Theor Appl Genet (2002) 104:261–272

S. Peñuela · D. Danesh · N.D. Young

Targeted isolation, sequence analysis, and physical mapping of nonTIR NBS-LRR genes in soybean

Received: 8 May 2001 / Accepted: 30 June 2001

Abstract Most cloned plant disease resistance genes (R-genes) code for proteins belonging to the nucleotide binding site (NBS) leucine-rich repeat (LRR) superfamily. NBS-LRRs can be divided into two classes based on the presence of a TIR domain (*Toll* and interleukin receptor-like sequence) or a coiled coil motif (nonTIR) in their N-terminus. We used conserved motifs specific to nonTIR-NBS-LRR sequences in a targeted PCR approach to generate nearly 50 genomic soybean sequences with strong homology to known resistance gene analogs (RGAs) of the nonTIR class. Phylogenetic analysis classified these sequences into four main subclasses. A representative clone from each subclass was used for genetic mapping, bacterial artificial chromosome (BAC) library screening, and construction of RGA-containing BAC contigs. Of the 14 RGAs that could be mapped genetically, 12 localized to a 25-cM region of soybean linkage group F already known to contain several classical disease resistance loci. A majority of the genomic region encompassing the RGAs was physically isolated in eight BAC contigs, together spanning more than 1 Mb of genomic sequence with at least 12 RGA copies. Phylogenetic and sequence analysis, together with genetic and physical mapping, provided insights into the genome organization and evolution of this large cluster of soybean RGAs.

Keywords Soybean · Nucleotide binding site (NBS) · Resistance gene analog (RGA)

Communicated by M.A. Saghai Maroof

S. Peñuela · D. Danesh · N.D. Young (✉)
Department of Plant Pathology,
University of Minnesota, 495 Borlaug Hall,
1991 Upper Buford Circle, St. Paul, MN 55108, USA
e-mail: neviny@tc.umn.edu
Fax: +1-612-6259728

Introduction

In soybean, a genomic region rich in resistance (R) genes is found on molecular linkage group F (MLG-F). Genetic studies have placed more than seven R-loci in this region, including *Rpg*1, an R-locus conferring resistance to *Pseudomonas syringae* pv. *glycinea* (Ashfield et al. 1998) and *Rps*3, an R-locus against *Phytophthora sojae* (Diers et al. 1992). In addition, quantitative trait loci (QTL) against root-knot nematode *Meloidogyne javanica (Mj)* (Tamulonis et al. 1997a) and *Meloidogyne arenaria (Ma)* (Tamulonis et al. 1997b) have also been mapped to this region. Moreover, three virus R-loci map in this area; *Rsv*1, a locus conferring resistance to soybean mosaic virus (Yu et al. 1994), *Rpv*1, a locus conferring resistance to peanut mottle virus (Roane et al. 1983), and another against peanut stripe virus (Choi et al. 1989).

Several cases of R-gene clusters originally described by classical mapping have now been confirmed by sequence analysis of genomic DNA. Notable examples of R-genes that exist as clustered arrays include *Cf-4* and *Cf-9*, which confer resistance to different races of *Cladosporium fulvum* in tomato (Parniske et al. 1997), *Pto* in tomato (*Pseudomonas syringae* pv *tomato*), *Xa21* in rice (*Xanthomonas oryzae*) and *N* in tobacco (tobacco mosaic virus) (Martin et al. 1993; Song et al. 1995; Whitham et al. 1994). There are now more than 20 R-genes of known function that have been cloned and analyzed, and with the exception of *Hm1* of maize (Johal and Briggs 1992) and *Mlo* of barley (Buschges et al. 1997), they all seem to encode components of signal transduction systems (Michelmore and Meyers 1998).

A major class of cloned plant R-proteins contains nucleotide-binding site (NBS) and leucine-rich repeat (LRR) domains. NBS domains occur in diverse proteins with ATP or GTP binding activity (Saraste et al. 1990; Traut 1994). Nucleotide triphosphate binding is thought to alter the interaction between R-gene products and other members of the defense signal transduction pathway (Bent 1996). The NBS of plant R-proteins has a central

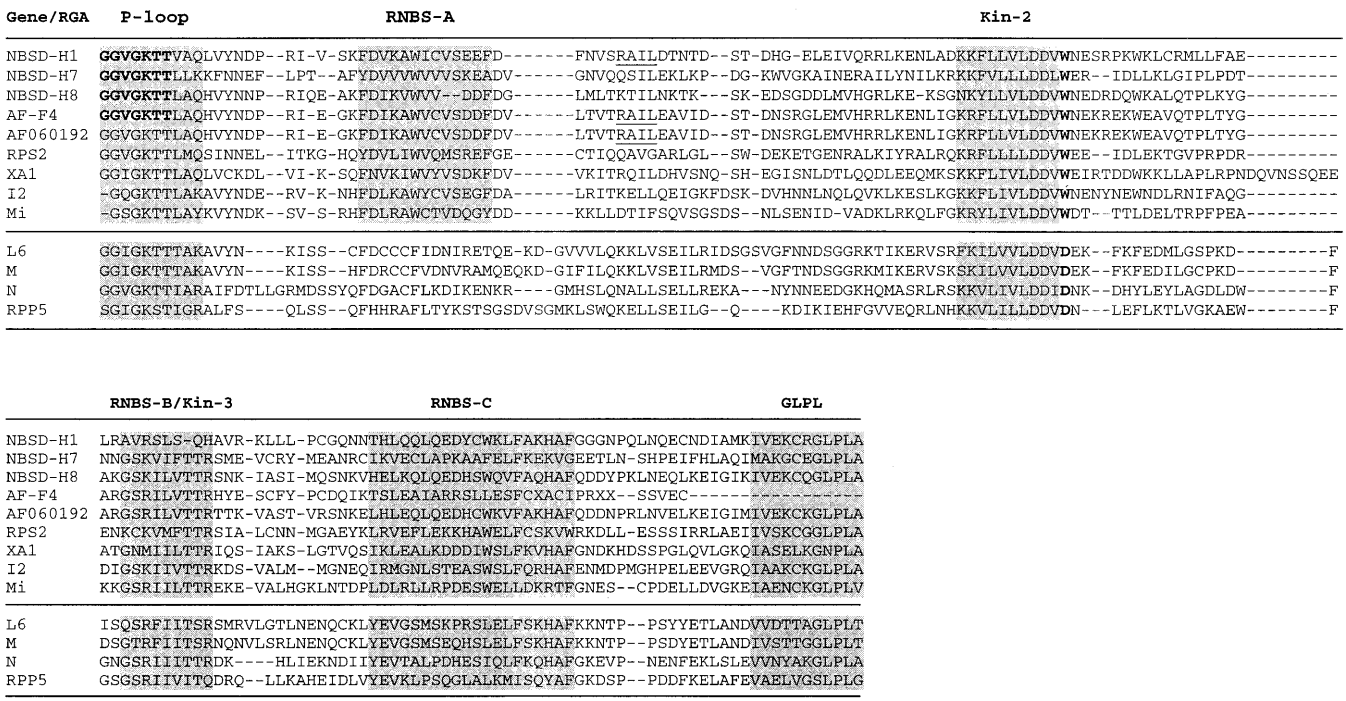

Fig. 1 Multiple sequence alignments (performed by CLUSTALW) of known R-proteins and representative soybean nonTIR RGAs: alignment of translated sequences of soybean RGAs (AF and NBSD) with nonTIR R-genes *RPS2, Xa1, I2, Mi* (top) and TIR genes *L6, M, N,* and *RPP5* (bottom). Amino acid sequences correspond to the NBS region from P-loop to GLPL motifs. Motif location highlighted according to Meyers et al. (1999). Residues of the P-loop motif in bold represent sequences derived from the forward primer used for isolation. *Underlined* residues show the 'RAIL' motif, unique to a subset of nonTIR sequences. *Letters in bold* indicate the tryptophan (*W*) and aspartic acid (*D*) residues characteristic of nonTIR and TIR sequences, respectively

region that constitutes the ATP- or GTP-binding pocket. Structurally related to regulators of animal apoptosis, including human Apaf-1 and nematode CED-4, this region functions as a module for protein-protein interactions (van der Biezen and Jones 1998). The LRR domain is a serial repeat motif of an approximately 24 amino acid motif with leucines and other hydrophobic residues at regular intervals, leading to a tertiary structure resembling a curved spring. The conserved leucines project into the hydrophobic core, while the other residues form a solvent-exposed surface involved in ligand binding (Kobe and Deisenhofer 1994), with a postulated role in plant-pathogen specificity (Ellis et al. 1997).

NBS-LRR proteins can be further subdivided into TIR and non-TIR proteins based on the presence or absence of an amino-terminal (N-terminal) TIR domain (Parker et al. 1997), named for its sequence similarity to the cytoplasmic signaling domains of animal innate immunity factors, *Toll* and interleukin-1 receptor (Rock et al. 1998). Plant NBS-LRR proteins without a TIR, termed nonTIR-NBS-LRR, have been reported to contain a coiled-coil motif (Pan et al. 2000). Coiled-coil structures form homo- or hetero-oligomeric associations facilitating interactions between proteins and possibly playing a role in the interaction of R-proteins with molecules downstream in the signal transduction pathway (Torii et al. 1998). A subset of nonTIR-NBS-LRRs have a leucine zipper (LZ), a specific example of the coiled-coil structure consisting of heptad repeat sequences with interspersed hydrophobic residues (Alber 1992).

TIR- and nonTIR-NBS-LRR sequences are distinguishable by amino acid motifs internal to their NBS domains. While motifs such as the P-loop, Kin-1a, and GLPLA signatures are present in both classes, motifs RNBS-A-TIR (LQKKLLSKLL) and RNBS-D-TIR (FLHIACFF) are found exclusively in the TIR class, while RNBS-A-nonTIR (FDLxAWVCVSQxF) and RNBS-D-nonTIR (CFLYCALFPED) are found exclusively in the nonTIR class (Meyers et al. 1999). Moreover, it is possible to distinguish the two classes with 95% accuracy by the final amino acid in motif Kin-2; a tryptophan (W) in nonTIRs and an aspartic acid (D) in TIRs (Fig. 1) (Meyers et al. 1999).

To isolate R-gene candidates in plants, well-conserved regions of the NBS domains have been used to design degenerate primers that amplify resistance gene analogs (RGAs). This approach has been used successfully in *Arabidopsis* (Aarts et al. 1998; Speulman et al. 1998), maize (Collins et al. 1998), bean (Creusot et al. 1999; Rivkin et al. 1999), potato (Leister et al. 1996), cereals (Leister et al. 1999), lettuce (Meyers et al. 1998) and several other plant species. Soybean was one of the first plant systems where this technique was applied (Kanazin et al. 1996; Yu et al. 1996). Degenerate primers were designed to amplify sequences from the P-loop to GLPLA (Kanazin et al. 1996) or from the P-loop to Kin-3 (Yu et al. 1996), uncovering many soybean RGAs. The products were split into 9 and 11 classes, respective-

Table 1 Oligonucleotide primers used to amplify non-TIR-NBS-LRR sequences

Degenerate primer	Sequence 5[a]´-3´	Motif consensus	T_m
LM638-For[b]	GGI GGI GTI GGI AAI ACI AC	GGVGKTT	58°
RNBSD-Rev	GGR AAI ARI SHR CAR TAI VIR AARC	CFLYCALFP	53°
Specific primer	**Sequence 5´–3´**	**Position in AF060192**	**T_m**
AF-For	GTT GGG AAG ACA ACG CTT GC	10	63°
AF-Rev	CAA CTC AAC ATT CAA CCG AGG	462	61°

[a] Degenerate IUB code: I, inosine; R, A or G; S, C or G; H, A or C or T, V, A or C or G
[b] Primer originally designed by Kanazin et al. (1996)

ly. Surprisingly, only two examples of nonTIR-NBS-LRRs were described (Yu et al. 1996).

RGAs are useful in physical mapping and as gene candidates in positional cloning. Bacterial artificial chromosome (BAC) libraries are frequently used in these studies because they are relatively simple to construct, inserts are easy to extract, and there is a low incidence of chimeras (Wang et al. 1995). Marek and Shoemaker (1997) identified BACs containing RGAs that belong to the TIR-NBS-LRR class in soybean. RGA-containing BACs were arranged into contigs through fingerprinting and DNA blot analyses, and the BACs were mapped to MLG-J of soybean. Between 9 and 12 RGA copies were found per contig, indicating these RGAs are clustered among several classical R loci (Polzin et al. 1994).

Another example of an R-gene cluster physically mapped in detail using BACs is the *Dm* gene cluster for resistance to downy mildew in lettuce (Meyers et al. 1998). BAC clones were identified using duplicated restriction fragment length polymorphism (RFLP) markers from the region. After fingerprinting and contig construction, 22 members of the RGC2 family were characterized from the BAC clones. The 22 RGC2 members were found to form a cluster spanning 3.5 Mb. A detailed study of this cluster indicated that complex rearrangements shaped this region, creating a highly diverse group of genes with both functional and non-functional gene family members (Meyers et al. 1998).

In order to isolate a representative sample of the important and underrepresented nonTIR NBS-LRR group of soybean, we used a targeted RGA amplification strategy based on motifs specific to nonTIR-NBS-LRRs. The resulting RGAs were subjected to sequence analysis and phylogenetic studies to elucidate the characteristics and relationships of this novel sub-class of RGAs in soybean. Selected sequences were then used for genetic mapping, BAC library screening, and physical mapping, providing insight into the genome organization and evolution of nonTIR-NBS-LRRs in soybean.

Materials and methods

Oligonucleotide primer design

Based on sequence alignments reported by Meyers et al. (1999), two amino acid motifs characteristic of the nonTIR-NBS-LRR class were used to design degenerate oligonucleotide primers for polymerase chain reaction (PCR) amplification of soybean genomic DNA. These motifs are present in the NBS-encoding region of several cloned R-genes such as *I2* (Simons et al. 1998), *Xa1* (Yoshimura et al. 1998), *RGC2B/Dm3* (Meyers et al. 1998), *RPS2* (Bent et al. 1994), *RPM1* (Grant et al. 1995), *RPS5* (Warren et al. 1998), *RPP8* (McDowell et al. 1998), and *Mi/Meu1* (Milligan et al. 1998). Primer LM638 from Kanazin et al. (1996) based on the P-loop motif of the NBS sequence (consensus GGVGKTT) was used as the forward degenerate primer (5´-GGIGGIGTIGGIAAIACIAC-3´). The reverse primer (5´-GGRAAIARISHRCARTAIVIRAARC-3´) was designed based on the RNBS-D-nonTIR motif (consensus CFLYCALFP) (Meyers et al. 1999) (Table 1). Clones derived from this primer set were designated with the abbreviation "NBSD".

Another set of RGAs was generated using a pair of specific primers designed to amplify a soybean RGA of the class nonTIR-NBS-LRR reported previously in GenBank, accession number AF060192 (Dong and Chen, 1998, direct submission). The forward primer, AF-For (5´-GTTGGGAAGACAACGCTTGC-3´) and reverse primer AF-Rev (5´-CAACTCAACATTCAACCGAGG-3´) correspond to the P-loop motif and a region preceding the GLPL motif, respectively. This database accession did not contain sequence data from the RNBS-D-nonTIR region, but based on internal sequence motifs, Meyers et al. (1999) classified the sequence as a nonTIR-NBS-LRR. Clones derived from this primer set were designated with the abbreviation "AF".

PCR amplification

For degenerate PCR amplification, genomic DNA was extracted from soybean cv. Faribault using a modified method of Dellaporta et al. (1983) described in Concibido et al. (1996). Fifty nanograms of genomic DNA was used as template in a 50-µl degenerate PCR reaction containing 2.5 m*M* $MgCl_2$, 0.2 m*M* dATP, 0.2 m*M* dCTP, 0.2 m*M* dGTP, 0.2 m*M* dTTP, 1 µ*M* of each primer (LM638-For and RNBS-D-NonTIR-Rev), 2.5 U *Taq* polymerase, and 1×PCR buffer (Life Technologies, Gaithersburg, MD). Amplification was carried out in an MJ Research PTC-100 thermocycler with the following program: 94°C for 3 min; 35 cycles of 94°C for 1 min, 45°C for 30 s and 72°C for 30 s; plus an extra elongation period of 10 min at 72°C.

For specific PCR amplification, 10 ng of soybean (cv. Faribault) genomic DNA was used in a 25-µl PCR reaction containing 1.5 m*M* $MgCl_2$, 0.1 m*M* dATP, 0.1 m*M* dCTP, 0.1 m*M* dGTP, 0.1 m*M* dTTP, 0.2 µ*M* of each primer (AF-For and AF-Rev), 1.5 U *Taq* polymerase, and 1×PCR buffer. The following PCR program was used for amplification: 94°C for 1 min; 35 cycles of 94°C for 10 s, 56°C for 30 s and 72°C for 1 min and 30 s; plus an additional elongation period of 7 min at 72°C.

Cloning of PCR products

PCR products were run on a 1% low-melting-point agarose gel for visualization. Bands of the appropriate sizes (700 bp for the degenerate PCR and 450 bp for the specific PCR reaction) were excised from the gel and purified using a QIAquick gel extraction column (QIAGEN, Valencia, Calif.). Each purified DNA band, presumably consisting of a mixture of products of similar sizes, was cloned into a plasmid vector using either the TOPO T/A Cloning kit (Invitrogen, Carlsbad, Calif.) or the pGEM-T-Easy

cloning kit (Promega, Madison, Wis.). Ninety-six transformed bacterial colonies were grown, and plasmid DNA was extracted with the Millipore (Bedford, Mass.) "High yield plasmid miniprep protocol" using MultiScreen 96-well filter plates. Recombinant plasmids were digested with *Eco*RI in a 20-µl reaction containing 0.1 mg/ml bovine serum albumin, 200 ng of plasmid DNA, 1×*Eco*RI reaction buffer, and 20 U *Eco*RI (New England Biolabs, Beverly, Mass.) to verify the presence of the appropriate size inserts.

DNA purification and sequencing of selected NBSD and AF clones

A total of 32 selected clones were processed from glycerol stocks, leading to the isolation of sequence-grade DNA with either the Wizard miniprep system (Promega) or the QIAspin miniprep kit (QIAGEN). One microgram of DNA from each clone and 3.2 pmol of M13 Forward (–20) or Reverse primer (Invitrogen) were used for sequencing at the Advanced Genetic Analysis Center at the University of Minnesota. Sequencing reactions were performed with the ABI Prism BigDye Terminator cycle sequencing ready reaction kit (PE Applied Biosystems, Foster City, Calif.) according to manufacturer's instructions.

Sequence analysis

Selected clones were fully sequenced and both forward and reverse sequence reads were assembled and analyzed using SEQUENCHER 3.0 (Gene Codes Corporation) and GENEWORKS 2.5.1 (Intelligenetics). Edited RGA sequences were compared to protein sequences in the GenBank NR database using BLASTX (Altschul et al. 1990, 1997).

Multiple sequence alignments

Phylogenetic analyses of nucleotide and amino acid sequences were performed using CLUSTALX (Thompson et al. 1994) for multiple sequence alignments and NJPLOT (Saitou and Nei 1987) for elaboration of neighbor-joining phylogenetic trees, both with default parameters. Both programs were implemented using public CLUSTALX software (Thompson et al. 1997). The P-loop region of the sequences was originally primer-derived and may not accurately represent the true sequence of this motif.

For the analysis of structural motifs present in RGA sequences, the program MEME was used (Bailey and Elkan 1994). MEME – Multiple EM for Motif Elicitation – is a tool for discovering motifs in a group of related DNA or protein sequences. Once motifs were identified with MEME, sequences were further analyzed using PROSITE (Hofmann et al. 1999), BLOCKS (Henikoff and Henikoff 1994), and PANAL (http://mgd.ahc.umn.edu/panal).

RFLP mapping of RGAs

PCR reactions from each of four representative RGA clones were purified using a Microcon column (Millipore) and labeled with [^{32}P] using a random hexamer labeling kit (Life Technologies). Parental DNA from PI 209332 and cv. Evans was used to prepare restriction enzyme survey blots (Southern 1975). DNA was individually digested with ten different restriction enzymes, *Bcl*I, *Bst*NI, *Dra*I, *Eco*RI, *Eco*RV, *Hae*III, *Hin*dIII, *Pst*I, *Taq*I, and *Xba*I, according to the manufacturer's instructions (New England Biolabs; Promega). Digested DNA was electrophoresed on a 1% agarose gel (with approximately 3 µg of DNA per lane) and transferred onto a Hybond N+ nylon membrane (Amersham, Arlington Heights, Ill.). Hybridizations were performed as described in Concibido et al. (1996). The sequence analysis program SEQUENCHER 3.0 was used to determine which restriction enzymes did not have sites internal to the sequences of the four representative RGA probes used in this study. The selected enzymes were also used to estimate copy number of each RGA sequence based on Southern blot hybridizations.

For segregation analysis, scorable fragment length polymorphisms were examined for each RGA clone/restriction enzyme combination. DNA was extracted from 90 recombinant inbred lines of a $F_{4:7}$ population derived from the cross PI209332 × Evans, and Southern blots of progeny DNA were prepared as described previously (Concibido et al. 1996). Bands were scored as dominant alleles and analyzed on a molecular linkage map consisting of 147 markers. Mapping data were analyzed as described previously (Concibido et al. 1996). Five anchor markers from the published soybean map (Cregan et al. 1999), A517_1, A186_1, Satt114, K644_1 and A708_1, were used in a detailed analysis of MLG-F. Anchor markers K387_1, Bng007_1, Bng062_1, A235, A668_1, Bng224 and CwP238 were used for the same purpose in the analysis of MLG-K.

BAC library screening

Vector-free RGA probes were generated by digestion of plasmid DNA with *Eco*RI. Inserts were run on a 1% low-melting-point agarose gel and purified with the QIAquick gel extraction system before radioactive labeling.

A soybean BAC library was previously constructed (Danesh et al. 1998) from high-molecular-weight DNA of soybean cv. Faribault. This library was expanded to approximately 73,000 individual clones, with an average insert size of 120 kb, and stored in 384-well microtiter dishes (Danesh, unpublished results). The 7.3 genome equivalent library was replicated onto four high-density 22.5 cm×22.5 cm colony blot filters at the Clemson University Genomic Institute (Clemson, S.C.). Four vector-free RGA probes were used to screen the high-density colony filters containing the entire BAC library. Hybridization was performed as described previously (Danesh et al. 1998).

Positive BAC clones were isolated by culturing an aliquot of each BAC glycerol stock on Luria's broth (LB) agar plates with 12.5 µg/ml of chloramphenicol. LB-plates were incubated at 37°C for 24 h. Single colonies were isolated and used for BAC DNA extraction following the "Modified QIAGEN R.E.A.L Prep 96 BAC Miniprep protocol for rapid extraction" alkaline lysis plasmid minipreps (QIAGEN).

BAC fingerprinting

Due to the large number of positive BAC clones resulting from library screening, only BAC clones with strong hybridizing signals were selected for fingerprinting.

BAC DNA (from cv. Faribault) was digested with one of the restriction enzymes originally used to map the corresponding RGA probe and electrophoresed following a method adapted from Marra et al. (1997). Gels were then stained with SYBR GOLD (Molecular Probes, Eugene, Ore.) and photo documentation created with a Kodak Digital Science 1D imaging system (Eastman Kodak, Rochester, N.Y.). Gels were blotted onto Hybond N+ nylon membranes and hybridized to the corresponding radiolabeled RGA probe. The banding pattern observed in the autoradiographs made it possible to group subsets of BACs into potentially overlapping "contigs".

To confirm the designations of overlapping BAC contigs, we ran another set of fingerprinting gels under the same conditions described earlier, but this time all BACs were digested with *Eco*RI. Digestion with *Eco*RI separated the vector completely from the BAC insert and allowed a more precise examination of the restriction patterns of BAC inserts for fingerprinting and contig construction. BACs that were thought to belong to the same contig based on hybridization were loaded next to each other in the gel for easier visualization. Stained gels were scanned with a Molecular Dynamics Storm Imager 840 set up for Blue fluorescence/ Chemifluorescence scanning to obtain a digital image of

the fingerprinting gels. Band calling and contig construction were performed using Image and FPC V2.6 software from the Sanger Center, Cambridge, UK (Soderlund et al. 1997; Sulston et al. 1989). After image analysis, gels were transferred onto nylon membranes and hybridized with the corresponding radiolabeled RGA probe to confirm BAC contig designations and to create a visual record corresponding to ordered BAC groups for physical mapping of RGAs in each contig.

Mapping of BAC contigs onto the genetic linkage map

BAC DNA (cv. Faribault) was digested with all the restriction enzymes originally used to map the corresponding RGA probe and then electrophoresed in parallel with digested soybean genomic DNA from cv. Faribault and from the parents of the mapping population (Evans and PI209332). Digestion reactions were electrophoresed following the method described in the previous section. Gels were transferred to nylon membranes and hybridized with the corresponding radiolabeled RGA probe.

If the size of a hybridizing band from the BAC clone was identical to the genomic fragment hybridizing to the RGA probe, we inferred that the BAC clone was derived from the same segment of the soybean genome. Moreover, if the RGA had been previously mapped as a polymorphic band by RFLP analysis, the map location of the BAC contig could then be inferred.

Faribault is a cultivar derived from PI209332 and was used for the construction of the BAC library. Therefore, since the mapping data was obtained from Evans × PI209332, it was necessary to compare the banding pattern of Faribault to that of the mapping parents. Only those fragments that had the same size in Faribault and the mapping parent were considered for mapping the corresponding BAC hybridizing bands.

Analysis of BAC RGA sequences

To characterize specific DNA sequences of RGAs found on BAC contigs mapping to MLG-F of soybean, we processed one BAC representative of each mapped contig as follows. BAC clones were used as templates for PCR amplification with degenerate primers (LM638-For and RNBS-D-nonTIR-Rev) to amplify RGA sequences present on that BAC clone. PCR conditions, subcloning, and sequencing were the same as those described earlier.

Twelve selected BAC-derived RGA clones were subjected to double pass sequencing in both directions. The first 350-bp sequences were used for multiple sequence alignments and phylogenetic tree construction as described earlier. Bootstrapping analysis was performed to estimate the confidence level of a particular phylogenetic tree topology using the CLUSTALX program (Thompson et al. 1997). The program generated a series of 1,000 pseudosamples by resampling the sites in the sample data with replacement of alignment positions.

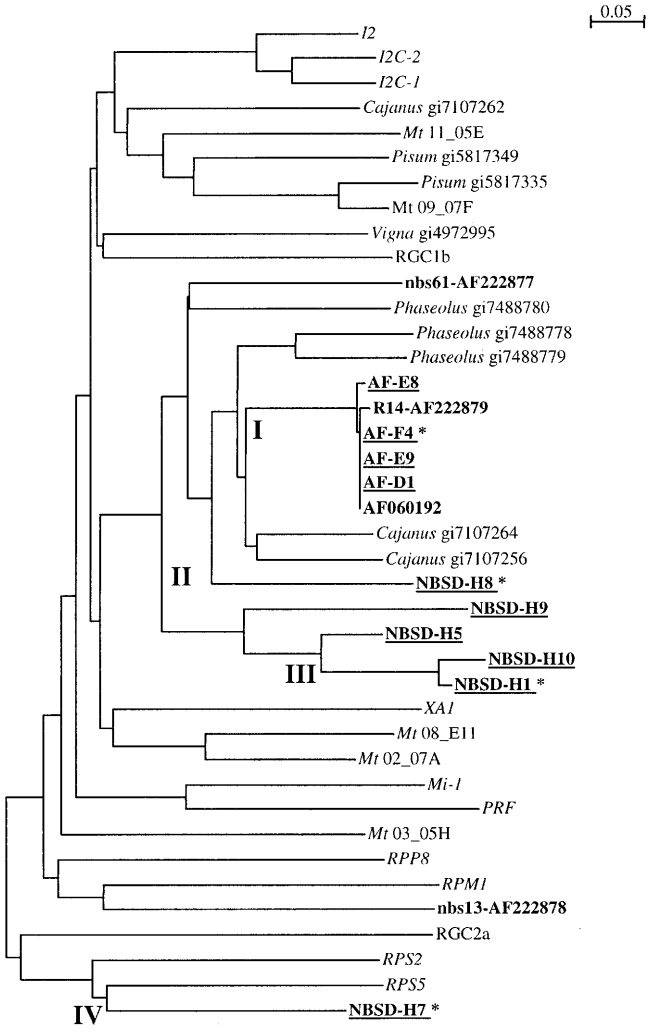

Fig. 2 Neighbor-joining tree based on CLUSTALW alignment of 40 nonTIR-NBS-LRR sequences from soybean, related legumes, and other plant species. Amino acid sequences from the P-loop to the Kin-3 (RNBS-B) motifs of the NBS domain were used for analysis. Branch lengths are drawn to scale. Nodes were supported by bootstrap values between 60% and 100% of 1,000 replicates (not shown). Sequences in *bold* correspond to soybean nonTIRs. *Roman numerals* indicate the four main classes of soybean nonTIRs uncovered in the present study (shown *underlined*). The four RGAs used for further studies are marked with *asterisks*

Results

Amplification and cloning of nonTIR-NBS-LRRs

Cloning of RGA amplification products led to 84 AF- and 108 NBSD clones. Of the 84 AF clones, 12 were picked randomly for verification of the correct size insert (approx. 450 bp). Four had inserts of the expected size and were selected for sequencing, while the remaining eight were discarded due to lack of inserts. Since these AF clones were not expected to exhibit substantial sequence diversity, only these four were pursued further: AF-D1, AF-E8, AF-E9, and AF-F4. From the NBSD clones, 24 random clones that had an insert of approximately 700 bp were sequenced.

Sequence analysis of RGA clones

Forty-nine single-pass sequences with an average edited length of 396 bp were obtained from 32 soybean NBS-LRR clones (including several sequenced in both directions). Phylogenetic analyses of all 32 clones classified them into four main classes of nonTIR-NBS-LRR sequences. A sample of ten RGA clones was fully sequenced for further analysis (Fig. 2).

Three out of these ten RGA clones had potential stop codons (NBSD-H5, NBSD-H9, and NBSD-H10), while the other seven were uninterrupted open reading frames (ORFs). All sequences had significant BLAST hits in the databases to R-genes and/or RGAs of the nonTIR-NBS-LRR class. As expected, the sequences showed only weak homology to previously cloned soybean TIR-NBS-LRR sequences (Kanazin et al. 1996; Yu et al. 1996). Therefore, the amplification strategy was successful at targeting sequences that contained both the P-loop and the nonTIR-NBS-LRR-specific RNBS-D motif. At the time of the final BLASTX analysis (April 6, 2000), four nonTIR soybean sequences were present in the public databases: AF060192 has been described earlier, AF222877 (clone nbs61) and AF222878 (nbs13) corresponded to two nonTIR-NBS-LRR RGAs reported by Yu et al. (1996); AF222879 (clone R14) had been reported by Hayes and Maroof (2000). Significant homology (E<1e-30) to accession AF222879 was observed with sequences NBSD-H1, NBSD-H8 and AF-F4. Accession AF222877 had somewhat lower homology values (E=approx. 1e-20) with clones AF-F4, NBSD-H1, and NBSD-H8. These two new soybean accessions were not the highest scoring BLASTX hits, and this may indicate that although related, they might belong to separate subclasses within the nonTIR-NBS-LRR sequences in the soybean genome. Accession AF222878 was not part of the BLASTX hits lists of any of the representative RGAs.

Nucleotide binding site motifs

As expected, alignments indicated regions of conserved homology (Fig. 1) corresponding to previously described motifs of NBS sequences, including P-loop, RNBS-A, Kin-2, RNBS-B (or Kin-3), RNBS-C, and GLPLA. The presence of these motifs was determined by visual inspection of alignments and confirmed by MEME analysis (Bailey and Elkan 1994).

In addition to these previously described motifs, a new conserved motif, "RAIL" (arginine, alanine, isoleucine, and leucine), was discovered between the RNBS-A motif and the Kin-2 motif. This "RAIL" signature is present in the sequences of AF soybean clones, AF060192, NBSD clones (except H7 and H8), and a modified version (RQIL) in *Xa*1 (Fig. 1). No putative function has been assigned to this particular motif according to searches in the PROSITE databases.

Similar analyses were performed to uncover putative functions for motif RNBS-D, since no previous reports were available in the literature for plant NBS genes. In this case, it was possible to search the motif and pattern databases with a longer motif that could be described as a 'pattern' rather than four specific amino acids, as was the case for "RAIL". The pattern most commonly used was C-[FV]-[LASPR]-Y-C-[TASG]-[LGIF]-[FY], which described the 'degenerate' version of the multilevel consensus CFLYCALFP. This pattern was matched 28 times in 28 sequences of TrEMBL (total number of Tr EMBL entries: 298,665), all corresponding to plant nonTIR-NBS-LRR R-genes or RGAs. This may indicate that the RNBS-D motif is a unique signature of nonTIR-NBS-LRR genes in plant species, but its function in signal transduction or other pathways related to disease resistance remains to be elucidated.

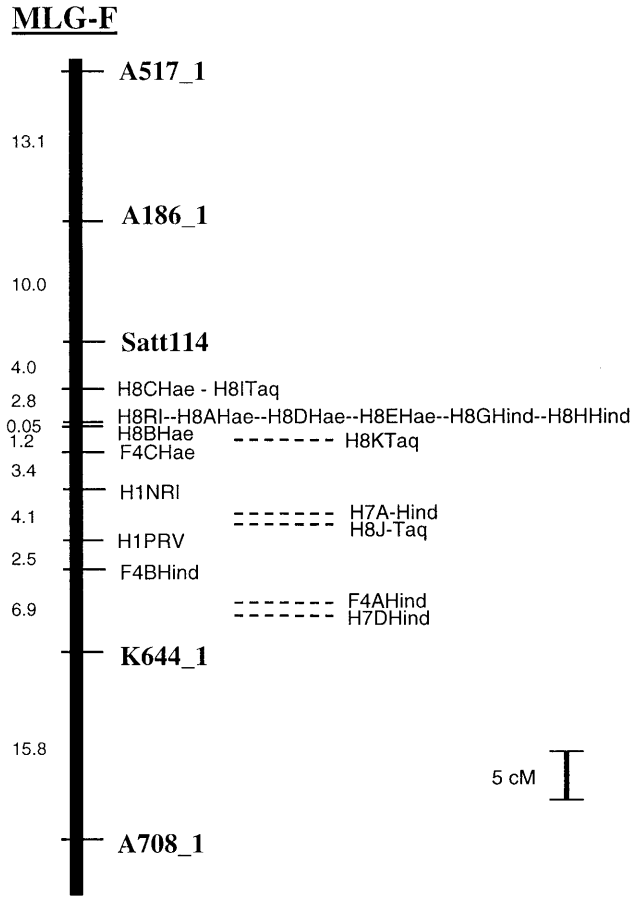

Fig. 3 RGAs on MLG-F of soybean. Eighteen NBS-LRR markers were mapped to 12 distinct loci on MLG-F of soybean (cosegregating markers are written in the *same line*). Markers connected to the *black bar* were mapped on 90 $F_{4:6}$ lines of a RIL population derived from PI209332 × Evans. Markers with *dotted lines* were binned to the anchor map with 44 lines of the mapping population

RFLP mapping

To represent the four main subclasses of nonTIR-NBS-LRR RGAs uncovered by sequence analysis, map locations of four chosen RGA probes (NBSD-H1, NBSD-H7, NBSD-H8 and AF-F4) were determined by segregation analysis in a soybean recombinant inbred population of 90 individuals. For each probe, several polymorphic bands were mapped using two or more restriction enzymes. Map locations for the RGA loci shown to map to MLG-F are illustrated in Fig. 3.

Altogether, there were 7 genomic copies of NBSD-H1, 5 of NBSD-H7, 7 copies of NBSD-H8, and 12 copies of AF-F4, for a total of at least 31 genomic copies of different intensities observed for the four RGA probes.

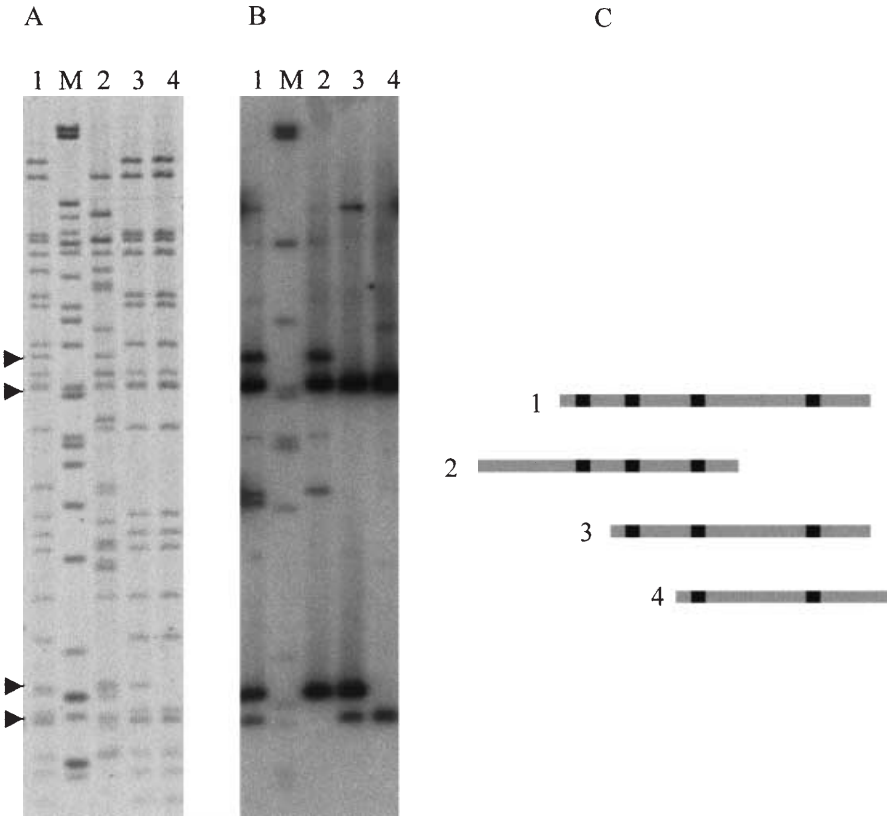

Fig. 4A-C Physical mapping of BAC contig NBSD-H1-ctgB. **A** shows a fingerprinting gel of the BACs cut with *EcoR*I, stained with SYBR-Gold, and scanned with STORM 840 Imager. *Lane M* DNA marker standard. Programs Image and FPC used common bands to build a tentative BAC contig. *Arrowheads* indicate bands that hybridized to the RGA probe. Panel B shows results from a hybridization of the Southern blot from panel A with radiolabeled NBSD-H1 probe. Four H1 copies are present in this contig. **C** A physical map of the contig, shows approximate locations of the RGA copies depicted as *black squares*. Approximate contig length: 214 kb

Of these, 14 loci could be placed on the genetic linkage map. Twelve of the mapped RGA loci were localized to MLG-F (Fig. 3) and two additional ones to MLG-K (data not shown).

BAC analysis of soybean RGAs

These same four nonTIR-NBS-LRR probes were also used to screen a 7.3× genome equivalent soybean BAC library; a total of 327 positive BACs were observed, 220 of which had strong hybridizing signals.

Compared to the number of genomic copies observed for each of the probes by Southern analysis, the number of positive BACs followed similar proportions. For example, H1 and H8 showed seven genomic bands on DNA blots and had around 90 positive BACs each, while F4 with 12 genomic bands uncovered 124 BACs. H7 hybridized only to 11 BACs, strongly suggesting that this RGA family is not widespread in the soybean genome. Furthermore, H7 did not exhibit significant homology to the other RGAs analyzed here.

As expected from sequence similarities observed among the four RGA probes, some cross-hybridization among probes was observed. Consistent with the high sequence similarity between H8 and F4 (77% nucleotide identity), these two probes shared the most positive BACs, with 82% of all H8 positives also being positive for H4. More moderate levels of cross-hybridization were observed among H1 and F4, and H8 and H1, ranging from 39% to 58% of shared BAC positives. Interestingly, 9% of the H8 positives were also positive for H7, a highly divergent RGA sequence. These common BACs represent 64% (7 of the 11) BACs positive for H7. This observation may be due to co-localization or to the fact that these two RGA probes (overall sequence similarity 3.5.%) have stretches of nucleotides towards the end of their sequences that appear to be identical in both RGAs (Fig. 1). The cross-hybridization results were taken into account during the process of contig building and were useful for mapping BAC contigs as discussed below.

Only a subset of the positive BACs with high intensity hybridization signals were chosen for further studies. Thirty-two BACs positive for NBSD-H1, 9 BACs positive for NBSD-H7, 50 BACs for NBSD-H8, and 32 BACs positive for AF-F4 were processed further. These 123 BACs represented a manageable sample size for the series of experiments described below.

After fingerprinting analysis, 14 contigs were constructed, incorporating a total of 66 BACs. From the remaining 57 BACs, 4 were either identical to another BAC in the contig or lacked a vector band after digestion with *Eco*RI. Fifty-three BACs did not belong to any contig as determined by fingerprinting and were not pursued further.

Image and FPC programs aided in the construction of physical configurations that represent the genomic organization of RGA copies in BAC contigs. Figure 4 shows the physical map of one BAC contig with a high density of RGAs and the tentative placement of RGA sequences.

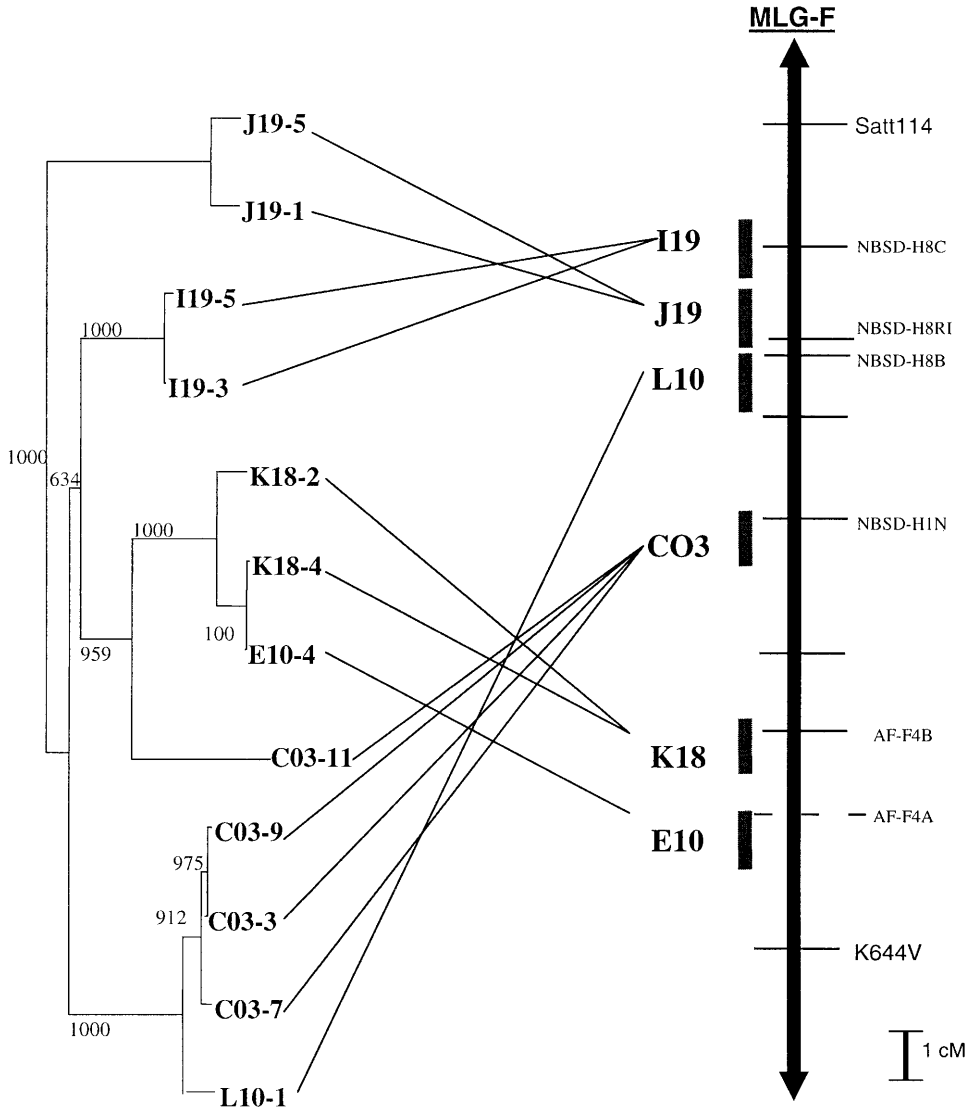

Fig. 5 Phylogenetic and physical relationships of soybean RGA copies in MLG-F. On the *left*, a neighbor-joining tree constructed using the 350-bp DNA sequences of the 12 cloned NBS PCR products amplified from six BACs representing the mapped contigs. On the *right*, the corresponding region of MLG-F map from Fig. 3, indicating the order of BACs in the map (not drawn to scale) and the corresponding RGA marker for each contig. Bootstrap values are shown as calculated for nodes supported with 60–100% of 1,000 replicates. Branch lengths were modified to fit figure

Numbers of RGAs per contig ranged from one to four (average, 2.1 RGAs), with contigs ranging in size from 82 kb to 224 kb (average, 156 kb). Calculating the number of RGAs per kilobase of genomic sequence, a minimum of one RGA per 175 kb and a maximum of one per 33 kb were observed among the BAC contigs.

Following procedures described in the Materials and methods section, seven BAC contigs were mapped to the genetic linkage map by matching the BAC RGA-hybridizing fragments to those of RFLP bands mapped previously (Fig. 3). Six contigs mapped to MLG-F (right panel of Fig. 5) and another one was mapped to MLG-K (data not shown).

Relation between physical and phylogenetic distance

To analyze the correspondence between physical and phylogenetic distance, we amplified RGA sequences directly from six BAC clones located in mapped contigs on MLG-F. A total of 21 RGA products was obtained, and the resulting subclones were sequenced and analyzed by BLASTX, confirming their identity as nonTIR-NBS-LRRs.

As a result of the experimental procedure used for the isolation of these RGA copies, some clones could have been duplicated sequences derived from the same RGA sequence in a BAC. In order to select only one sequence per RGA, preliminary phylogenetic analysis and pairwise comparisons were performed. This process reduced the number to 12 clearly unique sequences that were retained for further analysis, corresponding to 12 individual RGA copies from the BACs on MLG-F.

In the resulting neighbor-joining tree, there were four well-supported subfamilies (bootstrap values >90% of 1,000 replicates) and a single divergent member (C03–11) (Fig. 5). Overall, pairwise comparisons of the 12 RGA sequences revealed a high level of diversity within the RGAs present in the MLG-F cluster. Nucleotide sequence similarity ranged from 60% to 97%. Among subfamily members, similarity was especially high (93–97%), but in pairwise comparisons between

members of different subfamilies similarity ranged from 60% to 83%.

Comparing the results of phylogenetic data with map position demonstrated that phylogenetic relationship and physical position did not always show complete correspondence (Fig. 5). While RGA subclones from the contig with BACs I19 and J19 were 96–97% identical at the nucleotide level, contigs consisting of significantly divergent sequences were also observed. For example, subclone K18–4 was more similar to E10–4 (located on a neighboring contig) than to the other clone on the same BAC. Subclone C03–11 shared only 63–65% identity with neighboring RGAs on the same BAC and constituted a single divergent branch on the phylogenetic tree (Fig. 5). Subclone L10–1 also had a high percentage of similarity (>95%) to subclones from a different BAC, namely BAC C03. Given the uncertainties that remain in relating physical map positions to the genetic linkage map, some of these clones might still be examples of contiguous RGAs. For example, BACs K18, and E10, as well as L10 and CO3 are neighbors in the linkage map while still not grouping together into overlapping contigs. However, a lack of total correspondence between phylogenetic relationship and physical position of the RGAs has been observed previously with the *Dm3* cluster of lettuce (Meyers et al. 1998).

Discussion

Previous studies of RGAs in soybean uncovered several classes of TIR-NBS-LRR (Kanazin et al. 1996) but only two classes (b and j) of nonTIR-NBS-LRR (Yu et al. 1996). Jeong et al. (2001) uncovered additional members of class j derived from soybean genomic DNA and from a cDNA expression library. It is therefore significant that with the targeted approach used here for isolation of nonTIR RGAs, at least four new classes were successfully uncovered. Two additional classes, represented by NBSD-H5 and NBSD-H9, were also discovered but not analyzed further (Fig. 2).

Our results indicate that the approach of targeted PCR amplification was highly selective for nonTIR sequences, but may have targeted only a subset of all that exist in soybean. Therefore, it may be necessary to use a systematic approach with several different combinations of primers derived from various NBS motifs to uncover all nonTIR sequences present in the soybean genome. Such an approach was recently performed in tomato by Pan et al. (2000) who identified several classes of RGAs by systematically ‘scanning’ the genome with more than 17 different primer pairs.

Fifty-three singleton BACs that were not analyzed further could also contain additional copies of RGAs. Finally, 204 BAC clones beyond the 123 BACs analyzed in detail were identified as RGA-positives and could be used in future research for the isolation of additional RGAs in soybean.

Marek and Shoemaker (1997) used a different approach than the one described here to analyze RGAs on MLG-J of soybean. Representative BAC ends from contigs were obtained by cloning and used as RFLP probes for mapping with 56 individuals of a *Glycine max* × *G. soja* population. This approach faced the difficulty of having many RGA-specific BAC ends mapping to several positions on MLG-J, suggesting that the sequences were probably duplicated in this region. In our experiments, matching BAC bands were compared to all possible genomic bands previously mapped using a large set of enzymes. However, our strategy allowed for a more clear-cut localization of contigs and the possibility of confirming these locations when the BAC band matched several cosegregating bands (i.e., H8-ctgF). All BAC bands from the contigs were matched with a genomic band of the parental digest, confirming their identity as soybean RGA copies present somewhere in the genome. However, only those bands that matched polymorphic genomic bands could be mapped with confidence.

When a PCR-based approach is used to clone sequences that resemble disease R-genes, many products could potentially have no functional significance. Therefore, it is significant that the RGA clones selected for this study not only contained uninterrupted ORFs and high sequence homology to known R-genes of the nonTIR-NBS-LRR type, but also mapped to a region of soybean rich in R-loci (Fig. 3). Classic genetic studies have reported more than seven disease R-loci with map locations in MLG-F. Furthermore, other reports (Jeong et al. 2001; Yu et al. 1996) have confirmed the existence of a cluster of nonTIR sequences in this chromosomal region. It is possible that the RGA sequences described here correspond to members of these R-gene clusters and could be considered gene candidates. At the very least, the BAC contigs described here represents excellent starting points for positional cloning.

Physical organization of nonTIR-NBS-LRRs clusters

Previous reports of RGA organization indicated a wide variation in the spacing of NBS-LRR genes. Meyers et al. (1998) reported only one RGA copy per BAC (average 120 kb) in the *Dm3* region of lettuce, while Wei et al. (1999) found up to 11 copies in the 240-kb region of *Mla* in barley. RGAs in soybean of the type TIR-NBS-LRR were found in clusters as six RGAs over 300-kb and nine RGAs over 400-kb DNA fragments (Marek and Shoemaker 1997). *Arabidopsis thaliana* was reported to have a cluster of five NBS-LRRs in a 120-kb region (Aarts et al. 1998), which is similar to what was found in barley (average spacing of approx. 20 kb). Therefore, smaller genomes do not necessarily harbor more dense clusters, and even in the same species, there can be wide variation in genome organization of RGA clusters.

Examining the organization of the nonTIR-NBS-LRR RGAs among soybean contigs, it is clear that copy number varies and that the distribution of the RGAs in differ-

ent DNA segments do not follow distinctive patterns. In some cases, where contigs span two or more copies, sequences seem to be arranged in tandem (only a few kilobases apart) (Fig. 4), while others can be spaced by as much as 70 kb apart. Wider spacing is less common than tandem arrays, indicating that more tightly clustered RGA regions may predominate.

Mechanisms of R-gene evolution

The best known mechanisms for generating diversity are duplication and recombination, where a gene family is formed by chromosomal duplication and subsequent divergence of progenitor sequences. This process generates new loci, changes the number of family members, and provides homologous sequences for recombination and unequal crossing-over events (Richter and Ronald 2000). Although we focused on the NBS domain of R-genes, not the more highly variable LRR domain, we observed clusters composed of sequences with high sequence similarity as well as clusters composed of largely divergent sequences. In the case of RGAs arranged in tandem with high sequence similarity, such as J19–5/J19–1 (96% identical) and I19–3/I19–5 (97% identical), simple duplication may be involved. However, clusters with more divergent sequences suggest that other types of rearrangements may have also occurred. Still, the rate of sequence exchange among cluster members apparently did not lead to sequence homogenization, a predicted outcome of concerted evolution (Dover 1982). Ectopic (interlocus) recombination plus gene conversion could have also played an important role in the generation of sequence divergence among members of this gene family. Parniske and Jones (1999) reported this phenomenon among members of a gene family of homologues of the *Cladosporium fulvum* R-gene *Cf-9* (*Hcr9* genes) in tomato.

Potential applications of soybean RGAs

Several applications should now be possible with this report of nonTIR-NBS-LRR sequences in MLG-F of soybean. Research groups pursuing the positional cloning of the seven disease R loci already known to be located on MLG-F by classical genetic studies will benefit from the RGA clones as either candidate genes or tightly linked DNA markers. Still, the diversity of disease resistance loci present in this region of MLG-F presents a challenge for distinguishing among candidate genes potentially involved in resistance to different pathogens. Therefore, better mapping information is required for a more accurate determination of linkage between RGAs and defined R-loci. For this purpose it will be necessary to analyze the new RGAs in large populations segregating for the various disease reactions. In this regard, allele-specific markers that can differentiate polymorphisms between haplotypes should be designed and tested. These markers would be valuable in marker-assisted selection programs, providing rapid indication of the haplotype of the MLG-F cluster in a particular soybean breeding line, and would aid in the selection of desirable combinations of R-gene specificities. Such markers could also be used in high-resolution genetic mapping as an invaluable tool in map-based cloning efforts.

Finally, it is essential to determine the sequences flanking the P-loop-RNBS-D segment analyzed in detail here, thereby extending the analysis to the predicted coiled-coil and LRR regions. Obtaining full-length sequences of these RGAs would allow studies on the evolution of the diverse regions of R-gene sequences and the implications in plant-pathogen interactions. Eventually, extensive DNA sequencing of the entire region should determine the exact genomic structure and characteristics of this important cluster. In the present study, the first step towards the construction of BAC contigs that overlap the areas of interest has already been completed.

Acknowledgements We thank László Gyenis, Steve Cannon, Roxanne Denny, Joam Mudge, Dawn Foster-Hartnett, Karin Larson and Daniel Gulya for help and advice throughout the project. We also thank Dr. Deborah Samac for advice and comments about the manuscript.

This research was supported in part by grants U.S. Department of Agriculture-CSREE/98-3599-6168 and National Science Foundation DBI-9872664.

This paper is published as a contribution to the series of the Minnesota Agricultural Experiment Station.

References

Aarts MG, Hekkert B, Holub EB, Beynon JL, Stiekema WJ, Pereira A (1998) Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. Mol Plant-Microbe Interact 11:251–258

Alber T (1992) Structure of the leucine zipper. Curr Opin Genet Dev 2:205–210

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Altschul SF, Madden RL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Ashfield T, Danzer JR, Held D, Clayton K, Keim P, Saghai-Maroof MA, Webb DM, Innes RW (1998) *Rpg*1, a soybean gene effective against races of bacterial blight, maps to a cluster of previously identified disease resistance genes. Theor Appl Genet 96:1013–1021

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the second international conference on intelligent systems for molecular biology. Menlo Park, California. AAAI Press. pp 28–36

Bent AF (1996) Plant disease resistance genes: function meets structure. Plant Cell 8:1757–1771

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

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

Choi SH, Green SK, Lee DR (1989) Linkage relationship between two genes conferring resistance to peanut stripe virus and soybean mosaic. Euphytica 44:163–169

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

Concibido VC, Denny RL, Lange DA, Orf JH, Young ND (1996) RFLP mapping and marker-assisted selection of soybean cyst nematode resistance in PI 209332. Crop Sci 36:1643–1650

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean. Crop Sci 39:1464–1490

Creusot F, Macadre C, Ferrier-Cana E, Riou C, Geffroy V, Sevignac M, Dron M, Langin T (1999) Cloning and molecular characterization of three members of the NBS-LRR subfamily located in the vicinity of the *Co-2* locus for anthracnose resistance in *Phaseolus vulgaris.* Genome 42:254–264

Danesh D, Penuela S, Mudge J, Denny RL, Nordstrom H, Martinez JP, Young ND (1998) A bacterial artificial chromosome library for soybean and identification of clones near a major cyst nematode resistance gene. Theor Appl Genet 96: 196–202

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–21

Diers BW, Mansur L, Imsande J, Shoemaker RC (1992) Mapping *Phytophthora* resistance loci in soybean with restriction fragment length polymorphism markers. Crop Sci 32:377–383

Dover GA (1982) Molecular drive, a cohesive mode of species evolution. Nature 299:111–117

Ellis J, Lawrence G, Ayliffe M, Anderson P, Collins N, Finnegan J, Frost D, Luck J, Pryor T (1997) Advances in the molecular genetic analysis of the flax-flax rust interaction. Annu Rev Phytopathol 35:271–291

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW, Dangl JL (1995) Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. Science 269:843–846

Hayes AJ, Maroof MAS (2000) Targeted resistance gene mapping in soybean using modified AFLPs. Theor Appl Genet 100: 1279–1283

Henikoff S, Henikoff JG (1994) Protein family classification based on searching a database of blocks. Genomics 19:97–107

Hofmann K, Bucher P, Falquet L, Bairoch A (1999) The PROSITE database, its status in 1999. Nucleic Acids Res 27:215–219

Jeong SC, Hayes AJ, Biyashev RM, Saghai Maroof MA (2001) Diversity and evolution of a non-TIR-NBS sequence family that clusters to a chromosomal "hotspot" for disease resistance in soybean. Theor Appl Genet 103:406–414

Johal GS, Briggs SP (1992) Reductase activity encoded by the *HM1* disease resistance gene in maize. Science 258:985–987

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

Kobe B, Deisenhofer J (1994) The leucine-rich repeat: a versatile binding motif. Trends Biochem Sci 19:415–421

Leister D, Ballvora A, Salamini F, Gebhardt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nat Genet 14:421–428

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

Marek LF, Shoemaker RC (1997) BAC contig development by fingerprint analysis in soybean. Genome 40:420–427

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, MacDonald KM, Hillier LW, McPherson JD, Waterson RH (1997) High throughput fingerprint analysis of large-insert clones. Genome Res 7:1072–1084

Martin GB, Brommonshcenkel 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

McDowell JM, Dhandaydham M, Long TA, Aarts MG, Goff S, Holub EB, Dangl JL (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of *Arabidopsis.* Plant Cell 10:1861–1874

Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10:1817–1832

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

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113–1130

Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P, Williamson VM (1998) The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. Plant Cell 10:1307–1319

Pan Q, 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

Pan QL, Liu YS, Budai-Hadrian O, Sela M, Carmel-Goren L, Zamir D, Fluhr R (2000) Comparative genetics of NBS-LRR resistance gene homologues in the genomes of two dicotyledons: tomato and *Arabidopsis.* Genetics 155:309–322

Parker JE, Coleman M, Szabo V, Frost L, Schmidt R, Biezen Evd, Moores T, Dean C, Daniels MJ, Jones JDG (1997) The *Arabidopsis* downy mildew resistance gene *RPP5* shares similarity to the Toll and Inerleukin-1 receptors with *N* and *L6.* Plant Cell 9:879–894

Parniske M, Jones JDG (1999) Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family. Proc Natl Acad Sci USA 95:5850–5855

Parniske M, Hammond-Kosack KE, Goldstein C, Thomas CM, Jones DA, Harrison K, Wulff BBH, Jones JDG (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell 91:821–832

Polzin K, Lohnes D, Nickell C, Shoemaker RC (1994) Integration of *Rps2, Rmd,* and *Rj2* into linkage group J of the soybean molecular map. J Hered 85:300–303

Richter TE, Ronald PC (2000) The evolution of disease resistance genes. Plant Mol Biol 42:195–204

Rivkin MI, Vallejos CE, McClean PE (1999) Disease-resistance related sequences in common bean. Genome 42:41–47

Roane CW, Tolin SA, Buss GR (1983) Inheritance of reaction to two viruses in the cross 'York'x 'Lee 68'. J Hered 74:289–291

Rock FL, Hardiman G, Timans JC, Kastelein RA, Bazan JF (1998) A family of human receptors structurally related to *Drosophila* Toll. Proc Natl Acad Sci USA 95:588–593

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Saraste M, Sibbald PR, Wittinghofer A (1990) The P-loop – a common motif in ATP- and GTP- binding proteins. Trends Biochem 15:430–434

Simons G, Groenendijk J, Wijbrandi J, Reijans M, Groenen J, Diergaarde P, van-der-Lee T, Bleeker M, Onstenk J, Both Md, Haring M, Mes J, Cornelissen B, Zabeau M, Vos P (1998) Dissection of the *Fusarium I2* gene cluster in tomato reveals six homologs and one active gene copy. Plant Cell 10:1055–1068

Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. CABIOS 14:523–535

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

Southern E (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98: 503–517

Speulman E, Bouchez D, Holub EB, Beynon JL (1998) Disease resistance gene homologs correlate with disease resistance loci of *Arabidopsis thaliana*. Plant J 14:467–474

Sulston J, Mallet F, Durbin R, Horsnell T (1989) Image-analysis of restriction enzyme fingerprint autoradiograms. CABIOS 5: 101–105

Tamulonis JP, Luzzi BM, Hussey RS, Parrot WA, Boerma HR (1997a) DNA markers associated with resistance to javanese root-knot nematode in soybean. Crop Sci 37:783–788

Tamulonis JP, Luzzi BM, Hussey RS, Parrott WA, Boerma HR (1997b) DNA marker analysis of loci conferring resistance to peanut root-knot nematode in soybean. Theor Appl Genet 95: 664–670

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Torii KU, McNellis TW, Deng XW (1998) Functional dissection of *Arabidopsis* COP1 reveals specific roles of its three structural modules in light control of seedling development. EMBO 17:5577–5587

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

van der Biezen EA, Jones JDG (1998) The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Curr Biol 8:R226-R227

Wang GL, Holsten TE, Song WY, Wang HP, Ronald PC (1995) Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance locus. Plant J 7:525–533

Warren RF, Henk A, Mowery P, Holub E, Innes RW (1998) A mutation within the leucine-rich repeat domain of the *Arabidopsis* disease resistance gene *RPS5* partially suppresses multiple bacterial and downy mildew resistance genes. Plant Cell 10: 1439–1452

Wei F, Gobelman-Werner K, Morroll SM, Kurth J, Mao L, Wing R, Leister D, Schulze-Lefert P, Wise RP (1999) The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5 S (1HS) of barley. Genetics 153:1929–1948

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

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 YG, Saghai-Maroof MA, Buss GR, Maughan PJ, Tolin SA (1994) RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. Phytopathology 84:60–64

Yu YG, Buss GR, Saghai-Maroof MA (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proc Natl Acad Sci USA 93:11751–11756