The Plant Cell, Vol. 11, 2099–2111, November 1999, www.plantcell.org © 1999 American Society of Plant Physiologists

Pronounced Intraspecific Haplotype Divergence at the *RPP5* Complex Disease Resistance Locus of Arabidopsis

Laurent Noël,[1,2] Tracey L. Moores,[1] Erik A. van der Biezen,[1] Martin Parniske, Michael J. Daniels, Jane E. Parker, and Jonathan D. G. Jones[3]

Sainsbury Laboratory, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom

In Arabidopsis ecotype Landsberg *erecta* (L*er*), *RPP5* confers resistance to the pathogen *Peronospora parasitica. RPP5* is part of a clustered multigene family encoding nucleotide binding–leucine-rich repeat (LRR) proteins. We compared 95 kb of DNA sequence carrying the L*er RPP5* haplotype with the corresponding 90 kb of Arabidopsis ecotype Columbia (Col-0). Relative to the remainder of the genome, the L*er* and Col-0 *RPP5* haplotypes exhibit remarkable intraspecific polymorphism. The *RPP5* gene family probably evolved by extensive recombination between LRRs from an *RPP5*-like progenitor that carried only eight LRRs. Most members have variable LRR configurations and encode different numbers of LRRs. Although many members carry retroelement insertions or frameshift mutations, codon usage analysis suggests that regions of the genes have been subject to purifying or diversifying selection, indicating that these genes were, or are, functional. The *RPP5* haplotypes thus carry dynamic gene clusters with the potential to adapt rapidly to novel pathogen variants by gene duplication and modification of recognition capacity. We propose that the extremely high level of polymorphism at this complex resistance locus is maintained by frequency-dependent selection.

INTRODUCTION

Plant resistance to animal, fungal, bacterial, and viral pathogens often is governed by gene-for-gene interactions between host resistance (*R*) genes and corresponding pathogen avirulence (*Avr*) genes (Crute, 1994). In the absence of matching *R* genes, *Avr* gene products presumably enhance pathogen virulence (Collmer, 1998). Different pathogen strains carry different arrays of *Avr* genes (e.g., Holub and Beynon, 1996), which may indicate some functional redundancy. In crop monocultures, *R* genes impose strong selection on pathogen *Avr* genes for mutations to virulence; as a consequence, plant breeders must continuously recruit new *R* genes from wild relatives. In natural populations that are genetically and spatially more diverse, the evolutionary forces at work are less clear. However, the population dynamics of all plant–pathogen systems are strongly influenced by genetic variation in resistance and virulence determinants (Crute, 1994; Simms, 1996). Parallels have been proposed between the evolution of plant *R* genes and of genes in the major histocompatibility complex (MHC) of vertebrates (Dangl, 1992; Michelmore and Meyers, 1998). It is of major interest, therefore, to understand the molecular basis for the evolution of host–pathogen specificity and to characterize intraspecific natural variation at *R* gene haplotypes. Interspecific comparisons have been reported elsewhere (Parniske et al., 1997; Thomas et al., 1998).

The capacity to adapt rapidly to a broad spectrum of novel pathogen variants imposes a need for a mechanism that generates new *R* gene alleles. Many *R* genes are members of gene families residing at complex loci that can carry several distinct pathogen recognition specificities (Hammond-Kosack and Jones, 1997). *R* genes, therefore, probably evolve novel Avr recognition capacities through gene duplication, diversification, and subsequent selection. Most cloned *R* genes encode a leucine-rich repeat (LRR) domain (Staskawicz et al., 1995; Hammond-Kosack and Jones, 1997), a motif that has been associated with protein–protein interactions (Kobe and Deisenhofer, 1994). The LRRs in *R* gene products, therefore, are envisaged to specify pathogen recognition by direct or indirect interaction with Avr molecules (Jones and Jones, 1996). Indeed, in the porcine ribonuclease inhibitor protein, the solvent-exposed residues of a parallel β sheet form a surface for ligand interactions (Kobe and Deisenhofer, 1995). In several *R* gene products, the predicted solvent-exposed residues in the β-sheet regions are hypervariable and correlate with differential pathogen recognition (Parniske et al., 1997; Thomas et al., 1997; Botella et al., 1998; Dixon et al., 1998; McDowell et al., 1998; Meyers et al., 1998b; Ellis et al., 1999).

Variability in the solvent-exposed LRR residues probably

[1] These authors contributed equally to this work.

[2] Current address: Institute of Genetics, Martin-Luther-Universität, Weinbergweg 22, D-06120 Halle, Germany.

[3] To whom correspondence should be addressed. E-mail jonathan.jones@bbsrc.ac.uk; fax 1603-250024.

is accomplished through accumulation of and selection for point mutations. The excess of nonsynonymous substitutions over synonymous substitutions within codons encoding the predicted ligand-interacting LRR residues indicates that diversifying selection acts on these amino acids (Michelmore and Meyers, 1998). The same criterion was used to infer diversifying selection in the peptide binding region of the MHC class I genes (Hughes and Yeager, 1998). In contrast, purifying selection appears to act on regions encoding the presumed signal effector portion of the *R* gene product (Parniske et al., 1997; Wang et al., 1998; Botella et al., 1998; McDowell et al., 1998; Meyers et al., 1998b). The meiotic instability and the generation of novel *R* gene specificities at some *R* gene loci suggested that unequal recombination and gene conversion could also contribute to diversity (Ellis et al., 1997; Hulbert, 1997). This conclusion was supported recently by comparative sequence analysis of several *R* genes (Parniske et al., 1997; Botella et al., 1998; McDowell et al., 1998; Meyers et al., 1998b; Ellis et al., 1999).

The small crucifer Arabidopsis is a model for plant molecular genetics, and its genome sequence is being determined. The biotrophic oomycete pathogen *Peronospora parasitica* naturally infects and completes its life cycle in Arabidopsis (Holub and Beynon, 1996). In Arabidopsis ecotype Landsberg *erecta* (L*er*), a single gene, *RPP5*, on chromosome 4 (Parker et al., 1997) confers resistance to strain Noco2 of *P. parasitica.* In contrast, resistance to the same strain in ecotype Wassilewskija is conferred by the *RPP1* locus, which is located on chromosome 3. *RPP1* consists of at least three very similar genes, each probably recognizing a different *Avr* gene product (Botella et al., 1998). The *RPP5* and *RPP1* genes are members of the largest class of plant *R* genes that encode nucleotide binding (NB) sites and LRR domains (NB-LRR proteins; Staskawicz et al., 1995; Hammond-Kosack and Jones, 1997). Both *RPP5* and *RPP1* are further grouped into a TIR-NB-LRR subclass based on their N termini, which possess similarity to the cytoplasmic effector domains of the Drosophila and human Toll and interleukin-1 receptors (TIR domain, for Toll, interleukin-1 receptor, and R protein; Parker et al., 1997; Botella et al., 1998). The central part of all NB-LRR proteins comprises three motifs predicted to constitute an NB pocket and several short motifs without known function (Hammond-Kosack and Jones, 1997). This region is shared with the animal apoptosis regulatory proteins CED-4 and Apaf-1 and has been designated the NB-ARC domain (ARC, for Apaf-1, R protein, and CED-4; Van der Biezen and Jones, 1998a), and also the Ap-ATPase domain (Aravind et al., 1999). NB-LRR proteins have a C-terminal LRR domain that includes 21 LRRs in RPP5 (Parker et al., 1997) and either nine or 10 LRRs in the RPP1 proteins (Botella et al., 1998).

DNA gel blot analysis showed that all analyzed Arabidopsis ecotypes contain multiple, highly polymorphic, *RPP5*- and *RPP1*-related sequences (Parker et al., 1997; Botella et al., 1998). Therefore, it is of interest to study the evolutionary forces that generate this striking level of intraspecific polymorphism between complex disease resistance haplotypes. We determined the DNA sequence of the entire L*er RPP5* haplotype containing 10 homologs and compared it with that of the ecotype Columbia (Col-0), which contains eight homologs (Bevan et al., 1998). This analysis showed that the two *RPP5* gene clusters have diverged to an extraordinary degree compared with most other L*er* and Col-0 loci. The *RPP5* family members have different numbers of LRRs ranging from 13 to 23, and they could have evolved from a progenitor gene with only eight LRRs. In addition, the predicted ligand-interacting LRR residues are hypervariable and appear to be subject to diversifying selection, indicating that the *RPP5* family members function or functioned in recognition of different pathogen Avr products. Gene conversion, point mutation, retrotransposition, and unequal recombination all contributed to the high mutation rate responsible for the divergence of the *RPP5* haplotypes. This intraspecific DNA sequence comparison of two complex disease resistance haplotypes reveals that pathogens exert strong selective pressure for enhanced polymorphism at *R* gene loci. Possible mechanisms for this selection are discussed.

RESULTS

Structural Divergence of the L*er* and Col-0 *RPP5* Haplotypes

Analysis of the 95 kb of DNA sequence of the L*er RPP5* haplotype revealed the presence of 10 *RPP5* homologs: *La-A* (*RPP5*) to *La-J* (Figure 1). Apart from a truncated member at the telomeric end (*La-J*), all other homologs are oriented in the same direction. In addition, sequence analysis identified three *Ty1*/*Copia*–like retroelements (Bennetzen, 1996). Furthermore, one complete copy and one truncated copy of a mitochondrial 5S rRNA gene and eight truncated copies of a serine/threonine protein kinase gene are present (Figure 1). By contrast, the 90-kb sequence of the *RPP5* haplotype in Col-0, which is susceptible to infection by *P. parasitica* Noco2, consists of eight homologs (*Col-A* to *Col-H*); most of these also are oriented in the same direction (Bevan et al., 1998; Figure 1). As in L*er*, the Col-0 *RPP5* haplotype contains eight truncated copies of the same protein kinase pseudogene. It also contains a *Ty1*/*Copia*–like and a *Ty3*/*Gypsy*–like retroelement inserted in two *RPP5* homologs. The *Ty1*/*Copia*–like retroelements inserted in *La-D* and *La-G* are identical; the retroelements in *La-H* and *Col-D* differ from those in *La-D* and *La-G*, and from each other.

Structural comparison of the L*er* and Col-0 haplotypes shows that the regions flanking the gene clusters are almost identical, including *La-J* and *Col-H*, which are two truncated homologs (Figure 1). However, colinearity is almost completely absent within these boundaries. First, the relative location and number of homologs within each haplotype differ markedly, accounting for the high level of polymorphism re-

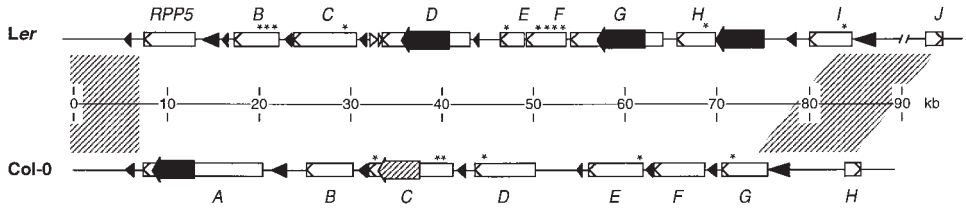

Figure 1. Representation of the DNA Sequence of the L*er* and Col-0 *RPP5* Haplotypes.

The *RPP5* homologs in L*er* and Col-0 are named from A to J and are shown as boxes, with the direction of transcription indicated by arrowheads. The hatched areas at the centromeric (left) and telomeric (right) ends indicate almost completely (96%) identical sequences. Asterisks indicate the positions of the open reading frame (ORF)–disrupting point mutations. Retroelement insertions are shown as black arrows (*Ty1*/*Copia* class) and as a hatched arrow (*Ty3*/*Gypsy* class). Similarity to a serine/threonine protein kinase pseudogene is shown as filled triangles; the two open triangles indicate sequences with similarity to a mitochondrial 5S rRNA gene.

vealed by DNA gel blot analysis (Parker et al., 1997). Second, the position in the cluster of a particular L*er* homolog does not correspond to the position of the most closely related member in the Col-0 haplotype (and vice versa). For example, *La-B* is most similar to *Col-C* (96% identity). Third, aside from the presence of truncated protein kinase sequences, limited sequence similarity exists in the intergenic regions. For example, no mitochondrial 5S rRNA gene is present in the Col-0 region. The extensive sequence differences between the L*er* and Col-0 *RPP5* gene clusters are in remarkable contrast with the generally observed low level of polymorphisms between these ecotypes (e.g., Bergelson et al., 1998). Therefore, we conclude that since the evolutionary separation of the ecotypes, the L*er* and Col-0 *RPP5* haplotypes have diversified extensively, whereas little divergence has occurred at the vast majority of other loci.

RPP5 Family Members Are Expressed but Have Disrupted Reading Frames

All homologs, including promoters, introns, and 3′ regions, have a similar overall structure. Three members have an eighth exon, including *RPP5*, *Col-A*, and *Col-F.* However, deletion of this last short exon does not compromise *RPP5* function (Parker et al., 1997) and therefore it may not be required for the other members. The most variable region encodes the LRRs, and this region often differs by multiple duplications and deletions (see next three sections). The nucleotide sequence identity between pairs of genes ranges from 74% for *Col-B*/*Col-G* to 99% for *La-J*/*Col-H.* Distance analysis does not preferentially group family members from one haplotype (data not shown). Thus, family members within one haplotype (paralogs) are not more similar to each other than they are to those from the other haplotype (orthologs). Phylogenetic analysis of the *RPP5* family conducted with full-length gene sequences, however, provides limited information, because considerable sequence exchange has occurred between homologs (see *RPP5* Family Members Are Mosaics of Shared Gene Segments, below).

Apart from *RPP5* itself, none of the other nine L*er* homologs is predicted to encode full-length open reading frames (ORFs). At the Col-0 haplotype, only two homologs (*Col-B* and *Col-F*) encode intact ORFs (Figure 1). Five members are severely truncated (*La-E, La-F*, *La-H*, *La-J*, and *Col-H*). One member (*La-D*) contains duplicated exon 2 and exon 3 sequences, possibly resulting from the insertion of the retroelement present at this location. However, most other homologs appear to be full length and carry only one ORF-disrupting mutation (Figure 1): two members carry a retroelement insertion (*La-G* and *Col-A*), and five members contain a single point mutation (*La-C*, *La-I*, *Col-D*, *Col-E*, and *Col-G*). The single mutations in *Col-A*, *Col-D*, and *Col-G* reside close to the 3′ end of the gene, beyond the region encoding the LRRs. Because a 3′ truncated *RPP5* transgene is still functional (Parker et al., 1997), it is possible that these three homologs also encode functional proteins. *La-B* and *Col-C* carry three point mutations, and the latter also carries a retroelement insertion.

RNA gel blot analysis of the expression of the *RPP5* homologs in healthy L*er* and Col-0 leaves revealed multiple mRNA transcripts (Figure 2). The *RPP5* transcript is 4.2 kb in length and is not abundantly expressed. A smaller (1.9 kb) and more abundant transcript in L*er* is encoded by *La-G*, and its corresponding cDNA was previously designated *SFH* (for sequence from homolog; Parker et al., 1997). The truncated transcript's presence and size show remarkable resemblance to transcripts resulting from alternative splicing of the structurally related *R* genes *N* from tobacco and *L6* from flax (Whitham et al., 1994; Ellis et al., 1997; Parker et al., 1997). The 3′ end of the corresponding cDNA consists of sequences belonging to the retroelement inserted in *La-G.* Several other rare *RPP5*-related transcripts are present in L*er*, but their corresponding genes have not been defined. In Col-0, transcripts of ~4.5 kb, slightly longer than that from *RPP5*, are present. Of the six Col-0 expressed sequence

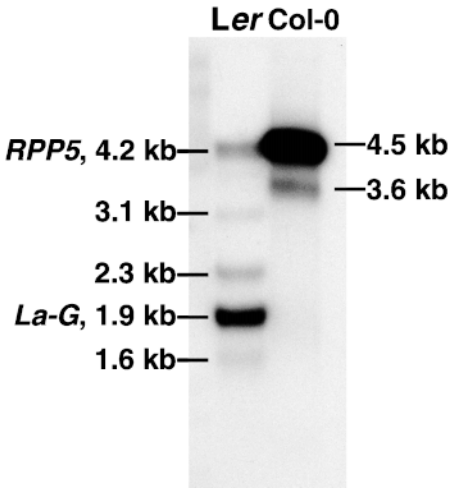

Figure 2. Transcript Analysis of *RPP5* Family Members in L*er* and Col-0.

Poly(A)$^+$ RNA gel blot analysis was conducted. Transcripts encoded by *RPP5* and *La-G* are indicated, with approximate transcript sizes indicated in kilobases.

tags present in the Arabidopsis AtEST database (September 1999), five correspond to *Col-F* (GenBank accession numbers AA86077, T41662, T03992, Z46716, and Z46715), and one corresponds to *Col-C* (N96078), indicating that at least these members are expressed.

RPP5 Family Members Are Mosaics of Shared Gene Segments

For determination of sequence relationships, the *RPP5* family members were fingerprinted by analysis of informative polymorphic sites (IPSs). IPSs are characteristic nucleotide substitutions shared between homologous sequences and distinguish these from other homologous sequences (Parniske et al., 1997). Sequence affiliations were identified when three or more consecutive IPSs are contained within homologous segments. For example, in exon 1 of *Col-F*, all IPSs are shared with those of *RPP5* (Figure 3), except between base pairs 234 and 243 in which four consecutive IPSs are shared with exon 1 of several other homologs (*La-D*, *La-I*, *Col-A*, *Col-E*, and *Col-G*). Overall, this analysis showed an extensive patchwork distribution of nucleotide polymorphisms between most family members (Figure 3), as also was inferred in a previous interspecific haplotype comparison at the *Cladosporium fulvum Cf-4*/*Cf-9* locus (Parniske et al., 1997). The IPS analysis clearly revealed traces of unequal recombination and/or gene conversion events against a background of homolog-specific polymorphisms. Three pairs of family members are almost identical in sequence: *Col-E*/*Col-G* (91%), *La-B*/*Col-C* (96%), and *La-J*/*Col-H* (99%).

From the number of distinct gene segments shared between the *RPP5* haplotypes, the minimum number of family members present before the separation of L*er* and Col-0 can be deduced. For example, five different exon 1 gene segments are shared between the L*er* and Col-0 haplotypes, suggesting that in the ancestral haplotype, the *RPP5* gene family consisted of at least five members. After the separation of the ecotypes, subsequent independent sequence exchanges between paralogs or with other *RPP5* haplotypes presumably contributed to further divergence of the gene clusters.

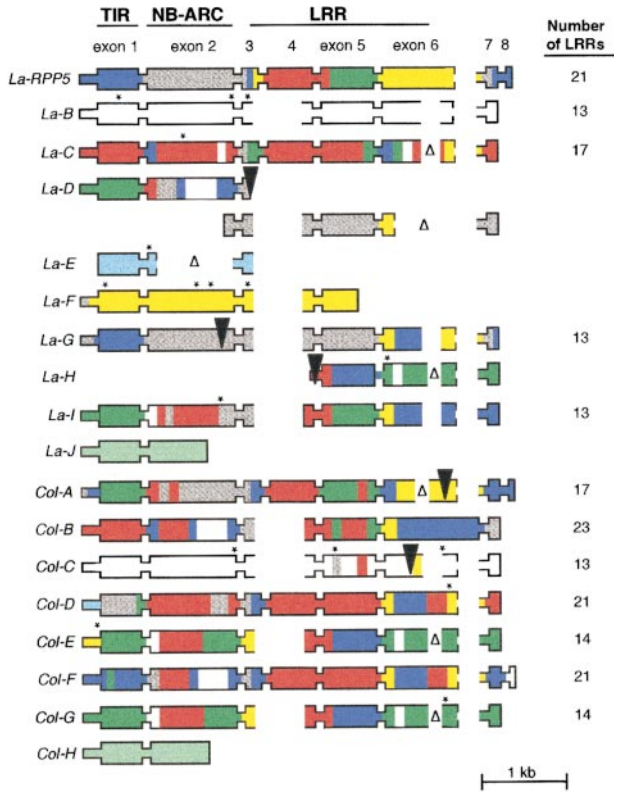

Figure 3. Sequence Exchange among the *RPP5* Multigene Family.

Gene structures are represented with exons as wide rectangles; promoters and introns are shown as narrow rectangles. Sequence affiliations were inferred from runs of at least three consecutive IPSs, with each shown in a different color. Asterisks indicate the position of the ORF-disrupting mutations, and the black arrowheads indicate the locations of the retroelement insertions. Deletions within the genes are indicated by open triangles. The spaces in some homologs are created because other homologs carry duplications in that region. The numbers of LRRs in the homologs are shown at the right.

RPP5 Family Members Evolve through Shuffling of LRRs

RPP5 encodes a predicted protein with 21 LRRs. These LRRs can be classified into eight groups (A to H) and consist of duplicated LRR-encoding segments (Parker et al., 1997). These duplications most likely result from successive unequal intragenic recombination events; those involving type B LRRs generated additional introns, which are contained in these gene segments. Experimental evidence for intragenic duplication within the LRR region was provided by plant FL-387 (Parker et al., 1997). This plant was identified in a fast neutron–mutagenized L*er* population and carries two unlinked mutations, one in *PAD4* (for *PHYTOALEXIN DEFICIENT 4*; necessary for camalexin synthesis; Zhou et al., 1998; J.E. Parker, unpublished data) and one in *RPP5.* This *RPP5-1* allele contains a 270-bp duplication encoding four LRRs, resulting in a predicted protein with 25 LRRs (Figure 4). Genetic separation from the *pad4* mutation showed that the *RPP5-1* allele fully retained its function in conferring resistance to *P. parasitica* Noco2.

The sequence analysis of the *RPP5* family members in Col-0 and L*er* suggests a model for *RPP5* homolog evolution. All homologs carry similar LRR duplications (or remnants thereof), as described for *RPP5.* However, *Col-D* and *Col-F* are the only members with 21 LRRs arranged identically to *RPP5.* Most members lack the duplication of four LRRs that generated exon 4 in *RPP5* (e.g., *Col-B* and *La-C*), and some members also lack the duplication of four LRRs that gave rise to exon 6 (e.g., *Col-C* and *La-B*) in *RPP5* and other members (Figure 3). Thus, the *RPP5* family can be interpreted as having evolved from ancestors with 13, 17, or 21 LRRs, which themselves probably were derived from a single *RPP5*-like gene that lacked the internal duplications and had only eight LRRs (Figure 4). Furthermore, many members carry additional and mostly unique LRR duplications and deletions. For example, *Col-B* and *Col-E* probably were derived from an *RPP5* ancestor with 17 LRRs, but they have both undergone subsequent independent rearrangements generating 23 and 14 LRRs, respectively (Figure 4). The different number of introns in the LRR regions correlates with the number of type B LRRs undergoing duplications or deletions.

Distance Analysis of Putative Structural and Signaling Domains within the *RPP5* Family

Comparison of the predicted proteins of the 12 nontruncated members (*RPP5*, *La-B*, *La-C*, *La-G*, *La-I*, and *Col-A* to *Col-G*) revealed considerable sequence conservation (Figures 5A

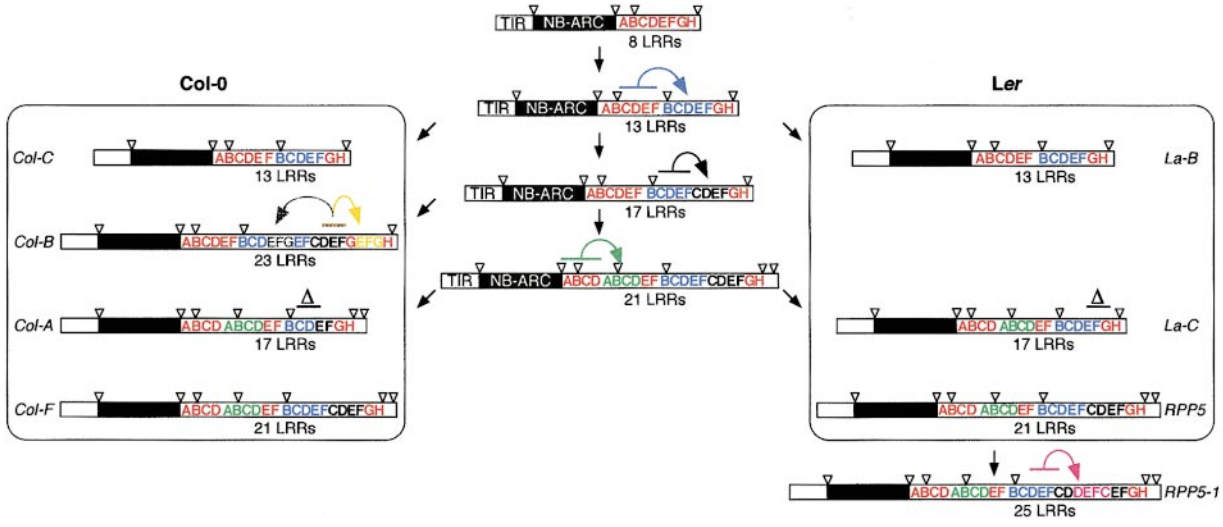

Figure 4. Model for the Evolutionary History of the *RPP5* Multigene Family.

The *RPP5* homologs are shown as rectangles, with the TIR domains (left, N-terminal) and LRR domains (right, C-terminal) as open rectangles, and the central NB-ARC domains as black rectangles. Open arrowheads indicate intron positions. All individual LRRs among the entire *RPP5* family can be classified into eight different groups (A to H). These LRRs are thought to be derived from an *RPP5* ancestor (top center) with eight progenitor LRRs (shown in red). A series of intragenic duplications may have led to an increase of these ancestral LRRs from eight to 13 LRRs (curved arrow and duplication in blue), to 17 LRRs (curved arrow and duplication in black), to 21 LRRs (curved arrow and duplication in green). The L*er RPP5* gene and the Col-0 homologs *Col-F* and *Col-D* have an identical LRR configuration with 21 LRRs. Most family members (most not shown), however, lack some of the ancestral duplications but carry other duplications and deletions in the LRR-encoded region; for example, *Col-B* carries two duplications (gray and yellow), and *Col-A* has undergone a deletion event (Δ). In L*er*, the functional *RPP5-1* allele with 25 LRRs was recovered (curved arrow and duplication in pink).

and 6). Most homologs have highly related exon 1–encoded TIR domains (Figure 5B), which is consistent with the TIR domain's predicted role as part of the effector portion of the protein (Whitham et al., 1994; Staskawicz et al., 1995). The exon 2–encoded NB-ARC (or Ap-ATPase) domain is homologous to the animal apoptosis proteins Apaf-1 and CED-4 and includes an NB site and several short conserved motifs with unknown function (Van der Biezen and Jones, 1998a; Aravind et al., 1999). Distance analysis shows extreme conservation of the predicted NB site (Figure 5C), which is consistent with the presence of a preserved ATP or GTP binding pocket in all NB-LRR proteins. The two similar members, *La-B* and *Col-C*, diverge significantly from the NB site consensus (Figure 5C). There appear to be three distinct domain classes within the second half of exon 2, designated the ARC domain (Figure 5D). One conserved ARC class is represented by *RPP5* and five other members. A second ARC class is represented by the only two Col-0 members, *Col-B* and *Col-F*, that have intact ORFs; this class also includes the diverged pair *La-B* and *Col-C*. A third ARC type is present in the two highly similar and diverged homologs, *Col-E* and *Col-G* (Figure 5D). The three ARC domain classes within the *RPP5* gene possibly reflect functional differences, such as interaction with different effector proteins.

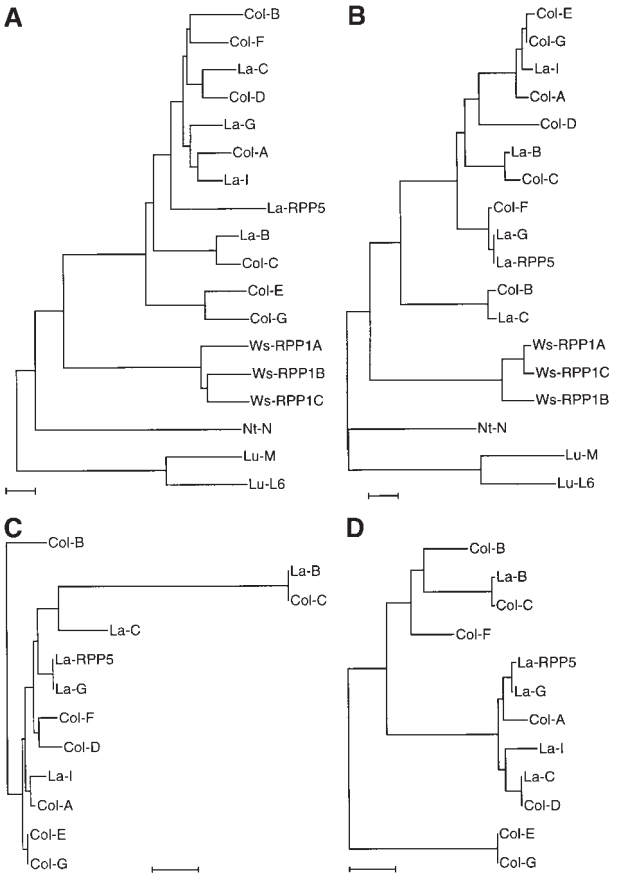

Figure 5. Dendrograms Showing Distance Relationships between Conceptual Protein Sequences of RPP5 Family Members from the L*er* and Col-0 Haplotypes.

Sequence distance trees were calculated using the neighbor-joining algorithm. Relative branch lengths (0.05) are indicated by bars below the trees. La, L*er*; Lu, *Linum usitatissimum*; Nt, *Nicotiana tabacum*; Ws, *Arabidopsis thaliana* land race Ws-0.

(A) Distance tree of entire gene products of the RPP5 family and other R proteins containing N-terminal TIR domains.

(B) Distance tree of TIR domains (residues 1 to 160 in RPP5) of the RPP5 family members and other R proteins.

(C) Distance tree of NB sites (residues 202 to 333 in RPP5) of the RPP5 family members.

(D) Distance tree of ARC domains (residues 334 to 518 in RPP5) of the RPP5 family members.

Solvent-Exposed Residues of LRRs Are Hypervariable

In RPP5, as in many NB-LRR proteins, the canonical LRRs consist of 23 or 24 residues, of which ~19 either are part of a predicted α-helical region connecting adjacent LRRs or are hydrophobic residues (mostly leucine residues) buried in the hydrophobic core (Jones and Jones, 1996). Each LRR also carries a short β-strand/β-turn region with the xxLxLxx motif in which the five interstitial residues (x) are predicted to be solvent exposed. Collectively, the β-strand/β-turn regions are predicted to form a parallel β sheet, which could create a surface for ligand interactions (Kobe and Deisenhofer, 1995). Comparisons between *RPP5* family members showed high conservation of most of the LRR residues, which conforms to the prediction that they serve a structural role. However, extensive amino acid variation occurs in the predicted solvent-exposed residues in the xxLxLxx motif. For example, comparison of RPP5 with Col-F (both of which possess 21 LRRs) showed that 49% (50 of 103) of these LRR residues are different, whereas only 11% (41 of 377) of different residues are present in the remainder of the LRRs. We then determined which residues within the entire family are hypervariable (greater than or equal to four different residues among the compared members; Figure 6). This analysis revealed that within the *RPP5* family, 23% (24 of 103) of the predicted solvent-exposed LRR residues are hypervariable as opposed to 0.8% (3 of 377) in the remaining structural LRR region. In several R proteins, hypervariability of solvent-exposed residues correlates with differential pathogen recognition (Parniske et al., 1997; Thomas et al., 1997; Botella et al., 1998; Dixon et al., 1998; Meyers et al., 1998b).

Positive Selection of Ligand-Interacting Residues in the LRRs

To examine further the idea that *RPP5* family members function or functioned in differential pathogen recognition, we

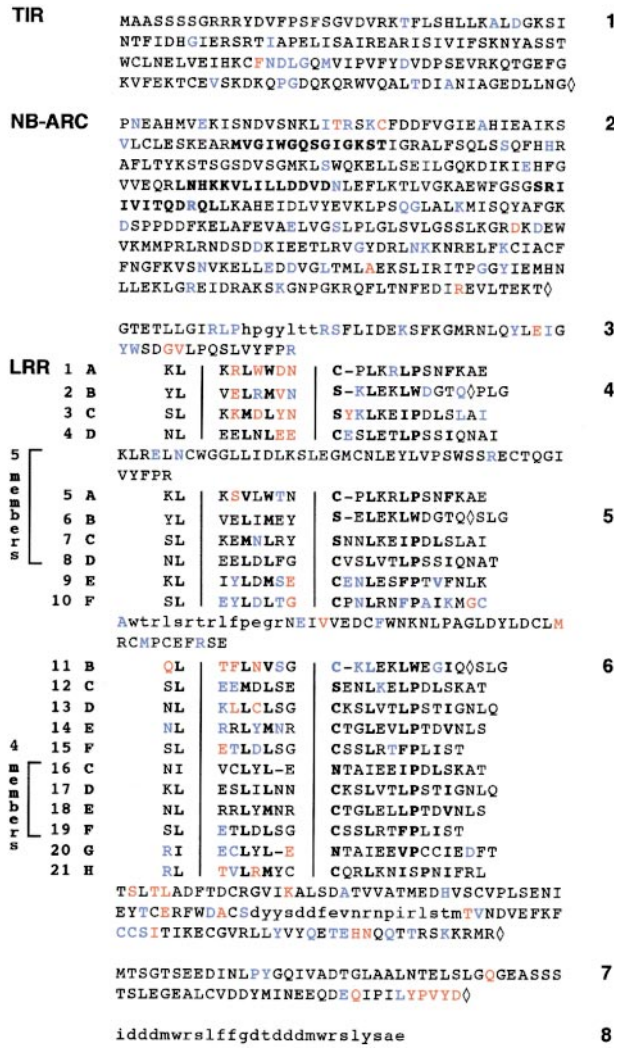

Figure 6. Conserved and Hypervariable Amino Acids among Members of the *RPP5* Multigene Family.

The RPP5 protein sequence and its predicted TIR, NB-ARC, and LRR structures are shown as given in Parker et al. (1997). Intron–exon boundaries are indicated by diamonds, and exon numbers are indicated at the right. Predicted proteins of 12 full-length family members are aligned, including those of *RPP5*, *La-B*, *La-C*, *La-G*, *La-I*, and *Col-A* to *Col-G.* Residues shown in lowercase letters either are part of highly variable regions and cannot be aligned or, like the residues that constitute exon 8, are absent from most members. Conserved amino acids (black) are defined as at most one different residue among the compared members. Hypervariable amino acids (red) are defined as four or more different types of residues among the compared members. Other amino acids are moderately variable (two or three different types of residues among the compared members) and are shown in blue. Because LRR-encoded regions were missing from some of the members, limited sequence comparison of LRRs 5 to 8 (five members) and LRRs 16 to 19 (four members) was possible. As a consequence, the hypervariable residues in these LRRs are possibly underestimated. Conserved kinase motifs that constitute an NB pocket and conserved hydrophobic residues within the LRRs are shown in boldface letters. Predicted solvent-exposed residues (x in the xxLxLxx motif) are shown between vertical lines.

subjected predicted functional domains to pairwise analysis for substitutions at synonymous (K_s) and nonsynonymous (K_a; Table 1) sites. Synonymous substitutions have no apparent selective advantage or disadvantage; therefore, the K_s values give an estimate for the background (random) substitution rate. However, if in a certain region more nonsynonymous than synonymous substitutions are found (i.e., $K_a/K_s > 1$), it provides solid evidence that these changes have been positively selected for (Parniske et al., 1997; Hughes and Yeager, 1998; Meyers et al., 1998b).

The exon 1–encoded TIR domains show low K_a/K_s ratios (on average $K_a/K_s = 0.6$), which is consistent with conservation of a proposed effector function. Low K_a/K_s ratios also were calculated for the regions encoding the NB site (on average, $K_a/K_s = 0.6$). The only significant deviation from this average is observed in the two members (*La-B* and *Col-C*) that possess more than one ORF-disrupting mutation. These homologs have a K_a/K_s ratio of 1.1, indicating that no selection pressure is acting on these members and classifying *La-B* and *Col-C* as pseudogenes. Consistent with this finding is the observation that these two highly similar genes diverge significantly from the rest of the family (see Distance Analysis of Putative Structural and Signaling Domains within the *RPP5* Family; Figure 5). Therefore, these pseudogenes serve as an internal control for the K_a/K_s analysis and reinforce the suggestion that the other members have been functional in the recent evolutionary past. Low K_a/K_s ratios also are observed for the LRR residues predicted to serve a structural role (on average $K_a/K_s = 0.6$; Table 1). The low K_a/K_s values calculated for the TIR domain, the NB site, and the structural LRR residues are consistent with the predicted structural and functional constraints of these domains and indicate that purifying selection operated such that the majority of amino acid substitutions are not tolerated.

In the region encoding the hypervariable solvent-exposed LRR residues, the K_a rate substantially exceeds the K_s rate, resulting in K_a/K_s ratios that average 2.8 (omitting the pseudogenes *La-B* and *Col-C*; Table 1). The generally high K_a/K_s ratios in this region suggest that diversifying selection acted on the predicted ligand-interacting LRR residues. Similar conclusions were drawn from K_a/K_s analyses of other *R* genes (Parniske et al., 1997; Wang et al., 1998; Botella et al., 1998; McDowell et al., 1998; Meyers et al., 1998b).

Stability of the L*er RPP5* Haplotype

To investigate the stability of the L*er RPP5* haplotype and the basis for the mutations in the L*er* homologs, we analyzed the La-0 ecotype, the progenitor of L*er* used in this

Table 1. Pairwise K_a/K_s Ratios in the Predicted Solvent-Exposed LRR Residues and Structural LRR Residues among the *RPP5* Family[a]

	RPP5	*La-B*	*La-C*	*La-G*	*La-I*	*Col-A*	*Col-B*	*Col-C*	*Col-D*	*Col-E*[b]	*Col-F*	*Col-G*[b]
RPP5	×	0.9	1.3	2.8	3.0	1.5	2.5	1.8	2.0	—	2.5	—
La-B	0.8	×	0.7	0.9	0.6	1.0	1.0	0.5	0.8	—	0.5	—
La-C	0.7	0.6	×	1.4	1.3	2.8	1.1	1.0	0.9	—	1.0	—
La-G	0.6	0.8	0.4	×	1.7	2.5	3.0	1.3	5.7	—	2.5	—
La-I	0.9	0.6	0.6	0.4	×	4.2	4.4	0.9	5.8	—	2.2	—
Col-A	0.6	0.6	0.6	0.3	0.8	×	4.7	1.8	6.6	—	4.1	—
Col-B	0.7	0.6	0.6	0.4	0.6	0.6	×	1.4	3.2	—	1.8	—
Col-C	1.1	0.5	0.7	0.7	0.8	0.8	0.8	×	1.1	—	1.2	—
Col-D	0.9	0.6	0.6	0.4	0.5	0.7	0.4	0.7	×	—	1.9	—
Col-E[b]	—	—	—	—	—	—	—	—	—	×	—	1.7
Col-F	0.9	0.5	0.7	0.4	0.6	0.8	0.4	0.6	1.4	—	×	—
Col-G[b]	—	—	—	—	—	—	—	—	—	0.5	—	×

[a] Solvent-exposed residues are shown above the diagonal, and structural residues are shown below the diagonal.
[b] *Col-E* and *Col-G* are highly similar but significantly diverge in their LRR-encoded DNA sequences from the rest of the family; as a consequence, the K_a/K_s analysis did not produce meaningful results (denoted by dashes).

study. DNA gel blots showed identical *RPP5*-hybridizing fragments in La-0 as is in L*er* (data not shown). All polymerase chain reaction (PCR) primers designed to detect the retroelement insertions unique to the L*er* haplotype amplified an identical product from the wild-type La-0 progenitor. In addition, the sequences of PCR-amplified fragments from the L*er* homologs show point mutations in La-0 identical to those in L*er* (data not shown). This analysis indicates that the loss of function of the L*er* members is not an artifact of recent domestication through propagation in greenhouses in the absence of selection for *P. parasitica* resistance.

To analyze further the genetic stability of the *RPP5* haplotype, we made an outcross of male sterile L*er* to Col-0 and tested the progeny for susceptibility to *P. parasitica* Noco2. Among ~7500 heterozygous progeny, no susceptible individuals were recovered, again suggesting that this locus is not highly meiotically unstable when homozygous. This experiment would have detected high instability (as at the maize *Rp1-G* disease resistance haplotype) but not low instability (as at the maize *Rp1-D* haplotype; Hulbert, 1997).

DISCUSSION

Intraspecific Haplotype Divergence at a Complex Disease Resistance Locus

Many different *R* genes have been isolated (Hammond-Kosack and Jones, 1997), and several *R* gene haplotypes have been compared. Interspecific comparisons of complex *R* gene haplotypes have been reported (Parniske et al., 1997; Thomas et al., 1997, 1998; Dixon et al., 1998), and intraspecific comparisons have been conducted with simple loci that contain only one *R* gene homolog (Grant et al., 1998; McDowell et al., 1998; Caicedo et al., 1999; Ellis et al., 1999; Henk et al., 1999). The unique availability of the complete DNA sequence of the L*er* and Col-0 *RPP5* haplotypes permits analysis of intraspecific variation at a complex *R* gene locus and a comparison of *R* gene locus polymorphism to the level of polymorphism at other loci.

The clearest conclusion from the L*er* and Col-0 *RPP5* haplotype comparisons is that this locus exhibits an extraordinary degree of intraspecific polymorphism. Variation was observed in the number and relative location of homologs, in the presence and location of retroelements, and in other DNA sequences (such as the 5S rDNA sequences) as well as in numerous point mutations. This pronounced lack of synteny over ~75 kb of sequence may have been the source of the significant recombination suppression noted in this region during the map-based cloning of *RPP5* (Parker et al., 1997). The meiotic stability of the homozygous L*er* *RPP5* haplotype is consistent with the results of crossing plants expressing *Cf9* to those expressing *Cf0* to create 20,000 testcross progeny in which no spontaneous disease-sensitive recombinants were recovered (Parniske et al., 1997). However, it is in contrast with the results of crossing several *Cf* gene transheterozygotes to disease-sensitive lines; in these experiments, recombination events were detected within the locus (Parniske et al., 1997) and between homologs (Dixon et al., 1998). These data suggest that when a complex haplotype is homozygous, the different homologs rarely mutate or exchange sequence information. This appears paradoxical, because *R* gene loci have been anticipated to be sites of frequent recombination that could give rise to new recognition specificities. However, the rate of recombination between *R* haplotypes can depend on the specific haplotype combination (Hulbert, 1997). The idea that some haplotypes pair better than others would have considerable consequences for the rate of sequence exchange

within and between *R* loci. In Arabidopsis, which reproduces predominantly by self-fertilization, haplotype mispairing is likely to be rare. Occasional outcrosses with genotypes carrying distinct haplotypes would create greater opportunity for mispairing to create evolutionary novelty (Hulbert, 1997; Parniske et al., 1997).

Mechanisms of DNA Sequence Divergence in the Col-0 and L*er RPP5* Haplotypes

We propose that the progenitor of the *RPP5* gene family contained only eight LRRs (Figure 4) and was flanked by a protein kinase gene. Gene duplication generated at least five *RPP5* homologs before the separation of the L*er* and Col-0 ecotypes. A series of intragenic duplications within the region encoding the LRRs presumably produced *RPP5* progenitors with 13, 17, and 21 LRRs. This progenitor *RPP5* multigene family diverged into dissimilar gene clusters with eight members in Col-0 and 10 in L*er*, which were interspersed with fragments from the ancestral protein kinase gene. In both haplotypes, members with 21 LRRs are present. However, most members lack one or more ancestral duplications. Furthermore, some members carry additional and mostly unique deletions and duplications within the region encoding the LRRs.

Both haplotypes carry an identical truncated copy, which is in inverse orientation relative to the other homologs. This curious structure could be a relic of the original duplication event, perhaps involving aberrant behavior of the DNA replication fork. In contrast to most of the Arabidopsis genome, the *RPP5* haplotypes have undergone gross structural rearrangements and accumulated numerous point mutations. This probably indicates that many of the *RPP5* homologs are under reduced selection for conservation or retention of function. An interesting corollary of this interpretation is that DNA sequence changes at the *RPP5* locus might reflect the mutagenic mechanisms that are constantly at work in the remainder of the genome, where their consequences are more rapidly purged by selection.

The *RPP5* multigene family is composed of a limited number of homologous but distinct sequence segments distinguished by runs of IPSs. The shared and exchanged gene segments most likely result from unequal recombination and/or gene conversion. Meiotic misalignment involving different homologs followed by intergenic or intragenic crossing-over would result in loss of one or more members on one chromosome and duplication of these members on the other. Intragenic crossing-over, in addition, would result in exchanges of gene segments. Unequal recombination also could involve mispairing of LRR-encoding regions followed by an intragenic crossing-over. Homologs with more than eight LRRs carry several duplications, which increase the probability of intragenic misalignments. Such events would generate additional duplications or deletions and probably account for the variable number of LRRs. Unequal intragenic recombination presumably gave rise to a novel *RPP5-1* allele with an internal duplication of four LRRs (Parker et al., 1997).

The predicted solvent-exposed LRR residues are hypervariable in the *RPP5* family (and other R proteins). Such changes are more likely to have arisen from point mutations than from sequence exchanges, and based on K_a/K_s ratios, this hypervariability is due to diversifying selection. Thus, although sequence exchange and LRR shuffling greatly contributed to diversification of the gene family members, point mutations in codons for the predicted solvent-exposed residues also must play a crucial role in generating novel recognition specificities. DNA shuffling enhances the potential of a point mutation to generate a useful recognition capacity, because it can be combined with other variants to create an allelic series of distinct, parallel β-sheet recognition surfaces.

"Birth and Death" of *RPP5* and Other Plant *R* Genes

It has been suggested that *R* genes at complex haplotypes evolve by mechanisms similar to those at work in the vertebrate MHC (Michelmore and Meyers, 1998). These multigene families are proposed to evolve through a birth and death process rather than through concerted evolution (Nei et al., 1997), explaining why, when different MHC haplotypes are compared, orthologs are usually more similar than paralogs.

There are more complete data available for haplotype comparisons at the MHC than for *R* gene loci. Comparisons of haplotypes at two complex *Cf* loci suggested significant sequence exchange between paralogs (Parniske et al., 1997; Thomas et al., 1997, 1998; Dixon et al., 1998). Many other *R* genes are also members of clustered multigene families (Whitham et al., 1994; Anderson et al., 1997; Song et al., 1997; Botella et al., 1998; Meyers et al., 1998a, 1998b; Simons et al., 1998), but complete comparisons of corresponding haplotypes in related species or ecotypes have not been reported.

At the Col-0 and L*er RPP5* haplotypes, some family members are more closely related to orthologs in the other ecotype than to paralogs. These include *La-B*/*Col-C*, *La-J*/*Col-H*, and possibly also *RPP5*/*Col-A* (Figures 1 and 3). The latter two pairs of orthologs (*RPP5*/*Col-A* and *La-J*/*Col-H*) reside at the proximal and distal borders of the cluster, respectively. However, orthologous relationships are difficult to discern between members located in the middle of the cluster. As noted at other loci (Michelmore and Meyers, 1998), if sequence exchange between paralogs were frequent, it would not be possible to detect orthologs when two haplotypes are compared. At the *RPP5* haplotypes, sequence exchange has been sufficiently frequent to obscure orthology, especially in the middle of the cluster, but has not lead to homogenization of the gene clusters. However, orthologous relationships are easier to discern when subdomains are compared (Figure 3; Meyers et al., 1998a). Given

the pronounced polymorphism at *R* gene loci in natural populations and the enhanced frequency of mispairing when haplotypes are combined in *trans*, it is probably unhelpful to maintain too rigid a distinction between paralogs and orthologs, because unequal crossovers could turn one into the other.

Selective Pressures on *RPP5* Haplotypes and Gene Family Members

The TIR domains, NB sites, and structural LRR residues within the *RPP5* family are highly conserved and show evidence for purifying selection. In contrast, the predicted ligand-interacting LRR residues are hypervariable and, based on their K_a/K_s ratios (Table 1), are subject to diversifying selection. The evidence for diversifying and purifying selection strongly suggests that even the apparently mutant *RPP5* homologs are and/or were functional.

Three distinct classes of ARC domains are present within the *RPP5* family. Interestingly, a similar divergence in the ARC domain has been observed in the L*er RPP8* gene and its homolog in Col-0 (McDowell et al., 1998). This domain may play a role in intramolecular signaling (Van der Biezen and Jones, 1998b) and therefore may coevolve with the diverging LRR domains. In the apoptosis proteins Apaf-1 and CED-4, the NB-ARC (or Ap-ATPase) domain functions as a conformation-dependent protein–protein interaction module (Srinivasula et al., 1998; Yang et al., 1998). The diverged ARC domains in R proteins, therefore, may indicate interactions with distinct proteins.

In addition to *RPP5* in L*er*, two family members in Col-0 have intact ORFs, and three have nearly full-length ORFs. Most of the other 12 family members either are truncated (five homologs) or encode ORFs that are disrupted at the 5′ end (seven homologs). However, eight members have ORFs that are only disrupted by a single mutational event (point mutation or retroelement insertion) and have not accumulated additional apparent deleterious mutations, suggesting that these genes have been functional in the recent evolutionary past. The five Col-0 homologs with intact or nearly full-length ORFs also may be functional *R* genes. If so, one of these Col-0 homologs could be the Col-0 *RPP4* gene that recognizes the *P. parasitica* strain Emoy2 (Tör et al., 1994). Apparently, nonfunctional *R* gene homologs also might serve a repository function and allow mutation, unequal recombination, and gene conversion to generate novel *R* gene variants upon which diversifying selection can operate.

Disease resistance has been associated with a fitness "cost" (Crute, 1994; Bergelson and Purrington, 1996; Simms, 1996). In the (temporary) absence of pathogen pressure, individual plants with mutated *R* genes therefore may have selective advantages. The high proportion of mutant *RPP5* family members supports the idea that some *R* genes can impose an evolutionary cost. Such costs might arise if the *R* gene confers some constitutive activity or if it weakly recognizes an endogenous ligand. Single *R* genes also can have a high proportion of mutant alleles (Grant et al., 1998; Caicedo et al., 1999; Henk et al., 1999; Stahl et al., 1999).

Maintenance of Extreme Polymorphism at *R* Gene Loci

Because pathogens such as *P. parasitica* can readily overcome a specific *R* gene by loss or mutation of the corresponding *Avr* gene (Holub and Beynon, 1996), it is likely that any pathogenicity function conferred by such *Avr* genes is redundant. Pathogens can escape recognition through loss of function of *Avr* genes, whereas novel *R* genes in plants evolve through presumably rarer gain-of-function mutations. In natural populations, if evolution to virulence exceeds evolution to resistance, how do plants tolerate pathogen pressure?

At least two theories have been proposed to explain the extreme polymorphism at the MHC, which, like complex *R* gene haplotypes, encodes multiple proteins that have the capacity to detect multiple nonhost recognition determinants (Hughes and Yeager, 1998). One theory maintains that heterozygote advantage (or overdominance) is key and that high levels of polymorphism are essential to ensure that most individuals are heterozygous. The alternative interpretation is frequency-dependent selection in which any advantageous MHC allele could be overcome by pathogens if it were at too high a frequency in the population; if an MHC allele is rare, there is little selective pressure on the pathogen to overcome it. In Arabidopsis, an inbreeder, the heterozygote advantage or overdominance theory is ruled out because most individuals are homozygous. It is interesting to speculate, however, that in agricultural systems, in which there is potential for one pathogen strain to cause an epidemic in a genetically uniform crop, overdominance at *R* gene loci (which are distributed throughout the genome) might contribute to F_1 hybrid vigor. Frequency-dependent selection due to minority advantage provides a simple explanation for the high level of natural intraspecific polymorphism between haplotypes at the *RPP5* locus and perhaps at other complex *R* gene loci. Recently, Stahl et al. (1999) also inferred frequency-dependent selection to explain polymorphism at the Arabidopsis *RPM1* locus.

METHODS

Construction of a Physical Cosmid Contig of the Entire Landsberg *erecta RPP5* Haplotype

Screening of an *Arabidopsis thaliana* ecotype Landsberg *erecta* (L*er*) genomic library with the C18 probe resulted in the isolation of seven cosmids, including pCLD29L17 containing the *RPP5* gene (Parker et al., 1997). To obtain additional cosmids, we screened the *abi3* yeast artificial chromosome (YAC) library by hybridization with the *RPP5* gene. This identified the YAC clone *abi3*-8D3 (180 kb) containing all

RPP5-hybridizing fragments observed in gel blots with genomic L*er* DNA (Parker et al., 1997). DNA of the *abi3*-8D3 clone was partially digested with Sau3A and ligated in the SLJ755I5 cosmid vector (www.uea.ac.uk/nrp/jic/s3d_plas.htm). Four thousand clones were screened by hybridization with *RPP5*, and 30 different cosmid clones were identified. These were assembled into physical contigs by using inverse polymerase chain reaction (PCR) products and direct end-sequencing, restriction digest, DNA fingerprinting, and DNA gel blot hybridizations with *RPP5*-derived sequences. Finally, a contig of eight cosmids covering ~110 kb was established.

DNA Sequencing

Sequencing templates were prepared from random-sheared DNA generated by sonication, repaired with T4 DNA polymerase (Pharmacia), and cloned into dephosphorylated SmaI-digested pUC18 plasmids (Pharmacia). The double-stranded plasmid DNA was sequenced using the PRISM Ready Reaction Terminator Cycle Sequencing System (Perkin-Elmer). The reaction products were separated on an Applied Biosystems 377 DNA sequencer (Foster City, CA), and the data were assembled using UNIX versions of the Staden package (R. Staden, Cambridge, UK). Single-stranded gaps were completed using specifically designed primers. Deletions in clones were generated to establish or confirm putative contigs. Double-stranded sequencing with an average of fourfold redundancy was achieved.

Sequence Analysis

Sequences of six cosmids were assembled into a contig of 92 kb covering the entire L*er RPP5* haplotype. The *RPP5* family member *La-J* was amplified by PCR from the L*er* cosmid *abi3*-8D3-21 by using primers based on the corresponding ecotype Columbia (Col-0) homolog *Col-H.* The Col-0 haplotype sequence was obtained from a 90.5-kb bacterial artificial chromosome clone IGF3D5 (Bevan et al., 1998) and the two cosmids cc33M3 and cc16N19. Analysis and alignments of assembled sequences were performed using Genetics Computer Group (Madison, WI) programs. Alignments were optimized manually. Putative splice sites were based on the sequence of *RPP5* and *Col-F* cDNA and reverse transcription–PCR products and confirmed by the NetPlantGene program (www.cbs.dtu.dk/NetPlantGene.html). Informative polymorphic sites (IPSs) were displayed using the Sequence Output program (B.G. Spratt, University of Sussex, Brighton, UK). Distance trees were calculated using the neighbor-joining algorithm from the Clustal X package (Thompson et al., 1997). Nonsynonymous (K_a) and synonymous (K_s) substitution ratios were calculated with NewDiverge (Genetics Computer Group) for each pair of sequences.

Nucleic Acid Manipulations

All DNA and RNA manipulations, including YAC and PCR analyses, were performed essentially as described by Parker et al. (1997). The autoradiogram of the RNA gel blot shown in Figure 2 was made from ~2 μg of poly(A)$^+$ RNA, which was size-fractionated through a vertical 1.4% agarose gel and blotted onto a Hybond N filter (Amersham), hybridized with a radiolabeled sequence encoding exons 1 and 2 of *Col-F* (NcoI-NsiI fragment, base pairs 1 to 1712) for 16 hr, stringently washed with 0.1 × SSC (1 × SSC is 0.15 M NaCl and 0.015 M sodium citrate), 0.1% SDS at 65°C, and exposed to Kodak X-Omat AR film for 2 days with intensifying screens at −70°C.

Plant and Pathogen Handling

Arabidopsis growth and *Peronospora parasitica* infections were as described previously (Parker et al., 1997).

Accession Numbers

The EMBL nucleotide sequence database accession numbers of the L*er RPP5* gene cluster (*La-A* to *La-I*) and the *La-J* homolog are AF180942 and AF180943, respectively.

ACKNOWLEDGMENTS

We thank David Baker and Patrick Bovill (Sainsbury Laboratory) for operating the sequencing equipment and Alan Cavill (Sainsbury Laboratory) for supportive management. Roger Innes (Indiana University, Bloomington) is thanked for critically reading the manuscript. We are grateful to Ian Bancroft and Mike Bevan (John Innes Centre) for supplying Col-0 cosmid libraries and bacterial artificial chromosome clones, and to Clare Lister and Caroline Dean (John Innes Centre) for providing the L*er* YAC library. Maarten Koornneef (Agricultural University, Wageningen, The Netherlands) is thanked for the gift of La-0 seeds. The determination and analysis of the sequence of the Col-0 *RPP5* haplotype were conducted under the European Scientists Sequencing Arabidopsis program sponsored by the European Community. E.A.v.d.B. and M.P. were supported by fellowships from the European Community. The Sainsbury Laboratory is supported by the Gatsby Charitable Foundation.

Received July 16, 1999; accepted September 13, 1999.

REFERENCES

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

Aravind, L., Dixit, V.M., and Koonin, E.V. (1999). The domains of death: Evolution of the apoptosis machinery. Trends Biochem. Sci. **24,** 47–53.

Bennetzen, J.L. (1996). The contributions of retroelements to plant genome organization, function and evolution. Trends Microbiol. **4,** 347–353.

Bergelson, J., and Purrington, C.B. (1996). Surveying patterns in the cost of resistance in plants. Am. Nat. **148,** 536–558.

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

Bevan, M., et al. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana.* Nature **391,** 485–488.

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.

Caicedo, A.L., Schaal, B.A., and Kunkel, B.N. (1999). Diversity and molecular evolution of the *RPS2* resistance gene in *Arabidopsis thaliana.* Proc. Natl. Acad. Sci. USA **96,** 302–306.

Collmer, A. (1998). Determinants of pathogenicity and avirulence in plant pathogenic bacteria. Curr. Opin. Plant Biol. **1,** 329–335.

Crute, I.R. (1994). Gene-for-gene recognition in plant–pathogen interactions. Philos. Trans. R. Soc. Lond. Biol. Sci. **346,** 345–349.

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

Dixon, M.S., Hatzixanthis, K., Jones, D.A., Harrison, K., and Jones, J.D.G. (1998). The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell **10,** 1915–1925.

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.

Ellis, J.G., Lawrence, G.J., Luck, J.E., and Dodds, P.N. (1999). The identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. Plant Cell **11,** 495–506.

Grant, M.R., McDowell, J.M., Sharpes, A.G., De Torres Zabala, M., Lydiate, D.J., and Dangl, J.L. (1998). Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis.* Proc. Natl. Acad. Sci. USA **95,** 15843–15848.

Hammond-Kosack, K.E., and Jones, J.D.G. (1997). Plant disease resistance genes. Annu. Rev. Plant Physiol. Plant Mol. Biol. **48,** 575–607.

Henk, A.D., Warren, R.F., and Innes, R.W. (1999). A new *Ac*-like transposon of *Arabidopsis* is associated with a deletion of the *RPS5* disease resistance gene. Genetics **151,** 1581–1589.

Holub, E.B., and Beynon, J.L. (1996). Symbiology of mouse ear cress (*Arabidopsis thaliana*) and oomycetes. Adv. Bot. Res. **24,** 227–273.

Hughes, A.L., and Yeager, M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. Annu. Rev. Genet. **32,** 415–435.

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 role of leucine-rich repeat proteins in plant defences. Adv. Bot. Res. **24,** 89–167.

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

Kobe, B., and Deisenhofer, J. (1995). A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature **374,** 384–386.

McDowell, J.M., Dhandaydham, M., Long, T.A., Aarts, M.G.M., Goff, S., Holub, E.B., and Dangl, J.L. (1998). Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of Arabidopsis. Plant Cell **10,** 1861–1887.

Meyers, B.C., Chin, D.B., Shen, K.A., Sivaramakrishnan, S., Lavelle, D.O., Zhang, Z., and Michelmore, R.W. (1998a). The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell **10,** 1817–1832.

Meyers, B.C., Shen, K.A., Rohani, P., Gaut, B.S., and Michelmore, R.W. (1998b). Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell **10,** 1833–1846.

Michelmore, R.W., and Meyers, B.C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res. **8,** 1113–1130.

Nei, M., Gu, X., and Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc. Natl. Acad. Sci. USA **94,** 7799–7806.

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.

Simms, E.L. (1996). The evolutionary genetics of plant–pathogen systems. Bioscience **46,** 136–145.

Simons, G., et al. (1998). Dissection of the Fusarium *I2* gene cluster in tomato reveals six homologs and one active gene copy. Plant Cell **10,** 1055–1068.

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.

Srinivasula, S.M., Ahmad, M., Fernandes-Alnemri, T., and Alnemri, E.S. (1998). Autoactivation of procaspase-9 by Apaf-1–mediated oligomerization. Mol. Cell **1,** 949–957.

Stahl, E.A., Dwyer, G., Maurico, R., Kreitman, M., and Bergelson, J. (1999). Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis.* Nature **400,** 667–671.

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.

Thomas, C.M., Jones, D.A., Parniske, M., Harrison, K., Balint-Kurti, P.J., Hatzixanthis, K., and Jones, J.D.G. (1997). Characterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in Cf-4 and Cf-9. Plant Cell **9,** 2209–2224.

Thomas, C.M., Dixon, M.S., Parniske, M., Golstein, C., and Jones, J.D.G. (1998). Genetic and molecular analysis of tomato *Cf* genes for resistance to *Cladosporium fulvum.* Philos. Trans. R. Soc. Lond. Biol. Sci. **353,** 1413–1424.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgens, D.G. (1997). The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25,** 4876–4882.

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.

Van der Biezen, E.A., and Jones, J.D.G. (1998a). The NB-ARC domain: A novel signaling motif shared by plant resistance gene products and regulators of cell death in animals. Curr. Biol. **8,** R226–R227.

Van der Biezen, E.A., and Jones, J.D.G. (1998b). Plant disease resistance proteins and the gene-for-gene concept. Trends Biochem. Sci. **23,** 454–456.

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.

Yang, X., Chang, H.Y., and Baltimore, D. (1998). Essential role of CED-4 oligomerization in CED-3 activation and apoptosis. Science **281,** 1355–1357.

Zhou, N., Tootle, T.L., Tsui, F., Klessig, D.F., and Glazebrook, J. (1998). *PAD4* functions upstream from salicylic acid to control defense responses in Arabidopsis. Plant Cell **10,** 1021–1030.

Pronounced Intraspecific Haplotype Divergence at the *RPP5* Complex Disease Resistance Locus of Arabidopsis

Laurent Noël, Tracey L. Moores, Erik A. van der Biezen, Martin Parniske, Michael J. Daniels, Jane E. Parker and Jonathan D. G. Jones

Plant Cell 1999;11;2099-2111
DOI 10.1105/tpc.11.11.2099