The Plant Cell, Vol. 10, 1861–1874, November 1998, www.plantcell.org © 1998 American Society of Plant Physiologists

Intragenic Recombination and Diversifying Selection Contribute to the Evolution of Downy Mildew Resistance at the *RPP8* Locus of Arabidopsis

John M. McDowell,[a,1] Murali Dhandaydham,[a,1] Terri A. Long,[a] Mark G. M. Aarts,[b] Stephen Goff,[c] Eric B. Holub,[d] and Jeffery L. Dangl[a,e,2]

[a] Department of Biology, C.B. 3280 Coker Hall, University of North Carolina, Chapel Hill, North Carolina 27599-3280
[b] Department of Molecular Biology, DLO Centre for Plant Breeding and Reproduction Research, Postbus 16, 6700AA, Wageningen, The Netherlands
[c] Biotechnology and Genomics Center, Novartis Crop Protection, Inc., Research Triangle Park, North Carolina 27709-2257
[d] Director's Research Group, Horticulture Research International, Wellesbourne, Warwickshire, CV35 9EF, United Kingdom
[e] Curriculum in Genetics and Molecular Biology, C.B. 3280 Coker Hall, University of North Carolina, Chapel Hill, North Carolina 27599-3280

Pathogen resistance (*R*) genes of the NBS-LRR class (for nucleotide binding site and leucine-rich repeat) are found in many plant species and confer resistance to a diverse spectrum of pathogens. Little is known about the mechanisms that drive NBS-LRR gene evolution in the host–pathogen arms race. We cloned the *RPP8* gene (for resistance to *Peronospora parasitica*) and compared the structure of alleles at this locus in resistant Landsberg *erecta* (L*er*-0) and susceptible Columbia (Col-0) accessions. *RPP8-L*er encodes an NBS-LRR protein with a putative N-terminal leucine zipper and is more closely related to previously cloned *R* genes that confer resistance to bacterial pathogens than it is to other known *RPP* genes. The *RPP8* haplotype in L*er*-0 contains the functional *RPP8-L*er gene and a nonfunctional homolog, *RPH8A.* In contrast, the *rpp8* locus in Col-0 contains a single chimeric gene, which was likely derived from unequal crossing over between *RPP8-L*er and *RPH8A* ancestors within a L*er*-like haplotype. Sequence divergence among *RPP8* family members has been accelerated by positive selection on the putative ligand binding region in the LRRs. These observations indicate that NBS-LRR molecular evolution is driven by the same mechanisms that promote rapid sequence diversification among other genes involved in non-self-recognition.

INTRODUCTION

A broad range of microorganisms have evolved the ability to use plants as a nutritional resource, and plants in turn have evolved multiple lines of defense against pathogen invasion (Hammond-Kosack and Jones, 1996a). Inducible defenses are mediated through gene-for-gene systems in which the plant carrying a particular resistance (*R*) gene allele responds to pathogens carrying a matching avirulence (*avr*) gene (Flor, 1971). Most plants contain large collections of highly specific *R* genes, which are thought to encode specialized receptors that recognize *avr* gene–dependent elicitors (Keen, 1990). If the *R* gene or the corresponding *avr* gene is not functional, then recognition does not occur, defenses are not activated, and the plant is susceptible to infection. Thus, pathogens can circumvent gene-for-gene resistance by alteration or loss of *avr* genes. This places the host under selective pressure to evolve new recognition capabilities. *avr* gene mutations and deletions occur at high frequency in nature (van Kan et al., 1991; Rohe et al., 1995; Sweigard et al., 1995; Joosten et al., 1997), but the host's response in this evolutionary arms race is not well understood.

Two themes have emerged from recent molecular characterization of *R* genes. *R* genes are often members of tightly linked multigene families, which can be functionally diversified (Hammond-Kosack and Jones, 1996b). A second, somewhat unexpected generality is that all *R* genes characterized to date, with one exception (Martin et al., 1993), encode proteins with long stretches of leucine-rich repeats (LRRs) (Jones and Jones, 1996). LRRs are present in a wide variety of proteins and participate in protein–protein interactions and ligand binding (Kobe and Deisenhofer, 1995). Crystal structure analysis has demonstrated that the LRRs of a ribonuclease inhibitor form a solvent-exposed β sheet structure that binds the ribonuclease (Kobe and Deisenhofer, 1993). By analogy, LRRs in plant R proteins are thought to bind pathogen-derived signal molecules and thereby mediate

[1] These authors contributed equally to this work.
[2] To whom correspondence should be addressed. E-mail dangl@email.unc.edu; fax 919-962-1625.

recognitional specificity (Dixon et al., 1996), although direct biochemical evidence is currently lacking.

Two superfamilies of LRR-encoding pathogen *R* genes have been defined by putative functional motifs and predicted localization of the encoded proteins (Dangl, 1995). One superfamily, represented by the tomato *Cf* genes (for resistance to the fungal pathogen *Cladosporium fulvum*) (Hammond-Kosack and Jones, 1996b) and the *Xa21* gene family in rice (for resistance to the bacterial pathogen *Xanthomonas campestris* pv *oryzae*) (Song et al., 1995), encodes proteins that are predicted to be membrane bound and composed primarily of extracytoplasmic LRRs. The Cf R proteins do not contain any recognizable signaling domain, whereas Xa21 contains extracytoplasmic LRRs fused to a cytoplasmic kinase domain.

The second and larger *R* gene superfamily (referred to as NBS-LRR) encodes proteins with a predicted nucleotide binding site followed by a variable number of C-terminal LRRs (Bent, 1996). NBS-LRR proteins do not contain a recognizable signal sequence and probably function inside the cell (Leister et al., 1996). Most NBS-LRR genes fall into one of two subclasses based on their N-terminal motifs (Bent, 1996). The TIR-NBS-LRR subclass is defined by an N-terminal region that resembles the cytoplasmic signaling domain of the Toll and interleukin1 transmembrane receptors (Parker et al., 1997). This subclass includes genes that specify resistance to a virus (*N* in tobacco) (Whitham et al., 1994), fungi (*L6* and *M* in flax) (Lawrence et al., 1995; Anderson et al., 1997), and oomycetes (*RPP5*, *RPP1A*, *RPP1B*, and *RPP1C* in Arabidopsis, where *RPP* signifies resistance to *Peronospora parasitica*) (Parker et al., 1997; Botella et al., 1998). The second subclass (LZ-NBS-LRR) contains a leucine zipper–like motif in place of the TIR domain and is represented by the genes *RPM1* (Grant et al., 1995), *RPS2* (Bent et al., 1994; Mindrinos et al., 1994), and *Prf* (Salmeron et al., 1996). These genes specify resistance to *Pseudomonas syringae* pathovars.

Recent comparative analyses of extracytoplasmic LRR gene clusters have provided insight into their evolution. The *Cf-4*/*9* gene cluster contains related but functionally distinct genes that are subject to positive diversifying selection in the LRRs (Parniske et al., 1997). Sequence exchanges appear to occur between linked *Cf-4*/*9* homologs; novel *Cf-4*/*9* haplotypes, which differ in gene copy number, can be generated by unequal crossovers at homologous intergenic regions. Evidence for gene duplications, intragenic recombination, and diversifying selection also has been reported for the *Xa21* gene cluster (Song et al., 1997; Wang et al., 1998). Thus, molecular evolution of gene clusters encoding extracytoplasmic LRR–containing R proteins is driven by the same mechanisms that generate diversity in other complex loci involved in non-self-recognition, such as the major histocompatibility complex (MHC) in animals (Dangl, 1992; Parham and Ohta, 1996; Hughes and Yeager, 1997).

Although NBS-LRR genes are widespread in plants and recognize many types of pathogens, little is known about the mode of NBS-LRR gene evolution. The available NBS-LRR sequences are very divergent from each other and provide no evolutionary insight other than definition of the conserved motifs described above. The structural differences between putative extracytoplasmic LRR proteins and NBS-LRR proteins imply that these two R protein superfamilies are biochemically distinct, and it is therefore of interest to determine whether they have evolved by different mechanisms.

We have used the Arabidopsis–*P. parasitica* (downy mildew) pathosystem for comparative analysis of *R* gene evolution. *P. parasitica* is a biotrophic oomycete and a prominent natural pathogen of Arabidopsis in Europe (Koch and Slusarenko, 1990; Holub and Beynon, 1996). A large number of Arabidopsis *RPP* genes have been defined using *P. parasitica* isolates from natural Arabidopsis populations (Holub et al., 1994; Tör et al., 1994). These genes are functionally polymorphic among Arabidopsis accessions, suggesting that coevolution of host and parasite has been rapid and dynamic. Thus, comparison of allelic variants will provide insight into *R* gene evolution. The Arabidopsis–*P. parasitica* pathosystem also provides the opportunity to examine *R* gene evolution in a naturally evolving interaction, thereby avoiding potential loss of genetic diversity from bottlenecks in selective breeding of crop species as well as phylogenetic artifacts caused by forced introgression of genes from wild species.

Four *RPP* genes recently have been shown to encode members of the TIR-NBS-LRR subclass (Parker et al., 1997; Botella et al., 1998). In contrast, we show in this study that the *RPP8* gene is a member of the LZ-NBS-LRR subclass. Furthermore, sequence comparisons of resistant and susceptible *RPP8* alleles provide evidence that intragenic recombination and positive selection interact to promote sequence diversification in NBS-LRR *R* gene evolution.

RESULTS

Genetic and Physical Definition of the *RPP8* Locus

The *RPP8* gene specifies resistance to the Emco5 isolate of *P. parasitica* in the Arabidopsis accession Landsberg *erecta* (L*er*-0). Emco5 is compatible with accession Columbia-0 (Col-0). Therefore, we used the Dean and Lister Col-0 × L*er*-0 recombinant inbred (RI) lines (Lister and Dean, 1993) to map *RPP8* genetically, as shown in Figure 1. When we used 100 RI lines, resistance to Emco5 segregated as a single locus (*RPP8*) on chromosome 5 in the interval between *Dfr* and *Lfy.* We identified 71 additional *Dfr–Lfy* recombinants from an additional 198 RI lines, and we used these recombinants to narrow the interval, defining *Spl2* and *Cra1* as closer markers on either side of *RPP8. Spl2* identified one recombinant centromeric to *RPP8*, and *Cra1* identified five recombinants telomeric to *RPP8* (Figure 1A).

Yeast artificial chromosome (YAC) end probes and genetically anchored molecular markers were used to construct a physical map of the *RPP8* interval (Figure 1A). The *Spl2* and

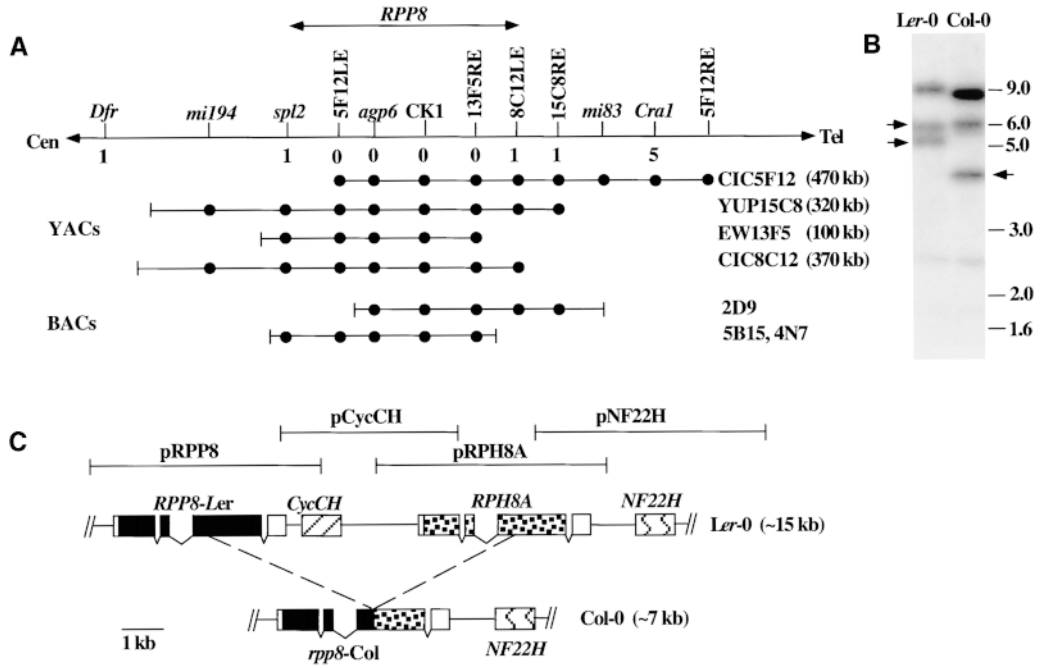

Figure 1. Genetic and Physical Map of the *RPP8* Region.

(A) Genetic map of the *RPP8* interval. Molecular markers are shown above the line, and the number of recombinants that separate each marker from *RPP8* are shown below the line. The minimum genetic interval of *RPP8* is shown between the arrowheads at top. *Dfr*, *spl2*, and *Cra1* are cleaved amplified polymorphic sequence markers. CK1 refers to the restriction fragment length polymorphism (RFLP) shown in **(B)**. The remaining markers are RFLPs derived from yeast artificial chromosome (YAC) ends. YAC and bacterial artificial chromosome (BAC) clones are depicted below the genetic map, with approximate lengths shown at right. YAC clones CIC4E12, CIC6F12, and EW6E5, which also map in the same region, are not shown. Cen, centromeric; Tel, telomeric.

(B) Gel blot of genomic DNA from Col-0 and L*er*-0 that was digested with EcoRV and probed with the CK1 candidate gene fragment at moderate stringency. The RFLP cosegregating with *RPP8* is shown by arrows. The Col-0 band that comigrates with the doublet in L*er*-0 segregated independently of the doublet. DNA length standards are shown at right in kilobases.

(C) Physical structure of the *RPP8* locus in L*er*-0 and Col-0. The L*er*-0 segment represents 15 of the 23 kb that were sequenced from the 9L9 cosmid. Genomic subclones are depicted above the physical map. *RPP8-L*er and *RPH8A* coding sequences are depicted by filled and stippled boxes, respectively, 5′ and 3′ untranslated regions are depicted by open boxes, and introns are represented by diagonal lines. The boxes labeled *CycCH* and *NF22H* represent regions of homology to rice cyclin C and a hypersensitive response–inducing gene (*NF22*) from tobacco, respectively. In Col-0, the region between *RPP8-L*er and *RPH8A* has been deleted, as depicted by the dashed lines.

Cra1 markers both mapped within the YAC contig, demonstrating that the contig spanned the *RPP8* locus. We genetically mapped four YAC ends as restriction fragment length polymorphisms (RFLPs) to refine further the *RPP8* interval. 5F12LE and 13F5RE RFLPs both cosegregated with *RPP8*, whereas 8C12LE and 15C8RE detected one recombinant telomeric to *RPP8.* The *Spl2*–8C12LE interval thus defined the smallest possible genetic interval in our mapping population. This genetic distance corresponds to a maximum physical distance of ~100 to 300 kb (Figure 1A).

Identification and Mapping of an *RPP8* Candidate Gene

A candidate for the *RPP8* gene (CK1, described by M.G.M. Aarts et al., 1998) was amplified with degenerate polymerase chain reaction (PCR) primers from conserved *R* gene motifs. CK1 hypothetically encodes an LRR sequence with ~30% identity and 40% similarity to segments of the *RPM1* gene and hybridized with a polymorphic multicopy family in both Col-0 and L*er*-0 (Figure 1B). We genetically and physically mapped an EcoRV RFLP, which consists of a double band (~5.5 to 6 kb) in L*er*-0 and a single ~4.5-kb band in Col-0 (Figure 1B). The L*er*-0 doublet cosegregated with resistance to Emco5 in the subset of RI lines that contained recombinations between *Dfr* and *Lfy.* Conversely, the 4.5-kb Col-0 band was always present in Emco5-susceptible RI lines and hybridized with all of the Col-0 YACs and bacterial artificial chromosomes (BACs) spanning *rpp8* (Figure 1A). The genetic and physical colocalization of the EcoRV RFLP with the *RPP8* phenotype, in combination with its sequence similarity to known *R* genes and potential copy

number polymorphism, implicated it as a candidate gene for *RPP8.*

Transgenic Complementation of *RPP8* Function

We isolated genomic cosmid clones containing the EcoRV doublet from L*er*-0 by using the CK1 probe. One cosmid (9L9) contains a 23-kb insert that includes both bands of the doublet. A second cosmid (25M19), which overlaps with 9L9 over ~17 kb, contains the upper band of the doublet and a fragment of the lower band (data not shown). Both cosmids were transformed into susceptible Col-0, and transgenic (T_1) seedlings were selected and allowed to set seed. T_2 progeny from multiple independent transformants were inoculated with Emco5 and assessed for resistance. Complementation experiments are summarized in Table 1. All 12 tested Col::9L9 transgenic lines segregated ~3:1 for resistant to susceptible in the T_2 generation, which is consistent with a single, dominant transgenic locus conferring Emco5 resistance. At least five of six tested Col::9L9 lines were independent transformants (data not shown). None of the seven tested Col::25M19 T_2 lines displayed resistance to Emco5 (Table 1), suggesting that the lower band of the doublet was necessary for Emco5 resistance. Neither 9L9 nor 25M19 provided resistance to the Madi1 or Noco2 isolates of *P. parasitica* (Table 1).

Only one CK1-hybridizing band was detectable in the Col-0 YACs and BACs spanning *rpp8* (data not shown), suggesting that only one Col-0 CK1 family member is present in this >470-kb interval. Furthermore, mapping of other CK1-hybridizing bands demonstrated that no other CK1 family members are closely linked to *RPP8* (described by M.G.M. Aarts et al., 1998). Cosmids containing additional CK1 family members conferred no resistance to any *P. parasitica* isolate in transgenic Col-0 (data not shown). These results suggest that resistance to Emco5 in L*er* is conferred specifically by one member of the CK1 gene family.

Table 1. Interaction Phenotypes of Col-0 Transgenic Plants and L*er*-0 *rpp8* Mutants with *P. parasitica* Isolates

Genotype	Emco5[a]	Madi1[a]	Noco2[a]
Col-0::9L9	R	S	S
Col-0::25M19	S	S	S
Col-0::pRPH8A	S		
Col-0::pRPP8	R		
Col-0::pCYCH	S		
Col-0::pNF22H	S		
rpp8-1	0.9 (0.4)[b]	R	R
rpp8-2	12.6 (1.2)[b]	R	R
rpp8-3	3.0 (0.9)[b]	R	R
rpp8-4	6.5 (1.0)[b]	R	R
rpp8-5	2.7 (0.5)[b]		
rpp8-6	2.3 (0.5)[b]		
Col-0 wild type	16.0 (0.9)[b]	S	S
L*er*-0 wild type	0.0[b]	R	R

[a] R, resistant; S, susceptible. Four to 12 transgenic lines were assayed for each construct. Consistent phenotypes were observed in each case.

[b] Quantitative disease ratings are expressed as the mean number of sporangiophores per cotyledon from 10 plants, with the standard error shown within parentheses. Quantitative assessments were conducted for three separate inoculations with similar results.

Two Closely Related Genes Are Present at the *RPP8* Locus in L*er*-0

Sequencing of the 9L9 cosmid insert revealed two highly similar NBS-LRR genes (Figure 1C). We constructed subclones to separate these two genes (Figure 1C). All of the four lines transgenic for pRPP8 were completely resistant to Emco5, whereas all of the four lines transgenic for pRPH8A were as susceptible to Emco5 as is wild-type Col-0 (Figure 2A and Table 1). Thus, a single NBS-LRR gene, referred to hereafter as *RPP8-L*er, is sufficient to provide Emco5 resistance in the Col-0 background. The second gene (named *RPH8A* for *RPP8* homolog A) is insufficient for transgenic complementation of resistance to Emco5 in Col-0.

*RPP8-L*er and *RPH8A* are separated by a 3.7-kb segment containing a putative open reading frame with 75% amino acid similarity to cyclin C from rice (Figure 1C). A fourth open reading frame ~1 kb downstream of *RPH8A* resembles (~50% amino acid similarity) the tobacco gene *NF22* (GenBank accession number U66266). *NF22* was identified by its ability to induce a hypersenstive response–like reaction when overexpressed (Karrer et al., 1998). Subclones of the *NF22* homolog or the cyclin C homolog conferred no resistance to Emco5 in transgenic Col-0 plants (Table 1).

The intron–exon structure of *RPP8-L*er was deduced by comparison to *RPP8* cDNAs and is diagrammed in Figure 1C. The *RPP8* coding sequence contains two introns: intron 1 (129 bp) splits codon 292, and intron 2 (675 bp) splits codon 341. A third intron (123 bp) begins 4 bp downstream of the stop codon in the *RPP8* cDNA. Sequence analysis of 11 independent *RPP8-L*er clones revealed variable polyadenylation sites ~450 bp downstream of the stop codon. The gene structure of *RPH8A* could not be confirmed by cDNA comparison because no *RPH8A* cDNAs were isolated, but it is probably identical because conserved intron–exon border sequences were found at identical locations in the *RPH8A* coding sequence. Interestingly, the 3′ ends of *RPP8-L*er and *RPH8A* are identical over an 898-bp stretch, from codon 837 to 688 bp downstream of the stop codon (including the intron, 3′ untranslated region, and downstream nontranscribed sequence). After this 898-bp stretch, similarity between the two genes is very low. The 5′ flanking sequences of *RPP8-L*er and *RPH8A* are almost completely dissimilar, except for a 90-bp stretch of 89% identity, which begins 473 and 692 bp upstream of the *RPP8-L*er and *RPH8A* start codons, respectively.

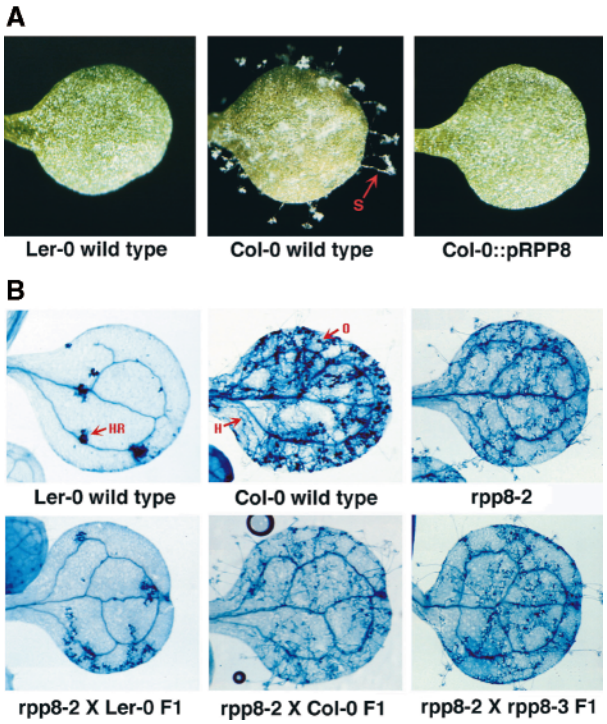

Figure 2. Interaction Phenotypes of Col-0::pRPP8 Transgenic Plants and L*er*-0 *rpp8* Mutants.

(A) *RPP8* from L*er*-0 confers resistance to Emco5 in transgenic Col-0 plants. At 7 days after inoculation with Emco5, wild-type Col-0 cotyledons support heavy asexual sporulation (S, sporangiophores), whereas no sporulation is visible on wild-type L*er*-0 or transgenic Col-0 seedlings containing the pRPP8 subclone from the 9L9 cosmid. **(B)** The *rpp8-2* mutant in L*er*-0 is susceptible to Emco5. The interaction phenotypes of cotyledons from F_1 progeny of various crosses demonstrate that *rpp8-2* is recessive to *RPP8-L*er and allelic to *rpp8*-Col and *rpp8-3.* Cotyledons were stained at 7 days after inoculation with trypan blue, which is retained by parasite structures (H, hyphae; O, oospores) and dead host cells (HR).

RPP8 Encodes a Member of the LZ-NBS-LRR Subclass

Figures 3 and 4 provide the primary structures of hypothetical proteins encoded by *RPP8-L*er (906 amino acids), *rpp8*-Col (908 amino acids), and *RPH8A* (907 amino acids). The latter two genes encode full-length hypothetical proteins that share 92 and 91% amino acid identity, respectively, with RPP8-L*er* (Figures 3 and 4, and Table 3). Several putative functional motifs present in known *R* genes are apparent in the encoded proteins (Figures 3 and 4). The C-terminal one-third of each gene is composed of 14 imperfect LRRs, which vary in length from 21 to 29 amino acids. A consensus nucleotide binding site and a hydrophobic domain conserved in all NBS-LRR genes are also apparent. Finally, a putative six-heptad leucine zipper is present near the N terminus. This motif clearly places RPP8 in a distinct structural subclass from the other RPP proteins that have been identified. RPP8-L*er* is more closely related to the RPM1 bacterial resistance protein from Arabidopsis (26% identity and 39% amino acid similarity) than it is to any other known R protein. RPP8 has no significant similarity with RPP5, RPP1A, RPP1B, or RPP1C, except in the functional domains that define the putative nucleotide binding site.

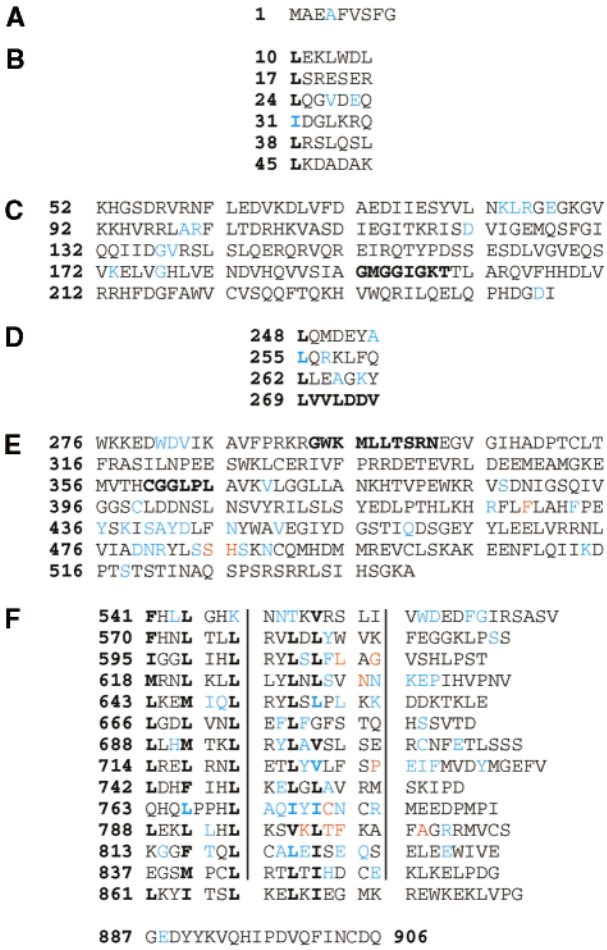

Figure 3. Deduced Amino Acid Sequence of *RPP8-L*er.

Domains A to F are based on putative functional motifs. Domains B and D contain putative leucine zippers. Domains C, D, and E contain the NBS motifs and a conserved hydrophobic domain, shown in boldface. Domain F contains 14 imperfect LRRs defined by the conserved residues shown in boldface. The LRR subdomain XXLXLXXXX, which encompasses the putative β strand/β turn region identified from the porcine ribonuclease inhibitor crystal structure, is framed between the solid lines. Blue residues represent positions in which either *RPH8A* or *rpp8*-Col encodes a different amino acid from *RPP8-L*er. Residues in red are different in all three proteins.

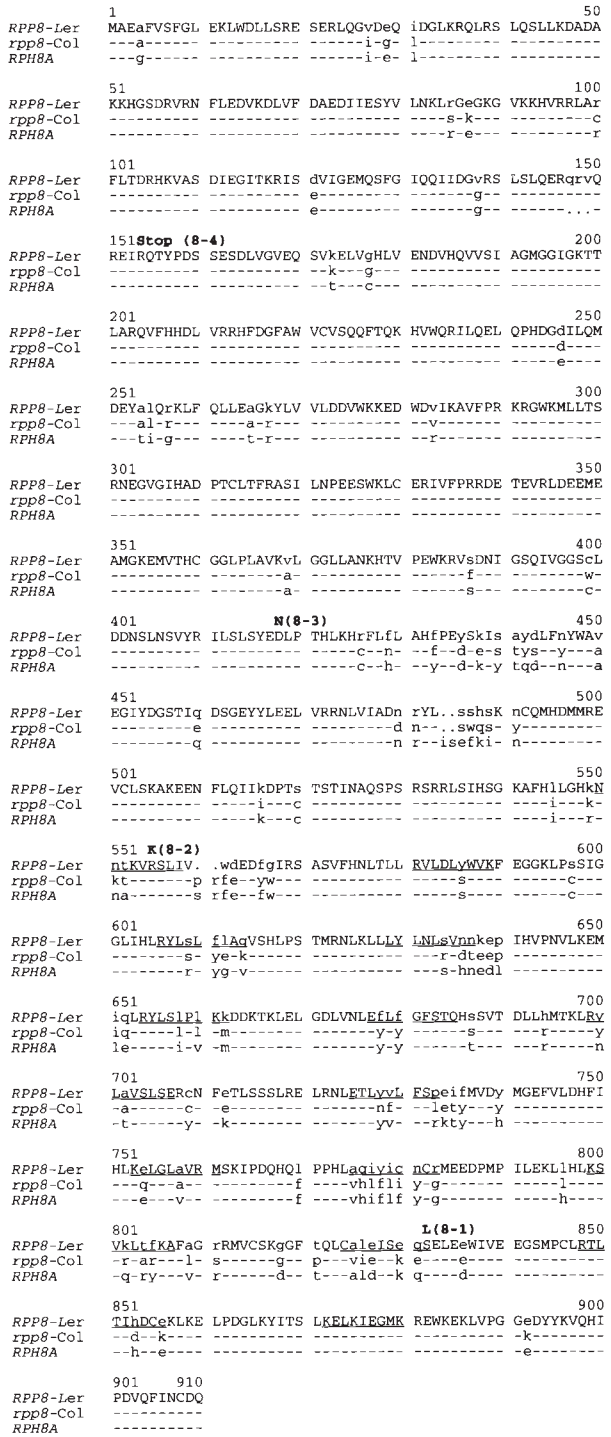

Figure 4. Amino Acid Sequence Alignment of *RPP8-L*er, *RPH8A*, and *rpp8*-Col.

Dashes represent identical amino acids, and dots represent deletions in *RPP8-L*er and *rpp8*-Col compared with *RPH8A*. Amino acid substitutions are shown as lowercase letters. Amino acid changes in L*er rpp8* mutants are shown above the *RPP8-L*er sequence in boldface. The XXLXLXXXX motifs are underlined. The corresponding nucleotide sequences have GenBank accession numbers AF089710 and AF089711 for *RPP8-L*er and *rpp8*-Col, respectively.

The *rpp8* Allele in Col-0 Is a Chimera of Progenitor Genes Related to *RPP8-L*er and *RPH8A*

As shown in Figure 1C, the structure of the *rpp8* locus in Col-0 is dramatically different from the *RPP8* locus in L*er*-0. Only one *RPP8* homolog (named *rpp8*-Col) is present at the Col-0 locus. The 5′ flanking sequence of *rpp8*-Col is almost identical to that of *RPP8*-L*er*, whereas the 3′ flanking sequence of *rpp8*-Col is almost identical to the segment extending from the end of *RPH8A* to the *NF22* homolog (Figure 1C). The segment that separates *RPP8* and *RPH8A* in L*er*-0, including the cyclin C homolog, is deleted in Col-0. *rpp8*-Col thus appears to be derived from a precise in-frame unequal crossover within an ancestral L*er*-like *RPP8* haplotype.

Seven insertion/deletion sites, shown in Figure 5, were used as landmarks to localize the most likely recombination breakpoint. *rpp8*-Col shares with *RPP8-L*er a 9-bp insertion (codons 147 to 149) and a 6-bp deletion (codons 484 to 485) relative to the *RPH8A* sequence (Figures 4 and 5). *rpp8*-Col also shares four additional indels with *RPP8-L*er in intron 2 (Figure 5). *rpp8*-Col shares a 6-bp insertion with *RPH8A*, relative to *RPP8-L*er, at codons 560 to 561. The recombination breakpoint thus appears to be located between codons 486 and 559, which includes the region just upstream of the LRRs as well as part of the first LRR (Figures 3 and 4). Interestingly, most of the indels encompass short direct repeats (Figure 5), which suggests that they could have been generated by transposon insertion and subsequent excision.

The pattern of nucleotide polymorphisms between *RPP8-L*er, *RPH8A*, and *rpp8*-Col is very complicated, as shown in Figure 6. We observed a lack of consistent sequence affiliation, based on shared nucleotide polymorphisms, between any pair of homologs. Instead, the three *RPP8* homologs exhibit a patchwork pattern of affiliations in their coding sequences. For example, the majority of polymorphisms (23 of 39) in the first 1000 bp support an affiliation between *RPP8-L*er and *rpp8*-Col, which is consistent with the hypothesis that the 5′ end of *rpp8*-Col was derived from an *RPP8-L*er–like ancestor. Similarly, the majority of 3′ polymorphisms support an affiliation between *rpp8*-Col and *RPH8A.* However, there are segments of contiguous polymorphisms that support different affiliations. For example, nucleotides 130 to 301 contain seven polymorphisms that affiliate *RPP8-L*er with *RPH8A* rather than *rpp8*-Col. This suggests that a recent exchange occurred between the two L*er*-0 genes. Alternatively, this affiliation could reflect the accumulation of contiguous point mutations in the Col-0 allele. Comparisons with other *RPP8* homologs are necessary to distinguish accurately between these possibilities.

```
Codons 147-149     RPP8-Ler   CAAAGGGTGCAGAGGGAGATCCGAb
                   rpp8-Col   CAAAGGGTGCAGAGGGAGATCCGA
                   RPH8A      CAA.........AGGGAGATCCGA

Intron 2 donor     RPP8-Ler   GGT....ACATATC
                   rpp8-Col   GGT....ACATATC
                   RPH8A      GGTacatACATATC

Intron 2-192       RPP8-Ler   GTAT................GGT
                   rpp8-Col   GTAT................GGT
                   RPH8A      GTATttacatatatattatctaa

Intron 2-336       RPP8-Ler   TAGTACAGTAAGTCAGTAATTA
                   rpp8-Col   TAGTACAGTAAGTCAGTAATTA
                   RPH8A      TAGaA.......atAGTAATTA

Intron 2-573       RPP8-Ler   CGTGATTACACGATTCCTTGTAATAAAGT
                   rpp8-Col   CGTGATTACACGATTCCTTGTAATAAAGT
                   RPH8A      CGTGATTACAC............acgAGT

Codons 484-485     RPP8-Ler   TCT......AAAAATTG
                   rpp8-Col   TCT......AAAtATTG
                   RPH8A      TtTaaaataAAAAATTG

Codons 560-561     RPP8-Ler   GTT... ...TggGAcGAAG
                   rpp8-Col   GTTcCG AGATTCGAGGAAG
                   RPH8A      GTTtCG AGATTCGAGGAAG
```

Figure 5. Location and Sequence of Insertion/Deletion Sites in *RPP8-L*er, *rpp8*-Col, and *RPH8A.*

Codon positions are numbered according to the multiple alignment in Figure 4. Intron positions are numbered from the first nucleotide of intron 2 in *RPP8-L*er. Direct repeats are emphasized by italics. Nucleotide substitutions are indicated by lowercase letters. Dots indicate gaps. The intron 2 splice donor site is underlined.

Several of the LRR-encoding segments are extremely divergent among the three genes (Figures 4 and 6). The degree of divergence among the LRRs is variable, with LRRs 11 and 12 exhibiting the highest divergence and LRRs 3, 10, 13, and 14 exhibiting the highest degree of conservation. Perhaps the divergent LRRs are directly involved in recognition specificity, whereas the conserved LRRs play a structural role. Two highly variable regions also are apparent outside the LRRs (amino acid residues 432 to 442 and 480 to 489). They do not fall within any recognizable functional motif. We predict that these regions define a new set of functionally relevant residues.

Analysis of *rpp8* Mutants in L*er*-0

We isolated six independent *rpp8* mutants from a screen of mutagenized L*er*-0 plants. Disease ratings of all mutants are shown in Table 1, and the phenotype of the *rpp8-2* mutant is presented in Figure 2B. The six mutants supported different levels of Emco5 sporulation, suggesting that they represent a series of alleles with mutations of differing severity. Although the levels of Emco5 growth varied somewhat between experiments, *rpp8-2* was generally the most susceptible mutant, whereas *rpp8-1* exhibited the weakest susceptibility phenotype. Typically, only 20 to 40% of *rpp8-1* cotyledons produced sporangiophores. Each mutant retained wild-type levels of resistance to the Madi1 (resistance provided by *RPP21*) and Noco2 (*RPP5*) isolates. None of the mutants exhibited any obvious developmental phenotype.

Results from genetic analysis of the *rpp8* mutants are summarized in Table 2. F_1 progeny from backcrosses of all six mutants to L*er*-0 were resistant to Emco5 (Table 2 and Figure 2B), demonstrating that the mutations were recessive. Trypan blue staining of backcross F_1 plants revealed occasional trails of necrotic host cells in the cotyledons (Figure 2B), suggestive of a slightly delayed defense response. This phenomenon was also observed in F_1 progeny from a cross of wild-type Col-0 × L*er*-0 (data not shown), suggesting that

```
                                                            11111111111111111111111111111111111111111 11111111
                  1122233334444444444445555666777777777777788889911111222333333333333344444444444444445555 66666666
           12678938559026712444444445123247233466668899044906036978901122222234471344555566666666781345 34557888
           1779610059861353123456789089172295810359392307813667076686290146786798082314680234567158679 36349012
cons       CacAACccAGCaaAaGcAGGGTGCAGAGTcccTGTtGCTCaaGGGGTgacCaCCTATGACTATTGAAtCCtAGGGCCAC------AccaAG AATACGAG
RPP8-Ler   **aG*A*******TgT************A***************A*****T***CT*T***G*****aT********C*********g**C T***----
rpp8-Col   ****G*ttTATgg***************************************tTG****G*****AGT**GaGAC*G*********Tt**T*  **G*****
RPH8A      Gg**************t---------CT*ttgCTGcAACGggAA*CGagt******A**AC*CG**********T*G*TAAAATA***c**  *G*G****

           11111 11 111111 11111111 1111 222 22222 22222222 22 2222222222 22222222222222 22222222222 22 2
           66666 77 888888 89999999 9999 000 00111 11112222 22 3333333333 34444444444444 44444444445 55 6
           88899 59 233334 70111115 5778 336 89023 57880001 67 1222333344 80000111222233 56777889990 56 7
           56979 70 723460 30027891 4285 390 18464 75254578 03 0489358917 74569013156703 31456181698 76 4
cons       TCGTT CG TACGGG tAAGACCA CCCT AAG GTGGG taTGCCAT GC CTCATCTTTG TCAAaACTtGTTqG GACTCAACagG CG G
RPP8-Ler   GGC*G AC *TTT** ***A**** ***A TT* A**** ****T*T* ** ACA*AA*GAC *A********C*a* ***C**G*g** ** *
rpp8-Col   ***A* ** *****A cCC***** **** *** ***** cgAT*A** C* ***T**A*** **GT*G*CaC*A*C *CT*A**G*** GA A
RPH8A      ***** ** G***A* ****TTTC GAG* **C *AAAA *******C *T ********** A***g*G******* A****C***aT ** *
```

Figure 6. Patchwork Distribution of Nucleotide Polymorphisms between *RPP8-L*er, *rpp8*-Col, and *RPH8A* Coding Sequences.

Polymorphic sites that distinguish the three genes are shown. Nucleotide positions, beginning from the start codon, are shown above the lines. Nucleotide positions for which all three genes are identical or for which each gene has a different nucleotide are omitted. A consensus (cons) sequence (two of three) is shown between the lines. Residues that conform to the consensus are represented by stars, and silent nucleotide substitutions are shown in lowercase letters. Gaps are represented by dashes. Colors indicate sequence affiliations based on shared polymorphisms. The 14 LRRs are separated by spaces, whereas nucleotides encoding amino acids in the XXLXLXXXX motif are shown by italics in the consensus sequence.

*RPP8-L*er is not completely dominant with respect to *rpp8*-Col. F_2 progeny from the backcrosses of *rpp8-2* and *rpp8-3* to L*er*-0 segregated ~3 resistant:1 susceptible, which is consistent with a single recessive mutation. F_2 progeny from the *rpp8-1* × L*er*-0 backcross did not segregate any individuals that supported sporulation. This most likely reflects the very weak effect of the *rpp8-1* mutation, as suggested by the weak and inconsistent Emco5 growth in the *rpp8-1* M3 seedlings described above.

Outcrosses of all of the six mutants to wild-type Col-0 as well as three intermutant crosses yielded susceptible F_1 progeny (Figure 2B and Table 2). Because *RPP8* is the only locus for Emco5 resistance that segregates between Col-0 and L*er*-0, the observed lack of complementation in F_1 progeny of these crosses strongly suggests that all seven mutations are in *RPP8.* F_2 segregation ratios from three tested outcrosses to Col-0 were consistent with this hypothesis. A significant proportion of F_2 progeny from the *rpp8-1* × Col-0 cross did not support sporulation, most likely because of the weak effect of the *rpp8-1* mutation. F_2 progeny from the intermutant crosses also segregated for disease-free individuals. This could reflect the additive effect of two partially functional mutations. Chi-square analysis (Table 2) strongly contradicts the hypothesis that the mutations are in unlinked second site loci (predicted 9 resistant:7 susceptible segregation in outcross and intermutant F_2 populations).

For further confirmation that these mutations are in the *RPP8* gene, we compared the *rpp8* coding sequence from four mutants with the wild-type *RPP8-L*er sequence. In *rpp8-1*, a C-to-T mutation in codon 827 caused an S-to-L substitution in LRR12 (Figure 4). In *rpp8-2*, a G-to-A mutation in codon 553 caused an R-to-K substitution in LRR1. In *rpp8-3*, a G-to-A mutation in codon 418 caused a D-to-N substitution. In *rpp8-4*, a C-to-T mutation in codon 151 created a stop codon. These sequence alterations confirm that the *R* gene candidate is indeed *RPP8.*

Nucleotide Substitution Patterns Suggest That Positive Selection Has Been Acting on *RPP8*

We determined that *RPP8* is under positive selection for amino acid diversification by comparing nonsynonymous (K_a) and synonymous (K_s) nucleotide substitutions in differ-

Table 2. Genetic Analysis of L*er*-0 *rpp8* Mutants

	F_1		F_2		
Cross	R[a]	S[b]	R	S	χ^2 (1 degree of freedom)
Backcross					
L*er*-0 × *rpp8-1*	10	0	40	0	13.3 ($P < 0.005$)[c]
L*er*-0 × *rpp8-2*	9	0	54	15	0.4 ($0.5 < P < 0.9$)[c]
L*er*-0 × *rpp8-3*	3	0	46	12	0.6 ($0.5 < P < 0.9$)[c]
L*er*-0 × *rpp8-4*	5	0			
L*er*-0 × *rpp8-5*	22	0			
L*er*-0 × *rpp8-6*	21	0			
L*er*-0 × Col-0	11	0	35	14	0.3 ($0.5 < P < 0.9$)[c]
Outcross					
Col-0 × *rpp8-1*	0	22	15	13	0.1 ($0.5 < P < 0.9$)[d]
Col-0 × *rpp8-2*	0	26	2	96	117.0 ($P < 0.005$)[d]
Col-0 × *rpp8-3*	0	16	0	35	45.0 ($P < 0.005$)[d]
Col-0 × *rpp8-4*	0	16	0	21	27.0 ($P < 0.005$)[d]
Col-0 × *rpp8-5*	1	16			
Col-0 × *rpp8-6*	0	10			
Intercross					
rpp8-1 × *rpp8-2*	0	2	8	22	10.7 ($P < 0.005$)[d]
rpp8-1 × *rpp8-3*	1	6			
rpp8-1 × *rpp8-4*	1	11	10	23	9.0 ($P < 0.005$)[d]
rpp8-2 × *rpp8-3*	0	17	6	27	19.4 ($P < 0.005$)[d]
rpp8-2 × *rpp8-6*	0	4			
rpp8-4 × *rpp8-2*	0	11	3	13	9.1 ($P < 0.005$)[d]
rpp8-4 × *rpp8-3*	1	8	6	34	27.7 ($P < 0.005$)[d]

[a] R, resistant.

[b] S, susceptible.

[c] χ^2 values were calculated for a hypothesized 3:1 ratio of resistance to susceptibility for F_2 segregation from the backcrosses to L*er*-0.

[d] χ^2 values were calculated for a hypothesized 9:7 ratio of resistance to susceptibility for F_2 progeny segregating from the outcrosses to Col-0 and the intermutant crosses.

ent segments of the *rpp8*-Col, *RPP8-L*er, and *RPH8A* protein coding regions. In most cases in which evolution is conservative, the number of synonymous substitutions greatly exceeds that of nonsynonymous substitutions, leading to a K_a/K_s ratio <1. A K_a/K_s ratio >1 indicates selection for amino acid diversification (Kreitman and Akashi, 1995).

Much of amino acid divergence among the three RPP8 family members was concentrated in a subdomain of the LRRs (XX[L]X[L]XXXX), where leucine, isoleucine, or valine residues are found at the conserved positions designated by an L (Figures 3 and 4). This motif encompasses a predicted β strand/β turn region in which hydrophobic side chains at the conserved positions are buried in the core, and the nonconserved, interstitial residues (designated by X) are solvent exposed (Dixon et al., 1996; Jones and Jones, 1996). Calculations of K_a and K_s (Table 3) support the hypothesis that positive selection is acting to diversify putative solvent-exposed residues. For example, K_a in the XX(L)X(L)XXXX codons was 15.8% between *rpp8*-Col and *RPP8-L*er, whereas K_s was only 7.8% ($K_a/K_s = 2.0$). In the remainder of the coding sequence, excluding the XX(L)X(L)XXXX codons, K_a was fivefold lower, and the K_a/K_s ratio was 0.8, indicating a more conservative mode of evolution. A similar trend was apparent in the other two pairwise comparisons (Table 3).

Table 3. Pairwise K_a and K_s and Nucleotide and Amino Acid Homology

Comparison		Framed[a]	Nonframed[b]	Homology[c]	%
RPP8-Ler/*rpp8*-Col	K_a	15.8	2.8	nt id[d]	96
	K_s	7.8	3.5	aa id	92
	K_a/K_s	2.0	0.8	aa sim	94
RPP8-Ler/*RPH8A*	K_a	11.9	3.3	nt id	96
	K_s	4.1	4.3	aa id	91
	K_a/K_s	2.9	0.8	aa sim	94
rpp8-Col/*RPH8A*	K_a	13.0	3.1	nt id	96
	K_s	5.9	5.2	aa id	91
	K_a/K_s	2.2	0.6	aa sim	93

[a] These values represent the percentage of divergence between the two indicated genes in codons encoding the XXLXLXXXX motif in the 14 LRRs. These residues are framed in Figure 3 by vertical bars. K_a represents nonsynonymous divergence; K_s represents synonymous divergence.

[b] The percentage of divergence between the indicated genes over the entire protein coding sequence, except for codons that encode the framed XXLXLXXXX motifs.

[c] Determined over the entire length of the protein coding sequence.

[d] nt, nucleotide; aa, amino acid; id, identity; sim, similarity.

DISCUSSION

Plants may have an inherent disadvantage in the gene-for-gene arms race, because loss-of-function mutations in pathogen *avr* genes are sufficient to disarm gene-for-gene resistance. In contrast, the host must respond with a corresponding gain of function (recognition), and accumulation of point mutations in preexisting *R* genes alone may not provide sufficient structural diversity for novel resistance specificities to evolve in a timely fashion. Below, we discuss the implications of our results that are relevant to this conundrum.

Structurally Distinct NBS-LRR Subclasses Can Function in *P. parasitica* Resistance

It seems likely that novel *R* genes are recruited from preexisting *R* genes. Genes at the *L* and *M* loci are highly related to each other, and the *Cf* genes in tomato have very similar structural features. Based on these precedents, one might predict that RPP8 is a member of the TIR-NBS-LRR subclass, like RPP5 and the RPP1 family members. However, *RPP8* encodes an LZ-NBS-LRR protein and is most closely related to the *RPM1* bacterial *R* gene, demonstrating that the TIR-NBS-LRR and the LZ-NBS-LRR subclasses can function in resistance to *P. parasitica.* Similarly, the *Xa1* NBS-LRR gene and the *Xa21* extracytoplasmic LRR gene specify resistance to different isolates of the same bacterial pathogen of rice (Yoshimura et al., 1998). These observations suggest that plants can recruit a wide range of R proteins to recognize structurally diverse elicitors from the same pathogen. This is likely to be a key adaptive mechanism, in view of the apparent ease with which pathogens can alter or discard certain *avr* genes (van Kan et al., 1991; Rohe et al., 1995; Sweigard et al., 1995; Joosten et al., 1997).

Recent genetic evidence suggests that *RPP8*-mediated resistance may operate through a different signaling pathway from *RPP1* and *RPP5.* The Arabidopsis *eds1* (for enhanced disease susceptibility) mutation abolishes the function of several *RPP* genes, including *RPP5* and *RPP1*; however, *eds1* has little or no effect on *RPP8* function (N. Aarts et al., 1998). Similarly, the *ndr1* mutation, which partially inactivates several *RPP* genes and completely inactivates the Arabidopsis LZ-NBS-LRR bacterial *R* genes, does not affect *RPP8* (N. Aarts et al., 1998). *RPP8* is the only cloned Arabidopsis *R* gene that does not require either *NDR1* or *EDS1* for function. *RPP8* may therefore define a novel resistance pathway, or alternatively, *NDR1* and *EDS1* could be functionally redundant in *RPP8*-mediated resistance. We are currently constructing lines to test these possibilities.

A Novel *rpp8* Haplotype Was Generated by an Unequal Crossover between Linked Genes

Genetic analyses of *R* gene clusters, such as *Rp1* in maize and *M* in flax, have indicated that recombination between

repeated sequences in *R* gene clusters is a critical mechanism in *R* gene evolution (reviewed in Ellis et al., 1997; Hulbert, 1997), and the chimeric structure of the *rpp8*-Col allele adds to a growing body of molecular data that supports this proposal. Intra-allelic recombinants have been discovered in mutational screens at the *M* and *RPP5* loci (Anderson et al., 1997; Parker et al., 1997). These recombinant alleles arose from ectopic recombination between LRR-encoding modules that caused expansions or contractions in LRR copy number, thereby inactivating the gene. Intragenic recombination also has been proposed to occur within a 5′ region that is highly conserved between genes in the *Xa21* cluster, resulting in "promoter swaps" with minimal alterations in the coding sequences (Song et al., 1997). Finally, expansions and contractions in gene copy number have been observed to occur in the *Cf-4*/*9* complex by unequal crossing over between homologous intergenic regions (Parniske et al., 1997). The structure of *rpp8*-Col expands on these observations: *rpp8*-Col was generated by unequal crossing over between linked, nonallelic genes, it encodes a chimeric protein that differs dramatically from both progenitors, and it was present in at least one natural Arabidopsis population from which the Col-0 accession was derived. The observations that recombination can produce coding sequence chimeras, promoter swaps, and expansion or contraction in gene number and LRR copy number collectively underscore the role of recombination as a potent and versatile force in *R* gene evolution.

The functional roles of *rpp8*-Col and *RPH8A* are currently unknown. Neither gene is sufficient for resistance to Emco5 in Col-0, but both genes encode predicted full-length proteins. The nonrandom pattern of substitutions in β strand/β turn LRR-encoding motifs of both genes suggest that they are functional and remain under selection. We did not find *RPH8A* cDNAs among the 25 that were isolated, but *rpp8*-Col is expressed, as evidenced by complete identity to the Col-0 expressed sequence tag clone T14073. Therefore, it seems likely that these genes recognize currently undefined pathogens, and experiments are under way to define their functions genetically.

It is also possible that the *rpp8*-Col and *RPH8A* genes are obsolete or superfluous. A potential analogy may exist in the MHC, which contains functional class 1a antigen presentation genes as well as class 1b genes, which evolved from class 1a genes by duplication (Klein and O'hUigin, 1994). Some class 1b genes are functional, whereas others are expressed at reduced levels and appear to be evolving into nonexpressed pseudogenes. Class 1 MHC genes are thought to undergo turnover through cycles of birth and death as inactive or obsolete genes are supplanted by more efficient copies arising from duplication and divergence (Nei and Hughes, 1992). This process also may operate in plant disease resistance loci, which typically contain duplicated genes with unknown functions (Martin et al., 1994; Anderson et al., 1997; Wang et al., 1998). A significant fraction of these genes could be "molecular fossils" arising from gene turnover during the host–pathogen arms race. Nonfunctional *R* gene homologs may still play an important role, however, as repositories of sequence variation, as is seen among class 1 MHC genes (Hughes, 1995). Indeed, close relatives of *RPP8*-*L*er and *RPH8A* served as sequence donors when *rpp8*-Col was generated.

RPP8 Sequence Diversity Arises from Positive Selection

The divergence between the Col-0 and L*er*-0 *RPP8* alleles is much higher than is divergence among other Arabidopsis alleles (typically $<0.01\%$) (Bergelson et al., 1998). Our analysis of nucleotide substitution patterns suggests that the divergence among *RPP8* family members has been accelerated by positive diversifying selection. Clear evidence for positive selection in molecular evolution has rarely been observed (Kreitman and Akashi, 1995). Interestingly, the majority of genes that appear to be under selection for protein diversification are involved in host–pathogen interactions (Endo et al., 1996). Members of the *Cf-4*/*9* and *Xa21* extracellular LRR gene families are under positive selection in the LRR subdomain that is predicted to form a β strand/β turn structure (reviewed in Jones and Jones, 1996), and *RPP8* appears to be evolving in an analogous fashion. The fact that two of four sequenced *rpp8* mutations are missense substitutions in the XXLXLXXXX motif underscores the functional importance of this domain. It appears that both superfamilies of LRR disease resistance proteins are subject to diversifying selection, potentially for altered ligand binding capabilities in the LRRs. Interestingly, the divergence among the *RPP8* family members is concentrated in a slightly longer motif than in the *Cf-4*/*9* homologs (XXLXLXX) (Parniske et al., 1997). This possibly reflects adaptations for interactions with structurally dissimilar ligands.

What mechanisms generate the mutations upon which selection acts? Point substitutions are undoubtedly a primary source; however, we found it intriguing that most of the insertion/deletion sites among the three genes, including three indels that encompassed two or three codons, comprise direct repeats of varying degeneracies (Figure 5). This direct repeat structure suggests target site duplication and subsequent imprecise excision of a transposable element(s). Perhaps transposon insertions occurred in *RPP8* immediately after the *RPP8*-*L*er/*RPH8A* duplication, allowing one homolog to compensate for loss of the other until the transposon was excised. Periods of decreased pathogen pressure also could provide windows of opportunity for transposon insertions (or other sequence rearrangements) to accumulate at no cost to the plant. Regardless of whether the *RPP8* indels were generated by transposons, their presence suggests alternative mutational mechanisms that augment diversification from point substitutions.

Recombination and gene conversion also may have generated sequence diversity at *RPP8.* Although these two mechanisms cannot create nucleotide substitutions, they

can reassort existing mutations and cause amino acid substitutions by creating novel codons at recombination breakpoints, as seen in the *Cf-4*/*9* complex (Parniske et al., 1997). The region of complete identity at the 3′ end of *RPP8-L*er and *RPH8A* is suggestive of a recent gene conversion or double crossover. The patchwork pattern of nucleotide polymorphisms among the three *RPP8* family members also suggests that sequence exchanges have occurred during their evolution. Strong evidence for sequence exchanges among MHC genes exists, and theoretical simulations of MHC evolution have suggested that gene conversion is particularly important for the acquisition of polymorphism under conditions of weak selection (Parham and Ohta, 1996). This may be particularly significant in interactions with biotrophic plant pathogens in which penalties to the host are subtle (Holub and Beynon, 1996).

In combination with other recent comparative analyses of *R* gene structure, our results have established clear mechanistic parallels between the evolution of the two *R* gene superfamilies and other loci that determine the outcome of interactions. A growing body of data suggests that genes mediating coevolutionary self- and non-self-interactions are subject to a mode and tempo of evolution that differ dramatically from most other types of genes. Future studies expanding our understanding of the interplay between mutation, recombination, and selection in the generation of novel pathogen *R* genes should provide insights of broad academic and agricultural significance.

METHODS

Emco5 Derivation and Pathogenicity Tests

The *Peronospora parasitica* isolate Emco5 was intentionally isolated for the purpose of cloning the *RPP8* allele from *Arabidopsis thaliana* Landsberg *erecta* (L*er*-0) (Holub and Beynon, 1996). *RPP8* was defined initially in L*er*-0 by mapping a locus involved in recognition of the isolate Emoy2 by using recombinant inbred (RI) lines from a cross between L*er*-W100 (L*er* carrying nine phenotypic markers) and Wassilewskija (Ws-0) (Reiter et al., 1992). However, detailed mapping of *RPP8* was difficult in this cross because segregation was complicated by the presence of two additional *R* genes: *RPP1* from Ws-0 on chromosome 3 and *RPP4* from L*er*-W100 on chromosome 4. The L*er* × Columbia (Col-0) RI mapping population could not be used to map *RPP8* because Col-0 also carries a functional *RPP4* allele. Consequently, a series of "baiting host lines" carrying a functional *RPP4* allele from either L*er*-0 or Col-0 were used to select natural recombinant variants that had lost the presumed *ATR4* gene. This screen was initiated with the natural oospore (sexual inoculum) population from which Emoy2 was originally derived. Emco5 was eventually isolated using a L*er*-0 × Col-0 RI line (LC175) carrying a L*er*-0 *RPP4* allele and a Col-0 *RPP8* allele. The asexual inoculum from this isolate was used to confirm that it was compatible with both Col-0 and Ws-0 and was detected by a single *RPP* locus from L*er*-0 in the two available RI mapping populations. This locus was closely linked to the phenotypic marker *TT3* in the L*er*-W100 × Ws-0 RI population that defined the *RPP8* locus for Emoy2 resistance and cosegregated with *agp6* in the Ler-0 × Col-0 RI population.

Pathogenicity tests and mutant screens were conducted by spraying 7-day-old seedlings with a suspension of asexual inoculum (5×10^4 conidiosporangia mL^{-1}). Seedlings were then covered with a transparent dome to maintain high humidity and to contain the isolate throughout the experiment. Seedlings were grown for 7 days at 16 to 18°C with an 8-hr photoperiod in a Percival Scientific growth chamber (Boone, Iowa). *P. parasitica* growth was assessed visually at 7 days after inoculation by counting sporangiophores on both sides of the cotyledon and classifying plants as either N (no sporangiophores), L (1 to 10 sporangiophores), M (11 to 19 sporangiophores), or H (20 or more sporangiophores). To calculate the mean sporangiophore production shown in Table 1, we used actual numbers (0 to 10) for N and L cotyledons and assigned values of 15 (M) and 20 (H). Hyphal growth was assessed by staining inoculated seedlings with lactophenol–trypan blue (Koch and Slusarenko, 1990).

Identification and Sequencing of *rpp8* Mutants

We mutagenized L*er*-0 seeds with 0.15% ethyl methanesulfonate for 8 hr. M_2 seed was collected from lots of ~50 M_1 plants. We inoculated 1500 to 2000 7-day-old M_2 seedlings from each lot with Emco5 and visually screened for asexual sporulation 7 days later. We screened 35 M_2 lots and identified Emco5-susceptible seedlings from nine lots. Mutants were rescued by treatment with a 1:400 dilution of Ridomil (0.1 g L^{-1} metalaxyl; Novartis Ltd., Basel, Switzerland) and transferred to a 16-hr photoperiod at 23°C. Three mutants exhibited very inconsistent resistance phenotypes and are not described.

We used L*er*-0 plants with the *ttg* marker in this screen to distinguish rogue seeds or outcross contaminants visually. In addition, we tested DNA from each mutant with a set of cleaved amplified polymorphic sequences and simple sequence length polymorphism markers from throughout the genome that distinguished polymorphisms between L*er*-0 and two Emco5-compatible accessions, Col-0 and Ws-0. A L*er*-0 pattern was observed for every marker tested in each mutant (data not shown), thereby demonstrating that the mutants were derived from the L*er*-0 background.

We determined the sequence of the mutant L*er rpp8* alleles by polymerase chain reaction (PCR) amplification and direct sequencing of the entire PCR product. Multiple amplification products were sequenced to check for misincorporations during the amplification. We designed PCR primers based on the sequence variation that exists between *RPP8-L*er and *RPH8A* to amplify specifically the *RPP8-L*er gene. Gene specificity of the primer sets was confirmed using pRPP8 and pRPH8A as controls. Primer sequences will be provided upon request.

Yeast and Bacterial Artificial Chromosome Manipulation

Yeast artificial chromosome (YAC) clones that hybridized with markers *Dfr* and *mi83* were kindly provided by R. Schmidt (Max Delbrück Laboratory, Cologne, Germany) and C. Dean (John Innes Centre, Norwich, UK) and assembled into a contig by hybridization with known nearby markers and YAC ends. The YAC ends were cloned by vectorette PCR (Matallana et al., 1992). TAMU bacterial artificial chromosome (BAC) clones (Choi et al., 1995) hybridizing with the 15C8RE and *Spl2* markers were identified, and their integrity was

confirmed by DNA gel blotting with several markers that span the contig.

cDNA and Genomic Clone Isolation

Two L*er*-0 cDNA libraries (Parker et al., 1997) were kindly provided by M. Coleman (John Innes Centre). One library was size selected for inserts >1.8 kb. A total of ~1.8 million plaque-forming units were screened with the CK1 probe at 65°C in 2 × Denhardt's solution (1 × Denhardt's solution is 0.02% Ficoll, 0.02% PVP, and 0.02% BSA). Filters were washed at 65°C in 2 × SSC (1 × SSC is 0.15 M NaCl and 0.015% sodium citrate). Twenty-five clones that gave signals of varying intensity were purified. Sequence was obtained from both ends of each clone, and the clone that had the longest insert was completely sequenced. The 9L9 and 25M19 cosmids were obtained by screening a L*er*-0 genomic DNA library in the pCLD04541 binary cosmid vector (kindly provided by M. Botella, John Innes Centre), as described above. Cosmid DNA was extracted by standard alkaline lysis procedures, and restriction digest patterns of each cosmid were compared with L*er*-0 genomic DNA on gel blots to check for rearrangements in the insert.

RPP8 Subclones

Fragments of the 9L9 cosmid were subcloned directly into binary plasmid vectors by standard procedures. pRPP8 contains a 5488-bp EcoRI fragment that includes the entire *RPP8-L*er coding sequence as well as 679 bp of the 5′ flanking sequence and 1288 bp of the 3′ flanking sequence. pRPH8A contains a 5672-bp EcoRI fragment that includes the entire *RPH8A* coding sequence, 1334 bp of the 5′ sequence, and 815-bp of the 3′ sequence. The cyclin C homolog was contained on a 4321-bp SacI subclone (2826 bp of the 5′ sequence and 639 bp of the 3′ sequence). The *NF22* homolog was contained on a 6576-bp SacI fragment (>2 kb of both the 5′ and 3′ sequences). The *CycH* and *NF22H* subclones are in the pGPTV-Kan binary vector (Becker et al., 1992). pRPP8 and pRPH8A are in pBAR1, which was derived from pGPTV-Bar by replacing the β-glucuronidase gene with the polylinker from pBluescript SK+ (Stratagene, La Jolla, CA) (B. Holt, D. Boyes, and J.L. Dangl, unpublished data).

Plant Transformation

Binary clones were transformed into *Agrobacterium tumefaciens* GV3101 by electroporation. Tetracycline at 12.5 mg mL^{-1} was not a reliable selection for transformed Agrobacterium, but kanamycin at 50 mg mL^{-1} worked consistently. We transformed plants using the vacuum infiltration method (Bechtold et al., 1993). We extracted cosmid DNA from an aliquot of each Agrobacterium culture used for transformation and checked for insert rearrangements by DNA gel blotting. Plants transformed with pGPTV-Kan were selected on Murashige and Skoog (Gibco BRL, Grand Island, NY) medium with 50 mg mL^{-1} kanamycin. pBar1 transformants were selected by spraying seedlings at 5, 6, and 7 days after germination with a solution of 0.01% BASTA (200 g L^{-1} glufosinate ammonium; AgrEvo USA, Wilmington, DE) and 0.01% Silwet L-77 (Lehle Seeds, Round Rock, TX). Plants that survived this selection were sprayed again at 14 and 15 days after germination.

DNA Sequencing

To obtain the L*er*-0 sequence, we fragmented the 9L9 cosmid and shotgun subcloned ~1-kb fragments into the M13 vector. Recombinant M13 clones that contained L*er*-0 DNA were identified by hybridization and sequenced with the M13 forward primer. These random sequences were assembled into contigs, and gaps were filled by primer walking. To determine the Col-0 sequence, we isolated two contiguous, *RPP8*-hybridizing, BglII subclones from the 2D9 BAC that spanned *RPP8.* We obtained most of the sequence from both inserts with a collection of primers derived from the L*er*-0 sequence. Gaps were filled by primer walking.

Sequence similarity searches were conducted using the BLAST program with default settings (Altschul et al., 1990). Conceptual translations, pairwise comparisons, and multiple alignments were performed with default settings using the Translate, Gap, and Pileup programs of the software package, version 9.1, from the Genetics Computer Group (Madison, WI). Nonsynonymous (K_a) and synonymous (K_s) substitutions were calculated using the Genetics Computer Group Diverge program, which corrects for multiple hits and unequal rates of transitions versus transversions. Residues at the position designated by an L in the XX(L)X(L)XXXX motif were omitted from the calculations of K_a and K_s, based on the rationale that they are under selection for conservation of function (Parniske et al., 1997). Polymorphic sites were displayed with the Sequence Output program (B.G. Spratt, University of Sussex, Brighton, UK).

ACKNOWLEDGMENTS

We gratefully acknowledge the following contributions: Miguel Botella, Mark Coleman, and Jonathan Jones for providing L*er*-0 genomic and cDNA libraries; Guillermo Cardon for providing the *Spl2* sequence; Canan Can and Patricia Chimot for assistance with *P. parasitica* isolate characterization and confirmation of pathology data; Claire Lister, Caroline Dean, and Renate Schmidt for assistance with genetic and physical mapping; and Ian Crute, Ben Holt, Susanne Kjemtrup, and Rheinhard Kunze for providing helpful comments on the manuscript. This work was supported by grants from the U.S. Department of Agriculture National Research Initiative Competitive Grants Program (to J.L.D.), the U.K. Biotechnology and Biological Sciences Research Council (to E.B.H.), and The Netherlands Technology Foundation (to M.G.M.A.). J.M.M. was supported by a postdoctoral fellowship from the National Institutes of Health.

Received June 23, 1998; accepted September 11, 1998.

REFERENCES

Aarts, M.G.M., te Lintel Hekkert, B., Holub, E.B., Beynon, J.L., Stiekema, W.J., and 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.

Aarts, N., Metz, M., Holub, E., Staskawicz, B.J., Daniels, M.J., and Parker, J.E. (1998). Different requirements for *EDS1* and

NDR1 by disease resistance genes define at least two *R* gene–mediated pathways in *Arabidopsis.* Proc. Natl. Acad. Sci. USA **95,** 10306–10311.

Altschul, S.F., Gish, W., Miller, W., Meyers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215,** 403–410.

Anderson, P.A., Lawrence, G.J., Morrish, B.C., Ayliffe, M.A., Finnegan, E.J., and Ellis, J.G. (1997). Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. Plant Cell **9,** 641–651.

Bechtold, N., Ellis, J., and Pelletier, G. (1993). *In planta Agrobacterium*-mediated gene transfer by infiltration of *Arabidopsis thaliana* plants. C. R. Acad. Sci. Ser. III Sci. Vie **316,** 1194–1199.

Becker, D., Kemper, E., Schell, J., and Masterson, R. (1992). New plant binary vectors with selectable markers located proximal to the left T-DNA border. Plant Mol. Biol. **20,** 1195–1197.

Bent, A.F. (1996). Plant disease resistance genes: Function meets structure. Plant Cell **8,** 1757–1771.

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.

Bergelson, J., Stahl, E., Dudek, S., and Kreitman, M. (1998). Genetic variation within and among populations of *Arabidopsis thaliana.* Genetics **148,** 1311–1323.

Botella, M.A., Parker, J.E., Frost, L.N., Bittner-Eddy, P.D., Beynon, J.L., Daniels, M.J., Holub, E.B., and Jones, J.D.G. (1998). Three genes of the Arabidopsis *RPP1* complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. Plant Cell **10,** 1847–1860.

Choi, S., Creelman, R.A., Mullet, J.E., and Wing, R.A. (1995). Construction and characterization of a bacterial artificial chromosome library of *Arabidopsis thaliana.* Plant Mol. Biol. **13,** 124–128.

Dangl, J.L. (1992). The major histocompatibility complex a la carte: Are there analogies to plant disease resistance genes on the menu? Plant J. **2,** 3–11.

Dangl, J.L. (1995). Pièce de résistance: Novel classes of plant disease resistance genes. Cell **80,** 363–366.

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.

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

Endo, T., Ikeo, K., and Gojobori, T. (1996). Large-scale search for genes upon which positive selection may operate. Mol. Biol. Evol. **13,** 685–690.

Flor, H.H. (1971). Current status of the gene-for-gene concept. Annu. Rev. Phytopathol. **9,** 275–296.

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.

Hammond-Kosack, K.E., and Jones, J.D.G. (1996a). Resistance gene–dependent plant defense responses. Plant Cell **8,** 1773–1791.

Hammond-Kosack, K.E., and Jones, J.D.G. (1996b). Plant disease resistance genes. Annu. Rev. Plant Physiol. Plant Mol. Biol. **48,** 575–607.

Holub, E.B., and Beynon, J.L. (1996). Symbiology of mouse ear cress (*Arabidopsis thaliana*) and oomycetes. Adv. Bot. Res. **24,** 228–273.

Holub, E.B., Beynon, J.L., and Crute, I.R. (1994). Phenotypic and genotypic characterization of interactions between isolates of *Peronospora parasitica* and accessions of *Arabidopsis thaliana.* Mol. Plant-Microbe Interact. **7,** 223–239.

Hughes, A.L. (1995). Origin and evolution of HLA class I pseudogenes. Mol. Biol. Evol. **12,** 247–258.

Hughes, A.L., and Yeager, M. (1997). Molecular evolution of the vertebrate immune system. Bioessays **19,** 777–786.

Hulbert, S.H. (1997). Structure and evolution of the *Rp1* complex conferring rust resistance in maize. Annu. Rev. Phytopathol. **35,** 293–310.

Jones, D.A., and Jones, J.D.G. (1996). The roles of leucine rich repeats in plant defences. Adv. Bot. Res. Adv. Plant Pathol. **24,** 90–167.

Joosten, M.H.A.J., Vogelsang, R., Cozijnsen, T.J., Verberne, M.C., and De Wit, P.J.G.M. (1997). The biotrophic fungus *Cladosporium fulvum* circumvents *Cf-4*–mediated resistance by producing unstable AVR4 elicitors. Plant Cell **9,** 367–379.

Karrer, E.E., Beachy, R.N., and Holt, C.A. (1998). Cloning of tobacco genes that elicit the hypersensitive response. Plant Mol. Biol. **36,** 681–690.

Keen, N.T. (1990). Gene-for-gene complementarity in plant–pathogen interactions. Annu. Rev. Genet. **24,** 447–463.

Klein, J., and O'hUigin, C. (1994). The conundrum of nonclassical histocompatibility complex genes. Proc. Natl. Acad. Sci. USA **91,** 6251–6252.

Kobe, B., and Deisenhofer, J. (1993). Crystal structure of porcine ribonuclease inhibitor, a protein with leucine rich repeats. Nature **366,** 751–756.

Kobe, B., and Deisenhofer, J. (1995). Proteins with leucine rich repeats. Curr. Opin. Struct. Biol. **5,** 409–416.

Koch, E., and Slusarenko, A. (1990). *Arabidopsis* is susceptible to infection by a downy mildew fungus. Plant Cell **2,** 437–445.

Kreitman, M., and Akashi, H. (1995). Molecular evidence for natural selection. Annu. Rev. Ecol. Syst. **26,** 403–422.

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.

Leister, R.T., Ausubel, F.M., and Katagiri, F. (1996). Molecular recognition of pathogen attack occurs inside of plant cells in plant disease resistance specified by the *Arabidopsis* genes *RPS2* and *RPM1.* Proc. Natl. Acad. Sci. USA **93,** 15497–15502.

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.

Martin, G.B., Frary, A., Wu, T., Brommonschenkel, S., Chunwongse, J., Earle, E.D., and Tanksley, S.D. (1994). A member of the tomato *Pto* gene family confers sensitivity to fenthion resulting in rapid cell death. Plant Cell **6,** 1543–1552.

Matallana, E., Bell, C.J., Lu, P.J., and Ecker, J.E. (1992). Genetic and physical linkage of the *Arabidopsis* genome: Methods for anchoring yeast artificial chromosomes. In Methods in Arabidopsis Research, C. Koncz, N.-H. Chua, and J. Schell, eds (Singapore: World Scientific), pp. 144–169.

Mindrinos, 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.

Nei, M., and Hughes, A. (1992). Balanced polymorphism and evolution by the birth and death process in the MHC loci. In Proceedings of the Eleventh Histocompatibility Workshop and Conference, K. Tsuji, M. Aizawa, and T. Suzuki, eds (Oxford, UK: Oxford University Press), pp. 27–38.

Parham, P., and Ohta, T. (1996). Population biology of antigen presentation by MHC class I molecules. Science **272,** 67–74.

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.

Parniske, M., Hammond-Kosack, K.E., Golstein, C., Thomas, C.M., Jones, D.A., Harrison, K., Wulff, B.B.H., and Jones, J.D.G. (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.

Reiter, R.S., Williams, J.G.K., Feldman, K.A., Rafalski, J.A., Tingey, S.V., and Scolnik, P.A. (1992). Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc. Natl. Acad. Sci. USA **89,** 1477–1481.

Rohe, M., Gierlich, A., Hermann, H., Hahn, M., Schmidt, B., Rosahl, S., and Knogge, W. (1995). The race-specific elicitor, NIP1, from the barley pathogen, *Rhynchosporium secalis*, determines avirulence on host plants of the *Rrs1* resistance genotype. EMBO J. **14,** 4168–4177.

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.

Song, W.-Y., Wang, G.-L., Chen, L.-L., Kim, H.-S., Pi, L.-Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.-X., Zhu, L.-H., Fauquet, C., and Ronald, P.C. (1995). A receptor kinase–like protein encoded by the rice disease resistance gene, *Xa21.* Science **270,** 1804–1806.

Song, W.-Y., Pi, L.-Y., Wang, G.-L., Gardner, J., Holsten, T., and Ronald, P.C. (1997). Evolution of the rice *Xa21* disease resistance gene family. Plant Cell **9,** 1279–1287.

Sweigard, J.A., Carroll, A.M., Kang, S., Farrall, L., Chumley, F.G., and Valent, B. (1995). Identification, cloning, and characterization of *PWL2*, a gene for host species specificity in the rice blast fungus. Plant Cell **7,** 1221–1233.

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.* Mol. Plant-Microbe Interact. **7,** 214–222.

van Kan, J.A.L., van den Ackerveken, G.F.J.M., and de Wit, P.G.J.M. (1991). Cloning and characterization of cDNA of avirulence gene *avr9* of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. Mol. Plant-Microbe Interact. **4,** 52–59.

Wang, G.-L., Ruan, D.-L., Song, W.-Y., Sideris, S., Chen, L., Pi, L.-Y., Zhang, S., Zhang, Z., Fauquet, C., Gaut, B.S., Whalen, M.C., and Ronald, P.C. (1998). *Xa21D* encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell **10,** 765–779.

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.

Yoshimura, S., Yamanouchi, U., Katayose, Y., Toki, S., Wang, Z.X., Kono, I., Kurata, N., Yano, M., Iwata, N., and 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.