Genome-wide characterization of the Rho family in cotton provides insights into fiber development

Man He (3022286964@qq.com)
CAAS CRI: Chinese Academy of Agricultural Sciences Cotton Research Institute
LIU Shang
CAAS CRI: Chinese Academy of Agricultural Sciences Cotton Research Institute
CHENG Hailiang
CAAS CRI: Chinese Academy of Agricultural Sciences Cotton Research Institute
ZUO Dongyun
CAAS CRI: Chinese Academy of Agricultural Sciences Cotton Research Institute
WANG Qiaolian
CAAS CRI: Chinese Academy of Agricultural Sciences Cotton Research Institute
LV Limin
CAAS CRI: Chinese Academy of Agricultural Sciences Cotton Research Institute
ZHANG Youping
CAAS CRI: Chinese Academy of Agricultural Sciences Cotton Research Institute
WANG Xingfen
Hebei Agricultural University
SONG Guoli
CAAS CRI: Chinese Academy of Agricultural Sciences Cotton Research Institute

Research

Keywords: Cotton fiber, Cell polarity, Rho family, association analysis, WGCNA

Posted Date: February 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1306831/v1

Abstract

Background: Cotton provides the most natural fibers globally, and fiber development is one of the most important biological processes. Fiber develops from a single cell, and polar growth plays a vital role in fiber development. Therefore, research about cell polarity is essential to reveal fiber development mechanisms. We got inspiration from research on eukaryotes, and a key factor controlling cell polarity, Cdc-42, was focused on because of the essential role in cell polarity.

Result: 2066 Rho members across eight cotton species were identified and grouped into 5 and 8 within A and D sub-genomes, respectively. Asymmetric selection of Rho members was observed among 5 tetraploids. Association analysis of previous resequenced data implied 34 Rho members in *G. hirustum* were related to fiber development. To verify the roles of these Rho members at the transcriptional level, previously published transcriptome data was utilized, and a weighted gene co-expression network was constructed to characterize genome-wide expression patterns. Twenty-two expressed Rho members from 5 modules were identified as functional Rho genes involved in fiber development. Interaction networks of 5 Rho members were constructed based on transcription abundance, and gene ontology enrichment showed these interacted genes participated in cell wall biosynthesis, fatty acid elongation, and other biological processes.

Conclusion: Our study characterized the Rho family in cotton based on previously resequenced data and transcriptome data, providing insights on the polar growth of cotton fiber and potential application in cotton fiber improvement.

Background

Cotton is the primary source of natural textile worldwide, and fiber development is one of the most important biological processes in the cotton functional genomics. Cotton fiber originated from the ovule epidermis and developed into a long trichome(Qin et al. 2011). 4 stages, initiation, elongation, secondary cell wall biosynthesis, and maturation, were divided according to the fiber morphologies in different days post-anthesis (DPA). During 5–25 DPA, fibers elong at a vigorous cell expansion rate (>2mm/day)(Kim et al. 2001, Ji 2003). Given that, cotton fiber is a suitable model for research about single-cell elongation(Haigler et al. 2012). Many investigations about the highly polarized manner possessed by cotton fiber were implemented (Zhang et al. 2017, Thyssen et al. 2017, Zhang et al. 2016, Stiff et al. 2016, Zeng et al. 2019, Wang et al. 2010). In these studies, several protein-coding genes were found to be involved in various metabolic processes, such as ethylene biosynthesis, long-chain fatty acids biosynthesis, and auxin transport during fiber development. However, metabolic processes such as long fatty acid elongation and cell wall synthesis are the results of cell polarized growth, and the key factors under the fiber polarity still remain unclear, indicating further investigations about cotton fiber polarity are demanded.

To detect critical effectors of cell polarity in cotton fiber, we got inspiration from researches about single-cell polarity in eukaryotes. Cdc-42, a highly conserved Rho-family GTPase, was the key element in establishing cell polarity in eukaryotes(Etienne-Manneville 2004, Farhan et al. 2016, Mack et al. 2014). Cdc-42 interacted with various downstream effectors to influence cell polarity by septin organization, actin reorganization, membrane trafficking, and microtubule polarization(Farhan and Hsu 2016, Goryachev et al. 2017, Moran et al. 2019, Woods et al. 2017, Woods et al. 2015, Watson et al. 2014). Although Cdc-42 has various pathways to influence cell polarity, it remains the center of polarity in a wide range of species, implying it may also control cell polarity in cotton fiber development(Etienne-Manneville 2004). Researches about Rho genes in cotton could be date back to the last two decades. Several Rho family members encoding small GTPase were cloned and found to involved in fiber development(Delmer et al. 1995). However, limited with sequencing technology, only a few Rho members were cloned and studied, a systematic investigation about the Rho family in cotton is still lacking. The explosion of genomic data provided the platform to identify family members genome-wide scale (Huang et al. 2020, Chen et al. 2020, Wang et al. 2019).

This study identified 2066 Rho members in three diploids and five tetraploids and classified five and eight groups in A and D (sub) genomes, respectively. These Rho members were classified according to sub-genomes. Furtherly, We found conservation of Rho genes among cottons were various and underwent a selection before domestication of *Gossypium hirustum*. Association analysis based on previous resequenced data confirmed that Rho members in cotton were related to fiber development (Du 2018). Subsequent transcriptome analysis demonstrated that Rho members related to fiber development could be classified into different modules, indicating Rho members regulate fiber cell polar growth in complex ways(Qin et al. 2019). Potential Rho-interacted genes identified in transcriptome analysis were enriched in biological processes such as cell wall biosynthesis and fatty acid elongation, consistent with the hypothesis that Rho members participate in the formation of single-cell polarity. In a word, this study identified Rho members in 8 cotton species and investigated Rho members in *G. hirustum* based on previous genomic and transcriptomic data. Results presented in this study shed insights on single-cell polarity during fiber development and could be applied in molecular breeding, especially for cotton fiber improvement.

Results

Rho family members in cotton

CDC42, as a small GTPase, plays an essential role in cell polarity in researches about vertebrates. Given that the cotton fiber has a polarity growth during fiber development, we infer that key factors homologous with CDC-42 in cotton could regulate fiber development. After gaining Rho HMM (Hiden Markov Model) file from Pfam platform, we identified Rho members among eight *Gossypium* species, including three diploids and five tetraploids (**Table 1**). Because of two sub-genomes contained by tetraploid, we separated Rho members in tetraploid cotton into two parts, members from A sub-genome and D sub-genome. Membership of identified Rho genes were furtherly validated by annotation from Uniprot (**Table**

S1). Finally, we got 139-170 Rho members in 8 cotton species. Among these Rho members, the protein sequence with the longest length was from *G. arboreum*, while the shortest length of proteins, 644 or 645, were from 4 tetraploids (**Table S2**). Variations in protein length implied high sequential diversity among Rho family members during cotton evolution.

Phylogeny analysis of Rho members in cotton

To investigate the diversity of Rho members furtherly, phylogenic analysis about Rho genes was performed. Rho genes were gathered and two phylogenic trees were built according to two sub-genomes, respectively (**Figure 1a, 1b**). Rho members from two A genomes (*G. herbaceum and G. arboreum*) and five A sub-genomes (A_{t1}-A_{t5}) were classified into five groups, while, members from D genome (*G. raimondii*) and five D sub-genomes (D_{t1}-D_{t5}) were classified into eight groups. As GTPase members related to cell polarity are identified in other species, Rho members participate in the basic biological processes and influence cotton evolution. We evaluate sequence conservation of Rho genes in 8 cotton species. (**Figure 2**). Firstly, Rho members from three diploids (A1, A2, and D5) were dealt with multi-alignment (**Figure 2a**). The results of multi-alignment identified that Rho genes in cotton had high diversity. Then, Rho genes from tetraploid and corresponding diploids were used to calculate identical site rate. The conservation of Rho members were evaluated by identical site rates (**Figure 2b**). We furtherly compared identical rates of Rho members among cottons and noticed that Rho members in *G. herbaceum* (A1-A1) had lower identical site rates than those in *G. arboreum* (A2-A2) (**Figure 2b**). What's more, we found that for A1 genome, inter-specific conservation (A1-At_{1-5}) were higher than intra-specific conservation (A1-A1) among Rho genes (higher identical site rates), indicating high conservation of A1 and 5 A sub-genomes **(Table S3)**. These result indicated that Rho members in *G. herbaceum* suffered less sequence diversification and was consistent with the previous conclusion that *G. herbaceum* is more relative with the donor of A sub-genome (Huang 2020). For the situation in D5 genomes, Rho members in *G. raimondii* were more conserved with those from D sub-genomes in *G. hirustum* (D5-Dt_1) and *G. tomentosum* (D5-Dt_3). (**Table S3**). From the results of phylogenic analysis about Rho genes, we found that Rho members had different sequence diversity, impling Rho may play important role in cotton evolution. More conservation between Rho genes from D5 genome and those from Dt_1 and Dt_3 showed Rho genes suffered a selection before domestication of *Gossypium hirustum* and may result in formation of long fiber.

Fiber length-related association analysis about Rho members

To check the potential role of Rho members in fiber development, we performed association analysis about fiber length based on previous sequenced data (Du 2018). From 419 resequenced cultivars, 30 with the longest fiber and 30 with the shortest fiber were selected as long/short fiber groups. By t-test, significant divergence of fiber length was observed between 2 groups (p=1.618e-30) (**Figure 3a** and **Table S4**). To detect candidate genomic regions, genome sequences were divided into windows

(windowsize=50000, stepsize=5000) and Fst of these windows were calculated.(**Figure 3b**). Finally, the windows with the top 5% Fst from 44626 windows were selected for gene detection. 6885 genes overlapped with the selected windows were extracted as candidate genes (**Table S5** and **Table S6**). The functions of 6885 genes were initially checked by GO enrichment analysis. We found that basic biological processes such as RNA binding (GO:0003723), mRNA splicing (GO:0000398), and several processes related to glucose metabolism, for instance, galactose metabolic process (GO:0006012), UDP-glucose 4-epimerase activity (GO:0003978) were enriched. Interestingly, we noticed that GTPase activity (GO:0003924) was also enriched, indicating the activities of GTPase could influence fiber length (**Figure 3c** and **Table S7**). 34 Rho members were found in the 6885 candidate genes which were focused on in the subsequent analysis (**Table S8**).

Expression patterns of Rho members in transcriptome data

From 34 Rho members gained by association analysis, only genes having transcriptional activation could be transcribed and translated into protein with biological functions. To evaluate the transcription ability of 34 Rho members, we proposed transcriptome profiling based on the previously published data. Raw transcriptome data of 8 samples were downloaded, consisting of 4 ovules and 4 fibers of 2 materials, 69-6025-12 with short fiber and 601 long stapled cotton with long fiber (**Table 2**). After raw data alignment and transcripts identification, we gained transcription abundance of each gene. 8678 and 298 differentially expressed genes (DEGs) from fiber/ovule and long/short fiber groups were identified, respectively(**Figure 4a,4b, Table S9**). We defined genes with the maximum TPM larger than five as expressed genes. For candidate 34 Rho genes, 22 of them were found to be expressed, and one of them, Ghir_D03G002970, was validated as an expressed gene by qRT-PCR (**Figure 4c** and **Table s10**). Among 22 expressed Rho members, 10 of them were found in DEGs from the fiber/ovule group (**Figure 4d** and **Table S10**). These 10 Rho members with diverge expression patterns between fiber and ovule epidermis cells may participate in polarity growth.

WGCNA analysis of transcriptome data

Rho members may regulate cell polarity with the involvement of other genes. To infer the potential interactions of Rho members during fiber development, we proposed a weighted gene co-expression network analysis (WGCNA) (**Figure 5a**). Genes whose maximum transcription abundance smaller than five were trimmed. The soft threshold of the network was set as 26, because R square was above 0.8, and mean connectivity was lower than 2000 under this threshold, indicating the network we built was scale-free (**Figure 5b**). After network construction, genes were classified into 13 modules according to the expression spectrum (**Figure 5c**).

Since the modules were generated by expression patterns of genes in each module, an association analysis between gene modules and phenotype data revealed potential functions of these modules. The phenotype data of 8 samples were classified as fiber/ovule and long /short fiber materials. We set 0.5 as

the threshold of Pearson correlations to link module to phenotype (**Figure 5d**). Among modules, pink, magenta, and tan modules were related to the long fiber trait, while the purple module was related to the short fiber trait. As for the fiber/ovule trait, we found blue and brown modules related to fiber development. On the other hand, pink, turquoise, purple, and tan modules were related to ovule tissue.

We checked the distribution of 22 expressed Rho members on 13 modules and found that 7, 3, 10, 1, 1 Rho members were contained by blue, brown, turquoise, green, and magenta modules, respectively (**Figure 5e**). To investigate the role of these modules, the genes with the top 10% membership were selected for GO enrichment analysis (**Table S11**). Results of GO enrichment of blue module showed that fatty acid biosynthetic process (GO:0006633), microtubule (GO:0005874), and microtubule-based movement (GO:0007018) were enriched (**Figure S1a**). The other ovule-related brown module contained different GO terms, such as glucose metabolic process (GO:0006006), cellulose microfibril organization (GO:0010215), and cell growth (GO:0016049) (**Figure S1b**). Although various biological processes were enriched in blue and brown modules, most of the enriched GO terms were metabolic pathways. While, the situation in the turquoise module were opposite. More basic biological processes such as RNA binding (GO:0003723), DNA binding (GO:0003677), and chromatin remodeling (GO:0006338) were enriched to these fiber-related modules (**Figure S1c**). The green module, another fiber-related module, had microtubule-based movement (GO:0007018), microtubule motor activity (GO:0003777), and small GTPase mediated signal transduction (GO:0007264) gathered (**Figure S1d**). The photosynthesis and light-harvesting (GO:0009765) was the most significant GO terms in magenta module, which was related to long fiber, implying the role of photosynthesis in fiber elongation (**Figure S1e**). Various GO terms in modules which harbour Rho genes indicated that complicated mechanisms underlying control of cell polarity by Rho members.

Interaction networks of Rho family members

The membership of the corresponding modules furtherly filtered expressed Rho members from 5 modules, and Rho genes with the highest membership were selected (**Table S12**). Finally, Ghir_A12G010690, Ghir_D08G007370, Ghir_D08G018200, Ghir_A11G010670, and Ghir_D03G002970 were selected as core genes from blue, brown, magenta, turquoise, and green modules (**Figure 6**). Among 5 core genes, 2 of them, Ghir_A12G010690 and Ghir_D08G007370, were DEGs. The genes possessing high transcriptional correlationship (pearson) with core genes were recorded as infered interacted genes which were used to construct interaction networks. GO enrichment of genes in networks were consistent with results from genes having high membership with corresponding modules (**Table S13**). In the network of a DEG, Ghir_A12G010690, genes were related to the fatty acid biosynthetic process (GO:0006633), microtubule-based process (GO:0007017), and small GTPase mediated signal transduction (GO:0007264). The fatty acid was essential for fiber elongation (Qin et al. 2011). Meanwhile, small GTPase mediated signal transduction was also enriched in this network, implying Ghir_A12G010690 may participate in signal transduction to activate polar growth and regulate fiber elongation through regulating fiber-related biological processes such as fatty acid biosynthesis (**Figure 6c** and **Figure 6d**). In

the other four networks, there were also other GO terms involved in fiber development, such as cell wall biogenesis (GO:0042546), calcium ion binding (GO:0005509) (**Table S13**). The diverse enriched GO terms of these five networks indicated that different Rho members played various roles in influencing cotton fiber cell polarity.

Discussion

In this study, we identified 2066 Rho members in 8 cotton species. *G. herbceum* (A_1) and *G. arboreum* (A_2) were the candidate A sub-genome donor in the past time until genome-wide comparison between them was implemented (Huang 2020). Identical sites of Rho family members in 2 A diploids showed that Rho members in *G. herbaceum* (A1) suffered less selection which means A1 had more relativeness with A donor in tetraploids. This conclusion is consistent with the previous findings (Huang 2020). Interestingly, for D sub-genomes, we also noticed that Rho members in *G. hirustum* and *G. tomentosum* were more conserved than those in the other three tetraploids, indicating a specific selection happend in *G. tomentosum* and *G. hirustum*. In the previous study, *G. hirustum* and *G. tomentosum* are two cotton species with the most relativeness (Chen 2020). Therefore, Rho members in D sub-genomes may undergo a strong selection before the divergence between *G. hirustum* and *G. tomentosum* (Chen 2020). This selection may result in long fiber length of *G. hirustum*.

In previous study, association analysis between genotype and phenotype data could dig out functional genes with high validity (Ma 2018). In this research, we performed an association analysis on 60 cultivars to verify potential functional Rho genes. Combined with transcriptomic analysis and association analysis, 22 Rho members were selected with high confidence. GO enrichment on WGCNA modules containing 22 candidate Rho genes implied a complicated mechanisms underlying Rho activities. Rho protein, as GTPase, participates in many essential pathways and has many factors to interact as reported in other species(Etienne-Manneville 2004, Farhan and Hsu 2016, Goryachev and Leda 2017, Moran 2019). The infered interaction networks provided us lights on how Rho control the fiber elongation. GO enrichment analysis about genes from a network constructed using a DEG, Ghir_A12G010690, as core gene showed that fatty acid biosynthesis process and small GTPase mediated signal transduction emerged in this network. We infered Rho genes could activate a signal transduction pathway and regulate fatty acid biosynthesis which was proved to activate fiber elongation. What's more, we noticed that cell wall biosynthesis, calcium ion binding also emerged in other Rho-based interaction networks. This result suggests that apart from influencing fatty acid biosynthesis, Rho genes may have multiple machenisms to regulate fiber development.

Conclusions

In this study we identified Rho members in 5 alloploidy cottons and their corresponding sub-genome donors. Rho members, as a family involved in basic biological processes, had undergone specific selection during cotton evolution. 5 Rho members potentially involved in fiber development were revealed by combinatin of association analysis and transcriptome profiling. Furtherly, infered interaction networks

were built. We believe that findings in this study could be utilized in cotton molecular breeding for fiber improvement.

Methods

Identification of Rho family members

The protein sequence of CDC42 in human was downloaded from Uniprot (https://beta.uniprot.org/) with entry ID P60953(Bateman et al. 2021). The sequence of CDC42 in human was aligned to protein domain database in Pfam (http://pfam.xfam.org/) to search for the corresponding HMM model file, Ras.hmm, with accession number as PF00071.23(Mistry et al. 2021). After getting HMM file of Ras domain, we searched all protein sequences of 8 *Gossypium* sequences (3 diploids, *G. arboreum, G.herbaceum,* and *G.raimondii*; 5 allopolyploids, *G. hirustum, G. barbadense, G.tomentosum, G. mustelinum,* and *G. darwinii*) by hmmscan programme in HMMER (v.3.3.2) to detect proteins with Ras domain(Eddy 2009). E value and domE value of hmmscan were both set as 1e-5. These proteins were aligned to the protein database in uniprot by blastp (v.2.9.0+) to further identify these detected proteins' functions(Altschul et al. 1990).

Multi-alignment within Ras members in cotton

Rho members identified in 8 Gossypium species were collected as the input of muscle (v.3.8.1551) to perform multi-alignments. Before aligning sequences, members grouping was implemented. Members from two A genomes (*G. arboreum* and *G.herbaceum*) were separately grouped into ten pairs with 5 At (A sub-genome of tetraploid), while members from D genome (*G.raimondii*) were grouped with five Dt (D sub-genome of tetraploid) into five pairs. Aligned fasta files from muscle results were used to evaluate the rate of identical sites. Firstly, in the aligned fasta file, the length of each sequence was equal, and for sites in the aligned sequence, sites absent in more than 95% sequences were abandoned. For remained sites, frequencies of site presenting within members were calculated and site presenting frequencies were compared between sub-genomes and putative donors by T-test through python package stats.

Phylogeny analysis on Rho members

Since tetraploids in cotton have two sub-genomes, we performed phylogeny analysis separately on A genomes (A sub-genomes) and D genomes (D sub-genomes). The protein sequences were aligned by muscle and put into fasttree (v.2.1.10) to construct phylogenic trees by default parameters(Price et al. 2009). The phylogenic tree was decorated by iTOL (https://itol.embl.de/).

Association analysis pipeline

Previous resequenced data was fetched from NCBI (https://www.ncbi.nlm.nih.gov/). The SRA accession number of raw data in fastq format was SRP115740 at PRJNA399050(Du 2018). The phenotype data was downloaded from http://cotton.hebau.edu.cn/. Phenotype data of fiber length was utilized in this study. The best linear unbiased prediction (BLUP) was applied to deal with fiber length from 419 cultivars in 12 environments. 30 cultivars with the longest fiber length and 30 cultivars with the shortest fiber length were selected for association analysis. Raw fastq data were aligned to the reference genome TM-1_HAU (https://cottonfgd.org/about/download.html, *G. hirustum*, HAU) by bwa (v0.7.17)(Wang 2019). Samtools (v1.9) transformed the results of bwa into a binary file, and the bcftools (v1.9) was used to call SNPs from binary alignment files(Li et al. 2009, Danecek et al. 2011). The genetic differentiation index (Fst) of 2 groups was calculated by vcftools (v0.1.16) with 50000 window size and 5000 step size(Danecek 2011).

Transcriptome analysis pipeline

Raw transcriptome data in fastq format were downloaded from SRA database on NCBI (https://www.ncbi.nlm.nih.gov/). Eight transcriptome data (SRR5992414 and SRR6379574-SRR6379580) from PRJNA400837 were obtained by sratoolkit (v.2.9.6). Quality control and the data trimming were performed by fastp (v.0.23.1) with the default parameters in pair-end mode, and the corresponding clean data were generated(Chen et al. 2018). We used the genome of Texas Marker-1 in association analysis as the reference genome in the subsequent analysis. Index of the reference genome was implemented by hisat2 (v.2.2.1). The clean data of 8 samples were aligned to the reference genome in a pair-end mode(Kim et al. 2019). The results of hisat2 alignment were transformed into binary alignment format file (bam), then the bam files were sorted. Transforming and sorting the files were both performed by samtools (v.1.9)(Li et al. 2009). Stringtie (2.1.7) was used to identify the transcription abundance of each gene with a gene feature file guided (-G) and transcription abundance of 8 samples were merged into one file(Pertea et al. 2015). Eight samples contained a long fiber and a short fiber cultivar. The transcription abundance of each gene between long fiber samples and short fiber samples was compared by t-test. The genes with significant results ($p<=0.05$) of t-test will be identified as differentially expressed genes (DEGs).

WGCNA analysis pipeline

Since transcription abundance of all genes in 8 samples were gained, a weighted gene co-expression network analysis (WGCNA) could be performed(Langfelder et al. 2012, Langfelder et al. 2008). Before network construction, genes in which the maximum TPM among eight samples was smaller than five were filtered. Apart from criterio about transcription abundance, genes absent in more than two samples and samples had more than 10% genes absent were removed. After data filtering, the co-expression network was constructed by R package WGCNA based on the remained genes. Different soft thresholds were applied, and finally, the network was evaluated by mean connectivity and R square. To construct a

scale-free network, deepSplit was set as two, and mergeCutHeight was 0.35. The minimum genes in a block were 30, while the maximum gene number in a block was 35000. All remained genes were divided into several modules by WGCNA, and the eigenvalue of each module was associated with phenotype through Pearson correlations. Membership of each gene and their potential interacted genes were recorded in the Cytoscape-like file for further visualization.

GO enrichment analysis

To perform GO enrichment analysis, genes collected from WGCNA were handed up to the CottonFGD (https://cottonfgd.org/analyze/)(Zhu et al. 2017). The significant level was set as 0.05, and the minimum gene number for enrichment analysis was set as 3. The result of significant GO enrichment will be recorded as a table, and the q-value of each GO term was used for visualization.

Cis-element analysis of Rho members

The promoter region was extracted according to gene name in CottonFGD (https://cottonfgd.org/analyze/). Cis-elements within these sequences were detected by PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/)(Lescot et al. 2002).

Data visualization methods

In all contents of this paper, data visualizations were implemented in various ways. Results of multi-alignments were plotted by python package matplotlib. Phylogenic trees were displayed by iTOL (https://itol.embl.de/)(Letunic et al. 2019). For visualization during DEG analysis, Venn plots were generated through an online tool Venny2.1.0 (https://bioinfogp.cnb.csic.es/tools/venny/index.html). All heatmaps in this study were plotted by python package seaborn. Figures about WGCNA were all generated by R package WGCNA except for association between modules and phenotype data. Python package pyecharts generated Sankey plot about Rhos and modules from WGCNA. Interactions of potential genes with Rho members were displayed by Cytoscape (v.3.7.2)(Shannon et al. 2003).

Abbreviations

GO: gene ontology; SNP: single nucleotide polymorphism; GWAS: genome-wide association analysis; TPM: transcripts per million; DEG: differentially expressed gene; WGCNA: weighted gene co-expressing network analysis.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The resequenced data in this manuscript could be downloaded from SRP115740 at PRJNA399050 in NCBI. The phenotype data of corresponding resequenced cultivars could be downloaded from the hyperlink http://cotton.hebau.edu.cn/. Transcriptome raw data could be fetched from PRJNA400837 in NCBI. Reference genomes used in this study were all obtained from https://cottonfgd.org/about/download.html.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Funds of the National Key Research and Development Program (2016YFD0101006, No. 2018YFD0100402), National Natural Science Foundation of China (No. 31621005 and No. 31901581), and Central Public-interest Scientific Institution Basal Research Fund(No.1610162021013). The funders had no role in the design of the study, collection, analysis or interpretation of the data, the writing of the manuscript or the decision to submit the manuscript for publication.

Authors contributions

WANG Xingfen, SONG Guoli conceived and designed the research. HE Man, LIU Shang, CHENG Hailiang and ZUO Dongyun performed the analysis and experiment. WANG Qiaolian, LV Limin, and ZHANG Youping downloaded the data and prepared samples. HE Man wrote the paper. WANG Xingfen, SONG Guoli, and ZUO Dongyun revised the manuscript. All authors read and approved the final manuscript.

Acknowledgments

Not applicable

References

1. Qin YM, Zhu YX. How cotton fibers elongate: a tale of linear cell-growth mode. CURR OPIN PLANT BIOL. 2011;14(1):106-11.
2. Kim HJ, Triplett BA. Cotton Fiber Growth in Planta and in Vitro. Models for Plant Cell Elongation and Cell Wall Biogenesis. PLANT PHYSIOL. 2001;127(4):1361-6, doi: 10.1104/pp.010724.
3. Ji SJ. Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. NUCLEIC ACIDS RES. 2003;31(10):2534-43, doi: 10.1093/nar/gkg358.
4. Haigler CH, Betancur L, Stiff MR, et al. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. FRONT PLANT SCI. 2012;3, doi: 10.3389/fpls.2012.00104.
5. Zhang Y, He P, Yang Z, et al. A Genome-Scale Analysis of the PIN Gene Family Reveals Its Functions in Cotton Fiber Development. FRONT PLANT SCI. 2017;8, doi: 10.3389/fpls.2017.00461.
6. Thyssen GN, Fang DD, Turley RB, et al. A Gly65Val substitution in an actin, GhACT_LI1, disrupts cell polarity and F-actin organization resulting in dwarf, lintless cotton plants. The Plant Journal. 2017;90(1):111-21, doi: 10.1111/tpj.13477.
7. Zhang M, Zeng J, Long H, Xiao Y, Yan X, Pei Y. Auxin Regulates Cotton Fiber Initiation via GhPIN-Mediated Auxin Transport. PLANT CELL PHYSIOL. 2016; w203, doi: 10.1093/pcp/pcw203.
8. Stiff MR, Haigler CH. Cotton fiber tips have diverse morphologies and show evidence of apical cell wall synthesis. SCI REP-UK. 2016;6(1):27883, doi: 10.1038/srep27883.
9. Zeng J, Zhang M, Hou L, et al. Cytokinin inhibits cotton fiber initiation by disrupting PIN3a-mediated asymmetric accumulation of auxin in the ovule epidermis. J EXP BOT. 2019 2019-06-28;70(12):3139-51, doi: 10.1093/jxb/erz162.
10. Wang LK, Niu XW, Lv YH, et al. Molecular cloning and localization of a novel cotton annexin gene expressed preferentially during fiber development. MOL BIOL REP. 2010;37(7):3327-34, doi: 10.1007/s11033-009-9919-2.
11. Etienne-Manneville S. Cdc42 - the centre of polarity. J CELL SCI. 2004;117(8):1291-300, doi: 10.1242/jcs.01115.
12. Farhan H, Hsu VW. Cdc42 and Cellular Polarity: Emerging Roles at the Golgi. TRENDS CELL BIOL. 2016;26(4):241-8, doi: 10.1016/j.tcb.2015.11.003.
13. Mack NA, Georgiou M. The interdependence of the Rho GTPases and apicobasal cell polarity. Small GTPases. 2014;5(2):e973768, doi: 10.4161/21541248.2014.973768.
14. Goryachev AB, Leda M. Cell Polarity: Spot-On Cdc42 Polarization Achieved on Demand. CURR BIOL. 2017;27(16):R810-2, doi: 10.1016/j.cub.2017.07.006.
15. Moran KD, Kang H, Araujo AV, et al. Cell-cycle control of cell polarity in yeast. J CELL BIOL. 2019;218(1):171-89, doi: 10.1083/jcb.201806196.

16. Woods B, Lew DJ. Polarity establishment by Cdc42: Key roles for positive feedback and differential mobility. Small GTPases. 2017;10(2):130-7, doi: 10.1080/21541248.2016.1275370.
17. Woods B, Kuo C, Wu C, et al. Polarity establishment requires localized activation of Cdc42. J CELL BIOL. 2015;211(1):19-26, doi: 10.1083/jcb.201506108.
18. Watson LJ, Rossi G, Brennwald P. Quantitative Analysis of Membrane Trafficking in Regulation of Cdc42 Polarity. TRAFFIC. 2014;15(12):1330-43, doi: 10.1111/tra.12211.
19. Delmer DP, Pear JR, Andrawis A, et al. Genes encoding small GTP-binding proteins analogous to mammalian rac are preferentially expressed in developing cotton fibers. Mol Gen Genet. 1995;248(1):43-51, doi: 10.1007/BF02456612.
20. Huang G, Wu Z, Percy RG, et al. Genome sequence of Gossypium herbaceum and genome updates of Gossypium arboreum and Gossypium hirsutum provide insights into cotton A-genome evolution. NAT GENET. 2020;52(5):516-24, doi: 10.1038/s41588-020-0607-4.
21. Chen ZJ, Sreedasyam A, Ando A, et al. Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. NAT GENET. 2020;52(5):525-33, doi: 10.1038/s41588-020-0614-5.
22. Wang M, Tu L, Yuan D, et al. Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. NAT GENET.. 2019;51(2):224-9, doi: 10.1038/s41588-018-0282-x.
23. Du X. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. 2018;50(6):803-813, doi: 10.1038/s41588-018-0119-7.
24. Qin Y, Sun H, Hao P, et al. Transcriptome analysis reveals differences in the mechanisms of fiber initiation and elongation between long- and short-fiber cotton (Gossypium hirsutum L.) lines. BMC GENOMICS. 2019;20(1):633, doi: 10.1186/s12864-019-5986-5.
25. Bateman A, Martin M, Orchard S, et al. UniProt: the universal protein knowledgebase in 2021. NUCLEIC ACIDS RES. 2021;49(D1):D480-9, doi: 10.1093/nar/gkaa1100.
26. Mistry J, Chuguransky S, Williams L, et al. Pfam: The protein families database in 2021. NUCLEIC ACIDS RES. 2021;49(D1):D412-9, doi: 10.1093/nar/gkaa913.
27. Eddy SR. A new generation of homology search tools based on probabilistic inference. Genome Inform. 2009;23(1):205-11.
28. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. J MOL BIOL. 1990;215(3):403-10, doi: 10.1016/S0022-2836(05)80360-2.
29. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. MOL BIOL EVOL. 2009;26(7):1641-50, doi: 10.1093/molbev/msp077.
30. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. BIOINFORMATICS. 2009;25(16):2078-9, doi: 10.1093/bioinformatics/btp352.
31. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. BIOINFORMATICS. 2011;27(15):2156-8, doi: 10.1093/bioinformatics/btr330.

32. Chen S, Zhou Y, Chen Y, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. BIOINFORMATICS. 2018;34(17):i884-90, doi: 10.1093/bioinformatics/bty560.
33. Kim D, Paggi JM, Park C, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. NAT BIOTECHNOL. 2019;37(8):907-15, doi: 10.1038/s41587-019-0201-4.
34. Pertea M, Pertea GM, Antonescu CM, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. NAT BIOTECHNOL. 2015;33(3):290-5, doi: 10.1038/nbt.3122.
35. Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. J STAT SOFTW. 2012;46(11).
36. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC BIOINFORMATICS. 2008;9:559, doi: 10.1186/1471-2105-9-559.
37. Zhu T, Liang C, Meng Z, et al. CottonFGD: an integrated functional genomics database for cotton. BMC PLANT BIOL. 2017;17(1):101, doi: 10.1186/s12870-017-1039-x.
38. Lescot M, Déhais P, Thijs G, et al. 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):325-7, doi: 10.1093/nar/30.1.325.
39. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. NUCLEIC ACIDS RES. 2019;47(W1):W256-9, doi: 10.1093/nar/gkz239.
40. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. GENOME RES. 2003;13(11):2498-504, doi: 10.1101/gr.1239303.

Tables

Table 1

The number of Rho family members in 8 cotton species. For tetrapploids, numbers of Rho family members within A and D sub-genomes were calculated seperately.

species	subgenome	number
G. herbaceum	A1	157
G. arboreum	A2	144
G. raimondii	D5	164
G. hirustum	At1	138
G. barbadense	At2	168
G. tomentosum	At3	169
G. mustelinum	At4	165
G. darwinii	At5	170
G. hirustum	Dt1	143
G. barbadense	Dt2	159
G. tomentosum	Dt3	164
G. mustelinum	Dt4	163
G. darwinii	Dt5	162

Table 2

Transcriptome data used in this study, the basic information about transcriptome data, including data accession, sample name, tissue, time of tissue and phenotype was recorded.

Data accession	Sample name	Time	Tissue	Phenotype
SRR6379580	69-6025-12	5 DPA	Fiber	Short fiber
SRR5992414	69-6025-12	0DPA	Ovule	Short fiber
SRR6379578	69-6025-12	10 DPA	Fiber	Short fiber
SRR6379579	69-6025-12	10 DPA	Ovule	Short fiber
SRR6379574	601 long stapled cotton	10 DPA	Fiber	Long fiber
SRR6379575	601 long stapled cotton	10DPA	Ovule	Long fiber
SRR6379576	601 long stapled cotton	0 DPA	Ovule	Long fiber
SRR6379577	601 long stapled cotton	5 DPA	Fiber	Long fiber

Figures

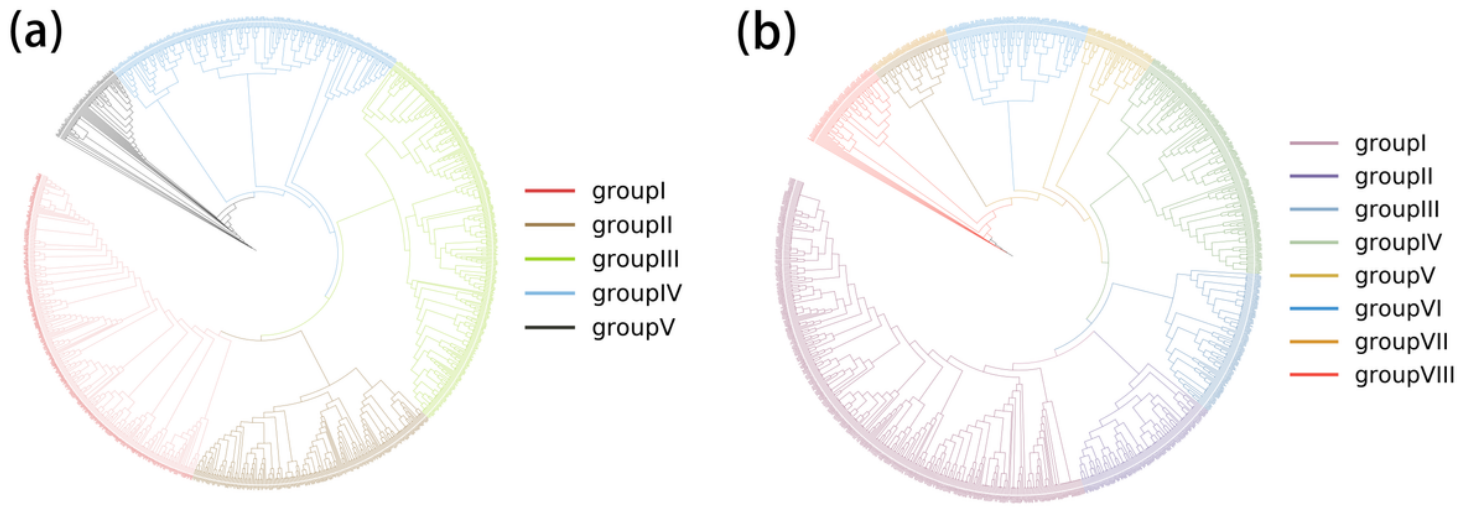

Figure 1

The phylogenic analysis of Rho members across 8 cotton species. (a) Phylogenic tree about Rho members from 2 A genomes and 5 A sub-genomes. (b) Phylogenic tree about Rho members from 1 D genome and 5 D sub-genomes.

Figure 2

Identical site in Rho members across 8 cotton species. (a) Multi-alignments of Rho members from A1, A2, and D5 with Rho members from 5 tetraploids, respectively. The orange and purple dots on the plot indicate identical sites in multi-alignments. From up to down for each diploid, the comparsion pairs were *G. herbaceum* (*G. arboreum, G. raimondii*) vs *G. herbaceum* (*G. arboreum, G. raimondii*), *G. herbaceum* (*G. arboreum, G. raimondii*) vs *G. hirustum*, *G. herbaceum* (*G. arboreum, G. raimondii*) vs *G. barbadense*, *G. herbaceum* (*G. arboreum, G. raimondii*) vs *G. tomentosum*, *G. herbaceum* (*G. arboreum, G. raimondii*) vs *G. mustelinum*, and *G. herbaceum* (*G. arboreum, G. raimondii*) vs *G. darwinii*. (b) The boxplots which display significance of identical sites in a comparsion pair, the t-test was applied for significance calculation.

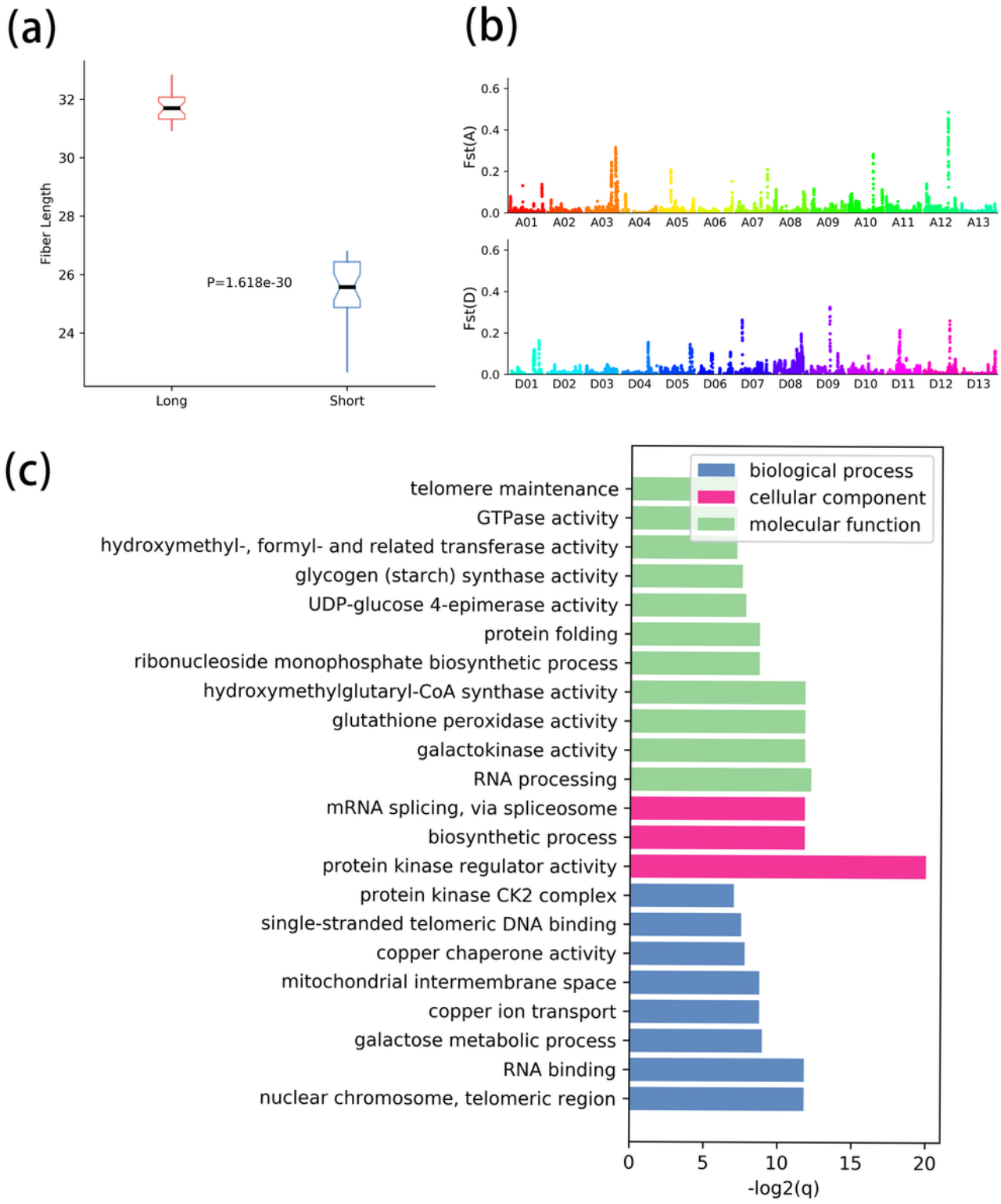

Figure 3

Results of genotype-phenotype association. (a) phenotype divergence between long fiber and short fiber cultivars. (b) Manhattan plot of Fst and location of window in association. (c) Go enrichment of genes overlapped with windows with the top 5% Fst values.

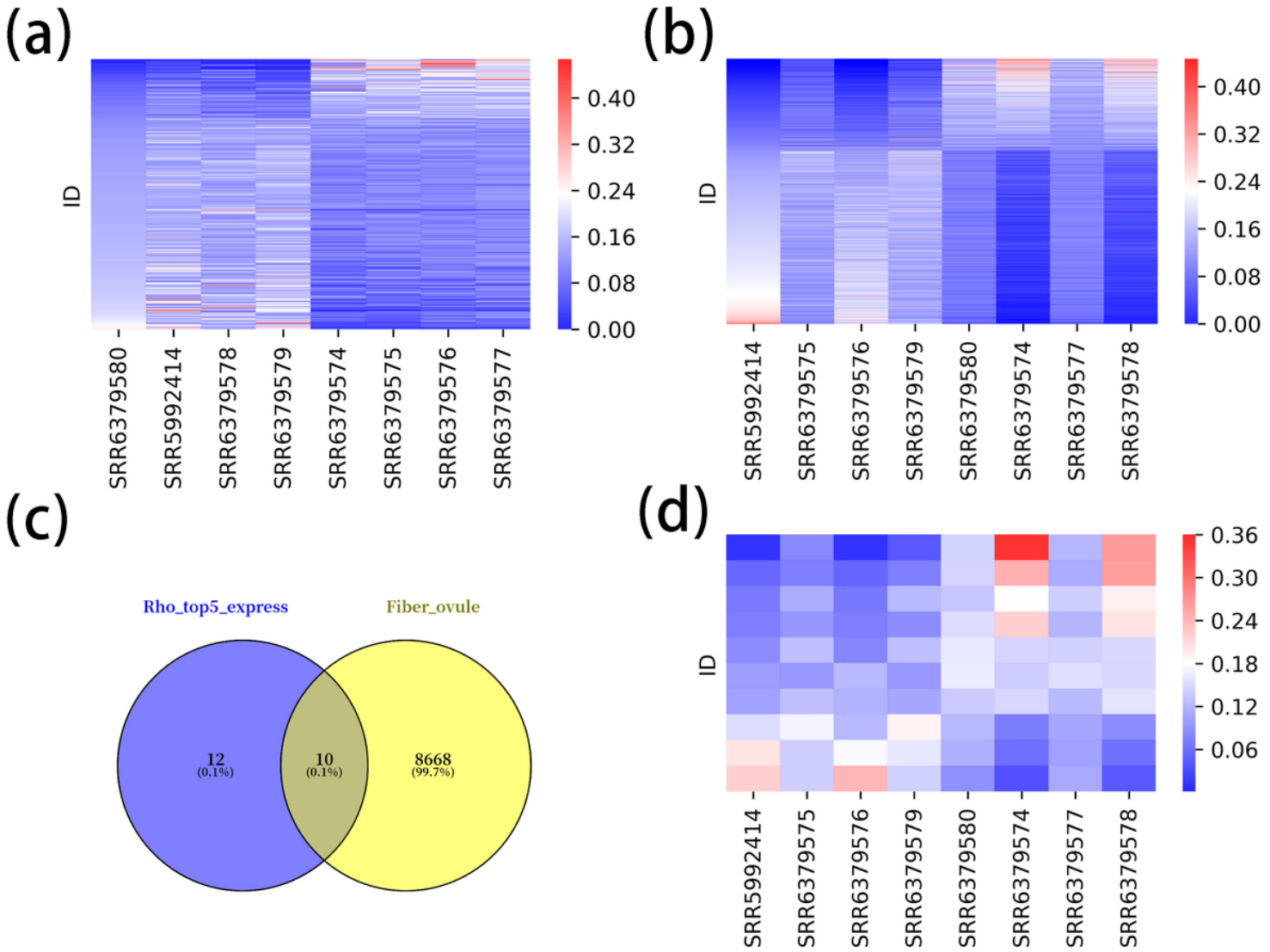

Figure 4

Characterization of DEGs in transcriptome data. (a) Relative transcription abundance of DEGs obtained from long fiber/short fiber cultivars. (b) Relative transcription abundance of DEGs obtained from fiber/ovule tissues. (c) Venn plot of expressed Rho members and DEGs in fiber/ovule tissues. (d) Relative transcription abundance of expressed Rho genes in fiber/ovule DEGs.

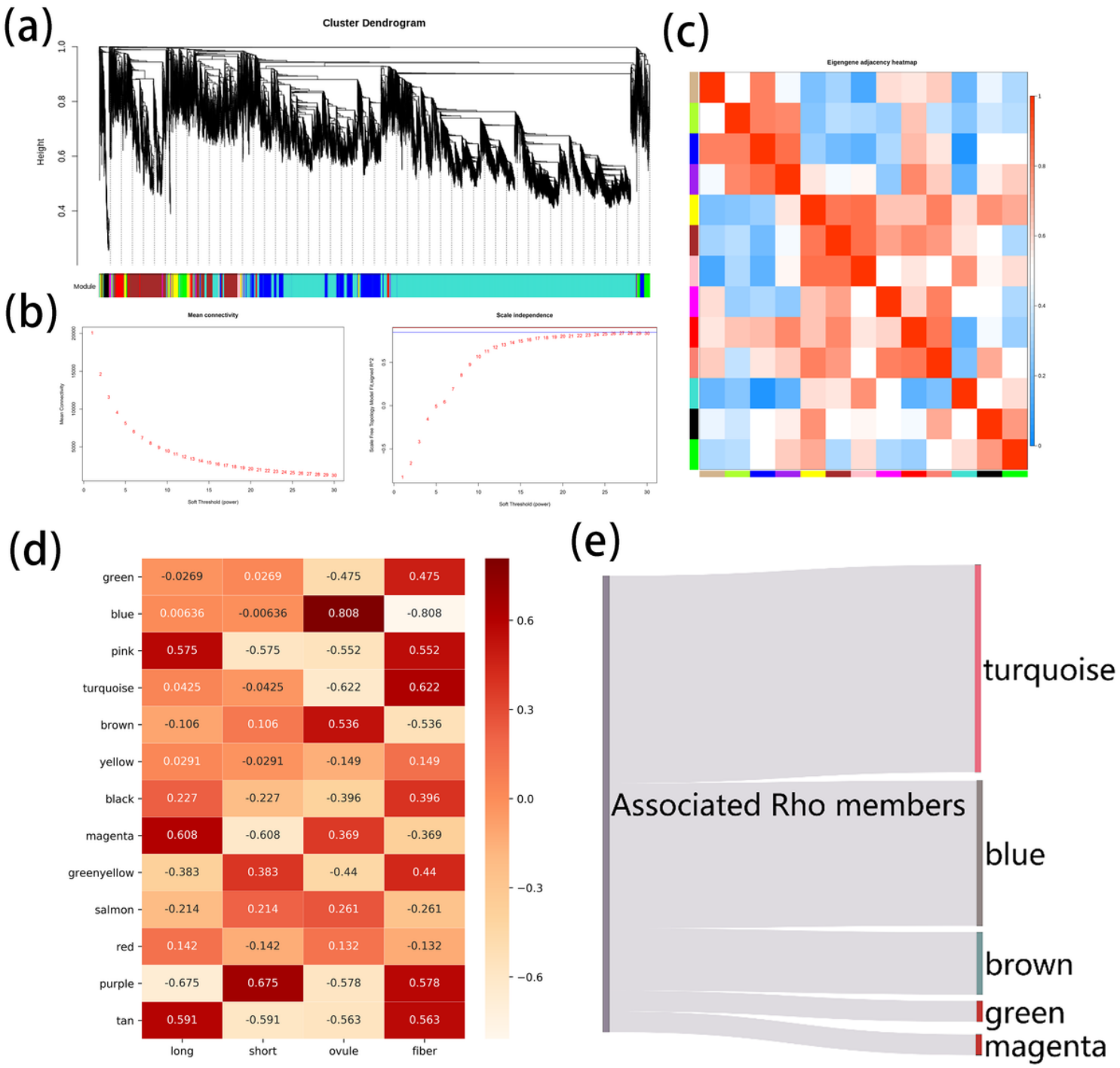

Figure 5

WGCNA for transcriptome data in the study. (a) results of gene clusteringin WGCNA. (b) Relationship of R-square and soft threshold, when softthreshold was set as 26, the R-square was larger than 0.8. Mean connectivity of network under different soft thresholds. (c) 13 modules classified from WGCNA. (d) Trait-module association results of WGCNA. (e) Sankey plot about distribution of associated Rho members in modules gained from WGCNA.

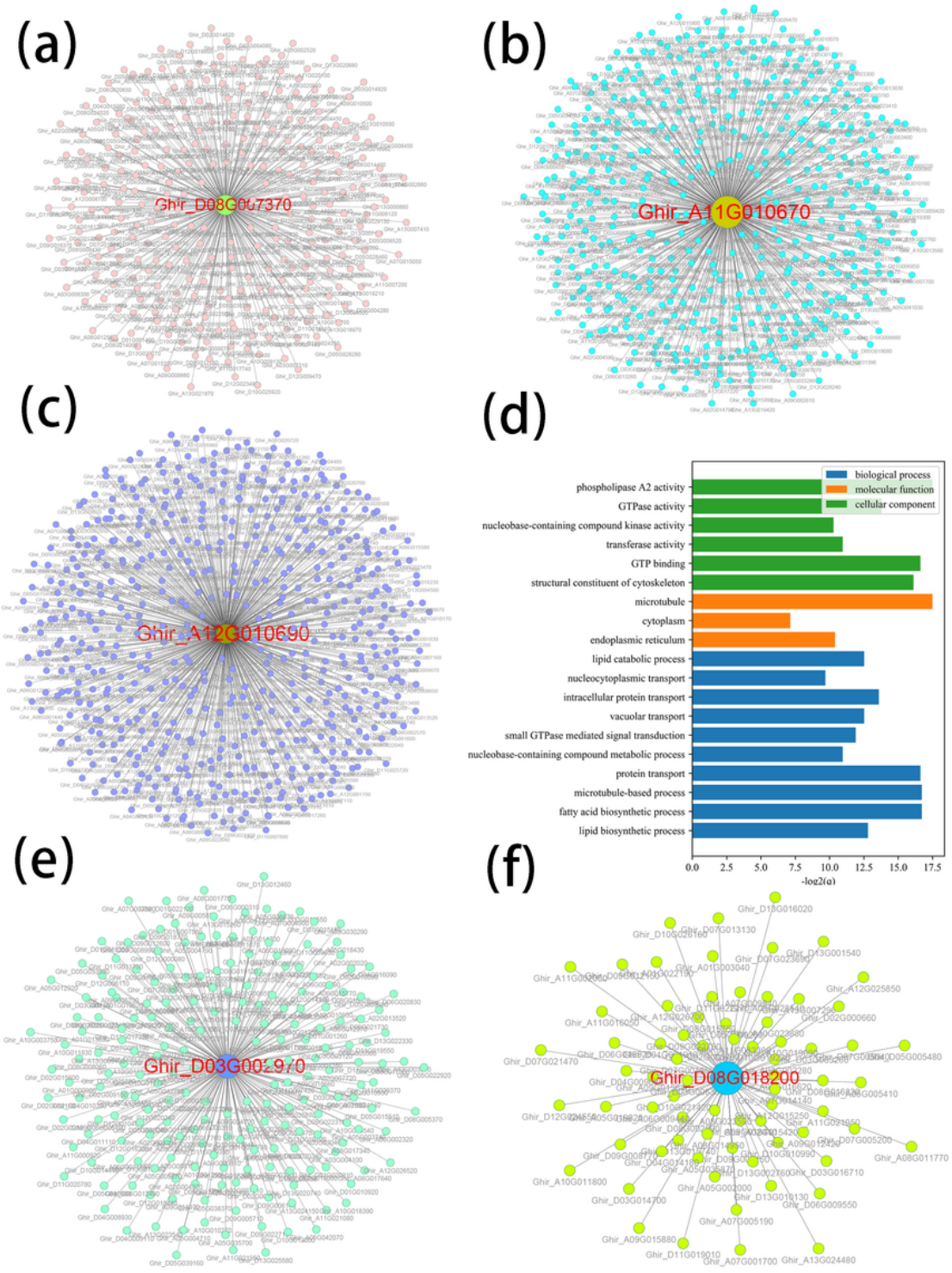

Figure 6

Interaction networks of 5 Rho members. (a) Interaction network of Ghir_D08G007370. (b) Interaction network of Ghir_A11GG010670. (c) Interaction network of Ghir_A12G010690. (d) GO enrichment of genes interacted with Ghir_A12G010690. (e) Interaction network of Ghir_D03G002970. (f) Interaction network of Ghir_D08G018200.

Supplementary Files

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

- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx
- TableS7.xlsx
- TableS8.xlsx
- TableS9.xlsx
- TableS10.xlsx
- TableS11.xlsx
- TableS12.xlsx
- TableS13.xlsx
- FigureS1.png