Environmental Sciences Europe

RESEARCH Open Access

GhNFYA16 was functionally observed positively responding to salt stress by genome-wide identification of *NFYA* gene family in cotton

Nan Xu[1†], Yupeng Cui[2†], Yuexin Zhang[1], Hong Zhang[1], Yapeng Fan[1], Xixian Feng[1], Hui Huang[1], Kesong Ni[1], Mingge Han[1], Xuke Lu[1], Xiugui Chen[1], Junjuan Wang[1], Delong Wang[1], Shuai Wang[1], Chao Chen[1], Lixue Guo[1], Lanjie Zhao[1] and Wuwei Ye[1*]

Abstract

Background: Nuclear transcription factor Y subunit A (*NFYA*) plays an important role in plant growth, development, and response to abiotic stress.

Results: This study systematically analyzed the *NFYA* gene family. Chromosome location analysis found that some *NFYA* genes in *Gossypium hirsutum* may have been lost during evolution. Collinearity analysis and selection pressure analysis indicated that the *GhNFYA* gene family underwent fragment duplication and whole genome duplication during evolution. At the same time, promoter *cis*-element analysis and gene interaction network analysis predicted that the expression of *GhNFYA* gene may be regulated by plant hormones and stress. To further explore the function of the gene, *Gossypium hirsutum* seedlings were treated with 4 °C, 37 °C, salt and PEG stress, respectively, found that the expression of *NFYA* is stimulated by multiple environments. By constructing a co-expression network, interactions between genes were found to defend against salt stress. Through virus-induced gene silencing experiments, it was found that plants that silenced the *GhNFYA16* gene were significantly more sensitive to salt stress.

Conclusions: This study found the relationship between the structure and function of *NFYA* gene family, provided a basis for the biological identification and functional verification of NFYA family members, and provided clues to clarify the specific roles of different types of NFYA proteins under different abiotic stress.

Keywords: *Gossypium hirsutum*, Nuclear transcription factor Y subunit A, Abiotic stress, Gene family, Virus-induced gene silencing

Introduction

Transcription factors, as known as trans-acting factors, have two modes of action: the first is to bind DNA, various *cis*-acting elements widely distributed in the promoter region of eukaryotic genes; The other is binding proteins, such as transcription factor proteins and other related proteins [1]. In this way, transcription factors can promote or inhibit the expression of downstream-related genes at the transcriptional level under specific conditions [2]. Plants grow under stress conditions, and

†Nan Xu and Yupeng Cui contributed equally to this work

*Correspondence: yew158@163.com

[1] Institute of Cotton Research of Chinese Academy of Agricultural Sciences/Zhengzhou Research Base, State Key Laboratory of Cotton Biology, School of Agricultural Sciences, Zhengzhou University, Anyang 455000, Henan, China
Full list of author information is available at the end of the article

stress signals will eventually stimulate the expression of transcription factors through a series of transmissions, and specifically regulate the transcription and expression of related response genes, respond to stress signals and adapt to the environment [3]. The activation and expression of transcription factor genes are regulated by environmental signals, such as high salt, drought and plant hormones, as well as plant self-development signals [4]. The activity of transcription factors when they function is influenced by post-translational modifications and their distribution in the cell. It is also affected by factors, such as the interaction between its own protein and other proteins [5].

Nuclear transcription factor Y (*NF-Y*) is a transcription factor widely existing in eukaryotes, which can specifically bind to CCAAT-box, and CCAAT-box, as a *cis*-element, exists in about 1/4 of eukaryotic gene promoters [6]. *NF-Y* is a trimer, which includes *NF-YA* (*CBF-B* or *HAP2*), *NF-YB* (*CBF-A* or *HAP3*), *NF-YC* (*CBF-C* or *HAP5*). In mammals, *NF-YB* and *NF-YC* are tightly bound together through a histone folding domain, and then combine with *NF-YA* in the nucleus to form the final trimer transcription factor [7], *NF-YA* can slide on the chromatin to look for the 25 bp CCAAT-box element. Once *NF-YA* is found, it will bind to it and insert it into the small groove of the DNA double helix, making the DNA double helix in a relaxed state, thereby improving favorably recruit RNA polymerase or other transcription factors, and then regulate downstream target genes positively or negatively [8].

As a conserved transcription factor, *NF-YA* plays a very important role in plant growth and abiotic stress [9]. Studies have found that *NF-YB* in *Zea mays* can increase the drought resistance of crops without affecting the yield reduction. Under drought conditions, compared with control plants, the yield of *ZmNF-YB2* overexpression *Zea mays* increases [10]; The drought resistance of *AtNF-YA5* transgenic plants is significantly higher than that of wild-type plants [9]; Overexpressed *Glycine max NF-YA3* gene reduces water evaporation from Arabidopsis leaves, thereby improving drought resistance. Meanwhile, *NF-YA3* responds to drought, high salinity, low temperature and ABA [11]. Under drought conditions, *BpNFYA5* can promote the accumulation of proline, enhance water retention capacity of the plant and improve the drought resistance of the *Brassica pekinensis*. In this process, *BpNFYA5* may also regulate some of the chlorophyll synthesis-related gene expression [2]; Overexpression of *GmNFYA3* can significantly improve the tolerance of *GmNFYA3* transgenic Arabidopsis to drought [11]. In addition, the *SiNF-YA6* gene can significantly improve the resistance of transgenic *SiNF-YA6* plants to low nitrogen by increasing the expression of nitrogen transport gene in *Setaria italica*. In *Medicago sativa*, the *NFYA1* gene is only expressed in a part of the root tip and regulates the development of metaphase root nodules. In conclusion, the functions of plant *NFYA* transcription factors are very complex, involving the flowering process, stress resistance and the regulation of growth and development of plants, which are very important in plants.

To further investigate the versatility of *NFYA* genes in plant development and defense response, the *NFYA* gene family was systematically investigated. This study provides potential candidate genes for functional studies of *Gossypium* genes and provides some molecular basis for *Gossypium* breeding. This may help to clarify the evolutionary mechanism of *NFYA* gene family in *Gossypium* and also provide us with further insights into stress response genes in *Gossypium*, and provide valuable information for cultivating stress-resistant *Gossypium*.

Materials and methods

Identification of NFYA family members

Four cotton genome files were downloaded from Cotton FGD: *G. hirsutum* (NAU version) [12] and *G. barbadense* (HAU version) [13], *G. arboreum* (CRI version) [14] and *G. raimondii* (JGI version) [15] (https://cottonfgd.org/) [16]. Through Pfam database analysis, it was found that the protein sequence conservative domain of *Arabidopsis* AT3G20910 was PF02045 (http://pfam.xfam.org/), and Hidden Markov Model (HMM) uses the local software HMMER to screen the gene with this conserved domain in four cotton genomes as the candidate gene of NFYA family genes. Used Phytozome [17] to search and download protein sequences of *Theobroma cacao* (*Theobroma cacao v1.1*), *Arabidopsis thaliana* (*Arabidopsis thaliana Araport11*), *Zea mays* (*Zea mays B73 RefGen_v3*), *Populus trichocarpa* (*Populus trichocarpa v3.0*), *Vitis vinifera* (*Vitis vinifera*) *v2.1, Oryza sativa* (*Oryza sativa v7.0*) and *Glycine max* (*Glycine max Wm82.a4.v1*). Gene sequences with incomplete domains were manually deleted using NCBI CD-Search (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). The physicochemical properties of the *GhNFYAs* genes were analyzed using the online website ExPASY (https://web.expasy.org/protparam/) [18]. TBtools software was used to draw *NFYA* chromosome location maps of 4 *Gossypium* species, *G. hirsutum*, *G. barbadense, G. arboretum* and *G. raimondii* [19].

Phylogeny analysis and sequence alignment

To study the evolutionary relationship of *NFYA* among different species, homologous genes of 11 species were obtained based on PF02045. The protein sequences of NFYA family members of 11 species were entered into MEGAX to construct an unrooted phylogenetic tree. The

maximum likelihood statistical method was used for the interspecific phylogenetic tree with a bootstrap duplication 1000, while the intraspecific phylogenetic tree used the neighbor-joining method with a parameter of bootstrap duplication 500 [20].

Collinearity analysis of *NFYA* genes in four *Gossypium* species

The *NFYA* genes of four *Gossypium* species were analyzed using TBtools software. Prepare gene ID, CDS sequence, chromosome length file and genome files and protein files for NFYA family genes. The MCScanX software [21] was used to collinearity analysis among the repetitive gene pairs of the four *Gossypium* species: *G. hirsutum*, *G. arboreum*, *G. raimondii* and *G. barbadense*, and TBtools software was used to visualize the results [22].

Calculation of duplicate gene pairs for selection pressure

The MEGAX comparison method was used to identify the selection pressure duplication gene pairs of *G. arboreum*, *G. hirsutum*, *G. raimondii*, and *G. barbadense*. The criteria for classification are as follows: the short sequence after alignment covers more than 80% of the long sequence, and the minimum homology of the alignment region is equal to or greater than 80% [23]. Selection pressure was investigated by calculated the non-synonymous to synonymous ratio (*Ka*/*Ks*) of duplicated genes.

Analysis of conservative protein motifs and gene structure

Multiple EM for Motif Elicitation (MEME) website (http://meme-suite.org/) was used to identify conserved protein motifs [24]. The MAST file of MEME website, the NEWICK (NWK) file of phylogenetic tree analysis and the GFF3 genome file of *G. hirsutum* were obtained, and then the gene structure of NFYA family members was analyzed by TBtools software [19].

Analysis of *GhNFYA* promoter region and differentially expressed genes

CottonFGD database (http://www.CottonFGD.org/) was used as a promoter to obtain a 2000 bp DNA sequence in the upstream region of *GhNFYA* [16]. The PlantCare website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to predict *cis*-regulatory elements in the promoter region of the *GhNFYA* gene. The *cis*-acting elements related to plant hormones, plant growth and development and abiotic stress were selected for further analysis. To study the expression patterns of the *GhNFYA* gene family, RNA-seq data (PRJNA490626) downloaded from the NCBI database of the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/) was used to analyze the expression levels of these genes under cold, heat, salt and PEG stress [25], the phylogenetic tree, *cis*-acting elements and expression caloric maps under different stress treatments were drawn by TBtools software [19].

Analysis of tissue specificity of NFYA family genes and salt stress expression in *G. hirsutum*

The gene expression data of the *G. hirsutum* family were obtained from the *Gossypium* Resources and Network Database (GRAND) established by the Cotton Research Institute[26], and the material used was the *G. hirsutum* material TM-1. All NFYA family genes were selected for tissue-specific analysis, and then the expression values of high-expressed genes in leaves were analyzed for 4 time periods (1 h, 3 h, 6 h and 12 h) under salt stress.

G. hirsutum cv H177 grown on sandy soil was cultivated in a constant temperature incubator with light at 28 °C for 16 h and darkness at 25 °C for 8 h. Cotton was treated with 100 mM NaCl stress when it reached the three leaf one heart stage.

Co-expression network analysis

Co-expression network analysis of *GhNFYA* genes under salt stress to determine the expression relationship between family members. Cytoscape software (version 3.8.0) was used to construct co-expression regulatory networks of genes [27]. Co-expression network plots were made with $p > 0.95$ as the threshold.

Subcellular localization and interaction network of NFYA in *G. hirsutum*

Using websites such as Wolf-PSORT [28] and ProtComp 9.0 [29] to predict the subcellular location of GhNFYA. To determine the subcellular localization of GhNFYA16, a GhNFYA16-GFP vector was constructed, and the recombinant GFP fusion protein was transiently overexpressed in tobacco leaves. The STRING (https://cn.string-db.org/) was used to predict the protein interaction of GhNFYA16 in *Arabidopsis* orthologous genes. According to the results, the interaction between GhNFYA16 and these genes in cotton can be inferred.

GhNFY's VIGS function verification

Total RNA was extracted using EASYspin Plus Plant RNA Kit [30]. RNA was reverse transcribed to cDNA using *TransScript*® All-in-One First-Strand cDNA Synthesis SuperMix for qPCR (One-Step gDNA Removal) (AT341). Real-time quantitative PCR detection was performed on an ABIPrism7500 Fast instrument, and the selected genes were amplified by a two-step method. The specific method refers to ChamQ Universal SYBR qPCR Master Mix (Q711, Vazyme Biotech Co., Ltd).

The 300 bp purified fragment was inserted into the cut pYL156 vector with restriction sites: *XmaI* and *XbaI*. The recombinant plasmid was transformed into *Agrobacterium* GV3101, and the *Agrobacterium* carrying pYL156 (empty), pYL156-*GhNFYA16*, pYL156-*CLA1* (positive control) and pYL192 (helper vector) were placed in the *Agrobacterium* containing 50 μg/mL kanamycin and 25 μg/mL rifampicin, cultured overnight at 28 °C, 200 rpm protected from light, shaken to OD600 = 1.2–1.5. Centrifuge at 6000 rpm for 15 min to collect the bacteria, and discard the supernatant. Adding an equal volume of sterile resuspension solution (10 mM MES, 200 μM AS, 10 mM $MgCl_2$, pH is about 5.8) to resuspend the bacteria, and let it stand for 4 h at room temperature in the dark. Choose *Gossypium* seedlings with just flattened cotyledons and water them before infection. Used a sterile needle to make a small hole in the epidermis of the cotyledon, and use a syringe to inject the bacterial solution into the epidermis of the cotyledon until the bacterial solution fills the entire cotyledon. After the injection, the *Gossypium* seedlings were protected from light for 24 h, and then cultured normally. The plants injected with pYL156 null control and pYL156-*GhNFYA16* were treated with drought stress after the albino phenotype appeared on pYL156-*CLA1 Gossypium* plants.

Results

Identification of NFYA family genes

A total of 94 family members were identified in the four cotton species. *G. hirsutum* mainly includes *GhNFYA1-30*, with 30 members. There are 32 members (*GbNFYA1-32*) in *G. barbadense*. There are 16 in *G. arboreum* (*GaNFYA1-16*). *G. raimondii* has 16 *GrNFYA1-16*. Other species were also classified according to their chromosomal location information (Additional file 1: Table S1). The amount of *NFYA* in tetraploid cotton (*G. hirsutum* and *G. barbadense*) was almost twice that in diploid cotton (*G. arboreum* and *G. raimondii*). It is proved again that the allotetraploid of cotton is formed by the hybridization of subgenome A and subgenome D to double chromosomes [15]. Analyzing the physicochemical properties of the genes of the NFYA family of *G. hirsutum*, the isoelectric point ranges from 6.84 (*GhNFYA11*) to 10.22 (*GhNFYA3*). The minimum number of amino acids is 175 (*GhNFYA3*) and the maximum is 999 amino acids (*GhNFYA11*). Molecular weights ranged from 19.40KDa (*GhNFYA3*) to 109.38KDa (*GhNFYA11*) (Additional file 1: Table S2).

Phylogenetic analysis of *NFYA* family genes

To study the evolutionary relationship of *NFYA* in plants, the protein sequences of 181 family members identified in 11 species were aligned and a phylogenetic tree was constructed. There are 30 genes in *G. hirsutum*, 32 in *G. barbadense*, 16 in *G. arboreum*, 16 genes in *G. raimondii*, 10 in *Arabidopsis thaliana*, 11 in *Oryza sativa*, 7 in *Theobroma cacao*, 18 in *Zea mays*, 13 in *Populus trichocarpa*, 21 in *Glycine max* and 7 *Vitis vinifera* (Additional file 1: Fig. S1). The identified NFYA family members were renamed according to their chromosomal location. Phylogenetic trees were constructed by maximum likelihood method using MEGAX. The unrooted phylogenetic tree was further beautified using the EvolView website (Fig. 1). Through the overall analysis of the phylogenetic tree, the NFYA family members are classified, divided into three classes A, B, and C, mainly by referring to the phylogenetic trees of *Arabidopsis thaliana* and four *Gossypium* species. Class A includes the *AtNFYA1*, *AtNFYA4*, *AtNFYA7* and *AtNFYA9* genes in *Arabidopsis thaliana* and the similar *NFYA* genes in cotton, *Populus trichocarpa*, *Glycine max*, *Zea mays* and other species. *NFYA* genes are classified according to maizeGDB (https://www.maizegdb.org/), TAIR (https://www.arabidopsis.org/index.jsp), JGI Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html) and other online sites. Class B includes the *AtNFYA3*, *AtNFYA5*, *AtNFYA6* and *AtNFYA8* genes of *Arabidopsis thaliana*, *Oryza sativa*, *Theobroma cacao*, and four *Gossypium* species with similar evolutionary relationships. Class C mainly includes *AtNFYA2*, *AtNFYA10* genes, *Glycine max*, *Zea mays*, *Oryza sativa* and a few *NFYA* genes of four *Gossypium* species.

As shown in Fig. 1A, the NFYA family members have two branches in class A. In each branch, the *NFYA* genes of the four *Gossypium* species have the closest evolutionary relationship. Meanwhile, the evolutionary relationship between cotton and *Theobroma cacao*, *Populus trichocarpa*, *Glycine max* and *Vitis vinifera* is also relatively similar. The *NFYA* genes in *Arabidopsis thaliana* and *Glycine max* are most closely related in evolution. Similar evolutionary relationships were also found in class B, and *NFYA* genes were more closely related in *Zea mays* and *Oryza sativa*. Combined with class C, it can be concluded that *Theobroma cacao*, *Populus trichocarpa* and *Glycine max* are the most closely related to cotton *NFYA* gene evolution. *Arabidopsis thaliana* is closest to *Glycine max* NFYA family genes. Maize and rice NFYA family genes have a relatively close evolutionary relationship. The number of NFYA family genes in *G. hirsutum* is more than twice that of *Arabidopsis thaliana*, *Oryza sativa* and *Theobroma cacao*, indicating that *Gossypium* has undergone significant gene amplification during its evolution [24]. The number of genes of NFYA family members in the four *Gossypium* species is 30 in *G. hirsutum*, 32 in *G. barbadense*, 16 in *G. arboreum*, and 16 family members in *G. raimondii*. During the evolution from

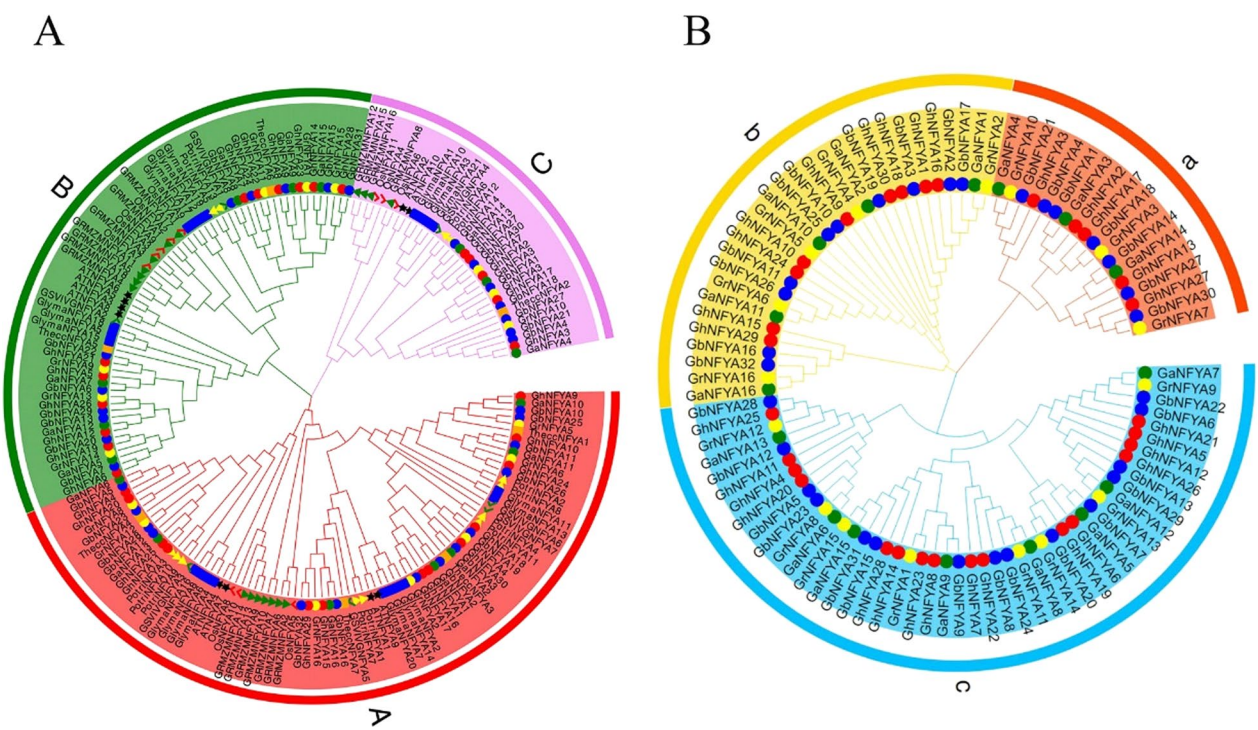

Fig. 1 Rootless phylogenetic tree constructed by the MEGAX maximum likelihood method, including the phylogenetic tree of *NFYA* genes of eleven species. **A** Construction of unroot phylogenetic tree of *NFYA* genes of eleven species; **B** constructed of the phylogenetic tree of four *Gossypium* species

diploid *Gossypium* to *G. hirsutum*, there are two genes are missing. By constructing the phylogenetic tree of NFYA family members of four cotton species, it is found that there are mainly three branches a, b and c (Fig. 1B).

Chromosome location analysis of NFYA family

To further analyze the characteristics of NFYA family genes, the chromosomes of NFYA family in four *Gossypium* species were mapped (Fig. 2). These genes were renamed according to their location on the chromosome. Both *G. hirsutum* and *G. barbadense* have two genes that are not located on chromosomes, and other genes of *NFYA* show uneven distribution on different chromosomes. The genes of *G. hirsutum* and *G. barbadense* have the same distribution on Chr03, Chr06, Chr07, Chr10, Chr11 and Chr13. One gene is distributed on Chr03 and Chr07. There are two genes distributed on Chr06 and Chr11 chromosomes. There are four genes located on Chr10 and Chr13 chromosomes. Meanwhile, the NFYA family genes are distributed in *G. hirsutum* with one more gene than *G. barbadense* GbAt-01 and GbDt-05. Upland cotton has one less gene than GbAt-02, GbDt-04, GbDt-08 and GbDt-09 on the corresponding chromosomes of sea island cotton. The results suggested that some *NFYA* genes of upland cotton may have been lost during evolution. Analyzing the distribution of *G. arboreum* chromosomes of this gene family, it is found that the distribution of the rest of the chromosomes tends to be the same on the A genome except for chromosome 2 which is more than that of the A genome of *G. hirsutum*. A large number of genes additions and deletions occurred in *G. raimondii*, and only the number of genes on chromosomes 3 and 13 is the same as that of *G. hirsutum* and *G. barbadense* D genome.

Collinearity analysis of NFYA family genes

Colinear analysis of *NFYA* genes in four cotton species was performed to understand the evolutionary relationship of *NFYA* gene family in cotton. The evolution of gene family generally goes through three processes, namely, fragment duplication, tandem duplication and whole genome duplication [31]. A joint analysis of the *NFYA* genes of *G. arboreum*, *G. raimondii*, *G. hirsutum* and *G. barbadense*, analyzed the gene duplication and collinearity between them. The *NFYA* genes of *G. arboreum* (Ga) and *G. raimondii* (Gr) were duplicated in *G. hirsutum* (Gh) and *G. barbadense* (Gb). It was shown that the two tetraploid genomes of upland cotton and sea island cotton were generated from the diploid genome in the process of genetic transformation. According to the

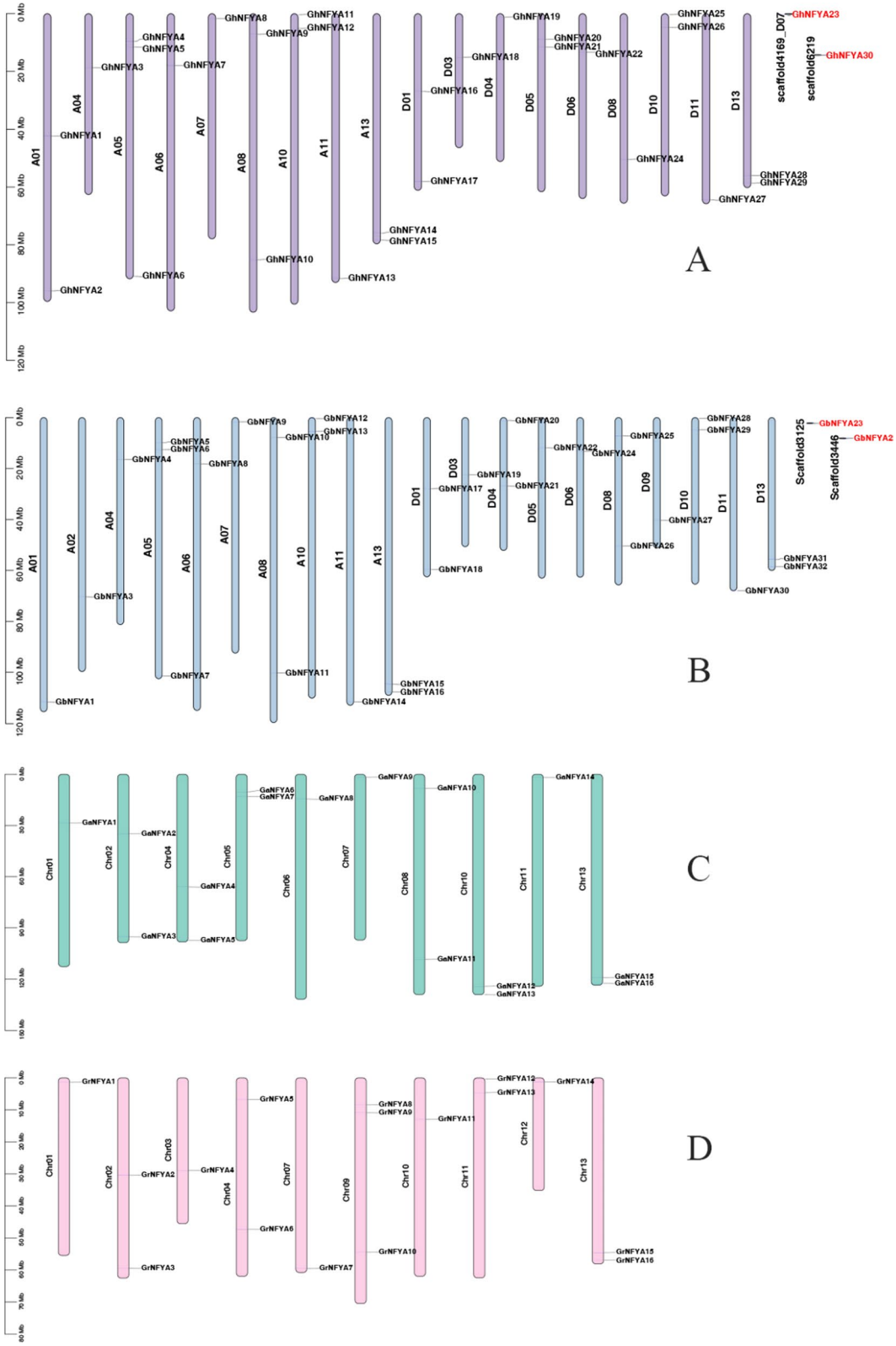

Fig. 2 Chromosome distribution of NFYA family genes in four *Gossypium* species, in order of **A** (*G. hirsutum*), **B** (*G. barbadense*), **C** (*G. arboretum*) and **D** (*G. raimondii*)

chromosomal distance, similarity and coverage of *NFYA* gene family members of diploid and tetraploid *Gossypium* species, tandem repeats and tandem repeats were identified, so as to identify the evolutionary relationship between NFYA family genes.

Genes linked together by collinear lines represent the same gene. In Fig. 3, it can be seen that many chromosomes in the GhAt/GhDt, GbAt/GbDt subgenomes and the GaA, GrD genomes are connected by lines of the same color. That is, the GhAt/GhDt and GbAt/GbDt subgenomes have *NFYA* homologous genes in the GaA and GrD genomes. It shows that these genomes/subgenomes are closely related in evolution, and most *NFYA* genes have been preserved in the evolution of polyploidy. Genes located in the same chromosome region (e value < 1e-5) are classified as tandem repeats, while the rest of the genes from the same genome are considered to be fragment repeats. In all collinearity analysis results, they come from 4 different *Gossypium* species. Genome and subgenome genes are usually classified as genome-wide duplication. Homologous/like homologous gene pairs were identified for the GhAt/GhDt and GbAt/GbDt subgenomes of two tetraploid *Gossypium* species. After homology analysis, it was found that several gene loci were highly conserved between the At and Dt subgenomes of the two tetraploid *Gossypium* species. As mentioned above, *G. hirsutum* and *G. barbadense* are derived from the ancestors of diploid *G. arboreum* and *G. raimondii* [32]. No tandem duplications were found in the *NFYA* gene family by analysis, with 98 fragment duplications and 266 genome-wide duplications. Based on these results, it is speculated that closely related gene pairs are usually generated through genome-wide duplication and fragment duplication, especially fragment duplication is

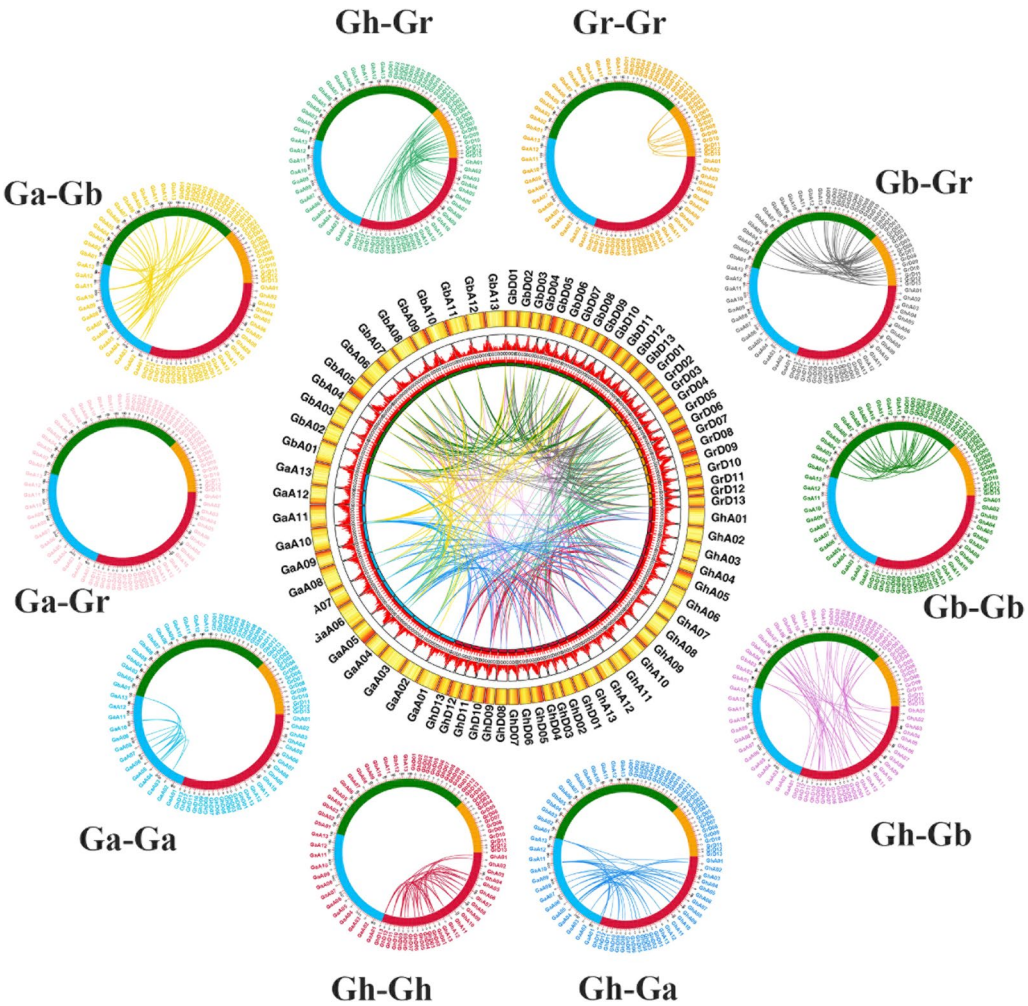

Fig. 3 Collinear relationship of repeated gene pairs in four *Gossypium* species of the NFYA family. Chromosomal lines represented by various colors indicates the syntenic regions around the *NFYA* genes

the most important factor in the evolutionary process (Fig. 3).

Ka/Ks analysis of selective pressure of NFYA family genes

The evolution of *NFYA* gene pairs in four cotton species was obtained by selection pressure analysis. It can be judged whether there is selective pressure acting on NFYA family genes. In the process of evolution, the duplicated gene pair may also diverge from its original function, which eventually leads to non-functionalization (loss of original function), sub-functionalization (division of original function), and new functionalization (acquisition of new function) [33]. To determine the nature and degree of selection pressure of repeated gene pairs and explore whether Darwin's positive selection is related to the divergence of repeated *NFYA* genes, Non-synonymous (Ka) and Synonymous (Ks) values were calculated for 314 duplicated gene pairs from 10 combinations of 4 cotton species. These combinations include *G. hirsutum* VS *G. hirsutum* (*Gh-Gh*), *G. hirsutum* VS *G. barbadense* (*Gh–Gb*), *G. hirsutum* VS *G. arboreum* (*Gh–Ga*), *G. hirsutum* VS *G. raimondii* (*Gh–Gr*), *G. barbadense* VS *G. barbadense* (*Gb–Gb*), *G. barbadense* VS *G. arboreum* (*Gb–Ga*), *G. barbadense* VS *G. raimondii* (*Gb–Gr*), *G. arboreum* VS *G. arboreum* (*Ga–Ga*), *G. arboreum* VS *G. raimondii* (*Ga–Gr*) and *G. raimondii* VS *G. raimondii* (*Gr–Gr*). According to the ratio of *Ka/Ks*, the selection pressure of duplicated gene pairs can be inferred. It is generally believed that $Ka/Ks = 1$ means neutral selection (pseudogene), $Ka/Ks < 1$ means negative purifying selection, and $Ka/Ks > 1$ means positive selection effect [34].

There are 314 duplicate gene pairs in the NFYA family genes in four cotton species. Among them, there are 13 gene pairs with positive selection effect, and 301 gene pairs with purifying selection. It indicates that the NFYA family genes are relatively conserved in the evolutionary process. Positive selection gene pairs appeared in Ga–Gr, Ga–Gb, and Ga–Gh were 2 pairs, 3 pairs and 2 pairs, respectively. It indicated that some *NFYA* genes had beneficial mutations in the process of hybridization of diploid cotton into allotetraploid. Likewise, there are also 1, 2 and 3 gene pairs in Gb–Gb, Gb–Gr, and Gh–Gr, respectively. There were 13 gene pairs that had beneficial mutations during evolution. There are 12 and 5 gene pairs with *Ka*/*Ks* values ranging from 0.49 to 0 in the Ga–Ga, Gr–Gr repeat gene pairs, respectively. Indicates that they were selected for complete purification (100%). Similarly, the number of *Gb–Gh*, *Gh–Gh* repeated gene pairs with *Ka/Ks* values between 0.99 and 0.5 are 8 and 1, respectively; the number of repeated gene pairs with *Ka*/*Ks* between 0.49 and 0 are 27 and 34.

In conclusion, 314 pairs of duplicate genes from four *Gossypium* species (*Gh*, *Gb*, *Ga* and *Gr*) were found in selection pressure analysis. Among them, 301 pairs (95.86%) of repeated gene pairs have *Ka/Ks* values less than 1, including 261 pairs of genes with *Ka/Ks* values less than 0.5 and 40 pairs of genes with a *Ka/Ks* value between 0.5 and 0.99, showing purification selection. Only 13 pairs (4.14%) of repetitive gene pairs have a *Ka/Ks* value greater than 1. These gene pairs may have undergone rapid evolution after repetition and have experienced positive selection pressure. Since most of the *Ka/Ks* values were less than 1.0, it was speculated that the *Gossypium* NFYA family genes has underwent strong purifying selection pressure and limited functional differentiation after fragment duplication and genome-wide duplication (Additional file 1: Fig. S2).

Analysis of motif and gene structure of NFYA family

Through the joint analysis of the phylogenetic tree, gene structure and motifs of the *NFYA* gene family, the characteristics of NFYA family members and their relationships were further understood. The phylogenetic tree of four cotton species was constructed using MAGAX software. Combined with the motif files obtained from the MEME website, the TBtools software was used to display the structure and taxonomic information of the four cotton species NFYA family (Fig. 4).

There are 10 motifs in *NFYA* genes in four cotton species. According to the phylogenetic tree and motif type, the four cotton species *NFYA* gene families were divided into three groups: I, II, III. The motifs of each class tend to be consistent and have obvious structural features. Class I includes all motif structures, and they are arranged in the order of motif8, motif9, motif5, motif6, motif4, motif2, motif10, motif1, motif3 and motif7. However, most of class II genes lack motif8, and few genes lack motif9. Compared with the class I *NFYA* genes, it is possible that a certain function will be lost. Class III contains fewer motif structures, lacking motif8, motif9 and motif6. In general, NFYA family genes contain motif5, motif4, motif2, motif10, motif1, and motif3, indicating that motif largely determine the similarity of family gene function and structure. From the point of view of gene structure introns and exons, all genes contain exons and introns. Meanwhile, the gene structures of class I, class II and class III have their own consistent characteristics. The class I of the *NFYA* gene introns and exons are compact and uniform, and the length of the exons are shorter. The exons of *NFYA* gene in class II were scattered. The exons of *GhNFYA11*, *GaNFYA13* and *GhNFYA25* genes are more dispersed, including a longer intron and exon. For the genes of class III, the exons are relatively short and only some of them contain longer exons, but the whole is consistent. In conclusion, NFYA family members have unique characteristics and obvious structural

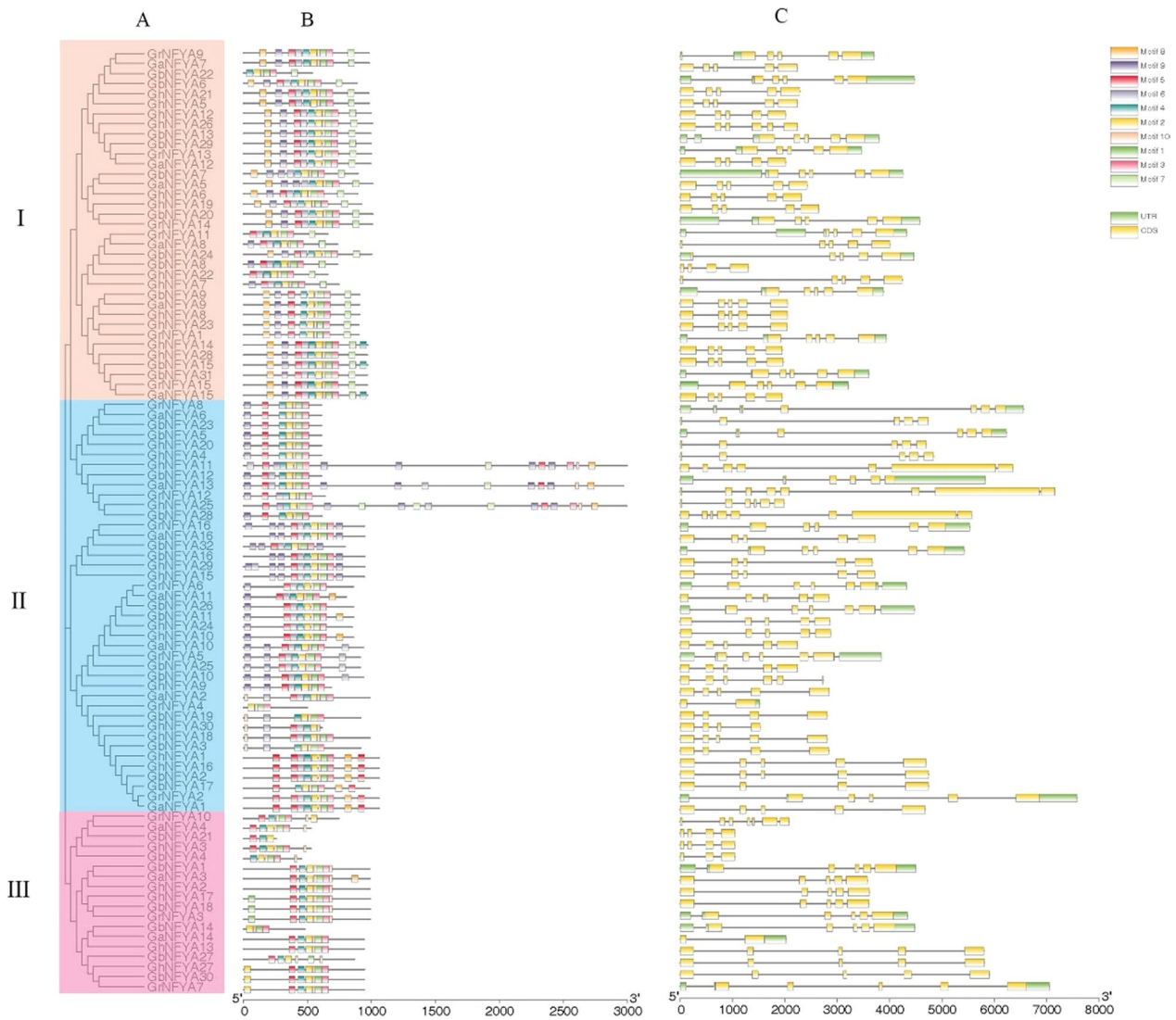

Fig. 4 **A** *GhNFYAs* gene phylogeny, gene structure and motif composition. A phylogenetic tree was constructed using MEGAX and NJ (Neighbor-Joining, NJ) methods, with a number of 1000 guided replications. **B** Schematic diagram of the conserved motifs in the *GhNFYAs* protein clarified by the TBtools software. Each colored box represents a motif in a protein, and its motif name is displayed in the box in the upper right corner. The length of the protein and the motif can be estimated using the ratio of the bottom. **C** Exon/intron organization of *GhNFYAs* gene. The green boxes represent exons, and the black lines represent introns. The size of exons and introns can be estimated using the ratio of the bottom

differences, which are relatively conservative in the process of evolution.

Analysis of differentially expressed genes of NFYA family in *G. hirsutum*

The members of the NFYA family play important roles in various important physiological and biochemical processes of plants [35]. In addition, *NFYA* is also involved in the response to various environmental stimuli [36]. To determine the function of *GhNFYA* gene in different environments, upland cotton was subjected to low temperature, high temperature, high salt and PEG stress. The expression level of *GhNFYA* gene during growth and development and its response to phytohormones were analyzed (Fig. 5). The *cis*-acting element is located in the promoter region of the gene and can be used as a reference for tissue specificity and stress response in different environments. The *cis*-acting elements of the *NFYA* gene family mainly include the *cis*-acting regulatory elements involved in the methyl jasmonate (MeJA) response, the *cis*-acting regulatory elements necessary for anaerobic induction, the MYB binding site involved in drought

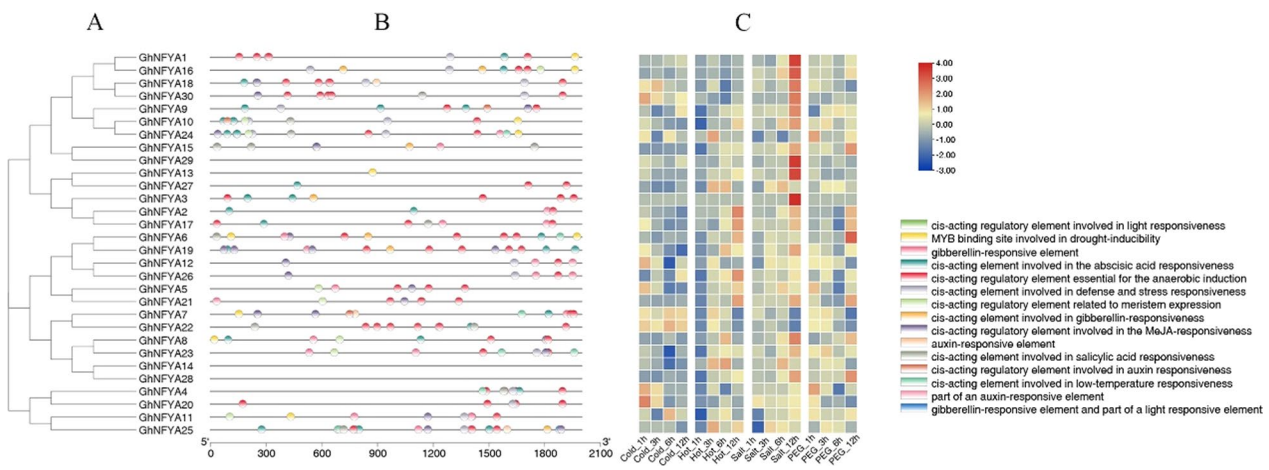

Fig. 5 Analysis of promoters and differentially expressed genes of *GhNFYA* family. **A** Phylogenetic tree of *GhNFYA* genes. **B** *Cis*-elements in promoters of *GhNFYA* genes. **C** Differentially expressed genes of *GhNFYA* genes under cold, hot, salt and PEG stress

induction, the *cis*-regulatory elements involved in meristem expression, *cis*-acting elements involved in low temperature response, *cis*-acting elements involved in defense and stress response, *cis*-acting elements involved in stress response, and phytohormone-related regulatory elements (salicylic acid, auxin, gibberellin and abscisic acid, etc.) (Additional file 2: Table S3). The number of *cis*-acting elements varied among genes, for example, *GhNFYA16* contained a *cis*-acting element for abscisic acid, a MYB binding site involved in drought induced, anaerobically induced action element, a *cis*-regulatory element involved in meristem expression, and a *cis*-acting element involved in stress response. In general, the NFYA family of *G. hirsutum* mainly contains *cis*-acting elements related to plant hormones, growth and development and adversity. It can be inferred that this gene family is related to adversity to a certain extent.

Gene expression patterns can provide an important reference for gene function analysis. It is related to the biological functions controlled by *cis*-acting elements. To explore the expression patterns of *GhNFYA* in *G. hirsutum* under different stress environments, the gene expression levels of cotton under four abiotic stress conditions of salt, cold, heat and PEG (1 h, 3 h, 6 h and 12 h) were analyzed [37]. The results showed that the genes of the NFYA family had different degrees of response to cold, heat, salt and PEG. It can be seen that the expression level was higher at 12 h of salt treatment, and the expression level of cold and heat stress also changed in a trend. Combining salt and PEG stress, it can be seen that *GhNFYA16* is differentially expressed under salt and PEG stress treatments, and the expression patterns of each gene are slightly different under stress treatments. These results further prove that *GhNFYAs* participate in the stress response of plants. Overall, it concludes that *NFYA* gene family has been influenced to more evolutionary events and extended. Moreover, some point mutations in exon regions and regulatory region of new family members might affect the function and expression of new family members [38, 39].

Tissue specificity of NFYA family genes in *G. hirsutum* and analysis of differentially expressed genes under salt stress

The tissue-specific presentation of 30 genes in the GhNFYA family showed that there were certain differences in the expression of 30 genes among different tissues (Fig. 6A). The expression levels of *GhNFYA5*, *GhNFYA20*, *GhNFYA23* and *GhNFYA28* genes were the highest in roots. 11 genes had the highest expression levels in the stem: *GhNFYA2*, *GhNFYA6*, *GhNFYA7*, *GhNFYA12*, *GhNFYA13*, *GhNFYA14*, *GhNFYA15*, *GhNFYA17*, *GhNFYA19*, *GhNFYA22* and *GhNFYA27*. The relative expression of 12 genes *GhNFYA1*, *GhNFYA8*, *GhNFYA9*, *GhNFYA10*, *GhNFYA11*, *GhNFYA16*, *GhNFYA18*, *GhNFYA21*, *GhNFYA24*, *GhNFYA26*, *GhNFYA29* and *GhNFYA30* were the highest in leaves, and the expression of *GhNFYA4* in roots and stems were higher than that in leaves. The expression levels of *GhNFYA3* and *GhNFYA25* did not differ among different tissues.

The most intuitive phenotypic changes occur in leaves when subjected to abiotic stress. Therefore, 12 highly expressed gene in leaves were selected, and the expression values were analyzed in periods (1, 3, 6 and 12 h) of NaCl stress (Fig. 6B, Additional file 3: Table S4), which provided support for subsequent virus-induced gene silencing experiments. The expression values of *GhNFYA1*, *GhNFYA18*, *GhNFYA29* and *GhNFYA30* were the

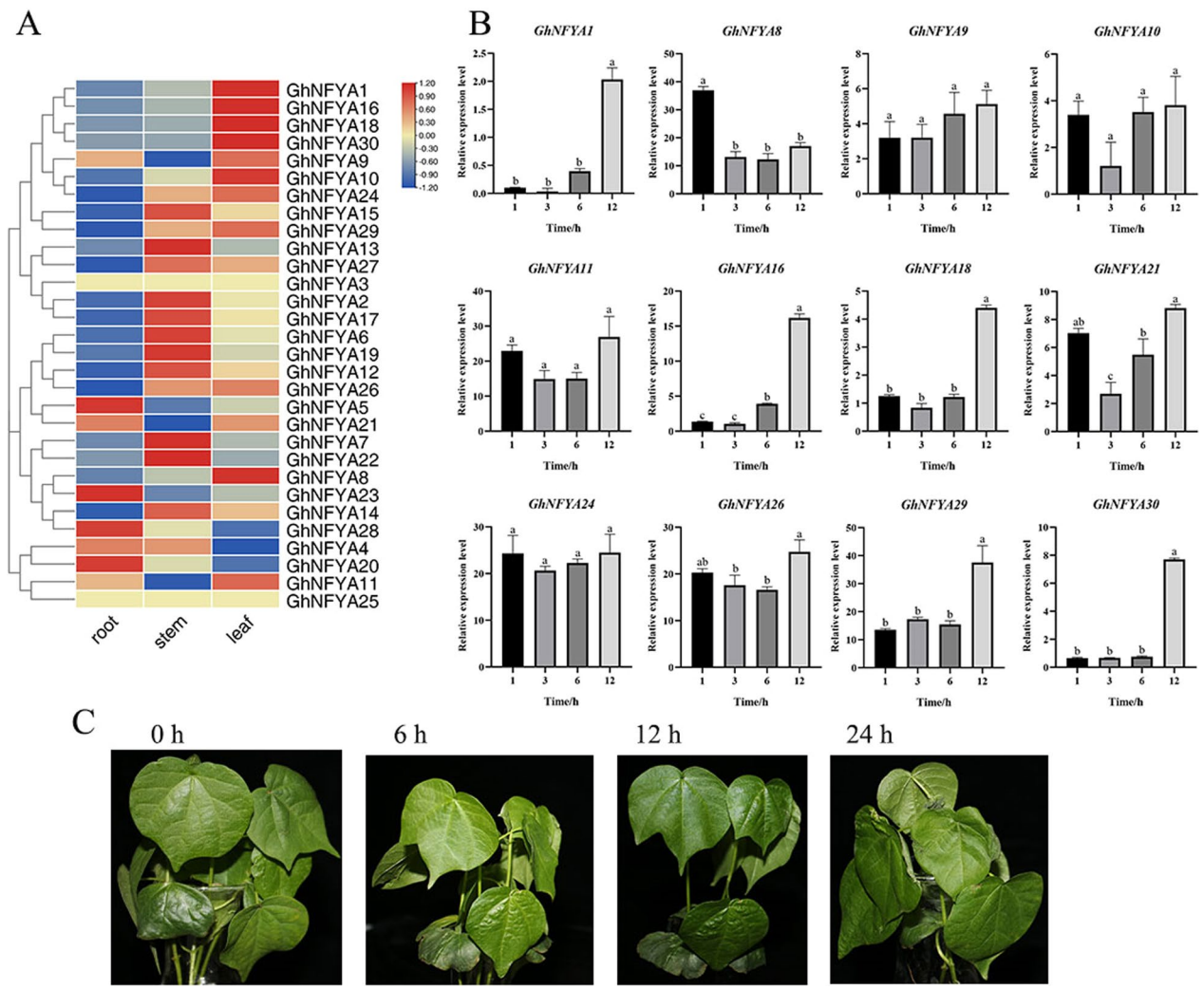

Fig. 6 **A** Specific expression of family genes in roots, stems and in leaves. **B** Select 12 genes specifically expressed in leaves, analyzed the differential expression under NaCl stress treatment in 4 time periods. Significance Level $a = 0.05$. **C** Phenotypic changes of *G. hirsutum* under salt stress at different time periods (0, 6, 12 and 24 h)

highest when treated with NaCl for 12 h. The expression of *GhNFYA8* decreased significantly after NaCl treatment for 3, 6 and 12 h compared with that after treatment for 1 h. The expression of *GhNFYA16* increased significantly after 6 and 12 h of NaCl treatment. The expression of *GhNFYA21* decreased significantly at 3 h compared with 1 h after NaCl treatment, and increased again at 6 h and 12 h. *GhNFYA26* was significantly decreased after NaCl treatment for 3 h and 6 h, while its expression level was significantly increased after NaCl treatment for 12 h. Obviously, some *GhNFYA* genes showed significant differential expression after NaCl treatment.

G. hirsutum was treated with 100 mM NaCl stress when it grew to the three leaf one heart stage. It was found that the cotyledons began to wilt and lose their luster after 6 h of treatment, and the wilting was more serious after 12 h of treatment. After 24 h of treatment, part of the cotyledons fell off, true leaves wilted, leaf edges curled, and new leaves wilted to death (Fig. 6C).

Co-expression network analysis of *GhNFYA* genes under salt stress

To further understand the role of *GhNFYA* genes in salt stress, the correlation network of family members based on Pearson correlation coefficients (PCCs) was analyzed [40]. Expression network analysis of genes under salt stress showed positive or negative correlations. A total of 142 gene pairs were positively correlated and 137 gene pairs were negatively correlated under stress (Fig. 7). Except for *GhNFYA1*, *GhNFYA11*, *GhNFYA24*

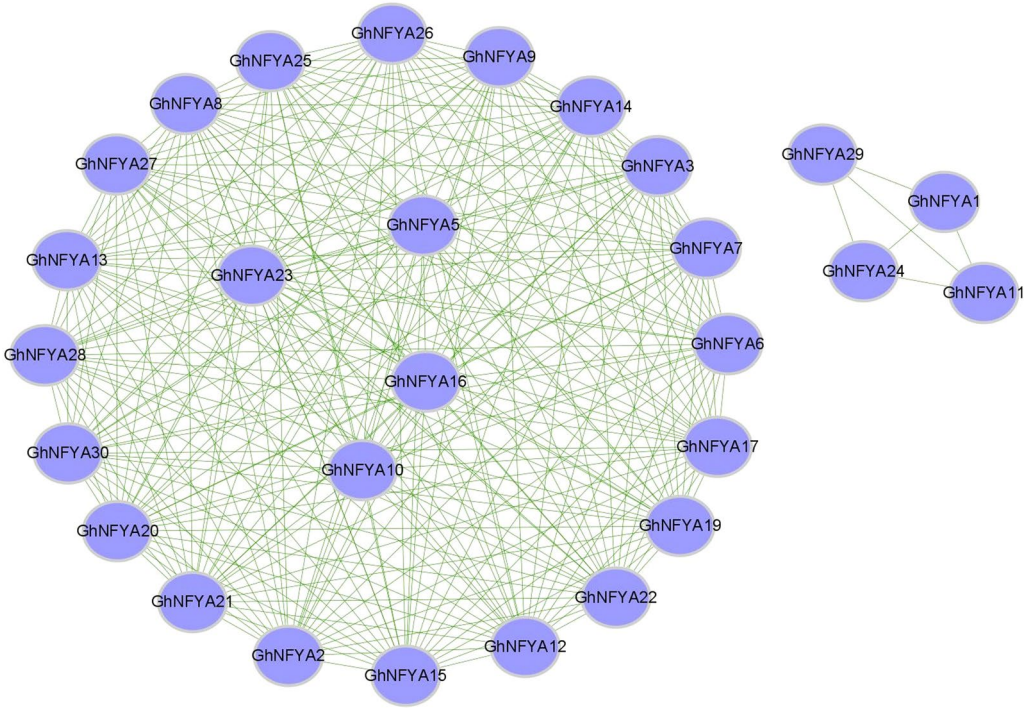

Fig. 7 Co-expression network of *GhNFYA* genes under salt stress treatment is based on Pearson's correlation coefficient (PCC) for these gene pairs was analyzed using RNA-Seq data. Lines indicate significance levels for co-expressed gene pairs. Co-expression network plots were made with $p > 0.95$ as the threshold

and *GhNFYA29*, other genes showed complex and highly similar functional relationships. 273 gene pairs that interact with each other during salt stress are involved in resilience. In conclusion, the expression network studies showed that *GhNFYAs* genes were closely related to each other in salt stress.

Subcellular localization of GhNFYA

The Programs website predicts that GhNFYA is most likely to be located in the nucleus and cytoplasm in *G. hirsutum*. According to the prediction results of the WoLF–PSORT website, GhNFYA is mainly located in the nucleus, cytoplasm, vacuole and chloroplast. For example, on the WoLF–PSORT website, the GhNFYA16 predicts subcellular localization in the nucleus and chloroplasts, while using the Programs website to predict that GhNFYA16 is localized in the nucleus. In general, subcellular localization verification focuses on the nucleus. Most of the GhNFYA are located in the nucleus, which may be related to their role as transcription factors that combine with NF–YB and NF–YC in the nucleus to form a trimer to regulate downstream target genes.

Combined with the differential expression of *GhNFYAs* genes under various stresses and the real-time expression of genes in 4 time periods under NaCl treatment, GhNFYA16 was selected for subcellular localization verification. The expression vector of GhNFYA16-GFP fusion protein was constructed. The recombinant plasmid containing the expression vector was injected into the epidermis of tobacco. 3 days later, the observe under a focusing microscope. The results showed that the green fluorescence signal of the fusion protein showed that GhNFYA16 was located in the nucleus (Fig. 8).

Analysis of GhNFYA protein interaction network

GhNFYA16 is an orthologous gene of *Arabidopsis thaliana NFYA1*. In this protein interaction network, Arabidopsis NF-YA1 interacts with NF-YB1, NF-YB2, NF-YB3, NF-YB6, NF-YC1, NF-YC2, NF-YC3, NF-YC4, NF-YC9 and NF-YC12. It is speculated that GhNFYA16 is closely related to the corresponding protein in cotton. Many functions were enriched in this protein network, such as positive regulation of photomorphogenesis, abscisic acid-activated signaling pathway, positive regulation of nitrogen compound metabolic process and regulation of developmental process and other Go enrichment. The functional diversification of NF-Y evolution in plants enables plants to actively respond to different abiotic stresses (Fig. 9).

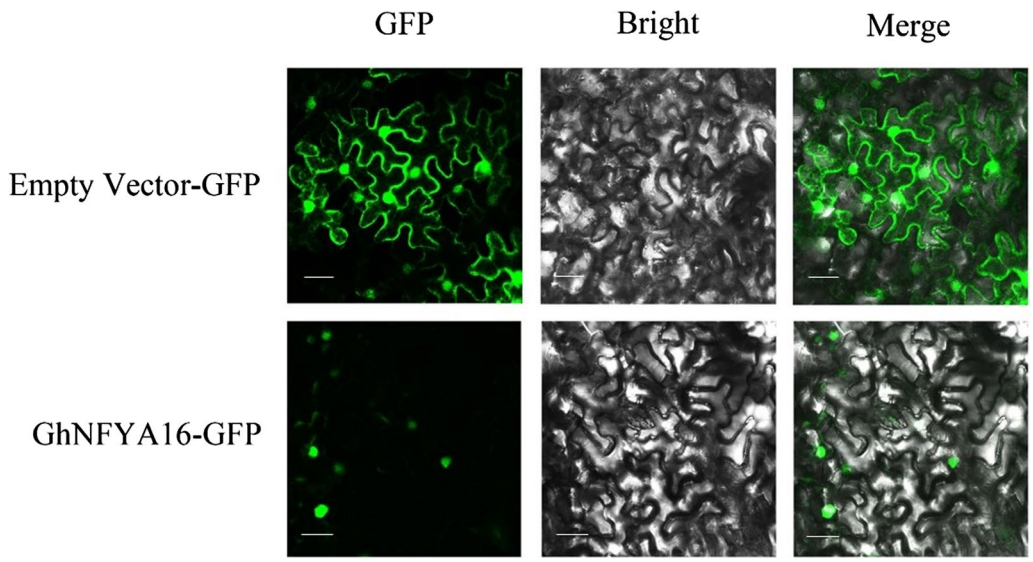

Fig. 8 Subcellular localization of GhNFYA16 (Bar = 75 μm)

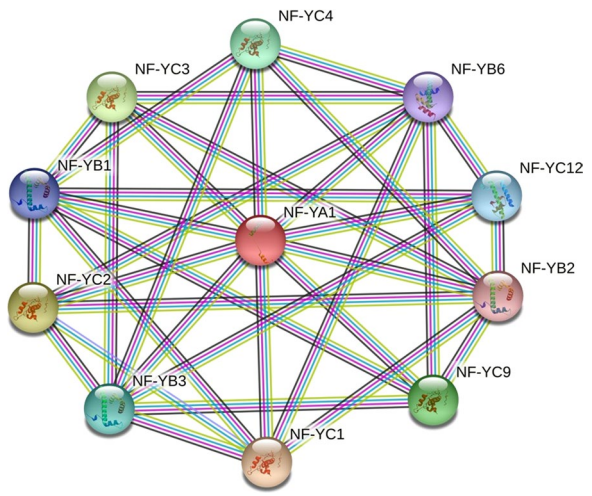

Fig. 9 Protein interaction network of GhNFYA16 homologous genes in *Arabidopsis* (NFYA1)

Silencing *GhNFYA16* reduced tolerance to salt stress in cotton

According to the analysis of the differentially expressed genes of the GhNFYA family, a highly expressed gene *GhD01G1179.1* (*GhNFYA16*) was screened out after NaCl stress for 6–12 h. To further study the function of the *GhNFYA16* gene, a VIGS experiment was performed on *GhNFYA16* with *Gossypium cv* H177 as the material. 2 weeks later, the *Gossypium* with pYL156: PDS showed albino phenomenon, indicating that VIGS silence was successful. The silencing effect of *GhNFYA16* gene was examined by quantitative real-time PCR (qRT-PCR). The results showed that the expression of pYL156: *GhNFYA16* was significantly lower than that of pYL156 in the control group. After salt treatment, the conductivity rate and chlorophyll content of the plants were measured. It was found that after silencing the *GhNFYA16* gene, the conductivity rate increased and the chlorophyll content decreased compared with the control group. Therefore, it can be inferred that *GhNFYA16* is involved in the adaptability of cotton to salt stress (Fig. 10).

Discussion

As an important economic crop, *G. hirsutum* is widely cultivated around the world and faces severe biotic and abiotic stress. The *NF-Y* transcription factor is closely related to adversity. Some studies have shown that *NF-YA* transcription factor participates in drought resistance, high salinity resistance, nitrogen–starvation responses, symbiotic nodule development, regulation of flowering time and embryogenesis [41]. *CmNF-Yb8* in chrysanthemum regulates the expression of *CmCIPK6* and *CmSHN3*, alters stomatal movement and cuticle thickness in leaf epidermis, thereby affecting plant drought resistance [42]. *TaNFYA-B1* plays a vital role in the development of wheat root system and the use of nitrogen and phosphorus [43]. Therefore, genome-wide identification of NFYA family genes was performed, focusing on the analyzing the relationship between NFYA family genes and their expression patterns under abiotic stress. This study lays a foundation for further exploring the role of *NFYA* transcription factors on *Gossypium* stress-related effects.

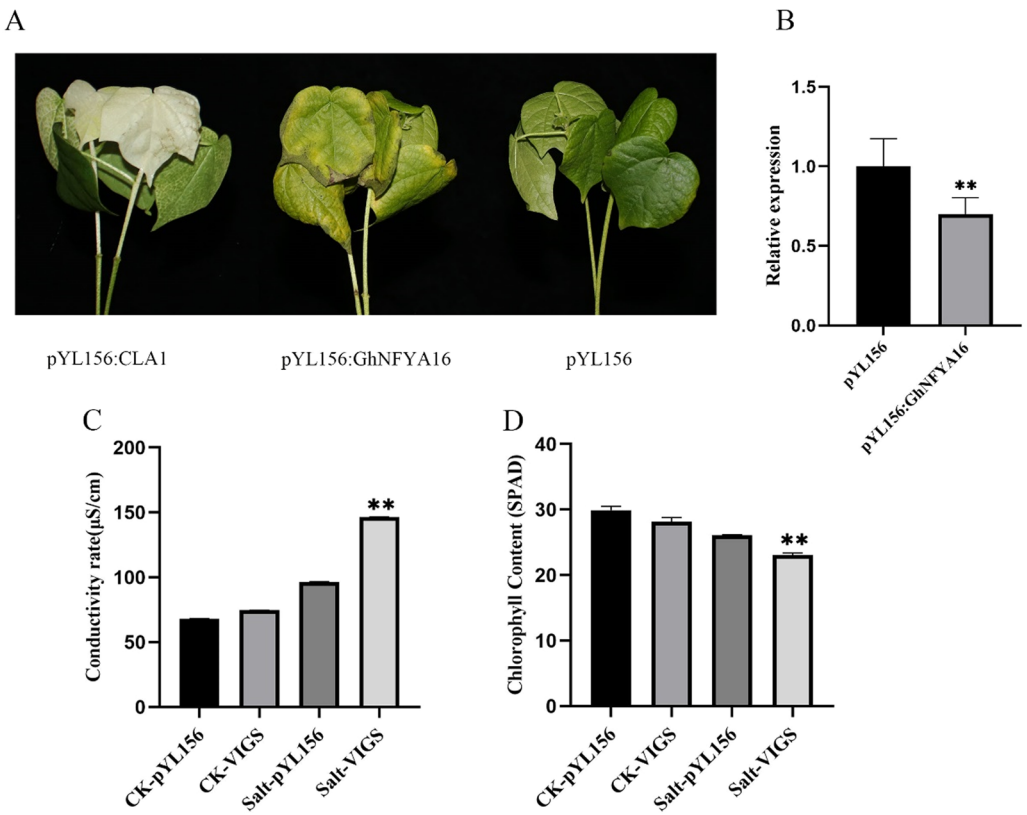

Fig. 10 Phenotype of *Gossypium* leaves after virus infection and expression analysis of *GhNFYA16* under NaCl stress. **A** Phenotype of *Gossypium* leaves after virus infection. **B** qRT-PCR for *GhNFYA16* under NaCl stress. **C** Determination of conductivity rate after virus infection. **D** Determination of chlorophyll content after virus infection. **: $p < 0.01$; the resulting values are expressed in relative units. The error bar in the figure is the standard deviation (SD) of the three biological replicates in each treatment group

Through the CottonFGD and JGI Phytozome websites, the genome sequences of the NFYA family in 11 species According to the number of members, it was again proved that the number of *NFYA* genes in tetraploid *Gossypium* was almost double that in diploid *Gossypium*. The analysis shows that the closest evolutionary relationship with the *Gossypium* NFYA family is *Theobroma cacao*, *Glycine max* and *Arabidopsis thaliana*. It can be seen that soybean and *Arabidopsis* can be used as a reference for the study of *NF-Y* transcription factors in cotton. At present, there are many studies on soybean nitrogen fixation and drought resistance. The information of Arabidopsis *NF-Y* is more detailed, which can be used as a good reference for the study of *NF-Y* in cotton. Among them, the *NFYA* genes in *Gossypium* and *Theobroma cacao* are closely related throughout the phylogenetic tree, which further validates the previous report that *Theobroma cacao* and *Gossypium* share a common ancestor [44].

After chromosomal mapping of the *NFYA* gene family in cotton, it was found that the gene of this family in *G. hirsutum* was lost or added on some chromosomes compared with sea island cotton. In the process of identifying the *NFYA* gene family, it was found that *G. hirsutum* had two less genes than *G. barbadense*, and the two missing genes were probably located on the chromosomes of GhAt-02, GhDt-04, GhDt-08 and GhDt-09. Meanwhile, the random and uneven distribution of genes further illustrates that gene loss may occur during evolution, or it may be the result of incomplete genome assembly [24].Gene duplication produces functional differences and is considered to be the most important factor in speciation and environmental adaptability [45]. For gene duplication, the aligned sequence should include at least 70% of the same part, and should cover more than 80% of their total length [46]. Two duplicate genes located on the same chromosome are probably due to tandem duplication, while the existence of the same subgenome on different chromosomes is mainly due to fragment duplication [47]. During evolution, apart from small-scale tandem duplications, most segment duplications contribute to the generation of new genes, thereby contributing to the complexity of plant genomes [48]. Our research has found 98 pairs of duplicated genes. Compared with the previous genes, the newly duplicated

genes are functionally redundant, and this redundancy is considered to be the basic driving force for evolutionary innovation [49]. During the amplification of family genes, the discovery of chromosomal interactions, polyploidy, evolutionary differences and the transfer of inheritance between genomes provides valuable information. Based on the above analysis, it is believed that the *GhNFYA* gene family has undergone fragment duplication and genome-wide duplication during the evolution process, which ultimately contributed to the expansion of the NFYA family genes.

To explore the effect of divergence after NFYA gene duplication and the extent of selective pressure on duplicating genes. By counting 314 duplicate gene pairs in four cotton species, 301 pairs (95.86%) of them were found to have a *Ka*/*Ks* ratio less than 1, indicating homozygous selection [50]. Only 13 pairs (4.14%) of duplicate pairs were generated by positive selection, which indicates that they may have undergone relatively rapid evolution after duplication. It is speculated that the *Gossypium NFYA* gene family has undergone strong purification selection. In addition, after whole genome duplication, functional differentiation is limited [51]. The functional differentiation of repetitive gene pairs is the source of plants generating new genes, injecting new impetus into the evolution of plant genomes. Changes in gene expression patterns are another important factor leading to functional differentiation [52].

Using the obtained NWK files and motif files, the gene structure analysis of NFYA family in four cotton species was carried out. It can be seen that the motif types contained in each class are almost identical, but individual genes are structurally distinct from other *NFYA* genes. Fragment duplication and whole genome duplication may lead to changes in gene structure. Importantly, motifs 5, 4, 2, 10, 1, and 3 are critical determinants of NFYA family gene function. During evolution, other motifs have been gradually added to make genes play different roles in plants.

The NFYA family genes include plant hormones, such as auxin, abscisic acid, gibberellin and other *cis*-acting elements related to stress response. The *cis*-acting elements contained in each gene vary. Some genes were found to have fewer *cis*-acting elements than others. Combined with gene structure analysis, it was found that *GhNFYA13* and *GhNFYA27* had fewer *cis*-acting elements and longer introns. It further verified that the gene structure determines the function of the gene to a certain extent. By comparing the *cis*-acting elements of the *GhNFYAs* gene with the gene differential expression heatmap under cold, heat, salt and PEG stresses, it was more verified that this gene family played an important role in adversity. For example, the *GhNFYA16* gene was significantly differentially expressed under both salt stress and PEG treatment. Through the analysis of tissue specificity and differentially expressed genes under NaCl stress, the role of NFYA family genes in stress was further verified. By constructing the co-expression network of *GhNFYAs* genes under salt stress, it was found that most of the genes play a positive or negative regulatory role, and the interactions between the genes jointly resist salt stress. As previously reported, GhNFYA10 and GhNFYA23 positively regulate salt stress [53]. VIGS experiments were performed on the *GhNFYA16* gene using *cv* H177 in *G. hirsutum*. Silenced plants were subjected to salt stress treatment. The study found that after silencing the *GhNFYA16* gene, the plants had obvious phenotypic changes. In addition, plants are more sensitive to salt stress. This study has provided potential candidate genes for the study of *Gossypium* gene function, and provided a certain molecular basis for *Gossypium* breeding. It helps us to deeply understand the biological and molecular functions of *NFYA* genes in *Gossypium*, as well as their antioxidant effects under various oxidative stresses.

Conclusions

In this study, for the first time, a comprehensive analysis of four *Gossypium* species *NFYA* transcription factors was performed, and a total of 30 *G. hirsutum NFYA* genes were identified. The *NFYA* gene is divided into three categories through the phylogenetic tree, and it is found that the *Gossypium NFYA* gene has a close evolutionary relationship with *Theobroma cacao* and *Populus tomentosa*. Chromosome location and gene duplication analysis showed that this gene family was amplified in *Gossypium* through fragment duplication and whole genome duplication. The analysis of gene conserved domains, promoters and differentially expressed genes further identified the role of the NFYA family in *Gossypium*, and found that the *GhNFYA* gene played a role in a variety of plant hormones and environmental stimuli. To further verify the role of NFYA family genes in *Gossypium*, tissue-specific and differentially expressed genes were analyzed under NaCl stress, and VIGS experiments were performed on the *GhNFYA16* gene, followed by salt stress after silencing. The results showed that *GhNFYA16* played a role in *Gossypium* under salt stress. The data provided in this study will further provide useful information for studying the function of *Gossypium NFYA* transcription factors.

Abbreviations

Ga: *Gossypium arboreum*; *Gb*: *Gossypium barbadense*; *Gh*: *Gossypium hirsutum*; *Gr*: *Gossypium raimondii*; HMM: Hidden Markov Model; NF-Y: Nuclear factor-Y; qRT-PCR: Quantitative real-time PCR; VIGS: Virus-induced gene silencing.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12302-022-00674-4.

Additional File 1: **Figure S1** Distribution of NFYA family members in 11 species. **Figure S2** Analysis of Non-synonymous (Ka) to Synonymous (Ks) ratio. **Table S1** *NFYA* family genes correspond to eleven species gene renames. **Table S2** Analysis of the physical and chemical properties of *NFYA* genes in *G.hirsutum*.

Additional File 2: **Table S3** *Cis*-acting elements of upland cotton NFYA family members.

Additional File 3: **Table S4** Expression values of 12 selected genes under salt stress.

Acknowledgements
Not applicable.

Author contributions
Conceptualization, NX; data curation, NX, HH, MH, JW, SW, CC and LG; methodology, NX and XC; project administration, WY; software, NX and YZ; supervision, WY; validation, NX, XF, KN, DW and LZ; visualization, NX; writing—original draft, NX; writing—review and editing, YC**,** HZ, YF and XL. All authors read and approved the final manuscript

Funding
This work was supported by China Agriculture Research System of MOF and MARA, and supported by National Natural Science Foundation of China (No. 31901509).

Availability of data and materials
The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors agreed to publish the paper.

Competing interests
The authors declare no competing or financial interests.

Author details
[1]Institute of Cotton Research of Chinese Academy of Agricultural Sciences/Zhengzhou Research Base, State Key Laboratory of Cotton Biology, School of Agricultural Sciences, Zhengzhou University, Anyang 455000, Henan, China. [2]Anyang Institute of Technology, Anyang 455000, Henan, China.

Received: 13 July 2022 Accepted: 11 September 2022
Published online: 22 September 2022

References

1. Francois M, Donovan P, Fontaine F (2020) Modulating transcription factor activity: interfering with protein-protein interaction networks. Semin Cell Dev Biol 99:12–19
2. Gao W, Liu W, Zhao M, Li WX (2015) NERF encodes a RING E3 ligase important for drought resistance and enhances the expression of its antisense gene *NFYA5* in *Arabidopsis*. Nucleic Acids Res 43:607–617
3. Su H, Cao Y, Ku L et al (2018) Dual functions of *ZmNF-YA3* in photoperiod-dependent flowering and abiotic stress responses in maize. J Exp Bot 69:5177–5189
4. Wu M, Wu S, Chen Z et al (2015) Genome-wide survey and expression analysis of the amino acid transporter gene family in poplar. Tree Genet Genomes. https://doi.org/10.1007/s11295-015-0908-4
5. Xie Q, Frugis G, Colgan D, Chua NH (2000) *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev 14:3024–3036
6. Quan S, Niu J, Zhou L, Xu H, Ma L, Qin Y (2018) Identification and characterization of NF-Y gene family in walnut (*Juglans regia L*). BMC Plant Biol. https://doi.org/10.1186/s12870-018-1459-2
7. Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. Gene 239:15–27
8. Nardini M, Gnesutta N, Donati G et al (2013) Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. Cell 152:132–143
9. Li WX, Oono Y, Zhu J et al (2008) The *Arabidopsis* NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. Plant Cell 20:2238–2251
10. Nelson DE, Repetti PP, Adams TR et al (2007) Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc Natl Acad Sci USA 104:16450–16455
11. Ni Z, Hu Z, Jiang Q, Zhang H (2013) *GmNFYA3*, a target gene of miR169, is a positive regulator of plant tolerance to drought stress. Plant Mol Biol 82:113–129
12. Zhang T, Hu Y, Jiang W et al (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum L*. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol 33:531–537
13. Yuan D, Tang Z, Wang M et al (2015) The genome sequence of Sea-Island cotton (Gossypium barbadense) provides insights into the allopolyploidization and development of superior spinnable fibres. Sci Rep 5:17662
14. Du X, Huang G, He S et al (2018) Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. Nat Genet 50:796–802
15. Paterson AH, Wendel JF, Gundlach H et al (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature 492:423–427
16. Zhu T, Liang CZ, Meng ZG et al (2017) CottonFGD: an integrated functional genomics database for cotton. BMC Plant Biol 17:101
17. Goodstein DM, Shu S, Howson R et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178-1186
18. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) The proteomics protocols handbook. Humana Press, Totowa, pp 571–607
19. Chen C, Chen H, Zhang Y et al (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant 13:1194–1202
20. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874
21. Wang Y, Tang H, Debarry JD et al (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40:e49
22. Krzywinski M, Schein J, Birol I et al (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645
23. Li J, Zhang Z, Vang S, Yu J, Wong GK, Wang J (2009) Correlation between Ka/Ks and Ks is related to substitution model and evolutionary lineage. J Mol Evol 68:414–423
24. Malik WA, Wang X, Wang X et al (2020) Genome-wide expression analysis suggests glutaredoxin genes response to various stresses in cotton. Int J Biol Macromol 153:470–491
25. Hu Y, Chen J, Fang L et al (2019) *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nat Genet 51:739–748
26. Zhang Z, Chai M, Yang Z, Yang Z, Fan L (2022) GRAND: an integrated genome, transcriptome resources, and gene network database for gossypium. Front Plant Sci 13:773107
27. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ (2019) Cytoscape StringApp: network analysis and visualization of proteomics data. J Proteome Res 18:623–632
28. Horton P, Park KJ, Obayashi T et al (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res 35:W585-587

29. Zeng R, Gao S, Xu L, Liu X, Dai F (2018) Prediction of pathogenesis-related secreted proteins from *Stemphylium lycopersici*. BMC Microbiol 18:191
30. Lian C, Li Q, Yao K et al (2018) Corrigendum: populus trichocarpa ptnf-ya9, a multifunctional transcription factor, regulates seed germination, abiotic stress, plant growth and development in *Arabidopsis*. Front Plant Sci 9:1403
31. Xu GX, Guo CC, Shan HY, Kong HZ (2012) Divergence of duplicate genes in exon-intron structure. Proc Natl Acad Sci USA 109:1187–1192
32. Li F, Fan G, Lu C et al (2015) Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol 33:524–530
33. Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet 3:827–837
34. Wang D, Zhang S, He F, Zhu J, Hu S, Yu J (2009) How do variable substitution rates influence Ka and Ks calculations? Genomics Proteomics Bioinform 7:116–127
35. Fan W, Zhang Z, Zhang Y (2009) Cloning and molecular characterization of fructose-1,6-bisphosphate aldolase gene regulated by high-salinity and drought in *Sesuvium portulacastrum*. Plant Cell Rep 28:975–984
36. Oelze ML, Muthuramalingam M, Vogel MO, Dietz KJ (2014) The link between transcript regulation and de novo protein synthesis in the retrograde high light acclimation response of *Arabidopsis thaliana*. BMC Genomics 15:320
37. Wang XG, Lu XK, Malik WA et al (2020) Differentially expressed bZIP transcription factors confer multi-tolerances in *Gossypium hirsutum* L. Int J Biol Macromol 146:569–578
38. Faraji S, Heidari P, Amouei H, Filiz E, Abdullah FS, Poczai P (2021) Investigation and computational analysis of the sulfotransferase (SOT) gene family in potato (Solanum tuberosum): insights into sulfur adjustment for proper development and stimuli responses. Plants. https://doi.org/10.3390/plants10122597
39. Heidari P, Abdullah FS, Poczai P (2021) Magnesium transporter gene family: genome-wide identification and characterization in *Theobroma cacao*, *Corchorus capsularis*, and *Gossypium hirsutum* of family Malvaceae. Agronomy. https://doi.org/10.3390/agronomy11081651
40. Schober P, Boer C, Schwarte LA (2018) Correlation coefficients: appropriate use and interpretation. Anesth Analg 126:1763–1768
41. Bohlenius H, Huang T, Charbonnel-Campaa L et al (2006) CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. Science 312:1040–1043
42. Wang T, Wei Q, Wang Z et al (2022) *CmNF-YB8* affects drought resistance in chrysanthemum by altering stomatal status and leaf cuticle thickness. J Integr Plant Biol 64:741–755
43. Qu B, He X, Wang J et al (2015) A wheat CCAAT box-binding transcription factor increases the grain yield of wheat with less fertilizer input. Plant Physiol 167:411–423
44. Li F, Fan G, Wang K et al (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet 46:567–572
45. Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9:938–950
46. Yang S, Zhang X, Yue JX, Tian D, Chen JQ (2008) Recent duplications dominate NBS-encoding gene expansion in two woody species. Mol Genet Genomics 280:187–198
47. He H, Dong Q, Shao Y et al (2012) Genome-wide survey and characterization of the *WRKY* gene family in *Populus trichocarpa*. Plant Cell Rep 31:1199–1217
48. Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol 4:10
49. Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. New Phytol 183:557–564
50. Wang J, Zhang Y, Xu N et al (2021) Genome-wide identification of *CK* gene family suggests functional expression pattern against Cd(2+) stress in *Gossypium hirsutum* L. Int J Biol Macromol 188:272–282
51. Hurst LD (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet 18:486
52. Wang Z, Zhou Z, Liu Y et al (2015) Functional evolution of phosphatidylethanolamine binding proteins in soybean and Arabidopsis. Plant Cell 27:323–336
53. Zhang Q, Zhang J, Wei H et al (2020) Genome-wide identification of NF-YA gene family in cotton and the positive role of GhNF-YA10 and GhNF-YA23 in salt tolerance. Int J Biol Macromol 165:2103–2115

Publisher's Note