RESEARCH ARTICLE

# Genome-wide identification and characterization of FORMIN gene family in cotton (*Gossypium hirsutum* L.) and their expression profiles in response to multiple abiotic stress treatments

Suronjeet Kumar Paul[1]☯, Md Shohel Ul Islam[iD][1]☯, Nasrin Akter[1], Fatema Tuz Zohra[2], Shuraya Beente Rashid[1], Md. Shakil Ahmed[3], Shaikh Mizanur Rahman[1], Md. Abdur Rauf Sarkar[iD][1]*
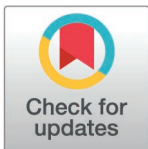
1 Laboratory of Functional Genomics and Proteomics, Department of Genetic Engineering and Biotechnology, Faculty of Biological Science and Technology, Jashore University of Science and Technology, Jashore, Bangladesh, 2 Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Rajshahi, Rajshahi, Bangladesh, 3 Department of Biochemistry and Molecular Biology, Faculty of Science, University of Rajshahi, Rajshahi, Bangladesh

☯ These authors contributed equally to this work.
* rauf.gebt@yahoo.com

## Abstract

FORMIN proteins distinguished by FH2 domain, are conserved throughout evolution and widely distributed in eukaryotic organisms. These proteins interact with various signaling molecules and cytoskeletal proteins, playing crucial roles in both biotic and abiotic stress responses. However, the functions of FORMINs in cotton (*Gossypium hirsutum* L.) remain uncovered. In this study, 46 FORMIN genes in *G. hirsutum* (referred to as *GhFH*) were systematically identified. The gene structures, conserved domains, and motifs of these *GhFH* genes were thoroughly explored. Phylogenetic and structural analysis classified these 46 *GhFH* genes into five distinct groups. *In silico* subcellular localization, prediction suggested that *GhFH* genes are distributed across various cellular compartments, including the nucleus, extracellular space, cytoplasm, mitochondria, cytoskeleton, plasma membrane, endoplasmic reticulum, and chloroplasts. Evolutionary and functional diversification analyses, based on on-synonymous (*Ka*) and synonymous (*Ks*) ratios and gene duplication events, indicated that *GhFH* genes have evolved under purifying selection. The analysis of *cis*-acting elements suggested that *GhFH* genes may be involved in plant growth, hormone regulation, light response, and stress response. Results from transcriptional factors TFs and gene ontology analysis indicate that FORMIN proteins regulate cell wall structure and cytoskeleton dynamics by reacting to hormone signals associated with environmental stress. Additionally, 45 putative ghr-miRNAs were identified from 32 families targeting 33 *GhFH* genes. Expression analysis revealed that *GhFH1*, *GhFH10*, *GhFH20*, *GhFH24*, and *GhFH30* exhibited the highest levels of expression under red, blue, and white light conditions. Further, *GhFH9*, *GhFH20*, and *GhFH30* displayed higher expression levels under heat stress, while *GhFH20* and *GhFH30* showed increased expression under

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

salt stress compared to controls. The result suggests that *GhFH20* and *GhFH30* genes could play significant roles in the development of *G. hirsutum* under heat and salt stresses. Overall these findings enhance our understanding of the biological functions of the cotton FORMIN family, offering prospects for developing stress-resistant cotton varieties through manipulation of *GhFH* gene expression.

## 1. Introduction

Light serves as a crucial energy source for photosynthesis and acts as a key signal in plant growth and development [1]. It primarily influences plant growth by regulating the activity of photosynthetic genes, which in turn contribute to the production of carbohydrates and other important secondary compounds in plants [2,3]. Plants possess photoreceptors such as phototropins, phytochromes, and cryptochromes which absorb light and initiate signaling pathways that affect plant physiology [4]. Due to their immobility, plants are exposed to various environmental stresses, including heavy metals, high salinity, drought, nutrient shortages, varying light levels, pesticide pollution, and extreme temperature conditions [5]. To adapt to these external conditions plants undergo rapid morphological changes in roots, leaves, and pollen, driven by cytoskeleton dynamics [6]. FORMINs are known to regulate actin cytoskeleton dynamics by facilitating actin polymerization [7]. These proteins enhance actin polymerization by stimulating filament nucleation and elongation [8,9]. In pants, three key FORMIN domains have been identified: FORMIN Homology 1 (FH1), FORMIN Homology 2 (FH2), and FORMIN Homology 3 (FH3) [10]. Among these, the FH2 domain is crucial for polymerase function and serves a specific marker for identifying gene families [11]. Plants FORMINs regulate actin filaments and control the dynamic remodeling actin cytoskeleton enabling plant cells to change their shapes and manage the cellular structure of plant tissues and organs [12]. FORMINs have been implicated in pathogen resistance and play a crucial role in male fertility in wheat (*Triticum aestivum*) [9,13]. The FH2 domain has been observed to influence actin polymerization dynamics through various mechanisms, including altering the rate of filament elongation and depolymerization, enhancing *de novo* filament nucleation, and preventing filament barbed-end capping by capping proteins [14].

Numerous FORMIN genes have been identified in various plant species, and comprehensive genome-wide analyses have been conducted in recent studies, such as 17 genes in rice (*Oryza sativa*) [15], 22 genes in *Arabidopsis* (*Arabidopsis thaliana*) [16], 25 genes in wheat (*Triticum aestivum*) [9]. In *Arabidopsis*, the functions of specific genes such as *AtFH1*, *AtFH8*, *AtFH6*, and *AtFH5* have been studied *in vivo*. For example, *AtFH1* influences pollen tube elongation, a polar cell growth process dependent on a precisely controlled actin cytoskeleton [16]. The fusion protein AtFH5-GFP exhibits a distinct concentration within the cell plate, which is crucial for cell division [17]. *AtFH8* affects root and root hair development by modifying the distribution of the actin cytoskeleton [18]. In rice, FORMIN gene expression patterns have been shown to respond to both drought and cadmium (Cd) stress [15].

Cotton is a significant crop, well known for its high-quality natural fibers that are essential to the global textile industry [19]. Additionally, cotton serves as a source of edible oil and plant proteins [20]. It is the largest genus in the Gossypieae tribe, with over 50 species, among which, *G. hirsutum* is the most widely cultivated. Native to southern Florida, the Caribbean, Mexico, and Central America [21]. *G. hirsutum* is distinguished by its broad adaptability and high yield, contributing to over 95% of global cotton production [22]. The species originated from a significant polyploidization event around 1–2 million years ago, resulting from the merging of the A and D genomes from *G. arboreum* and *G. raimondii*, respectively [23]. Despite the importance

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

of FORMIN proteins, their function has not been studied in *G. hirsutum*. Conducting wet lab experiments, to identify the FORMIN gene family, and analyze its expression is costly in terms of labor, time, and the need for well-equipped laboratories. In this study, we have systematically identified and characterized the members of the FORMIN gene family in *G. hirsutum* using integrated bioinformatics approaches to gain a better understanding of their functional roles under various physiological conditions. Each of *GhFH* members was further analyzed to determine their physiochemical properties, phylogenetic relationship, conserved domain, motifs, gene structures, *Ka/Ks* ratio, collinearity, synteny analysis, sub-cellular localization, transcription factors, protein-protein interactions, gene ontology, and *cis*-acting elements. Additionally, transcript profiling of identified *GhFH* members was conducted in response to three different light conditions and various abiotic stress conditions using RNA-seq data. Our study provides a foundation for further exploration of the FORMIN protein family in *G. hirsutum* and contributes to improving present cotton cultivar.

## 2. Materials and methods

### 2.1. Database search and retrieval of FORMIN protein sequences in *G. hirsutum* genome

The FH2 DNA-binding domains from *A. thaliana* were used to retrieve FH2 gene-encoding proteins in the *G. hirsutum* v3.1 genome from Phytozome v13 (https://phytozome-next.jgi.doe.gov/) using BLASTp (Protein-basic local alignment search tool) [24], with an expected (E) threshold value of −1, a comparison matrix (BLOSUM62), and other default parameters. The conserved FH2 domain was then identified in the retrieved amino acid sequences using the NCBI CDD (Conserved Domain Database) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) [25] SMART (Simple Modular Architecture Research Tool, (http://smart.embl-heidelberg.de/) [26] and the pfam database (http://pfam.xfam.org/) [27] at default settings. Redundant protein sequences that did not contain the FH-conserved domain were excluded from the list of candidates (S1 Data).

### 2.2. Determination of physio-chemical properties

The ProtParam online tool (http://web.expasy.org/protparam/) was used to determine the physicochemical properties of GhFH proteins, including their amino acid residue count, molecular weight, isoelectric point (pI), instability index, aliphatic index, and grand average of hydropathicity (GRAVY) [28].

### 2.3. Phylogenetic tree analysis

Protein sequences from *Gossypium hirsutum*, *Arabidopsis thaliana*, *Medicago truncatula*, *Oryza sativa*, and *Zea mays* were used to construct a phylogenetic tree (S2 Data). The MEGA11 software [29], was employed to construct the phylogenetic tree, and precise sequence alignment was performed using, the ClustalW program [30,31]. The maximum likelihood (ML) method was applied in MEGA11 software with default parameters, except for a bootstrap value of 1000 and Pearson correction. The final tree was then uploaded to iTOL v6 (https://itol.embl.de/) [32] for enhanced visual representation.

### 2.4. Analysis of gene structure

The gene structure of *GhFHs* was determined by retrieving genomic DNA and CDS sequences in FASTA format from Phytozome v13 (S3 and S4 Data). The Gene Structure Display Server (GSDS v2.0) (http://gsds.cbi.pku.edu.cn/) [33] was used for the analysis.

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

## 2.5. Conserved domain and motif analysis

The NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) was utilized to identify the typical conserved FH2 domain (pfam02181) with results displayed using the DOG2.0 software [34]. Structural motifs of GhFH protein sequences were explored using the MEME-suite tools (https://meme-suite.org/meme/meme_5.5.3/tools/meme) [35] with a maximum of 20 motifs, selected, and visualized using the TBtools software-v1.116 [36].

## 2.6. Prediction of the subcellular localization of GhFH proteins

The *in silico* subcellular localization of GhFH proteins was predicted using the WoLF PSORT online tool (https://wolfpsort.hgc.jp/) [37]. The predicted protein signals for each *GhFH* gene were visualized using RStiduo 4.2.1 software [38] with the following libraries: scales, extrafont, ggplot2, and reshape.

## 2.7. *Cis*-acting regulatory elements (CAREs) analysis of *GhFH* promoters

The 2000 bp upstream promoter region of each *GhFH* gene sequence was obtained from the Phytozome v13 database (S5 Data) for investigation of CAREs. The PlantCARE online tool (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [39] was used for CAREs analysis, with results visualized using RStiduo 4.2.1 software with the following libraries: scales, extrafont, ggplot2, and reshape.

## 2.8. *Ka/Ks* analysis

*Ks* and *Ka* values along with their substitution ratios for the *GhFH* gene family were calculated using the *Ka/Ks* calculation tool within TBtools software-v1.116. The molecular evolution rates for each set of paralogous genes were determined based on the *Ka/Ks* ratios. The formula, $T = Ks/2X$, where $X = 6.56 \times 10^{-9}$ [40], was used to calculate the duplication events and time of divergence (measured in million years ago, $MYA = 10^{-6}$) for the *GhFH* gene. Gene duplication predictions were performed using MCScanX within TBtools software-v1.116.

## 2.9. Gene ontology (GO) analysis

GO analysis was conducted to identify the functions of predicted *GhFH* genes in *G. hirsutum* using the Plant Transcription Factor Database (PlantTFDB, http://planttfdb.cbi.pku.edu.cn//) [41], with visualized using the online tool ChiPlot (https://www.chiplot.online/).

## 2.10. Collinearity and synteny analysis of *GhFH* gene family

Collinearity and synteny relationships within the cotton genome, as well as the FORMIN genes rice, maize, and *Arabidopsis* were visualized and analyzed using TBtools version-v1.116.

## 2.11. Transcription factor (TFs) analysis and regulatory network of *GhFH* genes

The online tool PlantTFDB4.0 (http://planttfdb.cbi.pku.edu.cn//) [41] was utilized to predict TFs associated with the candidate *GhFH* genes. The results were visualized using the online tool ChiPlot (https://www.chiplot.online/). The interaction network between the *GhFH* genes and TFs was constructed and illustrated using Cytoscape 3.9.1 software [42].

## 2.12. Protein-protein interaction (PPI) analysis

The STRING version 12.0 online program (https://string-db.org/) was used to predict and design the PPI network for GhFH proteins, utilizing homologous proteins from *Arabidpsois*.

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

The STRING tool was configured with specific parameters: network type-full STRING network, network edges meaning-evidence, a minimum required interaction score set at a medium confidence parameter (0.4), and a maximum display of no more than 10 interactions.

### 2.13. Identification of miRNAs targeting *GhFH* genes

CDS sequences of *GhFH* genes were uploaded to the online psRNATarget Server18 (https://www.zhaolab.org/psRNATarget/analysis?function=2) [43] with default parameters to predict putative miRNAs targeting *GhFH* genes. The interaction network between the predicted miRNAs and *GhFH* target genes was generated and visualized using Cytoscape software version 3.9.1

### 2.14. Expression analysis of *GhFH* genes in different light conditions

RNA-Seq data for *G. hirsutum* were obtained from the NCBI Sequence Read Archive (SRA) database [44] using the accession number SRA: PRJNA765172 [45]. Quality control and trimming of the RNA-Seq data were performed using the Trimmomatic v0.32 package [46]. The data then aligned to the *G. hirsutum* reference genome with STAR packages v2.7.11b [47]. Sequence alignment map (SAM) files were converted to binary alignment map (BAM) format and sorted using Samtools v1.20 [48]. FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values were calculated with the RSEM package v1.1.17 [49]. Due to significant variations in FPKM values across different *G. hirsutum* tissue samples, the values were log2 transformed. Heatmaps to visualize the expression profiles of *GhFH* genes in different light conditions were constructed using TB-Tools software v1.116.

### 2.15. Expression analysis of *GhFH* genes in various tissues and abiotic stresses

RNA-Seq raw reads for *G. hirsutum* were downloaded from the NCBI SRA database with the accession number SRA: PRJNA248163 [50]. Quality controlled and filtering of the RNA-Seq raw reads were performed using trimmomatic package version 0.32 and the reads were mapped to the *G. hirsutum* reference genome using the Bowtie2 package [51]. SAM files were transformed to BAM format and sorted using Samtools packages version 1.20. FPKM values were calculated using the RSEM package v1.1.17. Due to large differences in FPKM values among different tissues of *G. hirsutum*, the FPKM values were transformed to log2. Heatmaps to visualize the expression profiles of *GhFH* genes in different tissues and stress conditions were constructed using TB-Tools software v1.116.

## 3. Results

### 3.1. Analysis of physicochemical properties of GhFH

The FORMIN gene family is recognized as essential for the growth, function, and development of plants. By regulating both the microtubule and actin cytoskeleton, FORMINs contribute to vital cellular processes including cytokinesis, cell polarity, and cell migration [17,52]. The presence of the typical FH2 domain is used to identify the FORMIN family in plants [11]. In the *G. hirsutum* genome, 46 FORMIN genes were identified and designated as *GhFH1-GhFH46*. The physiochemical analysis of the protein sequences of *GhFH* genes revealed notable differences in the chemical and physical properties among the members of the GhFH family (Table 1). The amino acid count in GhFH proteins ranged from 402 to 4461. Significant variations in molecular weights were observed, ranging from 101138.61 kDa (GhFH34) to 375608.96 kDa (GhFH43). Furthermore, the isoelectric points (pI) of these 46 GhFH proteins exhibited diversity, ranging from 4.69 to 5.32.

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

**Table 1. List of 46 *GhFH* genes and their basic physiochemical characterization.**

| SI | Gene identifier | Gene name | Size (aa) | Mass (kDa) | pI | Instability index | Aliphatic index | GRAVY |
|---|---|---|---|---|---|---|---|---|
| 1 | Gohir.A02G007500 | *GhFH1* | 2799 | 237323.45 | 4.82 | 49.77 | 27.58 | 0.917 |
| 2 | Gohir.A03G143800 | *GhFH2* | 2376 | 196496.88 | 4.89 | 42.06 | 30.01 | 0.845 |
| 3 | Gohir.A04G022700 | *GhFH3* | 2778 | 231720.54 | 4.86 | 43.76 | 29.59 | 0.831 |
| 4 | Gohir.A04G089001 | *GhFH4* | 444 | 37134.29 | 5.32 | 25.36 | 29.73 | 0.493 |
| 5 | Gohir.A05G059300 | *GhFH5* | 3864 | 325089.78 | 4.78 | 45.98 | 28.67 | 0.815 |
| 6 | Gohir.A05G415300 | *GhFH6* | 2037 | 167794.83 | 4.93 | 44.33 | 30.09 | 0.811 |
| 7 | Gohir.A07G070200 | *GhFH7* | 3702 | 308059.04 | 4.80 | 45.28 | 27.93 | 0.776 |
| 8 | Gohir.A07G175778 | *GhFH8* | 444 | 37148.32 | 5.32 | 25.55 | 29.95 | 0.498 |
| 9 | Gohir.A07G205332 | *GhFH9* | 639 | 52886.37 | 5.26 | 26.69 | 29.42 | 0.503 |
| 10 | Gohir.A08G020008 | *GhFH10* | 2844 | 237904.89 | 4.86 | 44.84 | 29.01 | 0.798 |
| 11 | Gohir.A08G228101 | *GhFH11* | 3765 | 318539.36 | 4.77 | 45.53 | 26.75 | 0.836 |
| 12 | Gohir.A09G242700 | *GhFH12* | 4281 | 358550.88 | 4.75 | 50.02 | 28.15 | 0.825 |
| 13 | Gohir.A10G146300 | *GhFH13* | 2742 | 231853.47 | 4.82 | 48.62 | 26.55 | 0.879 |
| 14 | Gohir.A11G032500 | *GhFH14* | 1890 | 157618.91 | 4.91 | 55.96 | 28.10 | 0.904 |
| 15 | Gohir.A11G144300 | *GhFH15* | 3357 | 282986.80 | 4.79 | 43.17 | 25.89 | 0.820 |
| 16 | Gohir.A11G215800 | *GhFH16* | 2772 | 229193.24 | 4.86 | 46.65 | 28.93 | 0.817 |
| 17 | Gohir.A12G076200 | *GhFH17* | 2697 | 222787.26 | 4.86 | 44.49 | 29.55 | 0.834 |
| 18 | Gohir.A12G082833 | *GhFH18* | 444 | 37092.27 | 5.32 | 25.16 | 29.73 | 0.500 |
| 19 | Gohir.A12G135200 | *GhFH19* | 2466 | 202452.31 | 4.90 | 42.26 | 30.78 | 0.811 |
| 20 | Gohir.A12G179700 | *GhFH20* | 4149 | 348141.31 | 4.74 | 48.73 | 27.26 | 0.865 |
| 21 | Gohir.A13G108800 | *GhFH21* | 2424 | 200670.50 | 4.91 | 43.42 | 33.42 | 0.842 |
| 22 | Gohir.A13G141900 | *GhFH22* | 2844 | 236172.28 | 4.86 | 43.69 | 29.29 | 0.787 |
| 23 | Gohir.A13G184401 | *GhFH23* | 3543 | 295168.92 | 4.80 | 43.42 | 29.16 | 0.828 |
| 24 | Gohir.D02G007300 | *GhFH24* | 2796 | 237260.26 | 4.82 | 50.11 | 27.58 | 0.912 |
| 25 | Gohir.D02G110350 | *GhFH25* | 402 | 33565.70 | 5.25 | 36.66 | 33.58 | 0.871 |
| 26 | Gohir.D02G166900 | *GhFH26* | 2409 | 199719.82 | 4.88 | 42.35 | 29.80 | 0.853 |
| 27 | Gohir.D04G000900 | *GhFH27* | 2046 | 168576.68 | 4.93 | 45.12 | 29.96 | 0.808 |
| 28 | Gohir.D05G061366 | *GhFH28* | 1185 | 96453.43 | 5.09 | 37.45 | 31.81 | 0.715 |
| 29 | Gohir.D05G364800 | *GhFH29* | 2784 | 232389.07 | 4.86 | 41.91 | 29.13 | 0.816 |
| 30 | Gohir.D06G097600 | *GhFH30* | 3813 | 319654.53 | 4.79 | 45.91 | 28.93 | 0.819 |
| 31 | Gohir.D07G074800 | *GhFH31* | 3879 | 323119.72 | 4.78 | 45.42 | 28.44 | 0.802 |
| 32 | Gohir.D07G094100 | *GhFH32* | 3111 | 261853.51 | 4.81 | 48.85 | 28.22 | 0.871 |
| 33 | Gohir.D08G030700 | *GhFH33* | 2856 | 239230.54 | 4.85 | 44.01 | 28.68 | 0.798 |
| 34 | Gohir.D08G248366 | *GhFH34* | 1239 | 101138.61 | 5.08 | 32.93 | 31.96 | 0.714 |
| 35 | Gohir.D09G243400 | *GhFH35* | 3456 | 290291.18 | 4.78 | 53.75 | 28.12 | 0.886 |
| 36 | Gohir.D10G120700 | *GhFH36* | 2757 | 233618.67 | 4.82 | 49.47 | 27.02 | 0.890 |
| 37 | Gohir.D11G035801 | *GhFH37* | 3867 | 341003.22 | 4.69 | 72.58 | 23.95 | 1.080 |
| 38 | Gohir.D11G150700 | *GhFH38* | 3306 | 278574.76 | 4.80 | 44.20 | 25.86 | 0.818 |
| 39 | Gohir.D11G207500 | *GhFH39* | 2772 | 229257.48 | 4.86 | 45.77 | 29.08 | 0.824 |
| 40 | Gohir.D11G277200 | *GhFH40* | 2613 | 216470.99 | 4.89 | 40.47 | 31.19 | 0.805 |
| 41 | Gohir.D12G074600 | *GhFH41* | 2694 | 222436.00 | 4.86 | 44.62 | 29.92 | 0.844 |
| 42 | Gohir.D12G139300 | *GhFH42* | 2484 | 204158.41 | 4.89 | 42.76 | 30.68 | 0.816 |
| 43 | Gohir.D12G182300 | *GhFH43* | 4461 | 375608.96 | 4.72 | 49.42 | 26.97 | 0.869 |
| 44 | Gohir.D13G111700 | *GhFH44* | 2418 | 200135.98 | 4.91 | 42.58 | 33.58 | 0.847 |
| 45 | Gohir.D13G146600 | *GhFH45* | 2880 | 239035.56 | 4.86 | 43.70 | 29.24 | 0.790 |
| 46 | Gohir.D13G190601 | *GhFH46* | 3540 | 295031.65 | 4.80 | 43.10 | 29.18 | 0.824 |

MW, molecular weight (kDa); aa, number of amino acids; pI, theoretical isoelectric point; GRAVY, grand average of hydropathicity. Instability index (<40, the protein is stable >40, the protein is unstable) [53].

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

The instability index revealed that the GhFH18 protein had the lowest value at 25.16, while the GhFH37 protein showed the highest instability index value at 72.58. The aliphatic index values ranged from 23.95 to 33.58, with an average value of 29.09. Additionally, all GhFH proteins exhibited an average hydropathicity of less than 1, except for GhFH37, which scored 1.08.

## 3.2. Analysis of phylogenetic tree

To explore the conservation and evolutionary relationships of FORMIN proteins across different species, a phylogenetic tree was constructed using 46 proteins from *G. hirsutum*, 21 proteins from *A. thaliana*, 19 proteins from *M. truncatula*, 17 proteins from *O. sativa*, and 20 proteins from *Z. mays* (Fig 1). Proteins located closer together within a cluster on the phylogenetic tree, exhibit a higher degree of functional similarity [54]. The phylogenetic analysis of the 46 *G. hirsutum* proteins separated them into five distinct groups (A, B, C, D, and E). Group B contained the highest number of GhFH proteins with 19, while group D contained the lowest number of proteins with 3. Groups A, C, and E contained 4, 8, and 12 GhFH proteins, respectively (S6 Data). According to the phylogenetic tree GhFH6, GhFH2, GhFH3, and GhFH13 from G. *hirsutum* were closely related to MtFH13, MtFH3, MtFH1, and MtFH9 from M. *truncatula's*, respectively. Additionally, ZmFH10 exhibited a close relation with OsFH16, ZmFH18 with OsFH12, ZmFH7 with OsFH10, ZmFH11 with OsFH11, ZmFH9 with OsFH9, ZmFH20 with OsFH13, and ZmFH15 with OsFH2..

## 3.3. Analysis of *GhFH* gene structure

The ability to encode proteins and perform cellular function is determined by its structure [55]. The gene structure analysis of 46 *GhFH* genes revealed that all genes contained exons and introns but lacked upstream/downstream regions (Fig 2). The analysis showed that *GhFH46* had the longest gene length among the studied genes, while *GhFH4* and *GhFH8* had the shortest gene lengths. *GhFH7* and *GhFH11* had the highest number of exon sequences with 18 and intron sequences with 17 (S7 Data). Interestingly, *GhFH4* and *GhFH8* genes shared the same structure and length of approximately 2.0 kb. The genes in group A (*GhFH2*, *GhFH26*, *GhFH27*, *GhFH6*) were found to have similar structure.

## 3.4. Analysis of conserved domains in GhFH

Domain unravels the structure, function, and evolution of proteins [56]. In our analysis, we mainly focused on the typical FH2 domain along with PTEN_C2, PRIMA1, STE3, Remorin_c, and Mid2, respectively (Fig 3). The proteins of group A were found to contain only the FH2 domain.

## 3.5. Analysis of GhFH motifs

Motifs serve as representations of structural components, active sites, transcription binding sites, and splice junctions within genetic sequences [57]. In our analysis, 20 different motifs present in the GhFH peptide sequence were identified (S1 Fig). Closely related proteins within each phylogenetic group were found to share identical motif compositions (Fig 4). GhFH25 was found to contain only two motifs (Motif7, Motif5). Additionally, GhFH8, GhFH9, GhFH4, GhFH18, GhFH34, and GhFH28 exhibited different motifs from others present in group B, which may have some functional implications.

## 3.6. Analysis of subcellular localization of GhFH proteins

The cellular compartment in which a protein is present strongly influences its functions and actions [58]. In this study, GhFH proteins were mainly located in three major organelles: chloroplasts, mitochondria, and the nucleus. Almost all GhFH proteins are localized in
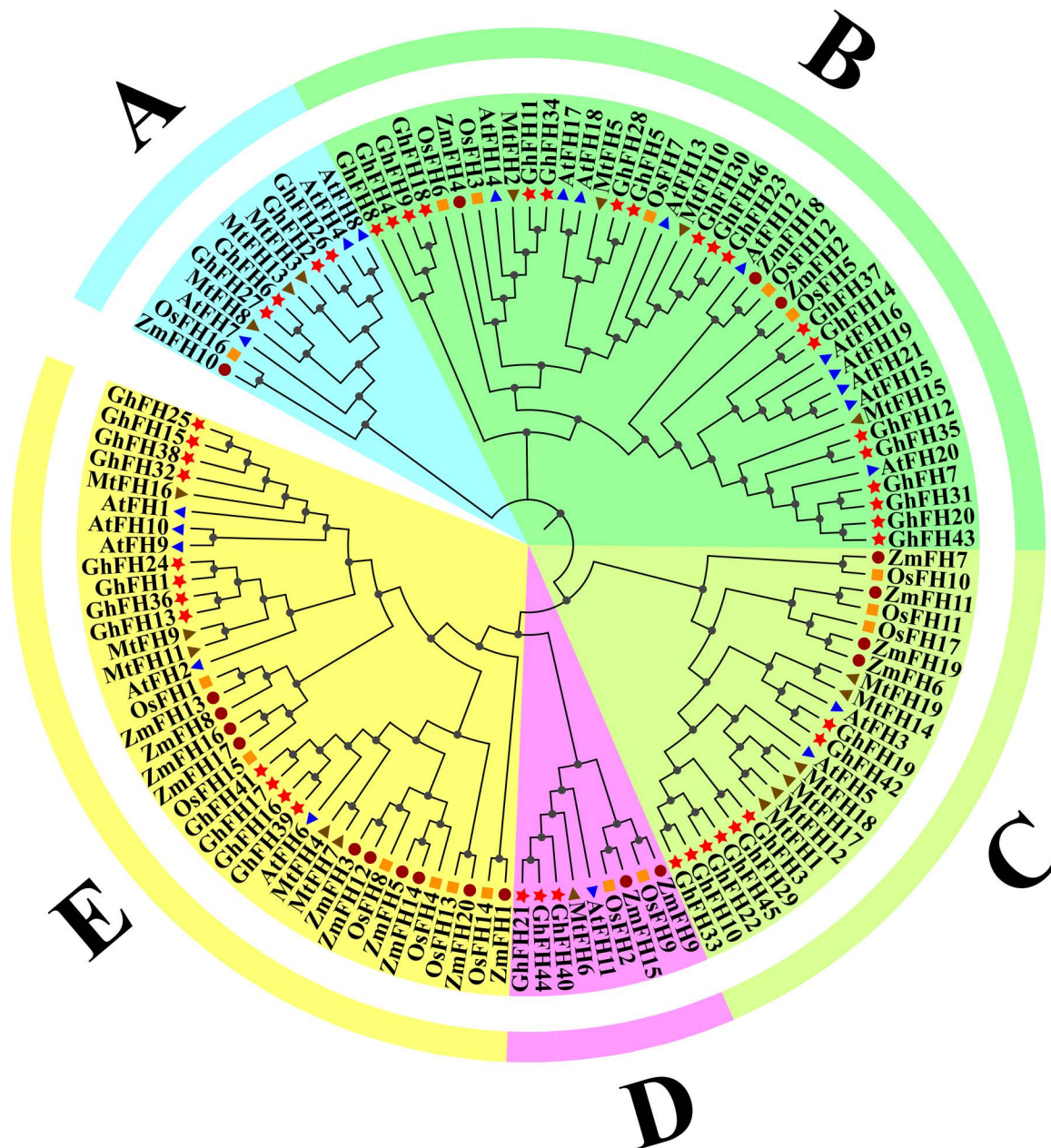
PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

**Fig 1. Phylogenetic tree analysis of FORMIN family members.** Phylogenetic tree displaying the evolutionary relationships of FORMIN proteins based on the FH2 domain from *G. hirsutum, A. thaliana, M. truncatula, O. sativa,* and *Z. mays.* All the FH members were divided into 5 groups and presented in different colors. The red star represented GhFH proteins, the blue triangle represented AtFH proteins, the brown triangle represented MtFH proteins, the orange square represented OsFH proteins and the pink circle represented ZmFH proteins.

https://doi.org/10.1371/journal.pone.0319176.g001

chloroplasts, except for GhFH1, GhFH5, GhFH10, GhFH19, GhFH24, GhFH30, GhFH33, GhFH35, GhFH40, GhFH42, and GhFH45 (Fig 5A). This analysis revealed that 76.5% of GhFH proteins were present in chloroplasts, 60.84% within the nucleus, 36.94% in the plasma membrane, 34.77% in the endoplasmic reticulum (E.R), 60.84% in mitochondria, 43.46% in the Golgi, 28.25% in the cytoplasm, 23.90% in the cytoskeleton, 47.81% in vacuoles, and 34.77% in the extracellular membrane (Fig 5B), providing a complete illustration of their distribution.
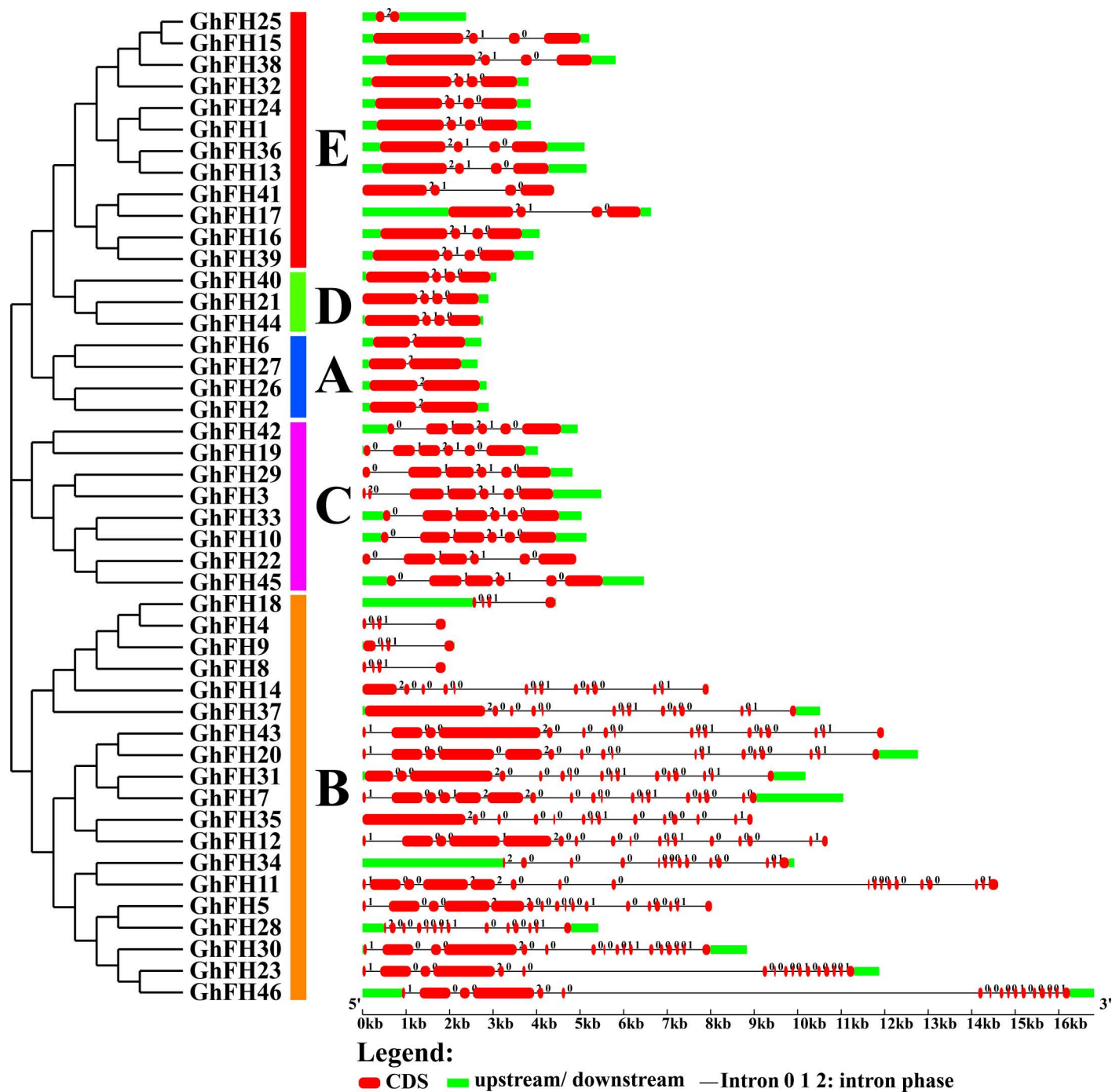
PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)



**Fig 2. Gene structure of *GhFH* genes.** Gene structure analyses for *GhFH* genes were carried out using the Gene Structure Display Server (GSDS 2.0, http://gsds.cbi.pku.edu.cn/index.php). The lengths of exons and introns for each *GhFH* gene are demonstrated proportionally. Gene groups are categorized and colored based on their phylogenetic relationships. For all *GhFH* genes, black lines represent introns, red-bold lines represent exons, and light-green lines represent 5' and 3' untranslated regions (UTR). The structure of each *GhFH* gene exon/intron is displayed proportionally according to the scale mentioned.
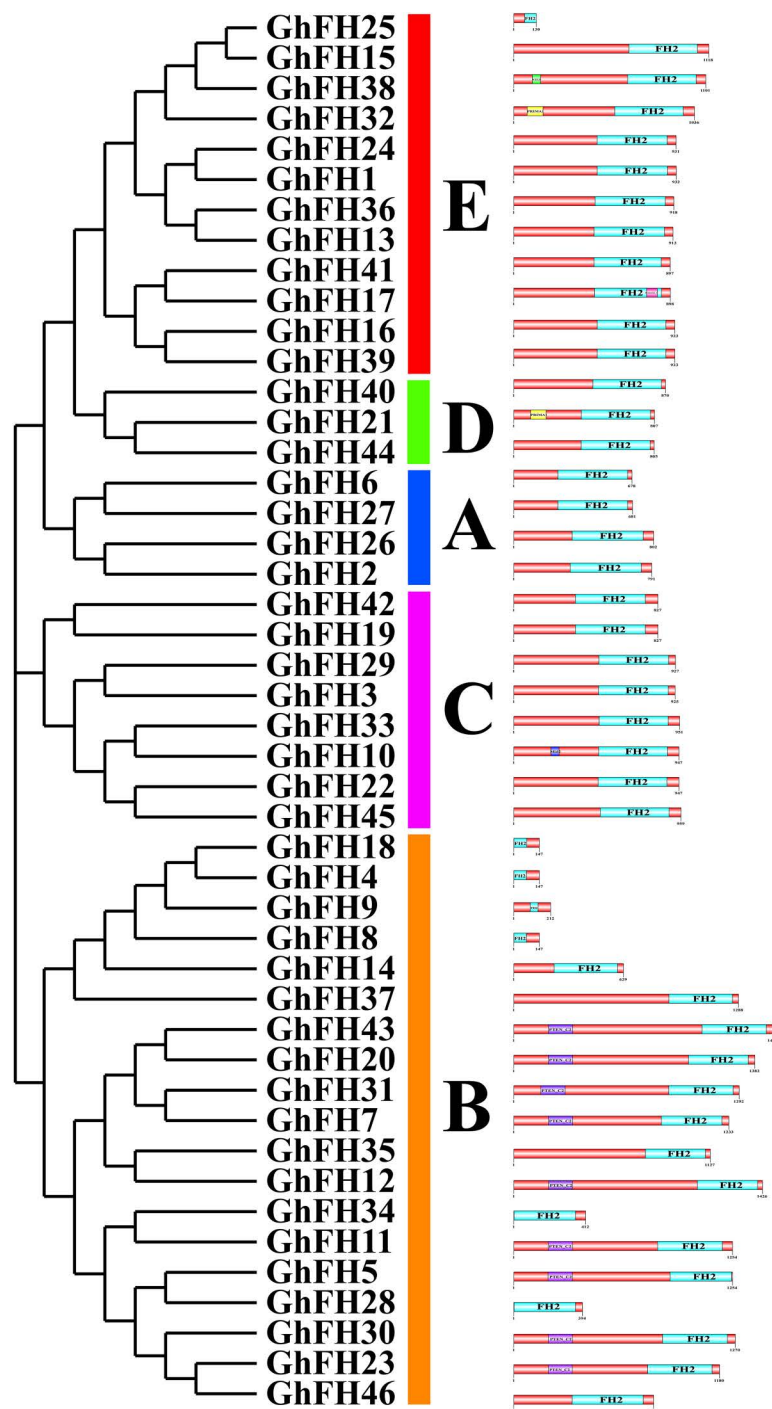
https://doi.org/10.1371/journal.pone.0319176.g002

**Fig 3. Conserved domains of GhFH proteins.** The positions of each conserved domain are demonstrated in differently colored boxes, with the domain names.

https://doi.org/10.1371/journal.pone.0319176.g003

## 3.7. Analysis of *cis*-acting regulatory elements (CAREs) of *GhFH* promoters

The 2000 base pairs upstream of the 5′ end of the 46 *GhFH* genes were analyzed to understand their possible regulatory mechanisms. A total of 62 CAREs that play pivotal roles in light
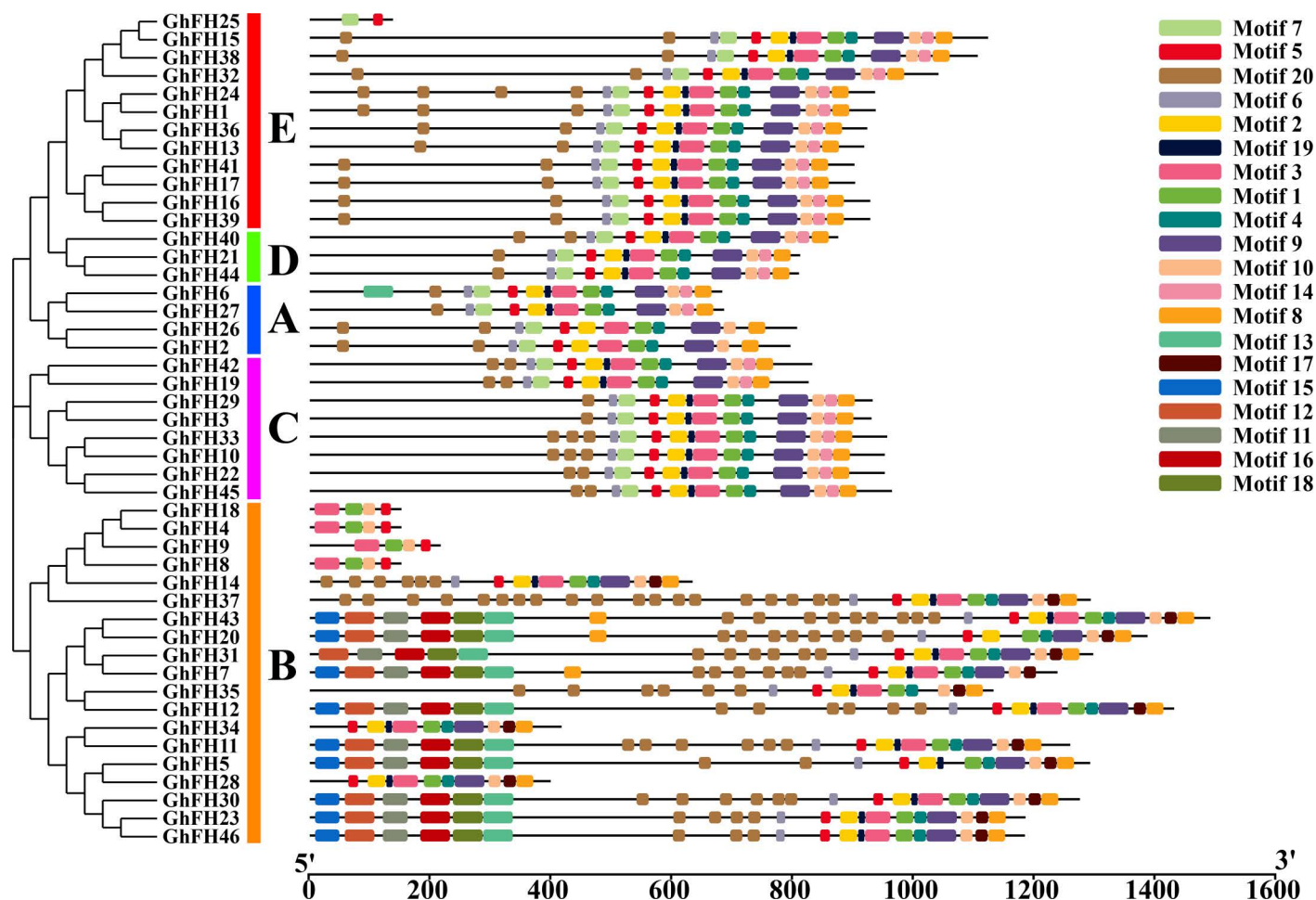
PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

**Fig 4. Distribution of conserved motifs of GhFH proteins.** The identification of conserved motifs in GhFH proteins was carried out using the Multiple EM for Motif Elicitation (MEME) (https://meme-suite.org/meme/tools/meme) tool, with a maximum of 20 motifs selected. Each motif is represented by a specific-colored box aligned on the right side of the figure. Different colors indicate individual motifs identified within each protein domain.

https://doi.org/10.1371/journal.pone.0319176.g004

response, phytohormone sensitivity, tissue-specific expression, and stress responsiveness were identified in the *G. hirsutum* genome (Fig 6). Among these CAREs, 5 were stress-responsiveness, including DREs, MBSs, LTRs, WUN motif-containing elements, and TC-rich repeats while 17 were related to tissue-specific expression. Notable tissue-specific elements including the 3-AF3 binding site, AT-rich element, Box II -like sequence, A-box, CAT-box, ARE, CCAAT-box, NON-box, circadian, MBSI, GCN4_motif, MSA-like, HD-Zip 1, HD-Zip 3, motif I, O2-site, and RY-element (S8 Data). AREs were found to be the most prevalent elements in the promoters of *GhFH* genes. Additionally, 11 phytohormone-related elements were identified, including ABREs, CGTCA motif-containing elements, AuxRR-core-containing elements, P-boxes, GARE motif-containing elements, SAREs, TCA elements, TATC-boxes, TGACG motif-containing elements, TGA-boxes, and TGA elements. These elements respond to abscisic acid (ABREs), auxin (AuxRR-core-containing elements, TGA motif-containing elements, and TGA-boxes), methyl jasmonate (MeJA) (CGTCA motif-containing elements and TGACG motif-containing elements), gibberellin (GARE motif-containing elements, TATC motif-containing elements, and P-boxes), and salicylic acid (SAREs and TCA elements). Moreover, 29 cis-regulatory elements were associated with light response, including
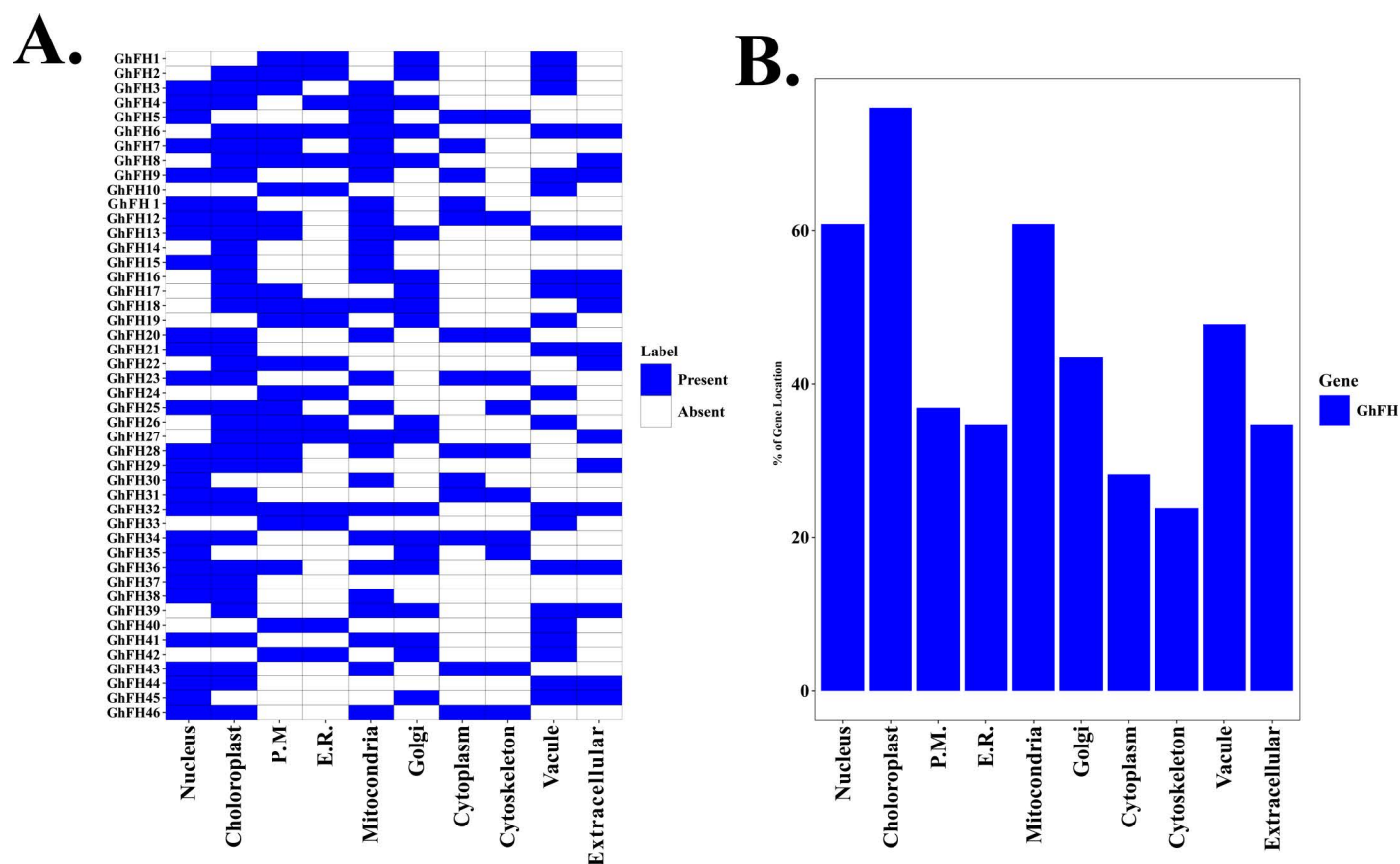
PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

**A.**

**B.**



**Fig 5. Subcellular localization of GhFH proteins.** The subcellular distribution of GhFH proteins is shown in a heatmap. A. The relevant cellular organelles are displayed at the bottom of the heatmap, and the names of each GhFH protein are displayed on the left side. The presence of protein signals corresponding to the genes is shown by a blue color on the heatmap. B. A bar diagram illustrates the percentage distribution of the GhFH protein signal across different cellular organelles. The percentages of protein signals that exist in various cellular organelles are displayed on the left side.

https://doi.org/10.1371/journal.pone.0319176.g005

C-boxes, Box 4 elements, G-boxes, and more. Notably, GT1 motif elements and G-boxes were particularly abundant among these light-responsive cis-regulatory elements, highlighting their significance in regulating gene expression under varying light conditions.

### 3.8. Analysis of *Ka/Ks* of *GhFH* genes

The *Ka/Ks* analysis is a crucial indicator for understanding the evolution of gene duplication after separation from ancestors. A *Ka/Ks* value of 1 signifies neutral selection, while a value below 1 implies purifying selection and a value above 1 indicates positive or Darwinian selection [59]. The number of *Ka* and *Ks* along with their ratio *Ka/Ks*, were analyzed for 20 duplicated gene pairs. *Ks* values for these gene pairs varied from 0.01 (*GhFH4-GhFH8*) to 0.68 (*GhFH15-GhFH32*) with an average *Ks* of 0.08 (Fig 7). Among the 20 duplicated pairs, 19 exhibited a *Ka/Ks* ratio below 1, indicating purifying selection, except for one exceptional pair (*GhFH12-GhFH35*) with a ratio of 1.06 (S9 Data), suggesting positive selection. Additionally, *Ks* values were used to estimate the timing of gene duplication events (*GhFH* genes) in the evolutionary history of the cotton genome. Segmental and tandem duplication events in cotton were estimated to have occurred over a period ranging from 0.822 to 52.50 MYA, averaging 6.10 MYA.
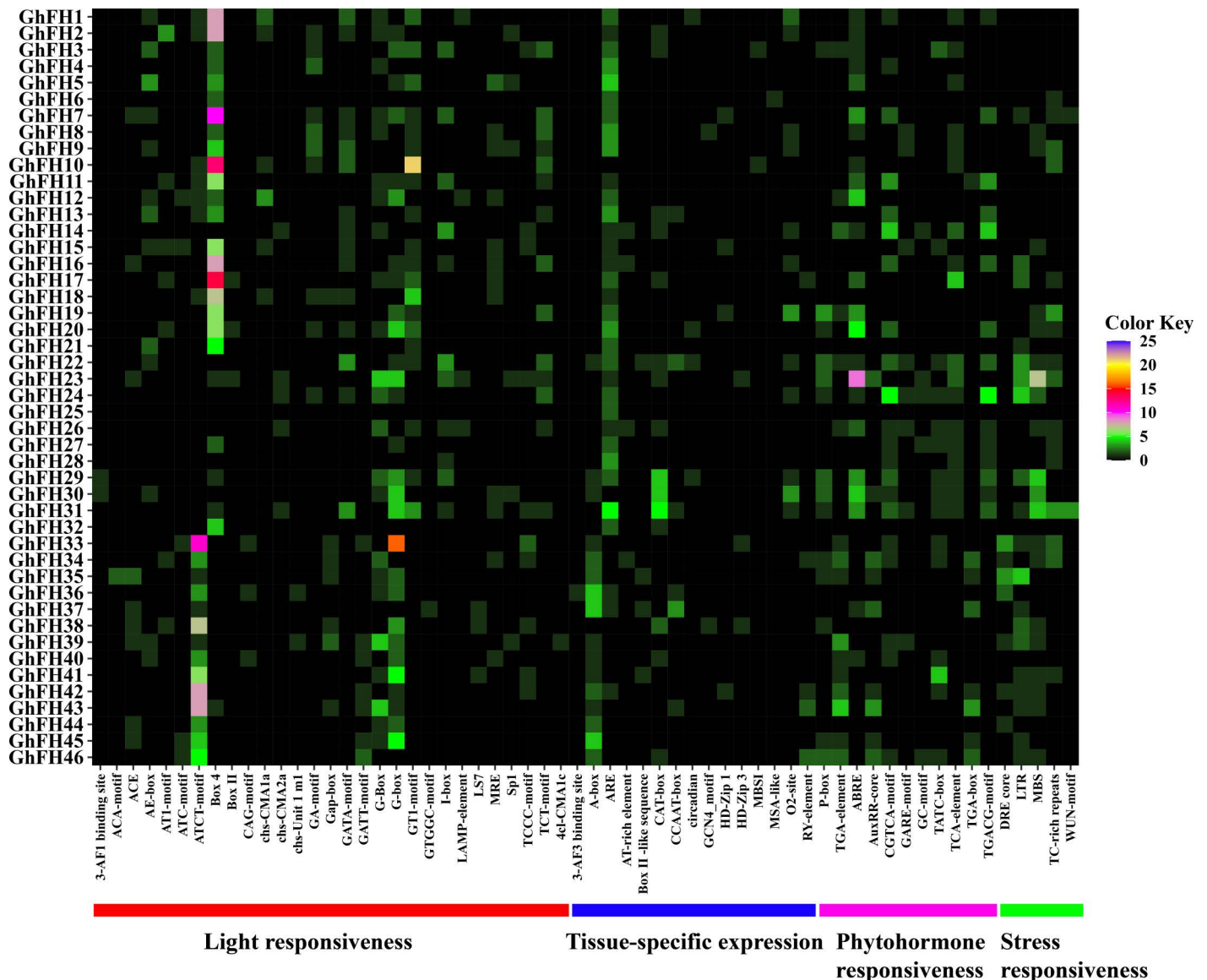
PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

**Fig 6. The distribution of CAREs in the 2.0 kb promoter region of *GhFH* genes is represented by a heatmap.** The names of each *GhFH* gene are shown on the left side of the heatmap. The number of putative CAREs for each *GhFH* gene is displayed on the bottom of the heatmap with a color scale (0–25) on the right side of the heatmap. Functions associated with CAREs of the corresponding genes, such as light responsiveness, tissue-specific expression, phytohormone responsiveness, and stress responsiveness, are indicated by bold lines in red, blue, pink, and green at the bottom of the heatmap, respectively.

https://doi.org/10.1371/journal.pone.0319176.g006

### 3.9. Analysis of gene ontology (GO) of *GhFH* genes

GO analysis was performed to investigate the regulatory pathways and functions of the identified *GhFH* genes. A total of 59 unique GO IDs were identified each with their respective *p*-value. The identified GO terms were categorized into three groups: molecular functions (F), biological processes (P), and cellular components (C) (S10 Data). Among these, the biological processes group contained 48 GO terms (Fig 8). For instance, (GO: 0051258, *p*-value: 0.0000000074) coordinates the process of protein polymerization. Additionally, (GO:0030036, *p*-value: 0.000000014) controls the organization of the actin cytoskeleton, a necessary component for intracellular mobility and the maintenance of cell shape. Organelle organization is directed by (GO: 0006996,
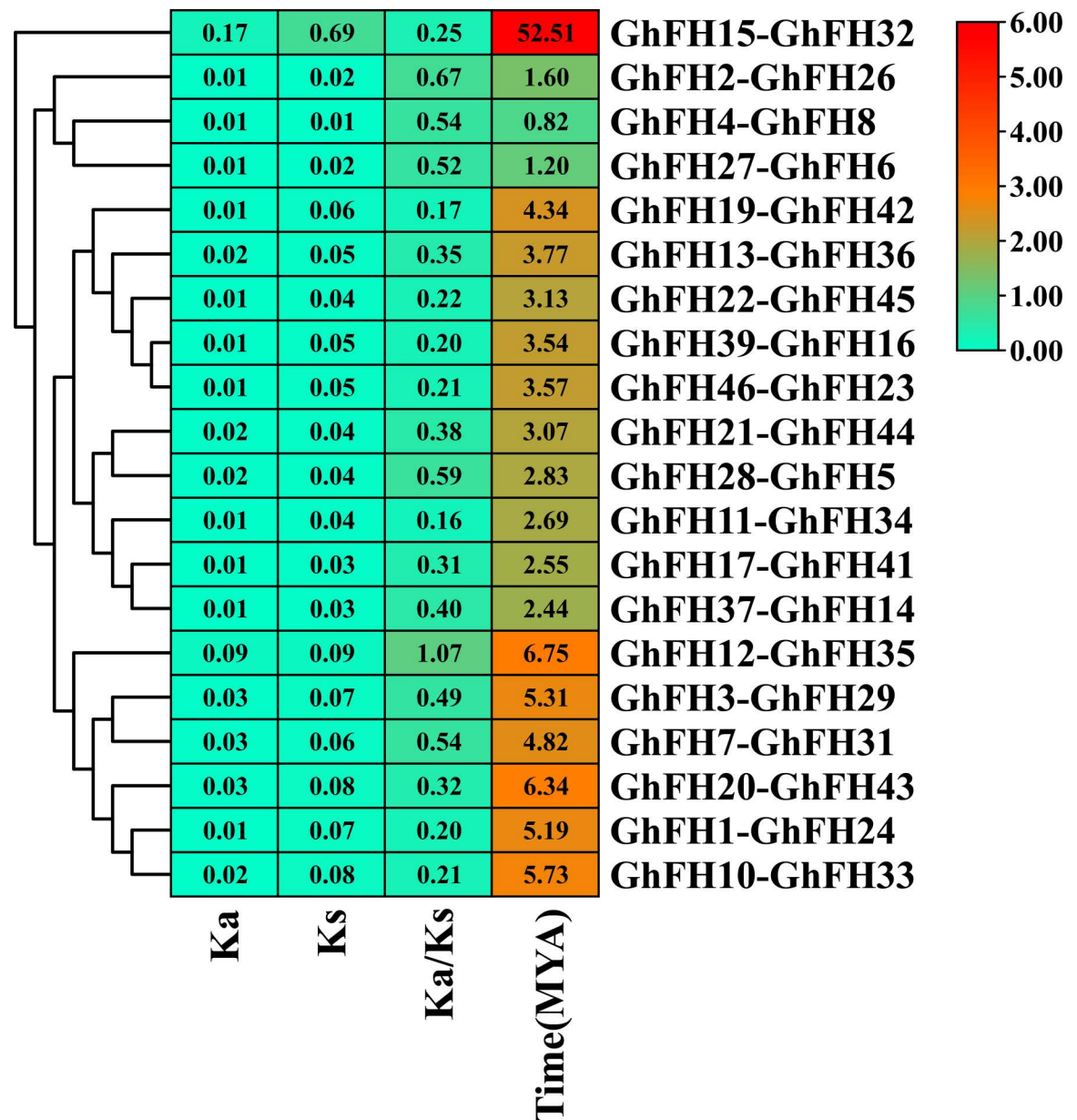
| Ka | Ks | Ka/Ks | Time(MYA) | |
|---|---|---|---|---|
| 0.17 | 0.69 | 0.25 | 52.51 | GhFH15-GhFH32 |
| 0.01 | 0.02 | 0.67 | 1.60 | GhFH2-GhFH26 |
| 0.01 | 0.01 | 0.54 | 0.82 | GhFH4-GhFH8 |
| 0.01 | 0.02 | 0.52 | 1.20 | GhFH27-GhFH6 |
| 0.01 | 0.06 | 0.17 | 4.34 | GhFH19-GhFH42 |
| 0.02 | 0.05 | 0.35 | 3.77 | GhFH13-GhFH36 |
| 0.01 | 0.04 | 0.22 | 3.13 | GhFH22-GhFH45 |
| 0.01 | 0.05 | 0.20 | 3.54 | GhFH39-GhFH16 |
| 0.01 | 0.05 | 0.21 | 3.57 | GhFH46-GhFH23 |
| 0.02 | 0.04 | 0.38 | 3.07 | GhFH21-GhFH44 |
| 0.02 | 0.04 | 0.59 | 2.83 | GhFH28-GhFH5 |
| 0.01 | 0.04 | 0.16 | 2.69 | GhFH11-GhFH34 |
| 0.01 | 0.03 | 0.31 | 2.55 | GhFH17-GhFH41 |
| 0.01 | 0.03 | 0.40 | 2.44 | GhFH37-GhFH14 |
| 0.09 | 0.09 | 1.07 | 6.75 | GhFH12-GhFH35 |
| 0.03 | 0.07 | 0.49 | 5.31 | GhFH3-GhFH29 |
| 0.03 | 0.06 | 0.54 | 4.82 | GhFH7-GhFH31 |
| 0.03 | 0.08 | 0.32 | 6.34 | GhFH20-GhFH43 |
| 0.01 | 0.07 | 0.20 | 5.19 | GhFH1-GhFH24 |
| 0.02 | 0.08 | 0.21 | 5.73 | GhFH10-GhFH33 |

**Fig 7. Estimation of gene duplication time for different gene pairs among *GhFH* genes, based on *Ks* and *Ka* values.** Gene duplication analyses were conducted using TBtools software version-v1.116. *Ka* represents the number of nonsynonymous substitutions per site, while *Ks* represents the number of synonymous substitutions per site. The ratio of nonsynonymous (*Ka*) to synonymous (*Ks*) changes is represented by *Ka/Ks*.

https://doi.org/10.1371/journal.pone.0319176.g007

*p*-value: 0.000054) and the regulation of biological processes is controlled by (GO: 0050789, *p*-value: 0.0072). The cellular components and molecular functions categories exhibited 5 GO terms each including (GO: 0071944, *p*-value: 0.000029), (GO: 0009524, *p*-value: 0.000074), (GO: 0005886, *p*-value: 0.00029), (GO: 0005618, *p*-value: 0.0004), (GO: 0030312, *p*-value: 0.0004) for cellular components, and (GO: 0005515, *p*-value: 0.00000016), (GO: 0051015, *p*-value: 0.000014), (GO: 0003779, *p*-value: 0.0011), (GO: 0032403, *p*-value: 0.0033), (GO: 0008092, *p*-value: 0.0072), and (GO: 0044877, *p*-value: 0.0076) for molecular functions. Notably, (GO: 0030312, *p*-value:
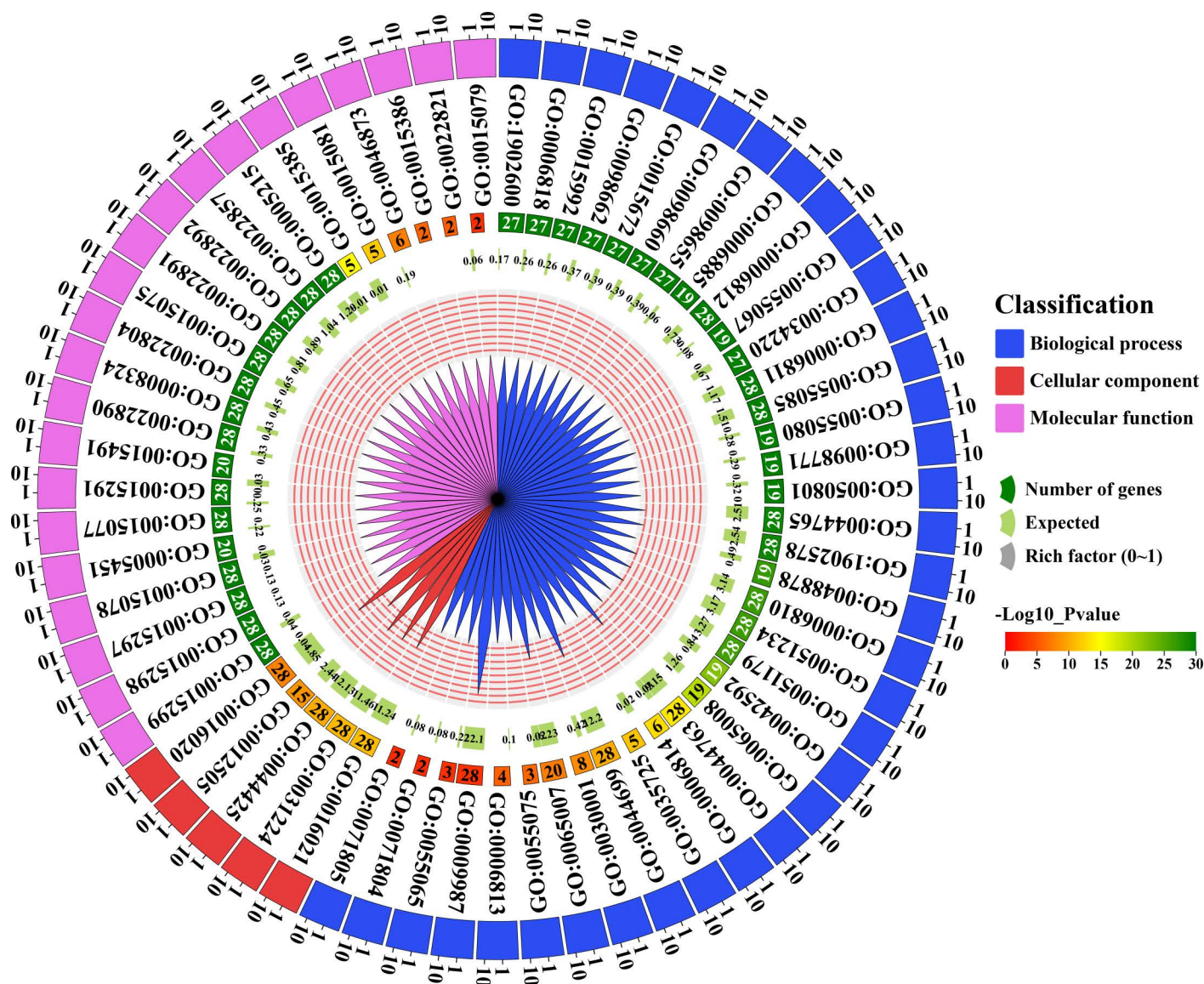
PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)



**Fig 8. GO analysis of the FORMIN gene family in *G. hirsutum*.** Circular heatmap for the predicted GO terms corresponding to the predicted *GhFH* genes presented for biological process, cellular components, and molecular function, whether the genes are associated or not. The *p*-value matching the GO terms is shown in the heatmap, using log10 (*p*-value).

0.0004) is associated with the external encapsulating structure while cytoskeletal protein binding is regulated by (GO: 0008092, *p*-value: 0.0072). Actin binding and actin filament binding are characterized by (GO: 0003779, *p*-value: 0.0011) and (GO: 0051015, *p*-value: 0.000014) respectively. Interestingly, 5 *GhFH* genes are linked to the cell periphery (GO:0071944, *p*-value: 0.000029) in the cellular components, and 12 *GhFH* genes can function as "protein binding (GO: 0005515, *p*-value: 0.00000016)" in the molecular functions category.

### 3.10. Collinearity analysis of *GhFH* genes

The origins of the *GhFH* gene in *G. hirsutum* and their relationship with other homologous genes were determined through collinearity analysis. 46 *GhFH* genes were distributed

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

unevenly across the 12 chromosomes. A total of 20 collinear pairs were identified among the 46 genes in *G. hirsutum* (Fig 9). Specifically, *GhFH2* was found to exhibit a collinear relationship with *GhFH26*, *GhFH16* with *GhFH27*, *GhFH3* with *GhFH29*, *GhFH7* with *GhFH38*, *GhFH4* with *GhFH8*, among. However, *GhFH30* located on chromosome 6, did not show any collinear pairing.

### 3.11. Analysis of synteny of *GhFH* genes with other plant species

To explore the potential evolutionary connection between FH members found in different plant species, a synteny analysis was conducted between *G. hirsutum, A. thaliana, Z. mays, and* O. *sativa* (Fig 10). The analysis revealed that *G. hirsutum* lacked orthologous FH genes but displayed 20 pairs of paralogous genes. In contrast,7 orthologous FH gene pairs were identified between *Z. mays* and *O. sativa*. For instance, *ZmFH10* showed a syntenic relationship with *OsFH16*, *ZmFH18* with *OsFH12*, *ZmFH7* with *OsFH10*, *ZmFH11* with *OsFH11*, *ZmFH9* with *OsFH9*, *ZmFH20* with *OsFH13*, and *ZmFH15* with *OsFH2*, respectively. Additionally, *A. thaliana* exhibited only 2 paralogous FH gene pairs, highlighting unique evolutionary patterns within this species.

### 3.12. Analysis of transcription factor (TFs) and regulatory network of *GhFH*

TFs are the master regulators that control gene expression by binding to specific DNA sequences [60]. These TFs play crucial roles in how plants respond to various environmental challenges, including biotic and abiotic stresses, and regulate essential processes such as metabolism, growth, and development. Additionally, TFs coordinate plant defense mechanisms against a wide range of microbial pathogens [61–65]. In plants, various TF families exist, including ERF, MYB, bZIP, GATA, CBF/DREB1, SBP, G2-like, NAC, LBD, HSF, E2F/DP, C2H2, AP2/EREBP, TALE, WRKY, Dof, MIKC_MADS, BBR-BPC, TGA6, C2H2, BOS1 families, and others. These TFs serve as the master regulators, coordinating gene expression in response to environmental challenges, developmental cues, and internal signals [63,66,67].

In this study, among the 1581 TFs, 43 distinct TFs were identified to regulate *GhFH* genes. These TFs belong to 11 different families (ERF, C2H2, GATA, LBD, MYB, TALE, E2F/DP, BBR-BPC, G2-likeHSF, and SBP) Among these the ERF, GATA, MYB, and LBD families, comprising 36 TFs are believed to play a crucial role in the regulation of *GhFH* genes. Notably, these four TFs families contained 28, 4, 2, and 2 TFs, respectively, accounting for 83.72% of the 43 identified TFs. The interactions between the predicted *GhFH* genes and the major TF families (ERF, C2H2, GATA, LBD, MYB, TALE, E2F/DP) were also analyzed (Fig 11A and S11 Data).

The fundamental mechanisms of biological activities, including organ formation and homeostasis [68], stress response [69], plant defense [70], and signal transduction [71] rely on interaction networks. ERF family was found to be linked to all of the *GhFH* genes, except *GhFH4* and *GhFH18*. The C2H2 family was the second most prevalent, with a single TF (Gh_A05G3867) interacting with 20 *GhFH* genes. Additionally, 17 *GhFH* genes were associated with the LBD protein family and 12 *GhFH* genes were involved in interactions with the GATA family. The TALE family interacted with 3 *GhFH* genes, while the MYB TF family connected with 5 *GhFH* genes. Furthermore, the E2F/DP family was found to interact exclusively with *GhFH4* and *GhFH18* (Fig 11B).

### 3.13. Analysis of protein-protein interaction (PPI)

A PPI network analysis of GhFH proteins was conducted based on known *Arabidopsis* proteins. GhFH proteins that share significant similarities with *Arabidopsis* proteins

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)
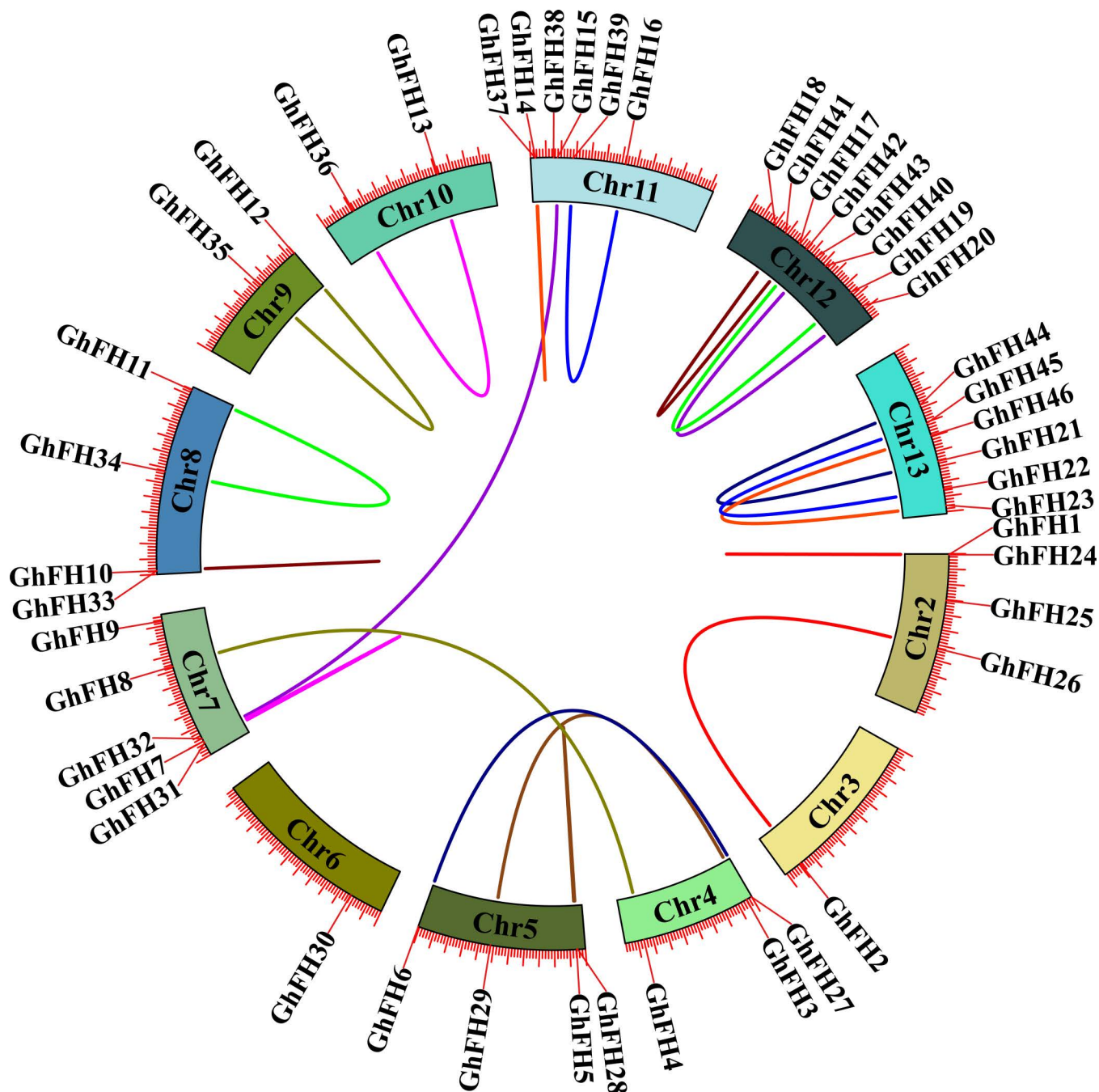


**Fig 9. Collinearity analysis of the *GhFH* gene.** Various colored rectangles represent chromosomes 2–13 within the *G. hirsutum* genome, while the colored lines linked between chromosomes represent segmental and tandem duplicated gene pairs.

https://doi.org/10.1371/journal.pone.0319176.g009

were designated as STRING proteins. All 46 GhFH proteins exhibited interactions with known *Arabidopsis* proteins. Notably, GhFH25, GhFH24, GhFH15, GhFH32, GhFH38, GhFH1, GhFH36, and GhFH13 were identified as homologous to AtFH1, interacting with FH5, FH2, FIM5, ARPC5A, PRF1, PRF2, PRF3, PRF4, and PRF5 proteins. Similarly, GhFH40, GhFH21, and GhFH44, homologous to AtFH11, displayed strong
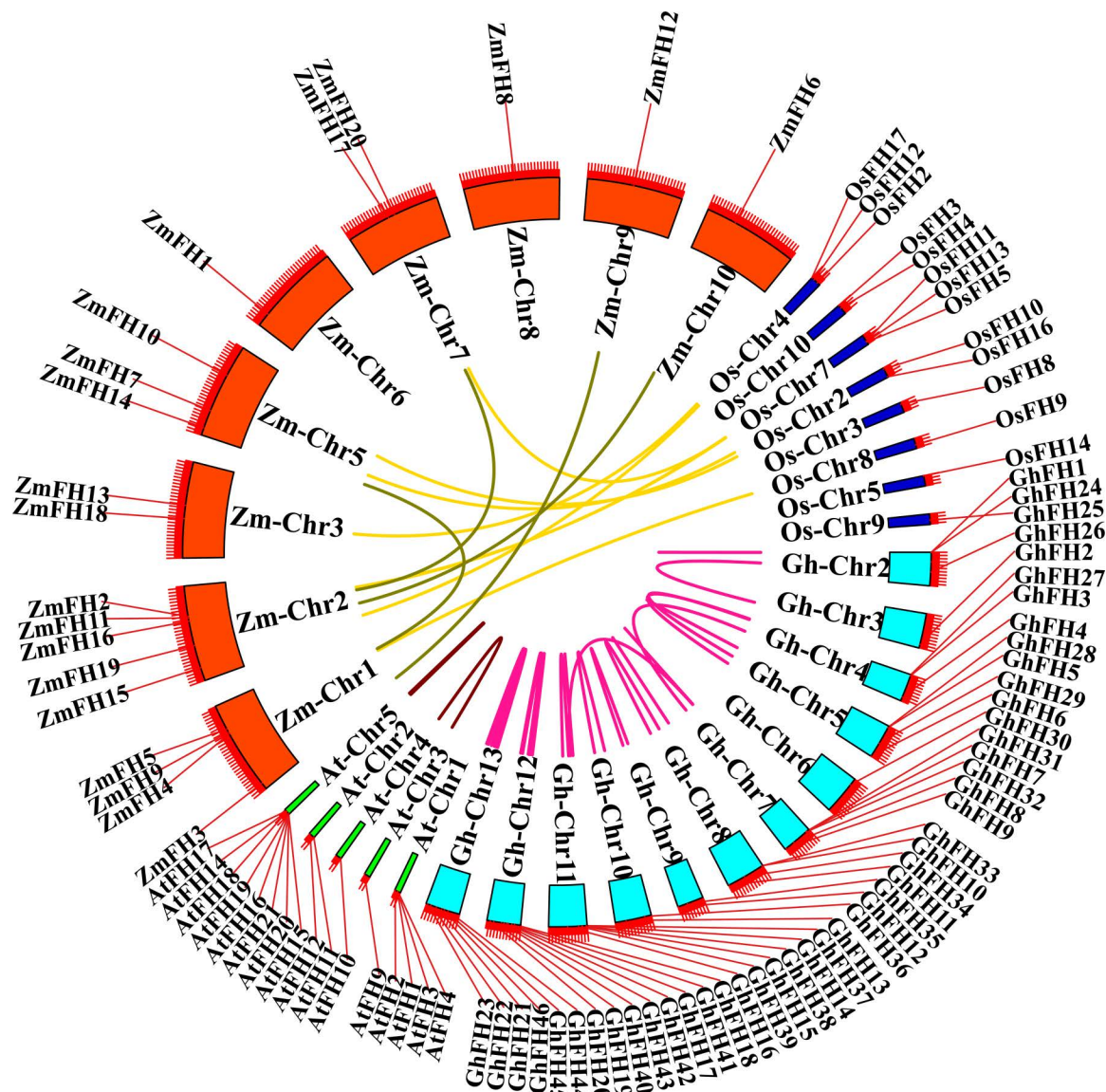
PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

**Fig 10. Synteny analysis of the FORMIN gene family among *G. hirsutum* and three other species, *A. thaliana, Z. mays, and* O.** *sativa.* The yellow-colored lines represent the 7 syntenic gene pairs between *O. sativa* and *Z. mays.*

interactions with FH12, ARPC5A, PRF1, PRF2, PRF3, PRF4, PRF5, and T14C9.40 proteins. GhFH30, GhFH23, and GhFH46, homologous to AtFH13, interacted with FH1, FH5, PRF1, PRF2, PRF3, PRF4, and PRF5 proteins ([Fig 12](#)). The pattern continued for other GhFH proteins, aligning with various AtFH proteins and forming distinct interaction groups. Additionally, GhFH43, GhFH14, GhFH31, GhFH37, GhFH20, GhFH18, GhFH7, GhFH9, GhFH8, GhFH35, GhFH4 and GhFH12 were homologous with AtFH20. GhFH29, GhFH42, GhFH33, GhFH19, GhFH22, GhFH3, GhFH10 and GhFH45 were homologous with AtFH5 and interact with FH20, FIM5, PRF1, PRF2, PRF3, PRF4, PRF5 proteins. Furthermore, GhFH41, GhFH16, GhFH39 and GhFH17 were homologous with AtFH6 and have a strong interaction between FH12, PRF1, PRF2, PRF3, PRF4, PRF5 proteins. GhFH6, GhFH26, GhFH2, GhFH27 were homologous with AtFH4.GhFH34,
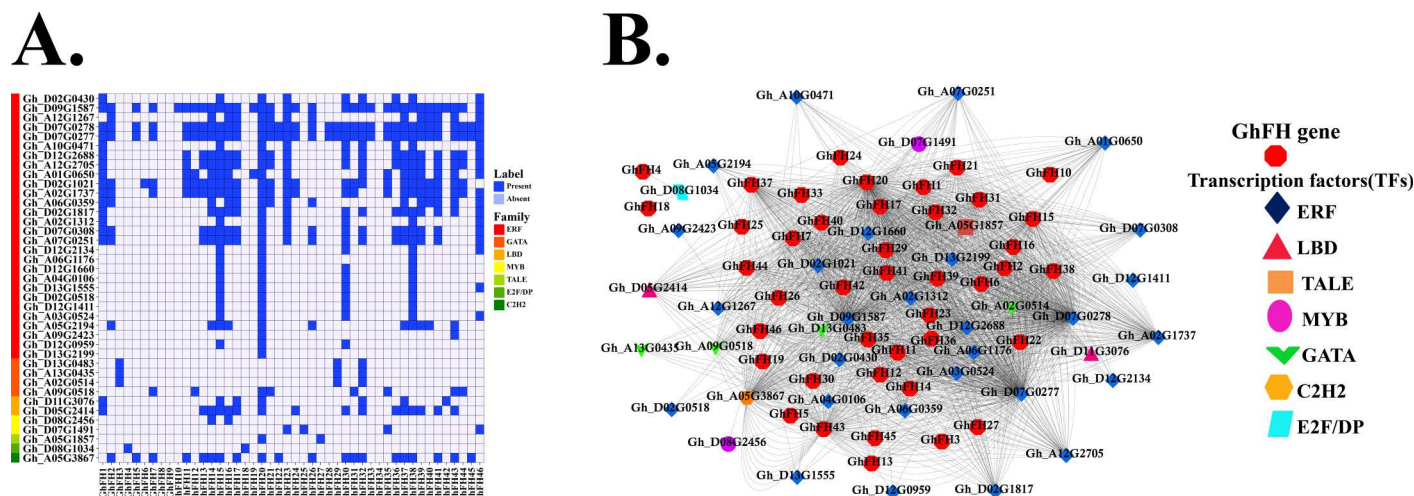
PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

# A.

# B.



**Fig 11. TFs analysis of *GhFH* genes.** A. Distribution of TFs of *GhFH* genes represented by a heatmap. Dark blue color denotes the present of TFs and light blue denotes their absence. B. Regulatory network among the TFs and the predicted *GhFH* genes. Nodes are colored based on *GhFH* genes and TFs. Genes are represented in red color, and the TFs are represented by different colors. Different node symbols are used for different TF families.

https://doi.org/10.1371/journal.pone.0319176.g011

GhFH11 were homologous with AtFH14 and GhFH5, GhFH28 were homologous with AtFH18. AtFH18 has a strong interaction with T6P5.20. The results of the PPI analysis were consistent with the phylogenetic relationships, as evidenced by AtFH proteins within the same phylogenetic group as GhFH proteins (e.g., GhFH6, GhFH26, GhFH2, GhFH27 being homologous with AtFH4 and situated in the same phylogenetic group A, interact with each other).

## 3.14. Identification of microRNAs (miRNAs) targeting *GhFH* genes

The role of miRNAs in gene regulation was investigated by identifying 45 putative ghr-miRNAs from 32 different families, targeting 33 *GhFH* genes. The regulatory mechanism of miRNAs in *GhFH* gene regulation was analyzed, with the findings illustrated through network diagrams ([Fig 13A](#) and [B](#) and [S12 Data](#)). Specific miRNA-family associations with *GhFH* genes were identified such as the ghr-miR390 family which targeted *GhFH10*, *GhFH16*, *GhFH29*, and *GhFH33*. Similarly, ghr-miR7484 targeted *GhFH36*, *GhFH24*, *GhFH13*, and *GhFH1*. The ghr-miR399 family influenced *GhFH16*, *GhFH38*, and *GhFH39*, while ghr-miR7495 targeted *GhFH7*, *GhFH35*, *GhFH31*, and *GhFH12*. Notably, the ghr-miR7502 family targeted five genes: *GhFH7*, *GhFH43*, *GhFH31*, *GhFH20*, and *GhFH19*. Other families such as ghr-miR7494, ghr-miR7504, and ghr-miR2950 showed similar patterns, targeting combinations of genes like *GhFH7*, *GhFH38*, *GhFH31*, *GhFH15* and *GhFH41*, *GhFH17*, *GhFH46*, *GhFH23*. Certain families targeted fewer genes for example, ghr-miR2949 and ghr-miR7492 each targeted *GhFH42* and *GhFH5*, respectively. Additionally, individual miRNAs like ghr-miR479, ghr-miR7491, and ghr-miR7508 play roles in regulating multiple genes, including GhFH43, GhFH34, GhFH3, GhFH11, and others. Unique miRNA-*GhFH* interactions were revealed, with miRNAs such as ghr-miR162a, ghr-miR7487, ghr-miR7488, ghr-miR7493, ghr-miR7505, ghr-miR7507, and ghr-miR7514 targeted genes such as GhFH26, GhFH11, GhFH46, GhFH31, and GhFH37. Notably, *GhFH7* was identified as the most targeted gene, with nine miRNAs from seven families regulating its expression. The ghr-miR7502 family was distinguished for targeting five genes: *GhFH7*, *GhFH43*, *GhFH31*, *GhFH20*, and *GhFH19*.
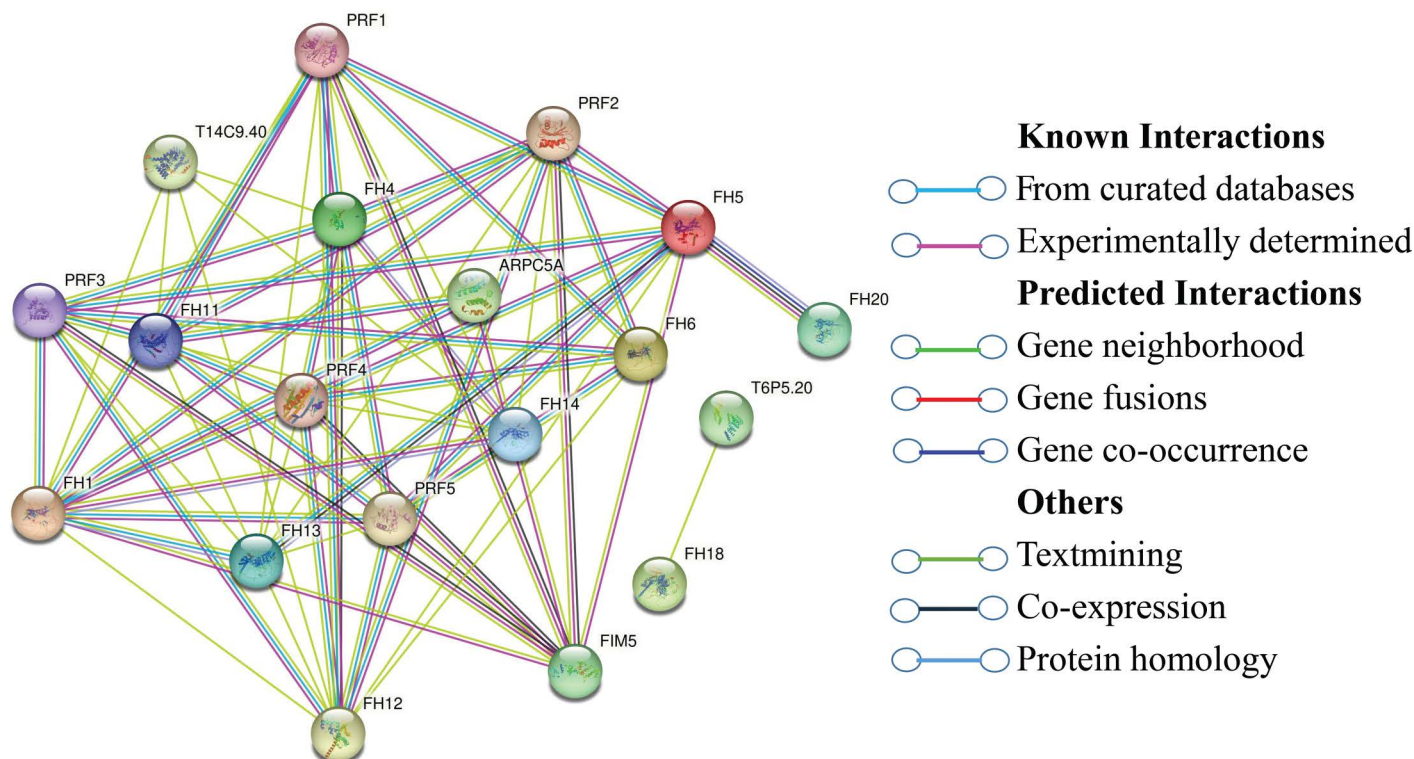
**Fig 12. Protein-protein interactions between GhFH proteins and *Arabidopsis* proteins.** The online STRING program was used to construct the network. Three-dimensional protein structures were shown at network nodes, and the colors of the lines indicate various data sources.

https://doi.org/10.1371/journal.pone.0319176.g012

### 3.15. Analysis of *GhFH* gene expression in various tissues

The expression of all identified *GhFH* genes was examined across various tissues to explore their potential functions in the growth and development of *G. hirsutum* cultivar Texas Marker-1 (TM-1) based on RNA-seq data. The results indicated significant variation in the expression of *GhFH* genes in specific tissues such as the root, stamen, stem, torus, leaf, petal, and pistil. All 46 *GhFH* genes were expressed in the pistil, with the highest expression observed in this tissue. Expression levels were also high in the root (43 genes, 93.48%), in the stamen (41 genes, 89.13%) torus, (40 genes, 86.96%), leaf (39 genes, 84.78%), and stem and petal (38 genes, 82.61%). The *GhFH20* and *GhFH34* genes were highly expressed across all selected tissues compared to other *GhFH* genes with particularly high expression levels, of *GhFH24* and *GhFH1* in the pistil (91.74 FPKM and 73.44 FPKM, respectively) ([Fig 14](#) and [S13 Data](#)).

### 3.16. Analysis of *GhFH* genes in response to different light conditions

The effect of light on the expression of the identified 46 *GhFH* genes was analyzed under three different light conditions: red, blue, and white. Among the 46 *GhFH* genes, 29 showed significantly higher expression under blue light compared to red and white light, while only 8 genes exhibited enhanced expression under white light ([Fig 15](#) and [S14 Data](#)). *GhFH20* had the highest expression, level under blue light, reaching 17.86 FPKM. *GhFH27* was not expressed under blue along with white lights, and *GhFH6* was not expressed under red and white lights. However, *GhFH1*, *GhFH10*, *GhFH20*, *GhFH24*, and *GhFH30* displayed a good level of expression under all three light conditions.
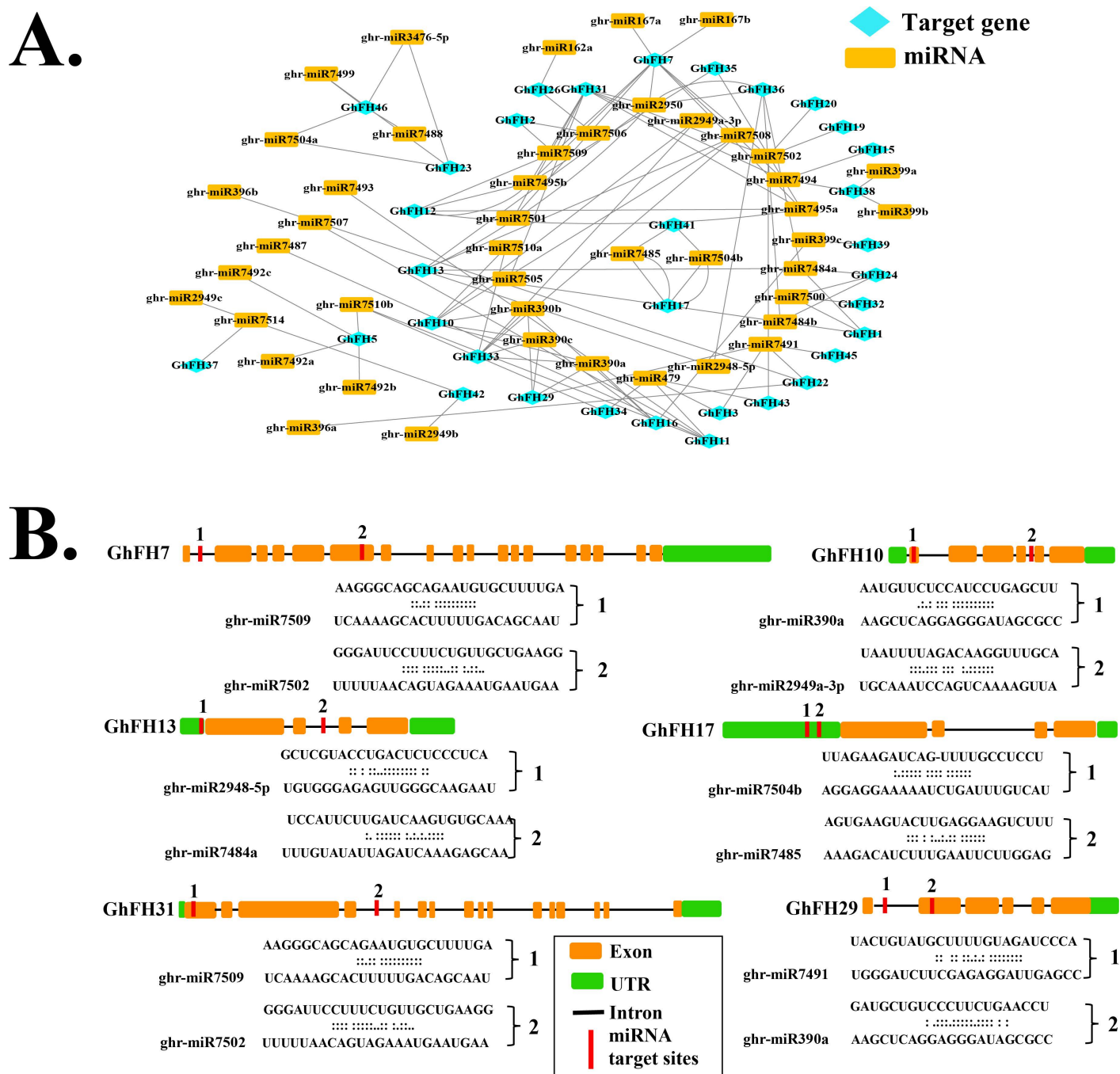
PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)



**Fig 13. Predicted miRNAs targeting *GhFH* genes.** A. The network diagram shows miRNAs predicted to target *GhFH* genes, with *GhFH* genes represented by rectangular shapes, and miRNAs by aqua ellipses. B. The schematic diagram displays *GhFH* genes targeted by miRNAs.

### 3.17. Analysis of *GhFH* genes in response to abiotic stress conditions

The response of identified *GhFH* genes in *G. hirsutum* to abiotic stresses (cold, hot, salt, and PEG) was analyzed using RNA-Seq data. Gene expression was studied in leaf tissue at different time points (1h, 3h, 6h, and 12h) following exposure to these stresses. The findings revealed
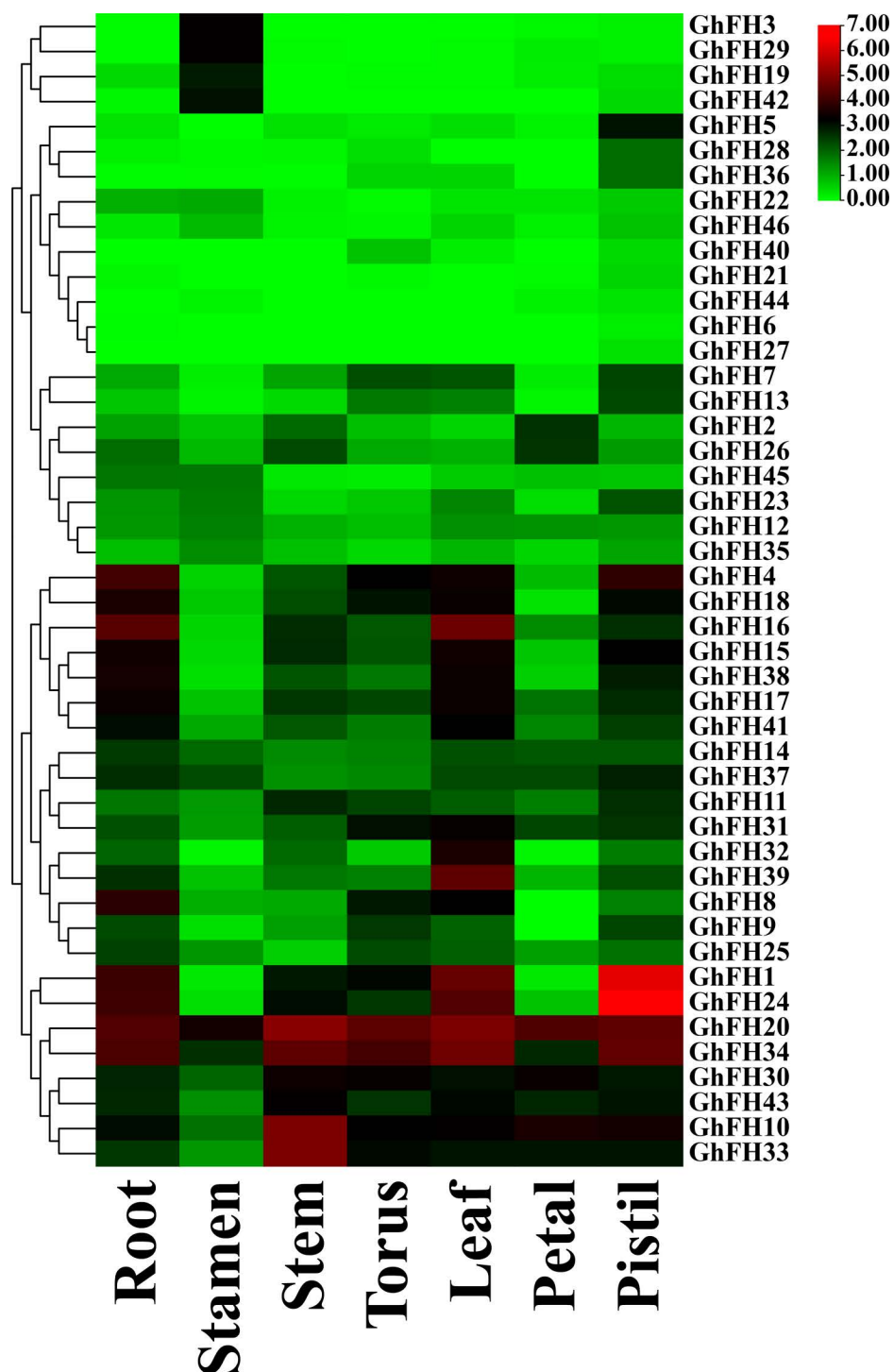
PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)



**Fig 14. Tissue-specific expression of *GhFH* genes.** A heatmap represents the expression profile of *GhFH* genes in root, stamen, stem, torus, leaf, petal, and pistil. *GhFH* gene names are listed on the right side of the heatmap, and tissue types are indicated at the bottom. Color intensity reflects the presence of protein signals corresponding to the genes.
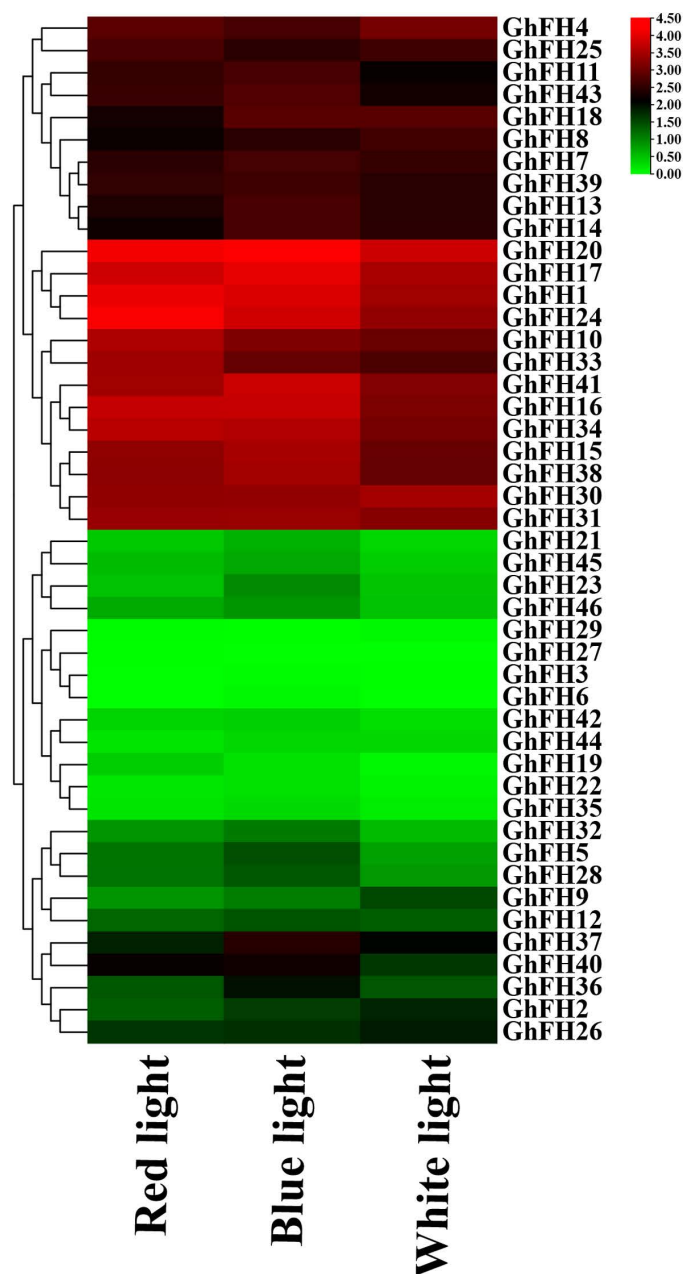
https://doi.org/10.1371/journal.pone.0319176.g014

**Fig 15. The expression pattern of *GhFH* under light conditions.** *GhFH* gene names are listed on the right side of the heatmap and light treatments (red, blue, and white) are represented at the bottom of the heatmap. The color gradient from green to red indicates the expression levels.

varied expression patterns, with some genes being up-regulated, and others down-regulated in response to stresses (Fig 16 and S15 Data). Based on the differential expression patterns, the *GhFH* genes were grouped into three categories: a) some *GhFHs* had very low expression levels, b) some had low to medium expression levels, and c) some had high expression levels throughout treatments. The results indicated that *GhFH9*, *GhFH20*, and *GhFH30* exhibited higher expression in hot environments compared to controls. *GhFH20* and *GhFH30* genes

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

also showed increased expression when exposed to salt. The expression of the *GhFH34* gene gradually increased under hot and salt treatment. *GhFH17*, *GhFH20*, and *GhFH33* were up-regulated during PEG treatment over treatment time. In contrast, all 46 identified *GhFH* genes showed decreased activity in leaf tissue under cold stress. *GhFH6* didn't express under any of the abiotic stresses(cold, hot, salt, and PEG).

## 4. Discussion

Cotton cultivation is recognized as the backbone of the economic prosperity in numerous nations. The textile industry primarily relies on cotton as its main source of natural fiber
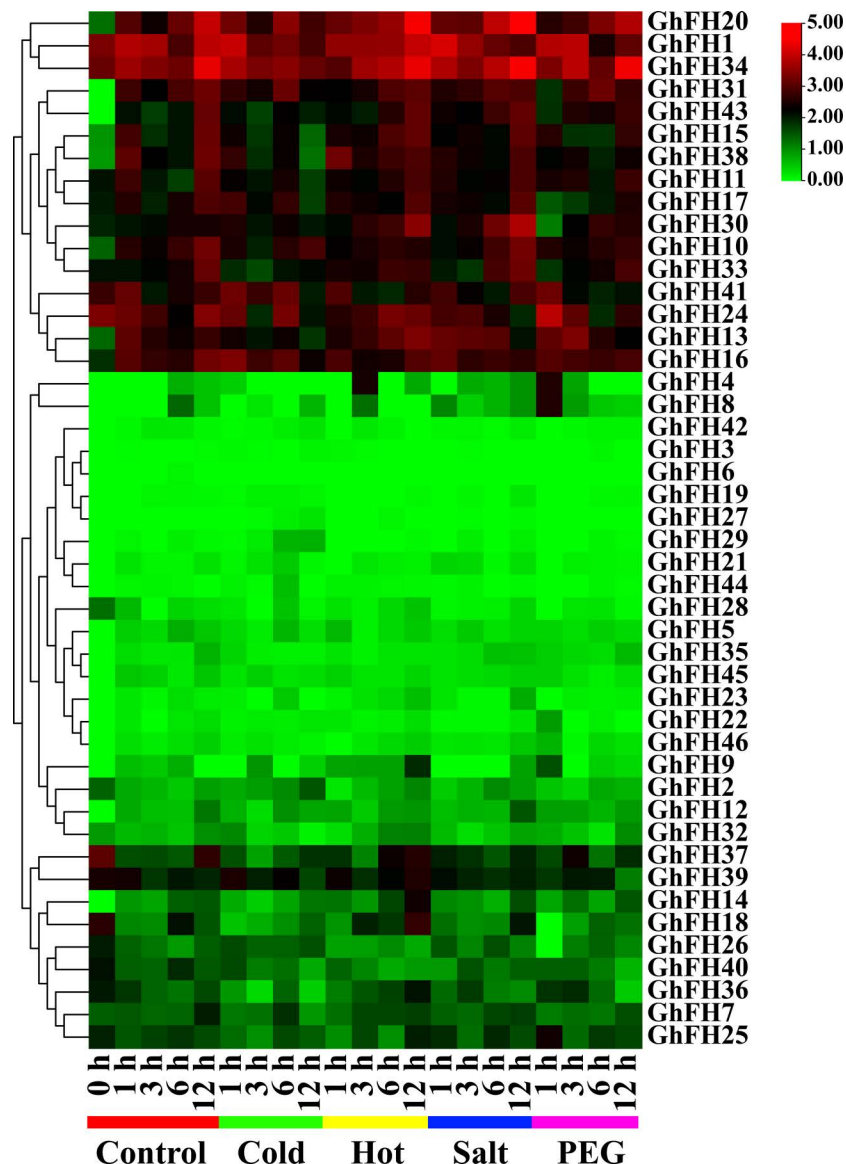


**Fig 16. Expression profiles of *GhFH* genes in leaf tissue under different abiotic stress conditions.** The FPKM values were converted to the Log2 format and compared to the control. Expression data were clustered and displayed using TBtools version-v1.116 with a color gradient indicating expression levels from low to high (ranging from green to red) depicted on the right side of the heatmap.

https://doi.org/10.1371/journal.pone.0319176.g016

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

source [72]. However, the production of this valuable crop is challenged by various abiotic stresses including salt, drought, cold, and heavy metal toxicity. These abiotic stress conditions significantly limit plant distribution, alter growth and development, and reduce crop productivity [73,74]. FORMIN proteins, which are crucial for cell growth and development, are ubiquitous in plants [17,75]. However, their involvement in mediating morphological changes in response to various environmental cues remains unclear [76]. FORMIN proteins have been successfully identified in *Arabidopsis* and rice, as well as in several angiosperms, including tobacco, sorghum, tomato, pea, wheat, and soybean [77].

In this study, *in silico* characterization of the identified FH genes in *G. hirsutum* was performed. A total of 46 FH genes were retrieved from *G. hirsutum*, distributed unevenly across the 12 chromosomes. It has been suggested that genes within the same family may be distributed on different chromosomes due to their involvement in various functions [78]. Proteins are classified as hydrophobic or hydrophilic based on their GRAVY value, with a positive value indicating hydrophobicity and a negative value implying hydrophilicity [79]. Based on the physio-chemical properties, all identified GhFH proteins except for GhFH37, with a value of 1.08, exhibited an average GRAVY value of less than 1, indicating that all proteins were hydrophobic (non-polar).

To further understand the evolutionary relationship among the 46 GhFH proteins a phylogenetic tree was constructed, including, 21 proteins from AtFH, 19 proteins from MtFH, 17 proteins from OsFH, and 20 proteins from ZmFH. These findings suggest that genes in *G. hirsutum*, *O. sativa*, and *Z. mays* remained highly conserved throughout evolution. The placement of exons and introns, essential for the evolution of gene families [80] was also analyzed. Genes carry the information required for reproduction and survival [81] and the analysis revealed a comparatively high level of structural variation among the *GhFH* genes. Genes clustered together within the phylogenetic tree shared significant similarities in their exon-intron structure [54]. The results of conserved domain and motif analysis showed that GhFH proteins within the same group, according to the phylogenetic tree, had similar motif distribution patterns and domain compositions. Besides the FH2 domain which was present in all GhFH proteins, certain individual domains were exclusively present in specific groups.

The subcellular locations of specific proteins play a crucial role in the biological processes and activities of plants [82,83]. FORMIN proteins have two key domains: the profilin-rich FH1 domain, which is essential for elongating actin filaments, and the FH2 domain, which is crucial for nucleating actin filaments [84]. Accordingly, our findings indicate that GhFH proteins are predominantly located in chloroplasts and the nucleus, suggesting a possible function as nucleation factors for actin filament nucleation. Proteins in chloroplasts are also involved in carbon fixation, amino acid biosynthesis, photosynthesis, and redox homeostasis. Chilling stress signals are also recognized by chloroplasts through membranes and photoreceptors, and their homeostasis is maintained and photosynthesis is promoted by regulating the state of lipid membranes [85]. The plant nucleus, which regulates gene expression is crucial for plants to adapt to abiotic stresses like drought, salinity, and extreme temperatures [74]. GhFH proteins present in the nucleus likely participate in the regulation of gene expression, signal transduction, sensing, or other essential nuclear processes crucial for cellular function. Additionally, 60.84% of GhFH proteins inhabit mitochondria suggesting a potential correlation with cellular respiration and energy metabolism.

CAREs are essential for coordinating cellular responses to environmental stimuli and developmental cues by regulating the TFs [86]. In this study, 62 types of CAREs (light-responsive, tissue-specific, stress-responsive, and phytohormone-responsive) were confirmed in the promoters of *GhFHs*. The most frequently observed light-responsive motifs in cotton *GhFH* genes were Box 4, GT1-motif, G-Box, GATA-motif, and TCT-motif. Photosynthesis,

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

a crucial physiological process related to light response, is typically observed in plant leaves. Early flowering, which can lead to high productivity, may be caused by a high photosynthesis rate [87]. Plant hormones or growth regulators (PGRs) play a vital role in plant seed germination, growth, development, and metabolic activities [88,89]. In this study, several important hormone-responsive CAREs were identified, including ABRE which is involved in abscisic acid responsiveness [90] and controls the expression of genes that respond to salt and dehydration in rice and *Arabidopsis* [91]. Other elements identified include the, GC motif and CGTCA-motif associated with anoxic-specific inducibility [92], AuxRR-core related to auxin responsiveness [88], TCA-element related to salicylic acid responsiveness [93] were identified. Additionally, LTR (involved in low-temperature response), TC-rich repeats (engaged in stress response and defense) and MBS (associated with drought inducibility) [91,94,95] were also found. These findings suggest that *GhFH* genes have a significant influence on responses to abiotic stress, phytohormone reactions, and defense-related signal transduction. The number of (*Ka*) and (*Ks*), along with their ratio, serves as a crucial tool for identifying proteins under selective pressure [96]. The calculated Ka/Ks ratio of all *GhFH* gene pairs was found to be less than 1 except for one exceptional pair (*GhFH12-GhFH35*) which had a ratio of 1.06. This analysis suggests that these genes may have undergone limited functional divergence and experienced strong purifying selection pressure during their evolutionary history. Gene duplication, including tandem and segmental duplication is regarded as a primary driving force in the evolution of genetic systems and genomes. It also enables organisms to adapt to their changing environments [97,98]. In this study, out of 46 *GhFH* genes identified in the *G. hirsutum* genome, 30 tandem duplications (65.28%) and 10 segmental duplications were observed. This detailed pattern highlights the role that tandem and segmental duplication events played in the development and expansion of the *GhFH* gene family.

To explore the functions of the identified *GhFH* genes, GO analysis was performed. A total of 59 unique GO IDs were identified, covering biological processes, molecular functions, and cellular components. Among these, biological processes exhibited the highest diversity with 48 essential terms, highlighting their crucial role in various biological functions. Distinct *p*-values were observed for each GO term, with GO:0044877 showing the highest number of *p*-value: 0.0076 and GO: 0030838 showing the lowest number of *p*-value: 0.00000000041, providing valuable insights into the significance of these terms in biological systems. Further investigation is needed to fully understand the functional significance of these findings.

Based on the results of collinearity and synteny analysis, it is predicted that the *GhFH* gene in *G. hirsutum* may have undergone duplication events during evolution, resulting in multiple copies of the gene in the genome. It was also demonstrated that the collinear gene pairs of *GhFH* genes have been maintained throughout cotton evolution, except *GhFH30*, located on chromosome 6, which did not exhibit such collinear gene pairing. Comparative synteny mapping revealed no syntenic pair of *GhFH* genes with other species *(Arabidopsis*, rice, and maize). Notably, the synteny analysis revealed the presence of 7 syntenic FH gene pairs between *Z. mays* and *O. sativa*. The presence of these syntenic gene pairs underscores the potential functional significance and evolutionary conservation of FH genes in plants, indicating possible functional similarities or shared evolutionary history among these FH genes.

TFs are bound to specific CARE regions in the promoters of target genes to regulate gene expression. Crucial regulators of numerous biological processes include ERF, C2H2, GATA, LBD, MYB, TALE, E2F/DP, and other plant TFs [99–101]. The Ethylene Response Factor (ERF) is essential for the response pathway in plants and for ethylene (ET) signaling. The ability of plants to endure challenging environments for extended periods is enhanced by the response of ERF to multiple plant hormones. PGRs such as abscisic acid (ABA) and ET can stimulate ABA-ET-dependent or independent stress-responsive (SR) genes through the action of certain AP2/ERF families [102].

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

The adaptability of tomatoes to salt and drought is improved by ERF, specifically Slerf5 (ERF5) [103]. Additionally, overexpression of Tsrf1, an ERF TF, has been shown to increase drought tolerance in rice [104]. MYB TFs are found in large quantities in plant systems; constituting approximately 9% of the entire TF family in *A. thaliana* [105]. MYB TFs influence numerous biological processes, including plant growth and development, cell shape and pattern creation, metabolism of physiological activities, and responses to biotic and abiotic stressors [106]. GATA TFs, a family of DNA-binding proteins found in many plant species, are connected to the regulation of transcription in plants that rely on light and nitrate [107]. The adaptability of GATA TFs is demonstrated by their interactions with biotic and abiotic stressors. Expression profiles show that GATA genes in rice, *Brassica juncea*, *Cucumis sativus*, and pepper respond to various abiotic stressors, including high temperatures, salinity, cold, and drought [108–111].

LBD TFs play pivotal roles in regulating the growth and development of various plant species. They are actively involved in secondary growth promotion, root, stem, leaf, and corolla growth, as well as the initiation and regulation of metabolic activities. Additionally, LBD genes contribute to the differentiation between terminal meristem primordia and lateral organ primordia. In higher plants, LBD genes have a significant influence on the development and maturation of both aerial and root. They are also essential for the metabolism of nitrogen and anthocyanins [112–114]. The C2H2 TFs family encodes proteins that are crucial for plant development, growth, and resistance to biological stress [115].

Within the TFs superfamily GARP (Golden2, ARR-B, and Psr1) domain, the G2-like proteins are unique members [116]. These G2-like TFs are critical for the development and maturation of chloroplasts [117–119] and have been associated with various defense mechanisms in various organisms, including response to biotic and abiotic stress [120–122]. Regulatory network analysis predicts a broad range of expression patterns for the *GhFH* genes and TFs in cotton. The results indicate that all genes interact with the ERF family except for two genes, providing strong evidence that these *GhFH* genes are associated with various plant hormones, which help plants survive in stressful environments for longer periods.

PPI network analysis reveals the activities of specific gene families associated with known proteins [123]. The study showed that 46 GhFH proteins share homology and establish strong interactions with 10 known *Arabidopsis* proteins, including AtFH1, AtFH11, AtFH13, AtFH14, AtFH18, AtFH20, AtFH4, AtFH5, and AtFH6. These FH proteins play crucial roles in *Arabidopsis*. For example, the AtFH5-GFP fusion protein, essential for cell division, accumulates in the cell plate. Additionally, AtFH6 regulates polarized growth by adjusting the assembly of actin cables [17]. Other reports indicate that AtFH8 influences root and root hair development by modifying the distribution of the actin cytoskeleton [18,124]. AtFH14 interacts with microtubules and microfilaments to regulate cell division [125]. The results suggest that GhFH family proteins may have similar functions. The study identified 45 putative ghr-miRNAs from 32 different families. Among these ghr-mir390a/b/c, ghr-miR2950, ghr-miR7491, ghr-miR7484a/b, and ghr-miR7502 were found to target most of the *GhFH* genes (Table 2).

These results suggested that the discovered gh-miRNAs may play important roles in overcoming various stresses by altering the transcriptional levels of *GhFH* genes in *G. hirsutum*, although further wet lab experiment is needed to confirm this theory.

Gene expression profiling provides important insights into determining gene functions [130]. In the current study, diverse expression levels of *GhFH* genes were observed among the selected tissues. All genes were expressed in the pistil, which is the female reproductive part of a flower, and responsible for receiving pollen and producing seeds [131]. This indicates that our identified genes play an important role in reproduction.

As the climate changes progress and arable land decreases, it is crucial to investigate how environmental stress affects the growth of important crops. Light serving as an essential

**Table 2. Information about abundant miRNA ID, functions, and their targeted *GhFH* genes.**

| miRNA ID | Functions | Targeted genes | References |
|---|---|---|---|
| ghr-mir390a/b/c | Responds to salt stress | GhFH10, GhFH16, GhFH29, GhFH33 | [126] |
| ghr-miR2950 | Role in the growth of cotton fiber | GhFH7, GhH13, GhFH31, GhFH36 | [127] |
| ghr-miR7491 | Involves in stress reactions, pest management, and fertility restoration | GhFH3, GhFH22, GhFH29, GhFH45 | [128] |
| ghr-miR7484a/b | Regulates cotton fiber development | GhFH1, GhFH13, GhFH24, GhFH36 | [127] |
| ghr-miR7502 | Role in response to salt stress | GhFH7, GhFH19, GhFH20, GhFH31, GhFH43 | [129] |

https://doi.org/10.1371/journal.pone.0319176.t002

energy source and a developmental cue for plants, can also induce stress and influence how plants respond to various stress factors [132]. Analysis of expression profiles from previous transcriptome data under red, blue, and white light conditions revealed that *GhFH* genes were highly expressed across all three conditions, strongly complementing the CARE analysis data. Specifically, *GhFH1*, *GhFH10*, *GhFH20*, *GhFH24*, and *GhFH30* exhibited the highest level of expression across the three light conditions. The tissue-specific expression also showed that 39 out of 46 *GhFH* genes had relatively high expression levels in the leaf, and almost all of those FORMIN genes demonstrated elevated expression levels under hot treatments. *GhFH9*, *GhFH20*, and *GhFH30* genes had greater expression profiles in hot conditions compared to controls. High expression of *GhFH17*, *GhFH20*, *GhFH30*, *GhFH33,* and *GhFH34* genes were high under salt and PEG treatments. Polyethylene glycol (PEG) an osmotic priming agent, helps to reduce the damage from abiotic stresses [133]. In conclusion, these findings indicate that these *GhFH* genes may play important roles in how plants respond to stress like heat, salt, and PEG and these patterns may be further explored in subsequent research.

## 5. Conclusions

In this study, the FORMIN gene family in *G. hirsutum* was systematically and scientifically identified and characterized. A total of 46 *GhFH*s were identified, distributed across 12 chromosomes. According to the phylogenetic tree, GhFHs were categorized into five groups and were found to be closely related to OsFHs and MtFHs. The gene structure analysis of the 46 *GhFH* genes revealed the structural diversity of genes in *G. hirsutum*. CAREs and GO analysis elucidated the functions of FORMINs in *G. hirsutum*, particularly in plant development and stress-related activities. *Ka/Ks* analysis revealed the evolutionary history of the *GhFH* genes (0.822- 52.50) MYA. Collinearity analysis showed that gene duplication events facilitated the expansion of the FORMIN family in *G. hirsutum*. The majority of *GhFH*s have ERF TFs that respond to various environmental stresses. Two genes, *GhFH20* and *GhFH30*, showed increased expression when exposed to heat and salt stresses. Additionally, the expression of GhFH34 gradually increased with hot and salt treatments. These genes were also highly expressed under light conditions. These findings provide a foundation for understanding the roles of *GhFH* genes and manipulating *GhFH* gene expression in *G. hirsutum* could potentially lead to the development of cotton varieties that are more resilient to environmental stresses.

## Supporting information

**S1 Data. Peptide sequences of *GhFH* gene family.**
(TXT)

**S2 Data. Peptide sequences of *AtFH*, *OsFH*, *MtFH*, *ZmFh,* and candidate *GhFH* gene families were used for the construction of a phylogenetic tree.**
(TXT)

**S3 Data. Genomic sequences of *GhFH* gene family.**
(TXT)

**S4 Data. CDS of *GhFH* gene family.**
(TXT)

**S5 Data. The promoter region of *GhFH* gene family.**
(TXT)

**S6 Data. Distribution of *GhFH* genes.**
(DOCX)

**S7 Data. Number of introns and exons in *GhFH* genes.**
(DOCX)

**S8 data. The predicted *cis*-acting regulatory elements of the upstream promoter region (2.0 kb genomic sequences) of *GhFH* gene family member.**
(XLSX)

**S9 Data. Time of gene duplication estimated for different paralogous pairs of *GhFH* genes based on Ka and Ks values.**
(XLSX)

**S10 Data. The details GO analysis of the predicted *GhFH* genes was performed using the Plant Transcription Factor Database (Plant TFDB,** http://planttfdb.cbi.pku.edu.cn/**).**
(XLSX)

**S11 Data. Identified the main 7 TF families associated with the regulation of identified *GhFH* genes.**
(XLSX)

**S12 Data. miRNA prediction of targeted GhFHs. The miRNA data was downloaded from psRNATarget Server18.**
(DOCX)

**S13 Data. Tissue-specific expression profiles of *GhFH* genes retrieved from NCBI (accession number SRA: PRJNA248163).**
(XLSX)

**S14 Data. Expression profiles of *GhFH* genes under light conditions (accession number SRA: PRJNA765172).**
(XLSX)

**S15 Data. Expression profiles of *GhFH* genes under abiotic stress like cold, heat, salt, and PEG (accession number SRA: PRJNA248163).**
(XLSX)

**S1 Fig. The sequence logos of the 20 motifs found in the GhFH proteins.**
(TIF)

## Acknowledgments

the reviewers and the members of the editorial panel for their valuable comments and critical suggestions for improving the quality of this manuscript.

## Author contributions

## References

1. Lee S-H, Tewari RK, Hahn E-J, Paek K-Y. Photon flux density and light quality induce changes in growth, stomatal development, photosynthesis and transpiration of Withania somnifera (L.) Dunal plantlets. Plant Cell Tissue Organ Cult. 2007;90:141–51.

2. Staneloni RJ, Rodriguez-Batiller MJ, Casal JJ Abscisic acid, high-light, and oxidative stress down-regulate a photosynthetic gene via a promoter motif not involved in phytochrome-mediated transcriptional regulation. Mol Plant. 2008;1(1):75–83.

3. Reis A, Kleinowski AM, Klein FRS, Telles RT, do Amarante L, Braga EJB, et al. Light quality on the in vitro growth and production of pigments in the genus Alternanthera. J Crop Sci Biotechnol. 2015;18:349–57.

4. Rao AQ, Ullah Khan MA, Shahid N, Din SU, Gul A, Muzaffar A, et al. An overview of phytochrome: an important light switch and photo-sensory antenna for regulation of vital functioning of plants. 2015;70(10):1273–83.

5. Bray EAJB, plants mbo. Responses to abiotic stresses; 2000. p. 1158–203.

6. Wang S, Kurepa J, Hashimoto T, Smalle JA. Salt stress–induced disassembly of Arabidopsis cortical microtubule arrays involves 26S proteasome–dependent degradation of SPIRAL1. Plant Cell. 2011;23(9):3412–27. https://doi.org/10.1105/tpc.111.089920 PMID: 21954463

7. Pruyne D, Evangelista M, Yang C, Bi E, Zigmond S, Bretscher A, et al. Role of formins in actin assembly: nucleation and barbed-end association. Science. 2002;297(5581):612–5. https://doi.org/10.1126/science.1072309 PMID: 12052901

8. Zweifel ME, Sherer LA, Mahanta B, Courtemanche N. Formin's nucleation activity influences actin filament length. bioRxiv 2021.06. 01.446650 [Preprint]. 2021.

9. Duan W-J, Liu Z-H, Bai J-F, Yuan S-H, Li Y-M, Lu F-K, et al. Comprehensive analysis of formin gene family highlights candidate genes related to pollen cytoskeleton and male fertility in wheat (Triticum aestivum L.). BMC Genomics. 2021;22(1):1–16.

10. Blanchoin L, Boujemaa-Paterski R, Henty JL, Khurana P, Staiger CJ. Actin dynamics in plant cells: a team effort from multiple proteins orchestrates this very fast-paced game. Curr Opin Plant Biol. 2010;13(6):714–23. Epub 2010 Oct 26. https://doi.org/10.1016/j.pbi.2010.09.013 PubMed PMID: 20970372

11. Cvrcková F, Novotný M, Pícková D, Zárský V. Formin homology 2 domains occur in multiple contexts in angiosperms. BMC Genomics. 2004;5(1):44. Epub 2004 Jul 17. https://doi.org/10.1186/1471-2164-5-44 PubMed PMID: 15256004; PubMed Central PMCID: PMCPMC509240

12. Vidali L, van Gisbergen PA, Guérin C, Franco P, Li M, Burkart GM, et al. Rapid formin-mediated actin-filament elongation is essential for polarized plant cell growth. Proc Natl Acad Sci U S A. 2009;106(32):13341–6. Epub 2009 Jul 28. https://doi.org/10.1073/pnas.0901170106 PubMed PMID: 19633191; PubMed Central PMCID: PMCPMC2726404

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

13. Qin L, Liu L, Tu J, Yang G, Wang S, Quilichini TD, et al. The ARP2/3 complex, acting cooperatively with Class I formins, modulates penetration resistance in Arabidopsis against powdery mildew invasion. Plant Cell. 2021;33(9):3151–75. Epub 2021 Jun 29. https://doi.org/10.1093/plcell/koab170 PubMed PMID: 34181022; PubMed Central PMCID: PMCPMC8462814

14. Higgs HN. Formin proteins: a domain-based approach. Trends Biochem Sci. 2005;30(6):342–53. Epub 2005 Jun 14. https://doi.org/10.1016/j.tibs.2005.04.014 PubMed PMID: 15950879

15. Li B, Du Z, Jiang N, He S, Shi Y, Xiao K, et al. Genome-wide identification and expression profiling of the FORMIN gene family implies their potential functions in abiotic stress tolerance in rice (Oryza sativa). Plant Mol Biol Rep. 2023;41(4):573–86. https://doi.org/10.1007/s11105-023-01387-5

16. Cheung AY, Wu HM. Overexpression of an Arabidopsis formin stimulates supernumerary actin cable formation from pollen tube cell membrane. Plant Cell. 2004;16(1):257–69. Epub 2003 Dec 11. https://doi.org/10.1105/tpc.016550 PubMed PMID: 14671023; PubMed Central PMCID: PMCPMC301409

17. Favery B, Chelysheva LA, Lebris M, Jammes F, Marmagne A, De Almeida-Engler J, et al. Arabidopsis formin AtFH6 is a plasma membrane-associated protein upregulated in giant cells induced by parasitic nematodes. Plant Cell. 2004;16(9):2529–40. Epub 2004 Aug 19. https://doi.org/10.1105/tpc.104.024372 PubMed PMID: 15319477; PubMed Central PMCID: PMCPMC520950

18. Yi K, Guo C, Chen D, Zhao B, Yang B, Ren H. Cloning and functional characterization of a formin-like protein (AtFH8) from Arabidopsis. Plant Physiol. 2005;138(2):1071–82. https://doi.org/10.1104/pp.104.055665 PMID: 15923338

19. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, et al. Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature. 2012;492(7429):423–7. Epub 2012 Dec 22. https://doi.org/10.1038/nature11798 PubMed PMID: 23257886

20. Sunilkumar G, Campbell LM, Puckhaber L, Stipanovic RD, Rathore KS. Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. Proc Natl Acad Sci USA. 2006;103(48):18054–9. https://doi.org/10.1073/pnas.0605389103 PMID: 17110445

21. Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM. Evolution and natural history of the cotton genus. In: Genetics and genomics of cotton; 2009. p. 3–22.

22. Ali I, Teng Z, Bai Y, Yang Q, Hao Y, Hou J, et al. A high density SLAF-SNP genetic map and QTL detection for fibre quality traits in Gossypium hirsutum. BMC Genomics. 2018;19(1):879. Epub 2018 Dec 14. https://doi.org/10.1186/s12864-018-5294-5 PubMed PMID: 30522437; PubMed Central PMCID: PMCPMC6282304

23. Grover CE, Gallagher JP, Jareczek JJ, Page JT, Udall JA, Gore MA, et al. Re-evaluating the phylogeny of allopolyploid Gossypium L. Mol Phylogenet Evol. 2015;92:45–52. https://doi.org/10.1016/j.ympev.2015.05.023 PMID: 26049043

24. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012;40(Database issue):D1178–86. Epub 2011 Nov 24. https://doi.org/10.1093/nar/gkr944 PubMed PMID: 22110026; PubMed Central PMCID: PMCPMC3245001

25. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 2020;48(D1):D265–8. Epub 2019 Nov 30. https://doi.org/10.1093/nar/gkz991 PubMed PMID: 31777944; PubMed Central PMCID: PMCPMC6943070

26. Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. 2021;49(D1):D458–60. Epub 2020 Oct 27. https://doi.org/10.1093/nar/gkaa937 PubMed PMID: 33104802; PubMed Central PMCID: PMCPMC7778883

27. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42(Database issue):D222–30. Epub 2013 Nov 30. https://doi.org/10.1093/nar/gkt1223 PubMed PMID: 24288371; PubMed Central PMCID: PMCPMC3965110

28. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999;112:531–52. Epub 1999 Feb 23. https://doi.org/10.1385/1-59259-584-7:531 PubMed PMID: 10027275

29. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38(7):3022–7. Epub 2021 Apr 24. https://doi.org/10.1093/molbev/msab120 PubMed PMID: 33892491; PubMed Central PMCID: PMCPMC8233496

30. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22(22):4673–80. Epub 1994 Nov 11. https://doi.org/10.1093/nar/22.22.4673 PubMed PMID: 7984417; PubMed Central PMCID: PMCPMC308517

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

31. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinform. 2002;Chapter 2:Unit 2.3. Epub 2008 Sep 17. https://doi.org/10.1002/0471250953.bi0203s00 PubMed PMID: 18792934.

32. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293–6. Epub 2021 Apr 23. https://doi.org/10.1093/nar/gkab301 PubMed PMID: 33885785; PubMed Central PMCID: PMCPMC8265157

33. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31(8):1296–7. Epub 2014 Dec 17. https://doi.org/10.1093/bioinformatics/btu817 PubMed PMID: 25504850; PubMed Central PMCID: PMCPMC4393523

34. Ren J, Wen L, Gao X, Jin C, Xue Y, Yao XJC. DOG 1.0: illustrator of protein domain structures. Cell Res. 2009;19(2):271–3. https://doi.org/10.1038/cr.2009.6 PMID: 19153597

35. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res. 2015;43(W1):W39–49. Epub 2015 May 09. https://doi.org/10.1093/nar/gkv416 PubMed PMID: 25953851; PubMed Central PMCID: PMCPMC4489269

36. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13(8):1194–202. Epub 2020 Jun 23. https://doi.org/10.1016/j.molp.2020.06.009 PubMed PMID: 32585190

37. Horton P, Park K-J, Obayashi T, Nakai K, editors. Protein subcellular localization prediction with WoLF PSORT. Proceedings of the 4th Asia-Pacific bioinformatics conference. World Scientific; 2006.

38. Giorgi FM, Ceraolo C, Mercatelli DJL. The R language: an engine for bioinformatics and data science. Life (Basel, Switzerland). 2022;12(5):648. https://doi.org/10.3390/life12050648 PMID: 35629316

39. Rombauts S, Déhais P, Van Montagu M, Rouzé P. PlantCARE, a plant cis-acting regulatory element database. Nucleic Acids Res. 1999;27(1):295–6. Epub 1998 Dec 10. https://doi.org/10.1093/nar/27.1.295 PubMed PMID: 9847207; PubMed Central PMCID: PMCPMC148162

40. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290(5494):1151–5. Epub 2000 Nov 10. https://doi.org/10.1126/science.290.5494.1151 PubMed PMID: 11073452

41. Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. 2017;45(D1):D1040–5. Epub 2016 Dec 08. https://doi.org/10.1093/nar/gkw982 PubMed PMID: 27924042; PubMed Central PMCID: PMCPMC5210657

42. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504. Epub 2003 Nov 05. https://doi.org/10.1101/gr.1239303 PubMed PMID: 14597658; PubMed Central PMCID: PMCPMC403769

43. Dai X, Zhuang Z, Zhao PX. psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Res. 2018;46(W1):W49–54.

44. Sayers EW, O'Sullivan C, Karsch-Mizrachi I. Using GenBank and SRA. In: Plant bioinformatics: methods and protocols. Springer; 2022. p. 1–25.

45. Shao D, Zhu Q-h, Liang Q, Wang X, Li Y, Sun Y, et al. Transcriptome analysis reveals differences in anthocyanin accumulation in cotton (Gossypium hirsutum L.) induced by red and blue light. 2022;13:788828.

46. Bolger AM, Lohse M, Usadel BJB. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England). 2014;30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

47. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England). 2013;29(1):15–21. Epub 2012 Oct 30. https://doi.org/10.1093/bioinformatics/bts635 PubMed PMID: 23104886; PubMed Central PMCID: PMCPMC3530905

48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics (Oxford, England). 2009;25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

49. Li B. Dewey CNJBb. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. 2011;12:1–16.

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

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

51. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

52. Evangelista M, Zigmond S, Boone C. Formins: signaling effectors for assembly and polarization of actin filaments. J Cell Sci. 2003;116(Pt 13):2603–11. Epub 2003 May 31. https://doi.org/10.1242/jcs.00611 PubMed PMID: 12775772

53. Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng. 1990;4(2):155–61. Epub 1990 Dec 01. https://doi.org/10.1093/protein/4.2.155 PubMed PMID: 2075190

54. Shen XX, Salichos L, Rokas A. A Genome-scale investigation of how sequence, function, and tree-based gene properties influence phylogenetic inference. Genome Biol Evol. 2016;8(8):2565–80. Epub 2016 Aug 06. https://doi.org/10.1093/gbe/evw179 PubMed PMID: 27492233; PubMed Central PMCID: PMCPMC5010910

55. Shaul O. How introns enhance gene expression. Int J Biochem Cell Biol. 2017;91(Pt B):145–55. Epub 2017 Jul 05. https://doi.org/10.1016/j.biocel.2017.06.016 PubMed PMID: 28673892

56. Aziz MF, Caetano-Anollés G. Evolution of networks of protein domain organization. Sci Rep. 2021;11(1):12075. Epub 2021 Jun 10. https://doi.org/10.1038/s41598-021-90498-8 PubMed PMID: 34103558; PubMed Central PMCID: PMCPMC8187734

57. Boeva V. Analysis of genomic sequence motifs for deciphering transcription factor binding and transcriptional regulation in eukaryotic cells. Front Genet. 2016;7:24. Epub 2016 Mar 05. https://doi.org/10.3389/fgene.2016.00024 PubMed PMID: 26941778; PubMed Central PMCID: PMCPMC4763482

58. Barberis E, Marengo E, Manfredi M. Protein subcellular localization prediction. Methods Mol Biol. 2021;2361:197–212. Epub 2021 Jul 09. https://doi.org/10.1007/978-1-0716-1641-3_12 PubMed PMID: 34236663

59. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18(9):486. Epub 2002 Aug 15. https://doi.org/10.1016/s0168-9525(02)02722-1 PubMed PMID: 12175810

60. Vidhyasekaran P, Vidhyasekaran P. Molecular manipulation of transcription factors, the master regulators of PAMP-triggered signaling systems. In: Switching on plant innate immunity signaling systems: bioengineering and molecular manipulation of PAMP-PIMP-PRR signaling complex; 2016. p. 255–358.

61. Khan SA, Li MZ, Wang SM, Yin HJ. Revisiting the role of plant transcription factors in the battle against abiotic stress. Int J Mol Sci. 2018;19(6). Epub 2018 Jun 03. https://doi.org/10.3390/ijms19061634 PubMed PMID: 29857524; PubMed Central PMCID: PMCPMC6032162

62. Latchman DS. Transcription factors: an overview. Int J Biochem Cell Biol. 1997;29(12):1305–12. Epub 1998 May 07. https://doi.org/10.1016/s1357-2725(97)00085-x PubMed PMID: 9570129

63. Lutova LA, Dodueva IE, Lebedeva MA, Tvorogova VE. [Transcription factors in developmental genetics and the evolution of higher plants]. Genetika. 2015;51(5):539–57. Epub 2015 Jul 04. PubMed PMID: 26137635

64. Sasaki K. Utilization of transcription factors for controlling floral morphogenesis in horticultural plants. Breed Sci. 2018;68(1):88–98. Epub 2018 Apr 24. https://doi.org/10.1270/jsbbs.17114 PubMed PMID: 29681751; PubMed Central PMCID: PMCPMC5903982

65. Shu Y, Liu Y, Zhang J, Song L, Guo C. Genome-Wide Analysis of the AP2/ERF Superfamily Genes and their Responses to Abiotic Stress in Medicago truncatula. Front Plant Sci. 2015;6:1247. Epub 2016 Feb 03. https://doi.org/10.3389/fpls.2015.01247 PubMed PMID: 26834762; PubMed Central PMCID: PMCPMC4717309

66. Mengiste T, Chen X, Salmeron J, Dietrich R. The BOTRYTIS SUSCEPTIBLE1 gene encodes an R2R3MYB transcription factor protein that is required for biotic and abiotic stress responses in Arabidopsis. Plant Cell. 2003;15(11):2551–65. Epub 2003 Oct 14. https://doi.org/10.1105/tpc.014167 PubMed PMID: 14555693; PubMed Central PMCID: PMCPMC280560

67. Meshi T, Iwabuchi M. Plant transcription factors. Plant Cell Physiol. 1995;36(8):1405–20. Epub 1995 Dec 01. PubMed PMID: 8589926

68. Cánovas FM, Dumas-Gaudot E, Recorbet G, Jorrin J, Mock HP, Rossignol M. Plant proteome analysis. Proteomics. 2004;4(2):285–98. Epub 2004 Feb 05. https://doi.org/10.1002/pmic.200300602 PubMed PMID: 14760698

69. Bracha-Drori K, Shichrur K, Katz A, Oliva M, Angelovici R, Yalovsky S, et al. Detection of protein-protein interactions in plants using bimolecular fluorescence complementation. Plant J. 2004;40(3):419–27. Epub 2004 Oct 08. https://doi.org/10.1111/j.1365-313X.2004.02206.x PubMed PMID: 15469499

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

70. Zhang Y, Gao P, Yuan JS. Plant protein-protein interaction network and interactome. Curr Genomics. 2010;11(1):40–6. Epub 2010 Sep 03. https://doi.org/10.2174/138920210790218016 PubMed PMID: 20808522; PubMed Central PMCID: PMCPMC2851115

71. Khan IK, Kihara D. Genome-scale prediction of moonlighting proteins using diverse protein association information. Bioinformatics. 2016;32(15):2281–8. Epub 2016 May 07. https://doi.org/10.1093/bioinformatics/btw166 PubMed PMID: 27153604; PubMed Central PMCID: PMCPMC4965633

72. Thyavihalli Girijappa YG, Mavinkere Rangappa S, Parameswaranpillai J, Siengchin S. Natural fibers as sustainable and renewable resource for development of eco-friendly composites: a comprehensive review. Front Mater. 2019;6:226.

73. Zhu J-K. Abiotic stress signaling and responses in plants. Cell. 2016;167(2):313–24. https://doi.org/10.1016/j.cell.2016.08.029 PMID: 27716505

74. Zhang H, Zhu J, Gong Z, Zhu J-K. Abiotic stress responses in plants. Nat Rev Genet. 2022;23(2):104–19. https://doi.org/10.1038/s41576-021-00413-0 PMID: 34561623

75. Kitayama C, Uyeda TQ. ForC, a novel type of formin family protein lacking an FH1 domain, is involved in multicellular development in Dictyostelium discoideum. J Cell Sci. 2003;116(Pt 4):711–23. https://doi.org/10.1242/jcs.00265 PubMed PMID: 12538772

76. Cvrčková F, Novotný M, Píčková D, Žárský V. Formin homology 2 domains occur in multiple contexts in angiosperms. BMC Genomics. 2004;5(1):1–18.

77. Zhang Z, Zhang Z, Shan M, Amjad Z, Xue J, Zhang Z, et al. Genome-wide studies of FH family members in soybean (Glycine max) and their responses under abiotic stresses. Plants (Basel, Switzerland). 2024;13(2):276. https://doi.org/10.3390/plants13020276 PMID: 38256829

78. Muhammad Ahmad H, Wang X, Fiaz S, Azhar Nadeem M, Aslam Khan S, Ahmar S, et al. Comprehensive genomics and expression analysis of eceriferum (CER) genes in sunflower (Helianthus annuus). 2021;28(12):6884–96.

79. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105–32.

80. Xu G, Guo C, Shan H, Kong H. Divergence of duplicate genes in exon-intron structure. Proc Natl Acad Sci U S A. 2012;109(4):1187–92. Epub 2012 Jan 11. https://doi.org/10.1073/pnas.1109047109 PubMed PMID: 22232673; PubMed Central PMCID: PMCPMC3268293

81. Polyak K, Meyerson M. Overview: gene structure. 6th ed. Holland-Frei Cancer Medicine; 2003.

82. Ehrlich JS, Hansen MD, Nelson WJ. Spatio-temporal regulation of Rac1 localization and lamellipodia dynamics during epithelial cell-cell adhesion. Dev Cell. 2002;3(2):259–70. Epub 2002 Aug 27. https://doi.org/10.1016/s1534-5807(02)00216-2 PubMed PMID: 12194856; PubMed Central PMCID: PMCPMC3369831

83. Glory E, Murphy RF. Automated subcellular location determination and high-throughput microscopy. Dev Cell. 2007;12(1):7–16. Epub 2007 Jan 03. https://doi.org/10.1016/j.devcel.2006.12.007 PubMed PMID: 17199037

84. Cheung AY, Niroomand S, Zou Y, Wu H-M. A transmembrane formin nucleates subapical actin assembly and controls tip-focused growth in pollen tubes. Proc Natl Acad Sci. 2010;107(37):16390–5.

85. Gan P, Liu F, Li R, Wang S, Luo J. Chloroplasts- beyond energy capture and carbon fixation: tuning of photosynthesis in response to chilling stress. Int J Mol Sci. 2019;20(20):5046. Epub 2019 Oct 17. https://doi.org/10.3390/ijms20205046 PubMed PMID: 31614592; PubMed Central PMCID: PMCPMC6834309

86. Schmitz RJ, Grotewold E, Stam M. Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. Plant Cell. 2022;34(2):718–41. Epub 2021 Dec 18. https://doi.org/10.1093/plcell/koab281 PubMed PMID: 34918159; PubMed Central PMCID: PMCPMC8824567

87. Lee HW, Cho C, Kim J. Lateral organ boundaries domain16 and 18 Act downstream of the AUXIN1 and LIKE-AUXIN3 auxin influx carriers to control lateral root development in Arabidopsis. Plant Physiol. 2015;168(4):1792–806. Epub 2015 Jun 11. https://doi.org/10.1104/pp.15.00578 PubMed PMID: 26059335; PubMed Central PMCID: PMCPMC4528759

88. Kaur A, Pati PK, Pati AM, Nagpal AK. In-silico analysis of cis-acting regulatory elements of pathogenesis-related proteins of Arabidopsis thaliana and Oryza sativa. PLoS One. 2017;12(9):e0184523. Epub 2017 Sep 14. https://doi.org/10.1371/journal.pone.0184523 PubMed PMID: 28910327; PubMed Central PMCID: PMCPMC5598985

89. Shariatipour N, Heidari B. Investigation of drought and salinity tolerance related genes and their regulatory mechanisms in Arabidopsis (). Open Bioinform J. 2018;11(1):12–28. https://doi.org/10.2174/1875036201811010012

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

90. Ezcurra I, Wycliffe P, Nehlin L, Ellerström M, Rask L. Transactivation of the Brassica napus napin promoter by ABI3 requires interaction of the conserved B2 and B3 domains of ABI3 with different cis-elements: B2 mediates activation through an ABRE, whereas B3 interacts with an RY/G-box. Plant J. 2000;24(1):57–66. Epub 2000 Oct 13. https://doi.org/10.1046/j.1365-313x.2000.00857.x PubMed PMID: 11029704

91. Maruyama K, Todaka D, Mizoi J, Yoshida T, Kidokoro S, Matsukura S, et al. Identification of cis-acting promoter elements in cold-and dehydration-induced transcriptional pathways in Arabidopsis, rice, and soybean. 2012;19(1):37–49.

92. Martin-Malpartida P, Batet M, Kaczmarska Z, Freier R, Gomes T, Aragón E, et al. Structural basis for genome wide recognition of 5-bp GC motifs by SMAD transcription factors. Nat Commun. 2017;8(1):2070. Epub 2017 Dec 14. https://doi.org/10.1038/s41467-017-02054-6 PubMed PMID: 29234012; PubMed Central PMCID: PMCPMC5727232

93. Kim SR, Kim Y, An G. Identification of methyl jasmonate and salicylic acid response elements from the nopaline synthase (nos) promoter. Plant Physiol. 1993;103(1):97–103. Epub 1993 Sep 01. https://doi.org/10.1104/pp.103.1.97 PubMed PMID: 8208860; PubMed Central PMCID: PMCPMC158951

94. Arias JA, Dixon RA, Lamb CJ. Dissection of the functional architecture of a plant defense gene pro-moter using a homologous in vitro transcription initiation system. Plant Cell. 1993;5(4):485–96. Epub 1993 Apr 01. https://doi.org/10.1105/tpc.5.4.485 PubMed PMID: 8485404; PubMed Central PMCID: PMCPMC160287

95. Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, et al. Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell. 2002;14(3):559–74. Epub 2002 Mar 23. https://doi.org/10.1105/tpc.010410 PubMed PMID: 11910004; PubMed Central PMCID: PMCPMC150579

96. He Q, Cai H, Bai M, Zhang M, Chen F, Huang Y, et al. A soybean bZIP transcription factor Gmb-ZIP19 confers multiple biotic and abiotic stress responses in plant. Int J Mol Sci. 2020;21(13):4701. Epub 2020 Jul 08. https://doi.org/10.3390/ijms21134701 PubMed PMID: 32630201; PubMed Central PMCID: PMCPMC7369738

97. Moore RC, Purugganan MD. The early stages of duplicate gene evolution. Proc Natl Acad Sci U S A. 2003;100(26):15682–7. Epub 2003 Dec 13. https://doi.org/10.1073/pnas.2535513100 PubMed PMID: 14671323; PubMed Central PMCID: PMCPMC307628

98. López-Maury L, Marguerat S, Bähler J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat Rev Genet. 2008;9(8):583–93. Epub 2008 Jul 02. https://doi.org/10.1038/nrg2398 PubMed PMID: 18591982

99. Erpen L, Devi HS, Grosser JW, Dutt M. Potential use of the DREB/ERF, MYB, NAC and WRKY transcription factors to improve abiotic and biotic stress in transgenic plants. Plant Cell Tissue Organ Cult. 2018;132(1):1–25. https://doi.org/10.1007/s11240-017-1320-6

100. Ohta M, Sato A, Renhu N, Yamamoto T, Oka N, Zhu JK, et al. MYC-type transcription factors, MYC67 and MYC70, interact with ICE1 and negatively regulate cold tolerance in Arabidopsis. Sci Rep. 2018;8(1):11622. Epub 2018 Aug 04. https://doi.org/10.1038/s41598-018-29722-x PubMed PMID: 30072714; PubMed Central PMCID: PMCPMC6072781

101. Luo P, Li Z, Chen W, Xing W, Yang J, Cui Y. Overexpression of RmICE1, a bHLH transcription factor from Rosa multiflora, enhances cold tolerance via modulating ROS levels and activating the expression of stress-responsive genes. Environ Exp Bot. 2020;178:104160. https://doi.org/10.1016/j.envexpbot.2020.104160

102. Xie Z, Nolan TM, Jiang H, Yin Y. AP2/ERF transcription factor regulatory networks in hormone and abiotic stress responses in Arabidopsis. Front Plant Sci. 2019;10:228. Epub 2019 Feb 28. https://doi.org/10.3389/fpls.2019.00228 PubMed PMID: 30873200; PubMed Central PMCID: PMCPMC6403161

103. Pan Y, Seymour GB, Lu C, Hu Z, Chen X, Chen G. An ethylene response factor (ERF5) promoting adaptation to drought and salt tolerance in tomato. Plant Cell Rep. 2012;31(2):349–60. Epub 2011 Nov 01. https://doi.org/10.1007/s00299-011-1170-3 PubMed PMID: 22038370

104. Quan R, Hu S, Zhang Z, Zhang H, Zhang Z, Huang R. Overexpression of an ERF transcrip-tion factor TSRF1 improves rice drought tolerance. Plant Biotechnol J. 2010;8(4):476–88. Epub 2010 Mar 18. https://doi.org/10.1111/j.1467-7652.2009.00492.x PubMed PMID: 20233336

105. Cao Y, Li K, Li Y, Zhao X, Wang L. MYB transcription factors as regulators of secondary metabolism in plants. Biology (Basel). 2020;9(3):61. Epub 2020 Mar 28. https://doi.org/10.3390/biology9030061 PubMed PMID: 32213912; PubMed Central PMCID: PMCPMC7150910

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

**106.** Ramya M, Kwon OK, An HR, Park PM, Baek YS, Park PH. Floral scent: regulation and role of MYB transcription factors. Phytochem Lett. 2017;19:114–20. https://doi.org/10.1016/j.phytol.2016.12.015

**107.** Reyes JC, Muro-Pastor MI, Florencio FJ. The GATA family of transcription factors in Arabidopsis and rice. Plant Physiol. 2004;134(4):1718–32. Epub 2004 Apr 16. https://doi.org/10.1104/pp.103.037788 PubMed PMID: 15084732; PubMed Central PMCID: PMCPMC419845

**108.** Gupta P, Nutan KK, Singla-Pareek SL, Pareek A. Abiotic stresses cause differential regulation of alternative splice forms of GATA transcription factor in rice. Front Plant Sci. 2017;8:1944. Epub 2017 Nov 29. https://doi.org/10.3389/fpls.2017.01944 PubMed PMID: 29181013; PubMed Central PMCID: PMCPMC5693882

**109.** Bhardwaj AR, Joshi G, Kukreja B, Malik V, Arora P, Pandey R, et al. Global insights into high temperature and drought stress regulated genes by RNA-Seq in economically important oilseed crop Brassica juncea. BMC Plant Biol. 2015;15:9. Epub 2015 Jan 22. https://doi.org/10.1186/s12870-014-0405-1 PubMed PMID: 25604693; PubMed Central PMCID: PMCPMC4310166

**110.** Zhang Z, Zou X, Huang Z, Fan S, Qun G, Liu A, et al. Genome-wide identification and analysis of the evolution and expression patterns of the GATA transcription factors in three species of Gossypium genus. Gene. 2019;680:72–83. Epub 2018 Sep 27. https://doi.org/10.1016/j.gene.2018.09.039 PubMed PMID: 30253181

**111.** Yu C, Li N, Yin Y, Wang F, Gao S, Jiao C, et al. Genome-wide identification and function characterization of GATA transcription factors during development and in response to abiotic stresses and hormone treatments in pepper. J Appl Genet. 2021;62(2):265–80. Epub 2021 Feb 25. https://doi.org/10.1007/s13353-021-00618-3 PubMed PMID: 33624251

**112.** Fan M, Xu C, Xu K, Hu Y. Lateral organ boundaries domain transcription factors direct callus formation in Arabidopsis regeneration. Cell Res. 2012;22(7):1169–80. https://doi.org/10.1038/cr.2012.63 PMID: 22508267

**113.** Rubin G, Tohge T, Matsuda F, Saito K, Scheible W-R. Members of the LBD family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in Arabidopsis. Plant Cell. 2009;21(11):3567–84. https://doi.org/10.1105/tpc.109.067041 PMID: 19933203

**114.** Shuai B, Reynaga-Pena CG, Springer PS. The lateral organ boundaries gene defines a novel, plant-specific gene family. Plant Physiol. 2002;129(2):747–61. https://doi.org/10.1104/pp.010926 PMID: 12068116

**115.** Cao H, Huang P, Zhang L, Shi Y, Sun D, Yan Y, et al. Characterization of 47 Cys2 -His2 zinc finger proteins required for the development and pathogenicity of the rice blast fungus Magnaporthe oryzae. New Phytol. 2016;211(3):1035–51. Epub 2016 Apr 05. https://doi.org/10.1111/nph.13948 PubMed PMID: 27041000

**116.** Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science. 2000;290(5499):2105–10. Epub 2000 Dec 16. https://doi.org/10.1126/science.290.5499.2105 PubMed PMID: 11118137

**117.** Jarvis P, López-Juez E. Biogenesis and homeostasis of chloroplasts and other plastids. Nat Rev Mol Cell Biol. 2013;14(12):787–802. Epub 2013 Nov 23. https://doi.org/10.1038/nrm3702 PubMed PMID: 24263360

**118.** Powell AL, Nguyen CV, Hill T, Cheng KL, Figueroa-Balderas R, Aktas H, et al. Uniform ripening encodes a Golden 2-like transcription factor regulating tomato fruit chloroplast development. Science. 2012;336(6089):1711–5. Epub 2012 Jun 30. https://doi.org/10.1126/science.1222218 PubMed PMID: 22745430

**119.** Rossini L, Cribb L, Martin DJ, Langdale JA. The maize golden2 gene defines a novel class of transcriptional regulators in plants. Plant Cell. 2001;13(5):1231–44. https://doi.org/10.1105/tpc.13.5.1231 PubMed PMID: 11340194; PubMed Central PMCID: PMCPMC135554

**120.** Murmu J, Wilton M, Allard G, Pandeya R, Desveaux D, Singh J, et al. Arabidopsis GOLDEN2-LIKE (GLK) transcription factors activate jasmonic acid (JA)-dependent disease susceptibility to the biotrophic pathogen Hyaloperonospora arabidopsidis, as well as JA-independent plant immunity against the necrotrophic pathogen Botrytis cinerea. Mol Plant Pathol. 2014;15(2):174–84. Epub 2014 Jan 08. https://doi.org/10.1111/mpp.12077 PubMed PMID: 24393452; PubMed Central PMCID: PMCPMC6638812

**121.** Savitch LV, Subramaniam R, Allard GC, Singh J. The GLK1 'regulon' encodes disease defense related proteins and confers resistance to Fusarium graminearum in Arabidopsis. Biochem Biophys Res Commun. 2007;359(2):234–8. Epub 2007 May 30. https://doi.org/10.1016/j.bbrc.2007.05.084 PubMed PMID: 17533111

**122.** Schreiber KJ, Nasmith CG, Allard G, Singh J, Subramaniam R, Desveaux D. Found in translation: high-throughput chemical screening in Arabidopsis thaliana identifies small molecules that reduce

PLOS ONE

Genome-wide identification and characterization of FORMIN genes in cotton (*Gossypium hirsutum* L.)

Fusarium head blight disease in wheat. Mol Plant Microbe Interact. 2011;24(6):640–8. Epub 2011 Feb 10. https://doi.org/10.1094/MPMI-09-10-0210 PubMed PMID: 21303209

123. Piya S, Shrestha SK, Binder B, Stewart CN Jr, Hewezi TJF. Protein-protein interaction and gene co-expression maps of ARFs and Aux/IAAs in Arabidopsis. Front Plant Sci. 2014;5:744. https://doi.org/10.3389/fpls.2014.00744 PMID: 25566309

124. Deeks MJ, Cvrcková F, Machesky LM, Mikitová V, Ketelaar T, Zársky V, et al. Arabidopsis group Ie formins localize to specific cell membrane domains, interact with actin-binding proteins and cause defects in cell expansion upon aberrant expression. New Phytol. 2005;168(3):529–40. https://doi.org/10.1111/j.1469-8137.2005.01582.x

125. Li Y, Shen Y, Cai C, Zhong C, Zhu L, Yuan M, et al. The type II Arabidopsis formin14 interacts with microtubules and microfilaments to regulate cell division. Plant Cell. 2010;22(8):2710–26. https://doi.org/10.1105/tpc.110.075507 PMID: 20709814

126. Chu Y, Bai W, Wang P, Li F, Zhan J, Ge XJIC. The mir390-GhCEPR2 module confers salt tolerance in cotton and Arabidopsis. Ind Crops Prod. 2022;190:115865. https://doi.org/10.1016/j.indcrop.2022.115865

127. Salih H, Gong W, He S, Xia W, Odongo MR, Du XJB. Long non-coding RNAs and their potential functions in Ligon-lintless-1 mutant cotton during fiber development. 2019;20(1):1–16.

128. Zhang B, Zhang X, Liu G, Guo L, Qi T, Zhang M, et al. A combined small RNA and transcriptome sequencing analysis reveal regulatory roles of miRNAs during anther development of Upland cotton carrying cytoplasmic male sterile Gossypium harknessii (D2) cytoplasm. 2018;18(1):1–17.

129. Deng F, Zhang X, Wang W, Yuan R, Shen FJB. Identification of Gossypium hirsutum long non-coding RNAs (lncRNAs) under salt stress. 2018;18(1):1–14.

130. Su H, Zhang S, Yuan X, Chen C, Wang X-F, Hao Y-JJP, et al. Genome-wide analysis and identification of stress-responsive genes of the NAM–ATAF1, 2–CUC2 transcription factor family in apple. 2013;71:11–21.

131. Shivanna KR. The pistil: structure in relation to its function. Reproductive ecology of flowering plants: patterns and processes; 2020. p. 41–50.

132. Roeber VM, Bajaj I, Rohde M, Schmülling T, Cortleven AJP. Light acts as a stressor and influences abiotic and biotic stress responses in plants. Plant Cell Environ. 2021;44(3):645–64. https://doi.org/10.1111/pce.13948 PMID: 33190307

133. Lei C, Bagavathiannan M, Wang H, Sharpe SM, Meng W, Yu JJA. Osmopriming with polyethylene glycol (PEG) for abiotic stress tolerance in germinating crop seeds: a review. Agronomy. 2021;11(11):2194. https://doi.org/10.3390/agronomy11112194