Identification of R-Gene Homologous DNA Fragments Genetically Linked to Disease Resistance Loci in *Arabidopsis thaliana*

Mark G. M. Aarts,¹ Bas te Lintel Hekkert,¹ Eric B. Holub,² Jim L. Beynon,³ Willem J. Stiekema,¹ and Andy Pereira¹

¹Department of Molecular Biology, DLO-Centre for Plant Breeding and Reproduction Research, Postbus 16, 6700 AA Wageningen, The Netherlands; ²Plant Pathology and Weed Science Department, Horticultural Research International-Wellesbourne, Warwickshire CV35 9EF, U.K.; ³Department of Biological Sciences, Wye College, University of London, Wye, Ashford, Kent TN25 5AH, U.K.

Accepted 30 December 1997.

Disease resistance in plants is a desirable economic trait. A number of disease resistance genes from various plant species have been cloned so far. The gene products of some of these can be distinguished by the presence of an N-terminal nucleotide binding site and a C-terminal stretch of leucine-rich repeats. Although these gene products are structurally related, the DNA sequences are poorly conserved. Only parts of the nucleotide binding site share enough DNA identity to design primers for polymerase chain reaction amplification of related DNA sequences. Such primers were used to amplify different resistance-gene-like (RGL) DNA fragments from *Arabidopsis thaliana* accessions Landsberg *erecta* and Columbia. Almost all cloned DNA fragments were genetically closely linked with known disease resistance loci. Most RGL fragments were found in a clustered or dispersed multi-copy sequence organization, supporting the supposed correlation of RGL sequences and disease resistance loci.

The plant disease resistance genes (R-genes) that have been molecularly characterized so far can be grouped into several classes based on similarities in the function or amino acid sequence of the proteins they encode. The class containing the majority of cloned R-genes is characterized by the presence of an N-terminal nucleotide binding site (NBS) and a C-terminal stretch of leucine-rich repeats (LRRs). A number of genes belonging to this class have been cloned, such as the *RPS2* (Bent et al. 1994; Mindrinos et al. 1994) and *RPM1* (Grant et al. 1995) genes (against *Pseudomonas syringae*) and the *RPP5* gene (against *Peronospora parasitica*) (Parker et al. 1997) from *Arabidopsis thaliana*, the *N* gene from tobacco (against tobacco mosaic virus) (Whitham et al. 1994), the *PRF* gene from tomato (against *Pseudomonas syringae*) (Salmeron et al. 1996), and the *L⁶* gene from flax (against *Melampsora lini*) (Lawrence et al. 1995). Although these R-genes are found in diverse plant species and are active against a range of pathogens, the conserved NBS-LRR features suggest a common function in the defense response against pathogen attack,

probably as part of the signal transduction pathway (Staskiewicz et al. 1995).

The isolation and characterization of more of these genes is important because they may provide clues about the complex mechanisms of resistance, the interactions involved in pathogen recognition, and the evolution of the R-genes. Furthermore, cloned genes can be transferred to other species (Rommens et al. 1995; Thilmony et al. 1995) to study the resistance mechanism in a completely different genetic background. It is known from hybridization experiments with the cloned NBS-LRR type of R-genes that several related sequences can be found, both in the same species as well as in other species (Lawrence et al. 1995; Mindrinos et al. 1994; Parker et al. 1997; Salmeron et al. 1996; Whitham et al. 1994). Isolation of these sequences by polymerase chain reaction (PCR) may be attainable. Despite the general lack in DNA sequence conservation between R-genes, there are a few conserved protein motifs present in the NBS. With degenerate primers based on these homologous regions, it is possible to amplify several resistance-gene-like (RGL) DNA fragments, as has been shown for soybean and potato (Kanazin et al. 1996; Leister et al. 1996; Yu et al. 1996). Some of these DNA fragments have been mapped in the vicinity of known disease resistance loci. We applied a similar approach in *A. thaliana* with a degenerate primer based on the conserved amino acid sequence of the so-called P-loop motif and another degenerate primer downstream based on the sequence of the conserved domain 5 (for definition of these conserved motifs see Lawrence et al. 1995). *A. thaliana* is a good model species to test the general applicability of a PCR-based isolation method, as many of the plant-pathogen interactions in this species have been studied (Crute et al. 1994), the genetic positions of many resistance loci have been determined (Kunkel 1996; Holub 1997), and some resistance gene homologues are represented in the extensive EST data base.

RESULTS

Amplification and cloning of genomic DNA with R-gene-based degenerated PCR primers.

The widely used and well-studied *A. thaliana* accessions Columbia (Col) and Landsberg *erecta* (Ler) were chosen for

Corresponding author: A. Pereira; Fax: +31-317-418094; E-mail: a.pereira@cpro.dlo.nl

PCR amplification of RGL sequences as they show resistance to a range of pathogens and, above all, are used to generate a Col×Ler population of recombinant inbred lines (RILs) available for mapping (Lister and Dean 1993). For the PCR, degenerate primers RG1 and RG2 were used whose sequences were based on the conserved P-loop and domain 5 region of the NBS in the *N*, *L⁶*, and *RPS2* R-genes from tobacco, flax, and *A. thaliana*, respectively (Fig. 1). In the absence of introns in RGL genes, the primers are expected to amplify DNA fragments of around 530 bp. From Col and Ler DNA, fragments of around 0.5 and 0.8 kb, respectively, were amplified, gel purified, and cloned. The cloned fragments were distinguished by restriction analysis and the DNA sequences of all different clones were determined.

Sequence comparison of R-gene homologous DNA fragments.

In total, four fragments from Col (C1 to C4) and four from Ler (L1 to L4) were sequenced. These fragments were grouped in three classes (C1 [0.5 kb]; L1, L2, C2, C3 [0.5 kb]; and L3, L4, C4 [0.8 kb]) based on their DNA sequence similarities. All sequences have the RG1 primer sequence on one end and the RG2 sequence on the other end. Upon screening GenBank with BLASTX (Altschul et al. 1990) for similar sequences, the derived amino acid sequence of all cloned fragments showed similarity to known plant R-gene products, such as from the *N*, *L⁶*, *PRF*, *RPS2*, and *RPM1* genes, confirming their identity as RGL sequences. The RGL fragments C1, L2, C2, and C3 of around 0.5 kb were found to encode one long open reading frame extending over the entire length of the fragment (Fig. 2). The deduced amino acid sequence of the 0.5-kb fragments was most similar to the *RPS2* (for L2, C2 and C3) or the *L⁶* or *N* gene products (for C1; Fig. 2). Fragment L1 was the only fragment of 0.5 kb without a full-length open reading frame, and although the coding parts shared similarity with the *RPS2* gene product, the presence of frame shifts and stop codons suggested that this fragment was not part of a functional gene and it was therefore not included in Figure 2. For the three 0.8-kb fragments L3, C4, and L4 the similarities were found in different reading frames, interrupted by nonhomologous sequences. Only the sequences with homology to R-genes were conserved between these three clones. The nonhomologous sequences are probably introns, as they are flanked by *A. thaliana* intron-exon motifs (Brown 1996). Upon removing these supposed introns, one open reading frame could be constructed for each fragment that corresponded in length to the open reading frames of the other RGLs. The deduced amino acid sequences of these 0.8-kb RGLs had most similarity to an *A. thaliana* gene product proposed to be a myosin heavy chain homologue and to the *RPM1* gene product. The myosin heavy chain homologue (GenBank accession no. U19616) shows only very poor homology to myosin heavy chains when tested with BLASTX ($P > 0.99$), but it displays all the characteristics of an NBS-LRR type of R-gene.

Cosegregation of R-gene DNA fragments with known resistance loci.

We decided to genetically map the cloned RGL DNA fragments to assess if any of these were genetically linked to a known R-gene locus. For mapping, a population of RILs made

from accessions Col and Ler (Lister and Dean 1993) was used. Most RGL probes revealed hybridizing signals of various intensities on genomic DNA blots, reflecting differences in homology (see for example Figure 3A and B). The number of RGL copies was predicted according to the number of hybridizing restriction fragments found with various restriction digests. The most discriminatory restriction fragment length polymorphism (RFLP) patterns were used for mapping the RGL fragments.

Fragment C3 mapped to chromosome 1 in the region where the *RPS5* locus against *Pseudomonas syringae* has been mapped previously (Simonich and Innes 1995) (Fig. 4). The C3 probe detected one major, Col-specific, DNA copy and a cosegregating RFLP with weaker hybridization signal in both Ler and Col. The C3 fragment turned out to be actually part of the cloned *RPS5* gene (R. W. Innes and R. F. Warren, *personal communication*). Another RFLP with a weak hybridization signal was visible as the major signal after hybridizing with the L2 or the C2 probe. However, these L2 and C2 fragments do not map to chromosome 1 but to chromosome 4 in the vicinity of the *RPP4* and *RPP5* loci against *Peronospora parasitica* in Col and Ler (Parker et al. 1993; Tör et al. 1994). The Col DNA sequence of this L2/C2 locus is part of a large piece of over 200 kb of genomic DNA known as the ATFC1 locus (accession no. Z97336), which sequence has been determined by the European Scientists Sequencing Arabidopsis (ESSA). Based on the sequence information, the C2 fragment is clearly not part of the closely linked *RPP5* gene or gene cluster

A

B

<i>N</i>	216	G	M	G	G	V	G	K	T	T	
<i>L</i> ⁶	265	G	M	G	G	I	G	K	T	T	
<i>RPS2</i>	182	G	P	G	G	V	G	K	T	T	
consensus		G	M	G	G	V	G	K	T	T	
RG1-primer	5'	ggiatgggiggigtiggiaaracnacn 3'									
<i>N</i>	384	G	L	P	L	A	L	K	V	W	G
<i>L</i> ⁶	435	G	L	P	L	T	L	K	V	I	G
<i>RPS2</i>	347	G	L	P	L	A	L	I	T	L	G
consensus		G	L	P	L	A	L	K	V	L	G
RG2-primer	3'	ccwraiggiraicgiraitycgaigaiacci 5'									

Fig. 1. Degenerate polymerase chain reaction (PCR) primers RG1 and RG2 used for the amplification of resistance-gene-like DNA sequences. **A**, Schematic model of the nucleotide binding site/leucine-rich repeat (NBS-LRR) type of disease resistance genes. Two conserved domains within the NBS region (gray box), used to design PCR primers RG1 and RG2 (arrows), are indicated with P (P-loop) and 5 (conserved domain 5). In the absence of introns, the distance between the primers is 0.5 kb. **B**, Amino acid sequences of the P-loop (top) and conserved domain 5 (bottom) of the *N*, *L⁶*, and *RPS2* gene products (according to Lawrence et al. 1995). Amino acid consensus sequences are used to design the RG1 and RG2 primers. I = inosine, R = A or G, N = A or T or C or G, Y = T or C, W = A or T.

(Parker et al. 1997). The cloned L2 and C2 fragments are both present as a single copy in Col and Ler.

The L4 and C4 fragments also mapped to chromosome 1, between the RFLP markers m213 and g4026 (Fig. 4). This is very close to where the *RPP7* locus against *P. parasitica* has been mapped in Col and Ler (C. Can, E. Holub, and J. Beynon, *unpublished*). The L4 and C4 probes revealed the same RFLP pattern for five restriction digests (e.g., for *EcoRI* see Figure 3A and B). All RFLP bands cosegregated among the 95 RILs tested. At this clustered multi-copy locus, Col and Ler contained, respectively, at least five and three L4/C4 homologous DNA copies.

Multiple segregating RFLP bands were detected by the L3 probe. The RFLP fragment with the most intense hybridization signal mapped only a few cM away from the *L4/C4* locus on chromosome 1 (Fig. 4). Col and Ler both contain one copy at this *L3A* locus. Based on the hybridization intensity, this is most likely the locus from which the cloned L3 fragment is derived. The other three loci mapped to chromosome 5. Loci *L3C* and *L3D* mapped 6 cM apart between markers g4028 and m435 (Fig. 4). The single-copy *L3D* locus, found in both Col and Ler, mapped close to where the resistance against *P. parasitica* isolate Aswal has been mapped in Ler (E. Holub, *unpublished*). The *L3C* locus, detected by a single, Ler-specific RFLP fragment, mapped near *RPP8* (Holub et al. 1994) and very close to where the resistance loci against *P. parasitica*

isolates Waco8, Waco11, and Boma1 were mapped in Ler (M. Aarts, *unpublished*). A genomic cosmid that contains L3/CK1 fragments from Ler confers resistance to Emco5 in transgenic Col-0 plants, which would ordinarily be susceptible to this isolate (J. Dangel, J. McDowell, and D. Murali, *personal communication*). Experiments are currently underway to determine whether the L3/CK1 fragment is indeed a part of the *RPP8* gene. The *L3B* locus, defined by a single, Col-specific RFLP fragment, mapped above the *L3C* and *L3D* loci on chromosome 5 between RFLP markers g4715b and m247. No resistance locus has yet been reported in this region. Apart from these polymorphic hybridizing restriction fragments, at least two additional nonpolymorphic fragments were detected with the L3 probe for each of the five restriction digests used.

Finally, the C1 fragment mapped between RFLP markers g4715b and m247 on chromosome 5, which is 5 cM below the *L3B* locus. Like *L3B*, the single-copy C1 locus was specific for Col and the probe detected not even a weak hybridization signal in Ler DNA.

DNA sequence extension.

Some of the RGL fragments were used to clone flanking DNA. This serves two purposes: to resolve the complexity of the gene copy number, and to obtain sequences at the 3' end of an RGL gene, which can be used to fish EST sequences from the data bases. Most EST sequences represent the 3' ends of

Fig. 2. Deduced amino acid sequences of *Arabidopsis thaliana* resistance-gene-like (RGL) polymerase chain reaction (PCR) fragments C1 to C4 and L1 to L4, compared with similar parts of cloned disease resistance gene products *L6*, *N*, *RPS2*, and *RPM1*, and a GenBank entry (accession no. U19616) deposited as a myosin heavy chain homologue (myosin h.c.). RGL sequences are grouped in classes according to their level of similarity. Level of amino acid identity (in percent) between members of a class is indicated at end of bottom panel. Domains found to be conserved among the *L6*, *N*, and *RPS2* gene products (Lawrence et al. 1995) are underlined. Overall consensus sequence similarity is indicated below each panel as dots (similar residue in ≥ 7 sequences) or colons (identical residue in ≥ 7 sequences, or similar residue in ≥ 10 sequences). Vertical arrows indicate positions of presumed intronic sequences in the L3, L4 and C4 RGL fragments.

genes and although there are ESTs from sequenced resistance genes like *RPS2* or *RPM1* present in the data base, these are not long enough to overlap with the PCR fragments we amplified from the 5' end of RGL genes.

We were most interested in extending the DNA sequence around the L3, L4, and C4 fragments, because they were present as both clustered and dispersed multi-copy sequences. With the use of only primer L3f, originally designed to amplify DNA toward the 3' end of the L3 RGL gene in combination with another primer (see Materials and Methods), a 1.3-kb fragment was PCR amplified from both Col and Ler (Fig. 5). The Ler fragment, called L3.1, was cloned and partially sequenced. The DNA sequence overlapped 100% with the 3' end of the L3 DNA sequence, starting with the L3f primer sequence and running toward the 3' end of the RGL gene, where it ended again with the L3f primer sequence. With the DNA sequence of this L3.1 fragment an EST clone (108E20) was identified in the EST data base, which showed 60% DNA identity in a 321-bp region overlapping with the L3.1 sequence (Fig. 5). This EST clone was requested from the Arabidopsis Biological Resources Center and its genetic position was determined with the RIL population. It did not map to the single-copy *L3A* locus, but was instead found to cosegregate with the *L4/C4* locus on chromosome 1.

In a parallel experiment, another DNA fragment mapping to the *L3* loci was obtained with a combination of the RG1 primer and the STK primer designed to fit to a conserved motif encoding domain IX of protein kinases like the *PTO* gene product involved in resistance against *P. syringae* in tomato (Martin et al. 1993). With this primer set, a number of fragments were amplified that were cloned and partially sequenced. Only the partial DNA sequence of a 1.6-kb Col fragment called CK1 (Fig. 5) showed similarity with the NBS-LRR class of R-genes. In particular, the similarity was found with the region encoding the LRR of the *RPM1* gene product. Apart from the RG1 primer sequence at the 5' end there was no homology to an NBS region, and apart from the kinase primer at the 3' end there was no homology to protein kinases.

Fig. 3. Clustered appearance of the L4 and C4 resistance-gene-like (RGL) fragments in Col and Ler. **A**, Col (C) and Ler (L) DNA digested with *EcoRI* and hybridized with the C4 probe. **B**, Same blot as in (A) hybridized with the L4 probe. **C**, DNA of BAC clones F24J19 and F21J10 (1 and 2) digested with *EcoRI* and hybridized with the L4 probe. **D**, Same blot as in (C) hybridized with the EST 108E20 probe. Hybridization to Col DNA (Col) is shown as a comparison.

Upon DNA blot hybridization, this fragment detected many RFLPs between Col and Ler. Surprisingly, all RFLPs cosegregated with the four *L3* loci. The RFLP patterns revealed with the CK1 and L3 probes had many bands in common, suggesting that the CK1 and L3 fragments detected different parts of the same genes. The DNA copy number per locus detected by the CK1 probe was similar as detected by the L3 probe, except for the *L3C* locus, which contained one additional copy in both Col and Ler. There was one nonpolymorphic DNA fragment detected with the CK1 probe that could not be assigned to a genetic locus. Based on the intensity of the hybridizing RFLP-fragments, the CK1 DNA fragment was derived from the *L3B* locus on chromosome 5 of Col (Fig. 4).

The sequence of the CK1 fragment overlapped with the 3' sequence of the L3.1 fragment, sharing about 60% DNA identity. With the CK1 partial DNA sequence, three EST clones (177N18, 46E10, and 221P17) were identified in the EST data base, which showed between 60 and 80% DNA sequence identity to the 3' end of the CK1 DNA sequence. Clone 177N18 cosegregated with the *L4/C4* locus on chromosome 1, but it did not detect the same restriction fragments detected by the previously identified *L4/C4* locus-specific EST clone 108E20. In addition, the partial DNA sequences of 108E20 and 177N18 showed only 75% DNA identity. The EST clones 46E10 and 221P17 cosegregated with the *L3C* locus on chromosome 5. These ESTs are very similar to each other, sharing about 94% sequence identity in the 86-bp overlap between the two sequences.

Isolation of an RGL gene cluster.

We screened a Col bacterial artificial chromosome (BAC) library with the L4 probe to extend the DNA sequence of the RGL gene cluster at the *L4/C4* locus. Two overlapping clones were isolated (F24J19 and F21J10), on which all copies of this RGL cluster were found, including cDNA fragment 108E20 (Fig. 3). The inserts of the BAC clones constituted approximately 120 kb of Col genomic DNA. On BAC F24J19, five copies of an RGL gene were found by restriction fragment analysis, including the gene from which EST108E20 is derived, demonstrating that at least one of these RGL genes is transcriptionally active. The gene from which EST clone 177N18 was derived was not present on the two isolated BAC clones, despite the lack of recombinants with either the *L4/C4* or 108E20 fragments.

DISCUSSION

Eight different bona-fide RGL PCR fragments have been obtained from *A. thaliana* accessions Col and Ler, using degenerate PCR primers based on the NBS-LRR type of previously cloned disease resistance genes (Staskawicz et al. 1995). These fragments mapped to eight different loci on the *A. thaliana* genome. Six of these loci are closely linked to previously characterized disease resistance loci. Two fragments are actually derived from a functional R-gene or from a member of an R-gene family. The DNA sequences of all RGL fragments showed similarity to the NBS-LRR type of R-genes, which demonstrates that the primers we used can amplify a range of RGL sequences. These primers were designed with a combination of inosines or multiple nucleotides at the third codon position, by which it is theoretically possible to amplify

all genes coding for the consensus amino acid sequences shown in Figure 1. However, not all NBS-LRR type of R-genes present in Col or Ler were represented in the cloned RGL fragments. For instance, we did not clone RGL fragments belonging to the *RPS2* or *RPM1* genes. An examination of the sequence of these R-genes at the RG primer sites showed that especially the RG2 primer did not sufficiently match with the template (for both *RPM1* and *RPS2*, 23% mismatch). Slight alterations at the 3' end of the primers, selecting for different amino acid variants, will probably lead to the amplification of many other RGL fragments.

With the RG set of primers, we have found a new type of RGL sequences represented by the L4, C4, and L3 fragments. These are about 300 bp longer than what is expected based on the sequence of the R-genes used to design the RG1 and RG2 primers. Comparison of the DNA sequences of these RGL fragments showed the presence of unique sequences in addition to homologous sequences, which are flanked by *A. thaliana* consensus sequences for exon-intron splice junctions (Brown 1996). This suggests the presence of introns in the NBS region of the corresponding RGL genes, which is supported by the presence of a continuous open reading frame after artificial splicing, and the continuous alignment with the derived amino acid sequences of other RGL fragments and of cloned gene products. For soybean and potato, similar ex-

periments also yielded longer RGL fragments than expected (Leister et al. 1996; Yu et al. 1996). However, the potato fragments were not reported to contain intronic sequences, while the soybean fragments were not analyzed.

The next step following the isolation of RGL sequences was to establish a correlation between an RGL gene sequence and disease resistance. Accessions Col and Ler have been used for the RGL sequence PCRs, so the cloned fragments can only be part of resistance genes known to be active in these two accessions. Kunkel (1996) and Holub (1997) previously summarized the map positions of many disease resistance loci determined so far. Not all disease resistance loci have been characterized as yet, and only a portion of the R-genes residing at these loci will belong to the NBS-LRR gene class. Nevertheless, genetic linkage between an RGL fragment and a disease resistance locus has been found for nearly all isolated RGL fragments. The best correlation between RGL fragment and R-gene was found for the C3 fragment, which turned out to be part of the *RPS5* gene against *Pseudomonas syringae* (R. W. Innes and R. F. Warren, *personal communication*).

For the other RGL fragments, the presence of an R-gene locus and a cosegregating RGL locus is often accession-dependent. For instance, the *L4/C4* locus and the linked *RPP7* locus on chromosome 1 have been found in both Col and Ler (Crute et al. 1994), as well as the closely linked *L3A* and

Fig. 4. Positions of eight resistance-gene-like (RGL) loci and four associated EST fragments on the genetic map of *Arabidopsis thaliana* in relation to previously mapped disease resistance loci involved in pathogen recognition. Genetic map is made with a core data set of restriction fragment length polymorphism markers for the Col×Ler population of recombinant inbred lines (Lister and Dean 1993). Relative positions of markers are indicated in centiMorgans on the left of the chromosomes. Loci that could not be genetically separated due to the absence of recombinant plants in the population are given one map position. RGL loci are underlined. The L3 fragment maps to four loci (A to D) of which the *L3A* locus harbors the genomic copy of the L3 fragment. Approximate positions of disease resistance loci mapped to chromosomes 1, 4, and 5 are as shown by Kunkel (1996). *RPS* = resistance against *Pseudomonas syringae*, *RAC* = resistance against *Albugo candida*, *RPP* = resistance against *Peronospora parasitica*, *RPW* = resistance against *Erysiphe cichoracearum* or *E. cruciferarum*, *HRT* = recognition of turnip crinkle virus.

EST177N18 loci. Also the *L2/C2* locus and the linked *RPP4* locus on chromosome 4 are present in both Col and Ler (Tör et al. 1994). Of the three segregating *L3* loci on chromosome 5, the *L3C* locus has an extra Ler-specific gene copy, which correlates with the linked *RPP8* locus from Ler (Holub et al. 1994) and the Ler-specific *P. parasitica* recognitions against isolates Waco8, Waco11, and Boma1. This region of chromosome 5 contains more disease resistance loci, which are not specifically found in Ler, such as the *RPS4* locus against *Pseudomonas syringae* pv. *pisi* (in accession Ws-3) (Hinsch and Staskawicz 1996), the *HRT* locus against turnip crinkle virus (not in Col or Ler) (Dempsey et al. 1997), the *TTR1* locus against tobacco ringspot virus (in Col) (Lee et al. 1996), and the *RAC2* locus against *Albugo candida* (in Ksk-2) (H. Borhan and J. Beynon, unpublished).

With a correlation established between the chromosomal positions of R-gene loci and RGL loci, supportive evidence that RGL fragments are part of resistance genes was found in their genetic organization. Often, R-genes have been found as multi-copy, clustered sequences such as the *PTO*, *FEN*, and *PRF* complex or the *Cf-9/Cf-4* and *Cf-2/Cf-5* clusters in tomato (Dixon et al. 1996; Jones et al. 1994), or as dispersed multi-copy loci such as the *L⁶* and *M* loci in flax (Lawrence et al. 1995). In contrast, the LRR receptor protein kinase genes such as *ERECTA*, *RLK1*, *RLK4*, *RLK5*, and *PR5K* are single- or low-copy genes (Torii et al. 1996; Walker 1993; Wang et al. 1996). Nearly all RGL fragments detected several hybridizing bands upon DNA blot hybridization and they often revealed an RFLP between accessions Col and Ler. An exception to the

rule is the *C1* fragment, which was present as a single-copy sequence in Col and not at all in Ler. A similar case was observed for the *RPM1* gene (Grant et al. 1995), which is present as a single-copy gene in resistant and absent in susceptible accessions.

Especially intriguing is the genome organization of the gene family to which fragments *L3*, *C4*, and *L4* belong. This family accommodates clustered and dispersed family members that appear to be correlated with the presence of *RPP* loci on two chromosomes. The DNA sequences and DNA blot hybridization patterns of the *L3*, *L4*, *C4*, *L3.1*, and *CK1* clones and the EST clones with RGL DNA sequence homology showed that all these fragments are part of genes coding for both NBS and LRR domains. The PCR fragments *L3.1* and *CK1* were produced after annealing of the *L3*, *RG1*, and *STK* PCR primers at unexpected positions. These annealing sites for the *L3* and *RG1* primers may be remnants of previous gene-shuffling events that are likely to have arisen during the generation of these multi-copy RGL genes.

Some of these genes are expressed, as demonstrated by the presence of the cosegregating EST sequences. For EST 108E20, this cosegregation was confirmed by partial DNA sequence analysis of BAC clone F24J19 containing part of the gene cluster at the *L4/C4* locus. Further analysis of the RGL copies present on this BAC clone may reveal clues about the organization and origin of this gene cluster and of the expression pattern of the individual genes in relation to disease resistance.

In conclusion, the analysis of RGL sequences from the well-studied plant species *A. thaliana* may provide an insight into the evolution of gene families coding for NBS-LRR type of proteins toward a function in disease resistance, and offers starting points for direct cloning of these disease resistance genes.

MATERIALS AND METHODS

Primer design, PCR amplification, and DNA sequencing.

Primers *RG1* (GGI-ATG-GGI-GGI-GTI-GGI-AAR-ACN-ACN) and *RG2* (ICC-IAG-IAC-YTT-IAR-IGC-IAR-IGG-IAR-WCC; R = A or G; Y = T or C; W = A or T) (Isogen Bioscience BV, Maarssen, NL) were designed based on the P-loop and another conserved domain in the NBS of the otherwise only structurally similar resistance genes *RPS2*, *N*, and *L⁶* (Fig. 1) (Bent et al. 1994; Mindrinos et al. 1994; Whitham et al. 1994; Lawrence et al. 1995). Degenerate primers were designed with inosine or more than one residue at the third codon position in order to fit the consensus amino acid sequence as shown in Figure 1. For PCR, 50 ng of genomic DNA from either Landsberg *erecta*-1 (Ler) or Columbia-5 (Col) was used in a 50- μ l reaction containing 240 ng of each primer. The PCR started with a hot start (3 min at 94°C) before *Taq* DNA polymerase was added, then followed by 30 cycles of 1 min at 94°C, 1 min at 50°C, and 2 min at 72°C. The reaction was terminated by a 5-min extension at 72°C. Two additional test PCRs were performed with annealing temperatures at 45 and 55°C, but only the products obtained at 50°C were cloned in pGEM-T (Promega, Madison, WI). Individual clones were distinguished by insert size determined by *SacI/SacII* or *EcoRI/XhoI* restriction analysis. The DNA sequences of different inserts were determined by automated DNA sequencing.

Fig. 5. Schematic drawing showing the correlation of cloned genomic DNA fragments and corresponding EST sequences belonging to the *L3/L4/C4* class of RGL genes, relative to a scheme of a nucleotide binding site-leucine-rich repeat type of resistance gene (R-gene). Relative positions of cloned R-gene-like (RGL) polymerase chain reaction fragments and EST fragments corresponding to genomic copies at several RGL loci are shown. For each fragment, the accession from which it is amplified (L = Landsberg *erecta*, C = Columbia) and the genetic locus at which it is mapped are indicated. Primers used to obtain the RGL fragments are shown below the genomic DNA. A primer name between single quotation marks means this is an unexpected annealing position for the indicated primer.

For the L3.1 fragment, only the L3f primer was used (TGC-GGA-TCC-AAC-ATG-TTT-TGC-C). For the CK1 fragment, the RG1 primer was used in combination with degenerate primer STK (CAA-CWC-CGA-AWG-ART-AAA-CAT-C; R = A or G; W = A or T) designed on the consensus sequence of conserved domain IX of serine/threonine protein kinases as in PTO (Martin et al. 1993). The conditions for both PCRs were as described above, with annealing temperature set at 50°C.

DNA sequence analysis.

DNA sequences of RGL fragments C1 to C4, L2 to L4, CK1, and L3.1 can be found in the GenBank data base under accession numbers AF039377–AF039387. RGL DNA sequences were pairwise compared and ordered in classes. Within each class, DNA sequences are at least 65% identical to each other. RGL fragment-derived amino acid sequences were obtained by taking the full-length open reading frame for the 0.5-kb fragments. For the 0.8-kb fragments, the full-length open reading frame after artificial splicing of putative intron sequences was used. Splicing was accomplished while regarding the *A. thaliana* exon-intron boundary consensus sequences determined by Brown (1996). Within the three classes of related amino acid sequences, sequences are at least 55% identical. Between classes, there is only around 25% identity. Sequence data base searches were accomplished with the BLASTN and BLASTX algorithms (Altschul et al. 1990). The EST clones identified in the data base searches and kindly obtained from the Arabidopsis Biological Resource Center (Columbus, OH) have accession numbers T22954 (108E20), H36320 (177N18), N64938 (221P17), and T14073 (46E10).

Genetic mapping.

To determine the presence of RFLPs between Col and Ler, blots containing DNA from both accessions, digested with *Bgl*III, *Dra*I, *Eco*RI, *Eco*RV, or *Hind*III, respectively, were hybridized with a ³²P-labeled RGL fragment probe and subsequently washed at 65°C with 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and 1% sodium dodecyl sulfate. A Col × Ler population of 95 RILs (Lister and Dean 1993), obtained through the Nottingham Arabidopsis Stock Centre, was used to collect RFLP segregation data. For each probe the enzyme giving most RFLP bands was used for mapping. RFLP bands were treated as dominant alleles, i.e., RILs were scored on the presence or absence of an RFLP-band. *Dra*I-digested RIL DNA was used for C1 hybridization. *Hind*III was used for C2 and L2. *Eco*RI, *Eco*RV, and *Hind*III were used for C3. *Eco*RI was used for L3. *Dra*I was used for L4 and C4. *Hind*III and *Dra*I were used for CK1. The genetic map positions of the cloned PCR fragments were determined relative to the original core RFLP dataset (Lister and Dean 1993) with the JoinMap mapping program (Stam 1993). Markers that cosegregated, meaning without recombinants between the markers, were placed at one position on the map in an arbitrary order. More information about the RFLP markers can be found on-line at Weeds World Volume 4(ii).

Segregation tests of the resistance loci acting against the *Peronospora parasitica* isolates Waco8, Waco11, and Boma1 (collected in Wageningen and Boxmeer, The Netherlands) were performed in duplicate on the same population of RILs. Upon inoculation with each isolate, Ler showed a necrotic fleck response. Col showed early heavy sporulation in re-

sponse to Boma1 and Waco11 and delayed medium sporulation in response to Waco8 (as defined by Holub et al. 1994). Propagation and inoculation of *P. parasitica* were as previously described (Dangl et al. 1992).

BAC library screening.

A colony blot filter containing the IGF-BAC library made from partially *Eco*RI-digested genomic DNA from Col-0 (T. Mozo, Max-Planck-Institut f. Molekulare Genetik, Berlin) was obtained through the ICRF-Reference Library-Database (MPIfMG, Berlin). This library was screened with the L4 fragment probe. Positive colonies were requested from the Reference Library and grown, and their BAC insert checked on DNA blots after *Eco*RI digestion with the same L4 probe. As a reference, *Eco*RI-digested Col-0 DNA was included in the blot. Overlapping BAC clones F24J19 and F21J10, which were identified as containing all L4-hybridizing DNA fragments, were selected.

ACKNOWLEDGMENTS

We want to thank Petra Wolters and Hans Sandbrink for critically reading the manuscript, and René Klein Lankhorst for providing us with the RG1 and RG2 primers. M. A. was financially supported by the Netherlands Technology Foundation; E. H. and J. B. are financially supported by the UK Biotechnology and Biological Sciences Research Council.

LITERATURE CITED

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Bent, A. F., Kunkel, B. N., Dahlbeck, D., Brown, K. L., Schmidt, R., Giraudat, J., Leung, J., and Staskawicz, B. J. 1994. *RPS2* of *Arabidopsis thaliana*: A leucine-rich repeat class of plant disease resistance genes. *Science* 265:1856-1860.
- Brown, J. W. S. 1996. *Arabidopsis* intron mutations and pre-mRNA splicing. *Plant J.* 10:771-780.
- Crute, I., Beynon, J., Dangl, J., Holub, E., Mauch-Mani, B., Slusarenko, A., Staskawicz, B., and Ausubel, F. 1994. Microbial pathogenesis of *Arabidopsis*. Pages 705-747 in: *Arabidopsis*. E. M. Meyerowitz and C. R. Somerville, eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Dangl, J. L., Holub, E. B., Debener, T., Lehnackers, H., Ritter, C., and Crute, I. 1992. Genetic definition of loci involved in *Arabidopsis*-pathogen interactions. Pages 393-418 in: *Methods in Arabidopsis Research*. C. Koncz, N. H. Chua, and J. Schell, eds. World Scientific Publishing, Singapore.
- Dempsey, D. A., Pathirana, M. S., Wobbe, K. K., and Klessig, D. F. 1997. Identification of an *Arabidopsis* locus required for resistance to turnip crinkle virus. *Plant J.* 11:301-311.
- Dixon, M. S., Jones, D. A., Keddie, J. S., Thomas, C. M., Harrison, K., and Jones, J. D. G. 1996. The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* 84:451-459.
- Grant, M. R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R. W., and Dangl, J. L. 1995. Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. *Science* 269: 843-846.
- Hinsch, M., and Staskawicz, B. 1996. Identification of a new *Arabidopsis* disease resistance locus, *RPS4*, and cloning of the corresponding avirulence gene, *avrRps4*, from *Pseudomonas syringae* pv. *persi*. *Mol. Plant-Microbe Interact.* 9:55-61.
- Holub, E. B. 1997. Organization of resistance genes in *Arabidopsis*. Pages 5-26 in: *The Gene-for-Gene Relationship in Plant-Parasite Interaction*. I. R. Crute, E. B. Holub, and J. J. Burdon, eds. CAB Int., Wallingford, UK.
- Holub, E. B., Beynon, J. L., and Crute, I. R. 1994. Phenotypic and genotypic characterization of interactions between isolates of *Perono-*

- spora parasitica* and accessions of *Arabidopsis thaliana*. Mol. Plant-Microbe Interact. 7:223-239.
- Jones, D. A., Thomas, C. M., Hammond-Kosack, K. E., Balint-Kurti, P. J., and Jones, J. D. G. 1994. Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266: 789-793.
- Kanazin, V., Marek, L. F., and Shoemaker, R. C. 1996. Resistance gene analogs are conserved and clustered in soybean. Proc. Natl. Acad. Sci. USA 93:11746-11750.
- Kunkel, B. N. 1996. A useful weed put to work: Genetic analysis of disease resistance in *Arabidopsis thaliana*. Trends Genet. 12:63-69.
- Lawrence, G. J., Finnegan, E. J., Ayliffe, M. A., and Ellis, J. G. 1995. The *L6* gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. Plant Cell 7:1195-1206.
- Lee, J.-M., Hartman, G. L., Domier, L. L., and Bent, A. F. 1996. Identification and map location of *TTR1*, a single locus in *Arabidopsis thaliana* that confers tolerance to tobacco ringspot nepovirus. Mol. Plant-Microbe Interact. 9:729-735.
- Leister, D., Ballvora, A., Salamini, F., and Gebhardt, C. 1996. A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nature Genet. 14:421-429.
- Lister, C., and Dean, C. 1993. Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J. 4: 745-750.
- Martin, G. B., Brommonschenkel, S. H., Chunwongse, J., Frary, A., Ganal, M. W., Spivey, R., Wu, T., Earle, E. D., and Tanksley, S. D. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432-1436.
- Mindrinis, M., Katagiri, F., Yu, G.-L., and Ausubel, F. M. 1994. The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell 78:1089-1099.
- Parker, J. E., Coleman, M. J., Szabò, V., Frost, L. N., Schmidt, R., van der Biezen, E. A., Moores, T., Dean, C., Daniels, M. J., and Jones, J. D. G. 1997. The *Arabidopsis* downy mildew resistance gene *RPP5* shares similarity to the Toll and Interleukin-1 receptors with *N* and *L6*. Plant Cell 9:879-894.
- Parker, J. E., Szabò, V., Staskawicz, B. J., Lister, C., Dean, C., Daniels, M. J., and Jones, J. D. G. 1993. Phenotypic characterization and molecular mapping of the *Arabidopsis thaliana* locus *RPP5*, determining disease resistance to *Peronospora parasitica*. Plant J. 4:821-831.
- Rommens, C. M. T., Salmeron, J. M., Oldroyd, G. E. D., and Staskawicz, B. J. 1995. Intergeneric transfer and functional expression of the tomato disease resistance gene *Pto*. Plant Cell 7:1537-1544.
- Salmeron, J. M., Oldroyd, G. E. D., Rommens, C. M. T., Scofield, S. R., Kim, H.-S., Lavelle, D. T., Dahlbeck, D., and Staskawicz, B. J. 1996. Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell 86:123-133.
- Simonich, M. T., and Innes, R. W. 1995. A disease resistance gene in *Arabidopsis* with specificity for the *avrPph3* gene of *Pseudomonas syringae* pv. *phaseolicola*. Mol. Plant-Microbe Interact. 8:637-640.
- Stam, P. 1993. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J. 3:739-744.
- Staskawicz, B. J., Ausubel, F. M., Baker, B. J., Ellis, J. G., and Jones, J. D. G. 1995. Molecular genetics of plant disease resistance. Science 268:661-667.
- Thilmony, R. L., Chen, Z., Bressan, R. A., and Martin, G. B. 1995. Expression of the tomato *Pto* gene in tobacco enhances resistance to *Pseudomonas syringae* pv. *tabaci* expressing *avrPto*. Plant Cell 7: 1529-1536.
- Tör, M., Holub, E. B., Brose, E., Musker, R., Gunn, N., Can, C., Crute, I. R., and Beynon, J. L. 1994. Map positions of three loci in *Arabidopsis thaliana* associated with isolate-specific recognition of *Peronospora parasitica* (downy mildew). Mol. Plant-Microbe Interact. 7:214-222.
- Torii, K. U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R. F., and Komeda, Y. 1996. The *Arabidopsis* *ERECTA* gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. Plant Cell 8:735-746.
- Walker, J. C. 1993. Receptor-like protein kinase genes of *Arabidopsis thaliana*. Plant J. 3:451-456.
- Wang, X., Zafian, P., Choudhary, M., and Lawton, M. 1996. The PR5K receptor protein kinase from *Arabidopsis thaliana* is structurally related to a family of plant defense proteins. Proc. Natl. Acad. Sci. USA 93:2598-2602.
- Whitham, S., Dinesh-Kumar, S. P., Choi, D., Hehl, R., Corr, C., and 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.
- Yu, Y. G., Buss, G. R., and Maroof, M. A. S. 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.