

RESEARCH ARTICLE

Open Access



Genome-wide analysis of the gene families of resistance gene analogues in cotton and their response to *Verticillium* wilt

Jie-Yin Chen^{1†}, Jin-Qun Huang^{2†}, Nan-Yang Li^{1†}, Xue-Feng Ma¹, Jin-Long Wang¹, Chuan Liu², Yong-Feng Liu², Yong Liang², Yu-Ming Bao¹ and Xiao-Feng Dai^{1*}

Abstract

Background: *Gossypium raimondii* is a *Verticillium* wilt-resistant cotton species whose genome encodes numerous disease resistance genes that play important roles in the defence against pathogens. However, the characteristics of resistance gene analogues (RGAs) and *Verticillium dahliae* response loci (VdRLs) have not been investigated on a global scale. In this study, the characteristics of RGA genes were systematically analysed using bioinformatics-driven methods. Moreover, the potential VdRLs involved in the defence response to *Verticillium* wilt were identified by RNA-seq and correlations with known resistance QTLs.

Results: The *G. raimondii* genome encodes 1004 RGA genes, and most of these genes cluster in homology groups based on high levels of similarity. Interestingly, nearly half of the RGA genes occurred in 26 RGA-gene-rich clusters (Rgrcs). The homology analysis showed that sequence exchanges and tandem duplications frequently occurred within Rgrcs, and segmental duplications took place among the different Rgrcs. An RNA-seq analysis showed that the RGA genes play roles in cotton defence responses, forming 26 VdRLs inside in the Rgrcs after being inoculated with *V. dahliae*. A correlation analysis found that 12 VdRLs were adjacent to the known *Verticillium* wilt resistance QTLs, and that 5 were rich in NB-ARC domain-containing disease resistance genes.

Conclusions: The cotton genome contains numerous RGA genes, and nearly half of them are located in clusters, which evolved by sequence exchanges, tandem duplications and segmental duplications. In the Rgrcs, 26 loci were induced by the *V. dahliae* inoculation, and 12 are in the vicinity of known *Verticillium* wilt resistance QTLs.

Keywords: Cotton, *Verticillium* wilt-resistant, Resistance gene analogues, RGA-gene-rich clusters, *Verticillium dahliae* response loci

Background

Resistance (*R*) genes play a central role in recognising effectors from pathogens and in triggering downstream signalling during plant disease resistance [1, 2]. To date, more than 112 *R* genes and 104,310 putative *R*-genes present in a wide variety of plants species and conferring resistance to 122 pathogens [3]. The known *R* proteins can be grouped into several super-families based on the presence of a few structural motifs, including nucleotide-binding sites (NBSs), leucine-rich repeat (LRR) domains, Toll/

Interleukin-1 receptor (TIR) domains, coiled-coil (CC) domains and transmembrane (TM) regions [4, 5]. Generally, the most prevalent *R* genes in plants are of the NBS-LRR type, which are divided into two sub-classes based on the presence of an N-terminal CC or TIR domain [6, 7]. For example, 480 NBS-LRR proteins are encoded by the rice genome [8].

Previous studies demonstrated that many *R* genes are clustered in plant genomes [9]. To date, clusters of *R* genes have been reported in several plant genomes, including *Arabidopsis* [7], rice [10], soybean [11], *Lotus japonicus* [12], *Medicago truncatula* [13] and *Phaseolus vulgaris* [14]. In *Arabidopsis*, the genome was found to encode 159 NBS-LRR genes, and 113 of these genes

* Correspondence: daixiaofeng@caas.cn

†Equal contributors

¹Laboratory of Cotton Disease, Institute of Agro-Products Processing Science & Technology, Chinese Academy of Agricultural Sciences, Beijing 100193, China
Full list of author information is available at the end of the article

occurred in 38 clusters [15]. A similar phenomenon was also found in the rice genome, in which 76 % of the rice NBS-LRR genes was arranged in 44 gene clusters, with the others occurring as singletons [8]. The lengths of RGA gene clusters varied from dozens of kilobases (kb) to several megabases (Mb). For example, RGA genes were tightly linked to the *RPP5* cluster in *Arabidopsis*, which covers less than 100 kb [16], while the RGA genes were distributed over several Mb of the *RGC2* locus in lettuce [17]. Different *R* genes from the same cluster can confer resistance to different pathogens or to different variants of a single pathogen [18, 19]. For example, the *Cf-9* gene cluster contains two *Cf-9* and *Cf-9B* homologues that recognise the Avr9 and Avr9B effectors, respectively, in *Cladosporium fulvum*, and contribute to the resistance against tomato leaf mould disease. Other homologous genes in the cluster may serve as a reservoir of variation for the generation of *R* genes with new specificities [20–22].

Previous research suggested that the evolution of RGA clusters is usually mediated by sequence exchange, tandem duplication, segmental duplication, or gene conversion [9, 23, 24]. Frequent sequence exchanges tend to homogenize the members of a gene family, like the *RGC2* genes in lettuce [25], the *RI* cluster in *Solanum demissum*, and the *Cf-9* cluster in tomato [26, 27]. Tandem and segmental genomic duplications are also important in the evolution of RGA genes [23], which frequently occur in NBS-LRR genes clusters, and led to the formation of the phylogenetic lineage of NBS-LRR genes in the *Arabidopsis* genome [7, 28]. The evolution of the *HcrVf* cluster in apple was primarily dependent on gene duplication, with four *HcrVf* genes originating from a single progenitor gene by two sequential duplication events [29]. RGA's evolution by gene conversion resulted in high levels of sequence similarity, close physical clustering, and the local recombination rate [15, 28, 30]. In conclusion, the plants employed a complicated mechanism on the RGA genes evolution to response the variations of pathogens.

Cotton is an important crop worldwide because of its natural fibres and oil seeds. The cotton acreage in China has reached 4.69 million hectares, which produced 6.83 million tons of cotton in 2012 (Data from the National Bureau of Statistics in China). At present, Verticillium wilt caused by *Verticillium dahliae* is the most destructive disease of cotton, and the survival structures produced by pathogens may remain viable in the soil, persistently threatening crops, for more than 20 years [31]. In some years, more than 50 % of the cotton acreage is affected by Verticillium wilt, significantly reducing the fibre quality and resulting in yield losses (National Cotton Council of America Disease Database). Because of its unique ecological niche in the plant's vascular, Verticillium wilt is

difficult to control using fungicides, chemicals and cultivation measures [32]. Improving genetic resistance is considered the best method to overcome Verticillium wilt, and at least 80 different Verticillium wilt resistance quantitative trait loci (QTLs) have been reported in cotton [33–37]. However, *Gossypium hirsutum* appears to lack genetic resistance against *V. dahliae* [38, 39].

Gossypium barbadense, which is a cultivated tetraploid cotton species, showed resistance or tolerance to Verticillium wilt [40]. To date, the transcriptomes and proteomes of this Verticillium wilt-resistant cotton's responses to *V. dahliae* have been analysed, and phytoalexin biosynthesis and hormone signalling were found to have important roles in pathogen defense [41–46]. Moreover, several genes that contribute to the defence response against Verticillium wilt have been reported, including *GbCAD1*, *GbSSI2* [43], *GbRLK* [47], *GbSTK* [48], *GbTLP1* [49] and *GbVe/GbVe1* [50, 51].

Recently, the genome sequence of a diploid cotton, *Gossypium raimondii*, which is a Verticillium wilt-resistant wild relative of cotton, was completed [52, 53]. It is commonly thought that the tetraploid cotton species *G. hirsutum* and *G. barbadense* were derived from a cross between a D-genome species as the pollen-providing parent and an A-genome species as the maternal parent, and that *G. raimondii* is the putative D-genome parent [54, 55]. Previous research showed that the cotton genome encodes numerous NBS domains and that some of these genes formed gene clusters [53, 56]. A transcriptome analysis showed that some RGAs are involved in the defence response against *V. dahliae* [42, 46]. However, there are no systematic studies of RGA genes in the cotton genome, and the genetic resistance to Verticillium wilt is unclear.

In this study, a global analysis, including sequence features, gene distribution and the evolution of RGA genes in the *G. raimondii* genome was performed. High-throughput RNA-seq was used to identify the RGA genes' transcriptome in a *V. dahliae*-resistant cultivar of *G. barbadense* and to screen for potential *Verticillium dahliae* response loci (VdRLs) in the gene clusters. Moreover, the association between the VdRLs and Verticillium wilt resistance QTLs were analysed to screen the Verticillium wilt-response loci in cotton.

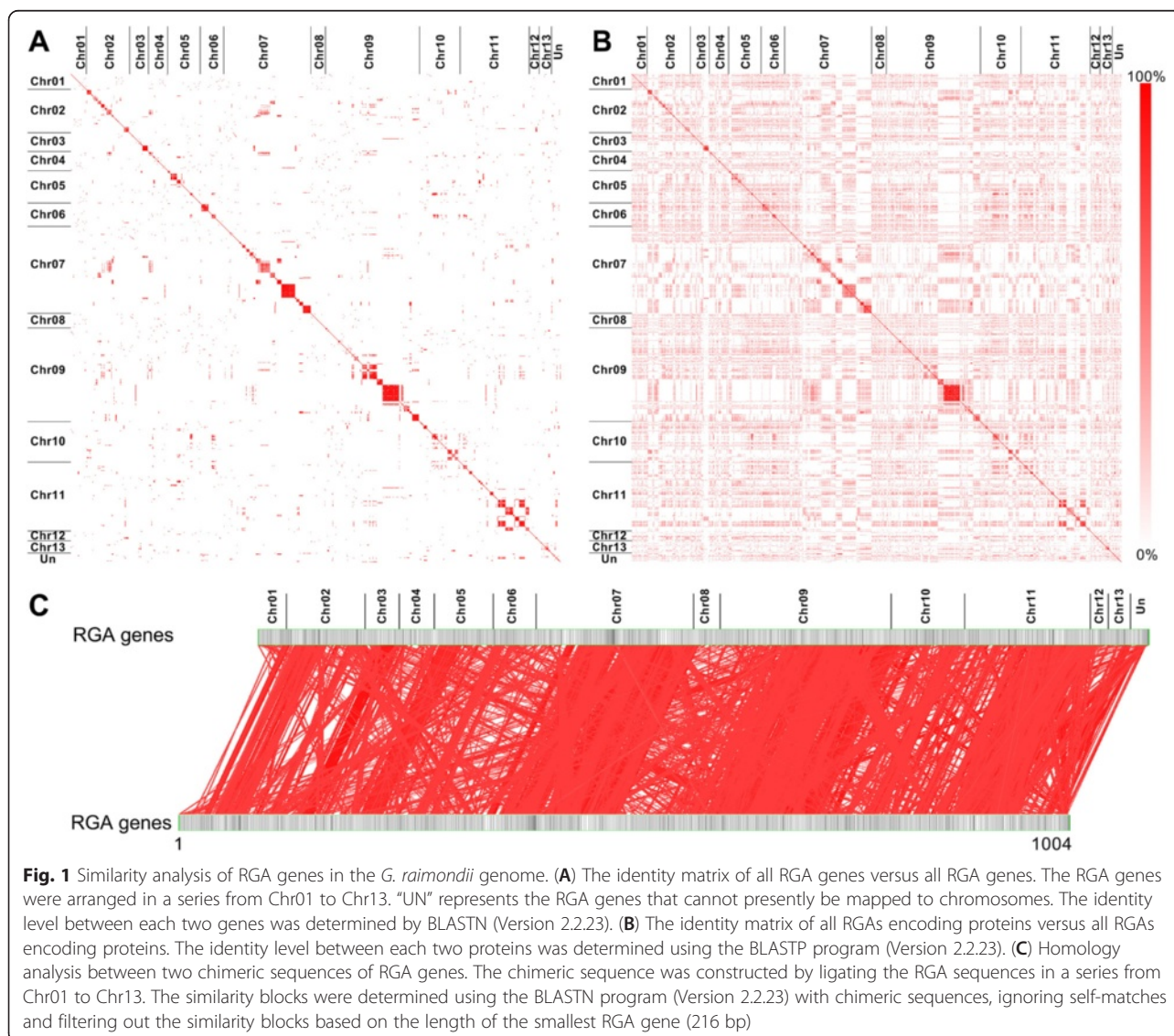
Results

Analysis of RGA genes in the *G. raimondii* genome

In this study, we focused on the RGA genes in the *G. raimondii* genome that probably participate in the disease resistance response. In total, 1004 RGA genes were classified into 11 families (R-I – R-XI) based on the integrated annotation of conserved motifs or domains in the *G. raimondii* genome [53]. The genome included 32 CC-NBS-LRR genes, 60 cysteine-rich receptor-like kinase (RLK) genes, 46 genes encoding disease resistance family

proteins/LRR family proteins, 58 genes encoding leucine-rich receptor-like protein kinase family proteins, 225 genes encoding LRR protein kinase family proteins, 44 genes encoding LRR receptor-like protein kinase family proteins, 78 genes encoding LRR transmembrane protein kinases, 79 genes encoding LRR and NB-ARC (Nucleotide-Binding adaptor shared by APAF-1, Resistance proteins and CED-4) domain-containing disease resistance proteins, 194 genes encoding NB-ARC domain-containing disease resistance proteins, 144 receptor-like proteins (RLP) genes and 44 TIR-NBS-LRR genes (Additional file 1: Table S1). A statistical analysis showed that more than half of the RGA genes were located on three chromosomes, with 194, 182 and 143 on Chr09, Chr07 and Chr11, respectively (Additional file 2: Figure S1). These results indicated that the cotton genome contains many RGA genes and numerous of them trend to enrich in several chromosome in cotton genome.

Generally, RGA genes contain conserved domains or motifs, such as NBSs and LRRs. In a comparative analysis, most of the RGA genes, and their encoded proteins, showed a high identity with one another (Fig. 1A, B), particularly RGA genes on Chr07 and Chr09, which shared high identities (up to 80 %) with one another (Additional file 1: Table S2). To investigate the correlation among all RGA genes, the similarity among RGA genes were compared according to the chimeric sequence which connected the RGA gene sequences from Chr01 to Chr13 in a series. Interestingly, the comparison of the chimeric sequence with itself showed a high similarity apart from small similarity blocks (less than the length of the smallest RGA gene, 216 bp) and self-match (Fig. 1C), indicating that many RGA genes are similar in the cotton genome. Moreover, the chimeric sequence segments from the same chromosome were



more similar than sequence segments from different chromosomes (Fig. 1C), indicating that RGA genes on the same chromosome were more closely related than genes on different chromosomes.

The homology clustering of RGA genes also indicated that RGA genes are conserved in cotton. Of the 1004 RGA genes, 974 could be divided into 45 homology groups (HG), with at least two genes in each HG, under the clustering conditions of match rate and identity being more than 33 % and 30 %, respectively. Of these, 838 were classified into 11 HGs, with HG13 containing the minimum 23 genes and HG17 containing the maximum 242 genes (Additional file 1: Table S3). Not surprisingly, most RGA genes in the same family could be clustered into a single HG based on a conserved feature. For example, five-sixths of the RGA genes in the R-II family were clustered into HG22. However, the genes of five RGA gene families were clustered into multiple groups, including R-I, R-V, R-VIII and R-IX. The RGA genes of the R-V family were clustered into two major HGs, HG17 and HG21 (Additional file 1: Table S3), indicating that the RGA gene families were not always clustered in one HG but could be clustered into different HGs. Moreover, the RGA genes could also be clustered into HGs using highly rigorous

conditions. The 306 RGA genes were divided into 104 HGs when the match rate and identity were more than 80 % for each gene (Additional file 2: Figure S2). The RGA genes in the same HGs are physically linked, such as 7 genes in the sub-HG of HG05 (HG05-04) that are closely linked in a small region that encodes 11 genes (Gorai.007G324100.1–Gorai.007G325100.1) (Additional file 1: Table S4). These results suggested that many RGA genes, which are probably multi-copy genes in cotton, are closely linked in the cotton genome.

The phylogenetic relationship analysis of RGA genes showed that most RGA genes could be arranged in clades in accordance with RGA gene families, such as R-II, R-III and R-IV (Fig. 2). These results also corresponded to the homology clustering, showing that the major HGs in an RGA gene family were arranged in a clade. For example, most R-II family genes were clustered into HG22, which was arranged in a single clade (Fig. 2; Additional file 1: Table S3). Although most of the R-V family genes could be arranged together in the phylogenetic tree, the R-V clade was split into three parts (Fig. 2), which indicated that variation occurred in the R-V family. More persuasive evidence showed four RGA gene families (R-I, R-VIII, R-IX and R-XI) which

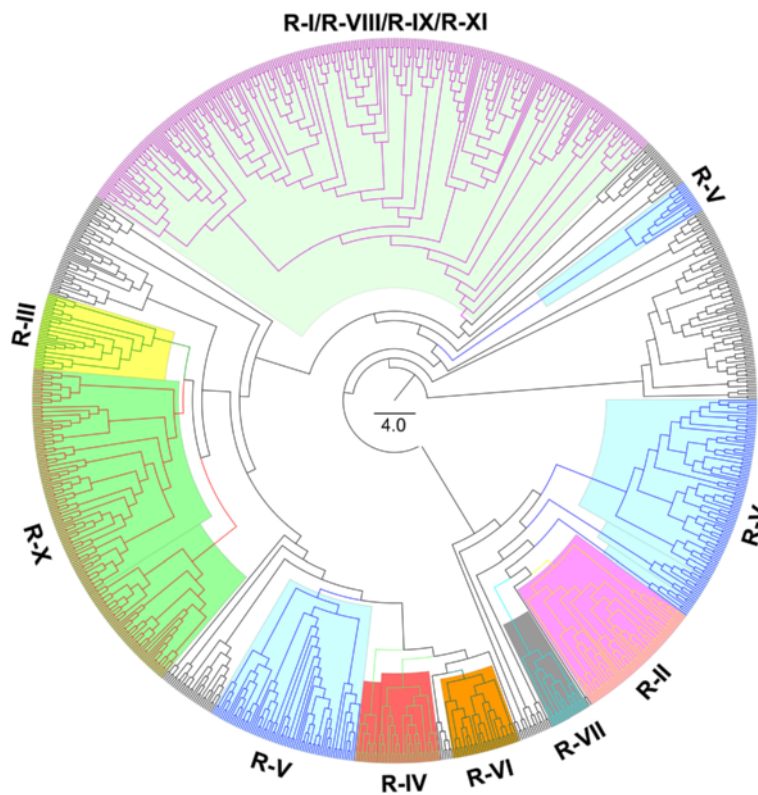


Fig. 2 Phylogeny analyses of RGA genes in the *G. raimondii* genome. The phylogenetic tree of RGA genes was constructed using the protein sequences by the neighbour-joining method, with 1000 bootstrap replicates. The branches of the mixed clade included four RGA gene families, which are marked in purple. Other conserved clades of RGA gene families are rendered in different colours

mainly contain the NBSs and LRRs domain were arranged in a mixed clade (Fig. 2). Together, these results indicated that the variation in RGA genes is as important as the conservation during the cotton genome's evolution.

Many RGA genes are deposited in gene clusters

In the *G. raimondii* genome, nearly half of the RGA genes were allocated to 26 Rgrcs (Fig. 3; Additional file 2: Figure S3). The total length of these Rgrcs is ~ 16.7 Mb, and there were 1148 genes, including 489 RGA genes. The average proportion of RGA genes in Rgrcs is significantly higher than in the whole genome, 42.6 % compared with 2.7 %. The average whole gene density was higher in Rgrcs (14.5 kb/gene) than in the whole genome (19.7 kb/gene) (Additional file 1: Table S5). Among these Rgrcs, Rgrc14 and Rgrc11 are the two largest clusters, which cover ~4.2 and 3.3 Mb, respectively, and contained 82 and 103 RGA genes, respectively (Additional file 1: Table S5). Most of the Rgrcs were located on Chr02, Chr07, Chr09, Chr10 and Chr11 (Fig. 3; Additional file 1: Table S5). Moreover, more than half of the RGA genes in the eight gene families occurred in these clusters, except those of RGA families R-IV, R-V and R-VII. Only 15.5 % of RGA genes in the R-V family occurred in Rgrc clusters (Additional file 1: Table S6). These results suggested that many RGA genes occur in gene clusters in the cotton genome.

To investigate how Rgrcs are related, all of the proteins encoded by Rgrcs were analysed using homology clustering. Clearly, most RGA genes are homologous to those clustered in the same HGs within the Rgrcs. This is also true for other genes in the Rgrcs that do not encode RGA genes, such as Rgrc2, Rgrc14 and Rgrc15. (Fig. 4). The homology of most genes within Rgrcs probably indicates that Rgrcs undergo tandem duplications

or sequence exchanges during their evolution. Moreover, most proteins encoded in different Rgrcs also clustered into same HGs (Fig. 4). Thus, the genes in different Rgrcs are homologous, indicating that some Rgrcs were probably generated from other Rgrcs by segmental duplications in cotton.

Homology analysis of the chimeric sequence, all the Rgrcs sequences connected in series from Chr01 to Chr13, showed that the Rgrcs was highly similar after apart from the small (less than the length of the smallest RGA gene, 216 bp) and self-matching similarity blocks (Additional file 2: Figure S4A). In total, 984 high similarity blocks in the chimeric sequence were matched to each other (up to 3 kb, ignoring self-match), except for the sequences of Rgrc4 and Rgrc20, and the identities of almost all the similarity blocks were close to 80 % (Additional file 2: Figure S4B/C). Of the similarity blocks, 589 belonged to "Rgrc-self-similarity", including 300 blocks within Rgrc14, and 78 blocks inside in Rgrc11 (Additional file 2: Figure S4B), indicating that the Rgrc sequences are similar by themselves, which could be the result of tandem duplication or sequence exchange. However, parts of the similarity blocks were also found among different Rgrcs, such as 42 matching blocks between Rgrc11 and Rgrc14, and 22 matching blocks between Rgrc11 and Rgrc24. (Additional file 2: Figure S4B), suggesting that some Rgrcs originated by segmental duplication in cotton.

RGA gene expression responses to *V. dahliae* infection

Analysis of RNA-seq data

In this study, *G. barbadense* cv. 7124, which is considered to be *V. dahliae*-resistant (Additional file 2: Figure S5), was inoculated with the highly aggressive defoliating *V. dahliae* strain Vd991. The inoculated root samples (2, 6, 12, 24, 48 and 72 h) were collected to identify differentially

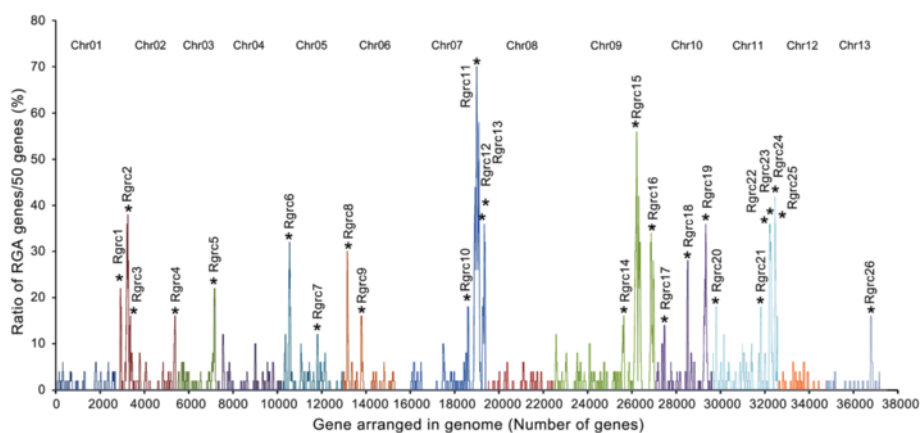


Fig. 3 The distribution of Rgrcs in the *G. raimondii* genome. All genes encoded by the *G. raimondii* genome were arranged in a series from Chr01 to Chr13. The ratio of RGA genes was calculated in the moving window (50 genes/window, walking forward 10 genes each time). RGA gene frequencies greater than 10 % were considered Rgrcs and clusters only containing 6 RGA genes in a window, but distributed evenly, were removed. The X-axis represents the number of genes in the cotton genome and the Y-axis represents the RGA gene ratio in the moving window

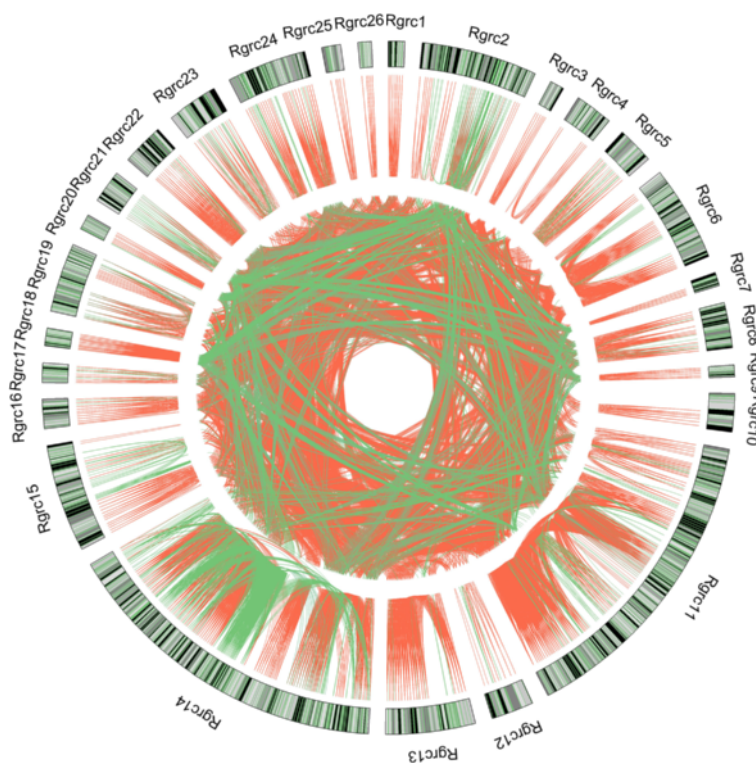


Fig. 4 Homology clustering of proteins encoded by genes in the Rgrcs of the *G. raimondii* genome. The homologous relationships were determined among proteins encoded by genes in the Rgrcs. The same homology groups of RGA genes are linked with red lines, while other genes in the same homology groups are linked with green lines. The outer ring represents the homology groups inside in Rgrcs, and the inner ring represents homology groups in different Rgrcs

expressed genes (DEGs) of RGAs using high-throughput RNA-seq. For extremely deep sequencing, ~200 million clean reads for each sample were generated, with quality control ($Q \geq 20$) (Additional file 1: Table S7). Of these reads, ~76 % matched the reference genome of *G. raimondii*, including ~140 million unique matched reads and ~13 million multi-position matched reads (Additional file 1: Table S7).

For DEG detection, the reads per exon kb per million mapped sequence reads (RPKM) was calculated for each gene and filtered using the false discovery rate (FDR) and with the p -value. In total, 28,360 DEGs were detected in the cotton genome at six inoculated time points, with 13,229 genes in common at different time points ($FDR < 0.001$, $p < 0.001$), 17,517 DEGs in all inoculated time points and 9811 genes in common ($FDR < 0.001$, $p < 0.001$, and $\log_2 \text{Ratio} \geq |1.0|$), 8122 DEGs in all inoculated time points and 5106 genes in common ($FDR < 0.001$, $p < 0.001$, and $\log_2 \text{Ratio} \geq |2.0|$) (Additional file 1: Table S8; Additional file 3: Table S9). The number of up-regulated DEGs peaked at 48 h after inoculation, and the number of down-regulated DEGs gradually decreased from 2 to 72 h (Additional file 2: Figure S6), which corresponded to the

important infection time point of 48 h in *V. dahliae*, for the penetration of hyphae into the roots was evident about two days [57–60].

DEGs of RGA genes

In the DEGs set, 723 RGA genes were induced in cotton inoculated with *V. dahliae*, with 319 RGA genes in common at six time points ($FDR < 0.001$, $p < 0.001$) (Additional file 1: Table S8). Real-time quantitative RT-PCR (qRT-PCR) showed that the fold-change of DEGs is reliable (Additional file 2: Figure S7). As with the DEGs in the whole genome, the DEGs of RGA genes were also obviously induced at 48 h after inoculation (Additional file 2: Figure S6). The statistical analysis of DEGs showed that all 11 RGA families could respond to the *V. dahliae* inoculation at all of the time points, although the proportion of DEGs in the RLP family was relatively small (Additional file 1: Table S10). These results suggested that RGA genes are involved in the cotton response to *V. dahliae*. The expression pattern analysis showed that RGA gene families that responded to *V. dahliae* could be classified into the early response stage (~2–12 h) and later response stage (~24–72 h). In the later response stage, the number of

RGA genes and their expression levels were induced more obvious than in the early response stage (Additional file 2: Figure S8). These results indicated that activating the later response stage is important to the resistant cotton plant's response to *V. dahliae*.

Many genes in the plant-pathogen interaction pathway are RGA genes, which play an important role in disease resistance. In this study, 451 differentially expressed RGA genes were induced in cotton inoculated with *V. dahliae*, and mapped to the plant-pathogen interaction pathway based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation (Fig. 5), including eight types of homologous genes, such as BAK1, FLS2 and EFR (Additional file 1: Table S11). Moreover, some genes homologous to signal factors in the plant-pathogen interaction pathway, which are not RGA genes, were also activated, such as protein kinases and transcription factors (Fig. 5). In addition, genes in the phytoalexin biosynthesis pathways, including those for phenylpropanoids, flavonoids and diterpenoids, were also induced in

cotton in response to *V. dahliae* (Additional file 2: Figure S9). Overall, the transcriptome results indicated that many RGA genes, which probably participated in the plant-pathogen interaction pathway and regulated the defence response, were induced in cotton.

DEGs in Rgrcs

The expression pattern analysis of DEGs in Rgrcs indicated that the RGA genes were up-regulated more often than other genes in Rgrcs (Additional file 2: Figure S10), which suggested that RGA genes were more sensitive to *V. dahliae* inoculation than the other genes in Rgrcs. To investigate the potential RGA gene responses to *V. dahliae* infection, highly rigorous conditions ($\log_2\text{Ratio} \geq |2.0|$, with more than one up-regulated post-infection time point) were used for screening in this study. In total, 168 differentially expressed RGA genes were identified as potential *Verticillium* wilt response genes. Of these genes, the proportion of potential *Verticillium* wilt resistance genes in R-II, R-III and R-IV families was

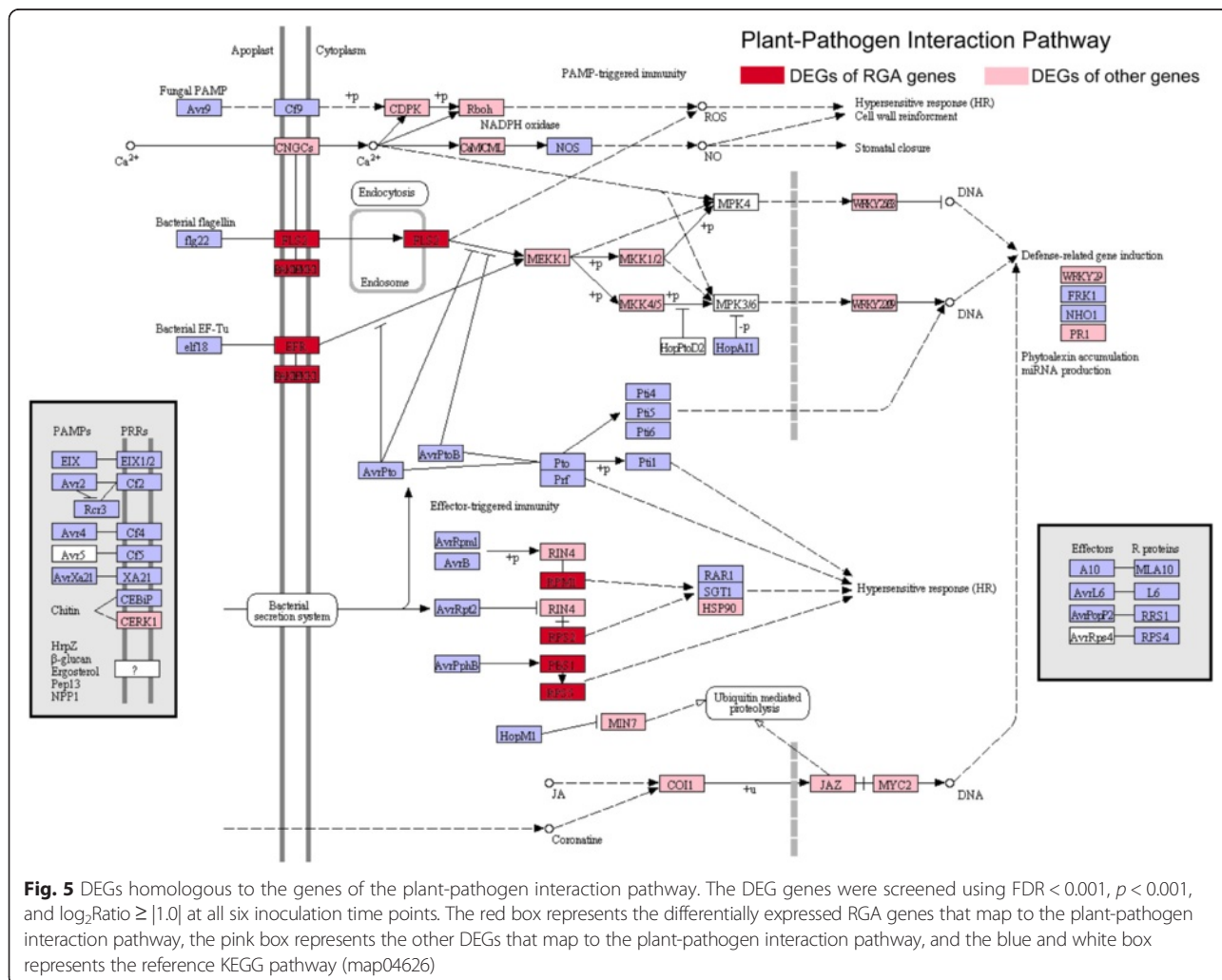


Fig. 5 DEGs homologous to the genes of the plant-pathogen interaction pathway. The DEG genes were screened using $FDR < 0.001$, $p < 0.001$, and $\log_2\text{Ratio} \geq |1.0|$ at all six inoculation time points. The red box represents the differentially expressed RGA genes that map to the plant-pathogen interaction pathway, the pink box represents the other DEGs that map to the plant-pathogen interaction pathway, and the blue and white box represents the reference KEGG pathway (map04626)

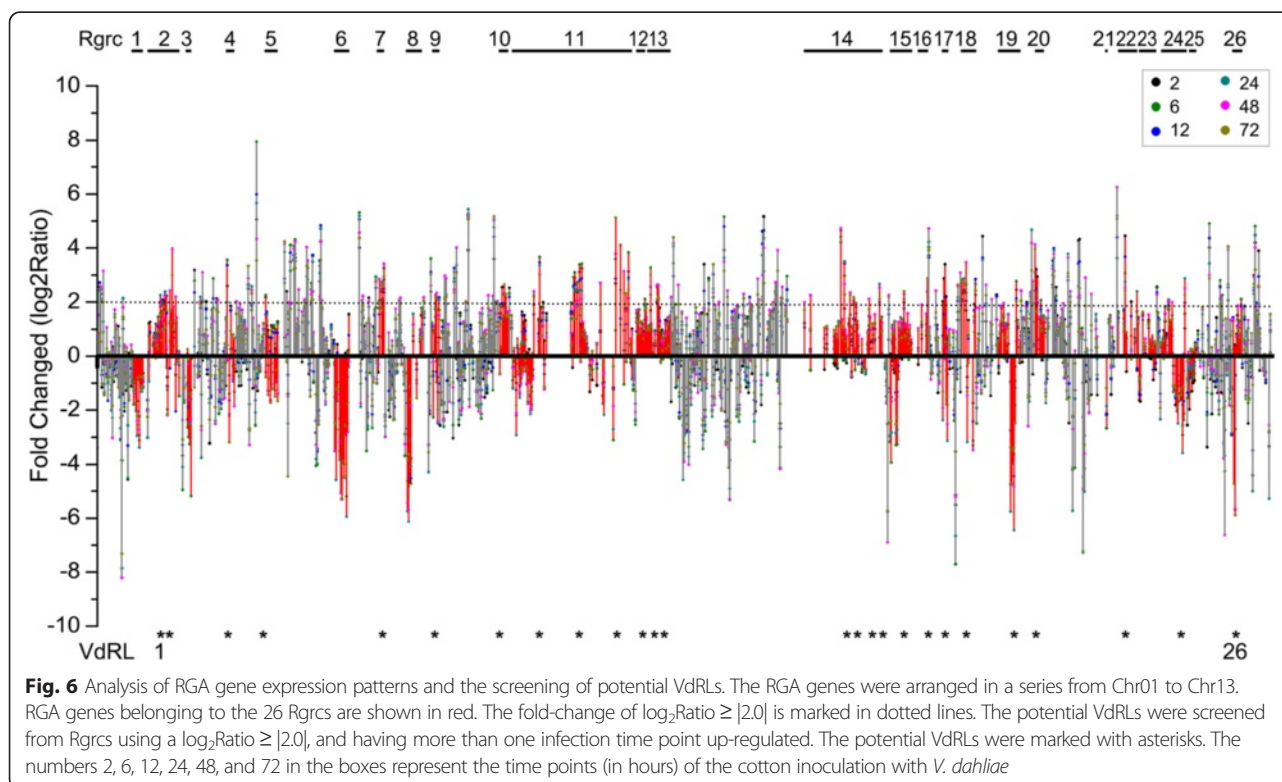
higher than in other families (Additional file 1: Table S12 and Table S13). Notably, 64 DEGs occurred in 19 Rgrcs, and 63 of them were distributed in the 26 small regions defined VdRL01 to VdRL26 (Fig. 6; Additional file 1: Table S12-S14). The total length of the VdRLs is ~2.4 Mb, and a minimum of 15 VdRLs contain at least two significantly differentially expressed RGA genes (Additional file 1: Table S14). A total of 39 differentially expressed RGA genes in the VdRLs belonged to the R-II, R-VII and R-IX families (Additional file 1: Table S12), indicating that these RGA genes were important to the cotton response to Verticillium wilt. Moreover, most VdRLs were primarily distributed in the small regions of a few chromosomes, particularly Chr07 and Chr09, which included seven and six VdRLs respectively (Additional file 1: Table S14). A further analysis showed that the RGA genes of nearly half of the VdRLs encoded NB-ARC domain-containing disease resistance proteins, and the RGA genes of the other VdRLs primarily encoded cysteine-rich RLKs, leucine-rich repeat protein kinase family proteins and RLPs (Additional file 1: Table S15). These results indicated that some RGA genes in the Rgrcs were strongly induced and a portion of them formed the VdRLs that participated in Verticillium wilt response in cotton.

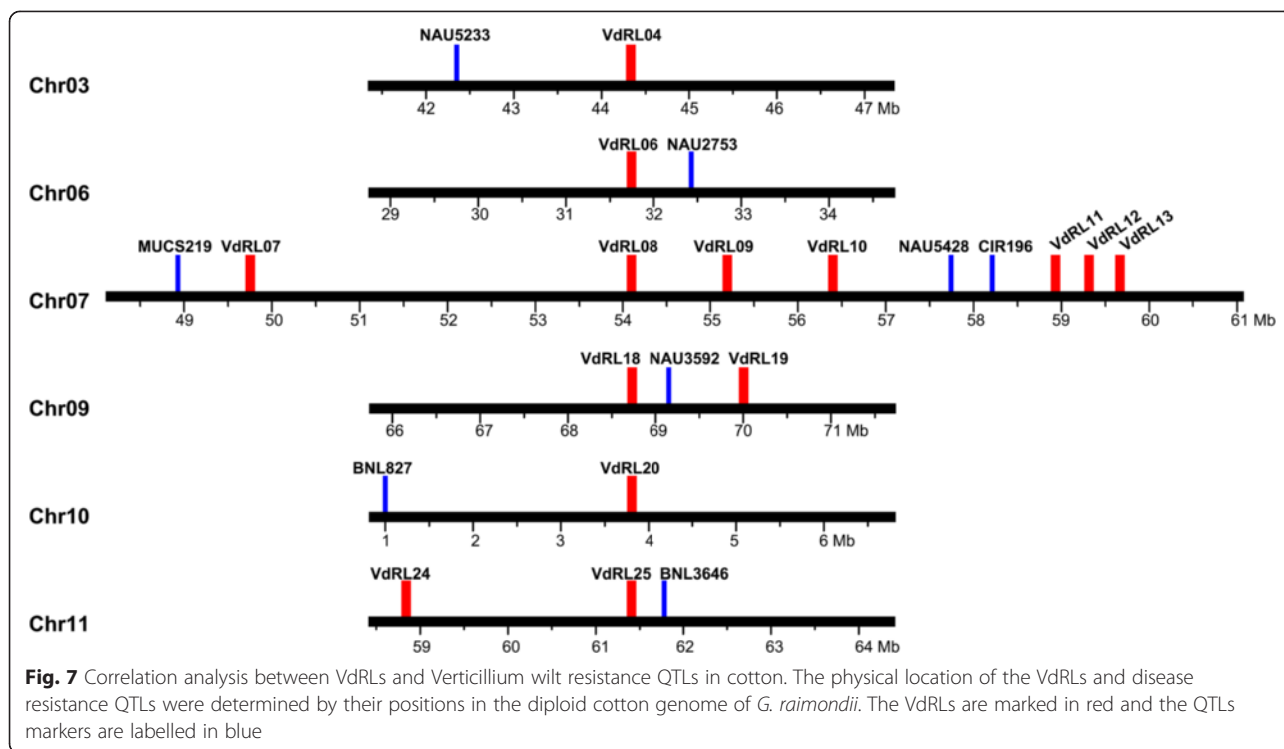
VdRLs adjacent to Verticillium wilt resistance QTLs

To detect the co-localization of VdRLs and QTLs, which had been identified to be associated with the Verticillium

wilt resistance in cotton [33–37], the locations of these QTLs in the diploid cotton genome were analysed based on the information provided by their corresponding markers. Among the 81 markers for these QTLs, 70 could be located on the diploid cotton genome (Additional file 1: Table S16), and 8 markers were adjacent to the VdRLs (Fig. 7; Additional file 1: Table S14). In total, 13 VdRLs were located on 6 chromosomes (3, 6, 7, 9, 10 and 11) with a physical distance of less than 3 Mb to the closest QTL marker, and 6 of them (VdRL06, VdRL07, VdRL11, VdRL18, VdRL19 and VdRL25) were less than 1 Mb from the closest marker (Fig. 7; Additional file 1: Table S14), suggesting that these VdRLs were positively correlated with the Verticillium wilt response. Moreover, the RGA genes in five VdRLs (VdRL07, VdRL11, VdRL12, VdRL13 and VdRL18) encoded NB-ARC domain-containing disease resistance proteins, of which three (VdRL07, VdRL11 and VdRL18) were close to Verticillium wilt resistance QTLs (Additional file 1: Table S14 and Additional file 1: Table S15).

Interestingly, six VdRLs (VdRL07 and VdRL09-VdRL13) located on Chr07 were found close to three Verticillium wilt resistance QTL markers (with a physical distance of less than 3 Mb), MUCS219, NAU5428 and CIR196 (Fig. 7; Additional file 1: Table S14). This region, in fact, extends about 10 Mb, which includes Rgrc10 and Rgrc11, and contains seven VdRLs (VdRL07-VdRL13). The physical distance between VdRL08 and the closest marker is ~3.66 Mb





(Fig. 7; Additional file 1: Table S14). Of these seven VdRLs on Chr07, five were enriched for the NB-ARC domain-containing disease resistance genes, and two (VdRL07 and VdRL13) were close to the Verticillium wilt resistance QTLs (less than 1 Mb) (Fig. 7; Additional file 1: Table S14). Overall, these results suggested that the VdRLs located on Chr07, which mainly encoded NB-ARC domain-containing disease resistance proteins, were closely associated with Verticillium wilt resistance in cotton.

Discussion

Plants have evolved a complicated and effective innate immune system to recognise, or respond to, many pathogenic organisms using *R* genes [1, 2]. At present, many *R* genes have been cloned from plants, and they can be divided into at least five classes based on conserved structural motifs, such as NBSs, LRRs and TIRs [4, 6]. In recent years, more than 20 plant genomes have been sequenced, and ~37,000 RGA genes were predicted based on conserved structural motifs [61]. Clearly, an analysis of the RGA genes in the genome will be useful for speculating on *R* gene evolution and for applying RGAs in cotton breeding. Recently, the genome of a diploid, *G. raimondii*, which is a Verticillium wilt-resistant wild relative of cotton, was sequenced [52, 53]. In this study, all probable RGA genes encoded by the *G. raimondii* genome were systematically analysed, and potential Verticillium wilt resistance loci/genes were identified using the bioinformatics analysis of transcriptome and QTL data.

In the *G. raimondii* genome, at least 300 genes encode NBS domains and most of these genes are of the CC-NBS or CC-NBS-LRR type [53, 56]. In this research, 1004 RGA genes were found in the *G. raimondii* genome based on an integrated annotation, and they were primarily distributed in Chr07, Chr09 and Chr11 (Additional file 2: Figure S1; Additional file 1: Table S1). As expected, the RGA genes showed a high similarity amongst themselves based on their conserved structural motifs, particularly when they occurred in small genomic regions of the same chromosome (Fig. 1, Additional file 1: Table S2). In contrast, some RGA genes in different families also showed similarities and were of the same phylogenetic lineage (Figs. 1 and 2). These results may indicate that the evolution of RGA genes in cotton had the dual characteristics of conservation and genetic variation, as did *RGC2* genes in lettuce [25]. RGA genes residing in clusters has been observed in many plant genomes [7, 10–14]. In *Arabidopsis thaliana*, more than 71 % of the NBS-LRR genes are arranged in 38 clusters [15], and the same characteristic is true of NBS-LRR genes in the rice genome [8]. As in other plants, the RGA genes in the *G. raimondii* genome reside in clusters (Fig. 3; Additional file 2: Figure S3; Additional file 1: Table S6). Previous studies have shown that the clustering of RGA genes is usually caused by tandem duplications [7, 62–64] or sequence exchanges [9], which have been detected in many RGA gene clusters [17, 19, 26, 65–67]. Similar results were found in the *G. raimondii* genome, where most of the RGA

genes are homologous and linked together to form the Rgrcs (Additional file 2: Figure S2; Additional file 2: Figure S4; Additional file 1: Table S4), indicating that tandem duplication or sequence exchanges could have occurred frequently in the evolution of RGA genes or Rgrcs. Segmental duplication is another evolutionary mechanism in RGA genes that could randomly translocate the genes in chromosomes, giving rise to a substantial number of RGA genes [9, 28, 68]. This was also found in our analysis (Additional file 2: Figure S4B), probably suggesting that the segmental duplication could happen in the RGA genes evolution. Together, these results probably indicated that tandem duplication, sequence exchange, and segmental duplication are important to the evolution of RGA genes and Rgrcs.

Verticillium wilt is the most destructive disease in cotton, and there are no effective methods to prevent this disease at present. Although improving genetic resistance is the direct method to combat Verticillium wilt, it has not been successful in *G. hirsutum*, which accounts for more than 90 % of the total cotton acreage in the world, because of the lack of genetic resistance [38]. *G. barbadense* is considered to be a resistant species, and many studies regarding Verticillium wilt resistance have been reported [36, 43, 47–51]. Recently, a transcriptome analysis showed that some RGA genes were induced in *G. barbadense* inoculated with *V. dahliae* [42, 46], indicating that the RGA genes contribute to the defence response in *G. barbadense*. In this study, the RGA genes in the cotton response to *V. dahliae* were analysed using RNA-seq. To overcome problems caused by the complicated genome and high identities between RGAs, an extremely deep RNA-seq strategy was applied in this study to produce reliable DEG screening (Additional file 1: Table S7). The results showed that more DEGs were identified in this study compared with previous studies on *G. barbadense* infected with *V. dahliae* (Additional file 1: Table S8; Additional file 2: Figure S6) [42, 46], which suggests that deep sequencing is useful for the transcriptome analysis of cotton and particularly for the analysis of homologous genes. However, it must point out that the DEGs also possibility reflect diurnal or developmental regulation for various times inoculated samples compared with a single mock-inoculated sample in our experiment. qRT-PCR validation between the inoculated samples and their corresponding mock-inoculated controls is necessary for screening the Verticillium wilt response genes.

Plant genomes encode many RGA genes, and some of these genes are transcriptionally activated in the plant's defence against pathogens [42, 46, 69–73]. Investigating the DEGs revealed that several hundred RGA genes, which belonged to different gene families, were induced in our experiment (Additional file 1: Table S10), and

many of them were homologous to genes in the plant-pathogen interaction pathway (Fig. 5; Additional file 1: Table S11), which suggests that these RGA genes could participate in the defence response against Verticillium wilt. Moreover, the RGA genes strongly responded from 24 to 72 h (Additional file 2: Figure S8), which is an important infection stage in *V. dahliae* [57–59]. These results suggest that the expression of RGA genes is important to the defence response of Verticillium wilt resistance.

RGA genes that are distributed in gene clusters usually act as genetic resistance sources in plants [9, 74]. In the *G. raimondii* genome, the RGA genes in the Rgrcs were also induced, which most likely indicated that the RGA genes formed clusters that were involved in Verticillium wilt resistance (Fig. 6), similar to the resistance clusters in many other plants [75–78]. In this study, at least 26 potential VdRLs, which included 63 RGA genes, were found to be strongly induced in *G. barbadense*, and half of these loci were on Chr07 and Chr09 (Fig. 6; Additional file 1: Table S12–S14), which is consistent with a previous finding that VdRLs were mainly distributed on Chr07 and Chr09 in upland cotton [36]. Among these VdRLs, half were enriched for NB-ARC domain-encoding RGAs (Additional file 1: Table S15), which are involved in a variety of processes, including apoptosis, transcriptional regulation and effector-triggered immunity [79, 80]. Moreover, some RGAs that clustered in several VdRLs are homologous to pattern recognition receptors (Fig. 5; Additional file 1: Table S15), which suggests that the VdRLs, like cysteine-rich RLKs and receptor-like proteins, participate in PAMP-triggered immunity [2, 81, 82]. These results suggested that the mechanisms of cotton resistance to *V. dahliae* are complicated and require the participation of multiple RGAs or loci for cotton Verticillium wilt resistance.

To date, at least 80 different Verticillium wilt resistance QTLs have been reported in cotton [33–37]. With the bioinformatics analysis of the RGA's distribution and expression after *V. dahliae* inoculation, at least 26 VdRLs were regarded as potential Verticillium wilt-response loci (Fig. 6). Interestingly, a correlation analysis showed that 12 VdRLs were less than 3 Mb (6 VdRLs were less than 1 Mb) from the closest Verticillium wilt resistance QTL, and 5 were of the NB-ARC gene cluster type (Fig. 7; Additional file 1: Table S14). An association analysis between disease resistance QTLs and NBS genes found that at least 32 NBS-encoding genes were adjacent to disease resistance QTLs in cotton [56], and there were similar results in other crops [56, 83–85]. Six of the VdRLs adjacent to Verticillium wilt resistance QTLs were located on the short region of Chr07 (Fig. 7; Additional file 1: Table S14), which again indicated that Verticillium wilt resistance QTLs clustered on chromosome D7 in cotton [36]. These results will be

beneficial for understanding the VdRLs in cotton and cloning the Verticillium wilt resistance gene.

Conclusions

In this study, the characteristics of RGA genes encoded in the *G. raimondii* genome were analysed, including the sequence structure, gene distribution and evolution. The *G. raimondii* genome encodes 1004 RGA genes, of which most are highly similar and could be clustered in HGs. Nearly half of the RGA genes occurred in 26 Rgrcs. Interestingly, many RGA genes are homologous, which results in most Rgrc sequences having a high similarity, indicating that sequence exchanges and tandem duplications frequently occurred in the evolution of RGA genes or Rgrcs. Moreover, the similarity among different Rgrcs suggests that some clusters may have evolved by segmental duplication. The RNA-seq analysis of the resistant cultivar *G. barbadense* showed that approximately half of the RGA genes were significantly induced by *V. dahliae* infection, and the portion of the RGA genes that formed 26 VdRLs in the Rgrcs were most likely involved in the Verticillium wilt response. A correlation analysis found that 12 VdRLs were adjacent to Verticillium wilt resistance QTLs, which strongly suggested that these loci respond during Verticillium wilt resistance in cotton.

Methods

Bioinformatics of RGA genes

Based on the integrated annotation of the *G. raimondii* reference genome from the DOE Joint Genome Institute (Cotton D V2.0, <ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Graimondii/>) [53], there were 11 classified RGA gene families, CC-NBS-LRR, cysteine-rich RLK, disease-resistance family protein/LRR family protein, leucine-rich receptor-like protein kinase family protein, LRR protein kinase family protein, LRR receptor-like protein kinase family protein, LRR transmembrane protein kinase, LRR and NB-ARC domain-containing disease resistance protein, NB-ARC domain-containing disease resistance protein, RLP and TIR-NBS-LRR.

The distribution of RGA genes in the *G. raimondii* genome was characterized by the number of RGA genes in the moving window (50 genes/window, walking forward 10 genes each time). The windows with RGA gene ratios that were greater than 10 % (considered Rgrcs) were collected and clusters only containing 6 RGA genes but distributed evenly were removed. Finally, the length of the Rgrcs was manually calculated based on the distribution of the RGA genes.

BLASTN and BLASTP programs (Version 2.2.23) were used to analyse the identities of the RGA genes ($e \leq 1e-10$), using the best hit results for each RGA gene. The filtered results were used to construct an RGA gene

matrix (total RGA genes versus total RGA genes) with a Perl script.

For the similarity analysis of RGA genes, a chimeric sequence was constructed by connecting RGA gene sequences in a series from Chr01 to Chr13. The similarities between segments of the chimeric sequences were analysed using the BLASTN program (Version 2.2.23), then small similarity blocks (less than the length of the smallest RGA gene, 216 bp) and the self-matching similarity blocks were removed. The homology between segments of the chimeric sequence was displayed using the ACT software [86]. The homology analysis of Rgrcs was performed using the same method, except similarity blocks less than 3 kb in length were filtered out.

In homology clustering, the reciprocal blast analysis of the proteins encoded by RGA genes (or encoding gene in Rgrcs) were conducted using the BLASTP program (Version 2.2.23) ($e \leq 1e-7$). The clustering of gene families was performed as previously described [87] and the software Solar (Version 0.9.6) was used to remove redundant members (match rate < 33 % or identities < 30 %). Three other rigorous conditions (match rate < 70 % and identities < 70 %, match rate < 80 % and identities < 80 %, and match rate < 90 % and identities < 90 %) were also used for high homology analyses. The software hcluster_sg (Version 0.5.0) was used for gene family clustering. The homologous relationships among genes in Rgrcs were depicted using the Circos program (Version 0.64) [88].

To construct the phylogenetic tree of RGA genes, the MUSCLE program (Version 3.8.31) was applied to create multiple alignments of protein sequences [89]. The unrooted tree was generated using the TreeBeST program (Version 1.9.2) by the neighbour-joining method, with 1000 bootstrap replicates [90].

Plant material and *V. dahliae* infection procedures

The resistant cultivar 7124 of *G. barbadense* L. was used as the experimental material. Cotton seeds were sown on commercial sterilised soil at 28 °C with a photoperiod of 14 h light/10 h dark for two weeks. Inoculations were performed using the high virulence V991 defoliating strain of *V. dahliae*. The strain was cultured on a potato-dextrose agar plate at 25 °C for one week. Spores were harvested from plates by eluting with sterile distilled water, then filtering through four layers of gauze and adjusted to 5×10^6 spores/ml with sterile distilled water. The cotton two-week-old seedlings were inoculated with *V. dahliae* using the root dip method. Seedlings were gently uprooted, rinsed in sterile water, inoculated into a spore suspension for 10 min, and then returned to new pots containing sterilised soil. Six individual seedling roots were collected at six time points, 2, 6, 12, 24, 48 and 72 h after inoculation. Control plants were treated with sterile distilled water in the same way, and roots samples were immediately

collected. All samples were immediately thrown into liquid nitrogen and stored at -80°C until further analysis.

Illumina sequencing

Total RNA was isolated from the root samples using an RNA kit according to the manufacturer's instructions (EASYspin for plant RNA, Beijing, China). The seven RNA samples, including the samples from the six inoculation time points and the mock-inoculated, were used for RNA-seq. RNA samples were digested with DNase I (Qiagen, Hilden, Germany), and the quality and quantity were determined using a NanoDrop 2000 (Thermo Scientific, NH, USA) and an Agilent 2100 (Agilent, Santa Clara, CA, USA) instrument. RNA of each sample was purified using oligo(dT)-attached magnetic beads from an mRNA-Seq Sample Prep Kit (Illumina, San Diego, CA, USA). The purified mRNA was used for preparing a non-directional Illumina RNA-seq library using a Small RNA Sample Prep Kit (Illumina, San Diego, CA, USA). The library's quality and quantification were analysed using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) and an ABI Step One Plus Real-Time PCR System (ABI, CA, USA). Each library was applied to an Illumina HiSeq 2000 (Illumina, San Diego, CA, USA) for single-end sequencing by the Beijing Genomics Institute (Shenzhen, China). Raw sequences were transformed into clean reads after data processing, leaving 49 nt tags.

Mapping of Illumina reads against the *G. raimondii* genome

The raw FASTQ format data sets were produced from the software CASAVA v1.8.2, with quality controls. Reads contaminated with Illumina adapters were detected and removed, and high-quality reads (Phred score ≥ 20) were collected for further analysis. The software SOAPaligner/SOAP2.0 [91] was used to map reads to the reference sequence of the *G. raimondii* genome (DOE Joint Genome Institute: Cotton D V2.0, <ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Graimondii/>) [53], with less than two mismatches allowed in the alignment.

Analysis of DEGs

The unique mapping read counts were normalised to RPKM, and the gene expression level was calculated using the RPKM method [92]. A strict algorithm was used to identify significant DEGs between mock-inoculated samples and inoculated samples. The FDR was set as 0.001 to determine the threshold of p -value (<0.001) in multiple tests, and the absolute value of \log_2 -Ratio was 1.0 [93]. The expression patterns were clustered using Cluster software [94]. The pathways were annotated based on the KEGG database [95] using BLASTX ($e \leq 1e-5$). KEGG mapper and iPath tools were used for the plant-pathogen interaction pathway and

the phytoalexin biosynthesis pathway analyses, respectively [95, 96].

Quantitative RT-PCR analysis

A qRT-PCR analysis was performed using a two-step Real-Time PCR system (ABI Biosystems, CA, USA). New treatment samples were collected at six time points of 2, 6, 12, 24, 48 and 72 h after inoculation and their corresponding mock-inoculated controls. First-strand cDNA synthesis was performed with $2.0 \mu\text{g}$ of purified total RNA using the Superscript Reverse Transcriptase (Invitrogen, CA, USA). Gene-specific primers for qRT-PCR were designed using the Primer3 software (<http://frodo.wi.mit.edu/primer3/>) (Additional file 1: Table S17). The constitutively expressed cotton 18S gene was used for normalisation. The expression levels of 15 RGA genes were analysed using qRT-PCR with a SYBR Green PCR Master Mix according to the manufacturer's instructions on an ABI 7500 Real Time PCR system (Applied Biosystems, CA, USA). The standard PCR cycles were as follows: 40 cycles at 95°C for 30 s, 60°C for 30 s, and 72°C for 15 s. Three technical replicates for each sample were performed, and the relative quantification of gene expression levels was determined using the comparative Ct method [97].

The correlation analysis between disease resistance QTL and VdRLs

The cotton Verticillium wilt resistance QTLs were retrieved from previous studies [33–37]. The primers and sequences of markers corresponding to these disease resistance QTLs were obtained from the Cotton Marker Database (<http://www.cottonmarker.org>). The physical locations of these QTLs in the diploid genome were determined by sequence mapping using PCR [98]. The physical distances between Verticillium wilt resistance QTLs and VdRLs in this study were calculated using their positions in the diploid cotton genome sequence mapping.

Availability of supporting data

All relevant supporting data can be found within the supplementary files accompanying to this article. The Raw RNA-seq data supporting the results of this article is available through the Sequence Read Archive under accession NO. SRP03537 at website: <http://www.ncbi.nlm.nih.gov/sra/?term=SRP035371>. Phylogenetic data supporting the results of this article are available in the TreeBASE repository, <http://purl.org/phylo/treebase/phylovs/study/TB2:S17448>.

Additional files

Additional file 1: Table S1. Statistics of RGA genes in the *G. raimondii* genome. R-I-R-XI represents the 11 RGA gene families. **Table S2.** Analysis of the identities of RGA genes in Chr07 and Chr09. **Table S3.** Homology clustering of RGA genes in the *G. raimondii* genome. **Table S4.** Information

regarding highly homologous genes screened by homology clustering. **Table S5.** Information on Rgrcs in the *G. raimondii* genome. **Table S6.** Statistical analysis of RGA genes in the Rgrcs. **Table S7.** Summary of sequencing yields and alignments. **Table S8.** Statistical analysis of DEGs in the *G. raimondii* genome and its RGA gene set. **Table S10.** Statistical analysis of DEGs in the 11 RGA gene families. **Table S11.** Information regarding differentially expressed RGA genes involved in the plant-pathogen interaction pathway. **Table S12.** Statistical analysis of potential DEGs in *G. barbadense* in response to *V. dahliae*. **Table S13.** Potential DEGs and VdRLs in *G. barbadense* in response to *V. dahliae*. **Table S14.** Information regarding VdRLs. **Table S15.** The RGA genes family enrichment in VdRLs. **Table S16.** Verticillium wilt resistance QTL information of cotton. **Table S17.** Primers used in this study.

Additional file 2: Supplementary figures. Figure S1. The statistics of RGA genes in *G. raimondii* chromosomes. The 11 families (R-I-R-XI) of RGA genes are marked in different colours. The X-axis represents the chromosomes of the *G. raimondii* genome (Chr01 to Chr13), and 'Others' represents the RGA genes that cannot be mapped to chromosomes at present. The Y-axis represents the number of genes. **Figure S2.** Homology clustering of RGA genes in the *G. raimondii* genome. Homology clustering was filtered using four conditions based on the match rates and identities among the RGA genes. Homology groups from HG01 to HG45 are arranged clockwise, the homology group intervals are differentiated by green and blue in series, according to the clustering conditions of match rate $\geq 33\%$ and identities $\geq 30\%$. **Figure S3.** A sketch map of coding genes in Rgrcs. The coding genes in the Rgrcs are marked with a red line based on the physical map of the *G. raimondii* genome. For the genetic structures of the Rgrcs, RGA genes are represented by red squares and other genes are represented by black squares. **Figure S4.** Homology analysis of Rgrcs in the *G. raimondii* genome. (A) Homology analysis of the Rgrcs' chimeric sequence. The chimeric sequence connected the Rgrc sequences in a series from Chr01 to Chr13 and was compared using the BlastN program (Version 2.2.23), ignoring self-matches and filtering out similarity blocks less than 3 kb in length. The forward-forward matches are marked with red lines, and the forward-reverse matches are marked with blue lines. (B) A statistical analysis of the similarity blocks among 26 Rgrcs. The lengths of the similarity blocks is greater than 3 kb. (C) Distribution of the identities of homology blocks. **Figure S5.** Cotton inoculated with *V. dahliae*. Two-week-old seedlings of the resistant cultivar *G. barbadense* cv. 7124 and the susceptible cultivar *G. hirsutum* cv. Jummian1 inoculated with the high virulence V991 defoliating strain of *V. dahliae* (5×10^6 spores/ml). The phenotypes were investigated three weeks after inoculation in this study. **Figure S6.** Statistical analysis of the DEGs in cotton inoculated with *V. dahliae*. The left side is the DEGs of all the genes in the *G. raimondii* genome and the right side is the RGA gene set. The X-axis represents numbers of DEGs. '12-172' represents the six inoculation time points (in hours). The numbers in the brackets from left to right represent the number of DEGs with more than a two-fold and a four-fold change, respectively, compared with mock-inoculated. Red represents up-regulation and green represents down-regulation. **Figure S7.** Differentially expressed RGA genes confirmed by qRT-PCR. In total, 15 differentially expressed RGA genes were randomly selected for qRT-PCR validation. The left side is the DEGs determined using RNA-seq and right side is the validation results using qRT-PCR. **Figure S8.** Expression pattern analysis of the response of 11 RGA gene families to *V. dahliae*. The filter conditions are $FDR < 0.001$ and $p < 0.001$. R-I-R-XI represents the 11 RGA gene families. '12-172' represents the six inoculation time points. **Figure S9.** Phytoalexin biosynthesis pathway of *G. barbadense* inoculated with *V. dahliae*. The DEGs used for the metabolism pathway analysis were screened by $FDR < 0.001$, $p < 0.001$, and $\log_2\text{Ratio} \geq |1.0|$ at all six inoculation time points. The thin lines represent the expression change of $\log_2\text{Ratio} \geq |1.0|$, and the thick lines represent the expression change of $\log_2\text{Ratio} \geq |2.0|$. **Figure S10.** Clustering of DEGs encoded in Rgrcs. (A) The expression pattern analysis of RGA genes in Rgrcs. (B) The expression pattern analysis of other genes, not encoding RGA genes in Rgrcs.

Additional file 3: Table S9. DEGs of *G. barbadense* inoculated with *V. dahliae*.

Abbreviations

RGA: Resistance Gene Analogue; Rgrc: RGA-gene-rich cluster; VdRL: *V. dahliae* response loci; HG: Homology groups; DEG: Differentially expression gene; QTL: Quantitative trait locus.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

D.X.F and C.J.Y conceived and directed the project. C.J.Y performed the genome data analysis. C.J.Y, H.J.Q, L.C, L.Y and L.Y.F performed RNA-seq data analysis. L.N.Y and M.X.F carried out the correlation analysis between disease resistance QTL and VdRLs. L.N.Y and W.J.L performed the qRT-PCR analysis. L.N.Y and B.Y.M prepared the DNA and RNA sample. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the Major State Basic Research Development Program of China (973 Program) (2011CB100705), the China Natural Scientific Foundation (No. 31,200,113), and the China Major Projects for Transgenic Breeding (2011ZX08005). The *G. barbadense* L. variety 7124 was supplied by National Medium-term Gene Bank of Cotton in China.

Author details

¹Laboratory of Cotton Disease, Institute of Agro-Products Processing Science & Technology, Chinese Academy of Agricultural Sciences, Beijing 100193, China. ²BGI-Shenzhen, Shenzhen, Guangdong 518083, China.

Received: 14 December 2014 Accepted: 27 April 2015

Published online: 19 June 2015

References

- Jones JD, Dangl JL. The plant immune system. *Nature*. 2006;444:323–9.
- Zipfel C. Pattern-recognition receptors in plant innate immunity. *Curr Opin Immunol*. 2008;20:10–6.
- Sanseverino W, Hermoso A, D'Alessandro R, Vlasova A, Andolfo G, Frusciantone L, et al. PRGdb 2.0: towards a community-based database model for the analysis of R-genes in plants. *Nucleic Acids Res*. 2012;41(D1):D1167–71.
- Martin GB, Bogdanove AJ, Sessa G. Understanding the functions of plant disease resistance proteins. *Annu Rev Plant Biol*. 2003;54:23–61.
- Joshi RK, Nayak S. Functional characterization and signal transduction ability of nucleotide-binding site-leucine-rich repeat resistance genes in plants. *Genet Mol Res*. 2011;10:2637–52.
- Dangl JL, Jones JD. Plant pathogens and integrated defence responses to infection. *Nature*. 2001;411:826–33.
- Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR encoding genes in *Arabidopsis*. *Plant Cell*. 2003;15:809–34.
- Zhou T, Wang Y, Chen JQ, Araki H, Jing ZQ, Jiang K, et al. Genome-wide identification of NBS genes in rice reveals significant expansion of divergent non-TIR NBS Genes. *Mol Genet Gen*. 2004;406:402–15.
- Hulbert SH, Webb CA, Smith SM, Sun Q. Resistance gene complexes: Evolution and utilization. *Annu Rev Phytopathol*. 2001;39:285–312.
- Bai J, Pennill LA, Ning J, Lee SW, Ramalingam J, Webb CR, et al. Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res*. 2002;12:1871–84.
- Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, et al. Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. *Plant Physiol*. 2008;148:1740–59.
- Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, et al. Genome structure of the legume, *Lotus japonicus*. *DNA Res*. 2008;15:227–39.
- Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu H, Roe B, et al. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol*. 2008;146:5–21.
- David P, Chen NW, Pedrosa-Harand A, Thareau V, Sévignac M, Cannon SB, et al. A nomadic subtelomeric disease resistance gene cluster in common bean. *Plant Physiol*. 2009;151:1048–65.
- Guo YL, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D. Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in *Arabidopsis*. *Plant Physiol*. 2011;157:757–69.
- Noël L, Moores TL, van Der Biezen EA, Parniske M, Daniels MJ, Parker JE, et al. Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis*. *Plant Cell*. 1999;11:2099–112.
- Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, et al. The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell*. 1998;10:1817–32.

18. Botella MA, Parker JE, Frost LN, Bittner-Eddy PD, Beynon JL, Daniels MJ, et al. Three genes of the *Arabidopsis* *RPP1* complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. *Plant Cell*. 1998;10:1847–60.
19. Ellis JG, Lawrence GJ, Luck JE, Dodds, PN: Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell*. 1999;11:495–506.
20. Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ, Jones JDG. Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science*. 1994;266:789–93.
21. Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, et al. Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf4/9* locus of tomato. *Cell*. 1997;91:821–32.
22. Laugé R, Dmitriev AP, Joosten MHAJ, De Wit PJGM. Additional resistance genes against *Cladosporium fulvum* present on the *Cf-9* introgression segment are associated with strong PR protein accumulation. *Mol Plant Microbe Interact*. 1998;11:301–8.
23. Leister D. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends Genet*. 2004;20:116–22.
24. Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, et al. *RPS2* of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science*. 1994;265:1856–60.
25. Kuang H, Woo S-S, Meyers BC, Nevo E, Michelmore RW. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell*. 2004;16:2870–94.
26. Van der Hoorn RA, Kruijt M, Roth R, Brandwagt BF, Joosten MH, De Wit PJ. Intragenic recombination generated two distinct *Cf* genes that mediate AVR9 recognition in the natural population of *Lycopersicon pimpinellifolium*. *Proc Natl Acad Sci U S A*. 2001;98:10493–8.
27. Kuang H, Wei F, Marano MR, Wirtz U, Wang X, Liu J, et al. The *R1* resistance gene cluster contains three groups of independently evolving, type I *R1* homologues and shows substantial structural variation among haplotypes of *Solanum demissum*. *Plant J*. 2005;44:37–51.
28. Baumgarten A, Cannon S, Spangler R, May G. Genome-level evolution of resistance genes in *Arabidopsis thaliana*. *Genetics*. 2003;165:309–19.
29. Xu M, Korban SS. Somatic variation plays a key role in the evolution of the *Vf* gene family in the *Vf* locus that confers resistance to apple scab disease. *Mol Phylogenet Evol*. 2004;32:57–65.
30. Mondragon-Palomino M, Gaut BS. Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. *Mol Biol Evol*. 2005;22:2444–56.
31. Klosterman SJ, Atallah ZK, Vallad GE, Subbarao KV. Diversity, pathogenicity, and management of *Verticillium* species. *Annu Rev Phytopathol*. 2009;47:39–62.
32. Kamal ME. Integrated control of *Verticillium* wilt of cotton. *Plant Dis*. 1985;69:1025–32.
33. Wang FR, Liu RZ, Wang LM, Zhang CY, Liu GD, Liu QH, et al. Molecular markers of *Verticillium* wilt resistance in upland cotton (*Gossypium hirsutum* L.) cultivar and their effects on assisted phenotypic selection. *Cotton Sci*. 2007;19:424–30.
34. Wang HM, Lin ZX, Zhang XL, Chen W, Guo XP, Nie YC, et al. Mapping and QTL analysis of *Verticillium* wilt resistance genes in cotton. *J inte PI Bio*. 2008;50:174–82.
35. Yang C, Guo W, Li G, Gao F, Lin S, Zhang T. QTLs mapping for *Verticillium* wilt resistance at seedling and maturity stages in *Gossypium barbadense* L. *Plant Sci*. 2008;174:290–8.
36. Jiang F, Zhao J, Zhou L, Guo WZ, Zhang TZ. Molecular mapping of *Verticillium* wilt resistance QTL clustered on chromosomes D7 and D9 in upland cotton. *Sci China C Life Sci*. 2009;52:872–84.
37. Zhao Y, Wang H, Chen W, Li Y. Genetic structure, linkage disequilibrium and association mapping of *Verticillium* wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. *PLoS One*. 2014;9(1):e86308.
38. Cai YF, He XH, Mo JC, Sun Q, Yang JP, Liu JG. Molecular research and genetic engineering of resistance to *Verticillium* wilt in cotton: a review. *Afr J Biotechnol*. 2009;8:7363–72.
39. Zhang J, Sanogo S, Flynn R, Baral JB, Bajaj S, Hughs SE, et al. Germplasm evaluation and transfer of *Verticillium* wilt resistance from Pima (*Gossypium barbadense*) to Upland cotton (*G. hirsutum*). *Euphytica*. 2011;187:147–60.
40. Wilhelm S, Sagen JE, Tietz H. Resistance to *Verticillium* wilt in cotton: sources, techniques of identification, inheritance trends, and the resistance potential of multiple cultivars. *Phytopath*. 1974;64:924–31.
41. Wang FX, Ma YP, Yang CL, Zhao PM, Yao Y, Jian GL, et al. Proteomic analysis of the sea-island cotton roots infected by wilt pathogen *Verticillium dahliae*. *Proteomics*. 2011;11:4296–309.
42. Xu L, Zhu LF, Tu LL, Liu LL, Yuan DJ, Jin L, et al. Lignin metabolism has a central role in the resistance of cotton to the wilt fungus *Verticillium dahliae* as revealed by RNA-Seq-dependent transcriptional analysis and histochemistry. *J Exp Bot*. 2011;62:5607–21.
43. Gao W, Long L, Zhu LF, Xu L, Gao WH, Sun LQ, et al. Proteomic and virus-induced gene silencing (VIGS) analyses reveal that gossypol, brassinosteroids, and jasmonic acid contribute to the resistance of cotton to *Verticillium dahliae*. *Mol Cell Proteomics*. 2013;12:3690–703.
44. Sun Q, Jiang HZ, Zhu XY, Wang WN, He XH, Shi YZ, et al. Analysis of sea-island cotton and upland cotton in response to *Verticillium dahliae* infection by RNA sequencing. *BMC Genomics*. 2013;14:852.
45. Zhang XY, Yao DX, Wang QH, Xu WY, Wei Q, Wang CC, et al. mRNA-seq analysis of the *Gossypium arboreum* transcriptome reveals tissue selective signaling in response to water stress during seedling stage. *PLoS One*. 2013;8:e54762.
46. Zhang Y, Wang XF, Ding ZG, Ma Q, Zhang GR, Zhang SL, et al. Transcriptome profiling of *Gossypium barbadense* inoculated with *Verticillium dahliae* provides a resource for cotton improvement. *BMC Genomics*. 2013;14:637.
47. Zhao J, Gao YL, Zhang ZY, Chen TZ, Guo WZ, Zhang TZ. A receptor-like kinase gene (*GbRLK*) from *Gossypium barbadense* enhances salinity and drought-stress tolerance in *Arabidopsis*. *BMC Plant Biol*. 2013;13:110.
48. Zhang Y, Wang XF, Li YY, Wu LZ, Zhou HM, Zhang GY, et al. Ectopic expression of a novel Ser/Thr protein kinase from cotton (*Gossypium barbadense*), enhances resistance to *Verticillium dahliae* infection and oxidative stress in *Arabidopsis*. *Plant Cell Rep*. 2013;32:1703–13.
49. Munis MF, Tu L, Deng FL, Tan JF, Xu L, Xu SC, et al. A thaumatin-like protein gene involved in cotton fiber secondary cell wall development enhances resistance against *Verticillium dahliae* and other stresses in transgenic tobacco. *Biochem Biophys Res Commun*. 2010;393:38–44.
50. Zhang BL, Yang YW, Chen TZ, Yu WG, Liu TL, Li HJ, et al. Island cotton *GbVe1* gene encoding a receptor-like protein confers resistance to both defoliating and non-defoliating isolates of *Verticillium dahliae*. *PLoS One*. 2012;7:e51091.
51. Zhang Y, Wang XF, Yang S, Chi JN, Zhang GY, Ma ZY. Cloning and characterization of a *Verticillium* wilt resistance gene from *Gossypium barbadense* and functional analysis in *Arabidopsis thaliana*. *Plant Cell Rep*. 2011;30:2085–96.
52. Wang KB, Wang ZW, Li FG, Ye WW, Wang JY, Song GL, et al. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet*. 2012;44:1098–103.
53. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*. 2012;492:423–7.
54. Sunilkumar G, Campbell LAM, Puckhaber L, Stipanovic RD, Rathore KS. Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proc Natl Acad Sci U S A*. 2006;103:18054–9.
55. Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, et al. Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol*. 2007;145:1303–10.
56. Wei HL, Li W, Sun XW, Zhu SJ, Zhu J. Systematic analysis and comparison of nucleotide-binding site disease resistance genes in a diploid cotton *Gossypium raimondii*. *PLoS One*. 2013;8:e68435.
57. Gold J, Robb J. The role of the coating response in *Craigella* tomatoes infected with *Verticillium dahliae*, races 1 and 2. *Physiol Mol Plant Pathol*. 1995;47:141–57.
58. Heinz R, Lee SW, Saparno A, Nazari RN, Robb J. Cyclical systemic colonization in *Verticillium*-infected tomato. *Physiol Mol Plant Pathol*. 1998;52:385–96.
59. Chen P, Lee B, Robb J. Tolerance to a non-host isolate of *Verticillium dahliae* in tomato. *Physiol Mol Plant Pathol*. 2004;64:283–91.
60. Zhao P, Zhao YL, Jin Y, Zhang T, Guo HS. Colonization process of *Arabidopsis thaliana* roots by a green fluorescent protein-tagged isolate of *Verticillium dahliae*. *Protein Cell*. 2014;5(2):94–8.
61. Kim J, Lim CJ, Lee BW, Choi JP, Oh SK, Ahmad R, et al. A genome-wide comparison of NB-LRR type of resistance gene analogs (RGA) in the plant kingdom. *Mol Cells*. 2012;33:385–92.
62. Michelmore RW, Meyers BC. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res*. 1998;8:1113–30.
63. Richly E, Kurth J, Leister D. Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. *Mol Biol Evol*. 2002;19:76–84.

64. Zhu H, Cannon SB, Young ND, Cook DR. Phylogeny and genomic organization of the TIR and non-TIR NBS-LRR resistance gene family in *Medicago truncatula*. *Mol Plant Microbe Interact*. 2002;15:529–39.
65. Song WY, Pi LY, Wang GL, Gardner J, Holsten T, Ronald PC. Evolution of the rice *Xa21* disease resistance gene family. *Plant Cell*. 1997;9:1279–87.
66. McDowell JM, Dhandaydham M, Long TA, Aarts MGM, Goff S, Holub EB, et al. Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of *Arabidopsis*. *Plant Cell*. 1998;10:1861–74.
67. Caicedo AL, Schaal BA, Kunkel BN. Diversity and molecular evolution of the *RPS2* resistance gene in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*. 1999;96:302–6.
68. Nobuta K, Ashfield T, Kim S, Innes RW. Diversification of non-TIR class NB-LRR genes in relation to whole-genome duplication events in *Arabidopsis*. *Mol Plant Microbe Interact*. 2005;18:103–9.
69. Li J, Zhang QY, Gao ZH, Wang F, Duan K, Ye ZW, et al. Genome-wide identification and comparative expression analysis of NBS-LRR-encoding genes upon *Colletotrichum gloeosporioides* infection in two ecotypes of *Fragaria vesca*. *Gene*. 2013;527:215–27.
70. Marathe R, Guan Z, Anandalakshmi R, Zhao H, Dinesh-Kumar SP. Study of *Arabidopsis thaliana* resistome in response to cucumber mosaic virus infection using whole genome microarray. *Plant Mol Biol*. 2004;55:501–20.
71. Coram TE, Wang M, Chen X. Transcriptome analysis of the wheat-*Puccinia striiformis* f. sp. *tritici* interaction. *Mol Plant Pathol*. 2008;9:157–69.
72. Joshi RK, Kar B, Nayak S. Survey and characterization of NBS-LRR (R) genes in *Curcuma longa* transcriptome. *Bioinformation*. 2011;6:360–3.
73. Bagnaresi P, Biselli C, Orrù L, Urso S, Crispino L, Abbruscato P, et al. Comparative transcriptome profiling of the early response to Magnaporthe oryzae in durable resistant vs susceptible rice (*Oryza sativa* L) genotypes. *PLoS One*. 2012;7:e51609.
74. Kruijt M, Kock MJ DE, de Wit PJ. Receptor-like proteins involved in plant disease resistance. *Mol Plant Pathol*. 2005;6:85–97.
75. Kawchuk LM, Hachey J, Lynch DR, Kulcsar F, van Rooijen G, Waterer DR, et al. Tomato *Ve* disease resistance genes encode cell surface-like receptors. *Proc Natl Acad Sci U S A*. 2001;98:6511–5.
76. Vinatzer BA, Patocchi A, Gianfranceschi L, Tartarini S, Zhang HB, Gessler C, et al. Apple contains receptor-like genes homologous to the *Cladosporium fulvum* resistance gene family of tomato with a cluster of genes cosegregating with *Vf* apple scab resistance. *Mol Plant Microbe Interact*. 2001;14:508–15.
77. Smith SM, Pryor AJ, Hulbert SH. Allelic and haplotypic diversity at the *rp1* rust resistance locus of maize. *Genetics*. 2004;167:1939–47.
78. Yue JX, Meyers BC, Chen JQ, Tian D, Yang S. Tracing the origin and evolutionary history of plant nucleotide binding site leucine rich repeat (NBS-LRR) genes. *New Phytol*. 2012;193:1049–63.
79. Takken FL, Albrecht M, Tameling WI. Resistance proteins: molecular switches of plant defence. *Curr Opin Plant Biol*. 2006;9:383–90.
80. Danot O, Marqueten E, Vidal-Ingigliardi D, Richet E. Wheel of life, Wheel of death: a mechanistic insight into signaling by STAND proteins. *Structure*. 2009;7:172–82.
81. Wrzaczek M, Brosché M, Salojärvi J, Kangasjärvi S, Idänheimo N, Mersmann S, et al. Transcriptional regulation of the CRK/DUF26 group of receptor-like protein kinases by ozone and plant hormones in *Arabidopsis*. *BMC Plant Biol*. 2010;10:95.
82. Thomma BPHJ, Nurnberger T, Joosten MHAJ. Of PAMPs and Effectors: The Blurred PTI-ETI Dichotomy. *Plant Cell*. 2011;23:4–15.
83. Madsen LH, Collins NC, Rakwalska M, Backes G, Sandal N, Krusell L, et al. Barley disease resistance gene analogs of the NBS-LRR class: identification and mapping. *Mol Genet Genomics*. 2003;269:150–61.
84. Bakker E, Borm T, Prins P, van der Vossen E, Uenk G, Arens M, et al. A genome wide genetic map of NB-LRR disease resistance loci in potato. *Theor Appl Genet*. 2011;123:493–508.
85. Kang YJ, Kim KH, Shim S, Yoon MY, Sun S, Kim MY, et al. Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. *BMC Plant Biol*. 2012;12:139.
86. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. ACT: the Artemis Comparison Tool. *Bioinformatics*. 2005;21:3422–3.
87. Zheng W, Huang L, Huang J, Wang X, Chen X, Zhao J, et al. High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nat Commun*. 2013;4:2673.
88. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639–45.
89. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
90. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. Ensembl/Comparative GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res*. 2009;19:327–35.
91. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*. 2009;25:1966–7.
92. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5:621–8.
93. Audic S, Claverie JM. The significance of digital gene expression profiles. *Genome Res*. 1997;7:986–95.
94. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics*. 2004;20:1453–4.
95. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*. 2012;40(Database issue):D109–14.
96. Letunic I, Yamada T, Kanehisa M, Bork P. iPath: interactive exploration of biochemical pathways and networks. *Trends Biochem Sci*. 2008;33:101–3.
97. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using Real-Time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods*. 2001;25:402–8.
98. Schuler GD. Sequence mapping by electronic PCR. *Genome Res*. 1997;7:541–50.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

