

Genome-wide analysis of LRR-RLK gene family in four *Gossypium* species and expression analysis during cotton development and stresses response.

SUN Ruibin¹, WANG Shaohui¹, MA Dan¹, LIU Chuanliang^{1,2,*}

¹State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agriculture Sciences, Anyang 455000, China

sunruibin@caas.cn (SUN R.); wangshaohui@caas.cn (WANG S.); madan@caas.cn (MA D.); liuchuanliang@caas.cn (LIU C.)

²Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou University

* Correspondence: liuchuanliang@caas.cn

Abstract: Leucine-rich repeat receptor-like kinases (LRR-RLKs) have been reported to play important roles in plant growth, development and stress responses. However, no comprehensive analysis of this family has been performed in *Gossypium*, which are important economic crop and suffer various stresses in growth and development. Here we conducted a comprehensive analysis of LRR-RLK family in four *Gossypium* species (*G. arboreum*, *G. barbadense*, *G. hirsutum* and *G. raimondii*). A total of 1641 LRR-RLK genes were identified in the four *Gossypium* species involved in our study. Maximum-likelihood phylogenetic tree revealed that all the LRR-RLK genes were divided into 21 subgroups. Exon-intron organization structure of LRR-RLK genes kept relative conserved in subfamilies and between *Arabidopsis* and *Gossypium*. Subfamilies XI and XII were found dramatically expanded in *Gossypium*. Tandem duplication acted as an important mechanism in expansion of *Gossypium* LRR-RLK gene family. Function analysis suggested that plant hormone signaling and plant-pathogen interaction pathway were enriched in *Gossypium* LRR-RLK genes. Promoters analysis and expression profiles analysis revealed that *Gossypium* LRR-RLK genes were extensively regulated by TFs, phytohormone and various environmental stimuli, and play key roles in stress defense and diverse development processes. Our study provided valuable information for further function study of *Gossypium* LRR-RLK genes.

Keywords: LRR-RLK family; *Gossypium*; expansion; phylogenetic analysis; gene expression profile; stress defense

1. Introduction

Receptor-like protein kinases (RLKs) represent a large number of transmembrane kinases which perceive stimulus at cellular surface and mediate the cellular signaling transduction via autophosphorylation and subsequent downstream phosphorylation for intercellular communication or response to extracellular environment [1,2]. In land plants, RLKs form a large family and expand extensively [3–5]. A common feature of the RLKs is that each usually has an N-terminal extracellular domain (ECD) that varies in structure, a transmembrane domain (TM), and a relatively conserved cytoplasmic protein kinase catalytic domain (KD) [6]. The ECD region, which is thought to act as a ligand-binding site, has a diverse structural feature, allowing to interact with proteins, polysaccharides, lipids, and other ligands [4,7]. The leucine-rich repeat RLKs (LRR-RLKs) comprise the largest group of plant RLKs [7,8], which contain a varying number of leucine-rich repeat (LRR) in the ECD region. The LRR is a 20-30 amino acid-residue

sequence motif, appears to provide the structural framework for recognition of ligands [9]. The number and arrangement pattern of LRRs may varies among different LRR-RLKs, partly contribute to the diversity of LRR-RLKs. A comprehensive LRR-RLKs analysis among a diverse of plants lineages classified plant LRR-RLKs into 19 subfamilies, phylogenetic analysis demonstrates much of the diversity of plant LRR-RLKs established in early land plants [10].

In plants, LRR-RLK genes play various important roles in plant development and response to biotic and abiotic stresses [10]. In terms of plant growth and development, the best-characterized LRR-RLK gene member in *Arabidopsis* is *CLAVATA1* (*CLV1*), which involves in shoot and flower apical meristem development. By combining with receptor-like protein *CLV2*, the LRR-RLK *CLV1* is dimerized and recognizes small secreted dodecapeptide *CLV3* as ligand to regulate expression of the downstream transcription factor *WUSCHEL* (*WUS*) which in return up-regulates the expression of *CLV3*, results in a feedback mechanism to regulate the meristem size [11,12]. The heterodimeric LRR-RLK complex *BAK1/BRI1* initiates the brassinosteroid signaling cascade [13,14]. *HAESA* (*HAE*) is involved in floral organ abscission [15], *RPK1* and *TOAD2* encode LRR-RLKs required for proper embryo morphogenesis [16]. In addition, some LRR-RLKs are defense related. In *Arabidopsis*, *FLS2* and *EFR* perceive bacterial antigen and mediate the defense against pathogen [17,18]. The rice LRR-RLK gene *Xa21*, encoding a LRR-RLK, is an effective rice bacterial blight resistance gene [19]. As the metabolic pathway intertwines, some LRR-RLKs function on several aspects. For instance, somatic embryogenesis receptor-like kinases (*SERKs*) participate in the process of microsporogenesis and embryogenesis, and enhance acquisition of embryogenic competence in culture regeneration [20–22]. Meanwhile, recent researches demonstrate that *SERKs* are indispensable in brassinosteroid signaling [23,24], even the rice *OsSERK1* host defense response against fungal infection by mediating defense signaling transduction [22].

As LRR-RLKs has many functional roles, genome-wide identification and analysis of LRR-RLK genes have been carried out extensively. Based on KD sequence phylogeny and gene structure, at least 213 identified LRR-RLK genes in *Arabidopsis thaliana* were classified into 15 groups [3,8], 309 identified in *Oryza sativa* were classified into 5 groups [25], 379 identified in *Populus trichocarpa* were classified into 14 groups [26], 303 identified in *Brassica rapa* were classified into 15 groups [27], 467 identified in *Glycine max* were classified into 14 groups, 234 identified in *Solanum lycopersicum* were classified into 10 groups [28]. To date, LRR-RLK genes have been identified and phylogenetically analyzed in more than 31 plant species [10,29]. However, no similar analysis has been conducted on cotton (*Gossypium spp.*), except *Gossypium arboreum* [29].

Cotton, which comprises several *Gossypium* genus species, is an important economical crop, produce large amounts of natural fibers for textile industry. For a long time, cotton suffers a variety of biotic and abiotic stresses in planting, and many efforts have been taken on the development of cotton fiber to improve the quality and yield. Considering the multifunction of LRR-RLK genes in plant defense response and development process, in this study, we conducted genome-wide identification and phylogenetic analysis of LRR-RLK genes on four genome-sequenced *Gossypium* genus species. In addition, the function and expression profiles of *Gossypium* LRR-RLK genes in several important developmental and stress response processes were analyzed, our investigations provide insights into the evolution of *Gossypium* LRR-RLK genes and the roles of LRR-RLK gene family in development and stress defense.

2. Materials and Methods

2.1 Identification of LRR-RLK genes in four *Gossypium* genus species

Proteomes data of *Gossypium arboreum*, *Gossypium barbadense*, *Gossypium hirsutum*, and *Gossypium raimondii* were download from public database (<ftp://bioinfo.ayit.edu.cn/downloads> [30]; <http://database.chgc.sh.cn/cotton/index.html> [31]; <https://phytozome.jgi.doe.gov/> [32]), the corresponding whole-genome gene annotations were downloaded as well. As LRR-RLKs are featured with KD domain, LRR domain and TM domain, the corresponding Hidden Markov Model (HMM) models of KD including Pkinase (PF00069), Pkinase_Tyr (PF07714), and HMM models of LRR including LRR_1 (PF00560), LRR_2 (PF07723), LRR_3 (PF07725), LRR_4 (PF12799), LRR_5 (PF13306), LRR_6 (PF13516), LRR_8 (PF13855), LRR_9 (PF14580), LRV (PF01816), were downloaded from Pfam database (<http://pfam.xfam.org/>) [33], and provided as queries to conduct homologues search (E-value $< 1 \times 10^{-10}$) against protein database of the four *Gossypium* species respectively by using HMMER 3.1b2 software [34]. The resulting hits that existed in both KD domain search results and LRR domain search results were collected for further filtering. To make sure that we get as close as the whole possible LRR-RLKs, the amino acid sequences of *Arabidopsis* LRR-RLKs members reported by Shiu et al. [7] were retrieved from The Arabidopsis Information Resource (TAIR) database v10.0 (<http://www.arabidopsis.org/>) [35] and served as query to perform a similarity search (evalue $< 1 \times 10^{-5}$, identity $> 50\%$) against protein database of the four *Gossypium* species using BLAST+ v.2.6.0 [36]. The sum total items of HMMER search result and BLAST+ result were used for subsequent validation analysis. InterProScan v.5.24-63.0 [37] was used to confirm the presence of KD domain and LRR domain and other characteristic domains. TMHMM server v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) [38] and Phobius (<http://phobius.binf.ku.dk/>) [39] were used for transmembrane (TM) domain prediction, either TMHMM server or Phobius server indicates a TM domain, we decided it TM domain contained. Proteins that contain both KD domain, LRR domain and TM domain were considered as LRR-RLKs.

2.2 Phylogenetic analysis of *Gossypium* LRR-RLKs

LRR-RLK genes identified in four *Gossypium* species and previously reported in *Arabidopsis* [8] were involved in phylogenetic analysis. Amino acid sequences were used to perform multiple sequence alignment by MUSCLE [40]. Maximum likelihood (ML) tree was constructed by FastTree 2 [41] with default arguments. Neighbor-joining (NJ) tree was constructed by MEGA 7 [42] with 1000 bootstrap.

2.3 Gene structure analysis

Exon-intron structure information of identified LRR-RLK genes were retrieved from the whole-genome gene annotations. LRR, TM and KD domain coordinates were derived from InterProScan annotation results and TM domain prediction results. TBtools [43] was used to display the gene exon-intron structure and domain coding regions.

2.4 Genomic distribution of LRR-RLKs and tandem duplication identification

The genomic coordinates of genes were extracted from genome annotation release data, and then used to map LRR-RLK genes on chromosome by TBtools [43]. Tandem duplication of LRR-RLK genes were identified when genes belong to the same LRR-RLK subfamily and separated by ten or less gene in 200 kb distance meanwhile.

2.5 Gene Ontology (GO) and pathway analysis of *Gossypium* LRR-RLKs

Gene Ontology (GO) and KEGG pathway annotation information for *G. barbadense*, *G. hirsutum* and *G. raimondii* were download from CottonFGD [44]. According to the gene function annotation method used in CottonFGD, GO annotation of *G. arboreum* LRR-RLKs were conducted by InterProScan. KEGG Orthology were assigned to *G. arboreum* LRR-RLKs using KEGG Automatic Annotation Server (KAAS) [45] and then mapped to KEGG pathways. GO and KEGG pathway enrichment analysis were conducted by R package clusterProfiler v3.6.0 [46].

2.6 Promoter and regulatory analysis of *Gossypium* LRR-RLK genes

The upstream 1.5 kb sequence of gene start codon was recognized and extracted as promoter region. Promoter sequences of all *Gossypium* LRR-RLK genes were submitted to PlantCARE database [47] to predict potential cis-acting regulatory elements. Transcription factor (TF) binding sites were predicted by Binding site prediction tools on PlantTFDB 4.0 [48].

2.7 Gene expression profiles analysis of *Gossypium* LRR-RLK genes

RNA sequencing data were accessed from NCBI Sequence Read Archive (SRA) database (PRJNA89721) and CottonFGD [44]. The fragments per kilobase million (FPKM) value for each gene was computed to represent gene's expression levels. Expression heatmap were drawn by R software package ComplexHeatmap (for k-means clustering) [49] and pheatmap (for hierarchy clustering) [50] based on log10-transformed FPKM values.

3. Results and Discussion

3.1 Identified LRR-RLK genes of four *Gossypium* genus species

298, 511, 515, 317 LRR-RLK genes were identified in *G. arboreum*, *G. barbadense*, *G. hirsutum* and *G. raimondii*, respectively. All of these identified LRR-RLK genes contained LRR domain, KD domain, and TM domain simultaneously, and the conserved protein domain arranged in the order of LRR-TM-KD from N-terminal to C-terminal (Figure S1). The number of *Gossypium* LRR-RLK genes account for 0.73%, 0.66%, 0.73%, 0.85% of whole genome protein coding genes in each species respectively. The proportions of LRR-RLK genes largely fit with the result of Liu et al. [10], which demonstrated this proportions in the eight studied angiosperm species were 0.67–1.39%.

3.2 Phylogenetic analysis and gene structure of *Gossypium* LRR-RLK genes

The amino acid sequences of 1641 LRR-RLK genes identified in present study and previously reported 213 *A. thaliana* LRR-RLK genes were used to construct phylogenetic tree. Refer to the classification of *A. thaliana* LRR-RLK genes, the maximum likelihood (ML) tree showed LRR-RLK genes from *Gossypium* were classified into 21 distinct clades (Figure 1). Clades were named to be correspondent subfamilies according to the nomenclature of *A. thaliana* LRR-RLK genes [8]. Most of LRR-RLK subfamilies of *Gossypium* were consistent with *A. thaliana*, while subfamilies VI, VII, VIII, XI were further divided into VI-1 and VI-2, VII-1 and VII-2, VIII-1 and VIII-2, XI-1, XI-2 and XI-3, respectively. Overall, subfamilies III, XI, XII show the highest members number of LRR-RLKs. While, in terms of *A. thaliana*, the majority of LRR-RLKs were distributed in subfamilies I, III, XI (Figure 3, Table S1). This is largely ascribed to that some subfamilies exhibit to be lineage specific expansion. For example, in subfamily XII only 7 of the whole 282 members belong to *A. thaliana*. Another example is the subfamily I, which is mainly (41 out of 61) composed of *A. thaliana* LRR-RLKs. Uncoordinated proportion of each subfamily between *Gossypium* and *A. thaliana* suggesting different expansion patterns of LRR-RLK genes between them. Besides, another tree constructed by MEGA 7 using the neighbor-joining (NJ) method was used to further validate the phylogenetic relationship of LRR-RLK genes in ML tree. Result showed that the topologies of two trees were somewhat different, while gene members assignment among different clades remained relative stable (Figure S2). Therefore, the subfamilies classification in ML tree was reliable and could be used for further analysis.

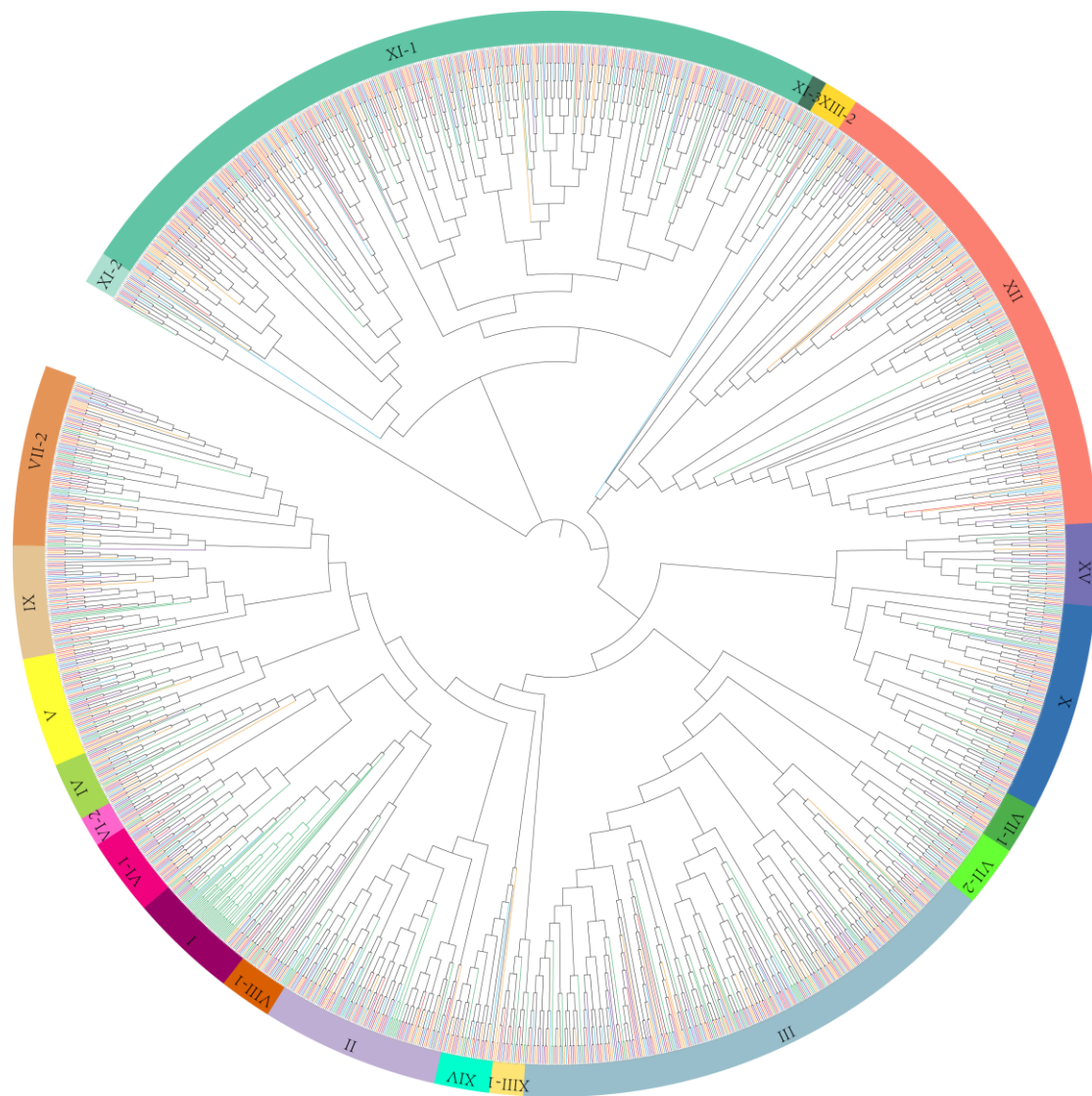


Figure 1. Phylogenetic tree of LRR-RLK genes from four *Gossypium* species and *A. thaliana*. The phylogenetic tree was constructed by maximum likelihood method based on kinase domain amino acid sequences of LRR-RLKs. All LRR-RLK genes were divided into 21 distinct clades, marked by bold curves with different colors. LRR-RLKs from *A. thaliana*, *G. arboreum*, *G. barbadense*, *G. hirsutum* and *G. raimondii* were represented by branch colored within green, red, yellow, purple, blue, respectively.

Gene exon-intron structures and characteristic domain organizations of *Gossypium* LRR-RLK genes were investigated (Figure S1). For each subfamily, gene structures of representative genes from each species were displayed and compared. Number of exons in ECD region varied both within and between subfamilies, while the number of exons in KD domain kept relatively constant. This implied that KD domain region was crucial for the function of LRR-RLKs and might subjected to strict purifying selection in evolution. Variable ECD regions might help LRR-RLKs perceive diverse environmental stimuli. As KD domain is more conserved than LRR domain in LRR-RLK family genes [10,29], we classified the *Gossypium* LRR-RLK genes into three groups based on the exon-intron organization of KD domain (Figure 2). LRR-RLK genes in Group A, which contains subfamilies VII-1, VII-2 and XV, had KD domains located on an integral exon (Figure 2A), while KD domains of LRR-RLK genes in Group B, which contains subfamilies

III, IV, IX, X, XI-1, XI-2, XI-3, XII, were separated by one intron (Figure 2B). In Group C, containing subfamilies I, II, V, VI-1, VI-2, VIII-1, VIII-2, XIII-1, XIII-2, XIV, KD domains of LRR-RLK gene members were separated into 3-6 exons by introns. Meanwhile, the ECD regions of Group C genes (except for subfamilies XIV and VI-1 members) showed highly discrete distributed exons (Figure 2C). As a result, the LRR domains of Group C gene members distributed in many different exons, on the contrary, majority of genes' LRR domains in Group A and B were located in an integral exon. Reference to subfamily division based on ML tree, exon-intron structures of LRR-RLK genes was variable between different subfamilies, while LRR-RLK genes in the same subfamily showed comparable gene structures. This indicated that the gene structures were relatively conserved within each subfamily, implying that the evolutionary relationships of these genes coincide with ML tree. For each subfamily, gene structures of LRR-RLK genes from both *A. thaliana* and four *Gossypium* species were almost identical, suggesting that exon-intron structures of LRR-RLK genes were conserved between *A. thaliana* and *Gossypium*.

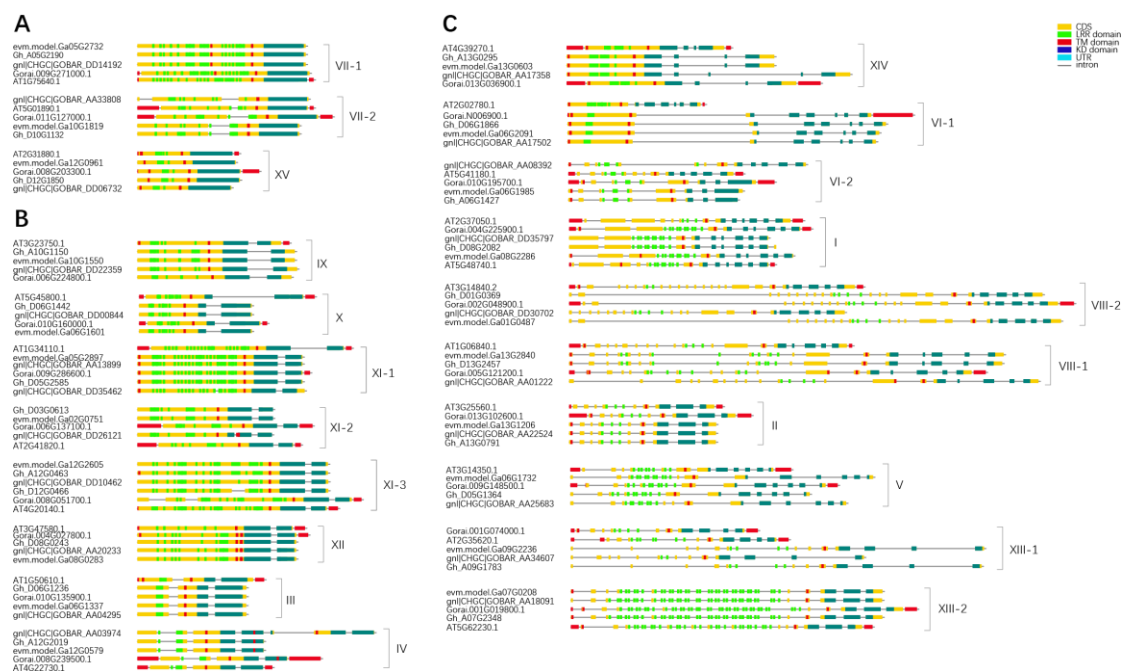


Figure 2. Exon-intron structures of representative LRR-RLK genes of each subfamily from four *Gossypium* species and *A. thaliana*. LRR, KD and TM domain coding regions were marked on exons by different colored rectangles. Based on the exon-intron structures of KD domain, *Gossypium* LRR-RLK genes were classified into Group A, Group B and Group C.

3.3 Dramatical expansion of LRR-RLK subfamilies XI and XII in *Gossypium* species

Statistics of the LRR-RLK genes distribution among different subfamilies were conducted in both *Gossypium* species and *A. thaliana*. In consideration of the genome size difference among different species, the proportion of each subfamily were compared instead of member number of each subfamily. In general, LRR-RLK genes show similar proportion in most of subfamilies between the *Gossypium* and *A. thaliana*. While *Gossypium* LRR-RLK genes have significant smaller proportion of subfamily I distribution, and larger proportion of subfamily XI and XII than *A. thaliana* (Figure 3). There were 41 LRR-RLK I members in *A. thaliana*, accounted for 19.2%

of all *A. thaliana* LRR-RLK genes, while the counterpart percentage was 0.9%-1.4% in *Gossypium* species. *A. thaliana* had about 13 times more LRR-RLK I members comparing to *Gossypium* species. On the contrary, *Gossypium* species showed 26.5%-27.4% and 11.8%-20.0% of LRR-RLKs distributed in subfamily XI and XII, respectively, the proportions were about twice to four times more compare to *A. thaliana*, in which the corresponding percentage were only 15.0% and 3.3%.

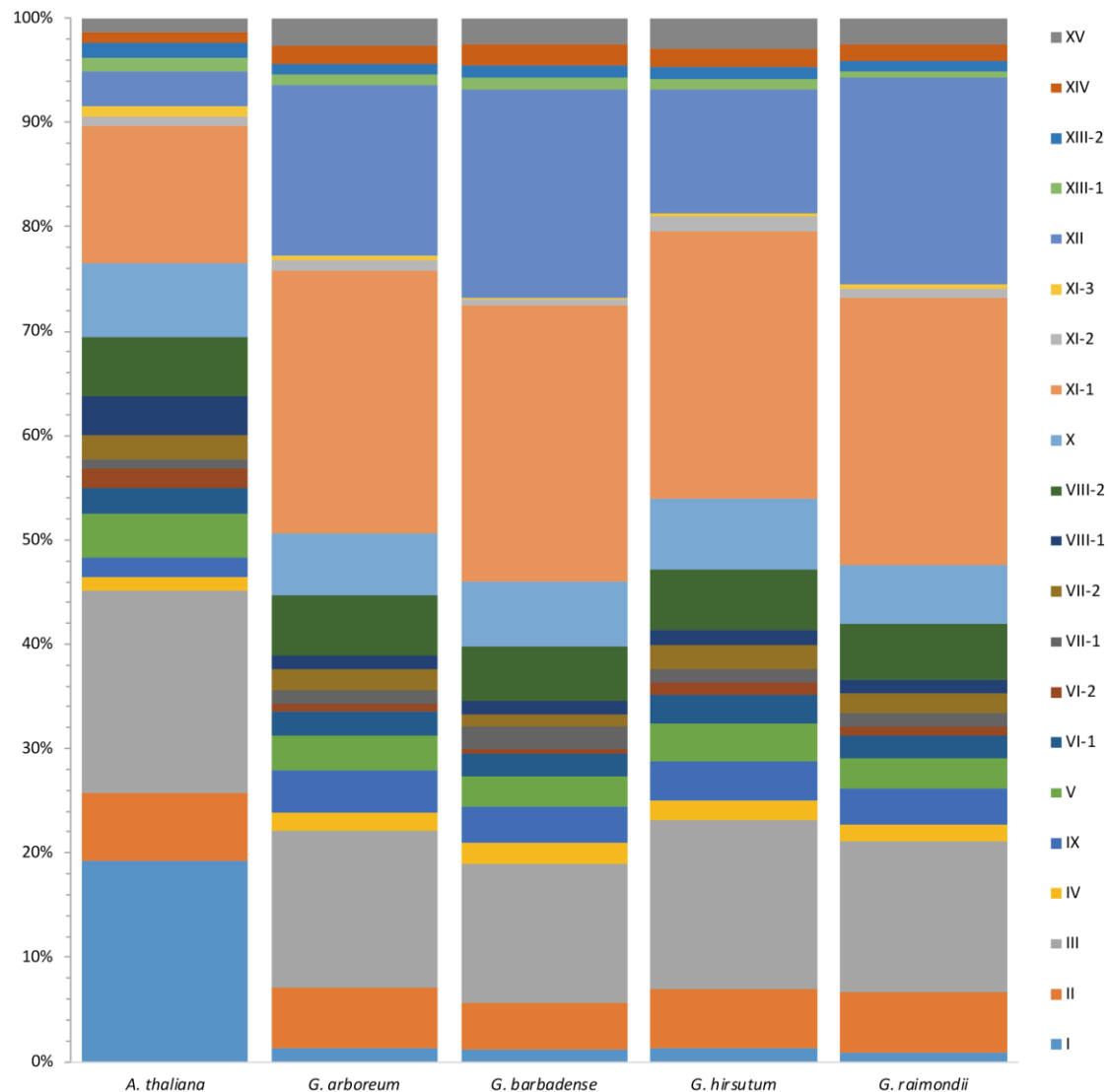


Figure 3. Comparison of LRR-RLK genes distribution among different subfamilies between *A. thaliana* and *Gossypium*. For each species, all the LRR-RLKs were divided into 21 subfamilies, represented by rectangle within different colors, the area of each rectangle within specific color represented the proportion of corresponding subfamily.

As the last common ancestor of angiosperms (LCAA) was estimated to contain about 7 LRR-RLK subfamily I members, *A. thaliana* LRR-RLK I subfamily shows dramatic expansion due to WGD of Brassicaceae [29]. The significant expansion of subfamily I in *A. thaliana* make it to be the largest subfamily. While a slight reduction of this subfamily was found in diploid *Gossypium*, which perhaps suggest the gene loss during the evolution. With respect to LRR-RLK subfamily XI and XII, comparing with LCAA, all four *Gossypium* species exhibit significant expansion in these

two subfamilies. As described in previous study [30], the LRR-RLK XI subfamily almost relatively keep stable member number in most angiosperms, as verified by *A. thaliana* in present study. While in *Gossypium*, subfamily XI expands dramatically (predominantly in XI-1), makes it to be the largest LRR-RLK subfamily, even accounts for about 25% of the total LRR-RLK genes. According to the current knowledges, lots of the *A. thaliana* LRR-RLK genes that fall into XI are involved in plant organ and tissue development. For example, *RGFR* genes are involved in root growth and development. *CLV1*, *BAM* are involved in shoot and floral meristem development and function. *PXY* is involved in vascular-tissue development. *HAE*, *HSL* controlling floral organ abscission. *IKU*, *GSO* are involved in embryo development. As homologues in same cluster may have similar functions, we think *Gossypium* has evolved more development-related LRR-RLK genes to regulate more complex development processes. Besides, some other *A. thaliana* XI gene members, such as *PEPR1* and *PEPR2* are defense and stress response-related, expanded subfamily XI might help *Gossypium* defend stresses. Likewise, subfamily XII expands greatly in *Gossypium*. As known, the expansion LRR-RLK XII is extensive in many different species [30]. Most of genes in subfamily XII are involved in biotic and abiotic stress response, so we suppose that the expansion of LRR-RLK XII help perceive and adapting diverse environment.

As allotetraploid species *G. hirsutum* and *G. barbadense*, known as derived from hybridization of diploid ancestors of *G. arboreum* and *G. raimondii* [51–53], the number of LRR-RLK members in each subfamily between the two allotetraploid species were compared. In general, *G. hirsutum* and *G. barbadense* have similar proportion in terms of most of subfamilies. While when regarding subfamily XII, *G. barbadense* significantly shows higher proportion than *G. hirsutum* (Figure 3). In details, *G. barbadense* has 102 LRR-RLK genes assigned into subfamily XII, as there were only 61 LRR-RLK subfamily XII members in *G. hirsutum*. As most subfamily XII members are defense-related, the fact that *G. barbadense* has better resistance than *G. hirsutum* may be attribute to more subfamily XII members in some degree. We suppose that *G. barbadense* is more likely to retain copies of LRR-RLK subfamily XII gene members form diploid ancestors, while *G. hirsutum* tends to lose some copies. Different retain/loss model of duplicates after polyploidization between *G. hirsutum* and *G. barbadense* may due to adaptation against different environment and different selection underwent.

3.4 Genomic distribution and gene duplicates of LRR-RLK genes in *Gossypium*

The genomic distribution of identified LRR-RLK genes form four *Gossypium* species was displayed respectively (Figure S3). Most of the LRR-RLK genes were mapped to chromosomes, with only 11(3.7%), 44(8.6%), 36(7.0%), 4(1.3%) LRR-RLK genes form *G. arboreum*, *G. hirsutum*, *G. barbadense*, *G. raimondii* were located on scaffolds. The LRR-RLK genes were distributed in all chromosomes but unevenly across different chromosomes. The distribution across different chromosomes was comparative in two polyploid species, chromosome A5, A10, D5, D10 contain highest proportion of LRR-RLK genes in both *G. barbadense* and *G. hirsutum*. The comparable genomic distribution pattern of this large gene family between *G. barbadense* and *G. hirsutum* suggesting the close phylogenetic relationship of these two allotetraploid species. While the distribution across different chromosomes was quite different between two diploid species *G. arboreum* and *G. raimondii* (Figure S3).

Mapping the LRR-RLK genes on chromosomes allow us to detect gene duplication. In our analysis, LRR-RLK genes fall into the same subfamily and separated by ten or less genes in 200 kb chromosomal distance were recognized as tandem duplication set [5,26], there were 26, 47, 33, 25 tandem duplication sets involving 79, 146, 92, 96 tandem duplicates found in *G. arboreum*, *G. barbadense*, *G. hirsutum*, *G. raimondii* respectively, gene numbers contained in each tandem duplication sets range from two to eleven. Furthermore, tandem duplication occurs unevenly among different subfamilies (Table 1). The dramatically expanded subfamilies XI and XII contain majority of all the tandem duplication sets (73.1%, 78.7%, 69.7%, 72.0% for *G. arboreum*, *G. barbadense*, *G. hirsutum*, *G. raimondii* respectively) and tandem duplicate genes (78.5%, 84.2%, 75.0%, 81.3% for *G. arboreum*, *G. barbadense*, *G. hirsutum*, *G. raimondii* respectively), subfamily VIII-2 contains averagely about 14.5% of tandem duplication sets and 11.7% of tandem duplicate genes, corresponding to the somewhat expansion of subfamily VIII-2. Other subfamilies (II, III, VII-1, IX) contain a few of tandem duplication. No tandem duplication was found in subfamilies I, IV, V, VI-1, VI-2, VII-2, VIII-1, X, XI-1, XI-3, XIII, XIV, XV) (Table 1, Table S2). When concerning the two dramatically expanded subfamilies XI and XII, we found that averagely about 40% of LRR-RLK subfamilies XI members and about 60% of LRR-RLK XII members are involved in tandem duplication (Table S2), implying that tandem duplication play an important role in greatly expansion of these subfamilies. Besides, as in subfamilies XI and XII, about half of the LRR-RLK subfamily VIII-2 members derived from tandem duplication (Table S2), suggesting the same important role of tandem duplication in expansion of subfamily VIII-2. We deduced that tandem duplication is extensively existed and acts as an important expansion mechanism in expanded LRR-RLK subfamilies.

Subfamily	Number of tandem duplication genes (gene sets)			
	<i>G. arboreum</i>	<i>G. barbadense</i>	<i>G. hirsutum</i>	<i>G. raimondii</i>
II	5(2)	2(1)	5(2)	5(2)
III	2(1)	4(2)	4(2)	2(1)
IX		2(1)		
VII-1		4(2)		
VIII-2	10(4)	11(4)	14(6)	11(4)
XI-1	31(11)	59(19)	42(16)	35(9)
XI-2			2(1)	
XII	31(8)	64(18)	25(6)	43(9)
Total	79(26)	146(47)	92(33)	96(25)

Table 1. Number of identified tandem duplicate LRR-RLK genes among different subfamilies in four *Gossypium* species. Number of tandem duplicate gene sets in each subfamily were indicated in brackets.

3.5 Functional and pathway annotation analysis of *Gossypium* LRR-RLK genes

Gene ontology (GO) annotation information were available from public *Gossypium* database, we conducted GO enrichment on all LRR-RLK genes from *G. arboreum*, *G. barbadense*, *G. hirsutum* and *G. raimondii*. Results showed that the molecular function of “protein kinase activity”, “protein binding”, “ATP binding” and biological process of “protein phosphorylation”

were significantly enriched in *Gossypium* LRR-RLK genes (Figure 4). The GO enrichment results well conformed with the basic kinase attributes of LRR-RLKs. KEGG Orthology were assigned to *Gossypium* LRR-RLKs by using KAAS (KEGG Automatic Annotation Server), involved pathway were annotated subsequently. Enrichment analysis showed that “plant-pathogen interaction” and “plant hormone signal transduction” were significantly enriched in all four *Gossypium* species (Figure 5), implying that *Gossypium* LRR-RLK genes may be more likely to involve in the process of biotic stress defense. As plant hormone participates in lots of developmental and abiotic stress response processes, such as fiber development, florescence, drought response, salt stress response et al., we think that *Gossypium* LRR-RLK genes have diverse functional roles in developmental, biotic and abiotic defense processes. “Toll-like receptor signaling” and “NOD-like receptor signaling pathway” pathways were found enriched in *G. arboreum* and *G. hirsutum*, as both Toll-like receptor and NOD-like receptor were infection defense and disease resistance related proteins [54,55], these results thus further illustrated that the key roles of LRR-RLK genes in disease defense may mediated by related signaling.

Function of the two dramatic expanded *Gossypium* LRR-RLK subfamilies XI and XII were further focused, results showed that the pathway “plant-pathogen interaction” and “MAPK signaling” were significantly enriched for *Gossypium* LRR-RLK subfamilies XII. The expansion of subfamily XII may confer *Gossypium* more vigorous pathogen response and better resistance to various environmental stresses.

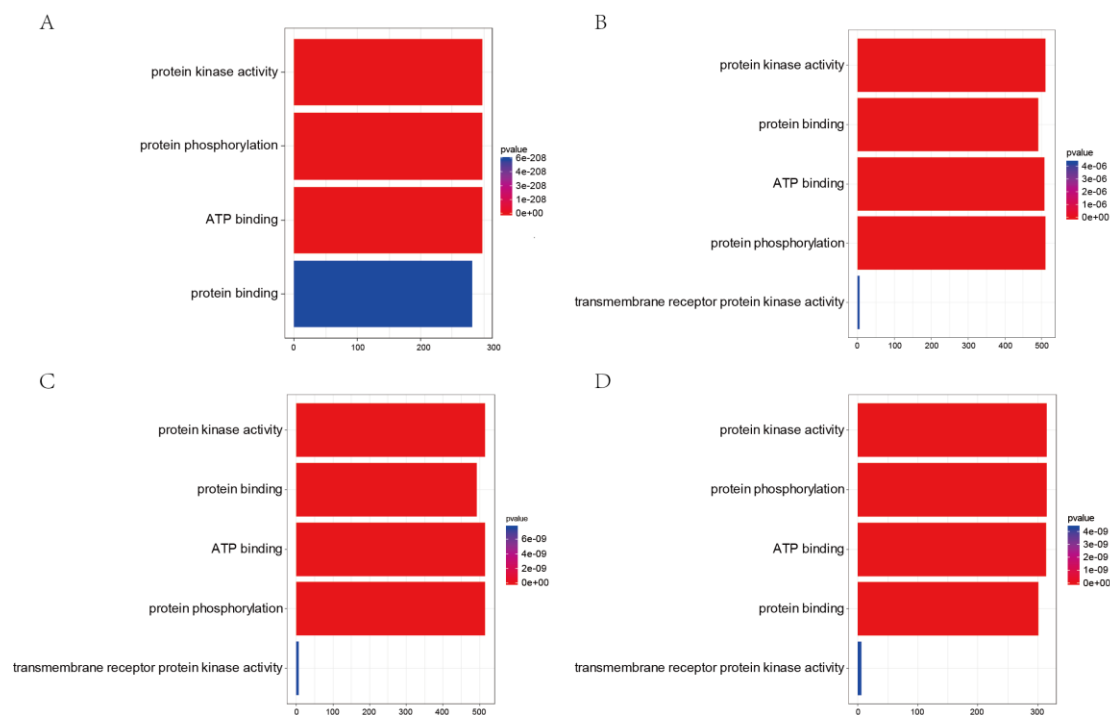


Figure 4. GO enrichment result of *Gossypium* LRR-RLK genes. Results of *G. arboreum*, *G. barbadense*, *G. hirsutum* and *G. raimondii* were shown by A, B, C, D, respectively.

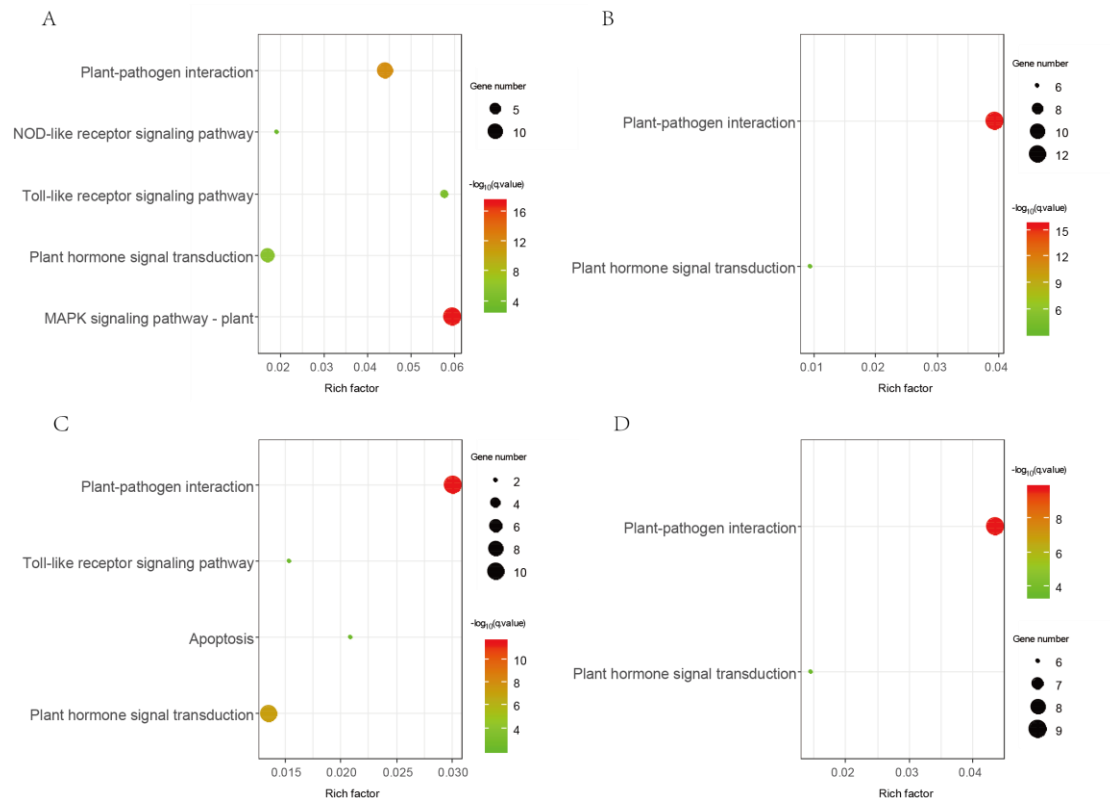


Figure 5. KEGG pathway enrichment result of *Gossypium* LRR-RLK genes. Results of *G. arboreum*, *G. barbadense*, *G. hirsutum* and *G. raimondii* were shown by A, B, C, D, respectively.

3.6 Cis-acting regulatory analysis of *Gossypium* LRR-RLK genes' promoters

Cis-acting regulatory elements in the promoter regions (1.5 kb sequence upstream of start codon) of *Gossypium* LRR-RLK genes were detected by searching PlantCARE database. Phytohormone, stresses defense, cell cycle related cis-acting regulatory elements were widespread in the promoter regions of *Gossypium* LRR-RLK genes (Table 2, Table S3). More than 80% of *Gossypium* LRR-RLK genes had water stress, drought stress and light response cis-acting regulatory elements in their promoters, more than half of *Gossypium* LRR-RLK genes had heat, osmotic stress, low pH, nutrient starvation, anaerobiosis stresses response, ethylene (ETH) and abscisic acid (ABA) response cis-acting regulatory elements in their promoters additionally. Methyl jasmonate (MeJA) response elements were existed in many *Gossypium* LRR-RLK genes' promoters (50.0%, 42.3%, 44.3%, 47.3% of *G. arboreum*, *G. barbadense*, *G. hirsutum* and *G. raimondii* LRR-RLK genes, respectively). Salicylic acid (SA) response elements were found in about 40% of *Gossypium* LRR-RLK genes' promoters. Gibberellin and auxin response elements were found in about 20% of *Gossypium* LRR-RLK genes' promoters. About half of *Gossypium* LRR-RLK genes had wounding and pathogen response elements in their promoters. TC-rich repeats (defense and stress response) and LTR (cold response) elements were found in promoters of about 30% of *Gossypium* LRR-RLK genes. Besides, more than 35% of *Gossypium* LRR-RLK genes had cell cycle and cell proliferation related elements in promoter regions. Some (about 5%) *Gossypium* LRR-RLK genes had heavy metal ions response related cis-acting regulatory element. Cis-acting regulatory elements analysis results revealed that the expression of *Gossypium* LRR-

RLK genes extensively regulated by phytohormone and other diverse abiotic and biotic environmental signals, implying the important roles of *Gossypium* LRR-RLK genes in stresses defense and development.

Element species (ID of CARE)	Number of elements in promoters of LRR-RLK genes(percentage)							
	<i>G. arboreum</i>		<i>G. barbadense</i>		<i>G. hirsutum</i>		<i>G. raimondii</i>	
core promoter / enhancer element (CAAT-box, TATA-box, AT-TATA-box, TATA, A-box)	298	(100.0%)	509	(99.6%)	515	(100.0%)	310	(97.8%)
water response (Myb, AT-rich)	285	(95.6%)	469	(91.8%)	492	(95.5%)	288	(90.9%)
drought response (MYC, as-1, MBS, DRE core, DREE1, ACTCATCCT sequence, MYB recognition site)	279	(93.6%)	450	(88.1%)	479	(93.0%)	274	(86.4%)
cold response (LTR)	79	(26.5%)	150	(29.4%)	168	(32.6%)	98	(30.9%)
heat, osmotic stress, low pH, nutrient starvation stresses response (STRE, TCA)	202	(67.8%)	337	(65.9%)	347	(67.4%)	177	(55.8%)
anaerobic response (ARE, GC-motif)	236	(79.2%)	378	(74.0%)	402	(78.1%)	229	(72.2%)
defense response (TC-rich repeats)	110	(36.9%)	167	(32.7%)	181	(35.1%)	108	(34.1%)
wounding and pathogen response (W box, WUN-motif, WRE3, box S)	149	(50.0%)	256	(50.1%)	252	(48.9%)	158	(49.8%)
light response (GT1-motif, TCT-motif, GATA-motif, MRE, AE-box, I-box, TCCC-motif, ATCT-motif, AT1-motif, GA-motif, chs-CMA1a, LAMP-element, 3-AF1 bindin site, ACE, Gap-box, Sp1, chs-CMA2a, ATC-motif, Box II, AAAC-motif, GTGGC-motif, CAG-motif, ACA-motif, chs-Unit 1 m1, L-box, LS7, 4cl-CMA2b, Pc-CMA2c)	149	(50.0%)	256	(50.1%)	252	(48.9%)	158	(49.8%)
circadian response (circadian, ERE, ABRE)	36	(12.1%)	83	(16.2%)	79	(15.3%)	52	(16.4%)
ETH response (ERE)	226	(75.8%)	354	(69.3%)	370	(71.8%)	242	(76.3%)
ABA response (ABRE, ABRE3a, ABRE4, ABRE2, AT-ABRE)	160	(53.7%)	268	(52.4%)	299	(58.1%)	175	(55.2%)
GA response (P-box, TATC-box, GARE-motif, CARE)	68	(22.8%)	117	(22.9%)	130	(25.2%)	68	(21.5%)
JA response (CGTCA-motif, TGACG-motif, JERE)	149	(50.0%)	216	(42.3%)	228	(44.3%)	150	(47.3%)
SA response (TCA-element)	133	(44.6%)	191	(37.4%)	218	(42.3%)	112	(35.3%)
auxin response (TGA-element, TGA-box, AuxRR-core, AuxRE)	60	(20.1%)	93	(18.2%)	87	(16.9%)	71	(22.4%)
cell cycle and cell proliferation response (Myb-binding site, CCGTCC motif, MSA-like, re2f-1, NON, dOCT, E2Fb)	116	(38.9%)	197	(38.6%)	200	(38.8%)	112	(35.3%)
Cd response (AP-1)	17	(5.7%)	22	(4.3%)	27	(5.2%)	16	(5.0%)
MYB related	171	(57.4%)	283	(55.4%)	299	(58.1%)	166	(52.4%)
tissue specific / preferential expressed (GCN4_motif, RY-element, AC-I, AC-II, AACA_motif, telo-box, motif I)	49	(16.4%)	66	(12.9%)	73	(14.2%)	42	(13.2%)

Table 2. Statistics of cis-acting regulatory element detected in promoter regions of *Gossypium* LRR-RLK genes. (*Cis-acting regulatory elements that have no functional description were not shown, see Table S3 for details.)

Transcription factors (TFs) play key roles in many cellular and biological processes by regulating expression of corresponding target genes. To investigate the possible regulation relations between TFs and *Gossypium* LRR-RLK genes, TF binding sites were predicted by online tools—Binding site prediction on PlantTFDB [48]. Result showed that *Gossypium* LRR-RLK genes could be regulated by 39 TF families (Figure 6, Table S4). Dof, MYC_MADS, MYB, AP2, C2H2, ERF were the most widely functioning TF families, could regulate majority of *Gossypium* LRR-RLK genes. Most of the top TF families were implicated in various aspects of plant development, hormonal signal transduction, plant defense and stresses response, suggesting that *Gossypium* LRR-RLKs might participate diverse plant development and stresses defense processes by TFs mediated regulation.

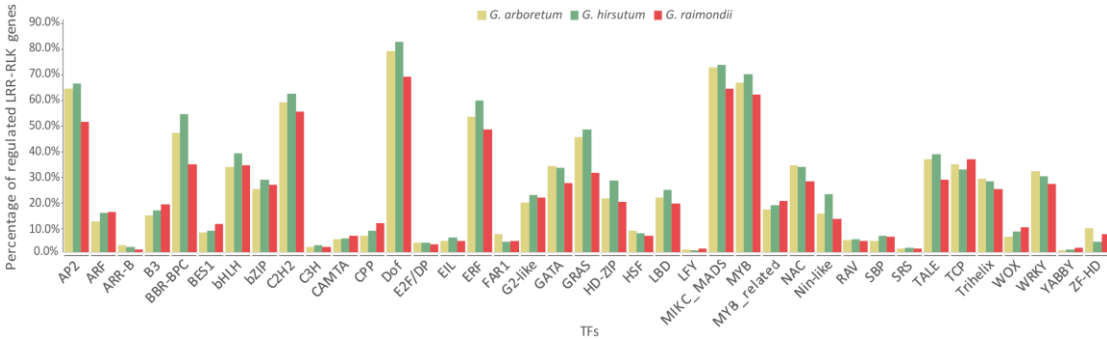


Figure 6. Statistics of *Gossypium* LRR-RLK genes regulated by different families of TFs (genes with TF binding sites were considered to be regulated by TFs).

3.7 Gene expression of *Gossypium* LRR-RLKs during developmental and stress defense processes

As the functional analysis showed the important roles of *Gossypium* LRR-RLKs in diverse developmental and defense processes, the expression profilers of *Gossypium* LRR-RLKs in several important developmental (fiber development, ovule development) and biotic stress (Verticillium wilt) defense and abiotic stress (cold, hot, drought, salt) defense processes were investigated. Fiber development is an important process in *Gossypium*, k-means clustering showed that *G. hirsutum* LRR-RLK genes were clustered into two groups, the majority of Group 1 gene members showed relative lower or moderate expression throughout the fiber development, while Group 2 gene members were highly expressed at the 5 and 10 dpa (day post anthesis) stages, followed by down-regulation at the 20 and 25 dpa (Figure 7A). The results suggested that LRR-RLK genes in Group 2 were important for the early development of fiber. There were 8 LRR-RLK genes assigned into Group 2 (Gh_A07G1471, Gh_D09G1268, Gh_A10G0460, Gh_D10G0477, Gh_A11G1546, Gh_D11G3486, Gh_A13G0257, Gh_D13G0274), except for Gh_A07G1471 and Gh_D09G1268, other 6 Group 2 members belong to 3 homologous pairs. Additionally, orthologous genes in Group 1 of Gh_A07G1471 and Gh_D09G1268 showed similar expression pattern with themselves. In conclusion, for LRR-RLK genes in Group 2, orthologous genes

between A and D sub-genomes showed identical expression patterns, suggesting the functional conservation of Group 2 LRR-RLK genes in fiber development between different *Gossypium* species.

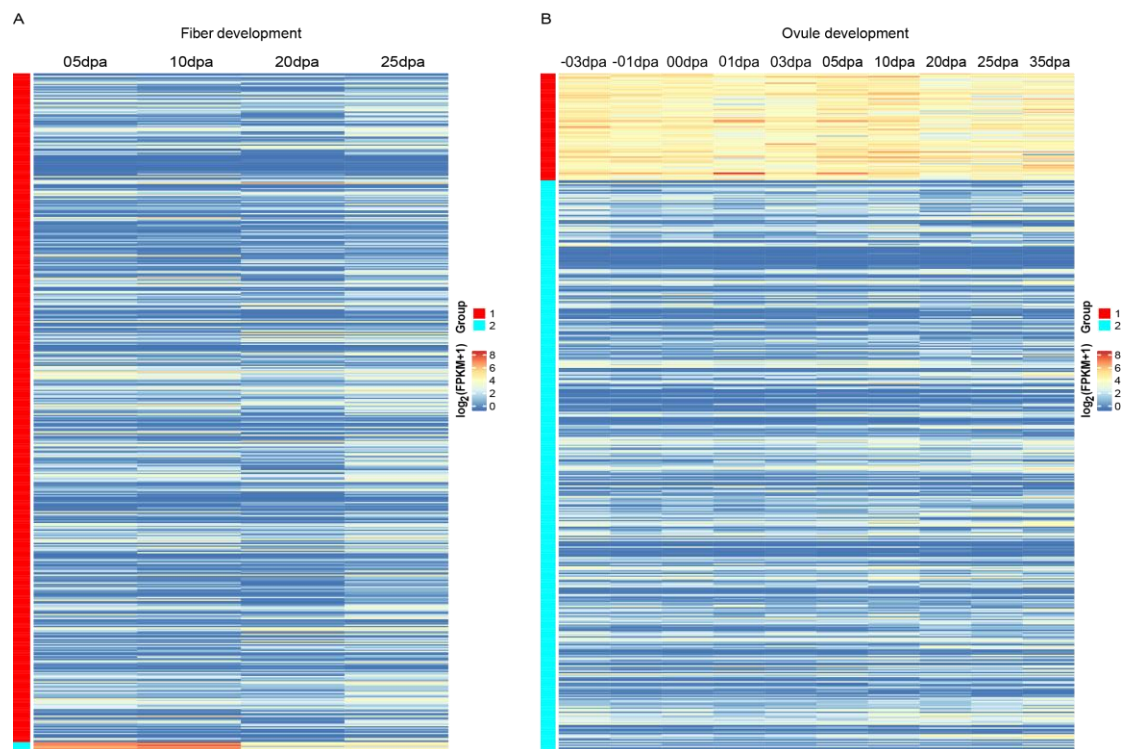


Figure 7. Expression patterns of *G. hirsutum* LRR-RLK genes in fiber and ovule development. Genes were clustered by k-means method.

In the process of ovule development, all *G. hirsutum* LRR-RLK genes were divided into two groups according to K-means clustering. Group 1 contain 81 LRR-RLK genes that showed highly expression at almost all stages of ovule development (Figure 7B), implying their important roles in ovule development. Moreover, the 6 LRR-RLK genes (Gh_A10G0460, Gh_D10G0477, Gh_A11G1546, Gh_D11G3486, Gh_A13G0257, Gh_D13G0274) that participated in fiber development as described above were also assigned to Group 1 in the ovule development K-means clustering. This implied the versatility of single *G. hirsutum* LRR-RLK gene, which might participate in multiple developmental processes.

Cotton is inevitably threatened by diverse abiotic stresses during its growth and development. Therefore, expression profiles of *G. hirsutum* LRR-RLK genes response to cold, hot, drought, salt stresses were analyzed. *G. hirsutum* LRR-RLK genes were clustered in four groups based on their expression profiles response to multiple abiotic stresses (Figure 8). About half of *G. hirsutum* LRR-RLK genes kept low expression during the stress treatments, as shown in Group blue. Majority of LRR-RLK genes in Group green showed relative lower expression level in control, most were down-regulated response to all the four abiotic stresses compared to control, some members subsequently up-regulated at late response stages. Genes in Group orange and Group red showed relative higher expression level in control. Most of Group orange genes down-regulated under all stresses, different genes down-regulated at different stages of stresses response suggested diverse mechanisms of LRR-RLK genes response to abiotic stresses. Genes in Group red kept highly expressed during all stress response processes. Most of them up-regulated at early stages of

cold and hot exposure while late stages of PEG and salt stress treatments. These results implied that *Gossypium* LRR-RLK genes were multi-functional, play important roles in multiple abiotic stresses response, and might help cotton adapt diverse abiotic environment.

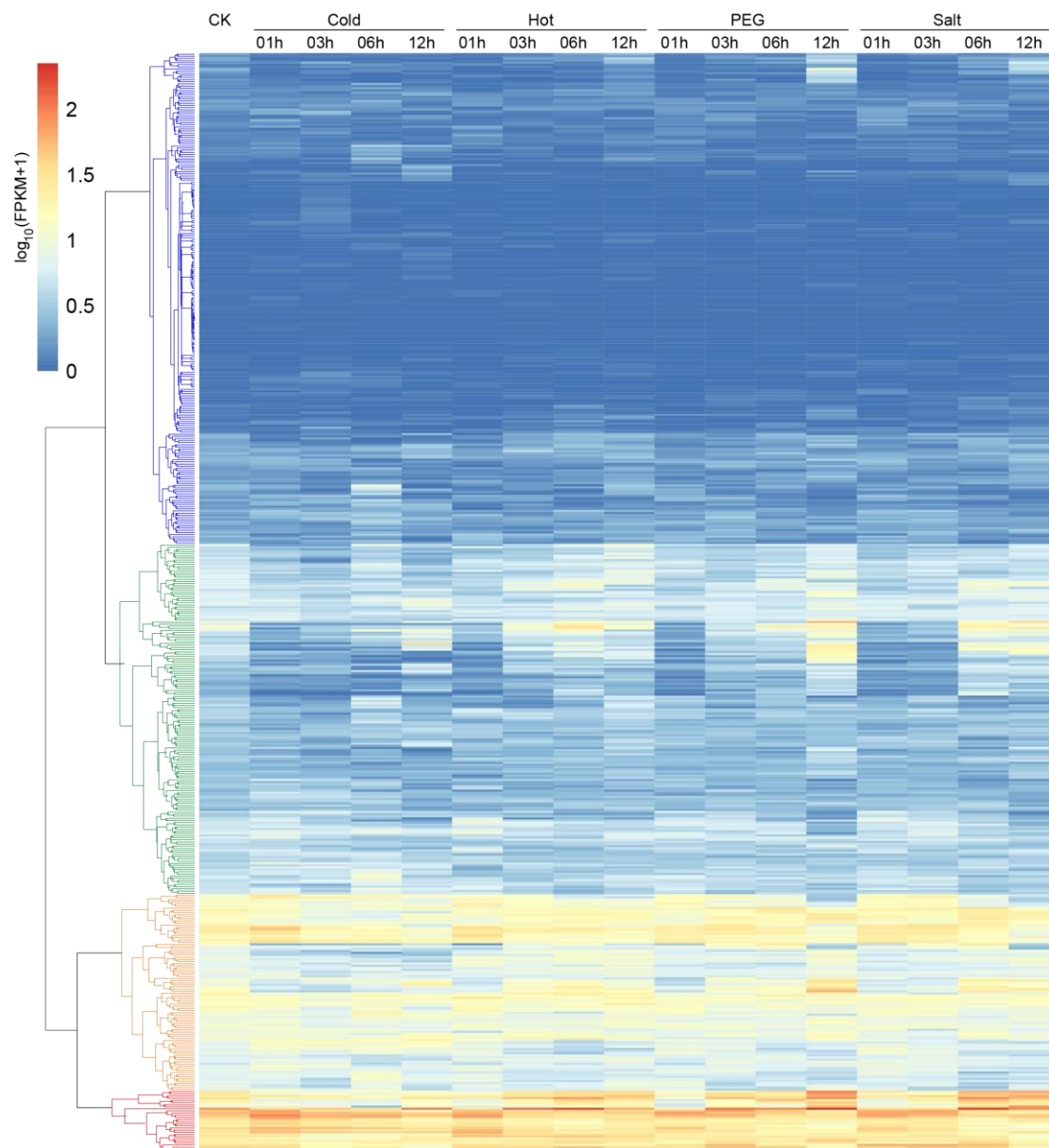


Figure 8. Expression patterns of *G. hirsutum* LRR-RLK genes in response to diverse abiotic stresses (cold, hot, drought simulated by PEG, salt). Hierarchical clustering analysis classified genes into four distinct groups (colored by blue, green, orange, red, respectively).

Verticillium wilt is one of the most intractable diseases in cotton growth. The expression profiles of *Gossypium* LRR-RLK genes under Verticillium wilt infection were investigated to analyze their response to biotic stress. The main cultivated cultivar *G. hirsutum* is susceptible to Verticillium wilt, while *G. barbadense* shows better resistance and immune to Verticillium wilt. Transcriptomic dynamics response to Verticillium wilt has been studied between *G. hirsutum* and *G. barbadense* [56]. There were 75 LRR-RLK genes showed significantly differentially expressed ($\log_2\text{FC} > 1$ and $\text{FDR} < 0.05$) in *G. barbadense* under Verticillium wilt infection. The counterpart

differentially expressed LRR-RLK genes number in *G. hirsutum* was 75 too. Among these differentially expressed LRR-RLK genes, subfamily XI-1, XII and VIII-2 accounted for the largest share (Figure 9). Comparing *G. barbadense* with *G. hirsutum*, we found that *G. barbadense* had more differentially expressed LRR-RLK genes belong to subfamily XI-1 and XII than *G. hirsutum* (Figure 9). Expression heatmap revealed that the majority of subfamily XI-1 and XII LRR-RLK genes down-regulated significantly when infected by *Verticillium* wilt, while subfamily VIII-2 LRR-RLK genes were more likely to be up-regulated. Additionally, *G. barbadense* LRR-RLK subfamily XI-1 and XII genes usually up-regulate more significant than *G. hirsutum* ones, as subfamily VIII-2 genes down-regulated more than *G. hirsutum* ones (Figure 10). As a result, there were more LRR-RLK genes showed low expression level in infected *G. barbadense* than in infected *G. hirsutum*. The difference of LRR-RLK expression regulation between *G. barbadense* and *G. hirsutum* may be associated with disease resistance difference of these two species.

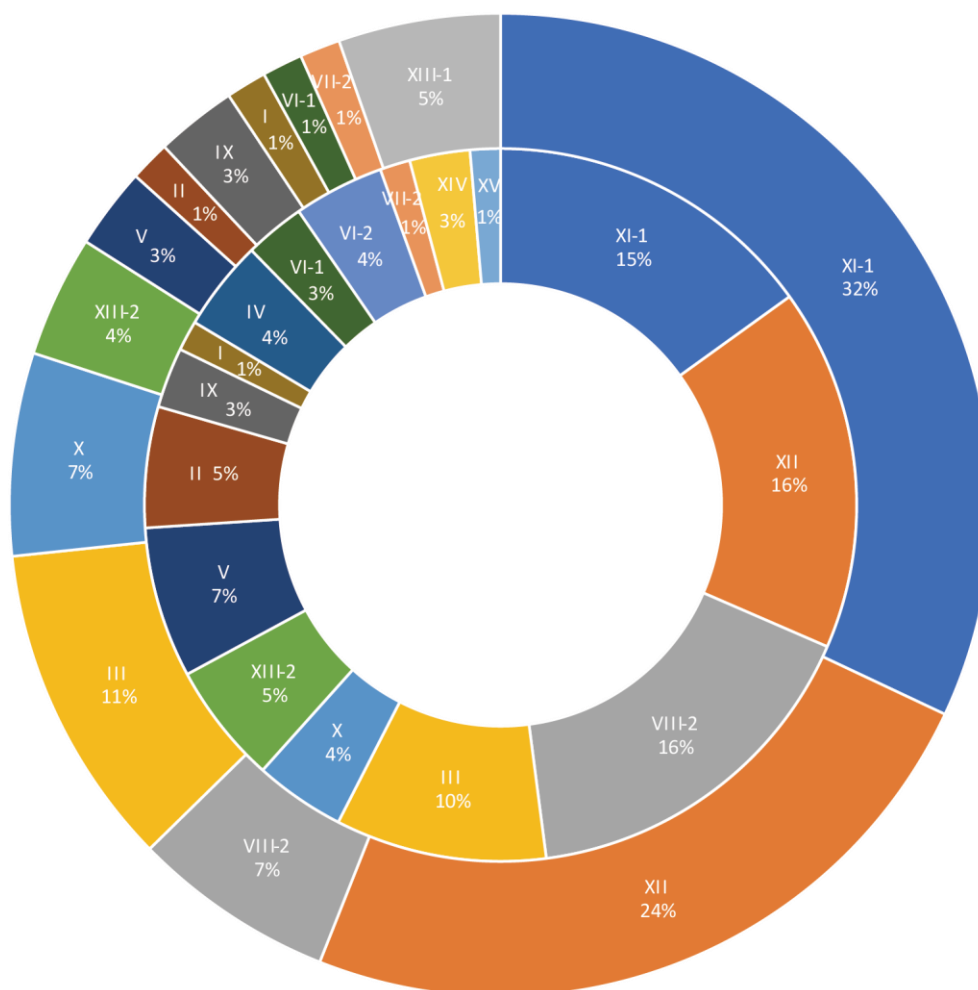


Figure 9. Distributions of differentially expressed LRR-RLK genes among different subfamilies in response to *Verticillium dahlia* infection in *G. barbadense* (outer ring) and *G. hirsutum* (inner ring). *G. barbadense* significantly had higher percentage of differentially expressed genes belong to subfamilies XI-1 and XII genes than *G. hirsutum*.

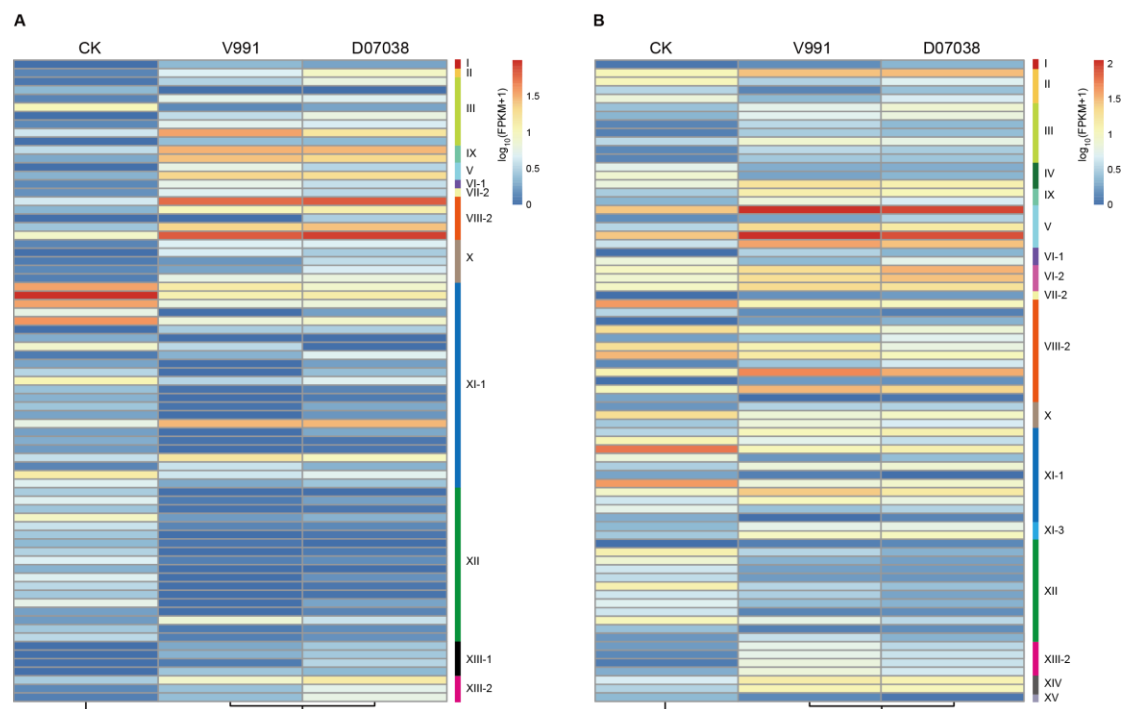


Figure 10. Expression patterns of differentially expressed LRR-RLK genes in response to *Verticillium dahliae* infection. For both *G. barbadense* (A) and *G. hirsutum* (B), the union sets of differentially expressed LRR-RLK genes detected in samples infected with *V. dahliae* strain V991 (highly toxic) and samples infected with *V. dahliae* strain D07038 (intermediately toxic) were used for heatmap drawing. The colored bars to the right of heatmaps indicated different LRR-RLK gene subfamilies.

4. Conclusions

The present study performed a comprehensive analysis of the large LRR-RLK gene family in four *Gossypium* species. The *Gossypium* LRR-RLK genes were classified into 21 distinct subfamilies. Subfamilies XI and XII were found dramatically expanded in *Gossypium*, tandem duplication were found act as important expansion mechanism in these expanded subfamilies. Functional and expression profiles analysis suggested that *Gossypium* LRR-RLK genes were involved in diverse developmental processes and stresses defense, the expansion of subfamily XI and XII could be associated with more complicated development and regulation process, and enhanced adaptability against various environment. Cis-acting regulatory elements analysis revealed that *Gossypium* LRR-RLK genes were extensively regulated by TFs and various abiotic and biotic stimuli. Our study provided valuable information for further function study of *Gossypium* LRR-RLK genes.

Supplementary Materials: Figure S1: Domain and exon-intron organization of identified LRR-RLK family members in *A. thaliana* and *Gossypium*, Figure S2: NJ tree constructed by MEGA 7 based on LRR-RLK family members of *A. thaliana* and *Gossypium*, Figure S3: Chromosomal

location of LRR-RLK genes from four *Gossypium* species, Table S1: Statistics of *A. thaliana* and *Gossypium* LRR-RLK genes distribution among different subfamilies, Table S2: Number of tandem duplication gene sets and involved genes in *Gossypium* LRR-RLK family, Table S3: Statistics of cis-acting regulatory elements found by PlantCARE in promoter regions of *Gossypium* LRR-RLK genes, Table S4: Statistics of TF binding sites predicted in promoter regions of *Gossypium* LRR-RLK genes.

Acknowledgements: This work was supported by National Science and Technology Major Project (2016YFD0101006).

Author Contributions: SUN Ruibin conceived the research, carried out the analysis and wrote draft manuscript. WANG Shaohui help prepare figures. Dr. MA Dan, WANG Shaohui and professor LIU Chuanliang supervised the research, and gave final approval of the version to be published. All the authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hubbard, S. R.; Miller, W. T. Receptor tyrosine kinases: mechanisms of activation and signaling. *Current Opinion in Cell Biology* **2007**, *19*, 117–123, doi:10.1016/j.ceb.2007.02.010.
2. Lemmon, M. A.; Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **2000**, *141*, 1117.
3. Shiu, S. H.; Bleecker, A. B. Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. *Plant Physiol* **2003**, *132*, 530–43, doi:10.1104/pp.103.021964.
4. Lehti-Shiu, M. D.; Cheng, Z.; Shiu, S. H. Origin, Diversity, Expansion History, and Functional Evolution of the Plant Receptor-Like Kinase/Pelle Family. *Signaling & Communication in Plants* **2012**, *13*, 1–22.
5. Lehti-Shiu, M. D.; Zou, C.; Hanada, K.; Shiu, S. H. Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. *Plant Physiol* **2009**, *150*, 12–26, doi:10.1104/pp.108.134353.
6. Walker, J. C. Structure and function of the receptor-like protein kinases of higher plants. *Plant Mol Biol* **1994**, *26*, 1599–609.
7. Shiu, S. H.; Bleecker, A. B. Plant receptor-like kinase gene family: diversity, function, and signaling. *Sci STKE* **2001**, *2001*, re22, doi:10.1126/stke.2001.113.re22.
8. Shiu, S. H.; Bleecker, A. B. Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci U S A* **2001**, *98*, 10763–8, doi:10.1073/pnas.181141598.

9. Kobe, B.; Kajava, A. V. The leucine-rich repeat as a protein recognition motif. *Current Opinion in Structural Biology* **2001**, *11*, 725.
10. Liu, P.-L.; Du, L.; Huang, Y.; Gao, S.-M.; Yu, M. Origin and diversification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in plants. *BMC Evolutionary Biology* **2017**, *17*, doi:10.1186/s12862-017-0891-5.
11. Clark, S. E.; Williams, R. W.; Meyerowitz, E. M. The CLAVATA1 gene encodes a putative receptor kinase that controls shoot and floral meristem size in Arabidopsis. *Cell* **1997**, *89*, 575–585, doi:10.1016/S0092-8674(00)80239-1.
12. Clavata, T.; Ganes, W.; Mayer, K. F. X.; Rens, G. J.; Laux, T. The Stem Cell Population of Arabidopsis Shoot Meristems Is Maintained by a Regulatory Loop between. *Cell* **2000**, *100*, 635–644.
13. Nam, K. H.; Li, J. BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. *Cell* **2002**, *110*, 203–12.
14. Li, J.; Wen, J.; Lease, K. A.; Doke, J. T.; Tax, F. E.; Walker, J. C. BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell* **2002**, *110*, 213–22.
15. Jinn, T. L.; Stone, J. M.; Walker, J. C. HAESA, an Arabidopsis leucine-rich repeat receptor kinase, controls floral organ abscission. *Genes Dev* **2000**, *14*, 108–17.
16. Nodine, M. D.; Yadegari, R.; Tax, F. E. RPK1 and TOAD2 are two receptor-like kinases redundantly required for arabidopsis embryonic pattern formation. *Dev Cell* **2007**, *12*, 943–56, doi:10.1016/j.devcel.2007.04.003.
17. Cyril, Z.; Gernot, K.; Delphine, C.; Anne, C.; Jones, J. D. G.; Thomas, B.; Georg, F. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. *Cell* **2006**, *125*, 749–760.
18. Gómez-Gómez, L.; Boller, T. FLS2 : An LRR Receptor-like Kinase Involved in the Perception of the Bacterial Elicitor Flagellin in Arabidopsis. *Molecular Cell* **2000**, *5*, 1003–1011.
19. Song, W. Y.; Wang, G. L.; Chen, L. L.; Kim, H. S.; Pi, L. Y.; Holsten, T.; Gardner, J.; Wang, B.; Zhai, W. X.; Zhu, L. H.; et al. A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* **1995**, *270*, 1804–6.
20. Albrecht, C.; Russinova, E.; Hecht, V.; Baaijens, E.; de Vries, S. The Arabidopsis thaliana SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASES1 and 2 control male sporogenesis. *Plant Cell* **2005**, *17*, 3337–49, doi:10.1105/tpc.105.036814.
21. Hecht, V.; Vielle-Calzada, J. P.; Hartog, M. V.; Schmidt, E. D.; Boutilier, K.; Grossniklaus, U.; de Vries, S. C. The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. *Plant Physiology* **2001**, *127*, 803.
22. Hu, H.; Xiong, L.; Yang, Y. Rice SERK1 gene positively regulates somatic embryogenesis of cultured cell and host defense response against fungal infection. *Planta* **2005**, *222*, 107–17, doi:10.1007/s00425-005-1534-4.
23. Gou, X.; Yin, H.; He, K.; Du, J.; Yi, J.; Xu, S.; Lin, H.; Clouse, S. D.; Li, J. Genetic evidence for an indispensable role of somatic embryogenesis receptor kinases in brassinosteroid signaling. *PLoS Genet* **2012**, *8*, e1002452, doi:10.1371/journal.pgen.1002452.

24. Kwaaitaal, M. A. C. J. SERK1 links somatic embryogenesis to brassinosteroid signalling. *Wur Wageningen Ur* **2007**.
25. Sun, X.; Wang, G. L. Genome-wide identification, characterization and phylogenetic analysis of the rice LRR-kinases. *PLoS One* **2011**, *6*, e16079, doi:10.1371/journal.pone.0016079.
26. Zan, Y.; Ji, Y.; Zhang, Y.; Yang, S.; Song, Y.; Wang, J. Genome-wide identification, characterization and expression analysis of populus leucine-rich repeat receptor-like protein kinase genes. *BMC Genomics* **2013**, *14*, 318, doi:10.1186/1471-2164-14-318.
27. Rameneni, J. J.; Lee, Y.; Dhandapani, V.; Yu, X.; Choi, S. R.; Oh, M. H.; Lim, Y. P. Genomic and Post-Translational Modification Analysis of Leucine-Rich-Repeat Receptor-Like Kinases in *Brassica rapa*. *PLoS One* **2015**, *10*, e0142255, doi:10.1371/journal.pone.0142255.
28. Wei, Z.; Wang, J.; Yang, S.; Song, Y. Identification and expression analysis of the LRR-RLK gene family in tomato (*Solanum lycopersicum*) Heinz 1706. *Genome* **2015**, *58*, 121–34, doi:10.1139/gen-2015-0035.
29. Fischer, I.; Dievart, A.; Droc, G.; Dufayard, J. F.; Chantret, N. Evolutionary Dynamics of the Leucine-Rich Repeat Receptor-Like Kinase (LRR-RLK) Subfamily in Angiosperms. *Plant Physiol* **2016**, *170*, 1595–1610, doi:10.1104/pp.15.01470.
30. 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. *Nature Genetics* **2018**, *50*, 796–802, doi:10.1038/s41588-018-0116-x.
31. Liu, X.; Zhao, B.; Zheng, H.-J.; Hu, Y.; Lu, G.; Yang, C.-Q.; Chen, J.-D.; Chen, J.-J.; Chen, D.-Y.; Zhang, L.; et al. *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Scientific Reports* **2015**, *5*, 14139.
32. Goodstein, D. M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R. D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **2012**, *40*, D1178–D1186, doi:10.1093/nar/gkr944.
33. Finn, R. D.; Coghill, P.; Eberhardt, R. Y.; Eddy, S. R.; Mistry, J.; Mitchell, A. L.; Potter, S. C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* **2016**, *44*, D279–D285, doi:10.1093/nar/gkv1344.
34. Finn, R. D.; Clements, J.; Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **2011**, *39*, W29–W37, doi:10.1093/nar/gkr367.
35. Berardini, T. Z.; Reiser, L.; Li, D.; Mezheritsky, Y.; Muller, R.; Strait, E.; Huala, E. The arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome: Tair: Making and Mining the “Gold Standard” Plant Genome. *genesis* **2015**, *53*, 474–485, doi:10.1002/dvg.22877.
36. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L. BLAST+: architecture and applications. *BMC Bioinformatics* **2009**, *10*, 421, doi:10.1186/1471-2105-10-421.

37. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240, doi:10.1093/bioinformatics/btu031.
38. Viklund, H.; Elofsson, A. Best α -helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. *Protein Science* **2004**, *13*, 1908–1917, doi:10.1110/ps.04625404.
39. Kall, L.; Krogh, A.; Sonnhammer, E. L. L. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Research* **2007**, *35*, W429–W432, doi:10.1093/nar/gkm256.
40. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **2004**, *32*, 1792–7, doi:10.1093/nar/gkh340.
41. Price, M. N.; Dehal, P. S.; Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **2010**, *5*, e9490, doi:10.1371/journal.pone.0009490.
42. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **2016**, *33*, 1870–1874, doi:10.1093/molbev/msw054.
43. 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**, doi:10.1101/289660.
44. Zhu, T.; Liang, C.; Meng, Z.; Sun, G.; Meng, Z.; Guo, S.; Zhang, R. CottonFGD: an integrated functional genomics database for cotton. *BMC Plant Biology* **2017**, *17*, doi:10.1186/s12870-017-1039-x.
45. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A. C.; Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **2007**, *35*, W182–W185, doi:10.1093/nar/gkm321.
46. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* **2012**, *16*, 284–287, doi:10.1089/omi.2011.0118.
47. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouzé, P.; Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327.
48. Jin, J.; Tian, F.; Yang, D.-C.; Meng, Y.-Q.; Kong, L.; Luo, J.; Gao, G. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research* **2017**, *45*, D1040–D1045, doi:10.1093/nar/gkw982.
49. Gu, Z.; Eils, R.; Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **2016**, *32*, 2847–2849, doi:10.1093/bioinformatics/btw313.
50. Kolde, R. pheatmap: Pretty Heatmaps. **2015**.
51. Chen, Z. J.; Scheffler, B. E.; Dennis, E.; Triplett, B. A.; Zhang, T.; Guo, W.; Chen, X.; Stelly, D. M.; Rabinowicz, P. D.; Town, C. D.; et al. Toward Sequencing Cotton (Gossypium) Genomes. *PLANT PHYSIOLOGY* **2007**, *145*, 1303–1310, doi:10.1104/pp.107.107672.
52. Li, F.; Fan, G.; Lu, C.; Xiao, G.; Zou, C.; Kohel, R. J.; 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. *Nature Biotechnology* **2015**, *33*, 524–530, doi:10.1038/nbt.3208.

53. 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. *Scientific Reports* **2015**, *5*, 17662.
54. Shaw, M. H.; Reimer, T.; Kim, Y.-G.; Nuñez, G. NOD-like receptors (NLRs): bona fide intracellular microbial sensors. *Current Opinion in Immunology* **2008**, *20*, 377–382, doi:10.1016/j.coi.2008.06.001.
55. Takeda, K.; Kaisho, T.; Akira, S. Toll-Like Receptors. *Annual Review of Immunology* **2003**, *21*, 335–376, doi:10.1146/annurev.immunol.21.120601.141126.
56. Quan, S.; Jiang, H.; Zhu, X.; Wang, W.; He, X.; Shi, Y.; Yuan, Y.; Du, X.; Cai, Y. Analysis of sea-island cotton and upland cotton in response to *Verticillium dahliae* infection by RNA sequencing. *Bmc Genomics* **2013**, *14*, 852–852.