Genome Wide Identification and Characterization of Light-Harvesting Chloro a/b Binding Genes Reveals their Potential Role in Enhancing Drought Tolerance in Gossypium hirsutum

Teame Gereziher  
CAAS-ICR  https://orcid.org/0000-0001-7749-1811

Yanchao Xu  
Chinese Academy of Agricultural Sciences Cotton Research Institute

Richard Odongo Magwanga  
Chinese Academy of Agricultural Sciences

Joy Nyangasi Kirungu  
Chinese Academy of Agricultural Sciences

Xiaoyan Cai  
Chinese Academy of Agricultural Sciences Cotton Research Institute

Yuqing Hou  
Chinese Academy of Agricultural Sciences Cotton Research Institute

Yuhong Wang  
Anyang Institute of Technology

Kunbo Wang  
Chinese Academy of Agricultural Sciences cotton research Institute

Zhongli Zhou  
Chinese Academy of Agricultural Sciences Harbin Veterinary Research Institute

Fang Liu (liufcri@163.com)  
Chinese Academy of Agricultural Sciences  https://orcid.org/0000-0002-6213-9572

Research

Keywords: Cotton, G. hirsutum, LHC genes, Gene expression, Drought tolerance

Posted Date: October 30th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-97630/v1

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

Background

Cotton is an important commercial crop for its valuable source of natural fiber. Its production has undergone a sharp failure because of abiotic stress influences, of significance is drought. Moreover, plants have evolved self-defense mechanisms against the effects of several ways of abiotic factors like drought, salt, cold among others. The evolution of stress responsive transcription factors such as the trihelix, a nodule-inception-like protein (NLP), the late embryogenesis abundant (LEA) proteins among others have shown positive response in improving resistance to several forms of abiotic stress features.

Results

Genome wide identification and characterization of the effects of Light-Harvesting Chlorophyll a/b binding (LHC) genes was carried out in cotton under drought stress conditions. A hundred and nine proteins encoded by the LHC genes were found in the cotton genome, with 55, 27, and 27 genes found to be distributed in Gossypium hirsutum, G. arboreum, and G. raimondii, respectively. The proteins encoded by the genes were unevenly distributed in various chromosomes. The Ka/Ks values were less than one, and an indication of negative selection of the gene family. differential expression arrangement of genes was showed with the majority of the genes being highly upregulated in the root tissues in relative to leave and stem tissues. Moreover, more genes were induced in M85 a relative drought tolerant germplasm.

Conclusion:

The results provide proof of the possible role of the LHC genes in improving drought stress tolerance, and can be explored by cotton breeders in releasing a more drought tolerant cotton germplasms.

Introduction

Over the course of the 21st century, food production has to match the increasing population (Beddington et al., 2012). However, Temperature increment and climate change have deepened the incidence and harshness of abiotic stresses that critically disturb the growth and development of crops (Nouri, Moumeni & Komatsu, 2015). Abiotic stress remains one of the key components of yield loss in plant production (Sasi et al., 2018). Moreover, abiotic stress has a major impact on plant growth and development compared to other forms of living organisms due to their immobility (He, He & Ding, 2018; Magwanga et al., 2018, 2019; Xu et al., 2019). Among the various forms of abiotic stress factors drought, heat, toxicity, and salinity do cause over-reduction of the electron transport chain (ETC) resulting in photooxidation (Nishiyama & Murata, 2014). Furthermore, in the chloroplasts, drought, high light, salinity, or extreme temperatures stresses do trigger a diminishing in CO₂ assimilation rates which in turn induce an upsurge in reactive oxygen species creation, which eventually leads to yield damage (Pintó-marijuan & Munné-bosch, 2014). It has been reported that abiotic stresses account for over 50% of losses in crop production
(Nath et al., 2013). Moreover, a decrease in photosynthesis results in a remarkable reduction in yield and yield quality in crops (Nouri, Moumeni & Komatsu, 2015).

Drought exposure alters the photosynthetic apparatus in the plants, and thus plants have evolved numerous coping mechanisms, one of which is the evolution of various plant transcription factors. (Hussain et al., 2018). The known plants genes with net effects on plant photosynthetic process includes, Ribulose bisphosphate carboxylase large chain (rbcL) (Berry & Yerramsetty, 2013), Cytochrome f (petA) (Xing et al., 2016), light-harvesting chlorophyll a/b-binding (LHC) (Zhao et al., 2020a), cytochrome p450 genes (Magwanga, Lu & Kirungu, 2019) among others. In crops, the LHCA and LHCB sub-families, which encode proteins founding the light-harvesting complex of photosystems I and II in LHC gene family (Fanna et al., 2016). The LHC proteins are the apoproteins of the Light-Harvesting complex of photosystem II (PSII), outer antenna complex which is perhaps the utmost ample membrane proteins in nature (Kró et al., 1995; Horton & Ruban, 2005; Xu et al., 2012). Moreover, studies have shown that LHCB1, LHCB2, LHCB3, LHCB4, LHCB5, or LHCB6, affects stomatal responsiveness to abscisic acid (ABA) influx, and therefore lowers the plant’s tolerance level to drought stress during their down-regulation (Xu et al., 2012). Furthermore, downregulation of the LHCB genes does cause ABA insensitive phenotypes in seed germination and post-germination growth (Liu et al., 2013a). In the recognition of the proteins encoded by the LHCB genes, 28 have been identified in Papaya carica (Zou et al., 2020a), 17 in Hordeum vulgare L. (Qin et al., 2017), 25 in Camellia sinensis (Li et al., 2020), and 35 genes in Manihot esculenta (Zou & Yang, 2019), However, the role of this important plant gene family concerning abiotic stress factors in cotton have not been studied. The complete sequencing of Gossypium hirsutum (Hu et al., 2019), Gossypium arboreum (Huang et al., 2020), and Gossypium raimondii (Wang et al., 2012; Agricultural et al., 2019), provided the needed information to carry out functional analysis of the proteins encoded by the LHC genes in the three cotton genomes.

Materials And Methods

Plant material and Hydroponics

The experiment was laid out in CRD design with three biological replications in Green house. A seed of Marie-galante 85 (M85), a race developed from Gossypium hirsutum species and comparatively tolerant to Abiotic stress (Chen et al., 2018). The seeds were treated by water soaking for one night and sow them in absorbent paper for germination, after a week then transplanted to a hydroponic set up that have Hoagland nutrient solution (Gene & Consortium, 2000), in the greenhouse, with 16 h/8h light-dark and temperature at 28 °C day/25 °C night (Zhao et al., 2020b). At three leaf stage, drought stress was imposed by supplementing the nutrient solution with 17% of PEG-6000 (Liu et al., 2013b). The leaf, stem, and root tissues were then collected for RNA extractions at 0 h, 3 h, 6 h, 9 h, 12 h, and 24 h after stress exposure.

Identification of the Light-Harvesting Chlorophyll a/b-binding Proteins in Cotton Species
The Light-Harvesting Chloro a/b bind domain number PF00504 was used as to search for the identifications of the cotton proteins encoded by the LHC genes. The LHC proteins for *G. hirsutum*, *G. raimondii* and *G. arboreum* were downloaded from the cotton functional genomics database (www.cottonfgd.org), while those for *Arabidopsis thaliana*, and *Theobroma cacao* were downloaded from phytozome (https://phytozome.jgi.doe.gov). The HMM profiles of the LHC functional domain PF00504 were retrieved from the Pfam database (http://pfam.xfam.org) and used for the identification of the putative ALDH proteins with the best domain e-value cutoffs \( f < 1 \times 10^{-4} \) (El-Gebali et al., 2019). Moreover, To get the, physicochemical traits of the gene family like Protein length (PL) molecular weight (MW) and molecular charge, isoelectric point (pI) and GRAVY value using the website of CottonFGD (www.cottonfgd.org).

**Phylogenetic tree and collinearity analysis**

Protein sequences of these three cotton species including *Arabidopsis thaliana* and *Theobroma cacao* were aligned by ClustalX in MEGA 7.0 for phylogenetic tree construction. We use Neighbor-joining (NJ) method to know the evolution distance, Jones–Taylor–Thornton (JTT) as substitution model of 1000 bootstrap replication (Tamura et al., 2011). To categorize the homologous genes of cotton species, the protein sequences of *G. hirsutum* were exposed to a BlastP search alongside the protein database of *G. arboreum* and *G. raimondii*; hits with E-values \( \leq 1 \times 10^{-5} \) and \( \geq 90\% \) similarity were enabled significant. The GFF3 file, linked file, and Gene ID were applied to construct the Collinearity analysis by TBtools software (Chen et al., 2018). Homologous genes of *G. hirsutum*, *G. raimondii* and *G. arboreum* were known from CottonFGD employing BLASTp with a threshold of >80% match and at least an 80% alignment ratio based on the protein length.

**Chromosome mapping, Gene Ontology, and Cis-regulatory elements analysis**

To know the distribution of Light-Harvesting Chloro a/b-bind genes in all the chromosomes of A, D, and AD cotton genomes, we used the GFF3 file from CottonFGD (www.cottonfgd.org) and gene ID of the genes. Then employed the TBtools software to show the genes on chromosome via amazing gene location from Gene Transfer Format/General Feature Format (GTF/GFF).

Cellular component (CC), biological process (BP) and molecular functions (MF) was used to determine the functional classification of genes by an online tool AgriGO (www.bioinfo.cau.edu.cn/agriGO) (Gene & Consortium, 2000). Analysis of the gene structure of the Light-Harvesting Chloro a/b-bind genes in *G. hirsutum*, *G. arboreum*, and *G. raimondii* was done by means of the Gene Structure Display Server – GSDS 2.0 (http://gsds.cbi.pku.edu.cn) online tool. While for the motif identification, an online tool MEME was employed (http://meme-suite.org/).

The 2000-bp upstream sequences of CAB genes of cotton species were downloaded from CottonFGD (http://www.cottonfgd.org/) to identify the cis-regulatory elements in the putative promoter regions. Thus the fasta file of the upstream sequence was submitted to Plant Care search
(https://pubmed.ncbi.nlm.nih.gov/11752327) for identifying the putative cis–regulatory elements among the promoter sequences (Lescot et al., 2002). The structure was visualized by TBtools.

**RNA extraction and RT-qPCR analysis**

At three leaf stages, drought stresses were forced by adding the nutrient solutions with 17% PEG-6000 solution as previously adopted by Magwanga et al [36,37]. Samples were then collected for RNA extraction at 0 h, 3 h, 6 h, 9 h, 12 h and 24 h of post stress treatment. Total RNA was extracted using TIANGEN, RNA preppure plant plus kit (www.tiangen.com) according to the manufacturer guidelines. Nano Drop 2000 was used to check the quality and concentration of RNA extracted with a standard of 260/280 which must be between 1.80–2.1 (Joshi et al., 2016). Thus, we convert the RNA to cDNA using TransGen Biotech kit Beijing, China (www.Transgen.com.cn), following the kit instructions. From the LHC gene family, we select 27 genes for RT-qPCR and design the primers (Table S1) using NCBI website (www). For the RT-qPCR analysis, we use 7500 fast real time system with 2µL, 2µL 6µL, and 10µL of cDNA, forward and reverse Primers, RNA free water, and SYBR solution respectively. Three biological and technical replications were used in the whole analysis with Ghactin7 as control. \( E = 2^{-\Delta\Delta Ct} \) formula uses to calculate the gene expression. (Schmittgen & Livak, 2008)

**Results**

**Identification of the Cotton LHC proteins**

109 proteins translated by the *LHC genes* were recognized in the three sequenced cotton genomes, with 55, 27, and 27 proteins in *G. hirsutum* (AD), *G. raimondii* (D) and *G. arboreum* (A), respectively (Table S2). The amounts of the proteins found in the *LHC genes* in the two diploid cotton species, *G. raimondii* and *G. arboreum* were less by one compare with the number of LHC proteins in *G. hirsutum*, may be due to AD emerged in the whole genome duplications between A and D genomes.

The evaluation of the physicochemical properties *G. hirsutum* of the Chloro a/b binding protein genes, the protein lengths for the *G. hirsutum* proteins stretched from 62 aa to 644 aa, molecular weights reached from 6.88 kDa to 72.66 kDa were scored respectively in *Gh_Sca017783G01* and *Gh_A02G1068*, a charge ranged from –8.5 (*Gh_A01G0519*) to 7(*Gh_A02G1068*), the isoelectric point (\( pI \)) ranged from 4.701 (*Gh_D06G2350*) to 10.228 (*Gh_D04G1505*) and finally the grand average of hydropathy (GRAVY) ranged from –0.529 (*Gh_A01G0519*) to 0.233 (*Gh_D06G2350*) (Table 1).
| Gene ID         | PL (aa) | MW(KDa) | Charge | IEP   | GRAVY |
|----------------|---------|---------|--------|-------|-------|
| Gh_D11G1504    | 272     | 29.46   | 4.5    | 9.273 | -0.089|
| Gh_A05G2108    | 291     | 30.991  | -2.5   | 5.416 | -0.005|
| Gh_A06G1447    | 279     | 30.514  | -1.5   | 5.793 | -0.002|
| Gh_A01G0976    | 268     | 29.312  | 5      | 7.942 | 0.051 |
| Gh_A10G0361    | 264     | 28.182  | -4.5   | 4.904 | -0.004|
| Gh_A04G0961    | 260     | 27.85   | 4      | 9.316 | 0.053 |
| Gh_D07G1663    | 265     | 28.609  | -3.5   | 5.084 | -0.001|
| Gh_D07G1659    | 265     | 28.609  | -3.5   | 5.084 | -0.001|
| Gh_A07G2182    | 265     | 28.608  | -2.5   | 5.335 | -0.001|
| Gh_D06G1791    | 282     | 30.791  | -1.5   | 5.793 | -0.02 |
| Gh_A04G0218    | 262     | 28.483  | -4.5   | 4.89  | 0.023 |
| Gh_D01G1508    | 291     | 31.443  | 0      | 6.504 | 0.11  |
| Gh_D05G1429    | 265     | 28.101  | -2.5   | 5.329 | 0.027 |
| Gh_D02G1996    | 259     | 27.33   | 2      | 8.425 | 0.126 |
| Gh_D05G3484    | 262     | 28.449  | -4.5   | 4.89  | 0.026 |
| Gh_D01G1028    | 268     | 29.391  | 6      | 8.203 | 0.01  |
| Gh_A01G1349    | 297     | 32.139  | 1.5    | 7.004 | 0.097 |
| Gh_A11G2259    | 166     | 18.205  | 2      | 8.741 | 0.07  |
| Gh_A07G2184    | 265     | 28.764  | -1.5   | 5.756 | -0.036|
| Gh_D01G2232    | 285     | 31.101  | -3     | 5.378 | -0.048|
| Gh_D04G1505    | 240     | 25.65   | 9      | 10.228| 0.037 |
| Gh_A07G1725    | 265     | 28.407  | -3.5   | 5.071 | 0.036 |
| Gh_D01G0531    | 261     | 28.405  | -4.5   | 4.89  | 0.011 |
| Gh_D12G1495    | 304     | 33.898  | 7      | 8.922 | 0.037 |
| Gh_A10G0616    | 285     | 30.716  | -1.5   | 5.805 | 0.021 |
| Gh_D07G0661    | 252     | 27.857  | 2.5    | 7.551 | -0.128|
| Gene ID                  | PL (aa) | MW(KDa) | Charge | IEP   | GRAVY  |
|-------------------------|---------|---------|--------|-------|--------|
| Gh_D05G2361             | 291     | 31.003  | 2.5    | 5.416 | 0.013  |
| Gh_A10G2108             | 273     | 29.857  | -2.5   | 5.793 | -0.059 |
| Gh_A07G0594             | 252     | 27.841  | 2.5    | 7.551 | -0.118 |
| Gh_A12G1617             | 252     | 27.946  | 3.5    | 7.968 | -0.117 |
| Gh_D10G0369             | 264     | 28.21   | -3.5   | 5.085 | -0.008 |
| Gh_D07G1929             | 265     | 28.407  | -3.5   | 5.071 | 0.036  |
| Gh_D05G0860             | 247     | 26.797  | 0.5    | 6.675 | -0.116 |
| Gh_A13G0222             | 246     | 26.76   | 0.5    | 6.675 | -0.164 |
| Gh_A05G1261             | 265     | 28.101  | -2.5   | 5.329 | 0.027  |
| Gh_D12G1757             | 252     | 27.953  | 2.5    | 7.6   | -0.124 |
| Gh_D10G0784             | 288     | 31.051  | -0.5   | 6.294 | -0.022 |
| Gh_A01G1972             | 285     | 31.107  | -3     | 5.378 | -0.071 |
| Gh_D13G0236             | 246     | 26.744  | 0.5    | 6.675 | -0.153 |
| Gh_A05G0725             | 247     | 26.797  | 0.5    | 6.675 | -0.116 |
| Gh_D03G0610             | 349     | 38.293  | 2.5    | 7.543 | 0.084  |
| Gh_D10G2385             | 273     | 29.883  | -1.5   | 5.789 | -0.073 |
| Gh_Sca123119G01         | 90      | 9.881   | -0.5   | 5.795 | -0.397 |
| Gh_Sca053293G01         | 143     | 15.811  | -3     | 5.138 | 0.158  |
| Gh_Sca017783G01         | 62      | 6.884   | -1     | 4.879 | 0.105  |
| Gh_D06G2350             | 192     | 20.154  | -4     | 4.701 | 0.233  |
| Gh_D06G2351             | 265     | 28.165  | -2.5   | 5.329 | 0.014  |
| Gh_A07G2366             | 281     | 31.181  | -1     | 6.128 | 0.064  |
| Gh_A03G2154             | 259     | 27.423  | 2      | 8.428 | 0.107  |
| Gh_D07G0125             | 261     | 28.763  | -2     | 5.7   | -0.029 |
| Gh_A11G1357             | 272     | 29.45   | 4.5    | 9.273 | -0.086 |
| Gh_A13G1282             | 166     | 18.169  | 2.5    | 8.417 | 0.051  |
| Gh_D06G2120             | 151     | 16.501  | 0.5    | 7.256 | -0.041 |
| Gh_A01G0519             | 466     | 51.35   | -8.5   | 5.178 | -0.529 |
| Gene ID      | PL (aa) | MW(KDa) | Charge | IEP  | GRAVY |
|-------------|---------|---------|--------|------|-------|
| Gh_A02G1068| 644     | 72.664  | 7      | 7.727| -0.2  |
| Ga01G0731  | 482     | 53.372  | -6     | 5.503| -0.377|
| Ga01G1437  | 265     | 28.978  | 2.5    | 7.705| 0.035 |
| Ga02G0756  | 610     | 68.741  | 9      | 8.188| -0.241|
| Ga02G1050  | 270     | 29.006  | 2      | 7.45 | 0.008 |
| Ga03G2284  | 228     | 24.599  | -1.5   | 5.716| 0.084 |
| Ga04G1033  | 114     | 12.823  | 5.5    | 9.897| -0.255|
| Ga05G0924  | 247     | 26.797  | 0.5    | 6.675| -0.116|
| Ga05G1596  | 265     | 28.101  | -2.5   | 5.329| 0.027 |
| Ga05G2647  | 291     | 30.991  | -2.5   | 5.146| -0.005|
| Ga05G4015  | 262     | 28.449  | -4.5   | 4.89 | 0.026 |
| Ga05G4018  | 262     | 28.495  | -4.5   | 4.89 | 0.042 |
| Ga06G2006  | 282     | 30.819  | -1.5   | 5.793| -0.011|
| Ga06G2455  | 167     | 18.453  | 4      | 8.634| -0.057|
| Ga07G0172  | 261     | 28.86   | 1.5    | 7.045| -0.048|
| Ga07G0768  | 252     | 27.841  | 2.5    | 7.551| -0.118|
| Ga07G1916  | 265     | 28.665  | -2.5   | 5.335| -0.02 |
| Ga07G1918  | 265     | 28.607  | -0.5   | 6.271| -0.002|
| Ga07G2205  | 265     | 28.407  | -3.5   | 5.071| 0.036 |
| Ga10G0035  | 273     | 29.897  | -1.5   | 5.793| -0.073|
| Ga10G2236  | 271     | 29.226  | -1.5   | 5.791| 0.043 |
| Ga10G2674  | 264     | 28.182  | -4.5   | 4.904| -0.004|
| Ga11G2486  | 272     | 29.477  | 4.5    | 9.273| -0.096|
| Ga12G1052  | 252     | 27.946  | 3.5    | 7.968| -0.117|
| Ga12G1366  | 300     | 32.84   | 6.5    | 8.712| -0.079|
| Ga13G0254  | 200     | 21.273  | 3.5    | 9.544| 0.167 |
| Ga13G0268  | 246     | 26.76   | 0.5    | 6.675| -0.164|
| Ga14G0061  | 285     | 31.03   | -2     | 5.724| -0.083|
| Gene ID          | PL (aa) | MW(KDa) | Charge | IEP   | GRAVY |
|-----------------|---------|---------|--------|-------|-------|
| Gorai.001G016400 | 261     | 28.749  | -2     | 5.696 | -0.029|
| Gorai.001G074300 | 252     | 27.857  | 2.5    | 7.551 | -0.128|
| Gorai.001G192000 | 265     | 28.575  | -3.5   | 5.084 | 0.003 |
| Gorai.001G192300 | 265     | 28.609  | -3.5   | 5.084 | -0.001|
| Gorai.001G220900 | 265     | 28.393  | -3.5   | 5.071 | 0.035 |
| Gorai.002G076400 | 262     | 28.505  | -4.5   | 4.89  | 0.027 |
| Gorai.002G132100 | 268     | 29.357  | 6      | 8.203 | 0.016 |
| Gorai.002G183400 | 292     | 31.461  | -1     | 6.114 | 0.12  |
| Gorai.002G263900 | 285     | 31.144  | -2     | 5.73  | -0.077|
| Gorai.003G092700 | 349     | 38.267  | 2.5    | 7.493 | 0.081 |
| Gorai.005G219000 | 259     | 27.33   | 2      | 8.425 | 0.126 |
| Gorai.007G163200 | 285     | 31.695  | 7.5    | 9.296 | -0.249|
| Gorai.008G165200 | 313     | 34.49   | 6      | 8.855 | 0.071 |
| Gorai.008G194000 | 252     | 27.983  | 2.5    | 7.5   | -0.114|
| Gorai.009G090600 | 247     | 26.821  | 1      | 6.79  | -0.155|
| Gorai.009G156900 | 265     | 28.101  | -2.5   | 5.329 | 0.027 |
| Gorai.009G262200 | 291     | 30.991  | -2.5   | 5.416 | -0.005|
| Gorai.009G430800 | 262     | 28.479  | -4.5   | 4.897 | 0.016 |
| Gorai.010G165000 | 192     | 20.226  | -4     | 4.701 | 0.244 |
| Gorai.010G165100 | 265     | 28.165  | -2.5   | 5.329 | 0.014 |
| Gorai.010G198600 | 282     | 30.764  | -2.5   | 5.391 | -0.008|
| Gorai.010G239300 | 151     | 16.55   | 0.5    | 7.254 | -0.026|
| Gorai.011G041600 | 264     | 28.21   | -3.5   | 5.085 | -0.008|
| Gorai.011G089000 | 217     | 23.895  | 0.5    | 7.235 | 0.048 |
| Gorai.011G285900 | 273     | 29.883  | -1.5   | 5.789 | -0.073|
| Gorai.012G141200 | 260     | 27.805  | 3      | 8.943 | 0.055 |
| Gorai.013G026000 | 246     | 26.744  | 0.5    | 6.675 | -0.153|
In the two diploid cotton species, the *G. arboreum* and *G. raimondii* LHC proteins physiochemical properties exhibited slight differences, in molecular weights, protein lengths, pI, molecular charge, and GRAVY values. The protein length stretched from 114 aa to 610 aa, and 151 aa to 349, molecular weights ranged from 12.823 to 68.741 KDa, and 16.55 to 38.267 KDa by a charge range of −6 to 9 and −4.5 to 7.5 in *G. arboreum* and *G. raimondii*, respectively (Table 1).

On the other hand, the values for pI and GRAVY was almost the same, pI ranges from 4.87 to 9.897, and 4.701 to 9.296, GRAVY −0.377 to 0.167 and −0.249 to 0.244 in order of *G. arboreum* and *G. raimondii*. In all cotton species, the GRAVY value was lower (positive and negative), which indicates all proteins may be a sign of the likelihood of enhanced relations with water that leads to hydrophilic nature.

**Phylogenetic Tree and Synteny block Analysis of the Cotton LHC Proteins**

The phylogenetic tree constructed grouped the cotton Light-Harvesting Chloro a/b binding proteins together with other plants into 12 clades. Numerous homolog gene pairs were formed among the several proteins encrypted by the cotton Light-Harvesting Chloro a/b binding genes (Fig. 1A).

The collinearity analysis among the three cotton species was analyzed, in which Circle gene viewer was applied to distinguish the collinear gene pairs with TBtools software (Chen et al., 2018). Finally, the collinearity analysis between the genetic map of At and Dt Subgenomes of *G. hirsutum*, *G. arboreum* and *G. raimondii* for their A Vs D; A vs At, and finally between D Vs Dt Subgenome relationships were observed. We found good collinearity between A vs D with 23 genes, A vs At with 20 genes, and finally between D vs Dt with 23 genes in the Subgenome (Fig. 1B).

**Gene Ontology Analysis**

Gene Ontology (GO) has a structure that allows powerful comparisons and inferences about gene functions in biological, cellular, and molecular levels (Gene & Consortium, 2000). Presumed functions of 109 genes in the *Gossypium* Light-Harvesting Chloro a/b-bind gene family, including biological processes (BP), molecular functions (MF), and cellular components (CC) were identified using agriGO online analysis.

In *G. hirsutum* biological processes (GO: 0008150), the functions included cellular and metabolic processes. Various cellular (GO: 0005575) functions were noted in the cell and cell part. Similarly, in *G. arboreum*, the biological (GO: 0008150) functions were responsible for stimuli, cellular and metabolic processes. In cellular component (GO: 00055750), the functions were focused on cell, macromolecular complex (Protein), and membrane related issues, whereas in molecular function (GO: 0003674), were related with binding function. In *G. raimondii* the biological process (GO: 0008150) was coined with cellular and metabolic processes, which is similar to *G. hirsutum*, whereas in cellular component (GO: 0005575), the function is related to membrane. In both *G. hirsutum* and *G. raimondii*, there is no significant GO term in molecular function (Fig. 2).
Gene Structure and Motif Identification of Chloro a/b-bind Proteins

Gene structural study is observed as a likely sign of the evolution of multigene families. To obtain additional evidence into the structural diversity of cotton Light-Harvesting Chloro a/b-bind genes, the exon/intron association in the full-length cDNAs was investigated in contrast with their equivalent genomic DNA sequences of distinct genes in *G. hirsutum*, and it was found that a higher proportion of the Light-Harvesting Chloro a/b-bind genes and their exons were extremely conserved inside the group. Gene structural diversity is regarded as a possible indicator of the evolution of multigene families. To gain further information into the structural diversity of cotton Light-Harvesting Chloro a/b-bind genes, the exon/intron organization in the full-length cDNAs was analyzed in comparison with their corresponding genomic DNA sequences of individual genes in *G. hirsutum*, and it was identified that a greater percentage of the Light-Harvesting Chloro a/b-bind genes and their exons were highly conserved within the group.

In the study of the gene structures, some of the Light-Harvesting Chloro a/b-bind gene structures were disturbed by introns. The maximum level of intron disruption of the Chloro a/b-bind gene structures was 11(Gh_A02G1068), 11(Ga02G0756), and 5 (Gorai.003G092700) for *G. hirsutum*, *G. arboreum* and *G. raimondii*, respectively. Light-Harvesting Chloro a/b-bind genes are mostly found with the occurrence of two exons and one intron. The highest number of exons and introns were found in Gh_A02G1068 (12 exons, 11 introns) and Gh_A01G0519 (10 exons, 9 introns). Remarkably, Exons and introns for diverse Light-Harvesting Chloro a/b-bind genes were observed to be dissimilar based on their lengths. For example, 18 genes had to have two exons and one intron and 7 genes three exons by two introns and seven genes with one exon and no intron. (Fig. 3).

On the other hand, in the diploid species, the maximum number of exon/intron were 12 exons, 11 introns (Ga02G0756) and 11 exons, 10 introns (Ga01G0731) in *G. arboreum*, 6 exons, 5 introns (Gorai.003G092700) and 6 exons, 5 introns (Gorai.009G262000) in *G. raimondii*, respectively. Similarly, the number of genes that have two exons with one intron is seven and ten in *G. arboreum* and *G. raimondii*. Genes with three exons and two introns as well as a single exon with no intron were five and three respectively in both species. To explore the structural evolution of LHC proteins, the patterns of motifs were analyzed. A total of 20 different motifs were detected by the MEME analysis (http://meme-suite.org/) in the three Gossypium species (Fig. 4). Based on the identified motifs, motif 3, motif 4 and motif 12 are the conserved motifs in the *G. hirsutum*, whereas motif 2 and 8 in *G. arboreum* and while motif 11 and 4 in *G. raimondii*, respectively.

Chromosomal Mapping Analysis of the Light-Harvesting Chloro a/b binding Genes

The *LHC* genes were evenly distributed across the various chromosomes of the *A*₂, *D*₅, and (AD)₁ cotton genomes. In the tetraploid (AD)₁ genome with At Subgenome, the highest gene loci were found on
chromosome A_t01, A_t05, and A_t10 with 3 genes, while At03, At08, and At09 chromosomes harbored none. Similarly, in the (AD)_{1}, Dt Subgenome, the highest gene loci were found in D_t07, D_t01, and D_t05 with 5, 4, and 4 genes, respectively, whereas A_t03, A_t08, and A_t09 had zero genes. The rest of the chromosome harbored between 1 to 3 genes (Fig. 5A and B). With the two diploid cotton species, A_{2} and D_{5} genomes, the gene distribution arrangement was different, In *G. arboreum*, the highest gene loci were observed on the chromosome, A_{2}05, and A_{2}07, with the same 4 genes while in *G. raimondii*, chromosome D_{5}01, D_{5}09, and D_{5}10 concealed the highest gene loci with 4 genes, respectively, while chromosome A_{2}04 and D_{5}06 harbored none (Fig. 5C and D).

Identification of Cis-regulatory elements

Cis-Acting regulatory elements are important molecular switches involved in the transcriptional regulation of a dynamic network of gene activities controlling various biological processes, including abiotic stress responses, hormone responses, and developmental processes. It encodes the genomic blueprints for coordinating spatiotemporal gene expression programs underlying highly specialized cell functions (Mao et al., 2020). In the plant Care analysis of Cis-regulatory elements ABRE, ARE, MRE, MYB, AT-rich elements, DRE, MBS, Box-4, and ACE were found related to drought stress in the three cotton species (Fig. 6). The major cis-acting elements, such as the ABA-responsive element (ABRE) and the dehydration-responsive element/C-repeat (DRE/CRT), that are a vital part of ABA-dependent and ABA-independent gene expression in osmotic and cold stress responses (Yamaguchi-Shinozaki & Shinozaki, 2005).

Evolution of LHC genes in Gossypium species

The Ks value in gene evolution was not affected by natural selection generally, but Ka does. The Ka/Ks value showed positive, neutral, and negative selection when the value was Ka/Ks > 1, Ka/Ks = 1, and Ka/Ks < 1 respectively (Zhao et al., 2020b). The distributions of Ka, Ks, and Ka/Ks among homologous pairs of Gossypium species were revealed similar results. (Fig. 7, Table S3) The Ka/Ks of GhAt-Ga ranged from 0–0.949034416, while for GhDt-Gr from 0–0.838286204. The Ka/Ks of GhAt-GhDt ranged from 0–0.523637063, whereas the Ka/Ks value of Ga-Gr was 0–0.755930549. In all the pairs, the Ka/Ks value was < 1 which indicated that the gene family was subjected to negative selection. The result suggested that the LHC of *G. hirsutum* genes derived from *G. raimondii* and *G. arboreum* experienced negative selection commands throughout the evolution.

RT-qPCR Validation of Light-Harvesting Chloro a/b binding genes under Water Deficit Conditions

Twenty-seven *LHC* genes expression profiles were carried out under drought stress conditions in different tissues and varying time intervals. The genes showed differential expression pattern on the tissues analyzed, in root tissues, the highly upregulated genes were *Gh_D10G2385, Gh_A13G0222, Gh_A05G0725, Gh_D05G0860, Gh_D07G0661, Gh_D01G1508, Gh_D12G1495, Gh_A07G2182*, and *Gh_A10G2108*, while in the leaf tissues, *Gh_A07G2184, Gh_D10G2385, Gh_D05G0860, Gh_D02G1996,
Gh_A13G0222, and Gh_A05G0725 showed higher upregulation after 12 h of stress exposure. Similarly, Gh_A13G0222, Gh_D06G1791, and Gh_A06G1447 genes were Up-regulated in stem tissues starting from 6 hours up to 24 hours (Fig. 8).

Most genes were Down-regulated mainly in leaf tissue followed by stem. Genes like Gh_A10G0361, Gh_D10G0369, Gh_A03G2154, and Gh_D03G0610 were Down-regulated in the three tissues of cotton in almost all time points. Generally, many genes were Up-regulated in the root tissue. Gh_A13G0222 (CAB6A) was Up-regulated in all tissue samples and Gh_D10G2385 (LHCB4), Gh_D05G0860 (CAB6A), and Gh_A05G0725 (CAB6A) also Up-regulated in Leaf and root tissues under drought stress. A detailed exploration of these genes will offer efficient information on considerate LHC genes in cotton (Gossypium) and its part in drought stress tolerance. Drought effect is first felt at the root zone, and the higher upregulation of various genes in the root tissues is in line with earlier results in which most of the LEA genes were upregulated in the root tissues in relative to leaf and stem tissues during drought stress situation (Magwanga et al., 2018).

**Discussion**

Drought is one of the key abiotic stresses that affect crop production worldwide. It also harshly affects the physiology and growth of many crops (Joshi et al., 2016). It was the main risk to a significant loss of cotton yield due to the ever-increasing shortage of water around the world (Hou et al., 2018). Drought stress damages photosynthetic pigments that usually begin with majorly stomatal effects at medium drought intensity, and come to an end in metabolic and structural alters caused by harsh drought stress. Photosynthesis stands for one of the greatest vital photo-chemical reactions in plants. Sunlight is transformed into chemical energy and is employed to change carbon dioxide, water, and minerals into oxygen and energy-rich organic composites then recycled as energy basis by heterotrophs (Gururani, Venkatesh & Tran, 2015).

Photosynthesis is the outcome of many steps and multipart developments that employs numerous biological pathways similar to photosynthetic electron transport system (PETs), makes sun light to transform into ATP and NADPH; in addition, CO₂ is fixed into carbohydrates, as well as assimilation, transport, and consumption of photo assimilates as the organic products of photosynthesis by Calvin-Benson cycle (Eberhard, Finazzi & Wollman, 2008; Foyer et al., 2012). Forming disorder of all photosynthesis mechanisms has the primary impact of abiotic stress on the activity of photosynthesis (Nouri, Moumeni & Komatsu, 2015). Photosynthetic reactions of mature crops and small seedlings to drought-stress are mainly diverse. In mature crops, efficient photosynthetic complexes are previously shaped and water-stress brings the creation of ROS due to surplus light absorption, which pressures the photosynthetic apparatus. Though, in water-stressed young seedlings, there is the likelihood to down-regulate Chl biosynthesis and slim down the production and gathering of light-harvesting complexes of PSI and PSII, and to acclimatize crops not to suck up surplus light, which is damaging (Dalal & Tripathy, 2018). Chloroplast was the main research area in the field of biology because it was the site for
photosynthesis. But it is also a very sensitive structure to biotic and abiotic stresses and indicates the real status in crops response to stress (Liu et al., 2013b; Li et al., 2020).

Light-harvesting chlorophyll a/b-binding (LHC) proteins contain a plant-specific superfamily comprised of photosynthesis and stress responses. Identifying genes of this family would help in studying the function and role of these genes in different crop species (Qin et al., 2017; Zou et al., 2020b). But we don’t get enough information in the cotton crop for this family. Previous studies in crops suggested that there was an important link between photosynthesis and final yield. Light-harvesting complex II (LHCII) is a central component of the photosynthesis, with fundamental parts in light harvest and acclimation to changing light (Longoni et al., 2015; Qin et al., 2017).

In our result, many genes were Up-regulated in the root tissue. \textit{Gh_A13G0222} (CAB6A) was Up-regulated in all tissue samples while \textit{Gh_D10G2385} (LHCB4), \textit{Gh_D05G0860} (CAB6A), and \textit{Gh_A05G0725} (CAB6A) were Up-regulated in leaf and root tissues under drought stress. A study from tea plants showed that two genes, CsCP1 and CsCP2, were found to affect phosphorylation/ dephosphorylation and GTP in the physiological regulation of PS II. The regulation of LHC protein stages allows chloroplasts to answer amenably and quickly to abiotic stresses (Li et al., 2020). Similarly, a finding in Papaya, plants treated with mannitol for drought stress after 10 days, three genes were upregulated (\textit{CpELIP}, \textit{CpLhcb7}, and \textit{CpPsbS}), for 15 days, five genes upregulated (\textit{CpELIP}, \textit{CpSEP2}, \textit{CpOHP2}, \textit{CpLhcb7}, and \textit{CpPsbS}) and for 20 days, 12 genes were meaningfully regulated with five genes upregulated (\textit{CpELIP}, \textit{CpSEP2}, \textit{CpOHP2}, \textit{CpLhcb7}, and \textit{CpPsbS}) (Zou et al., 2020a).

The evolution of LHC genes in Gossypium species indicated that the circulations of Ka, Ks, and Ka/Ks were similar among homologous pairs. The Ka/Ks of GhAt-Ga reached from 0–0.949034416, while GhDt-Gr reached from 0–0.838286204. The Ka/Ks of GhAt-GhDt ranged from 0–0.523637063, whereas the Ka/Ks value of Ga-Gr was 0–0.755930549. The result suggested that the LHC of \textit{G. hirsutum} genes derived from \textit{G. raimondii} and \textit{G. arboreum} experienced negative selection instructions throughout the evolution. In harmony with this finding, the Ka/Ks value of cassava light-harvesting chlorophyll a/b-binding (LHC) genes ranges from 0.0010–0.2507 (Zou & Yang, 2019).

LHCB family members positively regulate crops Abiotic stress tolerance by stomatal closure to ABA signaling starting from germination to final growth (Xu et al., 2012; Liu et al., 2013b). It is well-identified that ABA persuades stomatal closure in water shortage conditions, which hinders photosynthesis. Here, the genetic evidence provides that members of the LHCB family are certainly elaborated in guard cell signalling in response to ABA and so LHCB members have been found as new actors in ABA signalling in stomatal movement (Xu et al., 2012). The LHCB members were exposed to be targets of ABA-responsive WRKY-domain transcription factor, for an inducer that modifies LHCB expression at least through suppressing the WRKY transcription repressor in stressful conditions in collaboration with light, which permits crops to adjust to eco-friendly encounters (Liu et al., 2013b). Functional genomics trials will have desirable and be accommodating to demonstrate the biological and molecular function of \textit{LHC} genes and to make use of them in cotton improvement.
Conclusions

A hundred and nine proteins encrypted by the LHC genes were found in the cotton genome, with 55, 27, and 27 genes found to be distributed in *Gossypium hirsutum, G. arboreum, and G. raimondii*, respectively. The majority of LHC genes showed with high exon-intron connections. Collinearity analysis and chromosomal mapping showed that LHC genes were dispersed on chromosomes of three *Gossypium* species, with most genes clustering in the upper and lower arm of chromosomes. In the three cotton species, their GRAVY value was lower (positive and negative), which indicated that the protein was hydrophilic nature. In the RT-qPCR, many genes were Up-regulated in the root tissue. *Gh_A13G0222 (CAB6A)* was Up-regulated in all tissue samples and *Gh_D10G2385 (LHCB4), Gh_D05G0860 (CAB6A)*, and *Gh_A05G0725 (CAB6A)* also Up-regulated in Leaf and root tissues under drought stress. The Ka/Ks value showed that the LHC of *G. hirsutum* genes resulting from *G. raimondii* and *G. arboreum* experienced negative selection instructions throughout the evolution. Thus, a detailed investigation of these genes will offer efficient information on understanding LHC genes in cotton (*Gossypium*) and its part in drought stress tolerance.

Abbreviations

ABA: Abscisic acid

GO: Gene ontology;

LHC: Light-Harvesting Chlorophyll a/b binding;

Ka: Non-synonymous substitution rate;

Ks: Synonymous substitution rate

CottonFGD: Cotton Functional Genomics Database

RT-qPCR: Real Time Qualitative Polymerase Chain Reaction

Declarations

Consent for publication

Not applicable

Funding

This research was funded by the National Natural Science Foundation of China, grant number 31621005, 31530053, 31671745.
Availability of data and materials

All the related data and files are all presented including the primers sequences used in the genes expression profiling.

Author Contributions

T.G.M., Y.X and R.O.M., conducted the experiment and wrote the manuscript. X.C., JNK, Y.H, Y.W., and S.Y. assisted in data collection. K.W., Z.Z., and F.L. revised the manuscript. All authors reread and agreed the last manuscript.

Ethics approval and consent to participate

No ethical nor consent to contribute in this research was sought, this not application in this research work.

Acknowledgments

We honestly appreciate the provision given to us by our lab throughout the time of this research.

Competing interests

The authors declared that they have no competing interests.

References

1. Agricultural U, Germplasm C, Station C, Udall JA, Long E, Hanson C, Yuan D, Ramaraj T, Conover JL, Gong L, Arick MA, Grover CE, Peterson DG, Wendel JF. De Novo Genome Sequence Assemblies of Gossypium raimondii and Gossypium turneri. 2019;9:3079–3085. https://doi: 10.1534/g3.119.400392.

2. Beddington JR, Asaduzzaman M, Clark ME, Fernández Bremauntz A, Guillou MD, Howlett DJB, Jahn MM, Lin E, Mamo T, Negra C, Nobre CA, Scholes RJ, Van Bo N, Wakhungu J. Agriculture: What next for agriculture after Durban? Science. 2012;335:289–90. https://doi. 10.1126/science.1217941.

3. Berry JO, Yerramsetty P. Photosynthetic gene expression in higher plants Photosynthetic gene expression in higher plants 2013; https://doi:10.1007/s11120-013-9880-8.

4. Chen C, Chen H, He Y, Xia R. TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. bioRxiv: 2018; 289660. https://doi: 10.1101/289660.

5. Dalal VK, Tripathy BC. Water-stress induced downsizing of light-harvesting antenna complex protects developing rice seedlings from photo-oxidative damage. Scientific Reports 8:10–16. https://doi:
10.1038/s41598-017-14419-4.

6. Eberhard S, Finazzi G, Wollman F-A. The Dynamics of Photosynthesis. Annu Rev Genet. 2018;42:463–515. https://doi.10.1146/annurev.genet.42.110807.091452.

7. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database. Nucleic Acids Research. 2019;47:D427–D432. https://doi:10.1093/nar/gky995.

8. Fanna K, Yang Z, Peipei SUN, Min CAO, Hong LI. Identification of Light-Harvesting Chlorophyll a / b-Binding Protein Genes of Zostera marina L. and Their Expression Under Different Environmental Conditions. 2016;15:152–162. https://doi.10.1007/s11802-016-2688-3.

9. Foyer CH, Neukermans J, Queval G, Noctor G, Harbinson J. Photosynthetic control of electron transport and the regulation of gene expression. Journal of Experimental Botany. 2012;63:1637–61. https://doi.10.1093/jxb/ers013.

10. Gene T, Consortium O. Gene Ontology: tool for the. Gene Expr. 2000;25:25–9. https://doi.10.1038/75556.

11. Gururani MA, Venkatesh J, Tran LSP. Regulation of photosynthesis during abiotic stress-induced photoinhibition. Molecular Plant. 2015;8:1304–20. https://doi.10.1016/j.molp.2015.05.005.

12. He M, He C, Ding N. Abiotic Stresses: General Defenses of Land Plants and Chances for Engineering Multistress Tolerance. 2018;9:1–18. https://doi:10.3389/fpls.2018.01771.

13. Horton P, Ruban A. Molecular design of the photosystem II light-harvesting antenna: photosynthesis and photoprotection. 2005;56:365–373. https://doi:10.1093/jxb/eri023.

14. Hou S, Zhu G, Li Y, Li W, Fu J, Niu E, Li L, Zhang D. Genome-Wide Association Studies Reveal Genetic Variation and Candidate Genes of Drought Stress Related Traits in Cotton (Gossypium hirsutum L.). 2018;9:1–15. https://doi:10.3389/fpls.2018.01276.

15. Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J, Baruch K, Fang D, Liu X, Ruan Y, ling, Rahman M ur, Han J, Wang K, Wang Q, Wu H, Mei G, Zang Y, Han Z, Xu C, Shen W, Yang D, Si Z, Dai F, Zou L, Huang F, Bai Y, Zhang Y, Brodt A, Ben-Hamo H, Zhu X, Zhou B, Guan X, Zhu S, Chen X, Zhang T. Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. Nature Genetics. 2019;51:739–748. https://doi:10.1038/s41588-019-0371-5.

16. Huang G, Wu Z, Percy RG, Bai M, Li Y, Frelichowski JE, Hu J, Wang K, Yu JZ, Zhu Y. A-genome evolution. Nat Genet 2020;52. https://doi:10.1038/s41588-020-0607-4.

17. Hussain HA, Hussain S, Khaliq A, Ashraf U, Anjum SA, Men S, Wang L. Chilling and Drought Stresses in Crop Plants: Implications, Cross Talk, and Potential Management Opportunities. Front Plant Sci. 2018;9:1–21. https://doi.10.3389/fpls.2018.00393.

18. Joshi R, Wani SH, Singh B, Bohra A, Dar ZA, Lone AA, Pareek A, Singla-pareek SL. Transcription Factors and Plants Response to Drought Stress: Current Understanding and Future Directions. 2016;7:1–15. 10.3389/fpls.2016.01029.
19. Kró M, Spangfort MD, Huner NPA, Oquist G. Chlorophyll a / b-Binding Proteins, Pigment Conversions, and Early Light-Induced Proteins in a Chlorophyll 6-less Barley Mutant. 1995;873–883.

20. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van De Peer Y, Rouzé P, Rombauts S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002;30:325–7. https://doi. 10.1093/nar/30.1.325.

21. Li XW, Zhu YL, Chen CY, Geng ZJ, Li XY, Ye TT, Mao XN, Du F. Cloning and characterization of two chlorophyll A/B binding protein genes and analysis of their gene family in Camellia sinensis. Scientific Reports. 2020;10:1–9. https://doi. 10.1038/s41598-020-61317-3.

22. Liu R, Xu YH, Jiang SC, Lu K, Lu YF, Feng XJ, Wu Z, Liang S, Yu YT, Wang XF, Zhang DP. Light-harvesting chlorophyll a/b-binding proteins, positively involved in abscisic acid signalling, require a transcription repressor, WRKY40, to balance their function. Journal of Experimental Botany. 2013b;64:5443–56. https://doi:10.1093/jxb/ert307.

23. Longoni P, Douchi D, Cariti F, Fucile G, Goldschmidt-Clermont M. Phosphorylation of the light-harvesting complex II isoform Lhcb2 is central to state transitions. Plant Physiol. 2015;169:2874–83. https://doi.10.1104/pp.15.01498.

24. Magwanga RO, Lu P, Kirungu JN. Knockdown of Cytochrome P450 Genes Gh _ D07G1197 and Gh _ A13G2057 on Chromosomes D07 and A13 Reveals Their Putative Role in Enhancing Drought and Salt Stress Tolerance in Gossypium hirsutum. 2019; https://doi: 10.3390/genes10030226.

25. Magwanga RO, Lu P, Kirungu JN, Lu H, Wang X, Cai X, Zhou Z, Zhang Z, Salih H, Wang K, Liu F. Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton. BMC Genet 2018;19. https://doi: 10.1186/s12863-017-0596-1.

26. Mao H, Li S, Wang Z, Cheng X, Li F, Mei F, Chen N, Kang Z. Regulatory changes in TaSNAC8-6A are associated with drought tolerance in wheat seedlings. Plant Biotechnol J. 2020;18:1078–92. https://doi. 10.1111/pbi.13277.

27. Nath K, Jajoo A, Poudyal RS, Timilsina R, Park YS, Aro EM, Nam HG, Lee CH. Towards a critical understanding of the photosystem II repair mechanism and its regulation during stress conditions. FEBS Lett. 2013;587:3372–81. https://doi. 10.1016/j.febslet.2013.09.015.

28. Nishiyama Y, Murata N. Revised scheme for the mechanism of photoinhibition and its application to enhance the abiotic stress tolerance of the photosynthetic machinery. Appl Microbiol Biotechnol. 2014;98:8777–96. https://doi:10.1007/s00253-014-6020-0.

29. Nouri MZ, Moumeni A, Komatsu S. Abiotic stresses: Insight into gene regulation and protein expression in photosynthetic pathways of plants. Int J Mol Sci. 2015;16:20392–416. https://doi. 10.3390/ijms160920392.

30. Pintó-marijuan M, Munné-bosch S. Photo-oxidative stress markers as a measure of abiotic stress-induced leaf senescence: advantages and limitations. 2014;65:3845–3857. https://doi:10.1093/jxb/eru086.

31. Qin D, Dong J, Xu F, Ge S, Xu Q, Li M. Genome-Wide Identification and Characterization of Light Harvesting Chlorophyll a/b Binding Protein Genes in Barley (Hordeum vulgare L.). Advances in Crop
Science Technology 2017;05. https://doi: 10.4172/2329-8863.1000301.
32. Sasi S, Venkatesh J, Daneshi RF, Gururani MA. Photosystem ii extrinsic proteins and their putative role in abiotic stress tolerance in higher plants. Plants. 2018;7:1–15. https://doi.10.3390/plants7040100.
33. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. Nat Protoc. 2008;3:1101–8. https://doi. 10.1038/nprot.2008.73.
34. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28:2731–9. https://doi. 10.1093/molbev/msr121.
35. Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C, Li Q, Yuan Y, Lu C, Wei H, Gou C, Zheng Z, Yin Y, Zhang X, Liu K, Wang B, Song C, Shi N, Kohel RJ, Percy RG, Yu JZ, Zhu Y, Wang J, Yu S. The draft genome of a diploid cotton Gossypium raimondii. 2012;44. https://doi: 10.1038/ng.2371.
36. Xing S, Kang L, Xu Q, Fan Y, Liu W, Zhu C, Yan J. The Coordination of Gene Expression within Photosynthesis Pathway for Acclimation of C 4 Energy Crop Miscanthus lutarioriparius. 2016;7. https://doi: 10.3389/fpls.2016.00109.
37. Xu YH, Liu R, Yan L, Liu ZQ, Jiang SC, Shen YY, Wang XF, Zhang DP. Light-harvesting chlorophyll a/b-binding proteins are required for stomatal response to abscisic acid in Arabidopsis. J Exp Bot. 2012;63:1095–106. https://doi. 10.1093/jxb/err315.
38. Xu Y, Magwanga RO, Cai X, Zhou Z, Wang X, Wang Y, Zhang Z, Jin D, Guo X, Wei Y, Li Z, Wang K, Liu F. Deep transcriptome analysis reveals reactive oxygen species (ROS) network evolution, response to abiotic stress, and regulation of fiber development in cotton. Int J Mol Sci 2019;20. https://doi: 10.3390/ijms20081863.
39. Yamaguchi-Shinozaki K, Shinozaki K. Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci. 2005;10:88–94. https://doi. 10.1016/j.tplants.2004.12.012.
40. Zhao Y, Kong H, Guo Y, Zou Z. Light-harvesting chlorophyll a / b-binding protein-coding genes in jatropha and the comparison with castor, cassava and arabidopsis. 2020a; https://doi: 10.7717/peerj.8465.
41. Zhao L, Lü Y, Chen W, Yao J, Li Y, Li Q, Pan J, Fang S, Sun J, Zhang Y. Genome-wide identification and analyses of the AHL gene family in cotton (Gossypium). BMC Genom. 2020b;21:1–14. https://doi. 10.1186/s12864-019-6406-6.
42. Zou Z, Li M, Jia R, Zhao H, He P, Zhang Y, Guo A. Genes encoding light-harvesting chlorophyll a / b-binding proteins in papaya (Carica papaya L.) and insight into lineage-specific evolution in Brassicaceae. Gene. 2020a;748:144685. https://doi:10.1016/j.gene.2020.144685.
43. Zou Z, Yang J. Genomics analysis of the light-harvesting chlorophyll a/b-binding (Lhc) superfamily in cassava (Manihot esculenta Crantz). Gene. 2019;702:171–81. https://doi:10.1016/j.gene.2019.03.071.
Figures

Figure 1

A. Phylogenetic tree association of 12 subfamilies of LHC genes in Gossypium hirsutum, Gossypium arboreum, Gossypium raimondii, Arabidopsis thaliana and Theobroma cacao. The tree was done using MEGA 7.0. B. Synteny blocks formation among cotton species chromosomes. A: Chromosomes of Gossypium arboreum; D: Chromosomes of Gossypium raimondii, At and Dt: chromosomes of A and D Subgenome of the tetraploid cotton, Gossypium hirsutum. TBtools was used to visualize the figure.
Figure 2

Gene Ontology annotation of LHC genes, showing the biological processes, cellular component and molecular function A, Gossypium hirsutum B, Gossypium arboreum C, Gossypium raimondii, AgriGO online tool analysis was used to do the graphic analysis.
Figure 3

Gene structure Display of Gossypium hirsutum, Gossypium arboreum and Gossypium raimondii, GSDS 2.0 online tool was used to construct the graph using CDS and Genomic DNA sequence of the genes.
Figure 4

Motif Identification of LHC Proteins A, Gossypium hirsutum, B, Gossypium arboreum and C, Gossypium raimondii, the motifs were detected by the MEME online analysis.

Figure 5

Distribution of LHC genes in three Gossypium Species A, Gossypium hirsutum At Subgenome B, Gossypium hirsutum Dt Subgenome C, Gossypium arboreum D, Gossypium raimondii, GFF3 file and gene ID was used to construct the chromosomal mapping via TBtools.
Figure 6

Cis-regulatory elements analysis of LHC genes in three Gossypium Species A, Gossypium hirsutum B, Gossypium arboreum C, Gossypium raimondii. Upstream sequence of gene lists were used to analyze the regulatory elements by Plant Care website.
Figure 7

Gene evolution forms of Ka, Ks and Ka/Ks values for homologous LHC gene pairs A, Gossypium hirsutum At - Gossypium arboreum B, Gossypium hirsutum Dt - G. raimondii C Gossypium hirsutum At - Gossypium hirsutum Dt D, Gossypium arboreum - Gossypium raimondii

Figure 8

Differential expression profiling of LHC genes under drought stress, RT-qPCR Analysis of LHC gene family in Gossypium hirsutum A, Leaf tissue B, Root tissue and C, Stem tissue, The Heatmap was done by MeV
software with Log2 transformation.

**Supplementary Files**

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

- TableS1.doc
- TableS2.xlsx
- TableS3.xlsx