Genome-Wide Identification and Characterization of The MIOX Gene Family in Cotton

Zhaoguo Li

Zhengzhou University, Zhengzhou, Henan 450001, China

Zhen Liu

Research Base, Anyang Institute of Technology, State Key Laboratory of Cotton Biology, Anyang, Henan 455000, China

Yangyang Wei

Research Base, Anyang Institute of Technology, State Key Laboratory of Cotton Biology, Anyang, Henan 455000, China

Yuling Liu

Research Base, Anyang Institute of Technology, State Key Laboratory of Cotton Biology, Anyang, Henan 455000, China

Pengtao Li

Research Base, Anyang Institute of Technology, State Key Laboratory of Cotton Biology, Anyang, Henan 455000, China

Linxue Xing

Research Base, Anyang Institute of Technology, State Key Laboratory of Cotton Biology, Anyang, Henan 455000, China

Mengjie Liu

Research Base, Anyang Institute of Technology, State Key Laboratory of Cotton Biology, Anyang, Henan 455000, China

Quanwei Lu

Research Base, Anyang Institute of Technology, State Key Laboratory of Cotton Biology, Anyang, Henan 455000, China

Renhai Peng (✉ aydxprh@163.com)

Research Base, Anyang Institute of Technology, State Key Laboratory of Cotton Biology, Anyang, Henan 455000, China

Research Article

Keywords: Cotton, Myo-inositol oxygenase (MIOX), Gene family, Abiotic stress, Gene expression

Posted Date: December 29th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-130287/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The enzyme myo-inositol oxygenase (MIOX) catalyzes the myo-inositol into glucuronic acid. Previous studies indicated that MIOX may play an important role in plant responses to abiotic stresses. Cotton is a major source of natural and renewable textile. However, the MIOX gene family of cotton has not been systematically identified and characterized yet.

Results: In this study, 6 MIOX genes were identified from all of the three diploid cotton species (*Gossypium arboreum*, *Gossypium herbaceum* and *Gossypium raimondii*), 12 MIOX genes were identified from two domesticated tetraploid cottons *Gossypium hirsutum*, *Gossypium barbadense*, and 11 MIOX genes were identified from three wild tetraploid cottons *Gossypium tomentosum*, *Gossypium mustelinum* and *Gossypium darwinii*. The number of MIOX genes in tetraploid cotton genome is roughly twice that of diploid cotton genome. Members of MIOX family were classified into six groups based on the phylogenetic analysis. Integrated analysis of collinearity events and chromosome locations suggested that both whole genome duplication and segmental duplication events contributed to the expansion of MIOX genes during cotton evolution. The ratios of non-synonymous (Ka) and synonymous (Ks) substitution rates revealed that purifying selection was the main force driving the evolution of MIOX genes. Numerous cis-acting elements related to light responsive element, defense and stress responsive element were identified in the promoter of the MIOX genes. Expression analyses of MIOX genes based on RNA-seq data showed that MIOX genes within the same group shared similar expression patterns with each other.

Conclusions: In this work, we systematically analyzed MIOX genes from eight *Gossypium* genomes and the *Gossypioides kirkii* genome using a set of bioinformatics approaches. All of these results provide the foundation for further study of the biological functions of MIOX genes in cotton environmental adaptability.

Background

Myo-inositol (MI) is a small molecule that is important in many different developmental and physiological processes in eukaryotic cells[1]. MI participates in the production of stress-related molecules, cell wall biosynthesis, phytic acid biosynthesis, auxin transport and storage[2]. The Myo-inositol oxidation pathway (MIOP) effectively control MI homeostasis. In this pathway, Myo-inositol oxygenase (MIOX, E.C. 1.13.99.1) is a key enzyme that catalyzes the conversion of MI into D-glucuronic acid (D-GlcUA), which is subsequently activated to UDP-D-glucuronic acid (UDP-GlcA) and serves as an important precursor for plant cell wall polysaccharides[3, 4]. Furthermore, previous reports suggested that MIOX might also be involved in the production of ascorbate, consequently, protection from ROS-mediated injury[5]. Studies have shown that MIOX plays a role in the response to environmental stresses. In *Arabidopsis*, MIOX is encoded by a gene family of four members, it has been reported that *Arabidopsis* MIOX4 over express can increase the tolerance to cold, heat and salt[3, 6–8]. In addition, it has also been

reported that rice MIOX has a specific function in drought stress tolerance by decreasing oxidative damage[9].

Cotton (*Gossypium*) is one of the most economically important fiber crop plants throughout the world. It is a remarkably diverse genus, with over 50 species divided into eight diploid genomic groups, designated A-G and K, and one tetraploid genomic group, namely AD[10]. Divergence analysis suggests that the major diploid branches of the cotton genus diverged about 7 ~ 11 million years ago, and the polyploid clade originated circa 1 ~ 2 million years ago[11–14]. Due to the rapid development of next generation sequencing, the whole genomes have been reported for the diploid D-genome of *G. raimondii*[15], A-genome of *G. arboreum*[16], *G. herbaceum*[17], and allopolyploid AD genome species of *G. hirsutum*[18], *G. barbadense*[19], *G. darwinii*, *G. mustelinum* and *G. tomentosum*[20]. The sequence of these genomes provides useful resources for studying the gene families and a series of families have been genome-wide analyzed in cotton[21, 22].

Although cotton has better abiotic stress tolerance compared with other crops, its yield and fiber quality under saline-alkali, drought and heat stress are still under great threat. Considering the role of MIOX gene in various abiotic stress responses in plants and there is no genome-wide identification and characterization of MIOX gene family members in cotton. We conducted a systematic analysis of the cotton MIOX gene family by genome-wide search, the phylogeny evolutionary relationships, group classification, conserved motifs and the expression profile of MIOX were analyzed. Our results will provide good foundation for understanding the key roles of MIOX genes in cotton stress-responsive and other biological processes.

Results

Identification of the MIOX Gene Family

A total of 81 MIOX genes were identified from nine species, including 75 MIOXs in eight *Gossypium* genomes and 6 MIOXs in the *Gossypioides kirkii* genome. Except for one MIOX from *G. herbaceum* has significantly more amino acids (896 aa) and two MIOXs from *G. barbadense* has significantly less amino acids (101 aa and 113 aa), the others are relatively conservative (Additional file 1: Table S1). Furthermore, we found that all of the four diploid species contained 6 MIOX genes, the number of MIOX genes in tetraploid cotton is roughly twice that of diploid cotton, but the three wild tetraploid cottons (*G. tomentosum*, *G. mustelinum* and *G. darwinii*) missed a MIOX gene.

For convenience, we renamed the MIOX gene family. Gh, Gb, Ga, Gher, Gr, Gd, Gt, Gm, and Gkk were used as prefixes before the names of MIOX genes from *G. hirsutum*, *G. barbadense*, *G. arboreum*, *G. herbaceum*, *G. raimondii*, *G. darwinii*, *G. tomentosum*, *G. mustelinum* and *Gossypioides kirkii* respectively. we assigned the names to MIOX genes as (GhMIOX01- GhMIOX12) for *G. hirsutum*, (GbMIOX01 - GbMIOX12) for *G. barbadense* and so on.

Phylogenetic Analysis of the MIOX Gene Family

To reveal the evolutionary relationship of identified MIOX proteins, the amino acid sequences of 75 proteins from *Gossypium* species, 6 proteins from *Gossypioides kirkii* and 4 proteins from *Arabidopsis thaliana* were used to construct a phylogenetic tree (Fig. 1, Additional file 2: Figure S1). According to the result, 81 MIOX proteins from 8 *Gossypium* species and *Gossypioides kirkii* were divided into six groups of A-F, the number of MIOX genes was very stable in each evolutionary branch, diploid species containing 6 MIOXs had one distribution in each branch, while the other five tetraploid cottons containing 11 or 12 MIOX genes had one or two distribution in each branch. All of these genes showed one-to-one homology relationship among different genomes or subgenomes. Compared to the A~F groups, the group G~I only contained *A. thaliana* MIOX proteins.

Gene Structure and Protein Motifs of the MIOX Gene Family

To explore the structural diversity of MIOX genes, the intron-exon organization of each MIOX gene was analyzed (Fig. 2). The number of introns ranged from 3 to 21, most (52/81) MIOX genes contained 9 or 10 introns. In the same group, the intron numbers exhibited comparable exon-intron structure and intron numbers, while many *Gossypioides kirkii* MIOX genes were different from cotton, implying important evidence for the phylogenetic relationship among members of the MIOX gene family.

Furthermore, we investigated the conserved motifs in MIOX proteins to understand the diversity of motif compositions (Fig. 2). A total of 13 motifs, named Motif 1~Motif 13, were identified in MIOX proteins. The number of motifs varied from 3 to 12 in each MIOX protein and most MIOX proteins within the same group exhibited similar motif composition and arrangement, which indicates that the members of MIOX gene family that clustered in the same group may have similar biological functions. Motif 5 was found in 78 MIOX proteins, Motif 8 was completely missing only in group B, and Motif 13 was specific to group C. The gene structure and motif composition of the MIOX members from each group that obtained from phylogenetic analysis were similar, which indicates the classification was reliable.

Chromosomal Location and Gene Duplications of the MIOX Gene Family

To more intuitively understand the distribution of MIOX family genes on the chromosomes, we performed a chromosomal localization analysis. The result shows that MIOX genes were mapped onto 5/10 different chromosomes of diploid/tetraploid cotton. Each chromosome contained only one MIOX gene usually, while some chromosome contained 2 MIOX genes (Fig. 3). In addition, the distribution of *Gossypioides kirkii* MIOX genes showed similar to *Gossypium*.

Whole genome duplication, segmental duplication and tandem duplication provides major forces that drive the expansion of gene families. The number of MIOX genes in tetraploid *Gossypium* species was twice as much as that in diploid *Gossypium* species which indicates the expansion of MIOX gene family during polyploidization. We searched the segmental duplication using MCscanX within each genome, and identified 133 collinear genes pairs among the 72 MIOX genes. Of the 75 MIOX genes in eight *Gossypium* genomes, only 3 were located outside of the duplicated blocks, while 96% (72 of 75) were located in duplicated regions. In addition, 40, 38 duplication gene-pairs were found between diploid *G. arboreum* A-

genome and tetraploid *G. hirsutum*, *G. barbadense* A-subgenome respectively. 18, 17 duplication gene-pairs were found between diploid *G. raimondii* D-genome and tetraploid *G. hirsutum*, *G. barbadense* D-subgenome respectively (Fig. 4). According to Holub's description, a chromosomal region within 200 kb containing two or more genes was defined as a tandem duplication event[23]. Our results indicated that the MIOX genes of the nine species has no tandem duplication.

Cis-elements in the Promoter of MIOX Genes

Cis-elements in promoters play vital roles in regulating the expression of genes. To gain more insight into the functions of MIOX genes, the cis-regulatory elements were scanned in the 2000 bp upstream of the transcription start sites of cotton MIOX genes (Fig. 5, Additional file 3: Table S2, Additional file 4: Figure S2). The results showed that there were many kinds of response elements, such as light responsive element, defense and stress responsive element involved in drought and low temperature, and hormone responsive element associated with salicylic acid, abscisic acid, gibberellin and MeJA. All of the cotton MIOX genes contained more than one light responsive element, however only 29.33% (22/75) of the cotton MIOX genes contained auxins responsive element.

The cis-elements of MIOX genes in the same phylogenetic group was similar. In addition, within the same group, half of the cis-elements of MIOX genes in tetraploid *Gossypium* species are similar to the diploid A genome species (*G. herbaceum* and *G. arboreum*), and half are similar to the D genome species (*G. raimondii*). These results further indicated the expansion of MIOX gene family during polyploidization.

Tissue-Specific Expression Profiles of MIOX Genes

To study the tissue-specific expression patterns of the MIOX genes, we analyzed the expression profiles of the MIOX genes in different tissues. As shown in Fig. 6, MIOX genes showed different expression levels in different tissues. In *G. hirsutum*, the expression of GhMIOX02 and GhMIOX08 were higher in stem, leaf, torus, pistil, bract and sepal, while their expression was lower in root. The expression of GhMIOX04 and GhMIOX10 was higher in root, stem, leaf, torus and sepal, while their expression was lower in pistil and bract. In *G. barbadense*, GbMIOX02, GbMIOX08, GbMIOX04 and GbMIOX10 had similar expression profiles with *G. hirsutum*. In addition, GhMIOX03 of *G. hirsutum* showed higher transcription level only in root, however, GbMIOX03 of *G. barbadense* showed higher transcription levels in root, stem and pistil.

The expression levels of MIOX genes at ovule and fiber developmental stages were also investigated (Fig. 6). At most stages of development, the expression of GhMIOX08 from *G. hirsutum*, GbMIOX08 from *G. barbadense* were higher than that of the other genes; Several genes were expressed at high levels during specific developmental stages; for instance, GhMIOX03, GhMIOX09, GhMIOX10 from *G. hirsutum* and GbMIOX03, GbMIOX09, GbMIOX10 from *G. barbadense*. In contrast, the RNA transcript levels of GhMIOX05, GhMIOX11 from *G. hirsutum* and GbMIOX05, GbMIOX11 from *G. barbadense* were low at all stages and all tissues. These findings indicated the MIOX genes play differential roles in tissue development.

Stress-Induced Expression Patterns of MIOX Genes

The expression patterns of MIOX genes were further analyzed in *G. hirsutum* and *G. barbadense* exposed to different durations of cold, heat, salt, and drought stresses for different times by RNA-seq data downloaded from the public database (Fig. 7). Based on the clustering analysis, MIOX genes in group C and D general responded to all of these stress, but the genes in other groups were not significantly responded. In addition, it was observed that MIOX genes exhibited variations in expression in response to one or more stresses. In the case of heat and salt treatment, the expression of GhMIOX02 and GhMIOX08 showed up-regulation in the early time points (1h~3h), then down-regulation (3h~6h), and then another up-regulation (6h~24h). While GhMIOX03 was up-regulated in the early time points (1h~6h), but down regulated in the late time points (12h~24h) under code stress. Moreover, GbMIOX02 showed continuously down-regulation at all of the time points under salt stress. These results indicate that 3~6 hour is an important time point for MIOX genes respond to abiotic stress.

Discussion

MI is a crucial substance in various aspects of plant physiology. Plants maintain an MI pool at a basal level throughout their life cycle, and MIOX is used to control the metabolite level of MI in plants[1, 2, 5, 8]. MIOX proteins are conserved and present in nearly all eukaryotes. MIOX is crucial in abiotic stress tolerance and previous studies on the MIOX gene family have been performed in many plants[2, 5], including *Arabidopsis thaliana*[3, 4], *Oryza sativa*[9] and *Solanum lycopersicum*[24]. In this study, we performed a comprehensive identification of MIOX genes in 5 tetraploid cotton species, 3 diploid cotton species and a cotton closely related species *Gossypioides kirkii*, with an aim of understanding the important and diverse roles of this gene family in abiotic stress tolerance in plants.

There were 4 MIOX genes in *Arabidopsis thaliana* and they were not clustered with the groups from *Gossypium* species and *Gossypioides kirkii*. In addition, the number of MIOX genes in tetraploid cotton is roughly twice that of diploid cotton. This suggested that MIOX gene duplication events occurred in diploid cotton before the emergence of tetraploid cotton.

Transposable elements are major evolutionary forces which can cause genomic instability, including genome expansion by amplification using an RNA intermediate[25]. Besides genome expansion, transposable elements facilitate the creation of new candidate genes called retrogenes by means of retroduplication, in which spliced messenger RNA is captured and subsequently integrated into the genome by a retrotransposon[26]. The role of transposable elements in the expansion of many genes has been unexplored in many plants[27–29]. To investigate whether transposable elements played roles in expansion of the MIOX genes, LTR_retriever[30], LTRharvest[31], LTR_FINDER[32], RepeatModeler and RepeatMasker[33] were used to identify transposable elements. We then compared those results with the genome annotation to predict genes inside transposable elements in the 10 genomes. However, the results show that none of MIOX gene was amplified by transposable elements.

An analysis of gene expression patterns can be used to some extent to predict the molecular functions of genes involved in different processes. Our heatmap data showed that most MIOX genes within each group shared similar expression patterns. The MIOX genes in Group D in *G. hirsutum* (GhMIOX02 and GhMIOX08) and *G. barbadense* (GbMIOX02, GbMIOX08) were expressed in stem, leaf, torus, pistil, bract, sepal and responded to cold, heat, salt, and drought stresses. In contrast, the expression of MIOX genes in Group F in *G. hirsutum* (GhMIOX05, GhMIOX11) and *G. barbadense* (GbMIOX05, GbMIOX11) were barely observable. To further elucidate the selection pressure of functional diversification of each MIOX group, we have comprehensively evaluated the non-synonymous (Ka) and synonymous (Ks) substitution rates (Fig. 8). The results showed that the median Ka/Ks ratios for group D and F was 0.267 and 0.344 respectively, indicating that these groups of *Gossypium* MIOXs might play specialized roles in the adaptive evolution of cotton.

Conclusions

We performed a genome-wide analysis of the MIOX gene family in 8 cotton species. The MIOX family was divided into 6 groups based on the phylogenetic tree. To understand the expansion mechanism of MIOX family, the gene duplication events were investigated, and the result indicated that segmental duplication and whole genome duplication were the major driving forces of MIOX family expansion, no tandem duplication and transposable element amplification was found. The expression profiles of *G. hirsutum* and *G. barbadense* MIOX genes in different tissues and under abiotic stresses were also analyzed. The results indicated that MIOX genes within the same group shared similar expression pattern and they exhibited different expression levels in different groups. Furthermore, the Ka/Ks ratio suggesting that all groups experienced purifying selection pressure. Our results will provide clues for researchers regarding the evolution and biological functions of MIOXs.

Methods

Identification of MIOX Genes in Cotton

The genome sequences of *G. darwinii*, *G. mustelinum* and *G. tomentosum*[20] were downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/genome/>), the genome sequences of *G. hirsutum*[18], *G. barbadense*[19], *G. arboreum*[16], *G. herbaceum*[17], *G. raimondii*[15] and a cotton closely related species *Gossypioidea kirki*[34] were downloaded from COTTONGEN (<http://www.cottongen.org>). *Arabidopsis* MIOX proteins were used to search possible cotton MIOX sequences by BLASTP using a E value of 1.0E-3[35, 36]. The HMMER[37] program was used to identify MIOX proteins with the Hidden Markov Model of the MIOX domain (PF05153), which was downloaded from the Pfam database (<http://pfam.xfam.org>). 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 sequences. The number of amino acids, molecular weights and theoretical isoelectric points of MIOX proteins were calculated using the ExPASy online server tool (<https://www.expasy.org/>).

Phylogenetic, Gene Structure, Conserved Motif and Cis-elements Analysis of MIOXs

A neighbor-joining tree of MIOX proteins was constructed using MEGA-X with 1000 bootstrap replications[38]. The phylogenetic tree was drawn using EvolView[39]. The gene structure of MIOXs were analyzed with TBtools[40]. The conserved motifs of MIOX protein sequences were obtained using MEME[41]. To investigate putative cis-acting regulatory elements of MIOX genes, 2000 base pair genomic DNA sequences upstream of the initiation codon were retrieved and screened against the Plant CARE database[42].

Chromosomal Distribution and Gene Duplication of MIOX Genes

The chromosome distributions of MIOX genes were extracted from the GFF files and the Mapchart[43] software was used to visually map the chromosomal location. We made use of Multiple collinear scanning toolkits (MCScanX)[44] to detect the gene duplication events and tandem duplications were identified as previously described. The synonymous (Ks) and nonsynonymous (Ka) substitution rates of MIOX genes were calculated by KaKs_Calculator[45].

Expression Patterns of MIOX Genes

To study the expression patterns of MIOX genes, *G. hirsutum* and *G. barbadense* high-throughput transcriptome sequencing data including various tissues, developmental stages and stress treatments were downloaded from the NCBI SRA (PRJNA490626). Trimmomatic[46] was used to remove the adapters and to perform quality control. The program hisat2[47] was used to map the reads to the genomes, then Fragments Per Kilobase of transcript per Million fragments (FPKM) values of the MIOX genes were calculated by Cufflinks[48, 49]. Heat maps of gene expression profiles were drawn by TBtools[40].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Competing interests

The authors declare that they have no competing interests.

Funding

This work is partially support by College Student Innovation and Entrepreneurship Training Program in Henan Province(S202011330003), China Postdoctoral Science Foundation(2018M632761), the National Natural Science Foundation of China (No. 31471548), the National Key Research and Development Program of China (No. 2018YFD0100302) and the Science and Technology Development Project of Anyang City (2018-66-117).

Authors' contributions

RHP and ZL conceived, designed, and supervised this study. ZGL and PTL analyzed the data and wrote the paper, YYW, YLL, LXX, MJL, QWL, revised the paper. All authors reviewed and approved the manuscript.

Acknowledgements

We thank all colleagues in our laboratory for helpful discussions and technical assistance.

References

1. Ye W, Ren W, Kong L, Zhang W, Wang T: **Transcriptomic Profiling Analysis of *Arabidopsis thaliana* Treated with Exogenous Myo-Inositol**. *PLOS ONE* 2016, **11**(9):e161949.
2. Loewus FA, Murthy PPN: **Myo-Inositol metabolism in plants**. *Plant science (Limerick)* 2000, **150**(1):1–19.
3. Alford S: **Myo-Inositol oxygenase is required for responses to low energy conditions in *Arabidopsis thaliana***. *FRONT PLANT SCI* 2012, **3**.
4. Endres S, Tenhaken R: **Down-regulation of the myo-inositol oxygenase gene family has no effect on cell wall composition in *Arabidopsis***. *PLANTA* 2011, **234**(1):157–169.
5. Valluru R, Van den Ende W: **Myo-inositol and beyond-Emerging networks under stress**. *PLANT SCI* 2011, **181**(4):387–400.
6. Nepal N, Yactayo-Chang JP, Medina-Jiménez K, Acosta-Gamboa LM, González-Romero ME, Arteaga-Vázquez MA, Lorence A: **Mechanisms underlying the enhanced biomass and abiotic stress tolerance phenotype of an *Arabidopsis* MIOX over-expresser**. *Plant Direct* 2019, **3**(9):e165.
7. Chen C, Sun X, Duanmu H, Yu Y, Liu A, Xiao J, Zhu Y: **Ectopic expression of a glycine soja myo-inositol oxygenase gene (*GsMIOX1a*) in *Arabidopsis* enhances tolerance to alkaline stress**. *PLOS ONE* 2015, **10**(6):e129998.
8. Endres S, Tenhaken R: **Myoinositol oxygenase controls the level of myoinositol in *Arabidopsis*, but does not increase ascorbic acid**. *PLANT PHYSIOL* 2009, **149**(2):1042–1049.

9. Duan J, Zhang M, Zhang H, Xiong H, Liu P, Ali J, Li J, Li Z: **OsMIOX, a myo-inositol oxygenase gene, improves drought tolerance through scavenging of reactive oxygen species in rice (*Oryza sativa* L.).** *PLANT SCI* 2012, **196**:143–151.
10. Wendel JF, Cronn RC: **Polyploidy and the evolutionary history of cotton.** *ADV AGRON* 2003.
11. Chen Z, Feng K, Grover CE, Li P, Liu F, Wang Y, Xu Q, Shang M, Zhou Z, Cai X *et al*: **Chloroplast DNA structural variation, phylogeny, and age of divergence among diploid cotton species.** *PLOS ONE* 2016, **11**(6):e157183.
12. Chen Z, Grover CE, Li P, Wang Y, Nie H, Zhao Y, Wang M, Liu F, Zhou Z, Wang X *et al*: **Molecular evolution of the plastid genome during diversification of the cotton genus.** *MOL PHYLOGENET EVOL* 2017, **112**:268–276.
13. Wendel JF: **New World tetraploid cottons contain Old World cytoplasm.** *Proc Natl Acad Sci U S A* 1989, **86**(11):4132–4136.
14. Senchina DS, Alvarez I, Cronn RC, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF: **Rate variation among nuclear genes and the age of polyploidy in *Gossypium*.** *MOL BIOL EVOL* 2003, **20**(4):633–643.
15. Udall JA, Long E, Hanson C, Yuan D, Ramaraj T, Conover JL, Gong L, Arick MA, Grover CE, Peterson DG *et al*: **De novo genome sequence assemblies of *Gossypium raimondii* and *Gossypium turneri*.** *G3 (Bethesda)* 2019, **9**(10):3079–3085.
16. Du X, Huang G, He S, Yang Z, Sun G, Ma X, Li N, Zhang X, Sun J, Liu M *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**(6):796–802.
17. Huang G, Wu Z, Percy RG, Bai M, Li Y, Frelichowski JE, Hu J, Wang K, Yu JZ, Zhu Y: **Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution.** *NAT GENET* 2020.
18. Yang Z, Ge X, Yang Z, Qin W, Sun G, Wang Z, Li Z, Liu J, Wu J, Wang Y *et al*: **Extensive intraspecific gene order and gene structural variations in upland cotton cultivars.** *NAT COMMUN* 2019, **10**(1):2989.
19. Wang M, Tu L, Yuan D, Zhu, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G *et al*: **Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*.** *NAT GENET* 2019, **51**(2):224–229.
20. Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, Ding M, Ye W, Kirkbride RC, Jenkins J *et al*: **Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement.** *NAT GENET* 2020, **52**(5):525–533.
21. Li Y, Liu Z, Zhang K, Chen S, Liu M, Zhang Q: **Genome-wide analysis and comparison of the DNA-binding one zinc finger gene family in diploid and tetraploid cotton (*Gossypium*).** *PLOS ONE* 2020, **15**(6):e235317.
22. Ge Q, Cui Y, Li J, Gong J, Lu Q, Li P, Shi Y, Shang H, Liu A, Deng X *et al*: **Disequilibrium evolution of the Fructose-1,6-bisphosphatase gene family leads to their functional biodiversity in *Gossypium* species.**

BMC GENOMICS 2020, **21**(1):379.

23. Holub EB: **The arms race is ancient history in *Arabidopsis*, the wildflower.** *NAT REV GENET* 2001, **2**(7):516–527.
24. Munir S, Mumtaz MA, Ahiakpa JK, Liu G, Chen W, Zhou G, Zheng W, Ye Z, Zhang Y: **Genome-wide analysis of myo-inositol oxygenase gene family in tomato reveals their involvement in ascorbic acid accumulation.** *BMC GENOMICS* 2020, **21**(1):284.
25. Galindo-Gonzalez L, Mhiri C, Deyholos MK, Grandbastien MA: **LTR-retrotransposons in plants: Engines of evolution.** *GENE* 2017, **626**:14–25.
26. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O *et al*: **A unified classification system for eukaryotic transposable elements.** *NAT REV GENET* 2007, **8**(12):973–982.
27. Li M, Wang R, Liu Z, Wu X, Wang J: **Genome-wide identification and analysis of the WUSCHEL-related homeobox (WOX) gene family in allotetraploid *Brassica napus* reveals changes in WOX genes during polyploidization.** *BMC GENOMICS* 2019, **20**(1).
28. Kim S, Park J, Yeom SI, Kim YM, Seo E, Kim KT, Kim MS, Lee JM, Cheong K, Shin HS *et al*: **New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication.** *GENOME BIOL* 2017, **18**(1):210.
29. Jiang SY, Ramachandran S: **Genome-wide survey and comparative analysis of LTR retrotransposons and their captured genes in rice and sorghum.** *PLOS ONE* 2013, **8**(7):e71118.
30. Ou S, Jiang N: **LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons.** *PLANT PHYSIOL* 2018, **176**(2):1410–1422.
31. Ellinghaus D, Kurtz S, Willhoeft U: **LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons.** *BMC BIOINFORMATICS* 2008, **9**(1):1–14.
32. Xu Z, Wang H: **LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons.** *NUCLEIC ACIDS RES* 2007, **35**(Web Server issue):W265-W268.
33. Tempel S: **Using and understanding RepeatMasker.** *Methods Mol Biol* 2012, **859**:29–51.
34. Udall JA, Long E, Ramaraj T, Conover JL, Yuan D, Grover CE, Gong L, Arick MN, Masonbrink RE, Peterson DG *et al*: **The Genome Sequence of *Gossypioides kirkii* illustrates a descending dysploidy in plants.** *FRONT PLANT SCI* 2019, **10**:1541.
35. Nowicki M, Bzhalava D, BaLa P: **Massively parallel implementation of sequence alignment with basic local alignment search tool using parallel computing in Java library.** *J COMPUT BIOL* 2018, **25**(8):871–881.
36. Matsuda F, Tsugawa H, Fukusaki E: **Method for assessing the statistical significance of mass spectral similarities using basic local alignment search tool statistics.** *ANAL CHEM* 2013, **85**(17):8291–8297.
37. Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching.** *NUCLEIC ACIDS RES* 2011, **39**(Web Server issue):W29-W37.

38. Hall BG: **Building phylogenetic trees from molecular data with MEGA.** *MOL BIOL EVOL* 2013, **30**(5):1229–1235.
39. He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S: **Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees.** *NUCLEIC ACIDS RES* 2016, **44**(W1):W236-W241.
40. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R: **TBtools: an integrative toolkit developed for interactive analyses of big biological data.** *MOL PLANT* 2020, **13**(8):1194–1202.
41. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS: **MEME SUITE: tools for motif discovery and searching.** *NUCLEIC ACIDS RES* 2009, **37**(Web Server issue):W202-W208.
42. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S: **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**(1):325–327.
43. Voorrips RE: **MapChart: software for the graphical presentation of linkage maps and QTLs.** *J HERED* 2002, **93**(1):77–78.
44. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H *et al*: **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity.** *NUCLEIC ACIDS RES* 2012, **40**(7):e49.
45. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J: **KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies.** *Genomics Proteomics Bioinformatics* 2010, **8**(1):77–80.
46. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for illumina sequence data.** *BIOINFORMATICS* 2014, **30**(15):2114–2120.
47. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements.** *NAT METHODS* 2015, **12**(4):357–360.
48. Ghosh S, Chan CK: **Analysis of RNA-Seq Data Using TopHat and Cufflinks.** *Methods Mol Biol* 2016, **1374**:339–361.
49. Pollier J, Rombauts S, Goossens A: **Analysis of RNA-Seq data with TopHat and Cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cultures.** *Methods Mol Biol* 2013, **1011**:305–315.

Figures

Figure 1

Phylogenetic tree of MIOX proteins from 8 *Gossypium* species, *Gossypoides kirkii* and *Arabidopsis thaliana*. The different-colored arcs indicate different groups of MIOX proteins. The green circles and red boxes represent MIOX proteins from A genome and D genome, respectively, black triangles represent MIOX proteins from *Gossypoides kirkii*, blue stars represent MIOX proteins from *Arabidopsis thaliana*.

Figure 2

The exon–intron structures and conserved motifs of MIOX. The length of proteins and DNA sequence was estimated using the scale at the bottom, black line indicated the non-conserved amino acid or introns. (a) The conserved motifs of MIOX gene family; (b) The gene structure of MIOX gene family.

Figure 3

Distribution of the MIOX genes on the cotton chromosomes. The vertical bar located on the left side indicates the chromosome sizes in mega bases (Mb), the chromosome number is located above each chromosome.

Figure 4

Collinearity of MIOX genes between different cotton species. (a) Collinearity between A-genome of *G. arboreum* and A-subgenome of *G. hirsutum* and *G. barbadense*; (b) Collinearity between D-genome of *G. raimondii* and D-subgenome of *G. hirsutum* and *G. barbadense*.

Figure 5

Number of MIOX gene cis-elements in each group.

Figure 6

Expression profiles of MIOX genes in different tissues and during different developmental stages of ovule and fiber.

Figure 7

Expression profiles of the MIOX genes in response to different stresses.

Figure 8

The Ka/Ks ratio of MIOX orthologous gene pairs of each group.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterial.docx](#)
- [Additionalfile1TableS1.xlsx](#)
- [Additionalfile2FigureS1.pdf](#)
- [Additionalfile3TableS2.xlsx](#)
- [Additionalfile4FigureS2.pdf](#)