Comprehensive Identification and Expression Analysis of CRY Gene Family in Gossypium

Chaochen Huang
Center for Excellence in Molecular Plant Sciences
Pengbo Li
Shanxi Agricultural University
Junfeng Cao
Center for Excellence in Molecular Plant Sciences
Zishou Zheng
Center for Excellence in Molecular Plant Sciences
Jinquan Huang
Center for Excellence in Molecular Plant Sciences
Xiufang Zhang
Center for Excellence in Molecular Plant Sciences
Xiaoxia Shangguan
Shanxi Agricultural University
Lingjian Wang
Center for Excellence in Molecular Plant Sciences
Zhiwen Chen (b1301031@cau.edu.cn)
Center for Excellence in Molecular Plant Sciences

Research Article

Keywords: Cryptochrome, Cotton, Phylogenetics, Organ-specific expression, Abiotic stress

Posted Date: January 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1165543/v1

Version of Record: A version of this preprint was published at BMC Genomics on March 24th, 2022. See the published version at https://doi.org/10.1186/s12864-022-08440-9.

Abstract

Background: The cryptochromes (CRY) comprise a specific blue light receptor for plants and animals, which play crucial roles in physiological processes of plant growth, development, and stress tolerance.

Results: In the present work, a systematical analysis of CRY gene family from five allotetraploid cotton species, *G. hirsutum*, *G. barbadense*, *G. tomentosum*, *G. mustelinum* and *G. darwinii* together with seven diploid species. There were 18, 17, 17, 17, and 17 CRYs identified in *G. hirsutum*, *G. barbadense*, *G. tomentosum*, *G. mustelinum* and *G. darwinii*, respectively, whereas five to nine CRY genes were identified in the diploid species. Phylogenetic analysis of the protein-coding sequences revealed that CRY genes from the allotetraploids *G. hirsutum* and *G. barbadense*, three diploid cotton species (*G. raimondii*, *G. herbaceum*, and *G. arboreum*), and *Arabidopsis thaliana* could be classified into seven clades. Synteny analysis suggested that the homoeolog of *G. hirsutum Gh_A02G0384* has undergone an evolutionary loss event in the other four allotetraploid cotton species. *Cis*-element analysis predicated the possible functions of CRY genes in *G. hirsutum*. Public RNA-seq data were investigated to analyze the expression patterns of *G. hirsutum* CRY genes in various tissues as well as gene expressions under abiotic stress treatments.

Conclusion: These results indicated the possible functions of *G. hirsutum* CRY genes in differential tissues as well as in response to abiotic stress during the cotton plants life cycle.

Background

Cryptochromes (CRYs) are found in archaea, bacteria, algae, terrestrial plants, and humans, and they are photoreceptors for plants and animals [1-3]. Cryptochromes were first identified in Arabidopsis, named as HY4 or CRY1, which encodes a DNA photolyases protein [4, 5]. CRY1 gene was responsible for the blue-light inhibition of the hypocotyl elongation in Arabidopsis [5]. CRY1 contains flavin adenine dinucleotide (FAD), which is the binding domain of photolyases [6]. Cryptochromes contain two-domain structure: the highly conserved FAD-binding photolyase homology region (PHR) domain and the divergent CRY C-terminal extension (CCE) domain [7].

The numbers of cryptochromes vary among plant species, ranging from three in Arabidopsis to seven in soybean [8, 9]. At present, all higher plants studied have two diverged cryptochromes phylogenetically clades, CRY1 and CRY2 [10, 11]. Most cryptochromes in plants could regulate the gene transcription and expression of the plant life cycle [12-14]. Among them, CRY1 could inhibit the hypocotyl elongation of Arabidopsis [4], and inhibit the grain dormancy and germination in barley [15]. Arabidopsis cryptochrome 1 also controls photomorphogenesis through regulation of H2A.Z deposition [16]. CRY2 could interact with CIB1 or SPA1 to regulate the floral initiation in Arabidopsis [17, 18] as well as suppress the leaf senescence in soybean [19]. CRY1 and CRY2 together could stimulate the stomata opening and development in Arabidopsis [20, 21].

Among the many processes regulated by cryptochromes, responses to biotic and abiotic stresses, such as drought, salinity, heat, and so on, are one of the most active research topics in plant biology [22]. Studies have demonstrated that cryptochromes in Arabidopsis enhance plants resistance to *Pseudomonas syringae* [23] and drought [21]. In tomato, cryptochrome 1a could modulate water deficit and osmotic stress responses [24] as well as mediate long-distance signaling of soil water deficit [25]. In rice, suppression of cryptochrome 1b improved salt tolerance by down-regulation of the expressions of melatonin and brassinosteroid biosynthetic genes [26]. Overexpressing the wheat cryptochromes TaCRY1a and TaCRY2 into Arabidopsis led to more highly sensitive to high salt stress in the transgenic Arabidopsis plants [27]. In addition, the *cry1* mutant had a greater germination and seedling survival rate than the WT plants in response to salt conditions, showing more tolerance to salinity in Arabidopsis [28]. Obviously, these results provide evidences on proving the role of acting as negative regulator of cry in plant response to the salinity environments.

Cotton (from the genus *Gossypium*) is the most important fiber crop in the world. The genus contains more than 50 species [29, 30], recent advances in cotton genomics have produced the resources necessary to characterize the CRY gene family in *Gossypium.* Multiple high-quality genome sequences are available for several species, including diploid species, i.e., *Gossypium thurberi* (D_1), *G. raimondii* (D_5), *G. turneri* (D_{10}) [31-34], *G. herbaceum* (A_1; cultivated), *G. arboreum* (A_2; cultivated) [35-37], *G. longicalyx* [38], *G. australe* [39] and tetraploid *G. hirsutum* (AD_1; cultivated), *G. barbadense* (AD_2; cultivated), *G. tomentosum* (AD_3), *G. mustelinum* (AD_4) and *G. darwinii* (AD_5) [37, 40-47], and *Gossypium* sister genera *Gossypioides kirkii* [48]. These genome sequences will provide a good platform for dissecting gene functions by forward and reverse genetics for molecular breeding in cotton. Although genome sequencing has facilitated the functional characterizations of cotton genes, the CRY genes in *Gossypium* have not been extensively explored.

In the current study, we performed a genome-wide screening for CRY genes in cottons, based on data gathered from recent whole-genome sequencing results. We used in silico approach to identify and characterize CRY genes, and then focused on the characterization and phylogenetic relationships of *Gossypium* species. The CRY gene structures, conserved domains, synteny, as well as *cis*-elements were

explored in this systematical analysis. Moreover, the tissue-specific expression patterns and the transcriptional responses of GhCRYs to abiotic stresses were examined. Our data provide inspirations for molecular design of cotton cultivars with desired traits.

Methods

Identification of CRY family genes and CRY proteins in diploid and tetraploid Gossypium species

We downloaded the genome sequences of cotton species from the CottonGen database [62], including *G. raimondii* [31], *G. herbaceum* [37], *G. arboreum* [35], *G. hirsutum* [40], *G. barbadense*, *G. tomentosum*, *G. mustelinum* and *G. darwinii* [37, 40-47]. To identify all putative CRY transcription factor proteins in each genome assembly, the CRY protein conserved domains (PF00875 for DNA photolyase) were used to develop a Hidden Markov Model [63] profile matrix via the hmmbuild program from the HMMER package [64] using default parameters. This HMM profile matrix was used in conjunction with hmmersearch with default parameters against these *Gossypium* genome databases to identify putative CRY genes (*GhCRY*s). Previously identified *CRY* gene sequences from *Arabidopsis thaliana* (*AtCRY*s) were retrieved from the TAIR database [65] for phylogenetic comparison. The presence of conserved domains in each *Arabidopsis* and *Gossypium* gene was verified using the SMART conserved domain search tool [66] and Pfam databases [67].

Chromosomal location and gene structure analyses

Chromosomal locations for each of the above identified *GhCRYs* were extracted from the genome annotation gff3 file [40]. Chromosomal locations of the predicted *GhCRYs* was visualized using TBtools [68], and the exon-intron structure of each gene was displayed using the online tool GSDS 2.0 [69]. The number of amino acids, molecular weight (MW), and theoretical isoelectric point (pI) of putative *GhCRYs* proteins were determined using the ProtParam tool [70].

Sequence alignment, Ka, Ks and phylogenetic analyses

Complete protein-coding sequences for CRY genes from *Gossypium* and AtCRY were aligned using MAFFT with the G-INS-i algorithm [71]. The nonsynonymous substitutions rate (Ka) and synonymous substitution rate (Ks) were calculated using the DnaSP 6.0 [72]. The NJ phylogenetic tree was constructed using MEGA version 6.0 [73] by sampling 1000 bootstrap replicates.

Analysis of Cisacting element in promoter regions of GhCRYs

The upstream sequences (1.5 kb) of the *GhCRYs* genes were retrieved from *G. hirsutum* genome sequence based on the gene locations [40]. Then, the retrieved promoter sequences were submitted to PlantCARE [53] to identify the potential Cisacting element.

Chromosomal mapping and synteny analysis of CRY genes in diploid and allotetraploid Gossypium species

CRY genes were mapped on chromosomes using TBtools [68] software. Blastn was used to determine *CRY* gene synteny. Then, TBtools [68] software was applied to express the syntenic relationship of the homologous gene pairs.

Expression patterns of GhCRYs in different tissues and stress conditions

Raw RNA-Seq data for *G. hirsutum* seed, root, stem, leaf, torus, petal, stamen, ovary, calyx, ovule (-3 dpa, -1 dpa, 0 dpa, 1 dpa, 3 dpa, 5 dpa, 10 dpa, 20 dpa, 25dpa, 35dpa) and fiber (5 dpa, 10 dpa, 20 dpa, 25dpa) were downloaded from the NCBI Sequence Read Archive (PRJNA 248163) [40], represented by one library each. Reads were mapped to *G. hirsutum* genome [40] via HISAT2 software with default parameters, and read abundance with calculated via StringTie [74, 75]. Read counts were normalized in R3.2 using RUVSeq [76] and the internal control reference gene *GhUBQ7*, which is detected at relatively constant levels across different cotton samples [77]. Potential batch-effects were corrected by an improved version of ComBat, ComBat-seq [78]. Gene expression was estimated by Ballgown [86], using fragments per kilobase million (FPKM) values to calculate the gene expression levels across libraries. Expression levels of *G. hirsutum* leaf RNA-Seq data (in FPKM) for each *GhCRY* gene under drought, salt, heat and cold stress (time points: 0, 1, 3, 6, 12h) were retrieved from the ccNET database [79]. Genes were considered differentially expressed if expression varied more than two-fold change with a p-value of less than 0.05. TBtools [68] was used to display the gene expression patterns from the calculated FPKM values.

Plant cultivation and treatment

To generate new expression information via qRT-PCR, we grew representatives of *G. hirsutum, G. arboreum*, and *G. barbadense*. For *G. hirsutum*, seeds of *G. hirsutum* cv. R15 [80], were germinated in potting soil in a growth chamber, and the resulting seedlings were maintained in a controlled environment at 28°C day/20°C night, with a 16-h light/8-h dark photoperiod. Roots, stems, leaves, cotyledons, and hypocotyls were collected from the three-week old plants, and additional samples were collected from older, flowering plants; these include petal, stamen, pistil, sepals, torus, ovules (0 dpa (days post anthesis)) and fiber (6 dpa). Three biological replicates were collected for each sample, each

with three technical replicates. For salt treatment, 28-day old plants were sprayed with 200 and 500 mM NaCl solution after surfactant (Triton X-100) treatments. Leaves from salt-treated plants were collected at 0 (control), 1, 3, and 6 h post-NaCl treatment for further expression analyses. All plant tissues were frozen in liquid nitrogen immediately after collection and stored at -80°C until RNA extraction. All treatments were sampled at least three times.

Similarly, *G. arboreum* cv. Shixiya1 and *G. barbadense* H7124 were grown for qRT-PCR of salt-exposed leaf tissue timepoints only. For this experiment, seeds of *G. arboreum* cv. Shixiya1 were provided by Prof. Tianzhen Zhang and *G. barbadense* H7124 seeds were provided from the Esquel Group. These two *Gossypium* species were planted in the Damao field in Sanya, Hainan Province, China. For salt treatment, 50-day old plants were sprayed with 200 and 500 mM NaCl solution after surfactant (Triton X-100) treatment. Leaves from salt-treated plants were collected at 0 (control), 1, 3, and 6 h post-NaCl treatment as above.

RNA extraction, cDNA synthesis and qRT-PCR expression analyses

Total RNAs from cotton tissues were extracted using the RNAprep pure plant kit (TIANGEN, Shanghai, China) according to the manufacturer's protocol. The resulting RNAs were treated with DNase I prior to synthesizing cDNA with oligo (dT) primers and M-MLV Reverse Transcriptase (Invitrogen); these products were diluted 5-fold before use. For quantitative real-time PCR (qRT-PCR), Primer5 software was used to design gene-specific forward and reverse primers (Additional file 2: Table S2). As these primers are not homoeolog specific, both copies were amplified when retained in duplicate. Analyses were performed with SYBR-Green PCR Mastermix (TaKaRa) on a cycler (Mastercycler RealPlex; Eppendorf Ltd, Shanghai, China). The *G. hirsutum histone-3* (*GhHIS3*) and *GhUBQ7* genes were used as internal references, and the relative amount of amplified product was calculated following the 2-ΔΔCt method [81]. For the *G. hirsutum* samples, relative expression levels among different organs were normalized by calibrating with the root sample from that plant. The root sample was washed with DEPC sterile water three times before extracting the RNA.

G. hirsutum leaf microRNAs were isolated with miRcute Plant miRNA Isolation Kit (TIANGEN, Shanghai, China), reverse transcribed by the miRcute Plus miRNA First-Strand cDNA Kit (TIANGEN, Shanghai, China), and diluted 10-fold before use. The miRcute Plus miRNA qPCR Kit (TIANGEN, Shanghai, China) was used to perform quantitative real-time PCR (qRT-PCR) with U6 as the internal reference. The forward and reverse primers are listed in Additional file 2: Table S2.

Results

Identification and chromosomal location of CRY family genes in G. hirsutum

Cryptochrome (CRY) is a class of photolytic flavin proteins with UV-A/blue light receptors, which play an important role in plant growth and development [49]. These proteins are defined by the presence of a DNA photolyase domain [10], which is responsible for DNA binding to a flavin adenine dinucleotide (FAD) [50, 51]. Hmmersearch against the *G. hirsutum* genome database with the conserved domains (PF00875 for DNA photolyase domain) identified 18 CRY genes (Table 1). These 18 *GhCRY* genes are dispersed over 14 of the 26 *G. hirsutum* chromosomes, with most, but not all, homoeologs conserved (Fig. 1).

Table 1
Sequence characteristics of *GhCRY* (*Gossypium hirsutum* cryptochrome) genes and proteins.

Locus Name	Chr	Genomics Position	CDS	No. of Introns	Size (aa)	DNA Photolyase Domain	FAD binding 7 Domain	Cryptochrome C Domain	Hydrolase 4 Domain	MW	pI
Gh_A02G0384	A02	4840292-4843745	2,268	4	755	6-169	285-485			85.81	6.26
Gh_D02G0436	D02	5763605-5767015	2,268	4	755	6-170	285-485			85.49	6.41
Gh_A03G0120	A03	1871238-1873553	1,497	8	498	34-200				57.36	8.89
Gh_D03G1520	D03	44312391-44314703	1,497	8	498	34-200				57.38	8.89
Gh_A05G1941	A05	20388600-20391811	2,049	3	682	7-173	283-483	510-627		76.95	5.82
Gh_D05G2172	D05	20346471-20349578	2,049	3	682	7-173	283-483	510-627		76.98	5.55
Gh_A05G2282	A05	26904700-26907437	2,022	3	673	3-168	279-479	506-618		76.23	5.62
Gh_D05G2543	D05	25786819-25789574	2,022	3	673	3-168	279-479	506-618		76.17	5.62
Gh_A06G0969	A06	43955190-43958563	1,629	12	542	52-244	345-529			61.75	9.06
Gh_D06G1145	D06	26445874-26448973	1,380	10	459	82-268	368-456			52.29	9.43
Gh_A06G1059	A06	62119515-62125030	1,644	13	547	18-189	303-503			63.19	8.96
Gh_D06G2339	D06	67821-73312	1,644	13	547	18-188	303-503			63.32	8.97
Gh_A09G2012	A09	73305167-73307596	1,977	3	658	21-185	300-500			74.87	5.86
Gh_D09G2225	D09	49444930-49447313	1,935	3	644	7-171	286-486			73.59	6.08
Gh_A11G1040	A11	11750568-11752551	1,389	3	462	126-296				50.37	8.00
Gh_D11G1195	D11	11089391-11091357	1,392	3	463	126-296				50.63	7.52
Gh_A12G2401	A12	86483003-86486801	2,073	13	690	46-202			425-651	78.43	6.29
Gh_D12G2528	D12	58181188-58184994	2,073	13	690	46-195			425-650	78.35	7.63

Note: bp: Base pair, Chr.: Chromosome, aa: Amino acid, MW: Molecular weight, kDa: Kilodalton, pI: Isoelectric point.

Structural organization of GhCRY **genes**

Less than twofold variation in length was detected in the predicted coding sequences (CDS) for the recovered *GhCRY*s, from 1380 bp for *Gh_D06G1145* to 2,268 bp for *Gh_A02G0384/Gh_D02G0436* (Table 1), which translates to proteins ranging from 459 amino acids (aa) (52.29 kDa) to 755 aa (85.81 kDa). Predicted isoelectric points (pI) for members of this family also vary widely, from 5.55 to 9.43. All of the putative GhCRY proteins have DNA photolyase domain in the N-terminal region (Table 1). Twelve GhCRY proteins have FAD binding 7 domain, four proteins have cryptochrome C domain and two have hydrolase 4 domain in the C-terminal region, respectively (Table 1).

While all putative *GhCRY* genes contain introns (Fig. 2), they also exhibit considerable variations, in both length and number. In general, homoeologous *GhCRY* genes show highly similar intron patterns, however, intron structure among homoeologous pairs can exhibit variation in intron number (3 to 13) and length. One of the homoeologous gene pairs did exhibit divergence in structure, namely *Gh_A06G0969 vs*

Gh_D06G1145, which contain 12 and 10 introns, respectively. Characterization of parental (both 12 introns in the diploids) gene structure for the homoeologs suggests that this structural variation was descendant divergence rather than inherited. In addition, phylogenetic relationship of the *GhCRY* gene family was not consistent with the intron/exon structure characterized (Fig. 2).

Phylogenetic analysis of CRY family genes in Gossypium

The general conservation of CRY genes between the two subgenomes of allotetraploid prompted us to ask whether the minimal loss and/or gain occurred before or after the marriage of the two diploid progenitors. We specifically assessed this using the protein-coding sequences of 62 cotton CRY genes (*G. hirsutum*, 18; *G. barbadense*, 17; *G. raimondii*, 9; *G. arboreum*, 9; and *G. herbaceum*, 9) with 3 *Arabidopsis thaliana* CRY genes for phylogenetic analysis (Fig. 3). Seven clades (I–VII) were robustly supported with one *A. thaliana* gene associated with clades I, II and VI, respectively, and other four clades were composed of *Gossypium* CRY genes only.

Overall, the expected diploid-polyploid topology is reflected in the tree for each set of orthologous/homoeologous genes, indicating general preservation during diploid divergence and through polyploid evolution. That is, the number of CRY genes in tetraploids was generally additive with respect to the model diploid progenitors, with each homoeolog (A_t or D_t) sister to their respective parental copies. Clades I and II had the most CRY genes, and other five clades contain equal number (Fig. 3). In clades I and II, genes related to *AtCRY1* and *AtCRY2* exhibit duplication in *Gossypium* species, which indicate a duplication event in *Gossypium* compared to the *A. thaliana*. In addition, the *Gossypium* CRY genes of clade I have a sister relationship with AtCRY1, and clade II have the closest relationship with AtCRY2, and clade VI was classified with AtCRY3. Therefore, it is speculated that the function of CRY genes in these clades of cotton is similar to that of the corresponding CRY genes in Arabidopsis.

Although the CRY family exhibits general preservation, a few deviations were noted. For example, Clade II exhibits evidence of homoeolog loss; that is, the A copy of *GB_D02G0441* is missing from *G. barbadense* genome, whereas both copies (*Gh_A02G0384*/ *Gh_D02G0436*) exist in *G. hirsutum*. This gene loss might specific to *G. barbadense* after divergence of the two allotetraploid species.

Divergence of CRY genes in allotetraploid G. hirsutum and its diploid progenitors

The *CRY* genes in the two diploid species were then compared with *G. hirsutum* A_t- and D_t-subgenome homoeologs (Fig. 4, Additional file 1: Table S1). To explore the evolutionary relationship and possible functional divergence of *CRY* genes between the allotetraploid cotton and its extend diploid progenitors, the nonsynonymous substitution (*Ka*) and synonymous substitution values (*Ks*) and the *Ka*/*Ks* ratios for each pair of the genes were calculated (Additional file 1: Table S1). By comparing the *Ka* and *Ks* values of 18 orthologous gene sets between the allotetraploid and its diploid progenitor genomes, we found that the *Ka* and *Ks* values are higher in the D_t subgenome than in the A_t subgenome (Fig. 4a, b). These results indicate that *GhCRY* genes in the D_t subgenome tend to have experienced faster divergence than their A_t counterparts. However, the *Ka/Ks* ratios of D_t subgenome was lower than that of A_t subgenome (Fig. 4c), indicating that GhCRY genes in A_t subgenome were subjected to positive selection during the process of evolution and might cause diverged functions.

Dynamic evolution of CRY family genes in Gossypium

We further evaluated the general preservation of CRY genes in 12 *Gossypium* species, *Gossypioides kirkii* and *Arabidopsis thaliana* (Fig. 5). In *Arabidopsis thaliana*, only three *CRY* genes were identified; and the relative of *Gossypium*, *Gossypioides kirkii* have eight *CRY* genes. However, all *Gossypium* species surveyed recovered a minimum of five putative *CRY* genes in *G. australe* (G_2) and 18 in *G. hirsutum* (AD_1). Among D genome species, *CRY* gene copy number varied from the minimum of 7 in *G. thurberi* (D_1), to both 9 in *G. raimondii* (D_5) and *G. turneri* (D_{10}), respectively. The two cultivated diploid species of *G. herbaceum* (A_1) and *G. arboreum* (A_2) both have 9 copies. Another sister-species of A-genome G. longicalyx (F_1) contains 6 CRY genes. CRY copy numbers in the allotetraploid species surveyed varied from 17 putative *CRY* genes in four species: *G. barbadense* (AD_2), *G. tomentosum* (AD_3), *G. mustelinum* (AD_4), *G. darwinii* (AD_5) to 18 in *G. hirsutum* (AD_1). Notably, this high copy number in tetraploid is slightly more than double the copy number in diploid, likely reflective of the duplicated history of cotton. Comparatively, *G. hirsutum* (AD_1) CRY copy number is generally stable after polyploidization. The other four allotetraploid cotton species included here all appear to have undergone a homoeolog loss (17 versus18), where *G. hirsutum* has retained it on the A02 chromosome (Gh_A02G0384) (as stated above, Fig. 3).

Chromosomal distribution and synteny analysis of Gossypium CRY genes

Based on these *Gossypium* genomes, the location of *CRY* genes and the length of chromosomes from two diploid species and five allotetraploid species were used to analyze the chromosomal distribution and synteny analysis (Fig. 6). High similarity was found in the chromosomal distribution patterns of these seven cotton species. The CRY genes were unevenly arranged on chromosomes with divergence existed between the diploid and allotetraploid species. For instance, no CRYs were found on Chr 02 of two diploid progenitor species and Chr

A02 of four allotetraploid specie (*G. barbadense* (AD_2), *G. tomentosum* (AD_3), *G. mustelinum* (AD_4), *G. darwinii* (AD_5)), but existed on Chr A02 in *G. hirsutum*. There were totally 104 CRYs distributed throughout the 80 chromosomes comprising 38 located on the A or At subgenomes and 42 located on the D or Dt subgenomes. The majority of CRYs were located on the proximate or the distal ends of the chromosomes. In addition, there were nine collinear gene pairs between *G. raimondii* and *G. arboreum*, nine between A_t and D_t subgenomes of *G. hirsutum*, and eight for the other four allotetraploid subgenomes.

Cis **element analysis of CRY genes in** G. hirsutum

Cis-elements are responsive to corresponding stimulations to regulate the expression of genes [52]. In this study, a 1.5-kb upstream region from the start codon of each CRY gene in *G. hirsutum* was extracted to investigate putative *cis*-elements involved in the mediation of gene expression using the PlantCARE server [53]. We totally identified 581 *cis*-elements among 18 *GhCRY* genes, ranging from 22 in Gh_A02G0384 to 47 in Gh_D03G1520 (Fig. S1). Some *cis*-elements were predicted to be involved in phytohormone (ABRE) and stress (TC-rich repeats) responses (Fig. S1). 13 *GhCRY* gene promoters processed at least one abscisic acid responsiveness element (ABRE), and 11 *GhCRY* genes had at least one antioxidant response element (ARE). There were 9 GhCRYs processed the elements involved in the MeJA-responsiveness (CGTCA-motif and TGACG-motif), four had the *cis*-acting element involved in salicylic acid responsiveness (TCA element), two had gibberellin-responsive element (GARE-motif), and three had auxin-responsive element (TGA-element and AuxRR-core). 18 GhCRYs had light responsiveness element (Box 4 and G box) and at least two MYB binding site element (MYB). Three GhCRYs had meristem expression element (CAT-box) and four GhCRYs (Gh_D02G0436, Gh_D03G1520, Gh_A05G2282 and Gh_D05G2543) had low-temperature responsiveness element (LTR). The remaining elements related to stress, like defense and stress responsiveness (TC-rich repeats), recognized by WRKY transcription factors (W-box) and wound-responsive element (WUN-motif), MYB-binding site involved in drought inducibility (MBS), were also existed in these GhCRY genes.

Expression patterns of GhCRY **genes in different** G. hirsutum **tissues**

The expression profile of a gene family can provide valuable clues to possible functions of each gene. Analysis of 18 *GhCRY* genes showed that most genes have different spatial expression patterns. For instance, the expression levels of *Gh_A09G2012*, *Gh_D09G2225*, *Gh_A11G1040* and *Gh_D11G1195* in root, stem, leaf, torus, stamen, pistil and calycle were significantly higher than those of other *GhCRY* genes (Fig. 7a). *Gh_A06G1059* presented the highest expression level in petal (Fig. 7a). In addition, these four genes, *Gh_A09G2012*, *Gh_D09G2225*, *Gh_A11G1040* and *Gh_D11G1195*, still showed significant higher expression in seed, root and cotyledon samples at different time points after seed germination (Fig. 7b). In ovule samples of different development stages, *Gh_A09G2012*, and *Gh_D09G2225* had the highest expression levels followed by *Gh_A11G1040* and *Gh_D11G1195* (Fig. 7c). In different development periods of fiber samples, *Gh_A09G2012* had the highest expression in 20 and 25 dpa fiber cells (Fig. 7d), suggesting that this gene might play an important role at the cell wall thickening stage. *Gh_D09G2225* was preferentially expressed in 0 dpa ovule, implying a role at the initial stage of fiber development. The above two genes were the homoeologs of the Arabidopsis AtCRY2 (Fig. 3). *Gh_A11G1040* showed the highest expression in 5 and 10 dpa fiber cells, which suggests that it may play a role of regulation at the elongation stage of fiber development (Fig. 7d). Therefore, these three genes could be taken as candidate genes for subsequent transformation experiments to verify their functions in cotton.

Expression changes of GhCRY **genes in** G. hirsutum **under different stresses**

Cotton is often subjected to a variety of abiotic stresses during its growth and development. Therefore, we comprehensively analyzed the expression changes of CRY genes under simulated drought (PEG 6000), salt (NaCl), heat and cold abiotic stresses from RNA-seq data (Fig. 8). At different time points of PEG6000 simulation drought condition, expressions of most *GhCRY* gene were not changed ($|lg_2$ (Fold change)$| \geq 1|$ as the threshold of differentially expressed genes). For instances, expression of *Gh_A12G2401* gene was down-regulated after 1 h PEG treatment (Fig. 8a). The expression of *GhCRY* genes did not change after PEG treatment for 3 and 6h (Fig. 8a). After 12 hours of PEG treatment, the expression of one *GhCRY* gene (*Gh_A03G0120*) were repressed (Fig. 8a). These results indicated that few *GhCRY* genes were involved in response to drought stress in *G. hirsutum* plants.

Under salt stress, the expressions of *GhCRY* gene did not change after 1, 3 and 6h of NaCl treatment (Fig. 8b). However, 15 *GhCRY* genes were down-regulated after NaCl treatment for 12h (Fig. 8b), indicating that these *GhCRY* genes negatively regulate salt stress.

At the four time points of high temperature stress, the expression of *GhCRY gene* did not change after 1 and 3 h of high temperature treatment (Fig. 8c). After 6h of high temperature treatment, the expression levels of four *GhCRY* genes (*Gh_A05G2282*, *Gh_D06G1145*, *Gh_A12G2401* and *Gh_D12G2528*) were down-regulated (Fig. 8c). After 12h of high temperature treatment, the expression levels of nine *GhCRY* genes were down-regulated (Fig. 8c). Among which, *Gh_D06G1145* gene was both inhibited at 6 and 12h, which indicated that this gene might be a key factor in regulating the response to high temperature stress in *G. hirsutum*.

As for the four time points of low temperature stress in *G. hirsutum*, there were four *GhCRY* genes which were up-regulated after 1 h of low temperature treatment (Fig. 8d). At the 3 h after low-temperature treatment, the expressions of five *GhCRY* genes were induced, and two were reduced (Fig. 8d). After 6 hours of low temperature treatment, the expression levels of eight *GhCRY* genes were elevated and two were declined (Fig. 8d). After low temperature treatment for 12 h, the expression levels of five *GhCRY* genes were up-regulated, and five were down-regulated (Fig. 8d). Among them, the homoeologs of *AtCRY3*, *Gh_A06G0969* and *Gh_D06G1145* were low temperature induced at the four-time point stresses, indicating their positively regulating roles under low temperature stress. Therefore, these two genes can be prospective candidates for subsequent transformation experiments to verify their function in response to low temperature stress in cotton.

Discussion

CRYs in plants contain two conversed domains, the N-terminal photolyase homology-related (PHR) domain and a cryptochrome C-terminal extension (CCE) domain [54]. In Arabidopsis thaliana, three CRYs were identified (cry1, cry2, and cry3). Cry1 and cry2 are located in the nucleus and play a multifaceted role in various aspects of plant growth and development [1, 55]. For instance, cry1 primarily regulates photomorphogenic responses related to the inhibition of hypocotyl elongation, anthocyanin accumulation and cotyledon expansion, while cry2 plays a role in the hypocotyl inhibition, circadian clock and photoperiod-dependent flowering [56]. However, cry3 is a DASH protein located in chloroplasts and mitochondria [50], which works to repair UV-damaged DNA in a light-dependent manner [54]. Overall, cryptochrome-mediated photoresponses remain unclear with the existing differences of plant species as well as their physiological responses variations [57, 58].

It has been reported that plant cryptochromes were involved in the adversity stress response [22, 24, 59, 60]. In Arabidopsis, study has demonstrated that CRYs play an interesting role in drought stress tolerance [21]. They found that overexpressing the CRY1 protein exhibited excessive water loss whereas Arabidopsis double mutant *cry1cry2* plants were clearly more drought-tolerant than the wild type (WT). In addition, overexpressing *Triticum aestivum TaCRY1a* and *TaCRY2* into Arabidopsis plants exhibited a lower osmotic stress tolerance, including drought and salt stresses [27]. Meanwhile, overexpression of *Sorghum bicolor SbCRY1a* gene into Arabidopsis produced transgenic plants that were oversensitive to salt stress, however, the *cry1* mutant showed more tolerance to salinity [28]. In tomato (*Solanum lycopersicum* L.), cryptochrome 1a (cry1a) modulated the water deficit response under osmotic stress conditions, further to increase tomato growth by reduced malondialdehyde (MDA) and proline accumulation [24]. In addition, tomato *CRY1a* mutant plants could enhance its drought tolerance by increasing the leaf relative water content [25]. In *Brassica napus*, overexpression CRY1 resulted in plants that were very sensitive to osmotic stress whereas the transformed antisense silencing plants were more tolerant [59]. However, the relationship of CRYs and abscisic acid (ABA), or other blue light photoreceptors together to modulate water loss under drought or salt stresses are still very unclear, which need further excrements to clarify the interaction network.

Plant cryptochromes have also been implicated in adaptations to temperature variation, for instance, report showed that low temperatures would increase biological activity of CRY [61]. Although there have been many studies on the response of CRY in regulating plant growth and development as well as responses to adverse stresses, there are few studies on the exploration the potential roles of *CRYs* in cotton. Thus, *CRYs* manipulation may be an ideal biotechnological target to design the multiple stress-tolerant cotton varieties.

Conclusions

We systematically analyzed cotton CRY family genes using bioinformatic approaches and gene expression analyses. We analyzed gene structures, chromosomal locations, intron-exon organizations, phylogenetic relationships and expression profile patterns in different cotton tissues and under different stress condition to predict their possible biological functions. These *GhCRY* genes are variably expressed in different cotton tissues with particularly high expression in fibers. The decreased expressions of several *GhCRY* genes in response to multiple abiotic stress implies their involvement in the regulation of growth and development under the abiotic stress conditions in cotton. Together, our results provide candidate genes to facilitate the functional identification of the CRY genes in cotton to module plant growth, development and stress tolerance.

Abbreviations

CRYs: Cryptochromes; DPA: Days post anthesis; FPKM: Fragments per kilobase of transcript per million mapped fragments; *G. arboreum*: *Gossypium arboreum*; *G. hirsutum*: *Gossypium hirsutum*; *G. raimondii*: *Gossypium raimondii*; MW: Molecular weight; pI: Isoelectric point; qRT-PCR: Quantitative real-time polymerase chain reaction

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The genome sequences of cotton species and the genome annotation gff3 file in this manuscript were downloaded from the CottonGen database (https://www.cottongen.org/data/download) [61]. Raw RNA-Seq data for *G. hirsutum* seed, root, stem, leaf, torus, petal, stamen, ovary, calyx, ovule and fiber were downloaded from the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA248163) (NCBI Sequence Read Archive SRR1695173, SRR1695174, SRR1695175, SRR1695177, SRR1695178, SRR1695179, SRR1695181, SRR1695182, SRR1695183, SRR1695184, SRR1695185, SRR1695191, SRR1695192, SRR1695193,SRR1695194, SRR1768504, SRR1768505, SRR1768506, SRR1768507, SRR1768508, SRR1768509, SRR1768510, SRR1768511, SRR1768512, SRR1768513, SRR1768514, SRR1768515, SRR1768516, SRR1768517, SRR1768518 and SRR1768519) [30]. The conserved domain of CRY proteins (Pfam ID: PF00875) was downloaded from the Pfam databases (http://pfam.xfam.org/family/PF06507#tabview=tab3). All other data generated or analyzed during this study are included in this published article and its Additional files.

Competing interests

The authors declare that they have no competing interests

Funding

This work reported in this publication was supported by the National Natural Science Foundation of China through the Awards Nos. 31690092, 52071192, the Doctoral Research Initiation Foundation project of Shanxi Datong University (2019-B-03), the Platform and Base Project of Datong (2020190), the Open Project Foundation of the Shanxi Key Laboratory of Cotton Germplasm Resources Utilization and Molecular Design Breeding (MHKF202101), the China Postdoctoral Science Foundation through the Awards Nos. 2017M621546 and 2018T110411. The funding bodies did not participate in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

ZWC and PBL conceived the research. CCH, JFC, ZSZ, XFZ, and ZWC performed the experiments. JQH, XFZ, XXSG, LJW, and ZWC contributed materials and analyzed data. ZWC wrote the manuscript. All authors have read and approved the manuscript.

Acknowledgements

We thank Prof. Tian-Zhen Zhang for providing the RNA-seq expression data of the RPKM values and Prof. Xiao-Ya Chen for discussions.

References

1. Cashmore AR, Jarillo JA, Wu YJ, Liu D: **Cryptochromes: blue light receptors for plants and animals**. *Science* 1999, **284**(5415):760-765.
2. Han X, Chang X, Zhang Z, Chen H, He H, Zhong B, Deng XW: **Origin and evolution of core components responsible for monitoring light environment changes during plant terrestrialization**. *Mol Plant* 2019, **12**(6):847-862.
3. Wang Q, Lin C: **Mechanisms of cryptochrome-mediated photoresponses in plants**. *Annu Rev Plant Biol* 2020, **71**:103-129.
4. Ahmad M, Cashmore AR: **HY4 gene of A. thaliana encodes a protein with characteristics of a blue-light photoreceptor**. *Nature* 1993, **366**(6451):162-166.
5. Lin C, Ahmad M, Gordon D, Cashmore AR: **Expression of an Arabidopsis cryptochrome gene in transgenic tobacco results in hypersensitivity to blue, UV-A, and green light**. *Proc Natl Acad Sci U S A* 1995, **92**(18):8423-8427.
6. Lin C, Robertson DE, Ahmad M, Raibekas AA, Jorns MS, Dutton PL, Cashmore AR: **Association of flavin adenine dinucleotide with the Arabidopsis blue light receptor CRY1**. *Science* 1995, **269**(5226):968-970.
7. Partch CL, Clarkson MW, Ozgur S, Lee AL, Sancar A: **Role of structural plasticity in signal transduction by the cryptochrome blue-light photoreceptor**. *Biochemistry* 2005, **44**(10):3795-3805.

8. Zhang Q, Li H, Li R, Hu R, Fan C, Chen F, Wang Z, Liu X, Fu Y, Lin C: **Association of the circadian rhythmic expression of GmCRY1a with a latitudinal cline in photoperiodic flowering of soybean**. *Proc Natl Acad Sci U S A* 2008, **105**(52):21028-21033.
9. Huang Y, Baxter R, Smith BS, Partch CL, Colbert CL, Deisenhofer J: **Crystal structure of cryptochrome 3 from Arabidopsis thaliana and its implications for photolyase activity**. *Proc Natl Acad Sci U S A* 2006, **103**(47):17701-17706.
10. Lin C, Shalitin D: **Cryptochrome structure and signal transduction**. *Annu Rev Plant Biol* 2003, **54**:469-496.
11. Cao S, He S, Lv H, Zhang J, Aslam M, Cheng H, Hu A, Cao G, Zhang X, Yu Y *et al*: **Genome-wide analysis of the cryptochrome gene family in plants**. *Trop Plant Biol* 2020, **13**(1):117-126.
12. Wang Q, Zuo Z, Wang X, Liu Q, Gu L, Oka Y, Lin C: **Beyond the photocycle-how cryptochromes regulate photoresponses in plants?** *Curr Opin Plant Biol* 2018, **45**(Pt A):120-126.
13. Ma D, Li X, Guo Y, Chu J, Fang S, Yan C, Noel JP, Liu H: **Cryptochrome 1 interacts with PIF4 to regulate high temperature-mediated hypocotyl elongation in response to blue light**. *Proc Natl Acad Sci U S A* 2016, **113**(1):224-229.
14. Liu Y, Li X, Li K, Liu H, Lin C: **Multiple bHLH proteins form heterodimers to mediate CRY2-dependent regulation of flowering-time in Arabidopsis**. *PLoS Genet* 2013, **9**(10):e1003861.
15. Barrero JM, Downie AB, Xu Q, Gubler F: **A role for barley CRYPTOCHROME1 in light regulation of grain dormancy and germination**. *Plant Cell* 2014, **26**(3):1094-1104.
16. Mao Z, Wei X, Li L, Xu P, Zhang J, Wang W, Guo T, Kou S, Wang W, Miao L *et al*: **Arabidopsis Cryptochrome 1 Controls Photomorphogenesis through Regulation of H2A.Z Deposition**. *Plant Cell* 2021.
17. Liu H, Yu X, Li K, Klejnot J, Yang H, Lisiero D, Lin C: **Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in Arabidopsis**. *Science* 2008, **322**(5907):1535-1539.
18. Zuo Z, Liu H, Liu B, Liu X, Lin C: **Blue light-dependent interaction of CRY2 with SPA1 regulates COP1 activity and floral initiation in Arabidopsis**. *Curr Biol* 2011, **21**(10):841-847.
19. Meng Y, Li H, Wang Q, Liu B, Lin C: **Blue light-dependent interaction between cryptochrome2 and CIB1 regulates transcription and leaf senescence in soybean**. *Plant Cell* 2013, **25**(11):4405-4420.
20. Kang CY, Lian HL, Wang FF, Huang JR, Yang HQ: **Cryptochromes, phytochromes, and COP1 regulate light-controlled stomatal development in Arabidopsis**. *Plant Cell* 2009, **21**(9):2624-2641.
21. Mao J, Zhang YC, Sang Y, Li QH, Yang HQ: **From The Cover: A role for Arabidopsis cryptochromes and COP1 in the regulation of stomatal opening**. *Proc Natl Acad Sci U S A* 2005, **102**(34):12270-12275.
22. D'Amico-Damiao V, Carvalho RF: **Cryptochrome-related abiotic stress responses in plants**. *Front Plant Sci* 2018, **9**:1897.
23. Wu L, Yang HQ: **CRYPTOCHROME 1 is implicated in promoting R protein-mediated plant resistance to Pseudomonas syringae in Arabidopsis**. *Mol Plant* 2010, **3**(3):539-548.
24. D'Amico-Damiao V, Lucio JCB, Oliveira R, Gaion LA, Barreto RF, Carvalho RF: **Cryptochrome 1a depends on blue light fluence rate to mediate osmotic stress responses in tomato**. *J Plant Physiol* 2021, **258-259**:153374.
25. D'Amico-Damiao V, Dodd IC, Oliveira R, Lucio JCB, Rossatto DR, Carvalho RF: **Cryptochrome 1a of tomato mediates long-distance signaling of soil water deficit**. *Plant Sci* 2021, **303**:110763.
26. Hwang OJ, Back K: **Suppression of Rice Cryptochrome 1b Decreases Both Melatonin and Expression of Brassinosteroid Biosynthetic Genes Resulting in Salt Tolerance**. *Molecules* 2021, **26**(4).
27. Xu P, Xiang Y, Zhu H, Xu H, Zhang Z, Zhang C, Zhang L, Ma Z: **Wheat cryptochromes: subcellular localization and involvement in photomorphogenesis and osmotic stress responses**. *Plant Physiol* 2009, **149**(2):760-774.
28. Zhou T, Meng L, Ma Y, Liu Q, Zhang Y, Yang Z, Yang D, Bian M: **Overexpression of sweet sorghum cryptochrome 1a confers hypersensitivity to blue light, abscisic acid and salinity in Arabidopsis**. *Plant Cell Rep* 2018, **37**(2):251-264.
29. Wendel FJ, Grover C: **Taxonomy and evolution of the cotton genus, *Gossypium***, 2nd. edn: American Society of Agronomy Inc., Crop Science Society of America Inc., and Soil Science Society of America Inc., Madison; 2015.
30. Wang K, Wendel JF, Hua J: **Designations for individual genomes and chromosomes in Gossypium**. *Journal of Cotton Research* 2018, **1**(1):3.
31. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J *et al*: **Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres**. *Nature* 2012, **492**(7429):423-427.
32. Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S *et al*: **The draft genome of a diploid cotton *Gossypium raimondii***. *Nature genetics* 2012, **44**(10):1098-1103.

33. Udall JA, Long E, Hanson C, Yuan D, Ramaraj T, Conover JL, Gong L, Arick MA, Grover CE, Peterson DG *et al*: **De novo genome sequence assemblies of *Gossypium raimondii* and *Gossypium turneri***. *G3 (Bethesda)* 2019, **9**(10):3079-3085.
34. Grover CE, Arick MA, II, Thrash A, Conover JL, Sanders WS, Peterson DG, Frelichowski JE, Scheffler JA, Scheffler BE, Wendel JF: **Insights into the evolution of the New World diploid cottons (*Gossypium*, Subgenus Houzingenia) based on genome sequencing**. *Genome Biol Evol* 2018, **11**(1):53-71.
35. Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C *et al*: **Genome sequence of the cultivated cotton *Gossypium arboreum***. *Nature genetics* 2014, **46**(6):567-572.
36. Du X, Huang G, He S, Yang Z, Sun G, Ma X, Li N, Zhang X, Sun J, Liu M *et al*: **Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits**. *Nat Genet* 2018, **50**(6):796-802.
37. Huang G, Wu Z, Percy RG, Bai M, Li Y, Frelichowski JE, Hu J, Wang K, Yu JZ, Zhu Y: **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):516-524.
38. Grover CE, Pan M, Yuan D, Arick MA, Hu G, Brase L, Stelly DM, Lu Z, Schmitz RJ, Peterson DG: **The *Gossypium longicalyx* genome as a resource for cotton breeding and evolution**. 2020.
39. Cai Y, Cai X, Wang Q, Wang P, Zhang Y, Cai C, Xu Y, Wang K, Zhou Z, Wang C *et al*: **Genome sequencing of the Australian wild diploid species Gossypium australe highlights disease resistance and delayed gland morphogenesis**. *Plant Biotechnol J* 2020, **18**(3):814-828.
40. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM *et al*: **Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement**. *Nat Biotechnol* 2015, **33**(5):531-537.
41. Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, Ding M, Ye W, Kirkbride RC, Jenkins J *et al*: **Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement**. *Nat Genet* 2020, **52**(5):525-533.
42. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J *et al*: **Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution**. *Nat Biotechnol* 2015, **33**(5):524-530.
43. Liu X, Zhao B, Zheng HJ, Hu Y, Lu G, Yang CQ, Chen JD, Chen JJ, Chen DY, Zhang L *et al*: **Gossypium barbadense genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites**. *Sci Rep* 2015, **5**:14139.
44. Yuan D, Tang Z, Wang M, Gao W, Tu L, Jin X, Chen L, He Y, Zhang L, Zhu L *et al*: **The genome sequence of Sea-Island cotton (Gossypium barbadense) provides insights into the allopolyploidization and development of superior spinnable fibres**. *Sci Rep* 2015, **5**:17662.
45. Wang M, Tu L, Yuan D, Zhu, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G *et al*: **Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense**. *Nat Genet* 2019, **51**(2):224-229.
46. Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J *et al*: **Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton**. *Nat Genet* 2019, **51**(4):739-748.
47. Yang Z, Ge X, Yang Z, Qin W, Sun G, Wang Z, Li Z, Liu J, Wu J, Wang Y *et al*: **Extensive intraspecific gene order and gene structural variations in upland cotton cultivars**. *Nat Commun* 2019, **10**(1):2989.
48. Udall JA, Long E, Ramaraj T, Conover JL, Yuan D, Grover CE, Gong L, Arick MA, 2nd, Masonbrink RE, Peterson DG *et al*: **The genome sequence of *Gossypioides kirkii* illustrates a descending dysploidy in plants**. *Front Plant Sci* 2019, **10**:1541.
49. Blazquez M, Koornneef M, Putterill J: **Flowering on time: genes that regulate the floral transition. Workshop on the molecular basis of flowering time control**. *EMBO Rep* 2001, **2**(12):1078-1082.
50. Kleine T, Lockhart P, Batschauer A: **An Arabidopsis protein closely related to Synechocystis cryptochrome is targeted to organelles**. *Plant J* 2003, **35**(1):93-103.
51. Yu X, Shalitin D, Liu X, Maymon M, Klejnot J, Yang H, Lopez J, Zhao X, Bendehakkalu KT, Lin C: **Derepression of the NC80 motif is critical for the photoactivation of Arabidopsis CRY2**. *Proc Natl Acad Sci U S A* 2007, **104**(17):7289-7294.
52. Hernandez-Garcia CM, Finer JJ: **Identification and validation of promoters and cis-acting regulatory elements**. *Plant Sci* 2014, **217-218**:109-119.
53. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze 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**(1):325-327.
54. Mishra S, Khurana JP: **Emerging roles and new paradigms in signaling mechanisms of plant cryptochromes**. *Crit Rev Plant Sci* 2017, **36**(2):89-115.
55. Kleiner O, Kircher S, Harter K, Batschauer A: **Nuclear localization of the Arabidopsis blue light receptor cryptochrome 2**. *Plant J* 1999, **19**(3):289-296.
56. Yu X, Liu H, Klejnot J, Lin C: **The cryptochrome blue light receptors**. *Arabidopsis Book* 2010, **8**:e0135.
57. Kami C, Lorrain S, Hornitschek P, Fankhauser C: **Light-regulated plant growth and development**. *Curr Top Dev Biol* 2010, **91**:29-66.

58. Yang Z, Liu B, Su J, Liao J, Lin C, Oka Y: **Cryptochromes Orchestrate Transcription Regulation of Diverse Blue Light Responses in Plants**. *Photochem Photobiol* 2017, **93**(1):112-127.
59. Sharma P, Chatterjee M, Burman N, Khurana JP: **Cryptochrome 1 regulates growth and development in Brassica through alteration in the expression of genes involved in light, phytohormone and stress signalling**. *Plant Cell Environ* 2014, **37**(4):961-977.
60. Lopez L, Carbone F, Bianco L, Giuliano G, Facella P, Perrotta G: **Tomato plants overexpressing cryptochrome 2 reveal altered expression of energy and stress-related gene products in response to diurnal cues**. *Plant Cell Environ* 2012, **35**(5):994-1012.
61. Pooam M, Dixon N, Hilvert M, Misko P, Waters K, Jourdan N, Drahy S, Mills S, Engle D, Link J *et al*: **Effect of temperature on the Arabidopsis cryptochrome photocycle**. *Physiol Plant* 2021, **172**(3):1653-1661.
62. Yu J, Jung S, Cheng CH, Ficklin SP, Lee T, Zheng P, Jones D, Percy RG, Main D: **CottonGen: a genomics, genetics and breeding database for cotton research**. *Nucleic acids research* 2014, **42**(Database issue):1229-1236.
63. Wong DC, Schlechter R, Vannozzi A, Holl J, Hmmam I, Bogs J, Tornielli GB, Castellarin SD, Matus JT: **A systems-oriented analysis of the grapevine R2R3-MYB transcription factor family uncovers new insights into the regulation of stilbene accumulation**. *DNA Res* 2016.
64. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M: **Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions**. *Nucleic Acids Res* 2013, **41**(12):e121.
65. Poole RL: **The TAIR database**. *Methods in molecular biology* 2007, **406**:179-212.
66. Letunic I, Bork P: **20 years of the SMART protein domain annotation resource**. *Nucleic acids research* 2018, **46**(D1):D493-D496.
67. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al*: **The Pfam protein families database: towards a more sustainable future**. *Nucleic Acids Res* 2016, **44**(D1):D279-285.
68. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R: **TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data**. *Mol Plant* 2020, **13**(8):1194-1202.
69. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G: **GSDS 2.0: an upgraded gene feature visualization server**. *Bioinformatics* 2015, **31**(8):1296-1297.
70. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF: **Protein identification and analysis tools in the ExPASy server**. *Methods in molecular biology* 1999, **112**:531-552.
71. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability**. *Molecular Biology & Evolution* 2013, **30**(4):772.
72. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A: **DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets**. *Mol Biol Evol* 2017, **34**(12):3299-3302.
73. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: **MEGA6: Molecular Evolutionary Genetics Analysis version 6.0**. *Mol Biol Evol* 2013, **30**(12):2725-2729.
74. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements**. *Nat Methods* 2015, **12**(4):357-360.
75. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL: **Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown**. *Nat Protoc* 2016, **11**(9):1650-1667.
76. Risso D, Ngai J, Speed TP, Dudoit S: **Normalization of RNA-seq data using factor analysis of control genes or samples**. *Nat Biotechnol* 2014, **32**(9):896-902.
77. Wang M, Wang Q, Zhang B: **Evaluation and selection of reliable reference genes for gene expression under abiotic stress in cotton (*Gossypium hirsutum* L.)**. *Gene* 2013, **530**(1):44-50.
78. Zhang Y, Parmigiani G, Johnson WE: **Combat-seq: batch effect adjustment for RNA-seq count data**. *bioRxiv* 2020:2020.2001.2013.904730.
79. You Q, Xu WY, Zhang K, Zhang LW, Yi X, Yao DX, Wang CC, Zhang XY, Zhao XH, Provart NJ *et al*: **ccNET: Database of co-expression networks with functional modules for diploid and polyploid *Gossypium***. *Nucleic acids research* 2017, **45**(9):5625-5626.
80. Shan CM, Shangguan XX, Zhao B, Zhang XF, Chao LM, Yang CQ, Wang LJ, Zhu HY, Zeng YD, Guo WZ *et al*: **Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3**. *Nat Commun* 2014, **5**:5519.
81. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method**. *Methods* 2001, **25**(4):402-408.

Supplemental Table

Supplemental Table S2 is not available with this version

Figures

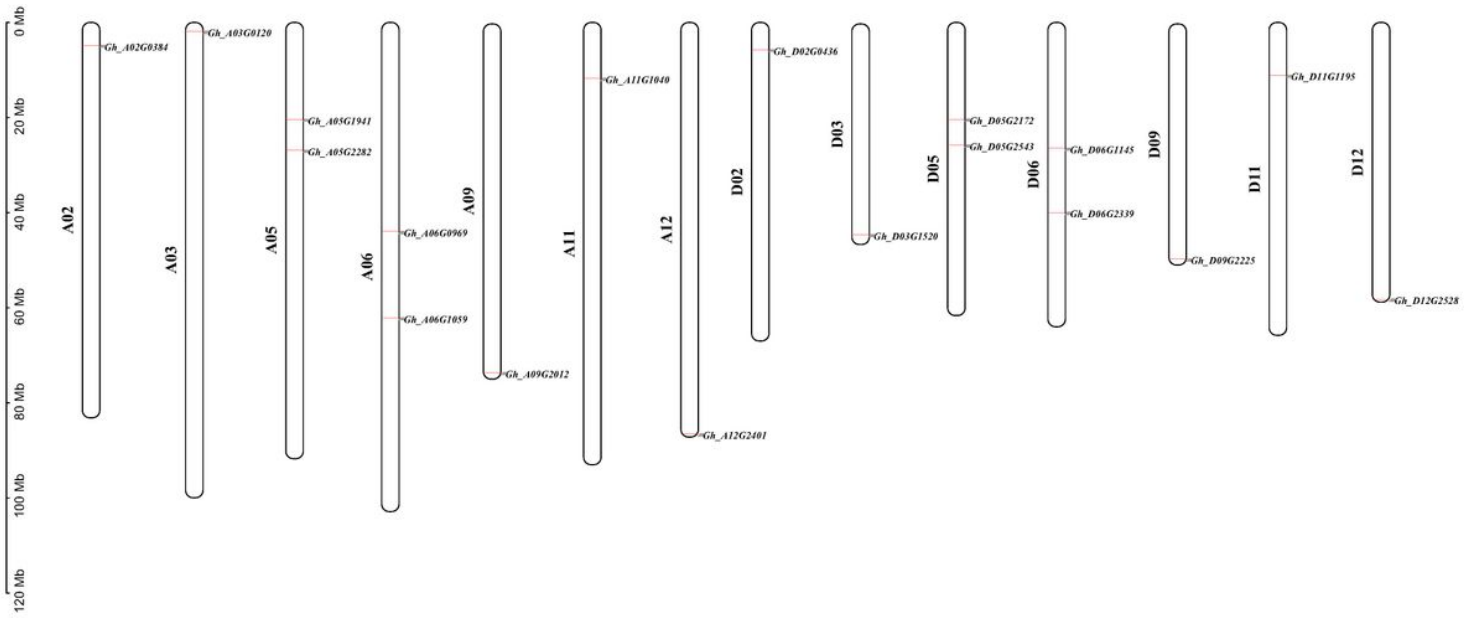

Figure 1

Dispersed distribution of CRY genes in *G. hirsutum* (AD_1) chromosomes. 18 *GhGRF* genes are scattered over 14 of the 26 *G. hirsutum* chromosomes.

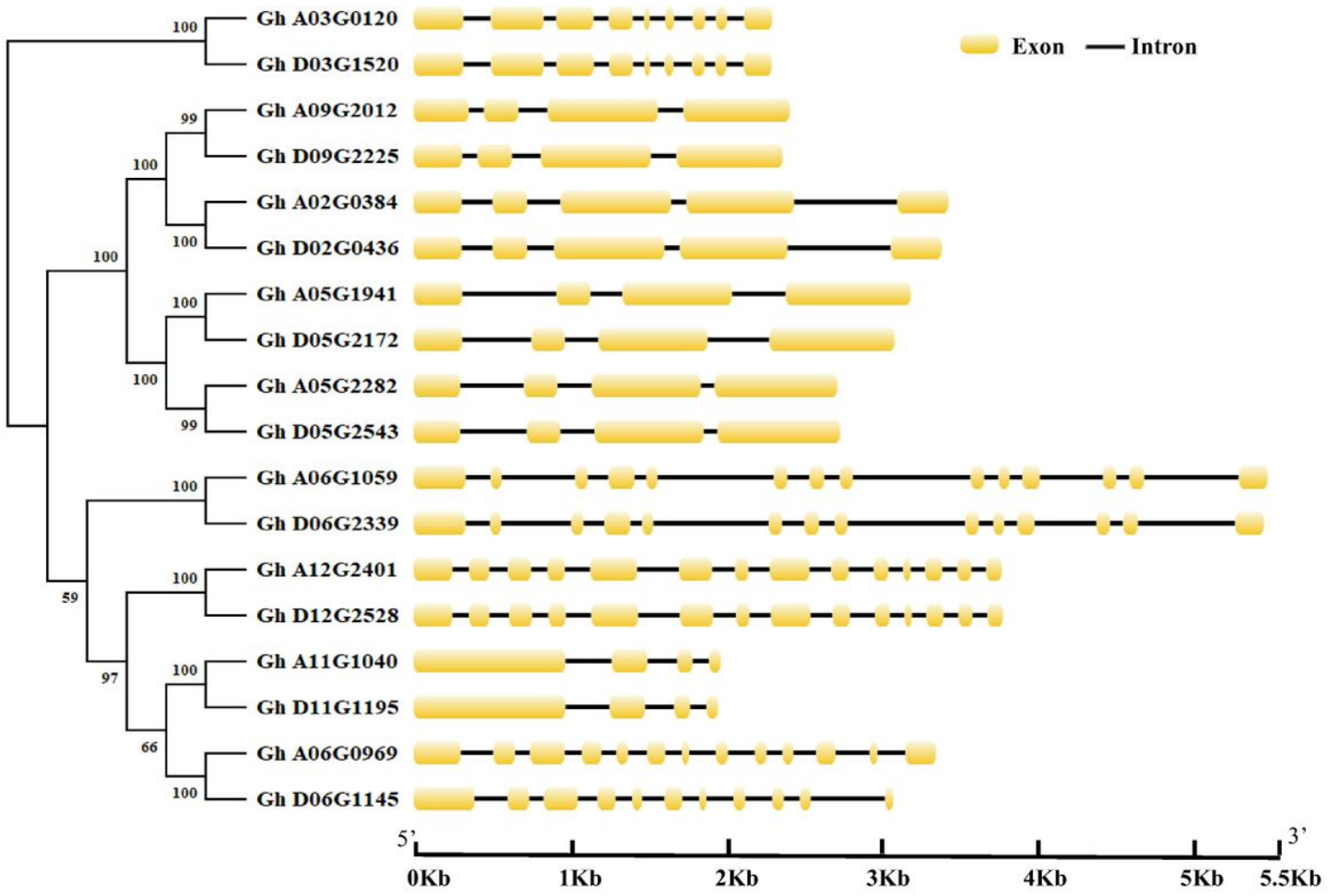

Figure 2

Phylogenetic tree and gene structure of CRY protein genes in *G. hirsutum*. Exons and introns are represented by yellow boxes and black lines, respectively.

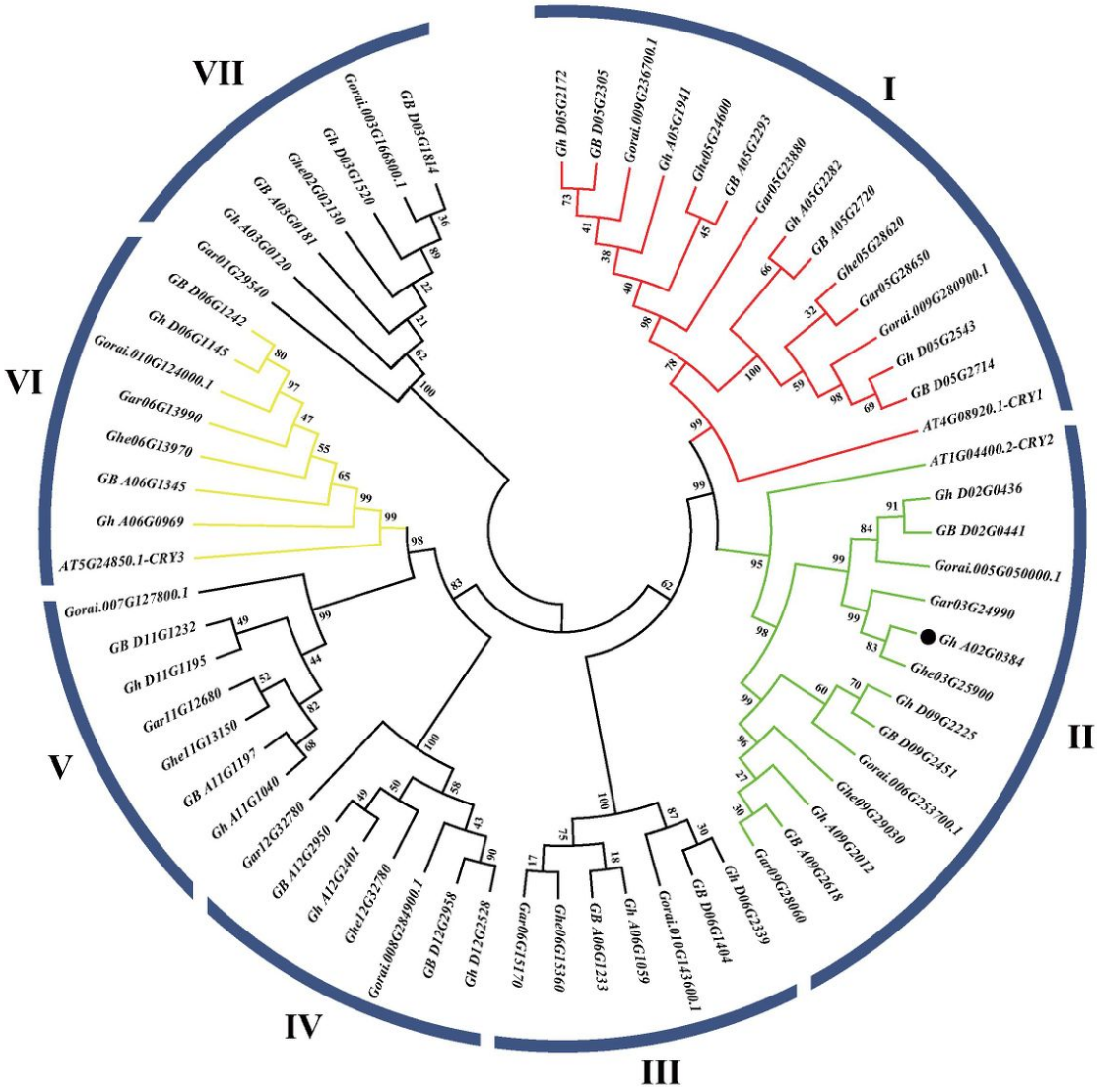

Figure 3

Phylogenetic analysis of CRY genes from five *Gossypium* species and *Arabidopsis thaliana*. The phylogenetic tree was established with entire protein-coding sequences with NJ methods. The numbers on the branches indicate bootstrap support values from 1000 replications.

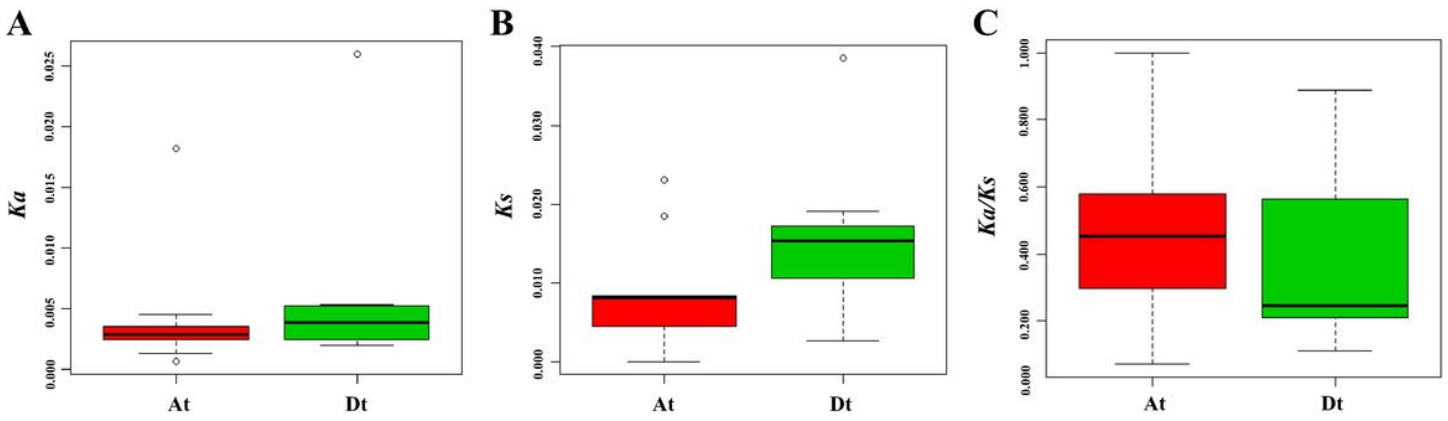

Figure 4

Distribution of Ka (A), Ks (B) and Ka/Ks (C) values of CRY genes between the A and D subgenomes versus their corresponding diploid progenitor homoeologs.

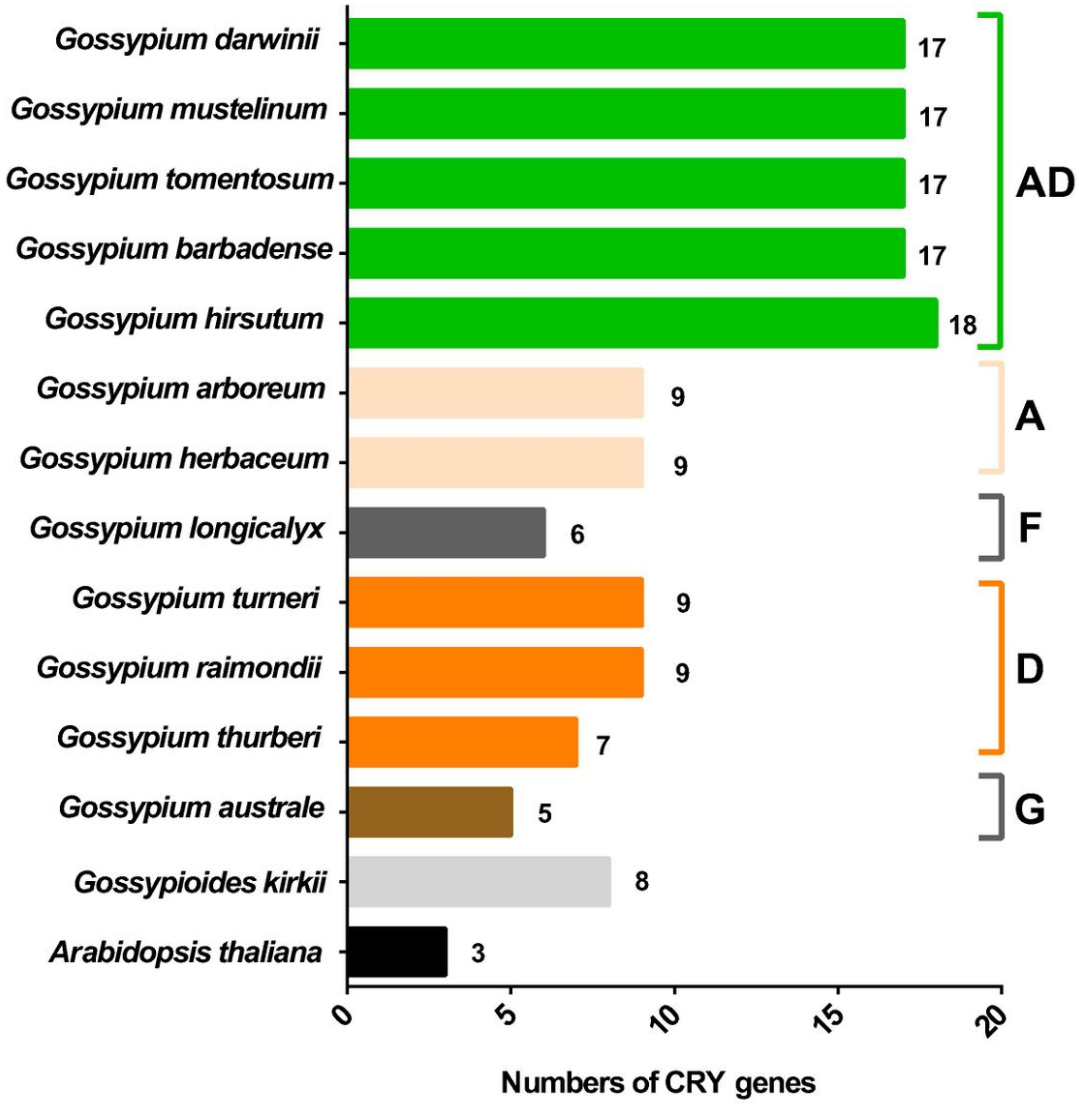

Figure 5

Dynamic evolution of the number of CRY family genes in 12 *Gossypium* species, *Gossypium kirkii* and *Arabidopsis thaliana*.

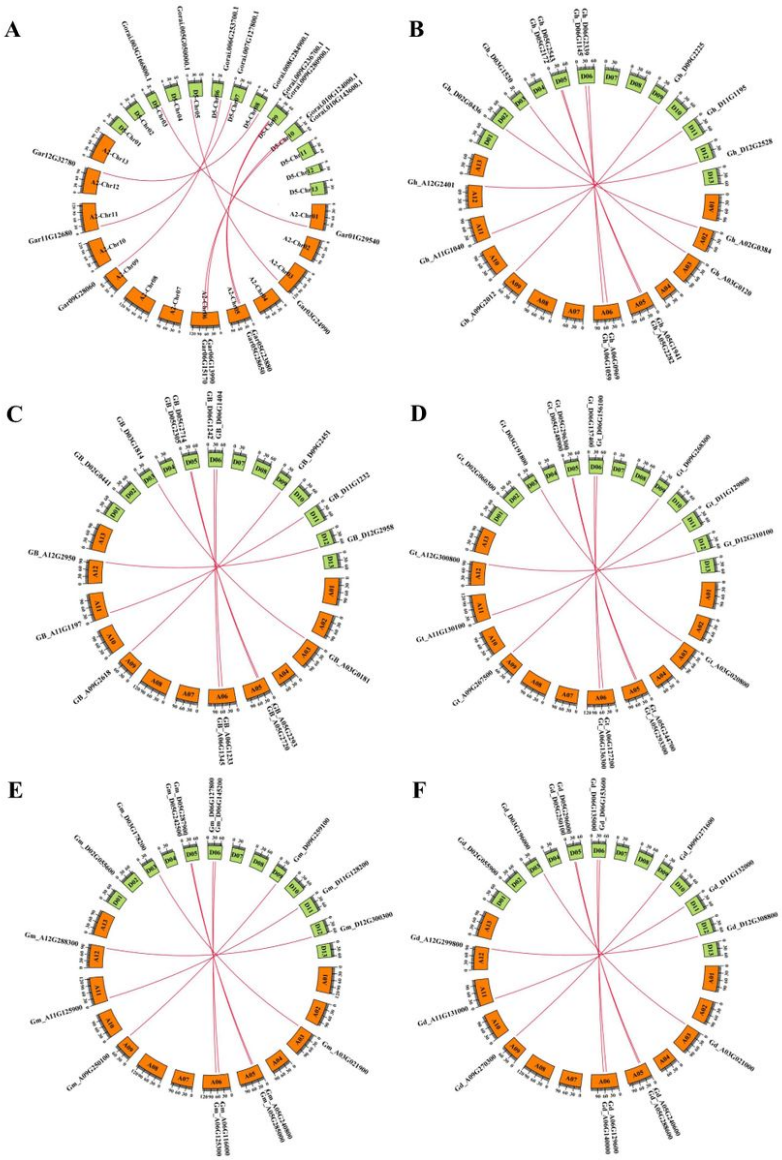

Figure 6

Syntenic analysis of the *Gossypium CRY* gene family. **A:** *G. arboreum vs G. raimondii*; **B:** *G. hirsutum*; **C:** *G. barbadense*; **D:** *G. tomentosum*; **E:** *G. mustelinum*; **F:** *G. darwinii*. The scale on the circle is in Megabases. The *CRY* gene IDs of each *Gossypium* species were on the chromosomes; the numbers of each chromosome of *Gossypium* species are shown inside the circle of each bar. The syntenic relationships of *CRY* gene are connected by red lines.

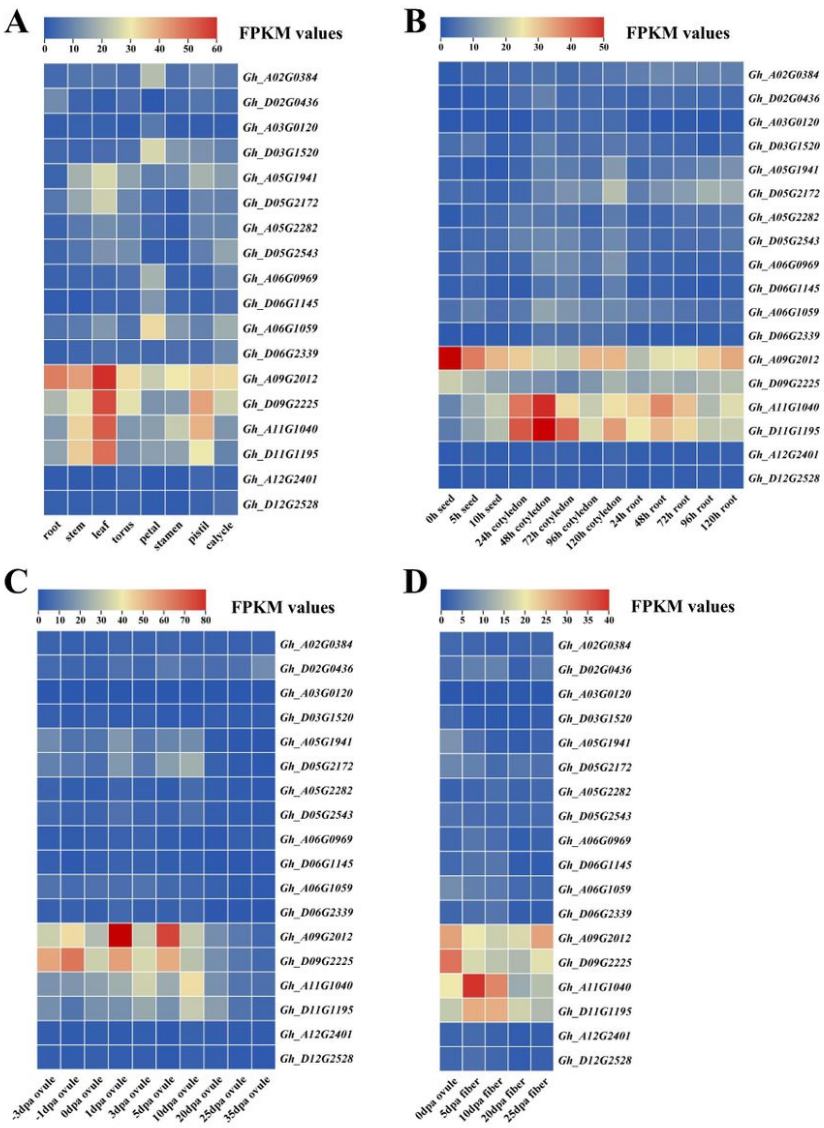

Figure 7

Expression patterns of *GhCRY* genes in different cotton tissues and fiber cells of different stages based on the RPKM values of RNA-seq data. A: Expression profiles of *GhCRY* genes in eight cotton tissues. B: Expression patterns of *GhCRY* genes in seed germination, cotyledons and roots after germination. C: Expression patterns of *GhCRY* genes in ovules of different stages. D: Expression patterns of *GhCRY* genes in fibers of different stages.

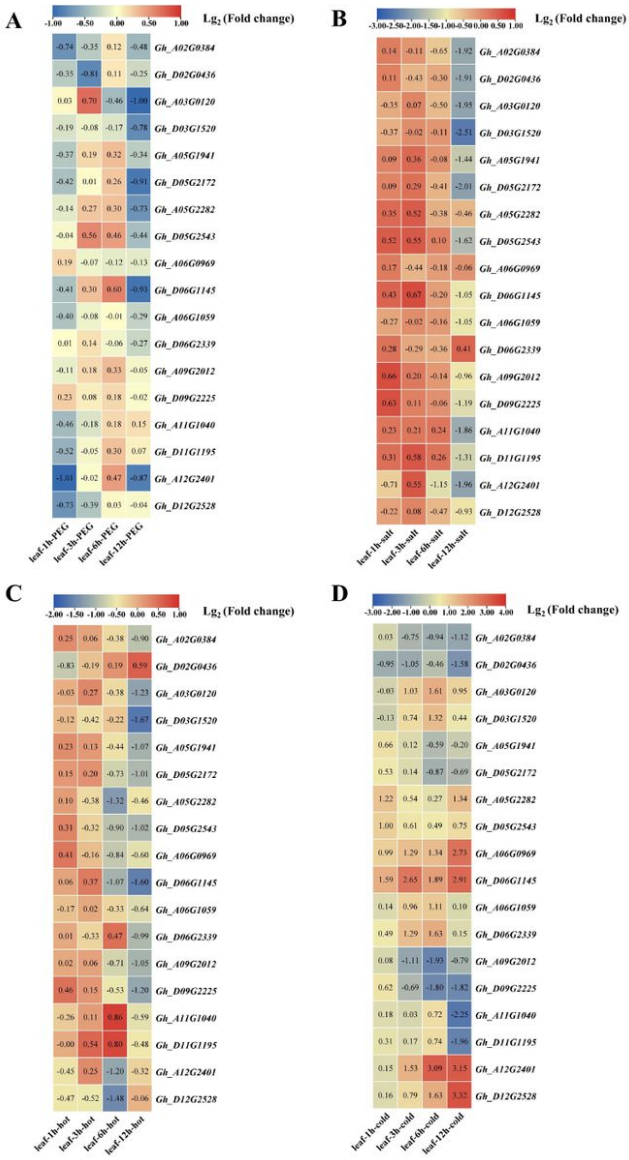

Figure 8

Expression patterns of *GhCRY* genes in response to different stresses from RNA-seq data. The RNA-seq data were downloaded from Zhang et al., 2015 and re-analyzed the RPKM values of five time points (0, 1, 3, 6 and 12 h) after stresses treatments. A: Drought stress; B: Salt stress; C: Hot stress; D: Cold stress.

Supplementary Files

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

- Additionalfile1TableS1.docx
- Additionalfile2FigureS1.pdf