

Genome-wide analysis of CrRLK1L genes and their expression profiling during fiber development in cotton

Dongyun Zuo

Chinese Academy of Agricultural Sciences Cotton Research Institute

Javaria Ashraf

Chinese Academy of Agricultural Sciences Cotton Research Institute

Hailiang Cheng

Chinese Academy of Agricultural Sciences Cotton Research Institute

Shang Liu

Chinese Academy of Agricultural Sciences Cotton Research Institute

Youping Zhang

Chinese Academy of Agricultural Sciences Cotton Research Institute

Qiaolian Wang

Chinese Academy of Agricultural Sciences Cotton Research Institute

Limin Lv

Chinese Academy of Agricultural Sciences Cotton Research Institute

Feng Xiaoxu

Chinese Academy of Agricultural Sciences Cotton Research Institute

Linlin Song

Chinese Academy of Agricultural Sciences Cotton Research Institute

Yuanyuan Liu

Chinese Academy of Agricultural Sciences Cotton Research Institute

Guoli Song (✉ sglzms@163.com)

<https://orcid.org/0000-0003-3236-9286>

Research article

Keywords: CrRLK1L, phylogenetic analysis, collinearity, gene duplication, fiber development, transcription factors

Posted Date: July 29th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-44129/v1>

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

[Read Full License](#)

Abstract

Background: *Catharanthus roseus* receptor-like kinase 1-like (*CrRLK1Ls*) proteins play important roles in cell growth, plant morphogenesis, reproduction, hormone signaling, plant immunity and stress responses in *Arabidopsis*. However, not much information is available about their functions during cotton fiber development.

Results: We identified a total of 125, 73 and 71 full-length putative *CrRLK1L* genes in *G. hirsutum*, *G. arboreum* and *G. raimondii*, which are much greater than that of the other plants. The phylogenetic and gene structure analysis divided the cotton *CrRLK1L* genes into six major groups, among which only group I and II contained *AtCrRLK1Ls* of *Arabidopsis*, suggesting that other groups (group III-VI) were expanded by gene duplication during cotton evolution. Genome collinearity analysis revealed that half of the At02 genes in *G. hirsutum* derived from A02 of *G. arboreum*, while the other half (*GhCrRLK1L6* and *GhCrRLK1L7*) originated from Dt03 and Dt02 of *G. raimondii*, indicating segmental duplication between noncorresponding chromosomes during polyploidization of *G. hirsutum*. In addition, expression and cis-element analysis revealed that only 22 *GhCrRLK1Ls* showed specific expression pattern during fiber development which are mainly due to the presence of binding sites for NAC, MYB and WRKY transcription factors.

Conclusions: This study provides a strong foundation to further explore the molecular mechanism of *CrRLK1L* genes during fiber development in upland cotton.

Background

Catharanthus roseus receptor-like kinase 1-like (*CrRLK1L*) protein is a subgroup of Receptor-like kinases (RLKs), containing an extracellular region with a lectin domain which binds to dimeric and oligomeric glucose [1]. These *CrRLK1L* proteins play versatile roles in cell growth, plant morphogenesis, reproduction, immunity, hormone signaling and stress responses [2].

In *Arabidopsis*, the *CrRLK1L* family contained a total of 17 members among which ten members have been functionally characterized. *THE1* was discovered as suppressor of the cellulose synthase deficient mutant *procuste1*, partially rescuing its short hypocotyl when grown in the dark which indicated its role to repress the growth of the damaged cells [3]. FER is another typical lectin domain containing *CrRLK1L* which regulates plant root development as its suppression caused the semi-dwarf plants and abnormal growth in root hair and trichomes [4]. Additionally, FER also regulates the communications between pollen tubes and female gametophytes [5]. In *fer* mutant, the pollen tube fails to rupture and continues to grow which results a nonfertilized female gametophyte. Another two members of *CrRLK1Ls* including *ANX1* and *ANX2* seems to be redundant in plant reproduction, as double mutant (*anx1anx2*) plant is sterile due to precocious rupturing of pollen tubes [6, 7]. In addition, BUPS1 and BUPS2 are the partially redundant *CrRLK1L* proteins which interact with *ANX1* and *ANX2* to regulate the pollen tube growth in plants [8]. Further, *ANX1*, *ANX2* and FER are also involved in plant immunity responses. Such as, FER

promotes the interaction between FLS2 or EFR and BAK1 by diminishing the FLS2/EFR-BAK1 complex formation which increased the PTI responses in plants [9]. Oppositely, ANX1 and ANX2 appear to be a negative regulator of plant immunity responses [10]. Further, EUR is a positive regulator of cell growth and is important for cytoplasmic NH⁴⁺ homeostasis [11]. Therefore, mutation of *eru* exhibits a short root hair phenotype that can be restored by growing plants in NH⁴⁺-depleted medium.

Cotton is an important economic crop that provides fiber for textile industry. Cotton fiber is a single cell differentiated from the epidermal ovule, and its development consists of four overlapping stages including fiber initiation, elongation, second cell wall biosynthesis and maturation[12], which are regulated by many genes[13]. For instance, *GbMYB2* was specifically expressed in the epidermal layer of ovule and fibers, and overexpression of *GbMYB2* in *Arabidopsis* displayed enhanced trichomes growth on leaves and longer roots [14]. Likewise, *GhMYB25* was specifically expressed in cotton fiber and its RNA interference resulted shorter fiber and reduction in seeds yield, while overexpressing plants showed increased number of trichomes and fiber initiation [15]. Similarly, *GhMYB25-like* (*GhMML3_A12*) gene is localized on A12 chromosome of *G. hirsutum* having active promoter at the 3' end which forms natural antisense transcripts (NATs). SiRNA derived from NATs participated in posttranscription gene silencing of *GhMML3_A12*, which leads to the fuzzless phenotype in N1 mutant [16, 17]. Besides these, silencing of *GhMYB212* or *GhSWEET12* in cotton results in the accumulation of sucrose and glucose in fiber, short fiber length and decreased lint percentage [18].

Although many genes belong to different transcription factors families have been functionally characterized during fiber development, but the functions of *CrRLK1Ls* is unknown. Here, we performed a genome-wide analysis of *CrRLK1Ls* genes in cotton to comprehensively explore their functions in fiber development. We found that some *GhCrRLK1Ls* regulated by NAC, MYB and WRKY transcription factors were played important roles in fiber development. Our results elaborated the origin of *GhCrRLK1Ls* and their regulation network in fiber development which is important to further explore their molecular mechanism in fiber development.

Results

Genome-wide identification of *CrRLK1Ls* in *G. raimondii*, *G. arboreum*, *G. hirsutum*

In this study, we identified 125, 73, and 71 full-length putative *CrRLK1L* genes in *G. hirsutum* (TM-1), *G. arboreum* (Shixiya 1) and *G. raimondii* respectively. According to their chromosomal locations, the identified *CrRLK1Ls* members of the three cotton species were designated as *GhCrRLK1L1* to *GhCrRLK1L125* in *G. hirsutum*, *GaCrRLK1L1* to *GaCrRLK1L73* in *G. arboreum*, and *GrCrRLK1L1* to *GrCrRLK1L72* in *G. raimondii*. The protein length of *GhCrRLK1L* genes ranged from 402 (*GhCrRLK1L111*, localized on Ghir_D11 chromosome) to 1360 (*GhCrRLK1L81*, localized on Ghir_D07 chromosome) amino acids (aa), while *GaCrRLK1Ls* varied from 476 aa (*GaCrRLK1L28*) to 1688 aa (*GaCrRLK1L1*), and *GrCrRLK1Ls* ranged from 581 aa (*GrCrRLK1L34*) to 1061 aa (*GrCrRLK1L60*). The detailed information of

CrRLK1Ls in different *Gossypium species* and their predicted subcellular location and signal peptide were shown in Table S2, S3 and S4.

Phylogenetic analysis of the *CrRLK1L* gene family

To examine the evolutionary relationships among *CrRLK1L* proteins, all 125 *GhCrRLK1L* protein sequences were used to construct phylogenetic trees using neighbor joining method, which contained six clades (I-VI) (Fig.1A). Group I and II were the largest which contained 40 and 37 *CrRLK1L* members, while groups IV and V and VI contained 13, 27 and 7 *CrRLK1L* members respectively. Interestingly, subfamily III contained only one member *GhCrRLK1L107* which was localized on Ghir_D11 chromosome and contained a protein of 518 aa with one malectin and phosphokinase domain. It was localized on the plasma membrane without a transmembrane domain, having a glycosyl-phosphatidyl inositol (GPI) anchor site in the 482 position of its protein sequence. We extracted the extracellular malectin domain of *GhCrRLK1L* genes sequences to generate a separate phylogenetic tree (Fig.1B) as it played important roles in their biology functions. These all 125 extracellular domain sequences were also divided into six groups, showing minor differences compared with the tree constructed by the full length protein sequences. These results indicated that the extracellular domain in *CrRLK1Ls* protein determined their biological functions.

Arabidopsis contained 17 *CrRLK1Ls* (*AtCrRLK1Ls*) members which played important roles in cell growth regulation, reproduction, stress responses. So, these 17 *AtCrRLK1Ls* protein sequences of *Arabidopsis* were downloaded from Ensembl Plants websites (<http://plants.ensembl.org>), and were used to construct phylogenetic tree (Fig.2) combined with the all *GhCrRLK1Ls* members, which were also divided into six groups (I-VI). However, all the 17 *AtCrRLK1Ls* were distributed in subfamily I and II which were further classified into six subgroup (I-a, I-b, I-c, II-a, II-b and II-c). Among them, 17 *AtCrRLK1Ls* were distributed in all subgroups except subgroup I-c, which is specific for cotton, containing only two *GhCrRLK1Ls* including *GhCrRLK1L21* and *GhCrRLK1L62*. Besides, subfamilies III, IV, V, and VI contained 48 *GhCrRLK1Ls* which were also specific for cotton. This data suggested that 50 *GhCrRLK1Ls* are specific for cotton, which might evolve in different directions and expanded to exhibit diverse functions among various cotton species.

Gene structure and conserved protein motifs analysis

During the evolution of gene families, the diversification of gene structure is responsible for evolving gene new function to acclimatize in the changing environment. Therefore, we explored the structural diversity of *GhCrRLK1L* genes, their exon-intron structure and conserved domains (Fig.3B-C). The *GhCrRLK1L* genes of group I and II contained 1 to 3 exons, except for *GhCrRLK1L124* which comprised of 7 exons. Most gene members in group II contained exons varying 1 to 4, except *GhCrRLK1L149*, *GhCrRLK1L58*, *GhCrRLK1L22* and *GhCrRLK1L81*. Specially, *GhCrRLK1L22* and *GhCrRLK1L81* comprised of 21 and 20 exons, respectively. The groups III to VI contained much more exons as compared to groups I and II, the number of exons in these subgroups varied 13 to 25, while *GhCrRLK1L111* in group V contained the least exon number (7 exons). The exon numbers were highly diverse among *GhCrRLK1Ls*, might indicating

their various functions. Based on the exon number, these six groups of *GhCrRLK1Ls* could be divided into two major groups, among which group I and II belong to first major group, while the rest of four groups belong to the second major group. As mentioned above, 17 *AtCrRLK1Ls* of *Arabidopsis* distributed in group I and II, belong to the first major group, while all members in groups III to VI were from *G. hirsutum*. These results illustrated that cotton evolved diverse exon number which may entitle with new functions for *GhCrRLK1L* family genes.

Furthermore, the protein domain analysis was performed to identify the conserved motifs present in *GhCrRLK1Ls*. All the *GhCrRLK1Ls* localized to the cytomembrane, so transmembrane helix was detected among all the members. Results showed that 8 out of the 125 *GhCrRLK1Ls* have no transmembrane helix domain, however, a GPI anchor site was discovered in each sequence of the 8 *GhCrRLK1Ls*. These result indicated that *GhCrRLK1Ls* localized in the plasma membrane by two methods. Signal peptide detection by the SignalP 5.0 [19] revealed that 89 out of the 125 *GhCrRLK1Ls* had signal peptides at their N-teiminal ends, and all the signal peptide belong to the secreted signal peptide I (SPI). These results indicated that these proteins may be secreted to the cytomembrane which was consistent with the results of their subcellular location. Furthermore, all the *GhCrRLK1L* proteins contained a phosphokinase domain and one or two malectin domains (Fig.3B), except for *GhCrRLK1L48*, *GhCrRLK1L54* and *GhCrRLK1L128* with a malectin-like domain instead of malectin domain. As the malectin domain is vital for *GhCrRLK1Ls* functions, we also analyzed the putative motifs in the extracellular domain using MEME programme which identified five conserved motifs (motif1 to motif5) (Fig.S1). These motifs were regularly distributed in the 125 *GhCrRLK1Ls*. The members of group I and II contained motif2, motif4 and motif5, while motif1, motif2 and motif3 were presented in group III, IV and V, except for *GhCrRLK1L109* and *GhCrRLK1L111* in group V which contained only motif2 and motif3. Six members of the group VI showed irregular distribution of motifs. Generally, motif1 and motif3 were present only in group III to V, and motif4 and motif5 were specific for group I and II. Collectively, these results suggested that *GhCrRLK1Ls* possessed similar gene structures, and motifs were clustered in the same group and might have similar functions during the evolution of *G. hirsutum*.

Distribution and gene duplication events of *CrRLK1Ls* in *Gossypiums*

The TBtools [20] was used to map the physical position of *CrRLK1Ls* (Fig.4). In *G. hirsutum*, 124 *GhCrRLK1Ls* were localized in 26 chromosomes (Fig.4A), while one gene (*GhCrRLK1L62*) was present on scaffold. The distribution of *GhCrRLK1Ls* on chromosomes was highly uneven *i.e.* At07 and Dt07 contained high number (13 and 11) of the *GhCrRLK1Ls* family members respectively. In contrast, A04 and D04 contained only 1 gene. In *G. arboreum*, 70 *GaCrRLK1Ls* were localized on 12 chromosomes (Fig.4B), with the exception of A02, while 3 genes (*GaCrRLK1L71- GaCrRLK1L73*) were localized on scaffolds. A07 of *G. arboreum* contained most (13) of the *GaCrRLK1Ls*, followed by A11 with 11 *GaCrRLK1Ls*. The gene numbers contained by the other chromosomes were varied from 2 to 7. A total of 70 *GrCrRLK1Ls* were mapped to the 13 chromosomes with one gene localized on scaffold. D07 and D01 of *G. raimondii* contained 13 and 11 *GrCrRLK1Ls* genes respectively (Fig.4C). The *CrRLK1L* gene numbers distributed on each chromosome of *Gossypiums* was showed in Table S5.

Fragment duplications in the genome region are vital for gene family expansion and occur along with plant genome evolution. In the present study, a gene duplication analysis was performed to investigate the expansion mechanism of the *CrRLK1L* gene family in the three *Gossypium* species. Previous studies have shown tandem and segmental duplications were two major events in gene duplication. Based on the open reading frame (ORF) sequence of all genes in each species, MCScan software [21] was used to detect the collinearity relationship between *CrRLK1Ls*. A total of 83, 8 and 11 gene-pairs with segmental duplication were discovered in *G. hirsutum*, *G. arboreum* and *G. raimondii*, respectively (Fig.5A, Fig.S2, Table S6, S8 and S10). As well, we identified 13, 9 and 6 tandem duplicated gene-pairs in corresponding species (Table S7, S9 and S11). While, there were 5 *GhCrRLK1Ls* tandem clustered localized on At07 from *GhCrRLK1L29* to *GhCrRLK1L33*, and from *GhCrRLK1L108* to *GhCrRLK1L111*, 4 *GhCrRLK1Ls* tandem clustered localized on Dt11 (Table S7). All these results indicated that both segmental and tandem duplication played important roles in the expansion of *CrRLK1L* family proteins in *G. arboreum* and *G. raimondii*, while segmental duplication was predominantly present in *G. hirsutum*. In addition, the relationship between two diploid cottons and *G. hirsutum* was discovered 57 *CrRLK1L* protein pairs between *G. raimondii* and *G. hirsutum* (Fig.5B, Table S13), and 26 homologous gene-pairs were found between *G. arboreum* and *G. hirsutum* (Table S12). These results suggested that most of the *CrRLK1L* family proteins in tetraploid *G. hirsutum* were derived from its ancestors diploid cottons, and most (57/72) of *CrRLK1Ls* in *G. raimondii* were inherited to the tetraploid cotton, while only 35.6% (26/73) *CrRLK1Ls* of *G. arboreum* inherited to *G. hirsutum*.

***GhCrRLK1L6* and *GhCrRLK1L7* genes derived from noncorresponding chromosomes**

It was reported that *G. raimondii* and *G. arboreum* were the ancestors of *G. hirsutum* [22-25]. According to the *CrRLK1Ls* distribution within the chromosomes of three *Gossypium* species, we found that there were two *GhCrRLK1Ls* including *GhCrRLK1L6* and *GhCrRLK1L7* localized on At02, while there was no *GhCrRLK1Ls* on the ancestor corresponding chromosome Ga-A02. In order to find out the origin of *GhCrRLK1L6* and *GhCrRLK1L7*, the collinearity of these two genes was analyzed. *GhCrRLK1L70* localized on Dt02 was found to have collinearity with *GhCrRLK1L6*. However, four genes such as *GhCrRLK1L49*, *GhCrRLK1L73*, *GhCrRLK1L105* and *GhCrRLK1L115* were identified to have collinear relationship with *GhCrRLK1L7*. Further, multiple sequence alignment was performed and phylogenetic tree was constructed to find their cognate relationship (Fig.S3). *GhCrRLK1L7* was found to be most homologous with *GhCrRLK1L73* which was localized on Dt03, speculated its duplication from *GhCrRLK1L73*. Besides, localized site of the *GhCrRLK1L7* and *GhCrRLK1L73* on chromosome were identified to have segmental duplication, which is similar with localized regions of *GhCrRLK1L6* and *GhCrRLK1L70* (Fig.6). These results indicated that *GhCrRLK1L6* of At02 chromosome duplicated from *GhCrRLK1L70* of Dt02 chromosome, and *GhCrRLK1L7* of At02 duplicated from *GhCrRLK1L73* of Dt03. Further, half of the At02 of tetraploid *G. hirsutum* originated from A02 of diploid *G. arboreum*, while other half originated from the other chromosomes Dt03 and Dt02.

Expression patterns of *GhCrRLK1Ls* in fibers

To explore the possible biological functions of *GhCrRLK1Ls* during cotton fiber development, the expression patterns of *GhCrRLK1Ls* in ovules and fiber at different fiber developmental stages were investigated. The ovules of *G. hirsutum* (TM-1) at different development, -3 days post anthesis (-3 DPA), -1, 0, 1, 3, 5, 7, 10, 15, 20 and 30 DPA were used for transcriptome analysis (unpublished). A total of 68 *GhCrRLK1Ls* were found with FPKM > 1 in at least one of the 11 development stages. Based on the previous reported method [26], 46 *GhCrRLK1Ls* were discovered to be expressed during all 11 stages without any fluctuate change (Fig.S4). However, 22 *GhCrRLK1Ls* were identified to express at specific fiber development stages. According to their expression patterns, 22 genes were clustered into three groups (A-C) (Fig.7). The group A contained only two members *GhCrRLK1L26* and *GhCrRLK1L84* belonging to group VI, which preferred to express at 1 DPA to 15 DPA. There were 11 *GhCrRLK1Ls* in cluster B, which tended to express in 15 DPA to 30 DPA. For instance, *GhCrRLK1L119* was specifically expressed at 15 DPA with almost no expression at the other stages, while *GhCrRLK1L52* was predominantly expressed only at 20 DPA. The 9 *GhCrRLK1Ls* in cluster C were inclined to express at the early ovules development stages (-3 DPA to 5 DPA), such as *GhCrRLK1L17* exhibited higher expression at -1 DPA. The *GhCrRLK1Ls* expression pattern suggested that genes in cluster A might play roles during cotton fiber initiation and elongation, while genes in cluster B might be involved in fiber secondary cell wall synthesis, and the cluster C genes were responsible for fiber initiation.

NAC, MYB and WRKY transcription factors might involve in regulating the specific expression of *GhCrRLK1Ls* during fiber development

To illuminate the regulation network of the 22 specific expressed *GhCrRLK1Ls*, the 1000 bps upstream regions from the transcript start site (TSS) of the 22 specific expressed and 46 nonspecific expressed *GhCrRLK1Ls* were extracted and analyzed by homer software. The promoters of the 46 nonspecific expressed *GhCrRLK1Ls* were used as a background, and 9 conserved motifs were identified in most of the target genes promoters and not in the background sequences. Among the 9 conserved motifs, 6 motifs were predicted to be NAC transcription factor binding sites, while 2 motifs were the MYB transcription factor binding sites and 1 motif was WRKY transcription factor binding site (Table S14). These results suggested that NAC, MYB and WRKY transcription factors might bind to the promoters of *GhCrRLK1Ls*, leading to their specific expression during fiber development. To further clarify the three transcription factors functions in regulating expression of *GhCrRLK1Ls*, we selected 2 of each transcription factor in the transcriptome data, and qRT-PCR was performed to identify the expression of the six transcription factors (*Ghir_D12G017660.1*, *Ghir_A13G007770.1*, *Ghir_A11G026140.1*, *Ghir_D01G006060.1*, *Ghir_A06G010870.1*, *Ghir_A03G015780.1*). All of the six transcription factors preferred to express at specific fiber development stage (Fig.8), which were coincident with the 22 *GhCrRLK1Ls*. This data indicated that *GhCrRLK1Ls* played important roles during fiber development and their functions are generally regulated by NAC, MYB and WRKY transcription factors.

Discussion

CrRLK1L genes were discovered in many plant species such as there were 23 and 26 *CrRLK1Ls* were identified in *Solanum lycopersicum*, and *Pyrus bretschneideri* respectively [27, 28], which was similar with the numbers of *CrRLK1Ls* in *Arabidopsis*. However, in the present study, we identified much more *CrRLK1Ls* in three *Gossypium* genomes *i.e.* 71, 73 and 125 *CrRLK1Ls* were identified in *G. raimondii*, *G. arboreum* and *G. hirsutum* respectively. These numbers were much greater than those reported in other species, indicating gene expansion occurred in the *CrRLK1L* gene family during the evolution of cotton genomes. Our research revealed that a large number of gene duplications played an important role in *CrRLK1L* gene family expansion. For example, a total of 83 and 13 segmental and tandem duplications were detected in *G. hirsutum*.

Phylogenetic analysis divided the *GhCrRLK1Ls* into six groups with the largest numbers in groups I and II. As well, all of the 17 *AtCrRLK1Ls* of *Arabidopsis* were localized in group I and II, indicating that group I and II were conserved among plants, and cotton have evolved its new *CrRLK1Ls* (group III, IV, V and VI) to adapt into the changing environment. Besides, conserved motif analysis revealed that motif4 and motif5 were specifically present in the group I and II and not in the other groups, while motif1 and motif3 were only conserved in the group III, IV and V. This data indicated that motif4 and motif5 might be conserved in plants among species, while motif1 and motif3 might be particular for *Gossypium* species, owning unique functions distinguished with the other species. All of the *GhCrRLK1Ls* exhibited the conserved phosphokinase domain, contained a serine and threonine phosphorylation site, which could phosphorylate different substrates triggering different cellular responses. Most of the *GhCrRLK1Ls* contained one or two extracellular malectin domains except for *GhCrRLK1L48*, *GhCrRLK1L54* and *GhCrRLK1L128*, with a malectin-like domain instead of malectin domain. In *Arabidopsis*, different *CrRLK1Ls* could sense various endogenous and exogenous ligands converged on MAPK cascades in plant growth, reproduction and defense [2, 29]. Similarly, different *GhCrRLK1Ls* owning various extracellular domains could sense a great diversity of stimuli, leading to the activation of intercellular phosphokinase domain, which could phosphorylate different substrates and trigger different biological responses. In this way, the messages from the environments or the neighbor cells perceived by the *GhCrRLK1Ls* and transmitted into cell, the cell initiated a series of biological processes to response the change outside the cell. The diverse functions of *GhCrRLK1Ls* might be determined by the extracellular malectin domain, which was consistent with the result that there was a minor difference in the *GhCrRLK1Ls* classification between full length protein sequences and the extracellular domain sequences.

It was reported that *G. raimondii* and *G. arboreum* were the ancestors of *G. hirsutum* [22-25], and polyploidization of diploid *G. arboreum* (A) and *G. raimondii* (D) formed the tetraploid *G. hirsutum* approximately 1-2 MYA [23]. However we found that there were two *GhCrRLK1Ls* (*GhCrRLK1L6* and *GhCrRLK1L7*) of *G. hirsutum* localized on At02 had no ancestral gene on corresponding chromosome Ga-A02. Further, based on the homology and segmental duplication analysis, we speculated the origin of *GhCrRLK1L6* and *GhCrRLK1L7*, *GhCrRLK1L6* from *GhCrRLK1L70* of Dt02, and *GhCrRLK1L7* duplicated from *GhCrRLK1L73* of Dt03. Genome collinearity analysis revealed that half of the At02 in *G. hirsutum* derived from A02 of *G. arboreum*, while the other half originated from Dt03 and Dt02. Our research

illuminated that segmental duplication between noncorresponding chromosomes occurred during polyploidization of tetraploid *G. hirsutum*, which was first discovered in upland cotton.

The functions of *CrRLK1Ls* had been clearly illuminated in *Arabidopsis*, most of these genes were involved in plant growth, root development, reproduction, stress responses [2, 29, 31]. In present study, the fiber development transcriptomic analysis of *GhCrRLK1Ls* revealed that 54% of the members expressed in ovules and fiber tissues, and 22 members owned fiber development stage specific expression pattern. For example, *GhCrRLK1L17* expressed only at the fiber initiation and *GhCrRLK1L119* expressed only at the fiber elongation. This data suggested that most of the *CrRLK1Ls* in cotton might play important roles in the other biological processes, and 22 of the members might involve in fiber development. Furthermore, analysis of the cis-elements in the promoter of 22 specifically expressed genes revealed that NAC, MYB and WRKY transcription factors binding sites were localized in all the promoters. This result indicted that MYB, NAC and WRKY transcription factors might regulate the expression of *GhCrRLK1Ls* during fiber development, which was consistent with the previous reports that MYB transcription factors such as *GhMYB109*, *GhMYB25*, *GhMYB25*-like and *GhMYB212* play vital roles in fiber development [15, 16, 18, 32]. The functions of NAC and WRKY transcription factors in fiber development need further research.

Conclusions

In summary, we identified a total of 125, 73 and 71 full-length putative *CrRLK1L* genes in *G. hirsutum*, *G. arboreum* and *G. raimondii*, which were much greater than that of the other plants, resulting from segmental duplications and tandem replications. We also found that gene duplications occurred between noncorresponding chromosomes, which is first discovered in cotton. Besides, we illuminated that *GhCrRLK1Ls*, regulated by NAC, MYB and WRKY transcription factors, plays important roles in fiber development. Our results provide a strong foundation to further explore the molecular mechanism of *CrRLK1L* genes during fiber development in upland cotton.

Methods

Plant materials

Cotton plants (*Gossypium hirsutum* cv TM-1) were grown in the experimental field at the Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang city, Henan province, China. Flowers were tagged on the day of anthesis (0 days post anthesis, DPA), and bolls were harvested at -3, -1, 0, 1, 3, 5, 7, 10, 15, 20 and 30 DPA. Cotton ovules were carefully separated from the bolls and dipped into liquid nitrogen immediately. Fiber was gently detached from ovules at 7, 10, 15, 20 and 30 DPA, and immediately dipped into liquid nitrogen.

Identification of *CrRLK1Ls* in cotton

Genomes data of *G. raimondii* (JGI_v2.1) [33], *G. arboreum* (CRI_v2.0) [34] and *G. hirsutum* acc. TM-1 (HAU_v1) [24] were obtained from CottonGen (<https://www.cottongen.org/>) [35]. The Hidden Markov

Model (HMM) profile of malectin (PF11721) was downloaded from Pfam (<https://pfam.xfam.org>) [36], and used as a query to identify the candidates *CrRLK1Ls* from the cotton genome protein database using HMMER. All candidates *CrRLK1Ls* were confirmed by motif scanning on Pfam (<https://pfam.xfam.org>), and the genes contained both malectin domain and phosphokinase domain were employed for further analysis.

The transmembrane domain, signal peptide and subcellular location were predicted using TMHMM Server v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) [37], SignalP 5.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>) [19] and ProtComp 9.0 (<http://www.softberry.com/berry.phtml?topic=protcomppl&group=programs&subgroup=proloc>), respectively.

Sequence alignment and phylogenetic analysis

Multiple alignments of all the *CrRLK1Ls* proteins sequences from the three *Gossypium* species and *Arabidopsis* were performed using ClustalW. Phylogenetic trees were constructed using the neighbor-joining method or maximum-likelihood method indicated in the text by MEGAX [38] and the trees were modified by the EvolView (<https://www.evolgenius.info/evolview>) [39].

Chromosomal mapping, gene structure and conserved motifs analysis

GFF (general feature format) information of the cotton *CrRLK1Ls* were obtained from the genome annotation files. The physical chromosome locations of all *CrRLK1Ls* members were visualized by TBtools software [20]. The genomic sequences and positions of exons and introns of *GhCrRLK1Ls* were visualized by Gene Structure Display Server (GSDS) [40] (<http://gsds.cbi.pku.edu.cn/>). The MEME program [41] was used to analyze the conserved motifs in the extracellular domain of *GhCrRLK1Ls*.

Gene duplication and synteny analysis in different *Gossypium* species

The collinearity analysis within or among different species were analyzed by MCScan and JCVI [21], respectively. The charts of synteny were drawn by Circos [42].

Analysis of cis-elements in the promoters

To determine the cis-elements in the promoters, the 1000 bp genomic sequences upstream of the transcription start site (TSS) were extracted. The promoters of *GhCrRLK1Ls* with specific expression pattern were referred as target sequences, while the promoters with nonspecific expression pattern were background sequences. Homer [43] was used to identify the specific cis-elements in the target sequences.

Gene expression pattern analysis

The TM-1 ovules at -3 DPA, -1 DPA, 0 DPA, 1 DPA, 3 DPA and 5 DPA and fibers at 7 DPA, 10 DPA, 15 DPA, 20 DPA, and 30 DPA were collected. All samples were immediately frozen in liquid nitrogen and stored at -80°C. Total RNAs were extracted from all the samples used for transcriptomic analysis. The

GhCrRLK1Ls with an FPKM > 1 at least in one stage of the fiber development were employed for further analysis. The genes were divided into specific and nonspecific expression by the method used in previous study [26]. The expression heatmaps were visualized by matplotlib [44].

RNA extraction and qRT-PCR analysis

The TM-1 ovules at -5, -3, -1, 0, 1, 3 and 5 DPA and fibers at 7 and 10 DPA were collected for total RNA extraction. RNA samples were used as templates for reverse transcription with the PrimeScript RT Reagent kit (Takara, Japan). Primers used in this study are listed in Table S1. The expression of histone-3 gene was used as the internal control. Real-time PCR was carried out with the SYBR Green PCR Master Mix system (Takara) on an Applied Biosystems QuantStudio 5 Real-time PCR System (ABI, Foster City, CA, USA). The PCR amplification conditions were performed according to the methods described by Cheng et al. (2016) [45]. Relative gene expression levels were calculated using the $2^{-\Delta\Delta C_t}$ method.

Abbreviations

CrRLK1L: *Catharanthus roseus* receptor-like kinase 1-like; FER: FERONIA; GPI: glycosyl-phosphatidyl inositol; DPA: days post anthesis; TSS: transcription start site; GFF: general feature format.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Key Research and Development Program of China (No. 2018YFD0100402) and National Natural Science Foundation of China (No. 31621005 and No. 31901581). The funders had no role in study design, data collection and analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

GL conceived and designed the research plans; DZ performed most of the experiments and wrote the manuscript. All authors performed data analysis and interpretation.

Acknowledgements

We are grateful to all the colleagues in our laboratory who provided technical assistance.

References

1. Li C, Wu HM, Cheung AY. FERONIA and her pals: functions and mechanisms. *Plant Physiol.* 2016, 171(4):2379-2392.
2. Franck CM, Westermann J, Boisson-Dernier A. Plant malectin-like receptor kinases: from cell wall integrity to immunity and beyond. *Annu Rev Plant Biol.* 2018, 69(1):301-328.
3. Hématy K, Sado PE, Van Tuinen A, Rochange S, Desnos T, Balzergue S, Pelletier S, Renou JP, Höfte H. A receptor-like kinase mediates the response of *Arabidopsis* cells to the inhibition of cellulose synthesis. *Curr Biol.* 2007, 17(11):922-931.
4. Duan Q, Kita D, Li C, Cheung AY, Wu HM. FERONIA receptor-like kinase regulates RHO GTPase signaling of root hair development. *Proc Natl Acad Sci USA.* 2010, 107(41):17821-17826.
5. Escobar-Restrepo JM, Huck N, Kessler S, Gagliardini V, Gheyselinck J, Yang WC, Grossniklaus U. The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. *Science.* 2007, 317(5838):656-660.
6. Boisson-Dernier A, Roy S, Kritsas K, Grobei MA, Jaciubek M, Schroeder JI, Grossniklaus U. Disruption of the pollen-expressed *FERONIA* homologs *ANXUR1* and *ANXUR2* triggers pollen tube discharge. *Development.* 2009, 136(19):3279-3288.
7. Miyazaki S, Murata T, Sakurai-Ozato N, Kubo M, Demura T, Fukuda H, Hasebe M. *ANXUR1* and 2, sister genes to *FERONIA/SIRENE*, are male factors for coordinated fertilization. *Curr Biol.* 2009, 19(15):1327-1331.
8. Ge Z, Bergonci T, Zhao Y, Zou Y, Du S, Liu MC, Luo X, Ruan H, García-Valencia LE, Zhong S et al. *Arabidopsis* pollen tube integrity and sperm release are regulated by RALF-mediated signaling. *Science.* 2017, 358(6370):1596-1600.
9. Stegmann M, Monaghan J, Smakowska-Luzan E, Rovenich H, Lehner A, Holton N, Belkhadir Y, Zipfel C. The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling. 2017, 355(6322):287-289.
10. Mang H, Feng B, Hu Z, Boisson-Dernier A, Franck CM, Meng X, Huang Y, Zhou J, Xu G, Wang T et al. Differential regulation of two-tiered plant immunity and sexual reproduction by ANXUR receptor-like kinases. *Plant Cell.* 2017, 29(12):3140-3156.
11. Bai L, Ma X, Zhang G, Song S, Zhou Y, Gao L, Miao Y, Song CP. A receptor-like kinase mediates ammonium homeostasis and is important for the polar growth of root hairs in *Arabidopsis*. *Plant Cell.* 2014, 26(4):1497-1511.
12. Qin YM, Zhu YX. How cotton fibers elongate: a tale of linear cell-growth mode. *Curr Opin Plant Biol.* 2011, 14(1):106-111.

13. Wang Z, Yang Z, Li F. Updates on molecular mechanisms in the development of branched trichome in *Arabidopsis* and nonbranched in cotton. *Plant Biotechnol J*. 2019, 17(9):1706-1722.
14. Huang Y, Liu X, Tang K, Zuo K. Functional analysis of the seed coat-specific gene GbMYB2 from cotton. *Plant Physiol Biochem*. 2013, 73:16-22.
15. Machado A, Wu Y, Yang Y, Llewellyn DJ, Dennis ES: The MYB transcription factor GhMYB25 regulates early fibre and trichome development. *Plant J*. 2009, 59(1):52-62.
16. Walford SA, Wu Y, Llewellyn DJ, Dennis ES. GhMYB25-like: a key factor in early cotton fibre development. *Plant J*. 2011, 65(5):785-797.
17. Wan Q, Guan X, Yang N, Wu H, Pan M, Liu B, Fang L, Yang S, Hu Y, Ye W et al. Small interfering RNAs from bidirectional transcripts of GhMML3_A12 regulate cotton fiber development. *New Phytol*. 2016, 210(4):1298-1310.
18. Sun W, Gao Z, Wang J, Huang Y, Chen Y, Li J, Lv M, Wang J, Luo M, Zuo. Cotton fiber elongation requires the transcription factor GhMYB212 to regulate sucrose transportation into expanding fibers. *New Phytol*. 2019, 222(2):864-881.
19. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019, 37(4):420-423.
20. Chen C, Xia R, Chen H, He Y. TBtools, a toolkit for biologists integrating various HTS-data handling tools with a user-friendly interface. *BioRxiv*. 2018:289660.
21. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012, 40(7):e49.
22. 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.
23. 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.
24. Wang M, Tu L, Yuan D, Zhu D, 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*. 2018, 51(2):224-229.
25. 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.
26. Wang K, Wang D, Zheng X, Qin A, Zhou J, Guo B, Chen Y, Wen X, Ye W, Zhou Y et al. Multi-strategic RNA-seq analysis reveals a high-resolution transcriptional landscape in cotton. *Nat Commun*. 2019, 10(1):4714.

27. Kou X, Qi K, Qiao X, Yin H, Liu X, Zhang S, Wu J. Evolution, expression analysis, and functional verification of *Catharanthus roseus* RLK1-like kinase (CrRLK1L) family proteins in pear (*Pyrus bretschneideri*). *Genomics*. 2017, 109(3):290-301.
28. Sakamoto T, Deguchi M, Brustolini OJ, Santos AA, Silva FF, Fontes EP. The tomato RLK superfamily: phylogeny and functional predictions about the role of the LRRIL-RLK subfamily in antiviral defense. *BMC Plant Biol*. 2012, 12:229.
29. Nissen KS, Willats WGT, Malinovsky FG. Understanding CrRLK1L function: cell walls and growth control. *Trends Plant Sci*. 2016, 21(6):516-527.
30. Gonneau M, Desprez T, Martin M, Doblaz VG, Bacete L, Miart F, Sormani R, Hématy K, Renou J, Landrein B et al. Receptor kinase THESEUS1 is a rapid alkalization factor 34 receptor in *Arabidopsis*. *Curr Biol*. 2018, 28(15):2452-2458.
31. Franck CM, Westermann J, Bürssner S, Lentz R, Lituiev DS, Boisson-Dernier A. The protein phosphatases ATUNIS1 and ATUNIS2 regulate cell wall integrity in tip-growing cells. *Plant Cell* 2018, 30(8):1906-1923.
32. Pu L, Li Q, Fan X, Yang W, Xue Y. The R2R3 MYB transcription factor GhMYB109 is required for cotton fiber development. *Genetics*. 2008, 180(2):811-820.
33. 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, 492:423-427.
34. 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.
35. 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 Res*. 2014, 42(Database issue):D1229-D1236.
36. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019, 47(D1):D427-D432.
37. Möller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*. 2001, 17(7):646-653.
38. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018, 35(6):1547-1549.
39. Subramanian B, Gao S, Lercher MJ, Hu S, Chen WH. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res*. 2019, 47(W1):W270-W275.
40. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 2014, 31(8):1296-1297.
41. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME suite: tools for motif discovery and searching. *Nucleic Acids Res*. 2009, 37:W202-W208.

42. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: An information aesthetic for comparative genomics. *Genome Res.* 2009, 19(9):1639-1645.
43. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010, 38(4):576-589.
44. Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng.* 2007, 9(3):90-95.
45. Cheng H, Lu C, Yu JZ, Zou C, Zhang Y, Wang Q, Huang J, Feng X, Jiang P, Yang W et al. Fine mapping and candidate gene analysis of the dominant glandless gene *Gl₂^e* in cotton (*Gossypium spp.*). *Theor Appl Genet.* 2016, 129(7):1347-1355.

Figures

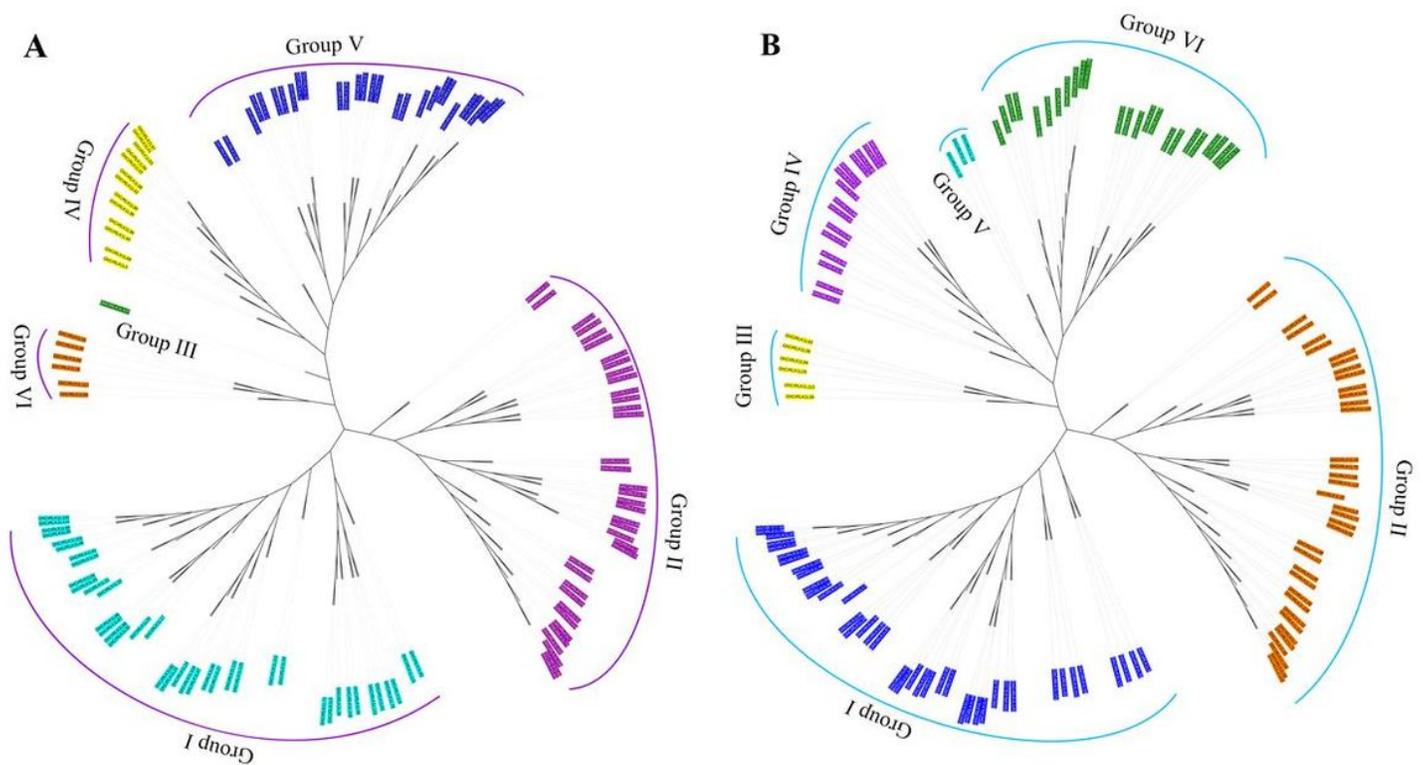


Figure 1

Unrooted phylogeny trees of 125 GhCrRLK1Ls genes of *G. hirsutum*. Unrooted phylogeny tree was constructed by neighbor-joining method based on (A) the full length amino acid sequence, and (B) the region of extracellular domain.

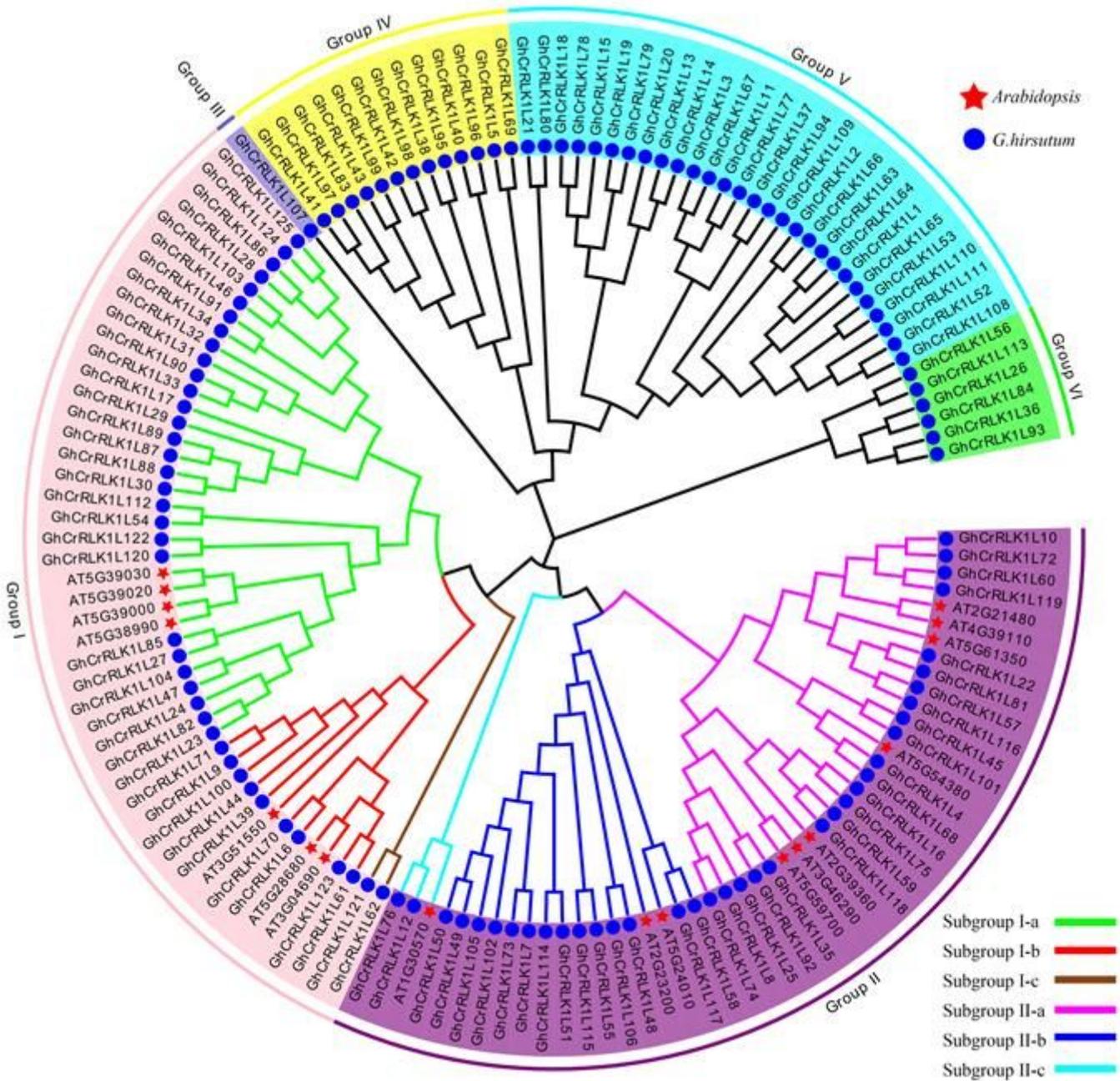


Figure 2

Phylogenetic tree of CrRLK1L proteins. 125 GhCrRLK1Ls from *G. hirsutum* marked by red pentagram and 17 AtCrRLK1Ls from *Arabidopsis* marked by blue cycle. The phylogenetic tree was generated using MEGAX via the neighbor-joining method. All 142 CrRLK1Ls were divided into six groups which were highlighted by different color. Group I and Group II were further clustered into six subgroup indicated by different color lines (subgroup I-a, subgroup I-b, subgroup I-c, subgroup II-a, subgroup II-b and subgroup II-c).

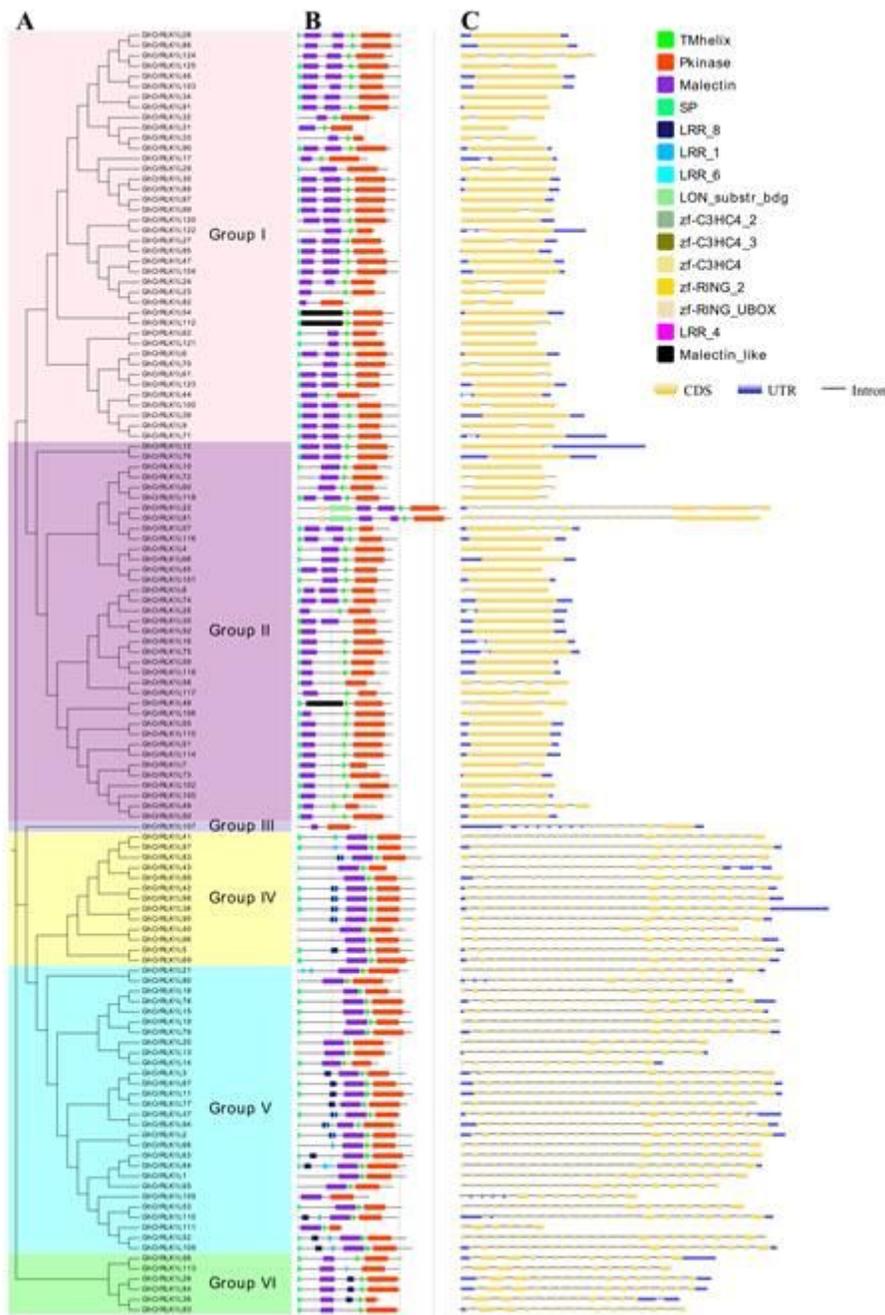


Figure 3

Phylogenetic relationships, protein domains and gene structure analysis of GhCrRLK1Ls. (A) Phylogenetic analysis of GhCrRLK1Ls proteins using MEGAX via the neighbor-joining method. (B) Conserved domain analyzed by the Pfam Scan, transmembrane helix analyzed by TMHMM Server v2.0 and signal peptide predicted by SignalP 5.0 Server. All domains were drawn by EvoView, different domains were marked by different color rectangles. (C) The gene structure of GhCrRLK1Ls. Gene structure analysis was performed by Gene Structure Display Server (GSDS). The CDSs, untranscribed regions (UTRs) and introns were indicated with yellow rectangles, blue rectangles, and black lines, respectively.

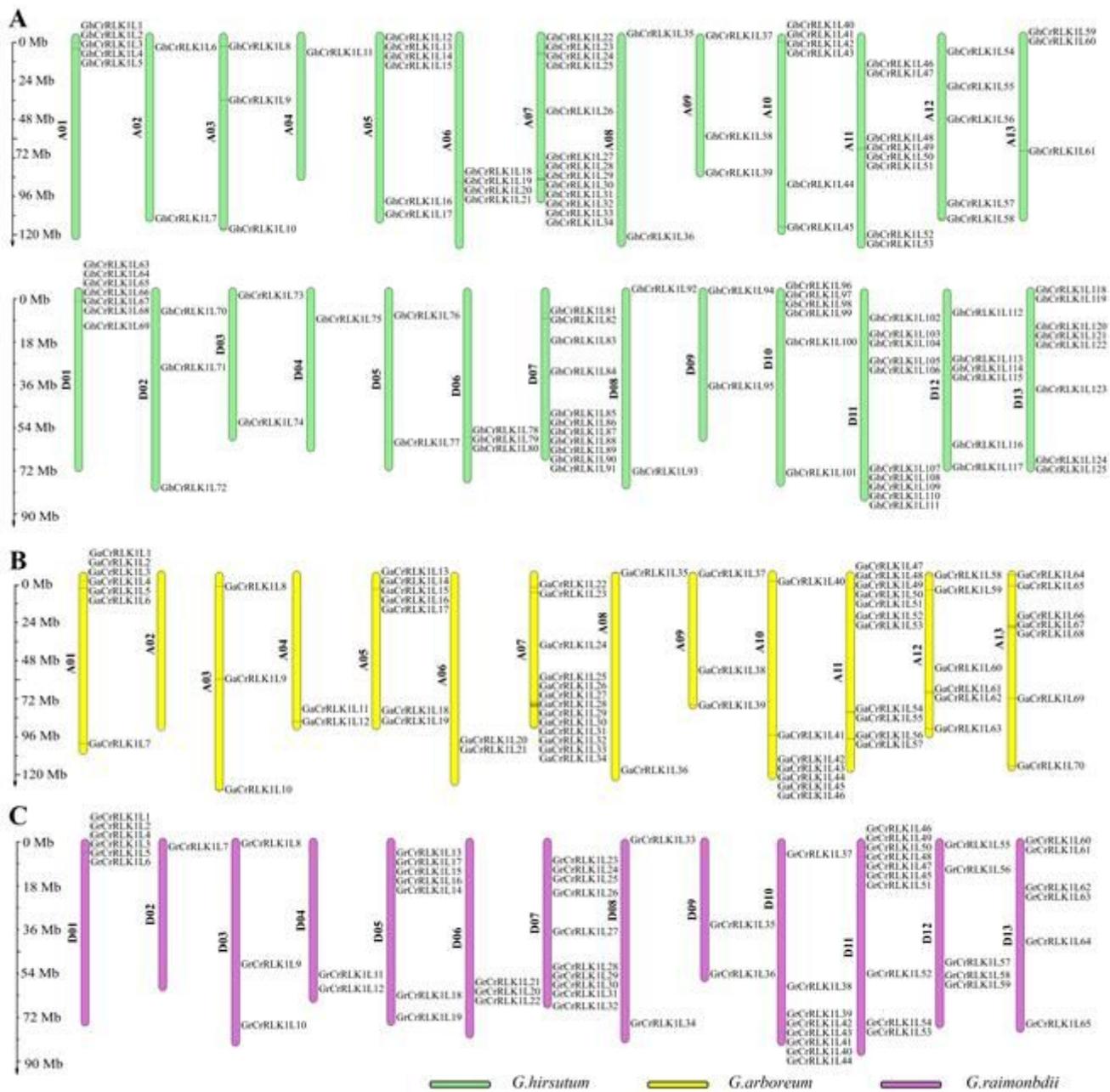


Figure 4

Chromosomal distribution of CrRLK1Ls in (A) *G. hirsutum*, (B) *G. arboreum*, and (C) *G. raimondii*. The scale represented megabases (Mb). The chromosome IDs were indicated beside each vertical bar. The putative CrRLK1Ls were indicated on different chromosomes. Green, yellow and purple bars represented the physical maps of chromosomes in *G. hirsutum*, *G. arboreum* and *G. raimondii*.

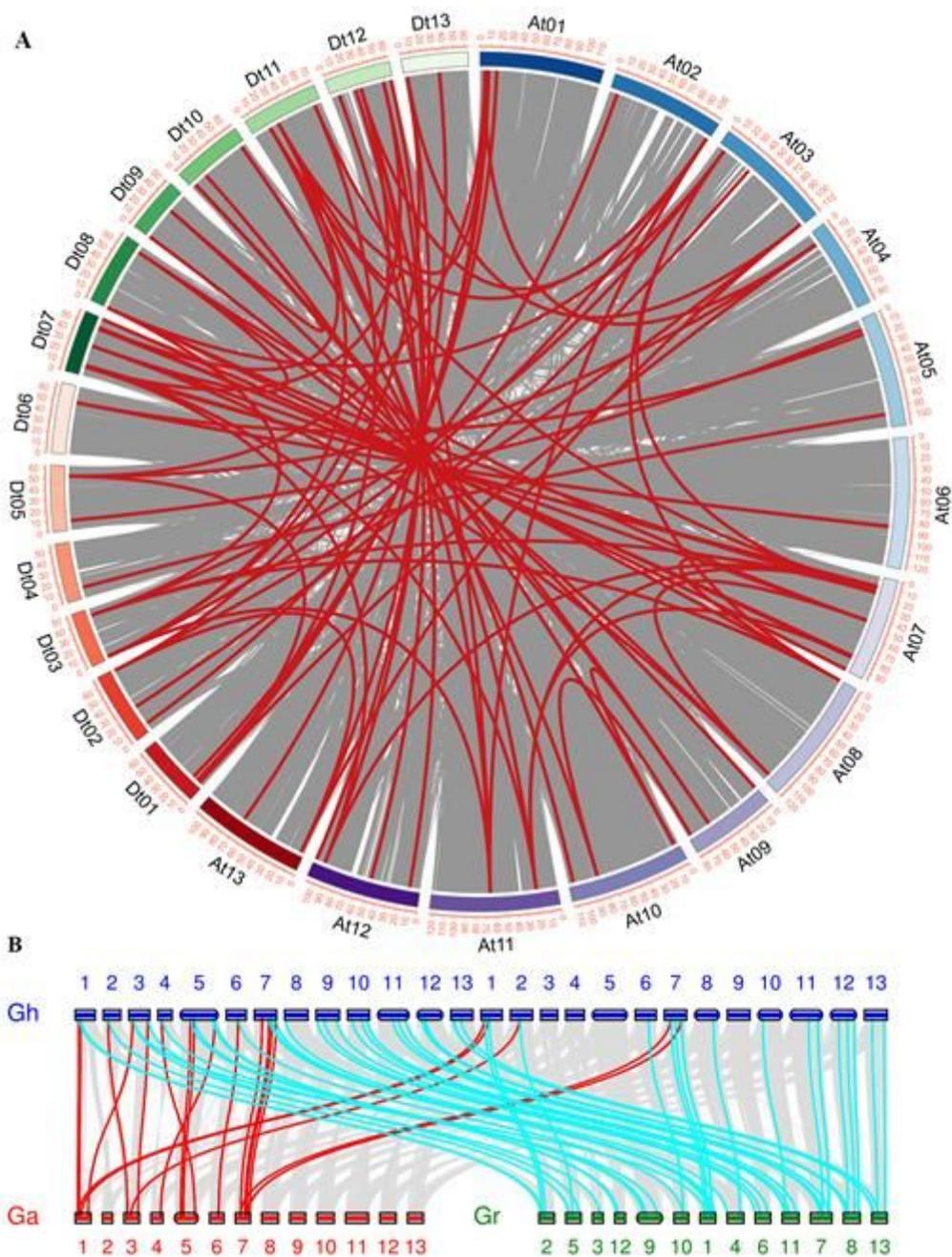


Figure 5

The syntenic relationship between the CrRLK1Ls of three *Gossypium* species. (A) The synteny of GhCrRLK1Ls in *G. hirsutum* marked with red lines. The grey lines indicated all the synteny relationships in *G. hirsutum* genome. (B) The synteny of GhCrRLK1Ls between *G. hirsutum* and *G. arboreum* labeled with red lines and the synteny of GhCrRLK1Ls between *G. hirsutum* and *G. raimondii* labeled with cyan lines. The grey lines indicated all the syntenic relationships in three *Gossypium* genomes.

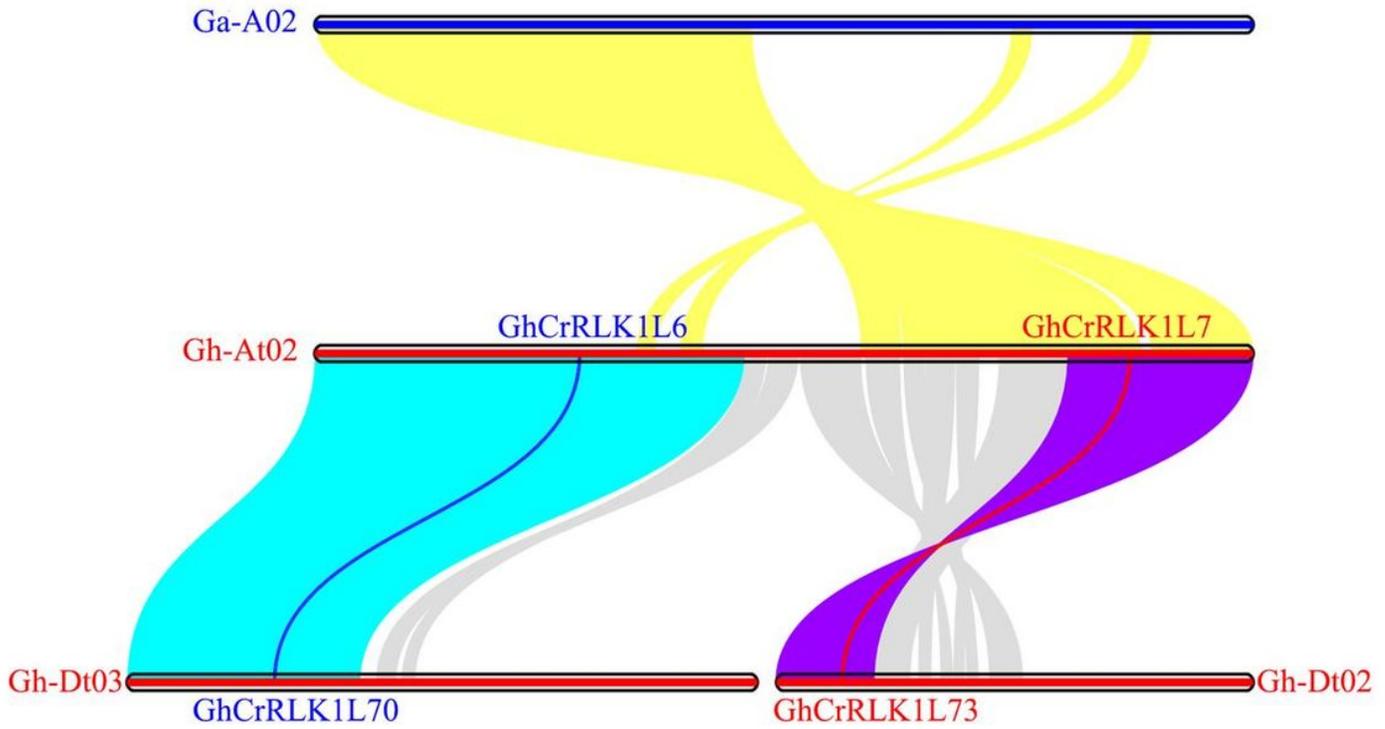


Figure 6

The localized origin of GhCrRLK1L6 and GhCrRLK1L7 on At02 chromosome of *G. hirsutum*. The syntenic relationships between At02 and A02 chromosomes of *G. hirsutum* and *G. arboreum* marked with yellow lines. The syntenic relationships between the regions of GhCrRLK1L6 and GhCrRLK1L70 marked with cyan lines, and purple lines showed the syntenic relationships between the regions of GhCrRLK1L7 and GhCrRLK1L73. The duplication relationships of GhCrRLK1L6 and GhCrRLK1L7 were indicated by blue and red lines.

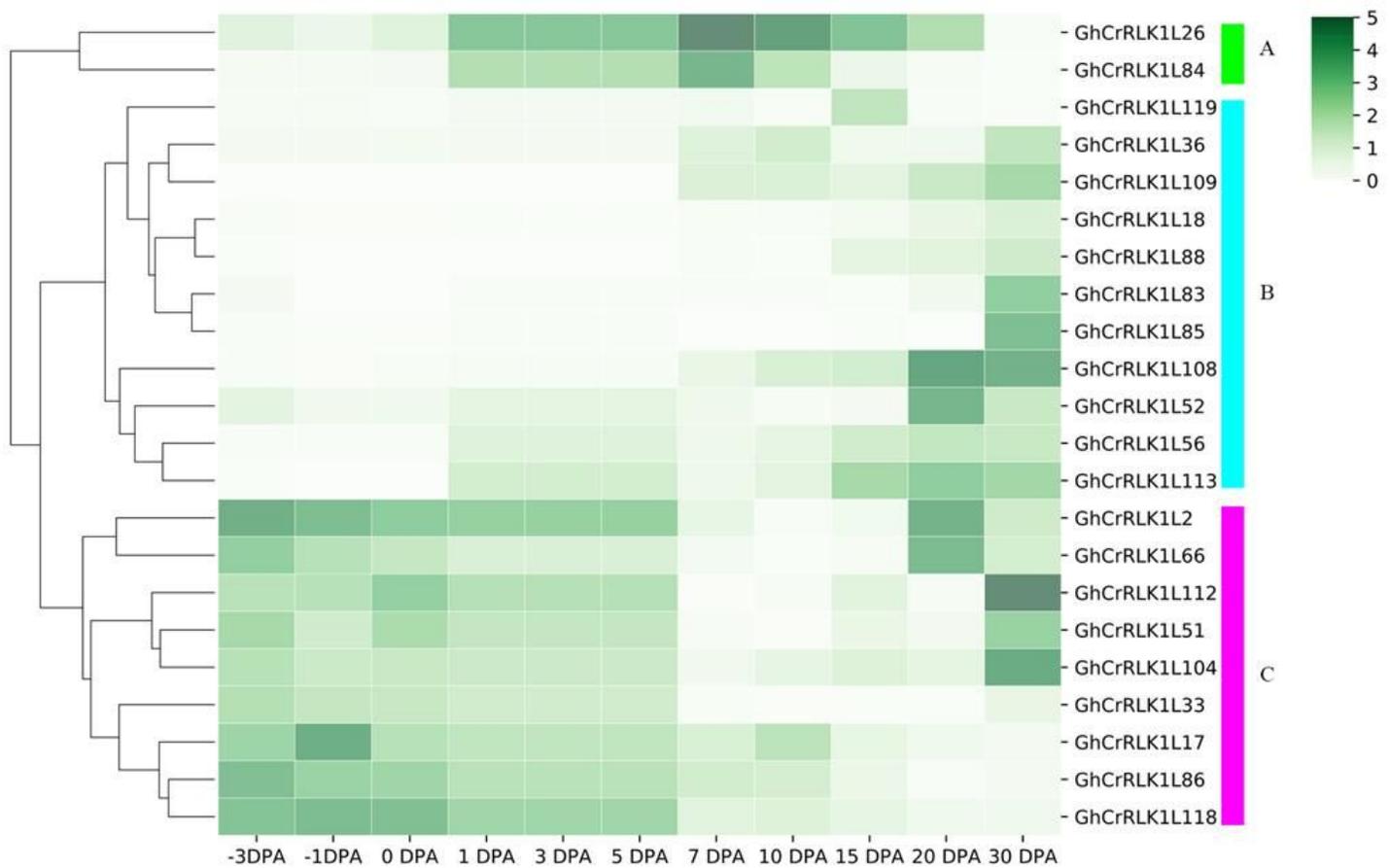


Figure 7

Expression profiles of GhCrRLK1Ls during fiber development. The fiber development stages were showed on the bottom; the gene names were showed on the right; and the phylogenetic relationships were showed on the left. This heatmap was clustered into three groups (A-C), which marked with different colors. Scale bars at the right represented log₂(FPKM+1), -3 DPA to 30 DPA indicated by -3, -1, 0, 1, 3, 5, 7, 10, 15, 20, 30 DPA.

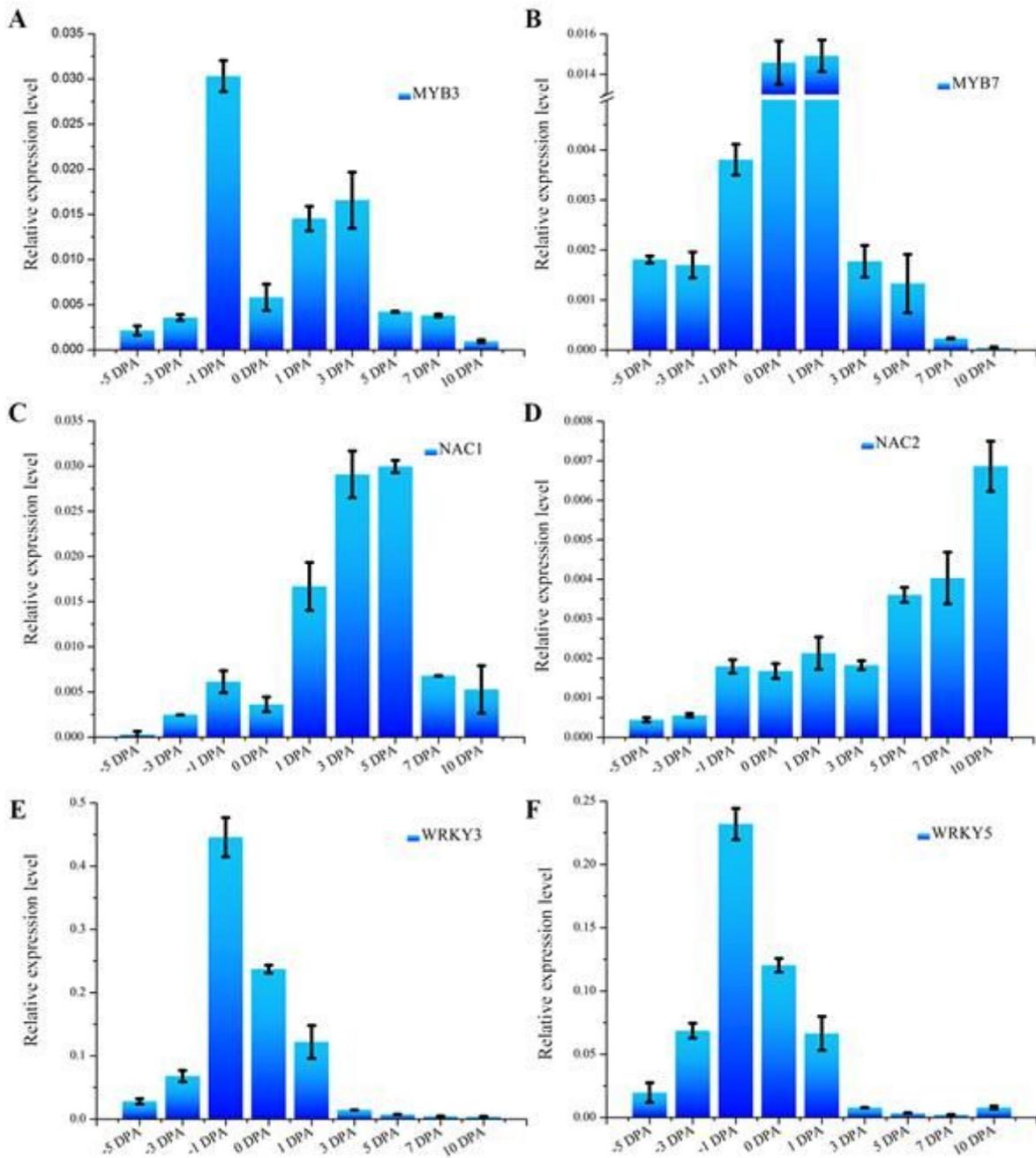


Figure 8

Expression patterns of MYBs, NACs and WRKYs in different fiber development stages via qRT-PCR. (A) and (B) represented the expression patterns of MYB genes Ghir_D12G017660.1 and Ghir_A13G007770.1 in different fiber development stages; (C and D) represented the expression patterns of NAC genes (Ghir_A11G026140.1 and Ghir_D01G006060.1) during different fiber development stages; and (E) and (F) represented the expression patterns of WRKY genes i.e. Ghir_A06G010870.1 and Ghir_A03G015780.1 in different fiber development stages. The error bars showed the standard deviation of three biological replicates.

Supplementary Files

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

- [TableS14.xlsx](#)
- [TableS13.xlsx](#)
- [TableS12.xlsx](#)
- [TableS11.xlsx](#)
- [TableS10.xlsx](#)
- [TableS9.xlsx](#)
- [TableS8.xlsx](#)
- [TableS7.xlsx](#)
- [TableS6.xlsx](#)
- [TableS5.xlsx](#)
- [TableS4.xlsx](#)
- [TableS3.xlsx](#)
- [TableS2.xlsx](#)
- [TableS1.xlsx](#)
- [FigureS4.tif](#)
- [FigureS3.tif](#)
- [FigureS2.tif](#)
- [FigureS1.tif](#)