Theor Appl Genet (2006) 113: 63–72
DOI 10.1007/s00122-006-0272-8

ORIGINAL PAPER

Xiao Wenkai · Xu Mingliang · Zhao Jiuren
Wang Fengge · Li Jiansheng · Dai Jingrui

Genome-wide isolation of resistance gene analogs in maize (*Zea mays* L.)

Received: 6 September 2005 / Accepted: 17 March 2006 / Published online: 11 April 2006
© Springer-Verlag 2006

Abstract Conserved domains or motifs shared by most known resistance (*R*) genes have been extensively exploited to identify unknown *R*-gene analogs (RGAs). In an attempt to isolate all potential RGAs from the maize genome, we adopted the following three methods: modified amplified fragment length polymorphism (AFLP), modified rapid amplification of cDNA ends (RACE), and data mining. The first two methods involved PCR-based isolations of RGAs with degenerate primers designed based on the conserved NBS domain; while the third method involved mining of RGAs from the maize EST database using full-length *R*-gene sequences. A total of 23 and 12 RGAs were obtained from the modified AFLP and RACE methods, respectively; while, as many as 109 unigenes and 77 singletons with high homology to known *R*-genes were recovered via data-mining. Moreover, *R*-gene-like ESTs (or RGAs) identified from the data-mining method could cover all RACE-derived RGAs and nearly half AFLP-derived RGAs. Totally, the three methods resulted in 199 non-redundant RGAs. Of them, at least 186 were derived from putative expressed *R*-genes. RGA-tagged markers were developed for 55 unique RGAs, including 16 STS and 39 CAPS markers.

Communicated by T. Lübberstedt

X. Wenkai · X. Mingliang (✉) · L. Jiansheng · D. Jingrui
National Maize Improvement Center of China,
China Agricultural University, 2 West Yuanmingyuan Road,
Beijing 100094, People's Republic of China
E-mail: mxu@cau.edu.cn
Tel.: +86-10-62733166
Fax: +86-10-62733808

Z. Jiuren · W. Fengge
Maize Research Center,
Beijing Academy of Agricultural Sciences,
Beijing, People's Republic of China

Introduction

Thus far, more than 50 resistance (*R*) genes have been cloned from a variety of plant species. These *R*-genes confer resistance against various pathogens, including bacteria, fungi, viruses, pests, and nematodes (Whitham et al. 1994; Buschges et al. 1997; Collins et al. 1999; Wang et al. 1999; Hulbert et al. 2001). Most isolated *R*-genes have conformed to the typical 'gene-for-gene hypothesis', except for the dominant *Hm1* gene in maize as well as several recessive *R*-genes, such as the *mlo* in barley and the *edr1* in *Arabidopsis* (Johal and Briggs 1992; Buschges et al. 1997; Frye and Innes 1998; Kim et al. 2002). Despite their broad resistance, most *R*-genes share limited conserved domains or motifs, such as nucleotide binding site (NBS), leucine-rich repeat (LRR), protein kinase (PK), transmembrane (TM), leucine zipper (LZ), and Toll-Interleukin-1 (TIR) (Traut 1994; Cooley et al. 2000; Liu and Ekramoddoullah 2004). Of 50 *R*-genes cloned so far, at least 37 share the NBS-LRR domain in their protein sequences, and forming the largest *R*-gene class (Bent et al. 1994; Lawrence et al. 1995; Kaloshian et al. 1998; Yoshimura et al. 1998). This *R*-gene class can be further divided into two sub-classes, TIR-NBS-LRR and CC-NBS-LRR, based on their N-terminal structure (Meyers et al. 1999). The second *R*-gene class equipped with the LRR-TM-PK domains integrates functions of both signal perception and transduction, for example the *Xa21* gene in rice (Song et al. 1995; Liu et al. 2002). The third *R*-gene class, *Cf-9* and *Cf-4* among others, is characterized with a large and symmetrical extra-cellular LRR domain that can combine with avr-polypeptide or other proteins for signal reception (Jones et al. 1994; Thomas et al. 1997). The fourth *R*-gene class contains only the PK domain, such as *Pto* and *Rpg1* (Martin et al. 1993; Brueggeman et al. 2002).

Most *R*-genes have been cloned via two approaches, transposon tagging and map-based cloning (or positional cloning). The first method depends on inactivation of the target gene caused by transposon insertion.

Currently, only a few transposons are suitable for tagging, and the likelihood that a transposon is inserted within the target *R* gene is very low (Brutnell 2002). The second approach requires saturation of the target *R* gene with molecular markers. In model plants with available genome sequences, such as *Arabidopsis* and rice, high-density markers can be readily developed (The *Arabidopsis* Genome Initiative 2000; Yu et al. 2002). In maize, however, development of high-density markers in the target *R* region is rather difficult, if not impossible, due to lack of availability of the complete genome sequence. Fortunately, most *R*-genes share limited conserved domains, and these domains can be exploited for designing degenerate primers to amplify other unknown *R*-gene analogs (RGAs) and mine EST databases in order to recover *R*-gene-like sequences (Kanazin et al. 1996; Graham et al. 2000; Dilbirligi and Gill 2003; Hunger et al. 2003; Madsen et al. 2003; Rossi et al. 2003). If sufficient RGAs are available to cover genome-wide *R*-genes for a particular plant species, this reservoir of RGAs can serve as a candidate *R*-gene resource for cloning of unknown *R*-genes. This candidate gene approach is particularly productive for cloning *R*-genes in plants having either large genomes or those with no available genome sequences such as wheat and maize.

So far, PCR amplification using degenerate primers designed based on conserved motifs is a common approach for identifying RGAs in various plant species. A total of 11 classes of non-cross-hybridizing sequences, having high-level amino acid identities with those of NBS-LRR *R*-genes, have been isolated and mapped in the maize genome (Collins et al. 1998). Based on both sequence comparisons and hybridization patterns, RGAs isolated from maize, rice, and wheat have been assigned into 13 classes, and 12 were mapped onto 17 loci on the barley genome using a doubled-haploid mapping population (Collins et al. 2001). In *Arabidopsis*, 42 non-redundant *R*-gene-like ESTs and three *Pto*-like sequences amplified with degenerate primers have been mapped onto 47 loci (Botella et al. 1997). In rice, 109 *R*-gene/defense-gene like ESTs were mapped to regions harboring major *R*-genes or QTLs using a doubled-haploid mapping population (Wang et al. 2001). To date, a sizable collection of ~800 RGAs have been obtained via PCR amplification from nearly 20 plant species, and one of these RGAs has already been confirmed to correspond to a functional *R*-gene (*DM3*) (Shen et al. 2002).

With advances in genomics programs, EST data have been steadily accumulating in many plant species, thus allowing for identification of RGAs from EST databases via the data-mining approach. Moreover, RGAs screened from EST databases are derived from expressed *R*-genes, and therefore are more desirable for cloning *R*-genes via the candidate gene approach. In wheat, 220 expressed *R*-gene candidates have been identified using the data-mining and modified RNA fingerprinting methods (Dilbirligi et al. 2003). In sugarcane, 88 *R*-gene-like ESTs have been identified via the data-mining approach (Rossi et al. 2003). Ultimately, genome-wide analysis of *R*-gene homologues has been conducted in *Arabidopsis* and rice whereby whole genome sequences are now available. Consequently, nearly 200 coding sequences, similar to NBS-LRR *R*-genes, have been identified in the *Arabidopsis* genome (Meyers et al. 2003). Approximately 500 NBS-LRR *R*-genes belonging to the non-TIR sub-class are present in the rice genome (Monosi et al. 2004).

The overall goal of this research is to isolate as many RGAs as possible from the maize genome using three different approaches, including a modified AFLP method to amplify RGAs from maize genome, a modified RACE method to isolate expressed RGAs from maize cDNA population, and data-mining of maize EST databases, available in MaizeGDB and NCBI, to obtain *R*-gene-like ESTs.

Materials and methods

Plant material

The parental lines '1145' and 'Y331' along with their BC_1 population (using 'Y331' as recurrent parent) were kindly provided by Prof. Shaojiang Chen (National Maize Improvement Center of China) and used in the modified AFLP method. The parent '1145' is a 'universal' resistant inbred line conferring resistance to various diseases, including rot stalk, head smut, Northern and Southern blights, and virus diseases, among others. Conversely, the parent 'Y331' is a rather susceptible inbred line with no reported resistance to any of the major diseases mentioned above. For the modified RACE method, the European inbred line FAP1360, conferring complete resistance against sugarcane mosaic virus (SCMV), was used. For the data-mining method, *R*-gene-like ESTs were separately cloned from two parental lines, '87-1' and 'Zong 3', from which a recombinant inbred line (RIL) mapping population has been developed.

The modified AFLP method

Extraction of genomic DNA was performed according to the CTAB procedure (Chen and Ling 1996). About 100 ng of genomic DNA was double-digested to completion with restriction enzymes *Mse*I/*Pst*I. Digested DNA fragments were ligated to *Mse*I- and *Pst*I-adapters. Ligation products were then diluted with $TE_{0.1}$ at a ratio of 1:10 for pre-amplification. PCR products were diluted with $TE_{0.1}$ at a ratio of 1:50 for selective amplification using the degenerate primer (designed on P-loop motif GVGKTT in the NBS domain) combined with each of the 64 P-xxx selective primers (x represents one of the four nucleotides).

The selective amplification was performed in 10 μl containing 2 μM of each of four dNTPs, 2 μM Mg^{2+}, 1 μl 10×PCR reaction buffer, 25 μM of each of the degenerate primer and of the selective primer, 0.5 U Taq DNA

polymerase, and 2.5 µl diluted pre-amplification product. The amplification conditions were denatured at 94°C for 2 min, then cycling for 10 times using a touch-down strategy (an initial cycle of 94°C for 30 s, 65°C for 45 s, 72°C for 1 min, then lowering the annealing temperature for each cycle by 0.7°C during the following 9 cycles), followed by 30 cycles at 94°C for 30 s, 56°C for 45 s, and 72°C for 1 min.

The amplification products were size-separated on 6% denaturing polyacrylamide gels. AFLP bands were sliced from the gel and boiled in 20 µl of TE for 10 min. After centrifugation at 13,000× *g* for 10 min, the supernatant was used for re-amplification of the isolated DNA band. The PCR mixture was prepared in 50 µl volume containing 25 µM corresponding primers, 5 µl 10×PCR buffer, 2 U Taq DNA polymerase, 2 µM Mg^{2+}, 2 µM each of four dNTP, and 0.5 µl supernatant. The PCR reaction was run at first step of denature at 94°C for 2 min, followed by 36 cycles of 94°C for 30 s, 56°C for 45 s, and 72°C for 1 min.

The modified RACE method

Leaf tissues of 2–3-week-old seedlings of inbred line FAP1360 were collected. Total RNA was isolated and precipitated using the LiCl method, and mRNA isolation was performed using PolyATract System III kit (Promega, Madison, WI, USA). The first-strand cDNA synthesis was catalyzed by PowerScript reverse transcriptase (BD SMART™ RACE cDNA Amplification kit, BD Biosciences, Palo Alto, CA, USA) according to the manual. The modified RACE reaction was performed using both the degenerate primer (designed based on the P-loop motif) and the 3′-UPM primer (provided in the SMART RACE kit). The PCR mixture, 50 µl, contains the following reagents: 5 µl 10×UPM primer, 2 µl P-loop degenerate primer (10 µM), 1 µl 50×BD Advantage 2 polymerase mix, 5 µl 10×BD Advantage 2 PCR buffer, 2 µl dNTP mixture (10 mM each), and 35 µl PCR-grade water. The following touch-down PCR program was used for the RACE reaction: the first 5 cycles at 94°C for 30 s, 68°C for 30 s, and 72°C for 3 min, the next 5 cycles at 94°C for 30 s, 65°C for 30 s, and 72°C for 3 min, followed by 25 cycles at 94°C for 30 s, 60°C for 30 s, and 72°C for 3 min.

The PCR products from the modified RACE reaction were subjected to electrophoresis on 1% agarose gel. The main bands, ranging in size from 1.5 to 2.0 kb, were excised from the agarose gel and used for identifying RGAs.

Data mining of *R*-gene-like ESTs

The amino-acid sequences of 48 known *R*-genes, covering the four major *R*-gene classes, were used to search for (tBLASTn) homologous ESTs in online available maize EST databases, including MaizeGDB (http://www.maizegdb.org) and NCBI (http://www.ncbi.nlm.nih.gov). Those ESTs with scores ≥ 100 and *E*-values ≤ 10^{-10} were deemed tentative *R*-gene-like ESTs, and all hypothetical *R*-gene-like ESTs were used to find their corresponding unigenes in MaizeGDB. The resulting unigenes and singletons were in turn used to search the GenBank database by tBLASTx to confirm their putative *R*-gene-like functions. In addition, all *R*-gene-like ESTs were used to identify possible genomic sequences in the MAGI database in the Schnable laboratory (http://schnablelab.plantgenomics.iastate.edu/). For each *R*-gene-like unigene or EST, the sequence was used to design primers for amplification of corresponding sequences from both parental lines '87-1' and 'Zong 3'.

Cloning, sequencing, and analysis of PCR products

Polymerase chain reaction products from the above three methods were purified using Nucleo Trap Gel Extraction Kit (BD Biosciences), and these were ligated into the pGEM-T vector (Promega) at 16°C for overnight. The ligation mixture, 10 µl, consisted of the following reagents: 5 µl 2×rapid ligation buffer, 1 µl pGEM-T vector, 1 µl T_4 DNA ligase, 2 µl purified PCR products, and 1 µl ddH_2O. The ligation product was transformed into *Escherichia coli* strain DH10B by electroporation, and incubated overnight at 37°C in a solidified LB medium containing 50 µg/ml ampicillin, 20 µg/ml X-gal, and 20 µg/ml IPTG.

The positive clones were identified by PCR amplification using the flanking T7 and SP6 primers, and sequenced at the HuaDa Sequencing Center (Beijing, China). The software, 'sequence assembly', and 'multiple sequence alignment' in DNAMAN Version 4.0, were used to confirm either identities or similarities for all sequences obtained from positive clones. Functional confirmation of obtained sequences were performed using tBLASTx (http://www.ncbi.nlm.nih.gov).

Development of RGA-tagged molecular markers

In maize, 3′-UTRs were reported to show more sequence diversity than coding regions (Rafalski 2002). Therefore, RGAs amplified from the two parental lines '87-1' and 'Zong 3' were carefully investigated for sequence divergence (InDels and SNPs) along their 3′-ends. The flanking regions of the InDels and SNPs were used to design primers to re-amplify both '87-1' and 'Zong 3'. For convenient separation of PCR products on 6% polyacrylamide gels, the distance between the forward and reverse primers was set at ~500 bp. If distinct and polymorphic PCR bands were observed between '87-1' and 'Zong 3', then an STS marker was deemed successfully developed for an InDel. Likewise, if polymorphic PCR bands were observed between '87-1' and 'Zong 3' following digestion with a particular restriction enzyme, then a CAPS marker was deemed successful for a SNP. These developed RGA-tagged markers are indispensable in anchoring corresponding RGAs along the maize genome.

Results

The RGAs generated from the modified AFLP method

For each primer pair, 40–50 AFLP bands ranging from 200 to 800 bp in sizes were visualized on a 6% polyacrylamide gel. Most bands appeared in both parental inbred lines, and only a few bands appeared in the 'universal' resistant parent '1145' and in resistant progeny, but not in the susceptible parent 'Y331'. All resistance-related bands and some of common bands were isolated. A total of 672 positive clones were sequenced.

Assembly of sequenced AFLP bands resulted in 284 non-redundant sequences, and these sequences were used to conduct a GenBank database search using tBLASTx to assign their putative protein functions. Of the 284 sequences, 23 were highly homologous to the maize rust *R*-gene *rp3-1* ($E \leq 10^{-13}$ to $E \leq 10^{-140}$) (Table 1), and 95 sequences contained putative domains similar to those in cloned *R*-genes, including 46 sequences containing NBS domains, 22 sequences harboring TM domains, and 27 sequences carrying PK domains. The remaining 166 sequences did not have any known domains whereby 132 sequences were highly homologous to maize genomic DNA and 34 sequences were highly homologous to genomic DNA of other plant species.

The RGAs from the modified RACE method

The PCR products from modified RACE amplifications were subjected to electrophoresis on a 1% agarose gel. As a result, PCR products appeared as smears, and ranging in size from 0.5 to 2.0 kb (Fig. 1). Considering the distance between the P-loop motif (where the degenerate primer was designed on) and poly-A tails of *R*-gene transcripts, the 1.5–2.0 kb bands most likely contained PCR products amplified from *R*-genes (Fig. 1). Therefore, bands of 1.5–2.0 kb in size were excised from 1% agarose gel, and cloned into the pGEM-T vector. A total of 672 positive clones were subjected to sequencing.

Following assembly of these 672 positive clones, 246 non-redundant sequences were identified. All the unique sequences were used to search for their homologous sequences in GenBank using tBLASTx to identify their hypothetical functions. Among these 246 sequences, 12 were highly homologous ($E \leq 10^{-12}$ to $E \leq 10^{-240}$) to the maize rust *R*-gene *rp3-1* (Table 1), 48 sequences contained putative NBS domains, 73 sequences contained TM domains, 25 sequences contained PK domains, and 88 sequences were maize mRNAs with unknown functions. Meanwhile, all 246 sequences were compared with the maize EST databases in both MaizeGDB and GenBank, and for every sequence, it's corresponding high-homologous EST was found. These results indicated that available maize EST databases have good coverage of genome-wide expressed sequences.

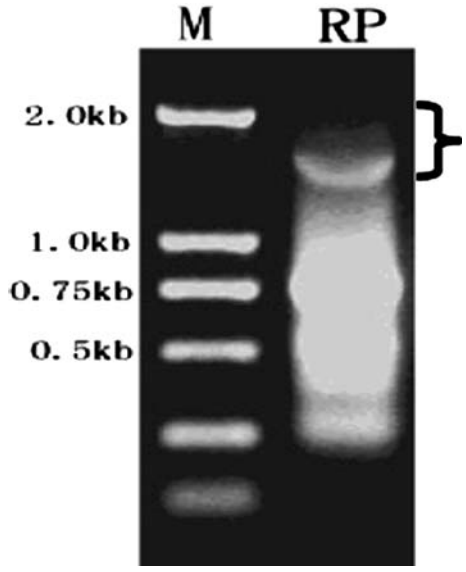

Fig. 1 The 3′-RACE products were subjected to electrophoresis on 1% agarose gel. *M* marker, DL2000; *RP* RACE products. The gel slice containing RACE products of 1.5–2.0 kb in sizes (as indicated in the *bracket*) was excised from the gel and cloned into pGEM-T vector

The RGAs from the data mining method

Full-length sequences of the 48 known *R*-genes were used to perform tBLASTn searches against the maize EST databases in MaizeGDB and GenBank. As shown, the

Table 1 Resistance gene analogs derived from both the modified AFLP and RACE methods

Methods	RGA names and sequence ID[a]
Modified AFLP	A-RGA1(DQ183069), A-RGA2(DQ183070), A-RGA3(DQ183071) A-RGA4(DQ183072), A-RGA5(DQ183073), A-RGA6(DQ183074) A-RGA7(DQ183075), A-RGA8(DQ183076), A-RGA9(DQ183077) A-RGA10(DQ183078), A-RGA11(DQ183079), A-RGA12(DQ183080) A-RGA13(DQ183081), A-RGA14(DQ183082), A-RGA15(DQ183083) A-RGA16(DQ183084), A-RGA17(DQ183085), A-RGA18(DQ183086) A-RGA19(DQ183087), A-RGA20(DQ183088), A-RGA21(DQ183089) A-RGA22(DQ183090), A-RGA23(DQ183091)
Modified RACE	R-RGA1(DQ183092), R-RGA2(DQ183093), R-RGA3(DQ183094) R-RGA4(DQ183095), R-RGA5(DQ183096), R-RGA6(DQ183097) R-RGA7(DQ183098), R-RGA8(DQ183099), R-RGA9(DQ183100) R-RGA10(DQ183101), R-RGA11(DQ183102), R-RGA12(DQ183103)

[a] All RGA sequences in the present study have been deposited in Genbank, and codes in parentheses are GenBank sequence IDs. All RGAs derived from the modified AFLP and RACE methods showed homologous sequences to maize *R*-gene *rp3-1*

same EST could be hit repeatedly by using different *R*-genes ($E \leq 10^{-10}$). Accordingly, all redundant ESTs and their corresponding *R*-genes were removed, thus resulting in over 600 *R*-gene-like ESTs. The *R*-gene-like ESTs identified were further used to identify their corresponding unigenes in the MaizeGDB database. The resultant unigenes together with singletons were in turn compared with the GenBank database to confirm their putative *R*-gene-like functions. A total of 109 unigenes and 77 singletons were identified as having amino-acid sequences similar to those of *R*-genes. These 186 non-redundant unigenes/ESTs were cloned and sequenced from both parental lines '87-1' and 'Zong 3'. Among these 186 unigenes/ESTs, 57 contained putative NBS-LRR domains, 53 contained putative LRR-TM domains, 66 contained putative PK, PK/PK, and PK-LRR domains, and ten contained putative TM domains (Table 2).

Integration of RGAs derived from the three methods

All 23 AFLP-derived RGAs and 12 RACE-derived RGAs were used to perform BLASTn search to identify their corresponding homologous ESTs in the maize EST database in MaizeGDB. With an *E* value setting of $E \leq 10^{-50}$, ten AFLP-derived RGAs and all 12 RACE-derived RGAs appeared to have corresponding *R*-gene-like ESTs previously recovered in the data-mining method. However, for the remaining 13 AFLP-derived RGAs, no corresponding ESTs were identified. Since the maize EST databases available at present have a good coverage of maize expressed genes, the *R*-gene-like ESTs identified from the data-mining method may cover the majority of *R*-genes present in the maize genome. The integration of RGAs recovered from the above three methods resulted in a total of 199 non-redundant RGAs available in the maize genome, of which at least 186 RGAs were derived from putative expressed *R*-genes.

Resistance gene analogs-tagged molecular markers

Sequence alignments between two parental lines '87-1' and 'Zong 3' revealed frequent occurrences of small InDels and SNPs. Based on small InDels, 16 RGA-tagged STS markers have been successfully developed, and each STS marker gave rise to distinct and polymorphic PCR bands between two parental lines (Table 3). Likewise, 39 RGA-tagged CAPS markers have been obtained from those SNPs related to restriction sites, and each CAPS marker showed unambiguous polymorphic bands following digestion with their corresponding restriction enzyme (Table 4).

Discussion

In an attempt to efficiently isolate RGAs, a modified AFLP method was developed in the present study whereby the *Pst*I restriction enzyme was used for digesting maize genomic DNA instead of *Eco*RI, and a degenerate primer based on the conservative P-loop was used for selective amplification. Of 284 unique sequences, only 23 were highly homologous to *rp3-1*. Although 95 other sequences contained *R*-gene-like domains, these were not confirmed to be derived from *R*-genes. In comparison with the commonly used PCR-based RGA isolation, this method was more flexible for isolating RGAs as only a single degenerate primer was needed. Nevertheless, there were some inherent drawbacks for this method. These included the following: (1) only a fraction of PCR products were *R*-related sequences, thus limiting applicability of this method for large-scale identification of RGAs, and (2) many *R*-genes might be missed if no *Pst*I/*Mse*I fragments could be generated from their NBS domains used for designing the degenerate primer.

Some inherent limitations are present in isolating RGAs from genomic DNA using degenerate primers. For plant species having small genome sizes such as *Arabidopsis*, the presence of RGAs lacking intron have been frequently observed (Pan et al. 2000), however, this is not often the case in most plant species of large genome sizes, such as barley (Leister et al. 1999). In addition, all family members in a resistance locus, regardless of the presence of either active genes or pseudogenes, are capable of generating RGAs. Conceivably, RGAs derived from non-expressed members will severely interfere with identification of active *R*-genes. A viable solution to these problems in RGA isolation can be achieved by replacing genomic DNA with cDNA. This switch of template DNA has been recently reported in wheat (Dilbirligi et al. 2004). In the present study, a modified RACE method using cDNA as template DNA has been set up in an attempt to uncover as many RGAs as possible from a single PCR reaction. Considering the distance between the C terminal poly-A and the P-loop motif in the NBS domain is about 2 kb in length for most *R*-genes (Mindrinos et al. 1994), RACE products of 1.5–2.0 kb in sizes are sliced from the agarose gel, and cloned into the pGEM-T vector. Sequence analysis has shown only 5% of cloned RACE products are highly homologous to the maize *R* gene *rp3-1*, and another 62% of RACE products contain domains similar to those in *R*-genes, although these could not be confirmed as *R*-genes derivatives.

To our surprise, both AFLP- and RACE-derived RGAs were highly homologous to a single *rp3-1* gene. The *rp3-1* locus was very complex and contained at least nine members (Webb et al. 2002). Sequence analysis indicated that this locus was a typical NBS-LRR *R*-gene. The complete sequence was deposited in MaizeGDB. As known, a majority of *R*-genes belong to the NBS-LRR class. In the present study, the degenerate primer was designed based on the P-loop in the NBS domain, and generated RGAs corresponding to the NBS-LRR class. This may explain the rapid identification of AFLP or RACE sequences homologous to *rp3-1*. There must be other RGAs corresponding to either other NBS-LRR *R*-genes or other resistance classes, however, their functions

Table 2 Resistance-gene-like unigenes/ESTs identified from the data-mining method

Plant species	*R*-genes and protein ID	Unigenes	Singleton ESTs	Putative domain
Pepper (*Capsicum annuum* L.)	*Bs2* (AAF09256)	ZMtuc03-08-11.15048, ZMtuc03-08-11.15672, ZMtuc03-08-11.5872, ZMtuc02-12-23.15809, ZMtuc03-08-11.5577, ZMtuc03-08-11.23619	AI372154.1	NBS-LRR
Tomato (*Lycopersicon esculentum* L.)	*Cf-2* (AAC15780)	ZMtuc02-12-23.16870, ZMtuc03-08-11.23525 ZMtuc03-08-11.11147, ZMtuc03-08-11.4171 ZMtuc03-08-11.1259, ZMtuc02-12-23.13447 ZMtuc02-12-23.6493, ZMtuc03-08-11.3195 ZMtuc03-08-11.204, ZMtuc03-08-11.6980 ZMtuc02-12-23.5884, ZMtuc03-08-11.2725	CB605462.1, CB885791.1, CD661783.1, CB331101.1 AI881878.1, CB833894.1, BG266585.1, CB885266.1 BQ703741.1, CD439101.1, AI861160.1, CF050131.1	LRR-TM
	Cf-4 (CAA05268)	ZMtuc03-08-11.8486, ZMtuc03-08-11.63 ZMtuc03-08-11.12207, ZMtuc03-08-11.13082 ZMtuc02-12-23.18367, ZMtuc03-08-11.28183	BU098845.1, BM074374.1, CB280878.1, AW067239 CD443188.1, BE638805.1, CD001549.1, BU080434.1	LRR-TM
	Cf-5 (AAC78591)	ZMtuc02-12-23.13447, ZMtuc03-08-11.14686 ZMtuc03-08-11.18407	CD449002.1, CD980669.1, CF007923.1, CD965824.1, CD965506.1	LRR-TM
	Cf-9 (CAA05274)	ZMtuc03-08-11.14284, ZMtuc03-08-11.10388 ZMtuc03-08-11.22359	BM501433.1, CD448503.1, BM073179.1, CB815939.1	LRR-TM
	Hero (CAD29728)	ZMtuc03-08-11.4092, ZMtuc03-08-11.6133	BM080497.1	NBS-LRR
	I2 (AAD27815)	ZMtuc02-12-23.16350, ZMtuc02-12-23.257 ZMtuc03-08-11.19518	CD432996	NBS-LRR
	Prf (AAC49408)		CF059864.1	NBS-LRR
	Pto (AAF76313)	ZMtuc03-08-11.11473, ZMtuc03-08-11.11091 ZMtuc03-08-11.23804, ZMtuc03-08-11.25777 ZMtuc03-08-11.11688, ZMtuc03-08-11.13789 ZMtuc03-08-11.3020, ZMtuc03-08-11.4448 ZMtuc02-12-23.15501, ZMtuc03-08-11.7711 ZMtuc03-08-11.26242, ZMtuc03-08-11.26241 ZMtuc03-08-11.16841, ZMtuc03-08-11.454	CD444610.1, CF052394.1 CF244010.1	PK
Wheat (*Triticum aestivum* L.)	*Cre3* (AAC05834)	ZMtuc03-08-11.22677, ZMtuc03-08-11.22900 ZMtuc03-08-11.14099	CD435787.1, BM661284.1	NBS-LRR
	Yr10 (AAG42168)	ZMtuc03-08-11.17424, ZMtuc03-08-11.6133	BM079360.1	NBS-LRR
Lettuce (*Lactuca sativa* L.)	*Dm3* (T30562)		BM500504.1, BM500372.1	NBS-LRR
Sugarbeet (*Beta vulgaris* L.)	*Hs1* (AAB48305)	ZMtuc03-08-11.25544	BG319898.1, BM500875.1	LRR-TM
Barley (*Hordeum vulgare* L.)	*Mla1* (AAG37356)	ZMtuc02-12-23.13704, ZMtuc02-12-23.17140 ZMtuc02-12-23.16075, ZMtuc03-08-11.6077	CF050069.1, AW258044, BG360799, CD432996, BE519242	NBS-LRR
	Mlo (P93766)	ZMtuc03-08-11.28961, ZMtuc02-12-23.13510 ZMtuc03-08-11.24960, ZMtuc02-12-23.1753 ZMtuc03-08-11.7848, ZMtuc03-08-11.12786	BF729342.1, CA452413.1 CF019983.1, BM080281.1	TM
	Rpg1 (AAM81980)	ZMtuc03-08-11.23446, ZMtuc03-08-11.2750 ZMtuc03-08-11.11139, ZMtuc03-08-11.27351 ZMtuc03-08-11.2480, ZMtuc03-08-11.16274 ZMtuc02-12-23.15501, ZMtuc03-08-11.12381 ZMtuc03-08-11.10390, ZMtuc02-12-23.10883 ZMtuc02-12-23.13808	CB179544.1, CF272760 CF033580.1	PK/PK
Arabidopsis (*Thaliana* L.)	*Pbs1* (AAG38109)	ZMtuc03-08-11.3970, ZMtuc03-08-11.13632 ZMtuc02-12-23.14468, ZMtuc03-08-11.24002 ZMtuc03-08-11.21457, ZMtuc03-08-11.12076 ZMtuc03-08-11.11101, ZMtuc03-08-11.3464 ZMtuc03-08-11.29933, ZMtuc03-08-11.16763 ZMtuc03-08-11.26771, ZMtuc03-08-11.16729 ZMtuc02-12-23.13762, ZMtuc02-12-23.16936 ZMtuc03-08-11.25238, ZMtuc03-08-11.10257 ZMtuc03-08-11.13547, ZMtuc02-12-23.16967 ZMtuc02-12-23.15404, ZMtuc03-08-11.21462 ZMtuc03-08-11.18230, ZMtuc03-08-11.14960 ZMtuc03-08-11.12767	AI770970.1, CB886188.1 AI854916.1, BQ578127.1 CF002281.1, CD439174.1	PK
	Rpp1 (AAC72977)	ZMtuc02-12-23.16870, ZMtuc03-04-07.15888	CD990401, BM074091.1	NBS-LRR
	Rps2 (AAA21874)	ZMtuc03-08-11.18256, ZMtuc03-08-11.18255		NBS-LRR
	Rpm1 (CAA61131)	ZMtuc03-08-11.19668	BM079685.1, BM080497.1	NBS-LRR

Table 2 (contd.)

Plant species	*R*-genes and protein ID	Unigenes	Singleton ESTs	Putative domain
Rice (*Oryza sativa* L.)	*Pib* (BAA76282)	ZMtuc02-12-23.16492	BM500195.1	NBS-LRR
	Pi-ta (AAK00132)	ZMtuc02-12-23.9851	BQ280307.1	NBS-LRR
	Xa21 (T10725)	ZMtuc02-12-23.16134, ZMtuc02-12-23.12796	BG321296, AA979732.1, AW061999.1, AI947508.1	PK-LRR
Maize (*Z. mays* L.)	*Rp1d* (AAD47197)	ZMtuc02-12-23.15207	CD445359.1, AI947572.1, BI992969.1, BM075603, BM349795, CF041780	NBS-LRR

All putative protein sequences of the known 48 *R*-genes have been used to perform tBLASTx search against the maize EST databases in MaizeGDB and GenBank. Normally, one EST could be hit repeatedly by using different *R*-genes. The redundant ESTs together with their corresponding *R*-genes were omitted, resulting in over 600 non-redundant ESTs. These *R*-gene-like ESTs were further used to identify their corresponding unigenes in the MaizeGDB database. The resultant unigenes together with singletons were in turn compared with the GenBank database to confirm their putative *R*-gene-like functions. Finally, 109 unigenes and 77 singleton ESTs emerged as having amino-acid sequences similar to those of *R*-genes. The names and types of unigenes/ESTs together with their 24 corresponding *R*-genes were listed in the table

Table 3 Primer sequences and sizes of polymorphic bands for the RGA-tagged STS markers

R-gene-like ESTs	Primer sequences (5′–3′)		Sizes (bp) 87-1/Zong3
	Forward	Reverse	
ZMtuc02-12-23.16967	TGGTGCAGCATAGCCGTAGGTA	AATAGCTTCCAGGGCCATAGGG	446/454
ZMtuc03-08-11.14686	TCCTCAACGGGAACAGGCTG	AGAAGTGCGGCGCCCAAGGC	521/536
ZMtuc03-08-11.18255	TGCGCGACAACTACTATATTCAGA	GAGGTTCAGCGATTCTAGGCAG	654/667
ZMtuc03-08-11.10257	CCCAACGGATACTTGTGTACGA	GTCACCTTGTTCTTGAGCCACTC	466/452
ZMtuc02-12-23.13447	GTCTTGACAGTACCGTCTGGTCC	TGGTTACAATCAGTTGACAGGCA	532/516
ZMtuc03-08-11.8486	CTGTCCTTGATCTGAGTCAG	CTCCAGTGCTTGAAGCTTGC	535/558
CD965506	GTTCTTTAGCGTTCCAAGTTCA	CTCCACAATAACAAGTTGACAGGA	587/579
BM500195	GGATGGTAGCAAAGCGAGGA	TCTGTCCCTGAAGAAGCGC	389/376
BG319898	GCACGTCGCTAACACCACTARG	ACTCTATCACCACAAGGCCACAG	450/427
CF244010	GATGAAGAAGCGGAGGAAGGTC	TCGAACTCGTGGTGCAGGTAGT	438/451
BU080434	GCCCAATTTCTCGTAACCTTGTT	CTGGAGCATCTTCAGTATCTGGAAC	678/696
AI770970	AGCACAGGCCAGGTTGTTG	CATGTGATTTGTCGCCAACTG	589/566
BM080497	TTCTGAATTTCCAATCTGCTGTGC	GTAGGCTGTTCCAGTCCGTGAC	486/514
CB605462	TTCTGGTAACCTGTTCGCTC	TGGCAATCATTTGACTGGGC	1,235/1,486
CD001549	GTGCCGACAGCGAGTTACGA	ATCATTTGACTGGGCCCATC	1,146/1,457
CD965824	TGAGAGTGCCTGTCAGGTGG	GGATATTGAACCTGAGGTGG	566/543

could not be confirmed by tBLASTx search due to lack of corresponding maize *R*-genes present in Genbank.

With a steady accumulation of ESTs, an efficient approach to mine expressed RGAs from EST databases has been reported in several plant species (Dilbirligi and Gill 2003; Rossi et al. 2003). In wheat, four data-mining strategies, including individual full-length search, domain search, consensus-sequence search, and multiple-motif search, were compared with each other in mining of RGAs from the wheat EST database, and the first strategy was proven to be the most successful (Dilbirligi and Gill 2003). Accordingly, in the present study, the individual full-length search using each of the 48 known *R*-genes was adopted to mine RGAs from the maize EST databases accessible in maizeGDB and NCBI websites. In addition, applying different cut-off values in BLAST searches would result in the recovery of different numbers of *R*-gene-like sequences. In sugarcane, adoption of a very stringent threshold value ($E \leq 10^{-50}$) resulted in identifying only 88 RGAs (Rossi et al. 2003). In contrast, large numbers of RGAs, including 243 NBS-LRR-type and 101 other type RGAs, were identified in wheat when applying a BLAST threshold value of $E \leq 10^{-1}$ (Dilbirligi and Gill 2003). In the present study, a moderate threshold value of $E \leq 10^{-10}$ was adopted to perform data mining, and the putative RGAs were in turn compared with the database in GenBank to confirm their identities, leading to the identification of 186 *R*-gene-like unigenes/ESTs. Since most known *R*-genes were used in the data-mining method, RGAs belonging to different *R*-gene classes could be recovered from the EST databases. As reported previously, the 88 RGAs screened from the sugarcane EST database represented three major classes of *R*-genes carrying respective NBS-LRR, LRR-TM, and Serine/Threonine kinase domains (Rossi et al. 2003). The 184 putative expressed *R*-genes from wheat consisted of 87 NBS-LRR genes, 16 receptor-like kinases, 13 *Pto*-like kinases, and 68 genes from other types (Dilbirligi et al. 2004). The number of *R*-gene-like ESTs identified in the present study was very close compared to those found in wheat. Of 186 ESTs/unigenes, 57 contained putative NBS-LRR domains, 53 contained putative LRR-TM domains, 66 contained putative PK, PK/PK, and PK-LRR domains, and ten contained putative

Table 4 Primer sequences and restriction enzymes for the RGA-tagged CAPS markers

R-gene-like ESTs and a RGA	Primer sequences (5′–3′)		Restriction enzymes
	Forward	Reverse	
ZMtuc02-12-23.15809	CTCTTTCAATAACCACAGCCGC	TTTCAGCATGCTTGTTGCAGAC	*Tsp*509I
ZMtuc03-08-11.11147	CGATGTCTCTTCTCTGATGAACTGC	TGCCACTGGTTTGCAATTCTTT	*Xsp*I
ZMtuc03-08-11.1259	CTGCGCCAGATATATTGAACGA	GAAGCATTCCAGAAGAGATCGG	*Alu*I
ZMtuc03-08-11.3195	CTAGAGGTTCAGGCATATGGCG	AGCTCCACAGGAATTCGTTGAG	*Xsp*I
ZMtuc03-08-11.12207	GCAAACGAGCAAGATGATAAGGA	GTCCAATAACAGACTCTCCGGC	*Nae*I
ZMtuc02-12-23.17140	ACGTACCCATCAGCCTAAAGCA	CAGATTGGACAGAGATTGCACCA	*Pvu*II
ZMtuc03-08-11.19518	ACTGCCAAGTGCATCTCACAGT	GAAAGCTCTTCCAACGAACGTG	*Xsp*I
ZMtuc02-12-23.14468	TCTTCCTGTCTCTCAGCATGGG	TTGGCTCACTTGCAGATCACCT	*Xsp*I
ZMtuc03-08-11.25238	AAATCACAGTCCGGATCGCTCT	CTTTGGCTACCATGCACCAGAA	*Taq*I
ZMtuc03-08-11.11091	CAGCAAGAGATGACTTTGAAGGC	ATCCTGAATATTTCCGGAGGCA	*Hae*III
ZMtuc03-08-11.26242	GCTACTGGCAAATTTGAGAGAAGAA	TGACACATGAGACCCATCCACT	*Spe*I
ZMtuc03-08-11.16841	CATCATCCACAGAGACGTGAAGG	AGCTACCACCAGGCTTGAACAG	*Kpn*I
ZMtuc03-08-11.12381	AAGGAGAAGGAAGGAAACATAGCAA	CCAGAGTACGCATACACCCTGA	*Taq*I
ZMtuc03-08-11.23446	CAACTTCGACGAGCTGAGAAAGA	AGGACGTTGCTGGACTTGATGT	*Xsp*I
ZMtuc03-08-11.5872	GGTGAAGGTCTAGTCCATCTTGAGA	AACCCTGAAGCTACCAATTGGG	*Bsp*1407I
ZMtuc03-08-11.4171	CGCGTTACTGCTCAGATCAAGC	GCAATTTACGGTGTCGGATGAA	*Xsp*I
ZMtuc03-08-11.6980	AGGCTCTCCAAGCTTCTGAGGT	TGTGTTGCAAGATCTCCCTTTCC	*Mse*I
ZMtuc02-12-23.18367	CATTCCTCCATGCTAAGAACC	CTCACCTCACTCCAGCGACT	*Tas*I
ZMtuc03-08-11.25544	ACAACGCAGTGCACAACGATAG	GTGGCCGTGAGAGGATTCTTCT	*Xsp*I
ZMtuc02-12-23.257	GAACCTGCACAAGATCTCGCTC	GACATCAGTCCGCTAGCCAGTC	*Xho*I
ZMtuc03-08-11.21457	CAGCTCTGCAAATCCTCCCTTT	AGTCATGGGCACCTACGGCTAC	*Bln*I
ZMtuc03-08-11.12767	TTCTCCACTTGAAAGGTTCTCCAT	GGCATGTTGAATGAAACTAGCGAT	*Alu*I
ZMtuc02-12-23.9851	CATTCCGGTGCCTAACTGAGA	TTGCCACTAGTACCCATGCG	*SnaB*I
ZMtuc03-08-11.4448	TCGGGACATCAAATCAGCAAAT	CGCCACCACTATGGATCTCATC	*Taq*I
ZMtuc03-08-11.26241	GGGACAGTTGGATACCTGGACC	AGCTCGTCGAAAGATGCTCCTC	*Cla*I
ZMtuc03-08-11.10390	TACTTGCATCACGATTGTTCGC	TCCAGCATCTGTACCACCCTGT	*Msp*I
ZMtuc02-12-23.16134	TGAATATCTGCCCATAGCATC	CATCACTCATTGACATCCTTTGGTC	*Hinf*I
A-RGA10	AAGCGAGAGGAGACCGATGG	GCAACAGCATAGTCGAACTG	*Alu*I
BM073179	AGCTTGCAAACCTGCCAGAACT	TGCCACGGAATAGTACAGCGAC	*Alu*I
CF050069	AGGTCCTGTGCTTCAAGAGCGT	TATCTGATGGCCTATGGGTTCG	*Msp*I
CD439174	CAAGGGCTTCTGCATCGAA	TTGCTCTACGTCCGGTCACA	*Msp*I
BM074091	GCTCAAGTACCTGGACATCTCGC	GAGTTTCCCTACGCTCGGTTTG	*Hae*III
AW061999	GCTCCAGAGTATGCTGAAGGCT	CCATCTTTGAGACTTTGATCAT	*Eco*T22I
CF050131	TAGCTTAGGCGATATCGACCCG	CCACAATAACAAGTTGACAGGACCA	*Msp*I
BQ703741	ACATTGAATTGCAAGAGGTGGG	TTATACCATCGGAAATCTCGAAGC	*Eco*T22I
AW258044	AAGTGACCTCAACACCTCTGGC	AAGCTGCGATATGGATGTGGAT	*Alu*I
BF729342	CGTGCGAACCAGAATTAGGAGT	CTTCTTCCGGCAGTTCTTCAGG	*Xsp*I
CB179544	AAGACTTGAAGGCTATTGTTGCC	CTTGTTCCCACAGGGTATTTCC	*Xsp*I
BM079685	TGCAAGTGAAGTCCACATACAACAA	TGAGCATGAACAGGGTCAGTAGTG	*Pst*I

TM domains. In contrast, RGAs amplified from genomic DNA or cDNA using degenerate primers could only recover those *R*-genes on which the degenerate primer was designed. Generally, degenerate primers were designed based on the conservative NBS domain, and these primers could only generate RGAs corresponding to the NBS domain.

Among the three different ways adopted in this study, the data-mining method is found to be the most efficient to isolate RGAs. The success of this data-mining method depends largely on the coverage of the EST database, but only if the EST database covers all active genes in the genome, will the majority of *R*-gene-like ESTs be obtained from a particular plant species. In maize, more than 550,000 EST entries (up to August 26, 2005) have been deposited in GenBank and MaizeGDB. All 246 unique sequences from the modified RACE method have corresponding ESTs in the above two EST databases, indicating that all 550,000 ESTs covered almost all expressed genes in the maize genome. While comparing RGAs obtained from the three different methods, all RACE-derived and nearly half AFLP-derived RGAs overlap with *R*-gene-like ESTs obtained from the data-mining method. For those13 AFLP-derived RGAs without corresponding ESTs, it is likely that they are derived from inactive *R*-gene sequences.

At this time, only two *R*-genes have been cloned from maize via the transposon tagging method, *Hm1* encoding a toxin reductase that detoxifies the fungal toxin of *Cochliobolus carbonum* race 1 (Johal and Briggs 1992) and *Rp1-D* confering resistance to rust in maize (Collins et al. 1999). Because of the presence of highly-repetitive sequences and lack of whole genome sequences for maize, it is very difficult to clone *R*-genes via a map-based cloning strategy. Considering the distinct *R*-gene structure and advances made in the maize genome, the candidate *R* gene approach via RGA is a very promising strategy for cloning *R*-genes in maize. This approach has been useful in cloning of the *R*-gene *Lr10* in wheat (Feuillet et al. 1997). We are attempting to use this candidate

gene approach in cloning of the *Scmv1* gene conferring resistance to the SCMV in maize. Among 20 RGAs discovered by Collins et al. (1998), one RGA, pic19, has been mapped to the region whereby one of two SCMV *R*-genes, *Scmv1*, resides (Xu et al. 1999). Three pic19 homologous sequences have been identified while screening of the B73 BAC library using pic19 (Quint et al. 2003). Based on conserved regions among pic19 and its three homologous sequences, full-size cDNAs have been obtained using RACE, and these cDNAs have been cloned downstream of the 35S promoter for complementary tests (unpublished results).

To date, only a small fraction of maize RGAs have been cloned and mapped onto the maize genome. However, it is impossible right now to clone other maize genes via the candidate gene approach. This is why we have pursued efforts to isolate genome-wide RGAs and developed RGA-tagged markers. In the next step, an RGA linkage map will be constructed based on RGA-tagged markers and a RIL mapping population. This RGA linkage map will be very useful in identifying a candidate *R*-gene(s) for a given disease if a certain RGA is co-located with a major resistance locus or QTL.

Acknowledgments This study was financially supported by the national outstanding youth foundation of China, the talent people's foundation of China Agricultural University, and the Beijing Agricultural Innovative Platform—Beijing Natural Science Fund Program, 'Construction of the standard DNA fingerprint database for Chinese maize germplasm and study on innovative fingerprinting techniques'.

References

Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Leung J, Staskawicz BJ (1994) RPS2 of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance gene. Science 265:1856–1860

Botella MA, Coleman MJ, Hughes DE, Nishimura MT, Jones JDG, Somerville SC (1997) Map positions of 47 *Arabidopsis* sequences with sequence similarity to disease resistance genes. Plant J 12:1197–1211

Brueggeman R, Rostoks N, Kudrna D, Kilian A, Han F, Chen J, Druka A, Steffenson B, Kleinhofs A (2002) The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. Proc Natl Acad Sci USA 99:9328–9333

Brutnell TP (2002) Transposon tagging in maize. Funct Integr Genomics 2:4–12

Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Topsch S, Vos P, Salamini F, Schulze-Lefert P (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell 88:695–705

Chen BJ, Ling XY (1996) Extraction genome DNA from *Z. mays*. J Anhui Agro tech Normal School 10:27–28

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogs in maize. Mol Plant Microbe Interact 11:968–978

Collins N, Drake J, Ayliffe M, Sun Q, Ellis J, Hulbert S, Pryor T (1999) Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants. Plant Cell 11:1365–1376

Collins N, Park R, Spielmeyer W, Ellis J, Pryor AJ (2001) Resistance gene analogs in barley and their relationship to rust resistance genes. Genome 44:375–381

Cooley MB, Pathirana S, Wu HJ, Kachroo P, Klessig DF (2000) Members of the *Arabidosis* HRT/RPP8 family of resistance gene confer resistance to both viral and oomycete pathogens. Plant cell 12:663–676

Dilbirligi M, Gill KS (2003) Identification and analysis of expressed resistance gene sequences in wheat. Plant Mol Biol 53:771–787

Dilbirligi M, Erayman M, Sandhu D, Sidhu D, Gill KS (2004) Identification of wheat chromosomal regions containing expressed resistance genes. Genetics 166:461–481

Feuillet C, Schachermayr G, Keller B (1997) Molecular cloning of a new receptor-like kinase gene encoded at the *Lr10* disease resistance locus of wheat. Plant J 11:45–52

Frye CA, Innes RW (1998) An *Arabidopsis* mutant with enhanced resistance to powdery mildew. Plant Cell 10:947–956

Graham MA, Marek LF, Lohnes D, Cregan P, Shoemaker RC (2000) Expression and genome organization of resistance gene analogs in soybean. Genome 43:86–93

Hulbert SH, Webb CA, Smith SM, Sun Q (2001) Resistance gene complexes: evolution and utilization. Annu Rev Phytopathol 39:285–312

Hunger S, Gaspero GD, Mohring S, Bellin D, Schafer-Pregl R, Borchardt DC, Durel CE, Werber M, Weisshaar B, Salamini F, Schneider K (2003) Isolation and linkage analysis of expressed disease-resistance gene analogues of sugar beet (*Beta vulgaris* L.). Genome 46:70–82

Johal GS, Briggs SP (1992) Reductase activity encoded by the *Hm1* disease resistance gene in maize. Science 258:985–987

Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ, Jones JD (1994) Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266:789–793

Kaloshian I, Yaghoobi J, Liharska T, Hontelez J, Hanson D, Hogan P, Jesse T, Wijbrandi J, Simons G, Vos P, Zabel P, Williamson VM (1998) Genetic and physical localization of the root-knot nematode locus Mi in tomato. Mol Gen Genet 257:376–385

Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. Proc Natl Acad Sci USA 93:11746–11750

Kim MC, Panstruga R, Elliott C, Muller J, Devoto A, Yoon HW, Park HC, Cho MJ, Schulze-Lefert P (2002) Calmodulin interacts with MLO protein to regulate defense against mildew in barley. Nature 416:447–451

Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG (1995) The *L6* gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene *RPS2* and tobacco viral resistance gene *N*. Plant Cell 7:1195–1206

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Graner A, Schulze-Lefert P (1999) RFLP and physical mapping of resistance gene homologues in rice (*O. sativa*) and barley (*H. vulgare*). Theor Appl Genet 98:509–520

Liu JJ, Ekramoddoullah AKM (2004) Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola Dougl. ex. D. Don.*). Mol Genet Genomics 270:432–441

Liu GZ, Pi LY, Walker JC, Ronald PC, Song WY (2002) Biochemical characterization of the kinase domain of the rice disease resistance receptor-like kinase *Xa21*. J Biol Chem 277:20264–20269

Madsen LH, Collins NC, Rakwalska M, Backes G, Sandal N, Krusell L, Jensen J, Waterman EH, Jahoor A, Ayliffe M, Pryor AJ, Langridge P, Schulze-Lefert P, Stougaard J (2003) Barley disease resistance gene analogs of the NBS-LRR class: identification and mapping. Mol Genet Genomics 269:150–161

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432–1436

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20:317–332

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant cell 15:809–834

Mindrinos M, Katagiri F, Yu GL, Ausubel FM (1994) The *A. thaliana* disease resistance rene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell 78:1089–1099

Monosi B, Wisser RJ, Pennill L, Hulbert SH (2004) Full-genome analysis of resistance gene homologues in rice. Theor Appl Genet 109:1434–1447

Pan QL, Wendel J, Fluhr R (2000) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J Mol Evol 50:203–213

Quint M, Dußle CM, Melchinger AE, Lübberstedt T (2003) Identification of genetically linked RGAs by BAC screening in maize and implications for gene cloning, mapping and MAS. Theor Appl Genet 106:1171–1177

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Rossi M, Araujo PG, Paulet F, Garsmeur O, Dias VM, Chen H, Van Sluys MA, D'Hont A (2003) Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. Mol Gen Genomics 269:406–419

Shen KA, Chin DB, Arroyo-Garcia R, Ochoa OE, Lavelle DO, Wroblewski T, Meyers BC, Michelmore RW (2002) Dm3 is one member of a large constitutively expressed family of nucleotide binding site-leucine-rich repeat encoding genes. Mol Plant Microbe Interact 15:251–261

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804–1806

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Thomas CM, Jones DA, Parniske M, Harrison K, Balint-Kurti PJ, Hatzixanthis K, Jones JDG (1997) Characterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in *Cf-4* and *Cf-9*. Plant Cell 9:2209–2224

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

Wang ZX, Yano M, Yamanouchi U, Iwamoto M, Monna L, Hayasaka H, Katayose Y, Sasaki T (1999) The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J 19:55–64

Wang Z, Taramino G, Yang D, Liu G, Tingey SV, Miao GH, Wang GL (2001) Rice ESTs with disease-resistance gene or defense response gene-like sequences mapped to regions containing major resistance genes or QTLs. Mol Genet Genomics 265:302–310

Webb CA, Richter TE, Collins NC, Nicolas M, Trick HN, Pryor T, Hulbert SH (2002) Genetic and molecular characterization of the Maize *rp3* rust resistance locus. Genetics 162:381–394

Whitham S, Dinesh-Kumar SP, Choi D, Hehl R, Corr C, Baker B (1994) The product of the tobacco mosaic virus resistance gene *N*: similarity to toll and the interleukin-1 receptor. Cell 78:1101–1115

Xu ML, Melchinger AE, Xia XC, Luebberstedt T (1999) High-resolution mapping of loci conferring resistance to sugarcane mosaic virus in maize using RFLP, SSR, and AFLP markers. Mol Gen Genet 261:574–581

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang ZX, Kono I, Kurata N, Yano M, Iwata N, Sasaki T (1998) Expression of *Xa1*, a bacterial blight resistance gene in rice, is induced by bacterial inoculation. Proc Natl Acad Sci USA 95:1663–1668

Yu J, Hu SN, Wang J, Wong GKS, Li SG, Liu B, Deng YJ, Dai L, Zhou Y, Zhang XQ, Cao ML, Liu J, Sun JD, Tang JB, Chen YJ, Huang XB, Lin WY, Tong W, Cong LJ, Geng JN, Han YJ, Li L, Li W, Hu GQ, Huang XG, Li WJ, Li J, Liu ZW, Liu JP, Qi QH, Liu JS, Li T, Wang XG, Lu H, Wu TT, Zhu M, Ni PX, Han H, Dong W, Ren XY, Feng XL, Cui P, Li XR, Wang H, Xu X, Zhai WX, Xu Z, Zhang JS, He SJ, Zhang JG, Xu JC, Zhang KL, Zheng XW, Dong JH, Zeng WY, Tao LY, Tan J, Ren XD, Chen XW, He J, Liu DF, Tian W, Tian CG, Xia HG, Bao QY, Li G, Gao H, Cao T, Zhao WM, Li P, Chen W, Wang XD, Zhang Y, Hu JF, Liu S, Yang J, Zhang GY, Xiong YQ, Li ZJ, Mao L, Zhou CS, Zhu Z, Chen RS, Hao BL, Zheng WM, Chen SY, Guo W, Li GJ, Liu SQ, Tao M, Zhu LH, Yuan LP, Yang HM (2002) A draft sequence of the rice genome (*Oryza sativa* L. *ssp indica*). Science 296:79–92