

RESEARCH

Open Access



# Genome wide identification and characterization of light-harvesting Chloro a/b binding (LHC) genes reveals their potential role in enhancing drought tolerance in *Gossypium hirsutum*

MEHARI Teame Gereziher<sup>1,2†</sup>, XU Yanchao<sup>1†</sup>, MAGWANGA Richard Odongo<sup>1,3</sup>, UMER Muhammad Jawad<sup>1</sup>, KIRUNGU Joy Nyangasi<sup>1</sup>, CAI Xiaoyan<sup>1</sup>, HOU Yuqing<sup>1</sup>, WANG Yuhong<sup>1</sup>, YU Shuxun<sup>1</sup>, WANG Kunbo<sup>1</sup>, ZHOU Zhongli<sup>1\*</sup> and LIU Fang<sup>1,4\*</sup> 

## Abstract

**Background:** Cotton is an important commercial crop for being a valuable source of natural fiber. Its production has undergone a sharp decline because of abiotic stresses, etc. Drought is one of the major abiotic stress causing significant yield losses in cotton. However, plants have evolved self-defense mechanisms to cope abiotic factors like drought, salt, cold, etc. The evolution of stress responsive transcription factors such as the trihelix, a nodule-inception-like protein (*NLP*), and the late embryogenesis abundant proteins have shown positive response in the resistance improvement to several abiotic stresses.

**Results:** Genome wide identification and characterization of the effects of Light-Harvesting Chloro a/b binding (*LHC*) genes were 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 on various chromosomes. The *Ka/Ks* (Non-synonymous substitution rate/Synonymous substitution rate) values were less than one, an indication of negative selection of the gene family. Differential expressions of genes showed that majority of the genes are being highly upregulated in the roots as compared with leaves and stem tissues. Most genes were found to be highly expressed in MR-85, a relative drought tolerant germplasm.

**Conclusion:** The results provide proofs 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 varieties.

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

\* Correspondence: [zhouzl@cricaas.com.cn](mailto:zhouzl@cricaas.com.cn); [liufcrl@163.com](mailto:liufcrl@163.com)

†Teame Gereziher MEHARI and Yanchao XU contributed equally to this work.

<sup>1</sup>State Key Laboratory of Cotton Biology/Institute of Cotton Research, Chinese Academy of Agricultural Sciences (ICR, CAAS), Anyang 455000, Henan, China

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Background

Cotton is an important commercial crop because of its useful natural fiber source and the fact that it can be grown in a variety of climates around the world. Cotton and its by-products are in higher demand than ever before, due to the increased use of this fiber in the textile industry and the use of cottonseed as a source of edible oil (Hassan et al. 2020). It is an important multiuse crop, which is highly sensitive to both biotic and abiotic stresses (Zahid et al. 2016). Its production has undergone a sharp decline because of abiotic stress influences, of significance is drought.

Over the course of the twenty-first century, food production has to match the increasing population (Beddington et al. 2012). However, temperature increments and climate change have deepened the incidence and harshness of abiotic stresses that critically disturb the growth and development of crops (Nouri et al. 2015). Abiotic stress remains one of the key components of yield losses in plants (Sasi et al. 2018). Moreover, abiotic stress has a major impact on plant growth and development as compared with other forms of living organisms due to their immobility (He et al. 2018; Magwanga et al. 2018; 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 and Murata 2014). Furthermore, in the chloroplasts, drought, high light, salinity, or extreme temperature stresses do trigger a diminishing in CO<sub>2</sub> assimilation rates, which in turn induce an upsurge in reactive oxygen species creation, which eventually leads to yield loss (Pintó-Marijuan and Munné-Bosch 2014). It has been reported that abiotic stresses account for over 50% yield loss in crops (Nath et al. 2013). Moreover, a decrease in photosynthesis rate results in a remarkable reduction of yield in crops (Nouri et al. 2015).

Drought exposure alters the photosynthetic apparatus in the plants, 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 are, Ribulose biphosphate carboxylase large chain (RBCL) (Berry et al. 2013), light-harvesting chlorophyll a/b-binding (LHC) (Zhao et al. 2020), and Cytochrome P450 genes (Magwanga et al. 2019). Using light-harvesting chlorophyll a/b-binding (LHC) proteins, photosynthesis requires the accumulation of light and the conversion of solar energy. In higher plants, the LHC gene family includes LHCA and LHCB sub-families, which encode proteins constituting the light-harvesting complex of photosystems I and II (Kong et al. 2016). The LHC proteins are the apoproteins of the Light-Harvesting complex of photosystem II (PSII), outer

antenna complex which are perhaps the most abundant membrane proteins in nature (Horton and Ruban 2005; Król et al. 1995; 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 against 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. 2013). In the recognition of the proteins encoded by the LHCB genes, 28 have been identified in *Papaya carica* (Zou et al. 2020), 17 in *Hordeum vulgare* L. (Qin et al. 2017), 25 in *Camellia sinensis* (Li et al. 2020), and 35 genes in *Manihot esculenta* (Zou and 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), *G. arboreum* (Huang et al. 2020), and *G. raimondii* (Wang et al. 2012) 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

Marie-galante 85 (MR-85), a race developed from *G. hirsutum* species and comparatively tolerant to abiotic stress, was used as an experimental material (Xu et al. 2020). The seeds were soaked in water overnight and put on the absorbent paper for germination. After 7 days of germination seedlings were transferred to a hydroponic setup composed of Hoagland nutrient solution, placed in the greenhouse, with 16 h/8 h light-dark and at 28 °C/25 °C day-night temperature (Zhao et al. 2020). At the third leaf stage, drought stress was applied by supplementing the nutrient solution with 17% of PEG-6000 (Liu et al. 2013). 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 post stress exposure. The experiment was conducted in a greenhouse located in Anyang, Institute of Cotton Research, CAAS, with a complete randomized design (CRD) with three biological replications.

### Identification of the LHC proteins in cotton species

The domain number PF00504 was used to search 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](http://www.cottonfgd.org)), while those for *Arabidopsis thaliana*, and *Theobroma cacao* were downloaded from phytozome (<https://phytozome.jgi.doe.gov>). The protein sequences of the LHC downloaded from the CottonFGD were uploaded to the Pfam database (<http://pfam.xfam.org>) for the

identification of the putative *LHC* proteins with the best domain e-value cutoffs  $f < 1 \times 10^{-4}$  (El-Gebali et al. 2019). Moreover, the CottonFGD website ([www.cottonfgd.org](http://www.cottonfgd.org)) was explored to obtain the physicochemical characteristics of the gene family, such as protein length (PL), molecular weight, molecular charge, isoelectric point (pI), and GRAVY value.

#### Phylogenetic tree and collinearity analysis

Protein sequences of *G. hirsutum*, *G. arboreum*, *G. raimondii*, *Arabidopsis thaliana* and *Theobroma cacao* were aligned by ClustalX (Larkin et al. 2007) 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 1 000 bootstrap replications (Tamura et al. 2011). To categorize the homologous genes of cotton species, the protein sequences of *G. hirsutum* were BlastP searched alongside the protein database of *G. arboreum* and *G. raimondii*; hits with E-values  $\leq 1 \times 10^{-5}$  and similarity  $\geq 90\%$  were considered significant. The GFF3 file, linked file, and Gene IDs were applied to construct the collinearity analysis by TBtools software (Chen et al. 2018). Homologous genes between *G. hirsutum*, *G. raimondii* and *G. arboreum* were sorted out from CottonFGD employing protein blast with a threshold of  $\geq 80\%$  match and at least an 80% alignment ratio based on the protein length.

#### Chromosomal mapping, gene ontology, and cis-regulatory elements analysis

To know the distribution of Light-Harvesting Chloro a/b-bind genes on all the chromosomes of A, D, and AD cotton genomes, we used the GFF3 file from CottonFGD ([www.cottonfgd.org](http://www.cottonfgd.org)) and gene ID of the genes. Then we used the TBtools software to show the genes distribution on chromosomes. 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 ([www.bioinfo.cau.edu.cn/agriGO](http://www.bioinfo.cau.edu.cn/agriGO)) online analysis (Ashburner et al. 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, an online tool (<http://gsds.cbi.pku.edu.cn/>). The motifs were identified by using an online tool MEME (<http://meme-suite.org>). The 2000-bp upstream sequences of *LHC* genes of cotton species were downloaded from CottonFGD (<http://www.cottonfgd.org>) to identify the cis-regulatory elements in the putative promoter regions. 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). TBtools was used to visualize the structure.

#### Gene evolution and subcellular localization prediction of LHC proteins

The coding sequences (CDS) and protein sequences of the homolog genes were downloaded from the CottonFGD website ([www.cottonfgd.org](http://www.cottonfgd.org)). CDS, protein sequences and gene IDs of the homolog genes were used to compute the *Ka/Ks* (Non-synonymous substitution rate/Synonymous substitution rate) value by TBtools software (Chen et al. 2018). For the estimation of protein subcellular localization of the *LHC* gene family, protein sequences of the three *Gossypium* species were downloaded from the CottonFGD (<http://www.cottonfgd.org>). The subcellular localization for the *LHC* proteins of *Gossypium* species was predicted using an online tool Wolf PSORT (<https://wolfsort.hgc.jp>).

#### RNA extraction and RT-qPCR analysis

Total RNA was extracted using TIANGEN, RNA-prep Pure Plant Plus Kit ([www.tiangen.com](http://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 (TransGen Biotech Co., Ltd., Beijing, China, [www.transgen.com.cn](http://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 (<https://www.ncbi.nlm.nih.gov>). For the RT-qPCR analysis, 7500 Fast Real-Time System with 2  $\mu$ L cDNA, 2  $\mu$ L forward and reverse primers, 6  $\mu$ L RNA-free water, and 10  $\mu$ L SYBR solution were used. Three biological and technical replications were used with *Ghactin7* as an internal control.  $E = 2^{-\Delta\Delta C_T}$  formula was used to calculate the gene expression levels (Schmittgen and Livak 2008).

## Results

#### Identification of the cotton *LHC* proteins

One hundred and nine 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 total number of the proteins found related to the *LHC* genes in the two diploid cotton species, *G. raimondii* and *G. arboreum*, were one less than 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 protein lengths for the *G. hirsutum* proteins stretched from 62 aa to 644 aa, molecular weights ranged from 6.88 kDa to 72.66 kDa, respectively, in *Gh\_*

*Sca017783G01* and *Gh\_A02G1068*, and the molecular 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 S1).

In the two diploid cotton species, the *G. arboreum* and *G. raimondii*, the physicochemical properties of the LHC proteins exhibited slight differences in molecular weights, protein lengths, pI, molecular charge, and GRAVY values. The protein length stretched from 114 aa, and 151 aa to 349 aa, respectively, and the 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 S1).

On the other hand, the values for the pI and GRAVY were almost the same, the pI ranges from  $4.87$  to  $9.897$ , and  $4.701$  to  $9.296$ , the GRAVY ranges from  $-0.377$  to  $0.167$  and  $-0.249$  to  $0.244$  in order of *G. arboreum* and *G. raimondii*, respectively. In all cotton species, the GRAVY values were low (positive and negative), which indicates the likelihood of enhanced relations with water that leads to hydrophilic nature.

**Phylogenetic tree and Synteny block analysis of the cotton LHC proteins**

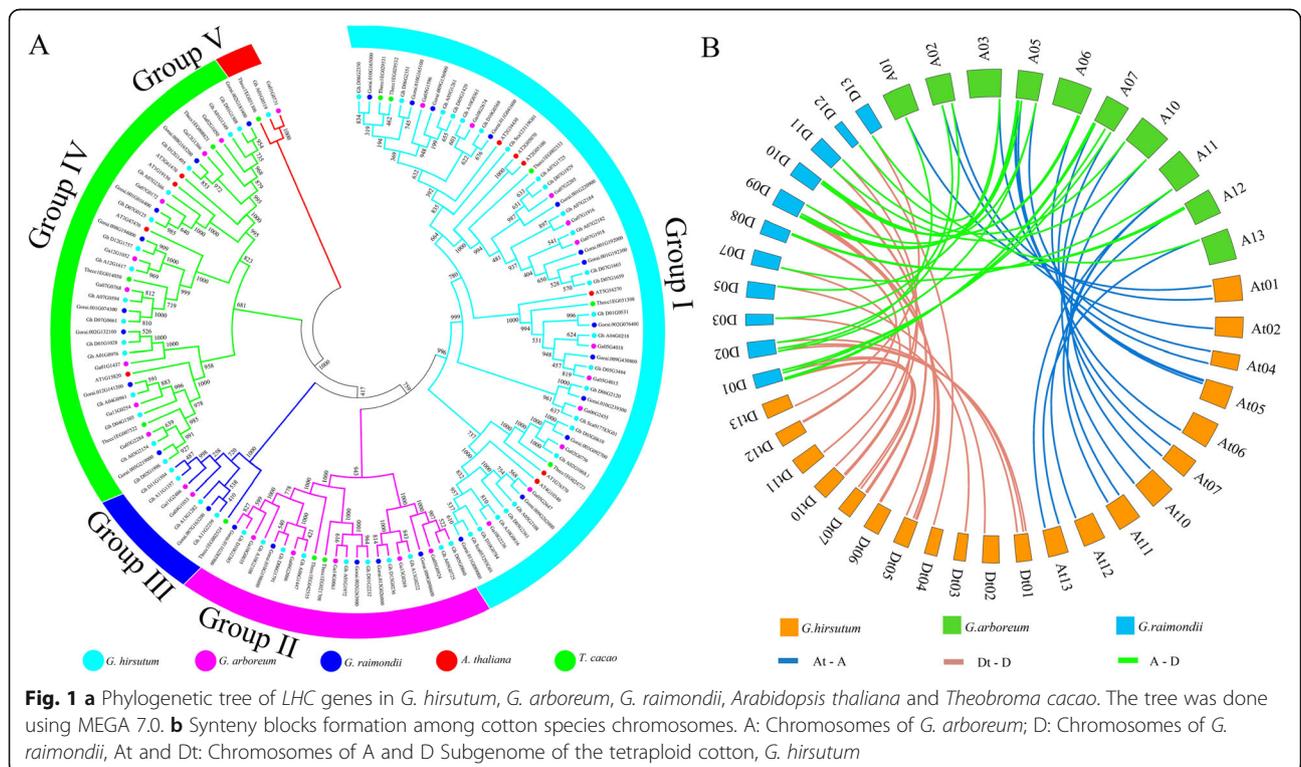
The phylogenetic analysis grouped the LHC proteins together with other plants into five clades. Numerous

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 done, in which Circle gene viewer was applied to distinguish the collinear gene pairs with TBtools software (Chen et al. 2018). The collinearity analysis between the physical map of At and Dt subgenomes of *G. hirsutum*, *G. arboreum* and *G. raimondii* for their A Vs D; A vs At, and between D Vs Dt subgenome relationships were observed. We found a good collinearity between A vs D with 23 genes, A vs At with 20 genes, and between D vs Dt with 23 genes in the subgenome (Fig. 1b).

**Gene ontology analysis**

According to the gene ontology analysis, in *G. hirsutum*, the functions in biological processes (GO: 0008150), were cellular and metabolic processes, various cellular (GO: 0005575) functions were noted in the cell and cell part. Similarly, in *G. arboreum*, the biological functions (GO: 0008150) were responsible for stimuli, cellular and metabolic processes. Whereas in cellular component (GO: 00055750), the functions focused on cell, macromolecular complex (protein), and membrane related issues, whereas molecular functions (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 was no significant GO term for molecular functions (Fig. 2).

#### Gene structure and motif identification of LHC proteins

Gene structural study is observed as a likely sign of the evolution of multigene families (Nei and Rooney 2005). To obtain additional evidence into the structural diversity of cotton Light-Harvesting Chloro a/b-bind genes, the exon/intron association in the representative transcripts were 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 structures of some Light-Harvesting Chloro a/b-bind genes possessed introns. The maximum number of introns observed for the Light-Harvesting Chloro a/b-bind gene structures were eleven (*Gh\_A02G1068*), eleven (*Ga02G0756*), and five (*Gorai.003G092700*) for *G. hirsutum*, *G. arboreum* and *G. raimondii*, respectively. The highest number of exons and introns were found in *Gh\_A02G1068* (twelve exons, eleven introns) and *Gh\_A01G0519* (ten exons, nine 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, eighteen genes have two exons and one intron and seven genes have three exons and two introns, whereas, seven genes with one exon but no intron (Fig. 3).

On the other hand, in the diploid species, the maximum number of exon/intron were twelve exons, eleven introns (*Ga02G0756*), eleven exons, and ten introns (*Ga01G0731*) in *G. arboreum*, whereas *G. raimondii* six exons, five introns (*Gorai.003G092700*) and six exons, five introns (*Gorai.009G262000*), respectively. Similarly, there are seven *G. arboreum* genes that have two exons and one intron while ten genes in *G. raimondii* have two exons and one intron. 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 were the conserved motifs in *G. hirsutum*, whereas motif 2 and 8 in *G. arboreum* and motif 11 and 4 in *G. raimondii*, respectively, were conserved, too.

#### Chromosomal mapping of the LHC genes

The LHC genes were unevenly distributed across various chromosomes of A2, D5, and (AD)1 cotton genomes. In

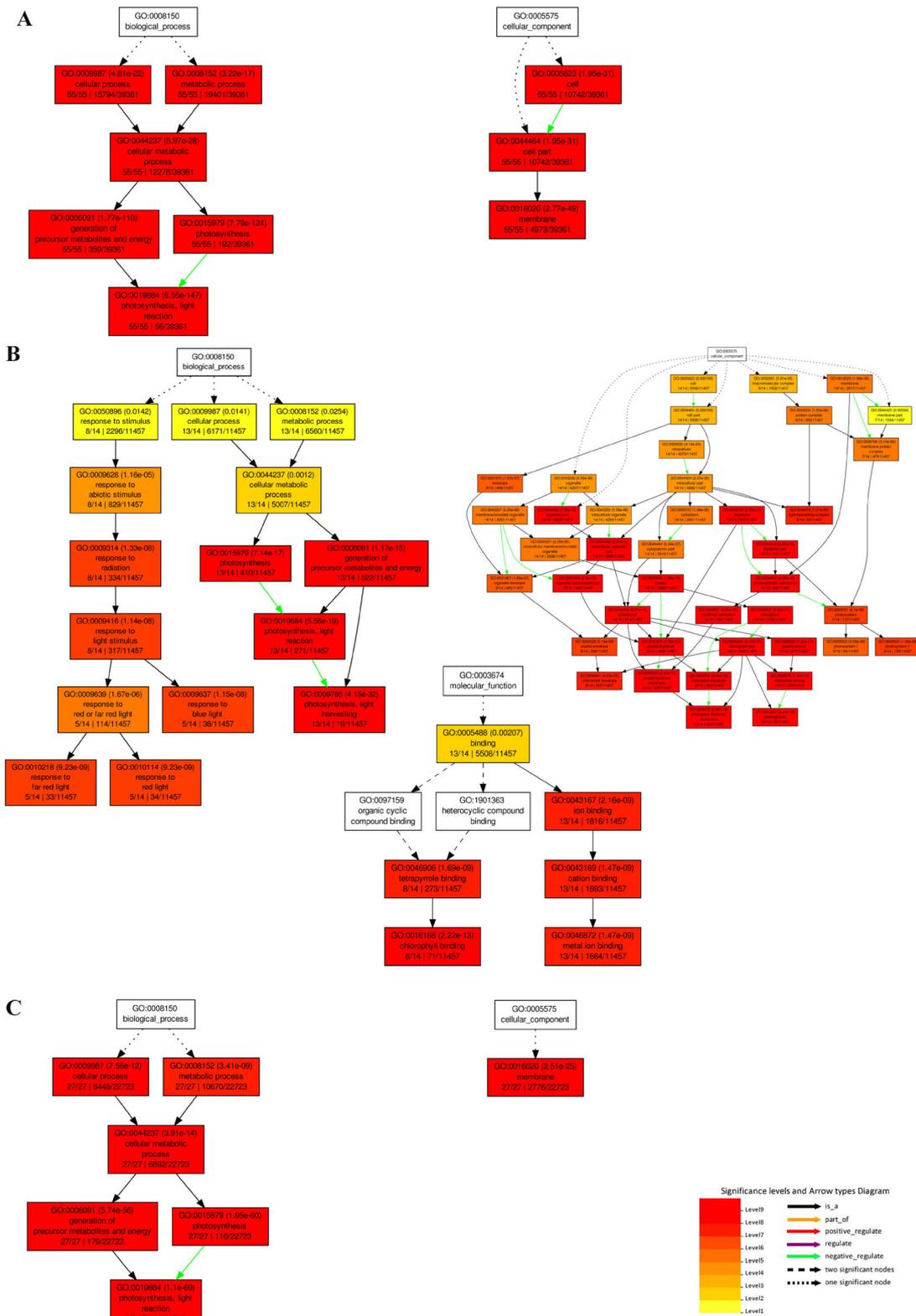
the tetraploid (AD)1 genome with At Subgenome, the highest number of genes were found on chromosome At01, At05, and At10 with three genes, while At03, At08, and At09 chromosomes harbored none. Similarly, in the (AD)1, Dt Subgenome, the highest number of genes were found on Dt07, Dt01, and Dt05 with five, four, and four genes, respectively, whereas At03, At08, and At09 had zero genes. The rest of the chromosomes harbored one to three genes (Fig. 5a and b). In the two diploid cotton species, A2 and D5 genomes, the gene distribution arrangement was different. In *G. arboreum*, the highest genes were observed on chromosomes, A2(05), and A2(07), with the four genes while in *G. raimondii*, chromosomes D5(01), D5(09), and D5(10) possessed the highest gene number with four genes, respectively, while chromosomes A2(04) and D5(06) harbored none. Some LHC genes have tandem duplications, although most are singletons dispersed along the genome (Fig. 5c and d).

#### Identification of cis-regulatory elements

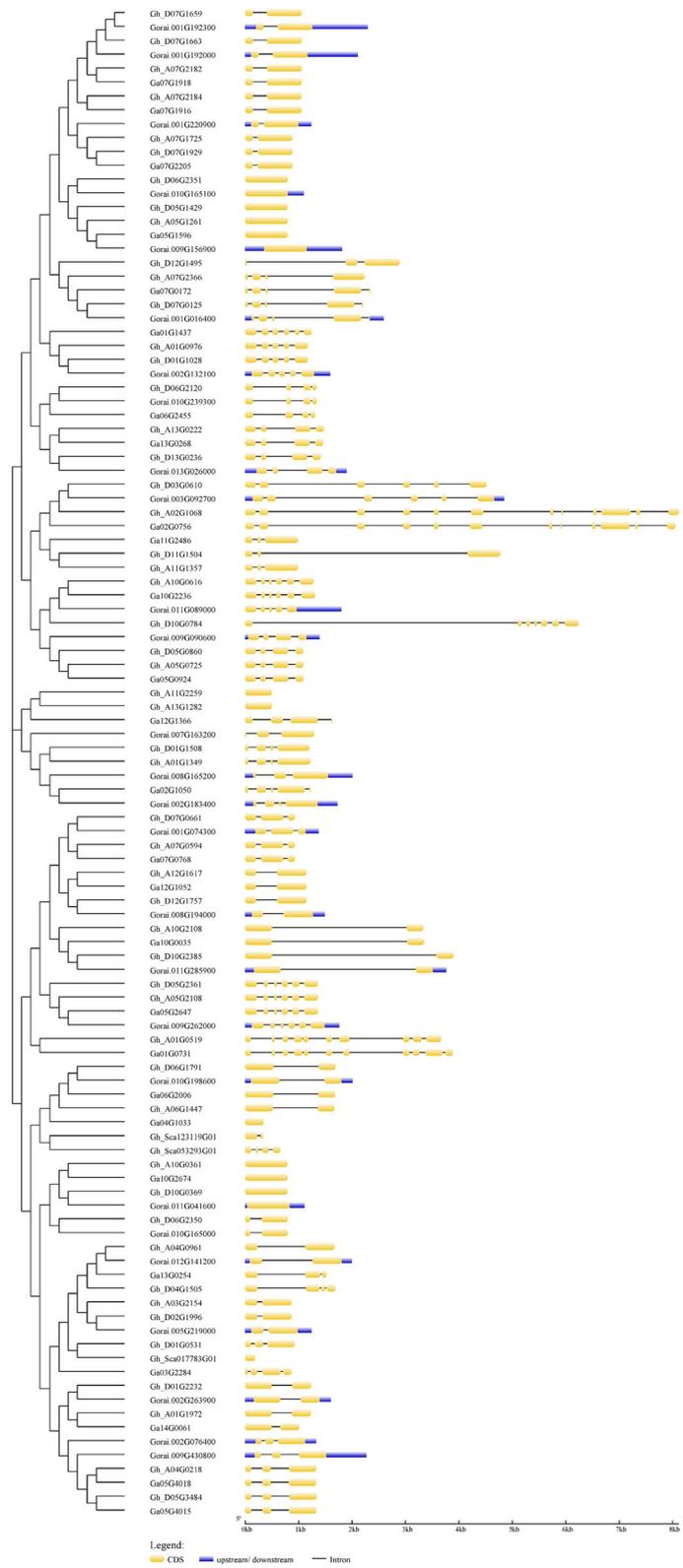
Cis-acting regulatory elements are important molecular switches involved in the transcriptional regulation of a dynamic network of gene activities controlling various biological processes, including abiotic stress responses, hormone responses, and developmental processes. These genes encode genomic blueprints for coordinating spatio-temporal gene expression programs underlying highly specialized cell functions (Mao et al. 2020). Analysis of cis-regulatory elements revealed that ABA-responsive element (ABRE), Antioxidant response elements (ARE), Metal response elements (MRE), Myeloblastosis (MYB), AT-rich elements (ATREs), Dehydration-responsive element (DRE), MBS, Box-4, and Angiotensin-Converting Enzyme (ACE) cis-regulatory elements were found related to drought stress in the three cotton species (Fig. 6). The ABA-responsive element (ABRE) and the dehydration-responsive element/C-repeat (DRE/CRT) are two major cis-acting elements involved in ABA-dependent and ABA-independent gene expression in osmotic and cold stress responses (Yamaguchi-Shinozaki and Shinozaki 2005).

#### Evolution of LHC genes in *Gossypium* species

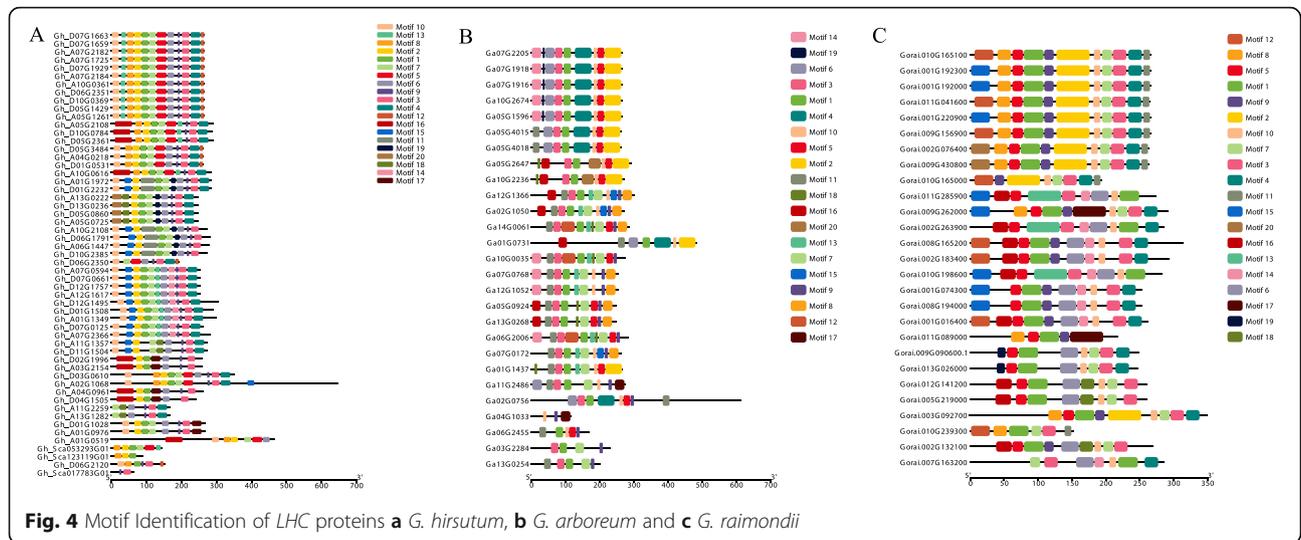
During the evolution, the  $K_s$  value of a gene is not affected by natural selection generally, but  $K_a$  value is affected. The  $K_a/K_s$  value show positive, neutral, and negative selection when,  $K_a/K_s > 1$ ,  $K_a/K_s = 1$ , or  $K_a/K_s < 1$ , respectively (Zhao et al. 2020). The distributions of  $K_a$ ,  $K_s$ , and  $K_a/K_s$  among homologous pairs of *Gossypium* species revealed similar results (Fig. 7, Table S3). The  $K_a/K_s$  for GhAt-Ga ranged from 0 to 0.949034416, while for GhDt-Gr  $K_a/K_s$  ranged from 0 to 0.838286204. The  $K_a/K_s$  of GhAt-GhDt ranged from 0 to 0.523637063, whereas the  $K_a/K_s$  value of Ga-Gr ranged from 0 to 0.755930549. In



**Fig. 2** Gene Ontology annotation analysis of LHC genes, showing their role in biological processes, cellular component and molecular function **a** *G. hirsutum* **b** *G. arboreum* **c** *G. raimondii*



**Fig. 3** Gene structure display using Gene Structure Display Server - GSDS 2.0 online tool for *G. hirsutum*, *G. arboreum* and *G. raimondii*



all the pairs, the *Ka/Ks* value was < 1 which indicated that the gene family was subjected to negative selection. The results suggested that the *LHC* genes of *G. hirsutum* derived from *G. raimondii* and *G. arboreum* experienced negative selection throughout the process of evolution.

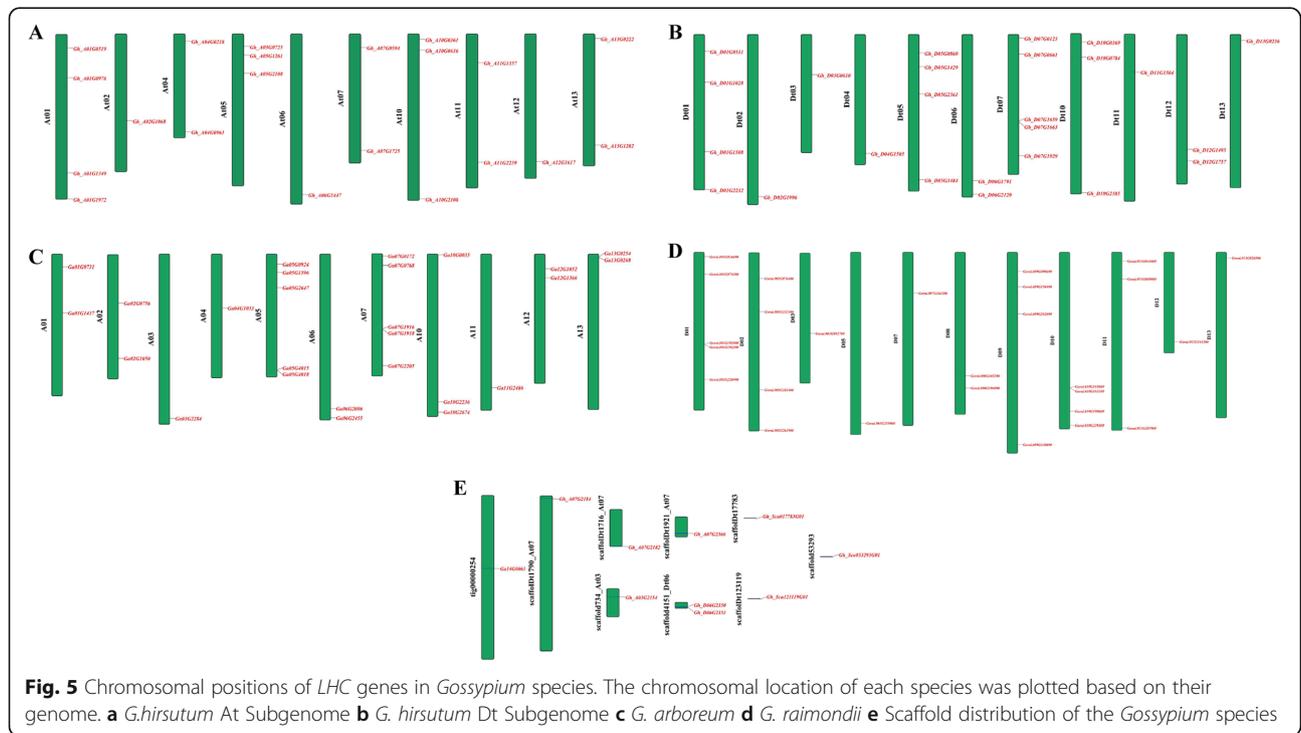
**Prediction of subcellular localization for *LHC* proteins in *Gossypium* species**

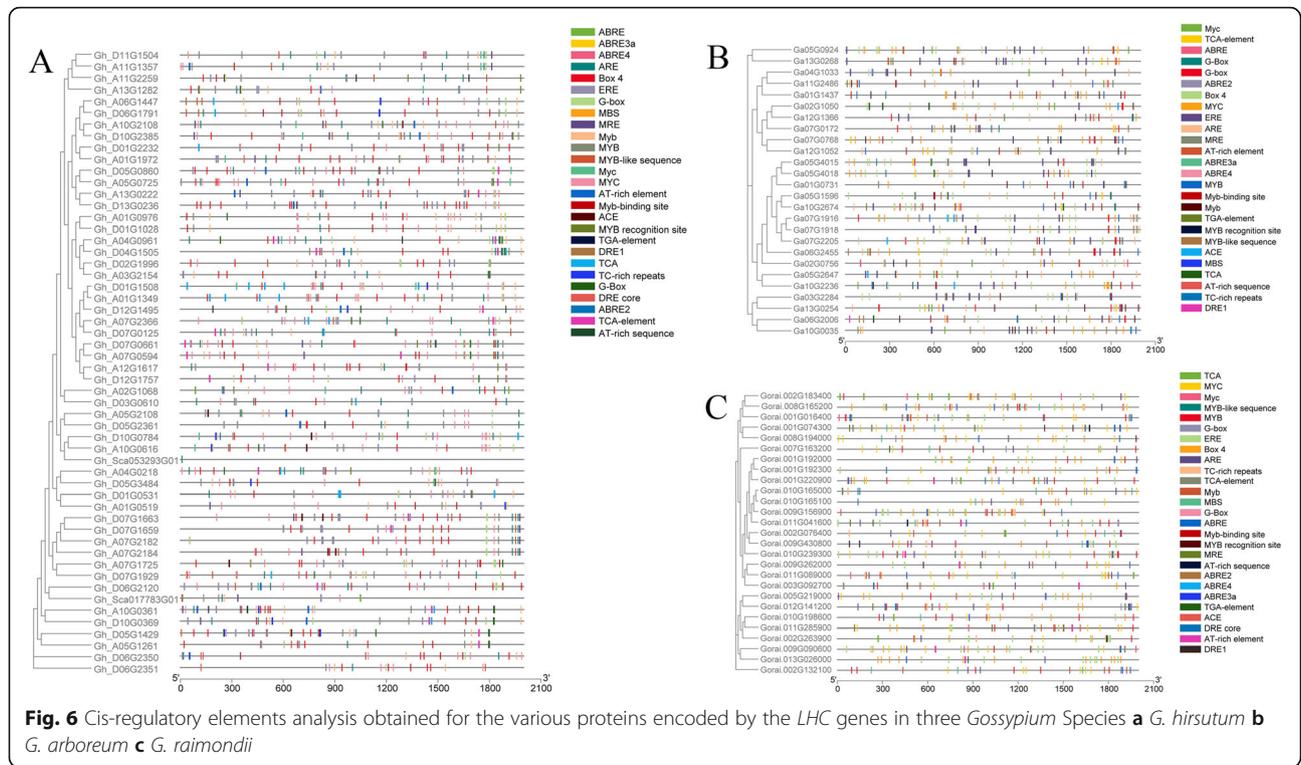
The results from the WOLF PSORT (<https://wolfsort.hgc.jp>) showed that the *LHC* proteins were localized in various cell parts including chloroplast, cytoplasm, endoplasmic reticulum, mitochondria, nucleus and vacuole

(Fig. 8). Based on the online analysis of the three *Gossypium* species, the *LHC* proteins were mainly localized in the chloroplast with 472.5 (72.1%), 251.5 (73.1%), 232 (73.3%) in *G. hirsutum*, *G. arboreum* and *G. raimondii*, respectively (Table S5).

**RT-qPCR validation of *LHC* genes under water deficit conditions**

Twenty-seven *LHC* genes expression profiles were carried out in different tissues and varying time intervals under PEG-6000 treatment. All genes showed differential expression patterns in the analyzed tissues (Table S6). *Gh*\_





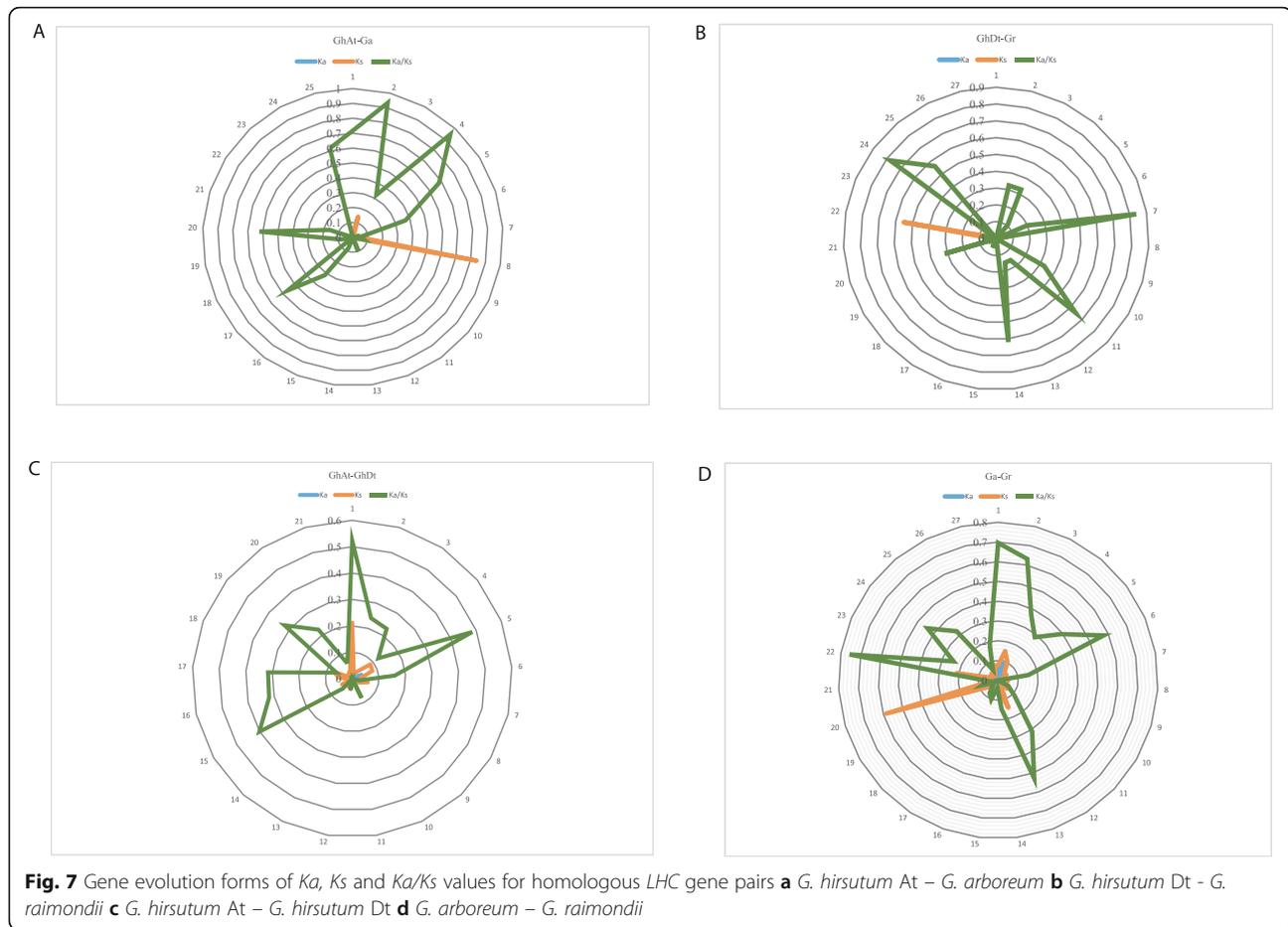
*D10G2385*, *Gh\_A13G0222*, *Gh\_A05G0725*, *Gh\_D05G0860*, *Gh\_D07G0661*, *Gh\_D01G1508*, *Gh\_D12G1495*, *Gh\_A07G2182*, and *Gh\_A10G2108* were found to be highly up-regulated in the roots, whereas, *Gh\_A07G2184*, *Gh\_D10G2385*, *Gh\_D05G0860*, *Gh\_D02G1996*, *Gh\_A13G0222*, and *Gh\_A05G0725* showed higher upregulation after 12 h of stress exposure in leaves. Similarly, *Gh\_A13G0222*, *Gh\_D06G1791*, and *Gh\_A06G1447* genes were upregulated in stem tissues starting from 6 h up to 24 h (Fig. 9).

Changes in genes expression at different time intervals and in different plant tissues was observed in the results. Most genes were downregulated mainly in leaf tissue followed by stem. Genes like *Gh\_A10G0361*, *Gh\_D10G0369*, *Gh\_A03G2154*, and *Gh\_D03G0610* were downregulated in the three tissues of cotton at almost all time points. Generally, more genes were upregulated in the root tissues. Expression of gene *Gh\_A13G0222* was significantly higher in root and stem tissues at 12 h and 24 h after treatment as compared with other time points, whereas in root tissues high expression was observed in almost all time points except at 12 h. *Gh\_D10G2385*, *Gh\_D05G0860*, and *Gh\_A05G0725* was also upregulated in leaf and root tissues under drought stress. A detailed exploration of these genes will offer efficient information on considerable *LHC* genes in cotton (*Gossypium*) and its part in drought stress tolerance. Drought effect comes first 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 is the main risk for 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, which typically start by affecting stomata at medium drought severity, causing metabolic and structural changes during harsh drought conditions. Photosynthesis is one of the greatest vital photochemical reactions that occurs 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 et al. 2015). It is the outcome of many steps and multipart developments that employs numerous biological pathways similar to photosynthetic electron transport system (PETs) makes sunlight to transform into ATP and NADPH; in addition, by the Calvin-Benson cycle, CO<sub>2</sub> is fixed into carbohydrates, as well as the assimilation, transport, and intake of photo assimilates as the organic products of photosynthesis (Eberhard et al. 2008; Foyer et al. 2012).

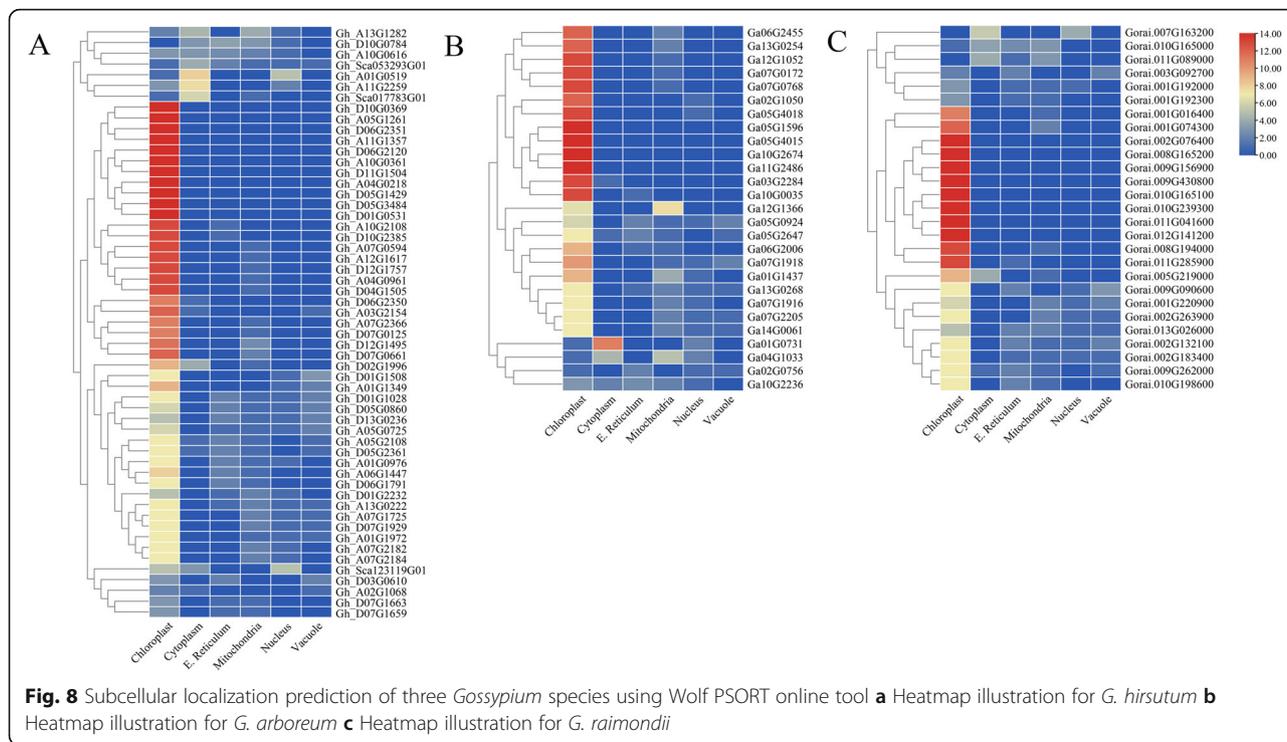


Forming disorder of all photosynthesis mechanisms has the primary impact of abiotic stress on the activity of photosynthesis (Nouri et al. 2015). Photosynthetic reactions of mature crops and small seedlings to drought-stress are mainly diverse. In mature crops, efficient photosynthetic complexes are previously shaped and water-stress brings the creation of reactive oxygen species due to surplus light absorption, which pressures the photosynthetic apparatus. Though, in young seedlings under water stress, there is the likelihood to down-regulate chlorophyll 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 and Tripathy 2018). Chloroplast is the main research area in the field of biology because it is 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 (Li et al. 2020; Liu et al. 2013).

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. 2020). However, not enough information is available 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 the fundamental parts in light harvest and acclimation to changing light (Longoni et al. 2015; Qin et al. 2017).

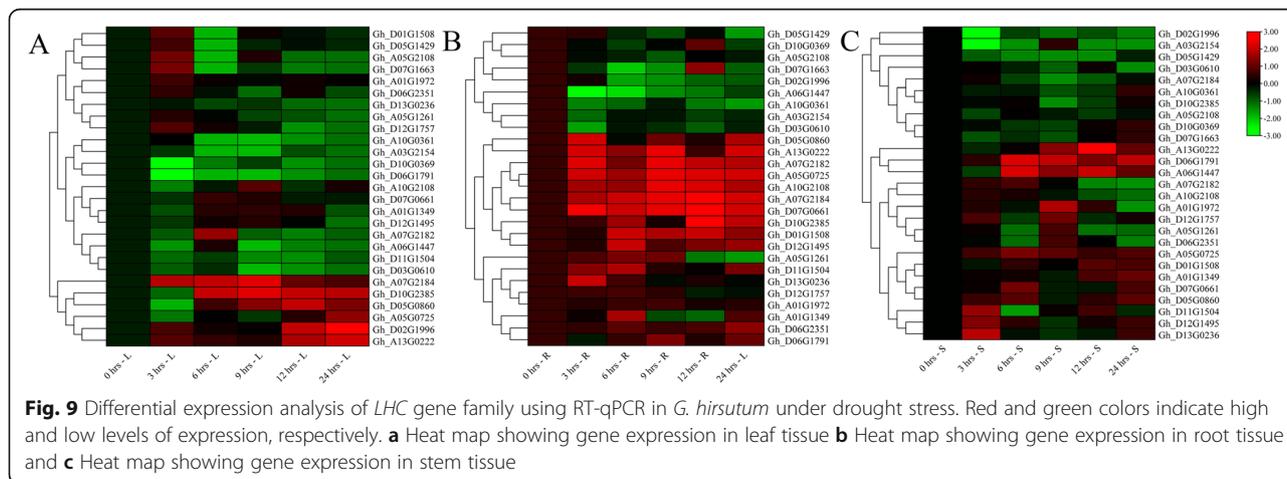
In our results, many genes were upregulated in the root tissue. *Gh\_A13G0222* was upregulated in all tissue samples while *Gh\_D10G2385*, *Gh\_D05G0860*, and *Gh\_A05G0725* were upregulated in leaf and root tissues under drought stress. A study from tea plants showed that two genes, *CsCPI* and *CsCP2*, were found to affect phosphorylation/ dephosphorylation and GTP in the physiological regulation of PSII. The regulation of LHC protein stages allows chloroplasts to answer amenably and quickly to abiotic stresses (Li et al. 2020). It was observed in papaya plants after treatment with mannitol for drought stress for 10 days, three genes were found to be upregulated (*CpELIP*, *CpLhcb7*, and *CpPsbS*). After 15 days, five genes were found to be upregulated



(*CpELIP*, *CpSEP2*, *CpOHP2*, *CpLhcb7*, and *CpPsbS*), and after 20 days, five genes were found to be highly upregulated (*CpELIP*, *CpSEP2*, *CpOHP2*, *CpLhcb7*, and *CpPsbS*) (Zou et al. 2020). The evolution of *LHC* genes in *Gossypium* species indicated that the distributions of *Ka*, *Ks*, and *Ka/Ks* were similar among homologous pairs. The *Ka/Ks* of GhAt-Ga ranged from 0 to 0.949034416, while GhDt-Gr ranged from 0 to 0.838286204. The *Ka/Ks* of GhAt-GhDt ranged from 0 to 0.523637063, whereas the *Ka/Ks* value of Ga-Gr was 0–0.755930549. The result suggested that the *LHC* of *G. hirsutum* genes derived from *G. raimondii* and *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 and Yang 2019).

*LHCB* family members positively regulates crops abiotic stress tolerance by stomatal closure to *ABA* signaling starting from germination to final growth (Liu et al. 2013; Xu et al. 2012). 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* so, *LHCB* members have been found as new actors in *ABA* signalling in stomatal movement (Xu et al.



2012). *WRKY* transcription factor helps to boost the expression of *LHCB* by suppressing the *WRKY* repressors under stress conditions (Liu et al. 2013). Functional genomics trials will be beneficial for the validation of *LHCB* genes function both at molecular and genetics levels thus, making *LHC* family useful for cotton improvement.

## Conclusions

The aim of this study is to investigate *G. hirsutum* *LHC* genes and their potential to drought stress tolerance. On the basis of family analysis, 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 on the upper and lower arm of chromosomes. RT-qPCR analysis reveals the upregulation of more genes in roots followed by stem and leaf tissues. *Gh A13G0222*, *Gh D05G0860*, and *Gh D10G2385* were found to be candidate genes linked to drought stress tolerance in cotton after being upregulated in post treatments examined in the current study. Therefore, we recommend a detailed investigation of candidate genes at both molecular and genetic levels to elucidate the underlying mechanisms in response to drought stress tolerance in cotton.

## 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

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42397-021-00090-8>.

**Additional file 1: Table S1.** Physiochemical properties of proteins in *G. hirsutum*, *G. arboreum* and *G. raimondii* Species.

**Additional file 2: Table S2.** List of primer details of RT-qPCR for *LHC* genes; Primers was designed by NCBI primer blast.

**Additional file 3: Table S3.** List of *LHC* genes in *G. hirsutum*, *G. arboreum*, and *G. raimondii*, respectively.

**Additional file 4: Table S4.** *Ka*, *Ks*, *Ka/Ks* values of *LHC* genes.

**Additional file 5: Table S5.** Subcellular localization prediction of *LHC* proteins in the three *Gossypium* species.

**Additional file 6: Table S6.** T-test for the RT-qPCR analysis of *LHC* genes using SAS software.

## Acknowledgments

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

## Authors' contributions

Mehari TG, Xu YC, Magwanga RO, and Umer MJ conducted the experiment and wrote the manuscript. Cai XY, Kirungu JN, Hou YQ, Wang YH, and Yu SX assisted in data collection. Wang K, Zhou ZL and Liu F revised the manuscript. All authors reread and agreed on the final manuscript.

## Funding

This research was funded by the National Natural Science Foundation of China, grant number 31621005, 31530053, 31671745, and The National Key R&D Program of China ((2021YFE0101200), PSF/CRP/18thProtocol (07).

## Availability of data and materials

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

## Declarations

### Ethics approval and consent to participate

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

### Consent for publication

Not applicable.

### Competing interests

The authors declared that they have no competing interests.

### Author details

<sup>1</sup>State Key Laboratory of Cotton Biology/Institute of Cotton Research, Chinese Academy of Agricultural Sciences (ICR, CAAS), Anyang 455000, Henan, China. <sup>2</sup>Ethiopian Institute of Agricultural Research, Mekhoni Agricultural Research Center, P.O. Box 47, Mekhoni, Tigray, Ethiopia. <sup>3</sup>School of Biological and Physical Sciences (SBPS), Main Campus, Jaramogi Oginga Odinga University of Science and Technology (JOUUST), Main Campus, P.O. Box 210-40601, Bondo, Kenya. <sup>4</sup>School of Agricultural Sciences, Zhengzhou University, Zhengzhou, Henan 450001, China.

Received: 23 October 2020 Accepted: 19 April 2021

Published online: 03 June 2021

## References

- Ashburner M, Ball C, Blake J, et al. Gene ontology: tool for the unification of biology. *Nat Genet.* 2000;25(1):25–9. <https://doi.org/10.1038/75556>.
- Beddington JR, Asaduzzaman M, Clark ME, et al. Agriculture: What next for agriculture after Durban? *Science.* 2012;335(6066):289–90. <https://doi.org/10.1126/science.1217941>.
- Berry JO, Yerramsetty P, Zielinski AM, Mure CM. Photosynthetic gene expression in higher plants. *Photosynth Res.* 2013;117(1):91–120. <https://doi.org/10.1007/s11120-013-9880-8>.
- Chen C, Chen H, He Y, et al. TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. *bioRxiv.* 2018. <https://doi.org/10.1101/289660>.
- Dalal VK, Tripathy BC. Water-stress induced downsizing of light-harvesting antenna complex protects developing rice seedlings from photo-oxidative damage. *Sci Rep.* 2018;8(1):10–6. <https://doi.org/10.1038/s41598-017-14419-4>.
- Eberhard S, Finazzi G, Wollman FA. The dynamics of photosynthesis. *Annu Rev Genet.* 2008;42(1):463–515. <https://doi.org/10.1146/annurev.genet.42.110807.091452>.
- El-Gebali S, Mistry J, Bateman A, et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* 2019;47(D1):D427–32. <https://doi.org/10.1093/nar/gky995>.
- Foyer CH, Neukermans J, Queval G, et al. Photosynthetic control of electron transport and the regulation of gene expression. *J Exp Bot.* 2012;63(4):1637–61. <https://doi.org/10.1093/jxb/ers013>.
- Gururani MA, Venkatesh J, Tran LSP. Regulation of photosynthesis during abiotic stress-induced photoinhibition. *Mol Plant.* 2015;8(9):1304–20. <https://doi.org/10.1016/j.molp.2015.05.005>.
- Hassan A, Ijaz M, Sattar A, et al. Abiotic stress tolerance in cotton. *Adv Cotton Res. IntechOpen.* 2020. <https://doi.org/10.5772/intechopen.89622>. Accessed 13 Aug 2020.

- He M, He CQ, Ding NZ, et al. Abiotic stresses: general defenses of land plants and chances for engineering multistress tolerance. *Front Plant Sci.* 2018;9:1771. <https://doi.org/10.3389/fpls.2018.01771>.
- Horton P, Ruban A. Molecular design of the photosystem II light-harvesting antenna: photosynthesis and photoprotection. *J Exp Bot.* 2005;56(411):365–73. <https://doi.org/10.1093/jxb/eri023>.
- Hou S, Zhu G, Li Y, et al. Genome-wide association studies reveal genetic variation and candidate genes of drought stress related traits in cotton (*Gossypium hirsutum* L.). *Front Plant Sci.* 2018;9:1276. <https://doi.org/10.3389/fpls.2018.01276>.
- Hu Y, Chen J, Fang L, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet.* 2019;51(4):739–48. <https://doi.org/10.1038/s41588-019-0371-5>.
- Huang G, Wu Z, Percy RG, et al. Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nat Genet.* 2020;52(5):516–24. <https://doi.org/10.1038/s41588-020-0607-4>.
- Hussain HA, Hussain S, Khaliq A, et al. Chilling and drought stresses in crop plants: implications, cross talk, and potential management opportunities. *Front Plant Sci.* 2018;9:393. <https://doi.org/10.3389/fpls.2018.00393>.
- Joshi R, Wani SH, Singh B, et al. Transcription factors and plants response to drought stress: current understanding and future directions. *Front Plant Sci.* 2016;7:1029. <https://doi.org/10.3389/fpls.2016.01029>.
- Kong F, Zhou Y, Sun P, et al. Identification of light-harvesting chlorophyll a/b-binding protein genes of *Zostera marina* L. and their expression under different environmental conditions. *J Ocean Univ China.* 2016;15(1):152–62. <https://doi.org/10.1007/s11802-016-2688-3>.
- Król M, Spangfort MD, Huner NP, et al. Chlorophyll a/b-binding proteins, pigment conversions, and early light-induced proteins in a chlorophyll b-less barley mutant. *Plant Physiol.* 1995;107(3):873–83. <https://doi.org/10.1104/pp.107.3.873>.
- Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947–8. <https://doi.org/10.1093/bioinformatics/btm404>.
- Lescot M, Déhais P, Thijs G, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 2002;30(1):325–7. <https://doi.org/10.1093/nar/30.1.325>.
- Li XW, Zhu YL, Chen CY, et al. Cloning and characterization of two chlorophyll A/B binding protein genes and analysis of their gene family in *Camellia sinensis*. *Sci Rep.* 2020;10(1):1–9. <https://doi.org/10.1038/s41598-020-61317-3>.
- Liu R, Xu YH, Jiang SC, et al. Light-harvesting chlorophyll a/b-binding proteins, positively involved in abscisic acid signalling, require a transcription repressor, WRKY40, to balance their function. *J Exp Bot.* 2013;64(18):5443–56. <https://doi.org/10.1093/jxb/ert307>.
- Longoni P, Douchi D, Cariti F, et al. Phosphorylation of the light-harvesting complex II isoform Lhcb2 is central to state transitions. *Plant Physiol.* 2015; 169(4):2874–83. <https://doi.org/10.1104/pp.15.01498>.
- Magwanga RO, Kirungu JN, Lu P, Yang X, Dong Q, Cai X, et al. Genome wide identification of the trihelix transcription factors and overexpression of *Gh\_A05G2067 (GT-2)*, a novel gene contributing to increased drought and salt stresses tolerance in cotton. *Physiol Plant.* 2019;167(3):447–64. <https://doi.org/10.1111/ppl.12920>.
- Magwanga RO, Lu P, Kirungu JN, et al. Whole genome analysis of cyclin dependent kinase (CDK) gene family in cotton and functional evaluation of the role of CDKF4 gene in drought and salt stress tolerance in plants. *Int J Mol Sci.* 2018;19(9):2625. <https://doi.org/10.3390/ijms19092625>.
- Mao H, Li S, Wang Z, et al. Regulatory changes in *TaSNAC8-6A* are associated with drought tolerance in wheat seedlings. *Plant Biotechnol J.* 2020;18(4): 1078–92. <https://doi.org/10.1111/pbi.13277>.
- Nath K, Kajoo A, Poudyal RS, et al. Towards a critical understanding of the photosystem II repair mechanism and its regulation during stress conditions. *FEBS Lett.* 2013;587(21):3372–81. <https://doi.org/10.1016/j.febslet.2013.09.015>.
- Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet.* 2005;39(1):121–52. <https://doi.org/10.1146/annurev.genet.39.073003.112240>.
- 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(21):8777–96. <https://doi.org/10.1007/s00253-014-6020-0>.
- 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(9):20392–416. <https://doi.org/10.3390/ijms160920392>.
- Pintó-Marijuan M, Munné-Bosch S. Photo-oxidative stress markers as a measure of abiotic stress-induced leaf senescence: advantages and limitations. *J Exp Bot.* 2014;65(14):3845–57. <https://doi.org/10.1093/jxb/eru086>.
- Qin D, Dong J, Xu F, et al. Genome-wide identification and characterization of light harvesting chlorophyll a/b binding protein genes in barley (*Hordeum vulgare* L.). *Adv Crop Sci Technol.* 2017;05(04). <https://doi.org/10.4172/2329-8863.1000301>.
- Sasi S, Venkatesh J, Daneshi RF, Gururani M. Photosystem II extrinsic proteins and their putative role in abiotic stress tolerance in higher plants. *Plants.* 2018; 7(4):1–15. <https://doi.org/10.3390/plants7040100>.
- Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C<sub>T</sub> method. *Nat Protoc.* 2008;3(6):1101–8. <https://doi.org/10.1038/nprot.2008.73>.
- Tamura K, Peterson D, Peterson N, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Bio Evol.* 2011;28(10):2731–9. <https://doi.org/10.1093/molbev/msr121>.
- Wang K, Wang Z, Li F, et al. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet.* 2012;44(10):1098–103. <https://doi.org/10.1038/ng.2371>.
- Xu Y, Magwanga RO, Cai X, et al. 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(8):1863. <https://doi.org/10.3390/ijms20081863>.
- Xu Y, Magwanga RO, Yang X, et al. Genetic regulatory networks for salt-alkali stress in *Gossypium hirsutum* with differing morphological characteristics. *BMC Genomics.* 2020;21(1):1–19. <https://doi.org/10.1186/s12864-019-6375-9>.
- Xu YH, Liu R, Yan L, et al. Light-harvesting chlorophyll a/b-binding proteins are required for stomatal response to abscisic acid in *Arabidopsis*. *J Exp Bot.* 2012;63(3):1095–106. <https://doi.org/10.1093/jxb/err315>.
- Yamaguchi-Shinozaki K, Shinozaki K. Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends Plant Sci.* 2005;10(2):88–94. <https://doi.org/10.1016/j.tplants.2004.12.012>.
- Zahid KR, Ali F, Shah F, et al. Response and tolerance mechanism of cotton *Gossypium hirsutum* L. to elevated temperature stress: a review. *Front Plant Sci.* 2016;7:1–13. <https://doi.org/10.3389/fpls.2016.00937>.
- Zhao L, Lü Y, Chen W, et al. Genome-wide identification and analyses of the *AHL* gene family in cotton (*Gossypium*). *BMC Genomics.* 2020;21(1):1–14. <https://doi.org/10.1186/s12864-019-6406-6>.
- Zou Z, Li M, Jia R, et al. Genes encoding light-harvesting chlorophyll a/b-binding proteins in papaya (*Carica papaya* L.) and insight into lineage-specific evolution in Brassicaceae. *Gene.* 2020;748:144685. <https://doi.org/10.1016/j.gene.2020.144685>.
- 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.org/10.1016/j.gene.2019.03.071>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

