Science in China Series C: Life Sciences 2006 Vol.49 No.6 530—542

DOI: 10.1007/s11427-006-2017-y

Isolation and characterization of resistance and defense gene analogs in cotton (*Gossypium barbadense* L.)

GAO Yulong, GUO Wangzhen, WANG Lei & ZHANG Tianzhen

National Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China

Correspondence should be addressed to Zhang Tianzhen (email: cotton@njau.edu.cn)

Received November 21, 2005; accepted December 2, 2005

Abstract Plant disease resistance gene (*R* gene) and defense response gene encode some conserved motifs. In the present work, a PCR strategy was used to clone resistance gene analogs (RGAs) and defense gene analogs (DGAs) from Sea-island cotton variety Hai7124 using oligonucleotide primers based on the nucleotide-binding site (NBS) and serine/threonine kinase (STK) in the *R*-gene and pathogenesis-related proteins of class 2 (PR2) of defense response gene. 79 NBS sequences, 21 STK sequences and 11 DGAs were cloned from disease-resistance cotton. Phylogenic analysis of 79 NBS-RGAs and NBS-RGAs nucleotide sequences of cotton already deposited in GenBank identified one new sub-cluster. The deduced amino acid sequences of NBS-RGAs and STK-RGAs were divided into two distinct groups respectively: Toll/Interleukin-1 receptor (TIR) group and non-TIR group, A group and B group. The expression of RGAs and DGAs having consecutive open reading frame (ORF) was also investigated and it was found that 6 NBS-RGAs and 1 STK-RGA were induced, and 1 DGA was up-regulated by infection of *Verticillium dahliae* strain VD8. 4 TIR-NBS-RGAs and 4 non-TIR-NBS-RGAs were arbitrarily used as probes for Southern-blotting. There existed 2—10 blotted bands. In addition, since three non-TIR-NBS-RGAs have the same hybridization pattern, we conjecture that these three RGAs form a cluster distribution in the genome.

Keywords: cotton, resistance gene analogs (RGAs), defense gene analogs (DGAs).

Plants in their natural environments are constantly challenged by a wide spectrum of pathogens including viruses, bacteria, fungi, nematodes, protozoans and parasites, etc. The interactions of plants and pathogens lead to either compatible or incompatible reaction, an outcome shaped by long-term co-evolution of the plant-pathogen system. The phenotypic manifestation of an incompatible reaction is disease resistance, a consequence of a series of complex signal transduction controlled by two classes of genes, namely resistance genes (*R* gene) and defense response genes (*DR* gene).

Up to 2004, nearly 50 resistance genes had been cloned from a wide range of different crops with the use of techniques such as transposon tagging and map-based cloning[1]. Sequence analyses of the predicted protein sequences of these *R* genes, which were cloned from a number of plant species and shown to confer resistance to a wide range of pathogens, uncovered several conserved structural domains. *R* genes, on the basis of conserved structural domains, can be grouped into the following classes[2−4]: 1) *R* gene encoding proteins containing a nucleotide binding site

and a leucine-rich repeat (NBS-LRR); 2) *R* gene encoding proteins with a serine/threonine kinase (STK) domain; 3) *R* gene encoding product containing a extracytoplasmic leucine-rich repeat (eLRR) and a C-terminal membrane-anchoring signal etc. NBS class *R* genes, based on the existence of an N-terminal *Drosophila* Toll/mammalian interleukin-1 receptor (TIR) homology region, can be further subdivided into two groups: TIR-NBS-LRR- and non-TIR-NBS-LRR- types. Of all the cloned *R* genes, 71%[1] contain the conserved structural domain of NBS. It has been proposed that the structural domains of NBS and STK etc. are essential for specific recognition between an *R*-gene product and the product of its corresponding *Avr* gene, and ensuing defense responses in plants as well[5].

Specific recognition between products of a plant *R* gene and the corresponding *Avr* gene of the pathogen triggers a cascade of signal transduction. The elicitation of *R* gene-mediated plant defense responses plays a key role in terms of inhibiting growth, proliferation, and restricting spread of pathogen in plant tissues. Based on their biological properties, pathogenesis-related proteins (PR proteins) can be divided into several groups[6,7]: (1) glucanase; (2) chitinase; and (3) peroxidase. Glucanase and chitinase are capable of hydrolyzing the polysaccharide structures of the fungal cell walls, thus leading to the death of a fungal pathogen. Furthermore, products derived from degraded fugal cell wall structures, in turn, will serve as elicitors to trigger further host resistance response[8−10]. In addition to their roles in local induced resistance, glucanase and chitinase are also known to be involved in systemic acquired resistance, an induced defense mechanism that plants use to confer immunity or increased resistance to the second infection of the same pathogen and other unrelated pathogens as well. Plant peroxidase is thought to exert its antimicrobial role by oxidizing the phenolic residues of the plant cell wall components in infected tissues, thus restricting the spread of the pathogen in host plants.

The PCR-based approach provides a convenient and efficient way of cloning genes with conserved functional domains, especially from crops with complex genomes to which transposon tagging and map-based gene cloning strategy are not readily applicable. *R* genes and defense genes are known to contain conserved structural domains which can be utilized to design degenerate primers to amplify both resistance gene analogs (RGAs) and defense gene analogs (DGAs). The amplified DNA fragments may represent part of *R*-gene and defense gene sequences or sequences linked to genetic loci of these two classes of genes[11−15].

Cotton, a major source of cotton fiber for the textile industry, is an important economic crop. A variety of pathogens can infect cotton. In China, cotton blight and cotton verticillium wilt (causative pathogen *Verticillium dahliae*) are especially severe and cause serious damage to cotton production annually. It is well documented that planting disease resistant cultivars is the most effective approach to controlling the disease. *Gossypium hirsutum* is the major cotton species planted in China and breeding of blight-resistant cultivars has been very fruitful. However, breeding of cultivars resistant to *V. dahliae* has proven to be unsuccessful, largely due to the lack of germplasm known to be immune or highly resistant to the fungal pathogen *V. dahliae*. The successful cloning of RGAs and DGAs in various crops prompted us to clone RGAs and DGAs sequences from cotton. It is hoped that isolation and characterization of RGAs and DGAs in cotton should help identify DNA markers closely linked to the *R* genes in cotton, and ultimately lead to the mapping and cloning of cotton *R* genes. It is expected that the isolation and identification of *R* genes against cotton verticillium wilt should facilitate breeding and development of cotton cultivars which are highly resistant to this devastating pathogen.

1 Materials and methods

1.1 Extraction of DNA and RNA

Gossypium barbadense L. cv. Hai7124, a widely used verticillium wilt-resistant cultivar in China for genetic study and breeding, was employed for cloning RGAs and DGAs. Genomic DNA was extracted from fresh leaves of Hai7124 using a cetyltrimethy ammonium bromide (CTAB) method[16]. Seedlings of Hai 7124 with two simple and one heart-shaped leaves were inoculated with the VD8 strain of *V. dahliae* by injecting determined amounts of inoculum suspensions to the bottom of plastic cultivation pots[17], and roots were collected 0, 24, 48, 96, and 144 h after inocula-

tion and total RNA was extracted individually by a modified hot borate method[18].

1. 2 Primer design and PCR amplification

All primers corresponding to conserved domains of cloned *R* gene and β-1,3-glucanase(PR-2) listed in Table 1 for this study were based on published literatures except for Y782F and Y782R which were designed using a Primer 5 software program. Forward and backward primers used for PCR amplification of RGAs of NBS class were designed based on the P-loop and the hydrophobic domain of cloned *R* genes, and the predicted PCR product is about 500 bp. Primers used for PCR amplification of RGAs of STK class were designed based on the conserved motifs of I, II, VII, IX, and the expected size of the PCR fragments is about 500 bp. The expected size is about 700 bp for amplification of DGAs using the primer pair of Glu-S and Glu-AS.

The total volume for each PCR reaction was 30 μL containing 1× reaction buffer, 0.2 mmol/L dNTPs, 1.5 mmol/L $MgCL_2$, 50 ng genomic DNA, 1 μmol/L of respective forward and reverse primers, 2U Taq DNA polymerase. PCR amplification was performed in a PTC-200 Thermal cycler (MJ Research, USA). PCR reaction conditions: 94℃ for 5 min; 35 cycles of 94℃ for 30 s, 50—55℃ for 1 min, and 72℃ for 1 min; 7 min at 72℃.

1.3 Electrophoresis, recovery, cloning, and sequencing of PCR products

Amplified DNA fragments were separated on a 1.5% agarose gel, and recovered with a DNA Gel Extraction Kit (*V* gene). The obtained DNA fragments were ligated to pGEM-T easy Vector System (Promega) overnight, and the resultant of ligation was trans-

Table 1 Primers used for PCR amplification of RGAs and DGAs from cotton

Classification	Primers	Sequences(5′—3′)	Conserved amino acid motifs	Reference
RGAs				
NBS class				
	Y782(F)	ATGGGNGGNATYGGNAARAC	P-loop	this study
	Y782(R)	ATDGCNARDGGSAGNCC	hydrophobic domain	
	P-loop2(F)	GGNGGNRTNGGNAAAACAAC	GGV/I/MGKTT(P-loop)	
	P-loop3(F)	GGNGGNRTNGGNAAGACGAC	GGV/I/MGKTT(P-loop)	[19]
	P-loop5(F)	GGNGGNRTNGGNAARACCAC	GGV/I/MGKTT(P-loop)	
	GLPL3(R)	CAANGCCAANGGCAANCC	GL/FPL/FAL/V (hydrophobic domain)	
	S1(F)	GGTGGGGTTGGGAAGACAACG	GGV/IGKTT (P-loop)	[20]
	AS1(R)	CAACGCTAGTGGCAATCC	GLPLA/TL (hydrophobic domain)	
	Gkt1(F)	GGNGTNGGNAARACNAC	GGVGKTT(P-loop)	
	Pal2(R)	ARIGCTARIGGIARICC	GLPL(S/A)L	[21]
	Gkt2(F)	GGIGGIGTIGGIAAIACIAC	GGV/IGKTT (P-loop)	
	Dmg1(R)	AAGATCTCGTCCCATATC	DMGRDL	
	F5	GGNGTNGGAAGACAAC	GVGKTT	
	F12	GGAATGGGNGGNGTNGGNAARAC	GMGGVGKT	
	F13	GGTGGGGTWGGKAARACNAC	GGVGKTT	[22]
	F14	GGNATGGGNGGNNTNGGNAARACNAC	GMGGVGKTT	
	R6	GAANGCCAANGGCAAACC	GLPLAL	
	R9	NACYTTNAGNGCNAGNGGNGNCC	GLPLALKV	
STK class				
	P3(F)	TNGGNSANGGNGKNTTYGG	STK subdomain II	[23]
	P2(R)	ACNCCRAANGARTANACRTC	STK subdomain IX	
	D(F)	GGIGGITTYGGIATHGTITWYAARGG	STK subdomain I	[24]
	D(R)	ARIARYTTIGCIARICCRAARTC	STK subdomain VII	
DGAs				
	Glu-S(F)	RYNGGWGTWTGYTAYGG	LGVCYG	[11]
	Glu-AS(R)	CADCCRCTYTCNGAYAC	VSESGW	
	Y91(F)	TTTTTAAACGCCGCAATGA		[25]
	Y91(R)	CAGAAAGCAATGGTGGGAAT		

F, R represent for forward and reverse primer respectively; codes for mixed bases:N=A/G/C/T; Y=C/T; R=A/G; D=A/G/T; S=C/G; W=A/T; K=G/T.

formed into DH5α competent cells by the 42℃ heat shock method, and potential transformants were selected on an LB plate containing ampicillin (50 μg/mL), X-gal and IPTG, at 37℃ for 12—16 h. Single white colony was picked and inoculated into liquid LB medium containing ampicillin (50 μg/mL) for 12h cultivation with vigorous shaking. Primer pairs T7 and SP6 were used for PCR amplification to determine the existence of insert of the expected size. Positive clones were sent to Shanghai Yingjun Biotechnology Co. Ltd. (China) for sequencing.

1.4 Sequence and phylogenetic analyses

Vector sequences were removed manually using DNAstar program from the sequence data of cloned RGAs and DGAs, and BLASTx searches of the protein database of the GenBank were performed for the edited sequences. BLASTn search in the GenBank was conducted if the BASTx search in the protein database failed to reveal homology with *R* genes or RGAs. The ORF finder program in NCBI was used to search ORF of cotton RGA. Multiple sequence alignments of nucleic acid and amino acid sequences deduced from nucleic acid having read-through ORF by DNAstar program were carried out with Clustal X program with default options, and Bootstrap N-J tree (1000 replicates) was drawn. The constructed N-J tree would be edited using MEGE3 program.

1.5 Southern blot

Southern blotting was performed according to Sambrook *et al*.[26] and instructions of manufacturers of DIG-Labeled Detection kit (Roche, Germany).

1.6 Primer synthesis and reagents

Primers were synthesized by Shanghai Yingjun Biotechnology Co., Ltd. Taq DNA polymerase and RNA reverse transcriptase kits were purchased from Dalian Bao Biotechnology Co. Ltd. DNA gel extraction kit was purchased from Hangzhou V-gene Biochemical Co., Ltd. Restriction enzymes were purchased from New England Biolabs (USA). DIG-Labeling and Detection Kits were from Roche (Germany). Other commonly used reagents were from Shanghai Shenggong Biotechnology Service Co. Ltd.

2 Results and analysis

2.1 Cloning and characterization of RGAs and DGAs in cotton

Using various combinations of primers listed in Table 1, RGAs and DGAs were amplified from genomic DNA of resistant variety *G. barbadense* L.cv. Hai7124. While some primer pairs resulted in amplified products displaying a diffuse band on a 1.5% agarose gel, PCR products from most primer sets appeared to be an apparent single band of the expected size (Fig. 1). Occasionally, we observed PCR products of the expected size from a single degenerate primer (Fig. 1(b)). To exclude such false positive amplified DNA fragments, we used single forward and backward degenerate primers respectively as controls for amplification of each primer pair, and PCR products of the expected size from a primer pair would be recovered for sequencing only if PCR amplification of either single forward or backward primer resulted in no products or products of unexpected sizes.

For each PCR fragment recovered, at least 5 positive clones would be sent for sequencing. A total of 277 clones were sequenced. Using individual cloned sequences as query, BLAST search of NCBI database revealed that 79 sequences have significant sequence

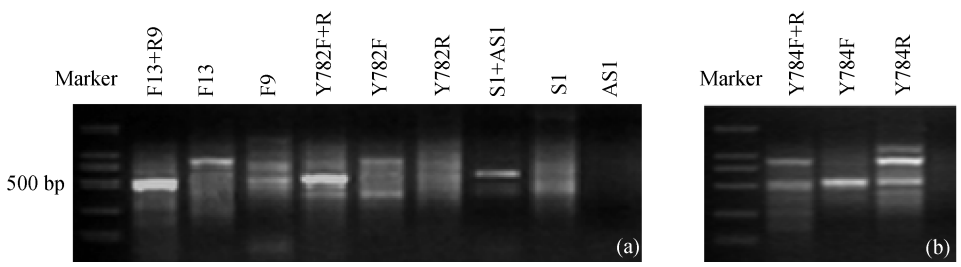

Fig. 1. PCR products of partial primers for RGAs. The names of primers were indicated at the top of gel.

homology with NBS class of disease resistance genes, 21 sequences show homology with protein kinase class of *R* genes, and 11 sequences are closely related to pathogenesis-related proteins.

2.2 Sequence analysis of RGAs and DGAs of cotton

(1) Sequence analysis of NBS-encoding RGAs in cotton. Previously 143 NBS-encoding RGAs were cloned from cotton by a PCR-based strategy[22,27−29]. Our current study further identified 79 novel NBS-encoding RGAs, among which 13 belonged to the new sequences relative to NBS-encoding RGAs cloned in cotton, the rest showed nucleotide identity from 85% to 99% with RGAs in cotton. Clustering analysis of all cloned RGAs in cotton (data not shown) revealed 10 subclasses on the basis of genetic distance with the smallest subclass containing only 1 member and the largest subclass consisting of 67 members. Significantly, all the 5 members of one such subclass were established by this study (Bootstrap value of 1000) containing L500, L501, L502, L518, L519, and the rest of our cloned sequences were assigned to other 5 subclasses.

Forty eight of the 79 cloned NBS-encoding RGAs have read-through ORFs and BLASTx search of the NCBI database revealed that the predicted proteins exhibit homology ranging from 61% to 79% with resistant proteins from cassava (*Manihot esculenta*), soybean (*Glycine max*), cacao (*Theobroma cacao*), coffee (*Coffea Arabica*), 53% to 66% similarity to products of identified *R* genes of tobacco *N*, tomato *I2*, Arabidopsis *RPS2*, *RPP5*, and flax *M*. Sequence similarity of these 48 NBS-coding RGAs among themselves ranges from 15% to 100% at amino acid level. Clustering analyses (Fig. 2) indicate that 20 of these RGAs could be classed into one group with TIR-NBS-encoding *R* genes of tobacco *N*, flax *L6*, *M*, Arabidopsis *RPP1*, *RPP5*, and the other 28 RGAs assigned to the group of non-TIR-NBS-encoding *R* genes of rice *XaI*, tomato *Ml-1*, Arabidopsis *RPS2*, *RPM1*, consistent with the results obtained by both Meyers *et al*.[30] and Tu *et al*.[22] that NBS-encoding RGAs can be classified into two distinct classes, namely TIR- and non-TIR classes.

In 1994, Traut *et al*.[31] identified two conserved nucleotide binding site motifs of P-LOOP and Kinase-2. Based on the sequence analyses of 481 sequences containing the isolated *R* genes from 26 plant species including Arabidopsis, rice, wheat, corn, soybean and the sequences of BAC, EST homology to *R* gene and RGAs cloned using methods of PCR, Meyers *et al*.[30] identified conserved motifs known as RNBS-A, RNBS-B, and RNBS-C etc. in the NBS region. It was also found that TIR- and non-TIR-containing *R* genes or RGAs displayed apparent sequence variations in the RNBS-A. The conserved sequence of TIR class RNBS-A is: FLENIRExSKKHGLEHLQKKLLSKLL, while that of the non-TIR class RNBS-A is: FDLx-AWVCVSQxF. NBS-encoding RGAs can be classified into two distinct groups on the basis of this conserved motif sequence. Alignment of multiple sequences of TIR- and non-TIR class RGAs identified 6 conserved motifs for each class respectively in cotton (Fig. 3). The consensus sequence of RNBS-A motif of TIR-encoding RGAs in cotton is: FL/IADI/VxxKC/HGLVS/CLQKQLLSQIF/L; while that of the non-TIR class RGAs is: FDS/IV/KxW/IA/VC/TVS/TE/QxF/I. The characterization of unique sequence features of these two classes of RGAs could provide the basis for classification of cotton *R* genes and related RGAs. In addition to the 6 conserved motifs mentioned above, NBS-encoding RGAs in cotton also display significant homology in other regions of the predicted proteins.

(2) Sequence analyses of STK-encoding RGAs. Using primer sets of P3, P2 and DF, DR, we isolated 21 STK-encoding RGAs in cotton. BLASTx search revealed that the best-matched sequence in the NCBI database for L444 and L455, with the identity of 98% and similarity of 95%, is the receptor-like protein kinase 3 from soybean, while L457 and L466, showing the lowest homology with sequences in the database, had the closest matched sequences being the putative receptor serine/threonine kinase PR5K and putative protein kinase *Xa21* from rice with 57% of identity, 80% and 78% of similarity respectively. The above results clearly indicate that STK-encoding RGAs from dicot plant species share a significantly higher homology. Twenty of the 21 STK-encoding RGAs have uninterrupted ORFs and their predicted products share 74%—98% homology with the identified protein kinases in the GenBank. The sequence similarities among STK-encoding RGAs isolated are

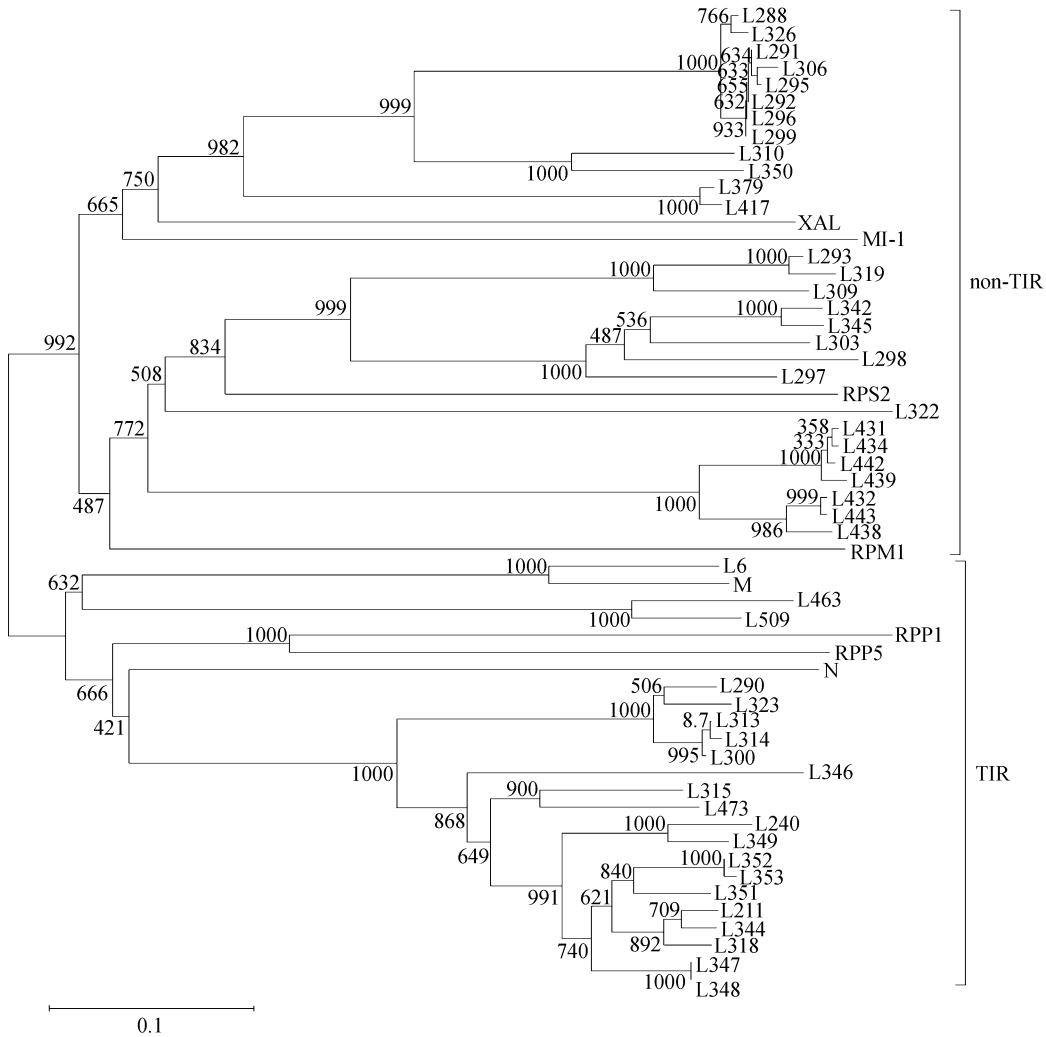

Fig. 2. Cluster analysis, based on the neighbor-joining method, of the deduced amino acid sequences of 48 cotton RGAs of the NBS class and cloned *R* genes: Arabidopsis *RPS2*, *PRM1*, *RPP1*, and *RPP5*, tomato *MI-1*, rice *XaI*, flax *L6* and *M*, tobacco *N*. The numbers above the branch represent bootstrap values (out of 1000).

35.7%—99.3% at the amino acid level. Based on the phylogenetic tree (Fig. 4) constructed with the predicted products of the above 20 sequences containing read-through ORFs, serine/threonine kinases from tomato, rice, and receptor-like kinases from tomato and rice, STK-encoding RGAs could be grouped into two subclasses: subclass A comprises 10 STK-encoding RGAs and serine/threonine kinases from rice, tamato; subclass B includes 10 STK-encoding RGAs and the receptor-like protein kinase-3 from soybean. Members of subclass L444, L455 and the receptor-like protein kinase-3 from soybean have the shortest phylogenetic distance, consistent with the results from the BLAST searches.

Alignment of predicted protein sequences of the 20 read-through ORFs of STK-encoding RGAs, tomato serine/threonine kinases and rice receptor-like protein kinases identified 9 conserved motifs termed I—VIII (Fig. 5) in the STK-encoding *R* genes.

(3) Sequence analysis of DGAs in cotton. Of the 11 DGAs isolated from cotton, the 265 bp stretch of the most 5′ part of L20, L268, and L269 displayed 98%—100% homology with genes encoding the identified *V. Dahliae*-induced PR proteins in cotton[32], but the matched sequences in the database contain only partial sequences of these three DGAs. BLASTx searches revealed that L279 exhibits 100% homology with 60 amino acids of a cowpea protease inhibitor,

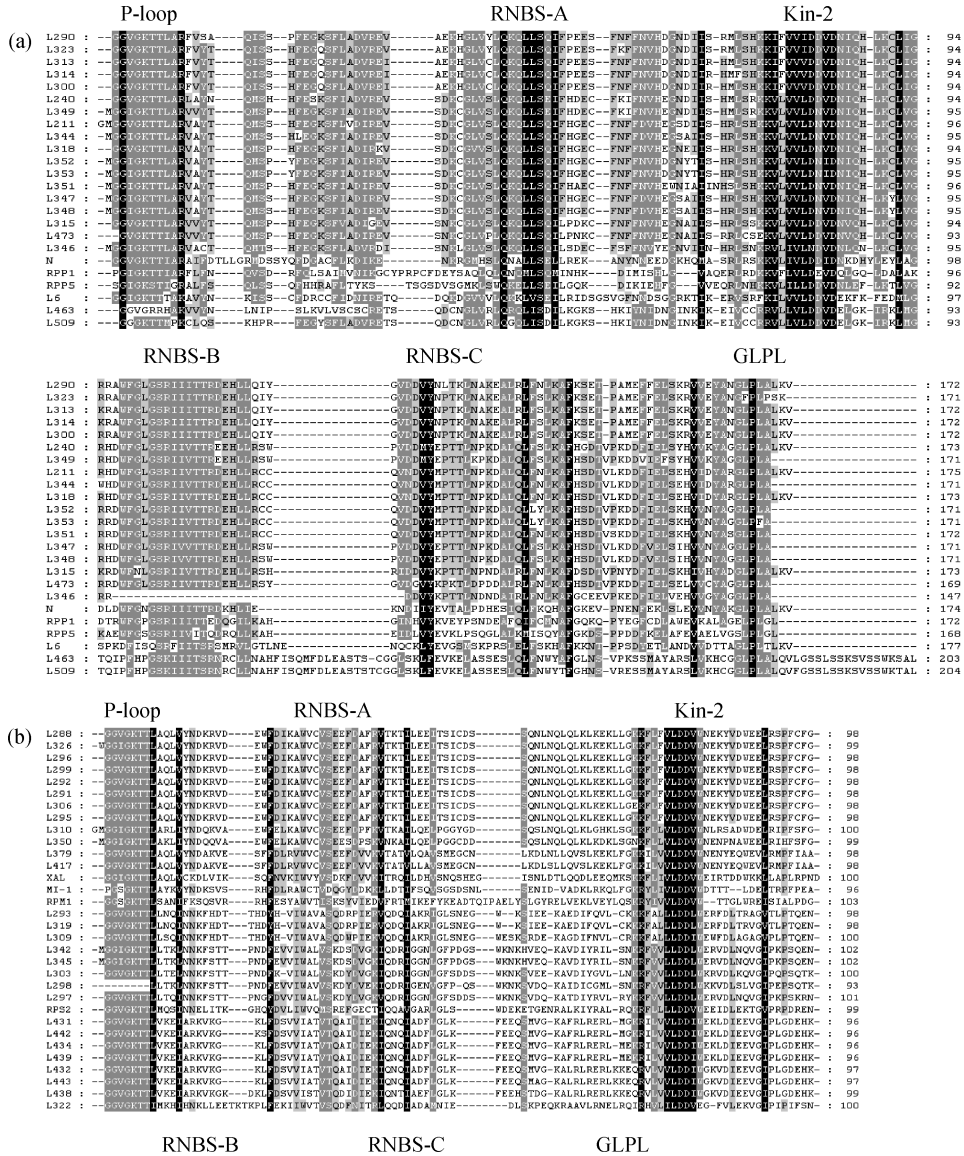

Fig. 3. Alignment of the deduced amino acid of the cotton NBS-encoding RGAs with the known NBS-encoding *R* genes (Arabidopsis *RPS2*, *PRM1*, *RPP1*, and *RPP5*, tomato *MI-1*, rice *XaI*, flax *L6,* tobacco *N*) and conserved motif analysis.

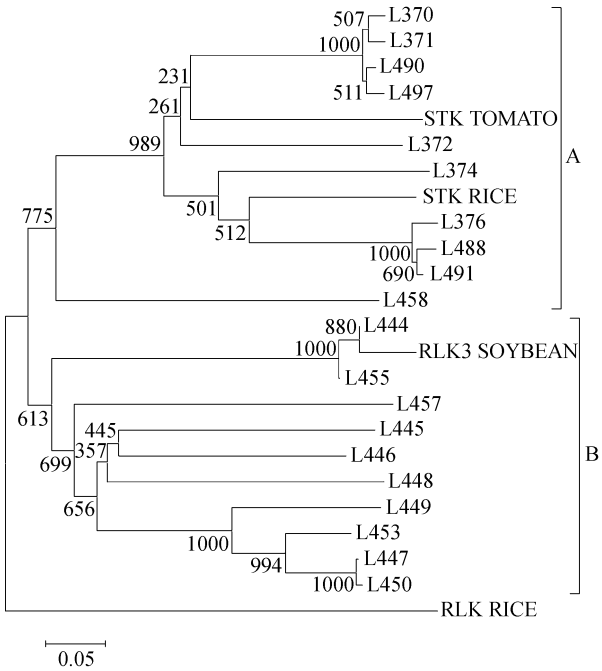

Fig. 4. Phylogenetic tree for 20 predicted amino acid sequences with the known protein kinases. STK TOMATO and STK RICE represent serine/threonine kinases of tomato and rice respectively, RLK3 SOYBEAN and RLK RICE represent receptor-like protein kinases of soybean and rice respectively. Bootstrap values (out of 1000) are noted above the branches.

and BLASTn searches showed that the identity is 100% between L279 and the region of 276 bp of a gene from the cotton curl virus. The shared homology is 85% or 86% between the predicted products of L354, L359, L361, L362 and the β-1,3-glucanase from citrus, the predicted L355 protein exhibits 81% homology with β-1,3-glucanase from the rubber trees, and the predicted products of L356 and L363 have significant homology with the precursors of β-glucanases. Of these 11 DGAs, L354, L355, L361 and L362 possess ORFs uninterrupted by stop codons. Alignments of multiple sequences with the translated amino acid sequences of these 4 DGAs and the β-1,3-glucanase from the orange and rubber trees (Fig. 6) showed that identities among L354, L355, L361, and L362 range from 56.4% to 98.7%, and that all four sequences display high homology with the rubber β-1,3-glucanase.

2.3 Cotton genome probed with NBS class of RGAs

NBS class of genes in plants is often found to be present as a cluster of a multigene family on the chromosomes. To assess the size of NBS-encoding RGAs multigene family, we arbitrarily used 4 clones, from TIR- and non-TIR-NBS-encoding RGAs respectively, as hybridization probes to probe the Hai7124 genome digested with various restriction enzymes. The hybridization patterns in Fig. 7 indicate that the Hai7124 genome has 3 copies of TIR-NBS class L326 and L379, and two copies of non-TIR-NBS class L314. The rest of the NBS class of RGAs have a copy number of 5—10. Thus, NBS class of RGAs resides as multigene families in the cotton genome. The fact that the hybridization patterns of the non-TIR-encoding sequences L344, L346, L473 are identical suggests that they are present as tandem arrays in the genome.

2.4 Expression analyses of RGAs and DGAs

Cotton verticillium wilt is still difficult to control and *Gossypium barbadense* L. cv Hai7124 is a well-known resistant variety. To identify RGAs and DGAs associated with resistance to verticillium wilt, specific primers (not listed) were designed respectively based on the sequences of 48 read-through ORFs of NBS-encoding RGAs, 20 read-through ORFs of STK-encoding RGAs, and 4 DGAs. Expression analyses were performed using cotton roots inoculated with the fungal pathogen *V. dahliae* on a time course of 0, 24, 48, 96, and 144 h. As shown in Fig. 8, the expression of 6 NBS class of RGAs, and 1 STK class of RGA was induced after infection, and the levels of expression of 1 DGA were up-regulated. RT primers of these genes were listed in Table 2. The expression levels of 7 RGAs were undetectable before the fungal pathogen challenge and the expression of L290 was detectable 24 and 48 h after inoculation. L296 expression was increased gradually with the progress of infection, and the expression pattern of L297 and L298 was consistent with no apparent changes in the amount of transcripts detected between 24 and 144 h after inoculation. The transcripts of L309, L347 and L477 were detectable at time points of 48, 96, and 144 h post-inoculation. The expression levels of L355 were strongly induced at the time points of 48 and 96 h after inoculation. It should be pointed out that no detection of transcripts cannot be viewed as evidence to show that the specific RGAs under study are not inducible

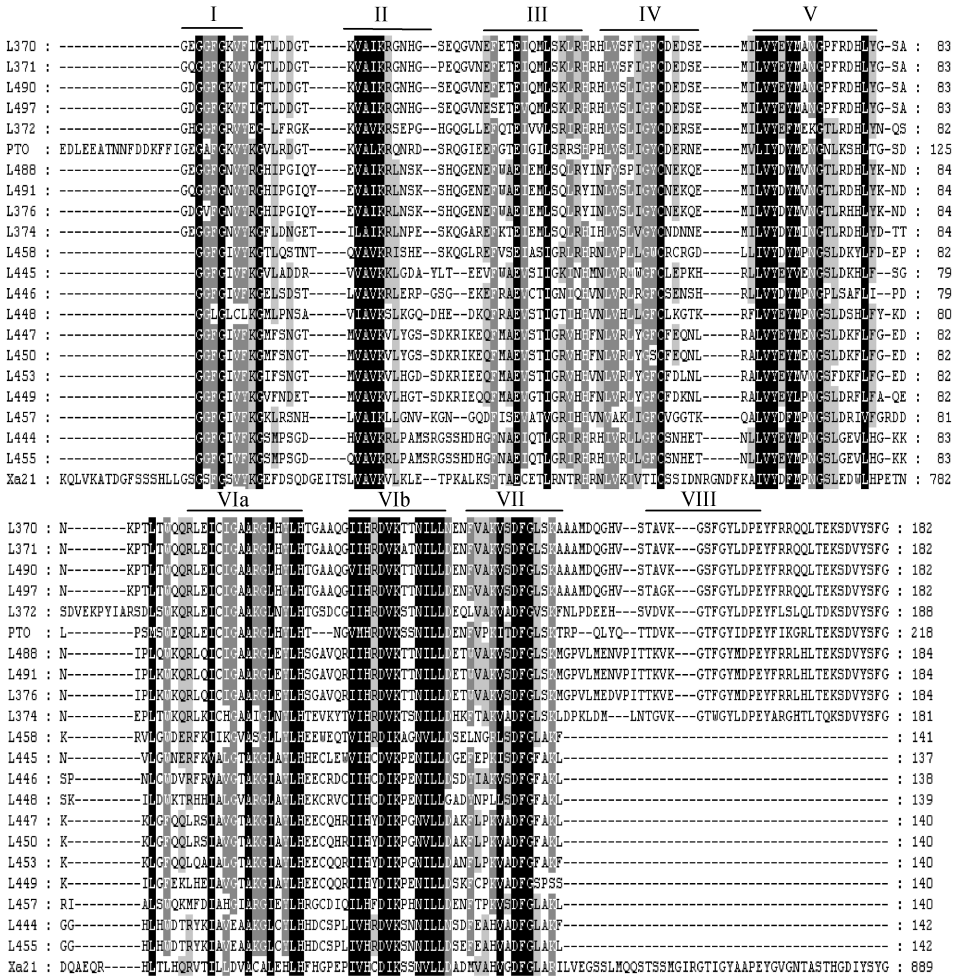

Fig. 5. Amino acid alignment of the cotton RGAs of STK class with homologues regions of *R* gene encoding kinase. Conserved motifs are marked at the top of sequences.

Fig. 6. Alignment of the deduced amino acid sequence of cotton DGAs with β-1,3-glucanase of orange and rubber trees.

by *V. dahliae* because of the uncertainty regarding the possibility that the studied RGAs could contain introns, and that the specific primers designed could be located within the intron regions.

3 Discussion

3.1 Isolation and analysis of RGAs and DGAs in cotton

PCR-based cloning of RGAs and DGAs from various plants has already attracted the attention of investigators in this field. It is evident that PCR-based cloning strategy is more convenient compared with transposon tagging and map-based cloning approaches. However, we have noticed occasional occurrence of interference of PCR products from single primer amplification likely due to the use of degenerate primers, especially given that our PCR amplification involves a polyploid cotton genome. We effectively eliminated

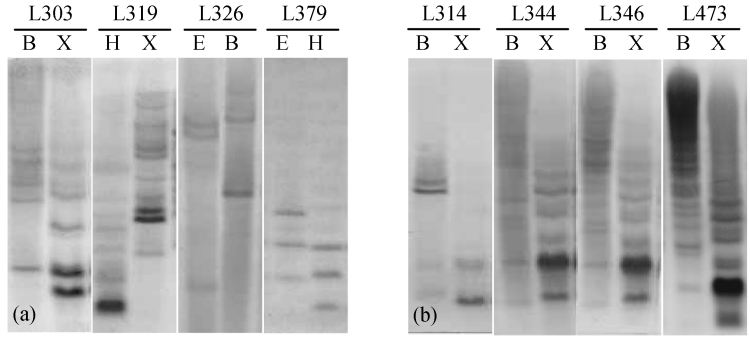

Fig. 7. Genome bloting of NBS-encoding RGAs of cotton. Probes of (a) and (b) come from TIR-NBS class and non-TIR-NBS class respectively. Restriction enzymes used: B, *Bam*HI, H, *Hin*dIII, E, *Eco*RI, X, *Xba*I.

Table 2 The RT primers of RGAs and DGAs

Sequence No.	F (5′—3′)	R (5′—3′)
L290	GGTGGGGTAGGTAAGACAAC	CCATTGGCATATTCCACTAC
L296	GGTGGGGTTGGTAAGACA	TAAGAGCAAGTGGGAGGC
L297	CAAGTTCAGCACCACACCG	AACCTTCAGAGCGAGGGG
L298	TCAGCGCGAGAGGAAGAC	GTTGTTATCTGGGCGGTG
L309	AACGACCCTCCTTAGCCA	GGGGAACGGGAGTTTCAA
L347	ATGGGGGGAATTGGAAAA	AGCCCACCAGCATAATTT
L447	GAAGAGCAGTTCATGGCGGA	ACAAAGCTTGGCTAATCCGA
L355	TGAAGTCCCTCCATCATCT	CCGAGTATAACGAATCCAAC

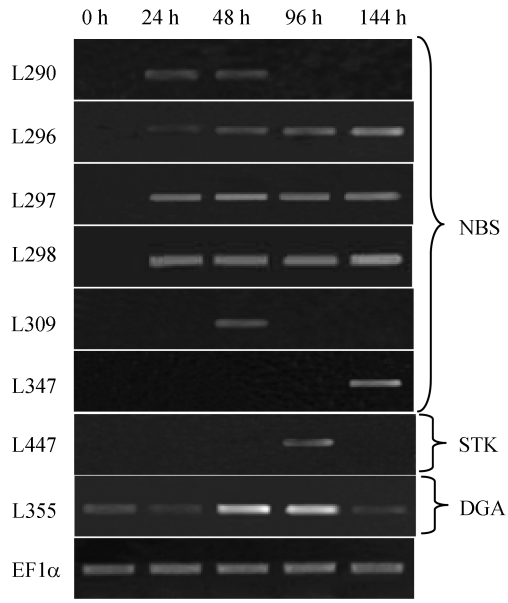

Fig. 8. RT-PCR expression analysis of RGAs and DGAs of cotton. *EF1α* is constituted expression gene of cotton. 0—144h indicate 0—144 h post inoculation.

this kind of interference by using single forward and backward primer as controls.

NBS class of *R* genes accounts for a significant proportion of the genomes of several plants. For example, it has been estimated that the Arabidopsis genome contains about 150 NBS-encoding genes, representing 0.5% of all its predicted ORFs[3,33]; in rice approximately 600 genes of NBS class have been identified, accounting for 1.5%[30,34] of all predicted ORFs of the rice genome. In cotton 143 NBS-encoding RGAs sequences have been cloned. Here we report the cloning of 79 novel NBS-encoding RGAs from *G. barbadense* L. cv. Hai7124. Clustering analysis of all cloned RGAs of NBS class resulted in the establishment of one novel subclass consisting of members cloned by this study only. An additional 5 subclasses also contain our isolated sequences. Therefore, our results have enriched both the number of the NBS-encoding RGAs and the number of its subclasses. All members of our identified novel subclass were isolated by PCR with the primer pair P-loop3(F) and GLPL3(R). P-loop3(F) is less degenerate than P-loop5 (F) and the difference between P-loop3(F) and P-loop2 (F) is that for the former the nucleotide A is replaced by G both at the third and the fifth positions of the 3′ end, arguing that it is more likely to obtain divergent PCR products when PCR amplification was performed with multiple less degenerate primers than a single highly degenerate primer pair.

Receptor-like protein kinases are known to be involved in a wide range of signal pathways, including

growth of meristem and leaf, regulation of abscission and self-incompatibility and transduction of brassinosteroid signals. In addition, STK also participates in plant defense responses. Of the cloned plant R genes, 5 encodes products related to the kinase class. There has been no report of sequences encoding proteins related to this class in cotton. This study identified 21 STK-encoding RGA sequences. The products of our isolated sequences share more than 74% homology with receptor-like protein kinases, and their shared homology with soybean protein kinase-3 shows identity as high as 98%. Based on the analysis of the translated amino acid sequence of 20 read-through ORFs, 2 subclasses were identified. One subclass contains the soybean receptor-like protein, and the other subclass includes the serine/threonine kinase from tomato. The above result indicates that STK-encoding genes of dicot plants have features of both structural conservation and diversity. All predicted products of cloned RGAs of STK class in cotton contain 9 (I—VIII) conserved consensus sequences. We also uncovered an STK-encoding RGA whose expression was shown to be induced after inoculation with *V. dahliae*. Successful isolation and mapping of *Pto* class of RGAs, and STK class of RGAs from wheat[35], soybean[36], grape[37], cocoa[38] etc. have been documented. Our current study clearly indicates that PCR-based strategy for isolation of STK-encoding RGAs in cotton is a viable option.

Defense response genes play a key role in signal transduction of plant host in response to pathogen infection. Some defense responses, such as hypersensitive reactions, produce host-encoding proteins, aptly termed pathogenesis related (PR) proteins, for example, glucanase, chitinase and peroxidase etc., to restrict the transport, and inhibit proliferation of the pathogen. Using degenerate primers based on the PR2 conserved motif, we isolated 11 DGAs which share high homology with PR proteins. Notably, part of the L279 sequence exhibits 100% homology with a gene from the cotton leaf curl virus. The temptation is to speculate that this observation might be the outcome of possible concerted evolution between plants and viruses. Similar instance has been observed in our lab (not published). Elucidation of the mechanism underlying these intriguing observations will necessitate future experimentations. Of the 4 sequences with read-through ORFs that have high homology with β-1, 3 glucanase, L355 was strongly induced at 48 and 96h post-inoculation. Zhen *et al*.[39] treated resistant and susceptible varieties of cotton with *V. dahliae* combined with salicylic acid, and it was shown that the expression levels of β-1, 3 glucanase were higher in the resistant variety than in the susceptible one, suggesting the potential causative association between *V. dahliae* resistance and the expression levels of β-1, 3 glucanase in cotton.

3.2 The distribution of RGAs of NBS class in cotton genome

R genes or RGAs conferring related functions but highly divergent in sequences are often found to be present in a cluster in the plant genomes[40]. Our analysis of some RGAs of the NBS class revealed that the same hybridization patterns were observed when the cotton genome was hybridized, with RGAs probes of 3 non-TIR-NBS class, indicating that these RGA sequences, with the related function, are clustered in the genome. Of the 8 RGAs studied, 5 sequences have more than 5 members, a result that is likely to be explained by the long-term concerted evolution to maintain the diversity of the resistance genetic loci. In addition, we also identified 6 NBS-encoding RGAs whose expression was shown to be induced by *V. dahliae* infection. It is possible that these RGAs could be associated with resistance to cotton verticillium wilt. Future experiments will be needed to test this possibility.

3.3 The relation between RGAs and DGAs with resistance loci in plant

One direct use of RGAs isolated by PCR is for genetic mapping of these sequences. There are many reported cases in which RGA sequences were localized in the regions near the resistance loci, for example, 3 RGAs from wheat were shown to be cosegregated with stripe rust resistance loci[41], and several RGAs in barley were identified in the vicinity of barley leaf rust resistance loci *Rph4*, *Rph7*, *Rph10*[42], and non-TIR-NBS-encoding RGAs in sunflower were shown to be linked to the sunflower downy mildew resistance locus *P15/P18*[43]. He *et al*.[29] reported the genetic mapping of 15 cloned NBS class of RGAs of

cotton to 7 chromosomes or gene clusters. Hinchliffe *et al*.[27] mapped 9 RGAs to partial homologous chromosomes 12 and 26 of cotton, and 6 of these RGAs were shown to be located within a genetic distance of 25.6 cM on chromosome 12, and the remaining 3 RGAs were localized within 40.5 cM on chromosome 26. Currently no resistance genes have been mapped in cotton. Thus, it is unclear whether any of these sequences is linked to cotton resistance loci. Nevertheless, with the increasing number of RGAs being isolated and mapped, it might be possible to identify specific disease resistance trait associated DNA markers or QTLs in cotton.

Breeding for *V. dahliae*-resistant cotton varieties is the long-sought goal of crop breeders. With the isolation of increasing numbers of RGAs and DGAs, it is reasonable to expect that genetic markers linked to disease resistance will be found and in turn the implementation of molecular marker-based or gene engineering-aided breeding strategy should greatly accelerate the process of developing cotton varieties highly resistant to various pathogens.

Acknowledgments This work was supported by the National Natural Science Foundation of China(Grant Nos. 30270806 and 30370899) and Program for Changjiang Scholars and Innovative Research Team in University in Ministry of Education in China.

References

1 Dilbirligi M, Erayman M, Sandhu D, et al. Identification of wheat chromosomal regions containing expressed resistance genes. Genetics, 2004, 166: 461－481
2 McDowell J M, Woffenden B J. Plant disease resistance genes: recent insights and potential applications. Trends Biotechnol, 2003, 21(4): 178－183
3 Dangl J L, Jones J D G. Plant pathogens and integrated defense responses to infection. Nature, 2001, 411: 826－833
4 Jones J D G. Putting knowledge of plant disease resistance genes to work. Curr Opin Plant Biol, 2001, 4: 281－287
5 Bendahmane A, Farnham G, Moffett P, et al. Constitutive gain-of-function mutants in a nucleotide binding site-leucine rich repeat protein encoded at the *Rx* locus of potato. Plant J, 2002, 32: 195－204
6 Lamb C J, Lawton M A, Dron M, et al. Signals and transduction mechanisms for activation of plant defense against microbial attack. Cell, 1989, 56(2): 215－224
7 Hammond-Kosack K E, Jones J D G. Resistant gene-dependent plant defense response. Plant Cell, 1996, 8: 1773－1791
8 Collinge D B, Slusarenko A J. Plant gene expression in response to pathogens. Plant Mol Biol, 1987, 4: 389－410
9 Cramer C L, Bell J N, Ryder T B, et al. Co-ordinated synthesis of phytoalexin biosynthetic enzymes in biologically-stressed cells of bean (*Phaseolus ulgaris* L.). EMBO J, 1985, 4: 285－289
10 Ham K, Kauffmann S, Albersheim P, et al. Host-pathogen interaction, XXXIX, A soybean pathogenesis-related protein with β-1,3-glucanase activity release phytoalexin elicitor-active heat-stable fragments from fungal walls. Mol Plant Microbe Interact, 1991, 4: 545－552
11 Pflieger S, Palloix A, Caranta C, et al. Defense response genes co-localize with quantitative disease resistance loci in pepper. Theor Appl Genet, 2001, 103: 920－929
12 Chen X, Soria M A, Yan G, et al. Development of sequence tagged site and cleaved amplified polymorphic sequence Markers for wheat stripe rust resistance gene *Yr5*. Crop Sci, 2003, 43(6): 2058－2064
13 Shi Z X, Chen X M, Line R F, et al. Development of resistance gene analog polymorphism markers for the *Yr9* gene resistance to wheat stripe rust. Genome, 2001, 44(4): 509－516
14 Radwan O, Bouzidi M F, Nicolas P, et al. Development of PCR markers for the *Pl5/Pl8* locus for resistance to *Plasmopara halstedii* in sunflower, *Helianthus annuus* L. from complete CC-NBS-LRR sequences. Theor Appl Genet, 2004, 109: 176－185
15 Zhang Z Z, Xu J S, Xu Q J, et al. Development of novel PCR markers linked to the BYDV resistance gene *Bdv2* useful in wheat for marker-assisted selection. Theor Appl Genet, 2004, 109: 433－439
16 Paterson A H, Lander E S, Hewitt J D. A rapid method for extraction of cotton (*Gossypium spp.)* genomic DNA suitable for RFLP or PCR analysis. Plant Mol Biol Rep, 1993, 11: 122－127
17 Ma Z Y, Wang X F, Zhang G Y, et al. Genetic studies of verticillium wilt resistance among different types of sea island cottons. Acta Agronomica Sinica (in Chinese), 2000, 26(3): 315－321
18 Wan C Y, Wilkins T A. A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). Anal Biochem, 1994, 223: 7－12
19 Noir S, Combes M C, Anthony F, et al. Origin, diversity and evolution of NBS-type disease-resistance gene homologues in coffee trees (*Coffea* L.). Mol Gen Genomics, 2001, 265: 654－662
20 Leister D, Ballvora A, Salamini F, et al. A PCR based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nat Genet, 1996, 14: 421－429
21 Liu J J, Ekramoddoullah A K M. Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola* Dougl. ex. D. Don.). Mol Gen Genomics, 2003, 270: 432－441
22 Tu L L, Zhang X L, Zhu L F, et al. Origin, diversity and evolution of NBS-type disease-resistance gene analogues in sea-island cotton (*Gossypium barbadense* L.). Acta Gene Sin (in Chinese), 2003, 30(11): 1071－1077
23 Vallad G, Rivkin M, Vallejos C, et al. Cloning and homology

modelling of a *Pto*-like protein kinase family of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet, 2001, 103: 1046－1058

24 Pflieger S, Lefebvre V, Caranta C, et al. Disease resistance gene analogs as candidates for QTLs involved in pepper-pathogen interactions. Genome, 1999, 42(6): 1100－1110

25 Pfaff T, Kahl G. Mapping of gene-specific markers on the genetic map of chickpea (*Cicer arietinum* L.). Mol Gen Genomics, 2003, 269: 243－251

26 Sambrook J, Fritsch E F, Maniatis F. Molecular cloning: A Laboratory Manual. 2nd ed. New York: Cold Spring Harbor Laboratory Press, 1989

27 Hinchliffe D J, Lu Y Z, Potenza C, et al. Resistance gene analogue markers are mapped to homeologous chromosomes in cultivated tetraploid cotton. Theor Appl Genet, 2005, 110: 1074－1085

28 Tan H, Callahan F E, Zhang X D, et al. Identification of resistance gene analogs in cotton (*Gossypium hirsutum* L.). Euphytica, 2003, 134: 1－7

29 He L M, Du C G, Covaleda L, et al. Cloning, characterization, and evolution of the NBS-LRR-encoding resistance gene analogue family in polyploid cotton (*Gossypium hirsutum* L.). Mol Plant Microbe Interact, 2004, 17(11): 1234－1241

30 Meyers B C, Dickerman A W, Michelmore R W, et al. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide binding superfamily. Plant J, 1999, 20: 317－332

31 Traut T W. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide binding-sites. Eur J Biochem, 1994, 222(1): 9－19

32 Mcfadden H G, Chapple R, Feyter R D E, et al. Expression of pathogenesis-related genes in cotton stems in response to infection by *Verticillium dahliae*. Physiol Mol Plant P, 2001, 58: 119－131

33 Meyers B C, Kozik A, Griego A, et al. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell, 2003, 15(4): 809－834

34 Goff S A, Ricke D, Lan T H, et al. A draft sequence of the rice genome (*Oryza sativa L.ssp. japonica*). Science, 2002, 296(5565): 79－92

35 Yan G P, Chen X M, Line R F, et al. Resistance gene-analog polymorphism markers co-segregating with the *Yr5* gene for resistance to wheat stripe rust. Theor Appl Genet, 2003, 106: 636－643

36 Collins N, Park R, Spielmeyer W, et al. Resistance gene analogs in barley and their relationship to rust resistance gene. Genome, 2001, 44(3): 375－381

37 Radwan O, Bouzidi M F, Vear F, et al. Identification of non-TIR-NBS-LRR markers linked to the *Pl5/Pl8* locus for resistance to downy mildew in sunflower. Theor Appl Genet, 2003, 106: 1438－1446

38 Hulbert S H, Webb C, Smith S M, et al. Resistance gene complexes: evolution and utilization. Annu Rev Phytopathol, 2001, 39: 285－312

39 Feuillet C, Reuzeau C, Kjellbom P, et al. Molecular characterization of a new type of receptor-like kinase (*wlrk*) gene family in wheat. Plant Mol Biol, 1998, 37(6): 943－953

40 Yamamoto E, Knap H T, Soybean receptor-like protein kinase genes: paralogous divergence of a gene family. Mol Biol Evol, 2001, 18(8): 1522－1531

41 Gaspero G D, Cipriani G. Nucleotide binding site/leucine-rich repeats, *Pto*-like and receptor-like kinase related to disease resistance in grapevine. Mol Gen Genomics, 2003, 269: 612－623

42 Lanaud C, Risterucci A M, Pieretti I, et al. Characterization and genetic mapping of resistance and defence gene analogs in cocoa (*Theobroma cacao L.*). Mol Breeding, 2004, 13: 211－227

43 Zhen X H, Li Y Z, Ultrastructural changes and location of β-1, 3-glucanase in resistant and susceptible cotton callus cells in response to treatment with toxin of *Verticillium dahliae* and salicylic acid. J Plant Physiol, 2004, 161(12): 1367－1377