhNCED proteins (60.9%, 14 of 23) in Group I, and five AtNCED proteins (55.6%, 5 of 9) were assigned to this group. These results suggest that the GhNCED proteins in this subgroup may have similar functions in AtNCED proteins. Intragenic-exon pattern analysis showed that most of the 23 genes contained only one exon and a few contained two or more exons. Numerous reports have shown that the upland cotton genome belongs to an AADD-type tetraploid plant. Although the A and D genomes are highly collinear [17, 18], they have undergone asymmetric evolution, which is well represented in our study. The 23 GhNCED genes were distributed on 8 chromosomes of the A genome, which is likely to be the result of the higher gene loss and higher inactivation frequency in the D sub-group than in the A subgroup.

Gene expression profiles often have some connection with gene function. To date, although the patterns of NCED gene expression have been determined in other plants, there have been
fewer detailed studies of the expression of cotton NCED genes. In the present study, the expression profiles of GhNCED genes in different periods were examined using qRT-PCR, and an HPLC method was established for determination of endogenous ABA content. The ABA content exhibited dynamic changes in a normal environment. The expression pattern of three genes (GhNCED5, GhNCED6 and GhNCED13) was similar to that of ABA, which may play a role in regulating ABA synthesis. Our study provides a theoretical foundation for further research on NCED genes, such as investigations of the mechanisms by which NCED genes respond to ABA signaling to regulate the growth and development of upland cotton.

**Conclusions**

In this study, a total of 23 GhNCED genes were identified from the cotton genome database, and the phylogenetic relationships, exon-intron structure, conserved motifs, chromosomal
localization, gene expression profiles and ABA content were comprehensively evaluated under normal growth. The bioinformatics results and analysis of NCED gene expression profiles provide a useful foundation for future studies aimed at understanding the potential role of each GhNCED gene in regulating growth and development of *Gossypium hirsutum* L.

Supporting information

S1 Table. Primer sequences.

(DOCX)

Author Contributions

Conceptualization: Ning Guo.

Formal analysis: JunShan Gao, DaHui Li, Xu Sun, Ning Guo.

Investigation: ShiZhou Li, Hao Zhou.

Methodology: QingHua Li, XianTao Yu, Long Chen, Na Sun.

Project administration: Ning Guo.

Resources: Long Chen, ShiZhou Li, Na Sun, YongFei Xie, Ning Guo.

Software: QingHua Li, Gang Zhao, Hao Zhou, Yu Dai, YongFei Xie.

Supervision: Gang Zhao, Yu Dai, JunShan Gao, DaHui Li, Xu Sun, Ning Guo.

Validation: XianTao Yu.

Visualization: XianTao Yu.

Writing – original draft: XianTao Yu.

Writing – review & editing: QingHua Li.

References

1. Zeevaart JA. Metabolism of Abscisic Acid and Its Regulation in Xanthium Leaves during and after Water Stress. Plant Physiol. 1983 Mar; 71(3):477–81. https://doi.org/10.1104/pp.71.3.477 PMID: 16662852.

2. Nambara E, Marion-Pol l A. Abscisic acid biosynthesis and catabolism. Annu Rev Plant Biol. 2005; 56:165–85. https://doi.org/10.1146/annurev.arplant.56.032604.144046 PMID: 15862093.

3. Schwartz SH, Tan BC, Gage DA, Zeevaart JA, McCarty DR. Specific oxidative cleavage of carotenoids by VP14 of maize. Science. 1997 Jun 20; 276(5320):1872–4. https://doi.org/10.1126/science.276.5320.1872 PMID: 9188535.

4. Priya R, Siva R. Analysis of phylogenetic and functional diverge in plant nine-cis epoxycarotenoid dioxygenase gene family. J Plant Res. 2015 Jul; 128(4):519–34. https://doi.org/10.1007/s10265-015-0726-7 PMID: 25929830.

5. Tan BC, Schwartz SH, Zeevaart JA, McCarty DR. Genetic control of abscisic acid biosynthesis in maize. Proc Natl Acad Sci U S A. 1997 Oct 28; 94(22):12235–40. https://doi.org/10.1073/pnas.94.22.12235 PMID: 9342392.

6. Thompson AJ, Jackson AC, Symonds RC, Mulholland BJ, Dadswell AR, Blake PS, et al. Ectopic expression of a tomato 9-cis-epoxycarotenoid dioxygenase gene causes over-production of abscisic acid. Plant J. 2000 Aug; 23(3):363–74. https://doi.org/10.1046/j.1365-313x.2000.00789.x PMID: 10929129.

7. Iuchi S, Kobayashi M, Taiji T, Naramoto M, Seki M, Kato T, et al. Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in Arabidopsis. Plant J. 2001 Aug; 27(4):325–33. https://doi.org/10.1046/j.1365-313x.2001.01096.x PMID: 11532178.
8. Burbidge A, Grieve TM, Jackson A, Thompson A, McCarty DR, Taylor IB. Characterization of the ABA-deficient tomato mutant notabilis and its relationship with maize Vp14. Plant J. 1999 Feb; 17(4):427–31. https://doi.org/10.1046/j.1365-313x.1999.00386.x PMID: 10205899.

9. Iuchi S, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K. A stress-inducible gene for 9-cis-epoxy-carotenoid dioxygenase involved in abscisic acid biosynthesis under water stress in drought-tolerant cowpea. Plant Physiol. 2000 Jun; 123(2):553–62. https://doi.org/10.1104/pp.123.2.553 PMID: 10859185.

10. Zhu G, Ye N, Zhang J. Glucose-induced delay of seed germination in rice is mediated by the suppression of ABA catabolism rather than an enhancement of ABA biosynthesis. Plant Cell Physiol. 2009 Mar; 50(3):644–51. https://doi.org/10.1093/pcp/pcp022 PMID: 19208695.

11. Yide Huang, Wei-Chuan Yao. Crop cultivation. China Agricultural University Press; 2002.

12. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, et al. Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. Nat Biotechnol. 2015 May; 33(5):524–30. https://doi.org/10.1038/nbt.3208 PMID: 25893780.

13. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, et al. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol. 2015 May; 33(5):531–7. https://doi.org/10.1038/nbt.3207 PMID: 25893781.

14. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007 Nov 1; 23(21):2947–8. https://doi.org/10.1093/bioinformatics/btm404 PMID: 17846036.

15. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research. 2009; 37:W202–W208. https://doi.org/10.1093/nar/gkp335 PMID: 19458158

16. Qiao L, Zhang X, Han X, Zhang L, Li X, Zhan H, et al. A genome-wide analysis of the auxin/indole-3-acetic acid gene family in hexaploid bread wheat (Triticum aestivum L.). Front Plant Sci. 2015 Sep 30; 6:770. https://doi.org/10.3389/fpls.2015.00770 PMID: 26483801.

17. Song A, Gao T, Li P, Chen S, Guan Z, Wu D, et al. Transcriptome-Wide Identification and Expression Profiling of the DOF Transcription Factor Gene Family in Chrysanthemum morifolium. Front Plant Sci. 2016 Feb 23; 7:199. https://doi.org/10.3389/fpls.2016.00199 PMID: 26941763.

18. Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, et al. Genome sequence of the cultivated cotton Gossypium arboreum. Nat Genet. 2014 Jun; 46(6):567–72. https://doi.org/10.1038/ng.2987 PMID: 24836287.