The Maintenance of Extreme Amino Acid Diversity at the Disease Resistance Gene, *RPP13*, in *Arabidopsis thaliana*

Laura E. Rose,*,[1] Peter D. Bittner-Eddy,†,[1] Charles H. Langley,* Eric B. Holub,† Richard W. Michelmore‡ and Jim L. Beynon†,[2]

**Center for Population Biology and* ‡*Department of Vegetable Crops, University of California, Davis, California 95616 and* †*Horticulture Research International, Wellesbourne, Warwick CV35 9EF, United Kingdom*

Manuscript received May 22, 2003
Accepted for publication December 17, 2003

ABSTRACT

We have used the naturally occurring plant-parasite system of *Arabidopsis thaliana* and its common parasite *Peronospora parasitica* (downy mildew) to study the evolution of resistance specificity in the host population. DNA sequence of the resistance gene, *RPP13*, from 24 accessions, including 20 from the United Kingdom, revealed amino acid sequence diversity higher than that of any protein coding gene reported so far in *A. thaliana*. A significant excess of amino acid polymorphism segregating within this species is localized within the leucine-rich repeat (LRR) domain of *RPP13*. These results indicate that single alleles of the gene have not swept through the population, but instead, a diverse collection of alleles have been maintained. Transgenic complementation experiments demonstrate functional differences among alleles in their resistance to various pathogen isolates, suggesting that the extreme amino acid polymorphism in *RPP13* is maintained through continual reciprocal selection between host and pathogen.

ALTHOUGH resistance (*R*) genes can be abundant and highly variable within a given plant species, little is known about the origin and maintenance of variation of *R* genes in natural plant populations (reviewed in Michelmore and Meyers 1998; Hulbert *et al.* 2001). Haldane (1949) hypothesized that coevolution between host and parasite could lead to the maintenance of variation within both organisms. Genetic models incorporating either overdominance (heterozygote advantage) or negative frequency-dependent selection (in which the fitness of an individual is negatively correlated with the frequency of its genotype within the population) corroborate Haldane's assertion that such dynamics could lead to either stable balanced polymorphisms or cycling of alleles in the host and parasite populations (reviewed in May and Anderson 1983). Alternatively, strong positive selection may lead to the fixation of a single allelic type at an *R*-gene locus within a host population. Depending on the strength of selection, the geographic distribution of the pathogen, and the host demography, the same allele may sweep to fixation in a number of host populations. When pathogens are variable at a geographic or temporal scale, polymorphism may be maintained at the species level, even in the face of strong directional selection within host populations. Analyses of sequence variation at *R*-gene loci in natural plant populations are necessary to distinguish between these different evolutionary scenarios.

Studies of allelic polymorphism at *R*-gene loci in *Arabidopsis thaliana* have been limited to a few previous examples. At both the *RPM1* and the *RPS5* loci, functional alleles and a null allele segregate between populations of *A. thaliana* (Grant *et al.* 1995, 1998; Tian *et al.* 2002). While the ratio of intraspecific nonsynonymous polymorphism to synonymous polymorphism within the coding regions of these two genes is much less than one (Bergelson *et al.* 2001), the linked intergenic regions of the resistant and susceptible (null-allele possessing) individuals show a large number of fixed differences. The authors interpreted this as evidence for the long-term maintenance of the resistant and susceptible haplotypes (Stahl *et al.* 1999; Tian *et al.* 2002).

The level of nucleotide polymorphism has also been determined at the *RPS2* *R*-gene locus in *A. thaliana* (Caicedo *et al.* 1999; Mauricio *et al.* 2003). Similar to the *RPM1* and *RPS5* loci, two different haplotype classes were found at the *RPS2* locus. Alleles of one haplotype were present in plants that were susceptible or only mildly resistant to *Pseudomonas syringae* pv. *tomato* (*Pst*) expressing *avrRpt2* (Mauricio *et al.* 2003). Alleles of the other haplotype class were present in resistant and partially resistant individuals as well as a susceptible lab-induced mutant and one susceptible plant with a nonsense mutation in the *RPS2* gene. Thus, the haplotype differentiation corresponded roughly with the observed phenotypic variation in the plants. These studies are

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY487208–AY487228 and AY487230–AY487236.

[1]These authors contributed equally to this work.

[2]*Corresponding author:* Horticulture Research International, Wellesbourne, Warwick CV35 9EF, United Kingdom.
E-mail: jim.beynon@hri.ac.uk.

consistent with the maintenance of variation between accessions or populations in self-compatible species, most likely through some form of negative frequency-dependent selection or geographic differentiation in pathogen selective pressures.

The *RPP13* gene in *A. thaliana* controls resistance to the oomycete pathogen, *Peronospora parasitica*, and encodes a protein containing a coiled-coil domain, a nucleotide-binding site (NBS), and a leucine-rich repeat region (LRR; Bittner-Eddy *et al.* 1999, 2000). *RPP13* is unique among characterized NBS:LRR *R* genes in that it retains full function in rar1, ndr1, eds1, pad4, npr1, and double eds1, ndr1 mutant plants (Bittner-Eddy and Beynon 2001). In contrast to the other *R*-gene loci that have been the focus of population genetic analyses, multiple recognition specificities to different pathogen genotypes are encoded by differentiated alleles at this single locus (Bittner-Eddy *et al.* 2000; P. Bittner-Eddy, unpublished data). This resistance gene is located on the bottom arm of chromosome 3 (At3g46530) and two *RPP13* paralogs have been identified in the Arabidopsis Col-0 genome. These paralogs are located on the third chromosome ~73 kb from the *RPP13* gene. They share 65 and 60% amino acid identity to the *RPP13* allele from the Columbia ecotype (GenBank accession nos. At3g46730 and At3g46710). The functions of these two distantly related paralogs are unknown.

In this study, we investigate 24 accessions of *A. thaliana* collected from 20 populations in the United Kingdom and 4 populations from elsewhere in northern Europe to determine whether the pattern of allelic variation at the *R* gene, *RPP13*, is consistent with a history of either balancing or directional selection. We observe extreme amino acid polymorphism in the LRR region of the protein. This level of variation is greater than that of 17 other loci in *A. thaliana*, suggesting a history of balancing selection. Furthermore, the *A. thaliana* individuals show different levels of resistance to three naturally occurring pathogen isolates, suggesting that multiple, functionally differentiated alleles have been maintained within *A. thaliana* through reciprocal plant-pathogen coevolution.

TABLE 1

Accessions for the *RPP13* study and reactions to three *P. parasitica* isolates (Maks9, Emco5, and Wela3)

Accession[a]	Town	Maks9	Emco5	Wela3
Duc-1	Dunvagen Castle	S	S	R
Pet-1	Peterculter	R	S	R
Ty-0	Taynult	R	S	S
Ci-1	Crief	S	R	S
Edi-2	Edinburgh	R	R	R
Crl-1	Carlisle	S	S	S
Asp-1	Aspatria	S	R	R
Cul-1	Culgaith	S	S	R
Coc-1	Cockermouth	S	R	R
Poo-1	Pooley Bridge	R	R	S
Ksk-2	Keswick	S	S	R
Ksk-1	Keswick	S	R	R
Bra-1	Braithwaite	R	R	R
Leg-1	Legburthwaite	S	S	R
Sna-1	Snape Malting	R	R	R
Frd-1[b]	Fordwich	S	R	R[c] (S)
Lha-1	Lower Harbledown	S	S	R
God-1	Godmersham	S	S	R
Sco-1	Scotney Castle	R	S	R
Hil-1	Hillier Arboretum	S	S	R
Col-5[b]	Lab stock	S	S	R[c] (S)
Nd-1[b]	Lab stock	R	R	R[c] (S)
Ws-2	Lab stock	R	S	R
Rld-2[b]	Lab stock	R[c] (S)	R[c] (S)	R

R, resistant; S, susceptible.

[a] The localities of these accessions are indicated on the map in Figure 1.

[b] The recognition specificity of the alleles isolated from these ecotypes to these three *P. parasitica* strains was tested by allelic complementation using stable transgenics as described in Bittner-Eddy *et al.* (2000).

[c] Although these ecotypes were resistant to the corresponding pathogen isolate, the *RPP13* allele isolated from these individuals did not confer this pathogen recognition. Other loci in these plants are implicated for conferring resistance to these isolates.

MATERIALS AND METHODS

Isolation and sequencing of alleles: Alleles of *RPP13* were isolated from single individuals from 24 different populations of *A. thaliana* (Table 1; Figure 1). Hybridization data indicated that each individual of *A. thaliana* studied contained a single *RPP13* gene (P. Bittner-Eddy, unpublished data). Four of the *A. thaliana* individuals were standard laboratory accessions from northern Europe (Nd-1, Ws-2, Col-5, and Rld-2), while the other individuals were collected from 20 natural populations across the United Kingdom. The methods for DNA isolation and PCR amplification from *A. thaliana* were as described in Bittner-Eddy *et al.* (2000). Primers specific to flanking noncoding sequence, coupled with internal primers, were used to generate three overlapping segments encompassing the entire *RPP13* gene. The PCR products from several amplification reactions were pooled and sequenced directly.

Orthologous and paralogous sequences of *RPP13* were also obtained from *A. arenosa* and *A. lyrata*, both described as sister species to *A. thaliana* (Price *et al.* 1994). The *A. arenosa* individual was from a population in Soubey, Switzerland, and the two *A. lyrata* individuals were from populations in Saugatuck, Michigan, and Mayodan, North Carolina, respectively. DNA was extracted from these individuals using a modified CTAB method (Doyle and Doyle 1987). In anticipation of sequence divergence between *A. thaliana* and these two species in potential primer binding sites flanking the *RPP13* locus, primers were instead designed in the peripheral coding regions of the gene. *Pfu* proofreading polymerase (Stratagene, La Jolla, CA) was used to minimize PCR artifacts. The PCR products were cloned into the TOPO TA cloning vector (Invitrogen, Carlsbad, CA) and multiple clones were sequenced to confirm the sequences. Sequences were edited by eye and assembled in Sequencher (Gene Codes, Ann Arbor, MI).

Data analyses: The amino acid sequences were predicted from the nucleotide sequences using MacClade (Maddison

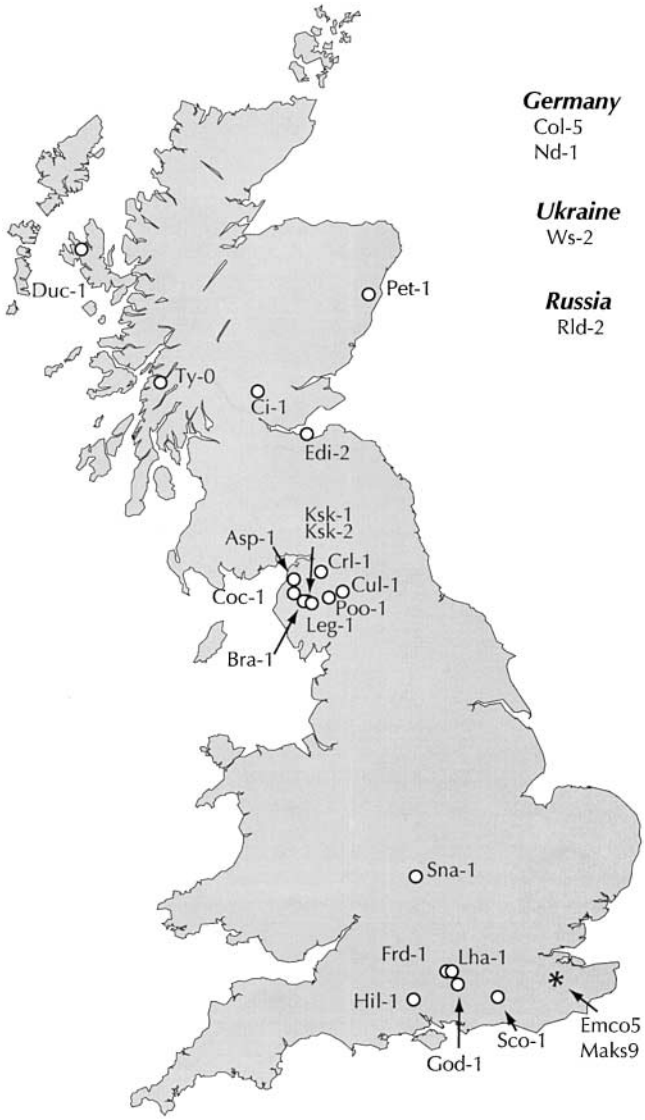

FIGURE 1.—Map of England showing origin of plants used in this study. The collection site of the two pathogen isolates, Maks9 and Emco5, is indicated by an asterisk.

and MADDISON 2000). Clustal X (THOMPSON *et al.* 1997) was used to align the predicted protein sequences. Minor adjustments to optimize this alignment were made by eye. Maximum parsimony, neighbor-joining, and bootstrapping analyses were completed using PAUP*4.0b8 (SWOFFORD 1999). DnaSP version 3.51 was used for intra- and interspecific genetic analyses and coalescent simulations (ROZAS and ROZAS 1999). The sliding window analysis was conducted as in AGUADÉ *et al.* (1992) on the basis of the estimation of silent and replacement substitutions proposed by NEI and GOJOBORI (1986). The sequence data for the interlocus comparisons were obtained from GenBank and the literature (INNAN *et al.* 1996; KAWABE *et al.* 1997, 2000; HENIKOFF and COMAI 1998; PURUGGANAN and SUDDITH 1998, 1999; KAWABE and MIYASHITA 1999; STAHL *et al.* 1999; KAMIYA *et al.* 2000; KUITTINEN and AGUADÉ 2000; AGUADÉ 2001; HAUSER *et al.* 2001; KLIEBENSTEIN *et al.* 2001; MAURICIO *et al.* 2003). Sawyer's Geneconv method was used to determine whether some regions of a pair of sequences had more consecutive identical polymorphic sites than expected by chance (SAWYER 1999). This test assumes that mutations are neutral and independently distributed and that there has been no history of recombination between sequences. Permutation of the sequences was used to assign *P* values to the observed shared fragments and to evaluate their statistical significance.

Phenotypic analyses: The reactions of all 24 *A. thaliana* accessions to three different *P. parasitica* isolates were determined. These isolates, Maks9, Emco5, and Wela3, were collected from naturally infected *A. thaliana* plants from Maidstone (United Kingdom), East Malling (United Kingdom), and Weiningen (Switzerland), respectively (KOCH and SLUSARENKO 1990; BITTNER-EDDY *et al.* 1999). The methods of inoculation and phenotyping were as described previously (HOLUB *et al.* 1994).

RESULTS

Intraspecific variation at *RPP13*: The length of the complete alignment of the 24 alleles was 2652 nucleotides. The predicted proteins encoded by individual alleles were between 820 and 843 amino acids in length. All alleles had the same overall domain structure and there were no obviously truncated genes (BITTNER-EDDY *et al.* 2000). The alleles showed numerous indel and nucleotide polymorphisms. In this 2.6-kb region, there were 32 indel polymorphisms and 365 nucleotide polymorphisms (see supplemental Figure 1 at http://www.genetics.org/supplemental/). One-half of all the indel polymorphisms occurred in the LRR region of the gene. Within the LRR, the indels were located most frequently in the putative β-α-connecting loop of the LRR, also described as the third subdomain in the repeat (BITTNER-EDDY *et al.* 2000).

The *RPP13* gene shows the greatest nucleotide polymorphism ($\pi = 0.043$; $\theta = 0.040$) of any gene surveyed to date from *A. thaliana* (Table 2; Figure 2). The average value of θ from *A. thaliana* across a sample of 17 other genes taken from the literature is 0.0085 (ranging from 0.0026 to 0.0206). Assuming even the highest of these θ-values ($\theta = 0.0206$ for ChiA; KAWABE *et al.* 1997), coalescent simulations with no recombination (the most conservative test) indicate that observing values of π or θ as high as those for the *RPP13* locus is unlikely ($P = 0.04$ and $P = 0.021$, respectively). The converse is also true. Assuming a genome-wide value of θ equal to that of *RPP13* ($\theta = 0.04$), even the highest levels of polymorphism in this sample from published surveys is improbable ($\theta \leq 0.0206$, $P = 0.026$; $\pi \leq 0.0109$, $P = 0.003$).

Not only is the synonymous and nonsynonymous polymorphism high across the entire gene ($\pi_{syn} = 0.049$; $\pi_{non} = 0.041$) compared to other genes in *A. thaliana*, but also the ratio of $\pi_{non}/\pi_{syn} = 1.5$ in the LRR. There is a significant excess of nonsynonymous polymorphisms per nonsynonymous site relative to synonymous polymorphisms per synonymous site ($\chi^2 = 3.92$, $P = 0.048$) in the LRR, suggesting balancing selection favoring amino acid variation in the LRR. In the non-LRR region, $\pi_{non}/\pi_{syn} = 0.44$ and there is a significant excess of synonymous polymorphisms per synonymous site ($\chi^2 = 13.6$,

TABLE 2

Patterns of nucleotide polymorphism and divergence at the *RPP13* locus of *A. thaliana*

	Total	Synonymous	Nonsynonymous
Entire gene			
No. of sites	2343	528.2	1814.8
Segregating sites	324	79	245
π	0.045	0.049	0.044
K^a	0.1	0.15	0.089
Tajima's *D*	0.092 (NS)		
First part of gene (excludes LRR)			
No. of sites	1608	356.1	1251.9
Segregating sites	119	43	76
π	0.023	0.041	0.018
K	0.074	0.15	0.058
Tajima's *D*	0.41 (NS)		
Second part of gene (includes LRR)			
No. of sites	735	172.1	562.9
Segregating sites	205	36	169
π	0.088	0.062	0.096
K	0.16	0.16	0.16
Tajima's *D*	−0.32 (NS)		

NS, not significant.

[a] Divergence relative to *RPP13* ortholog from *A. arenosa* (Aren1).

$P = 0.0002$). This suggests purifying selection against amino acid variants outside of the LRR.

While the number of segregating sites is 324, the estimated minimum number of mutations is 403, indicating that multiple hits have occurred at some positions. In some cases, three or more different amino acid residues were observed at a single codon position. In the first two-thirds of the gene, the region encoding the coiled-coil domain and the nucleotide-binding site, 71/542 (13%) of the codons exhibited nonsynonymous polymorphisms. Three or more amino acids were encoded at 6/71 of these polymorphic codons (8.4%).

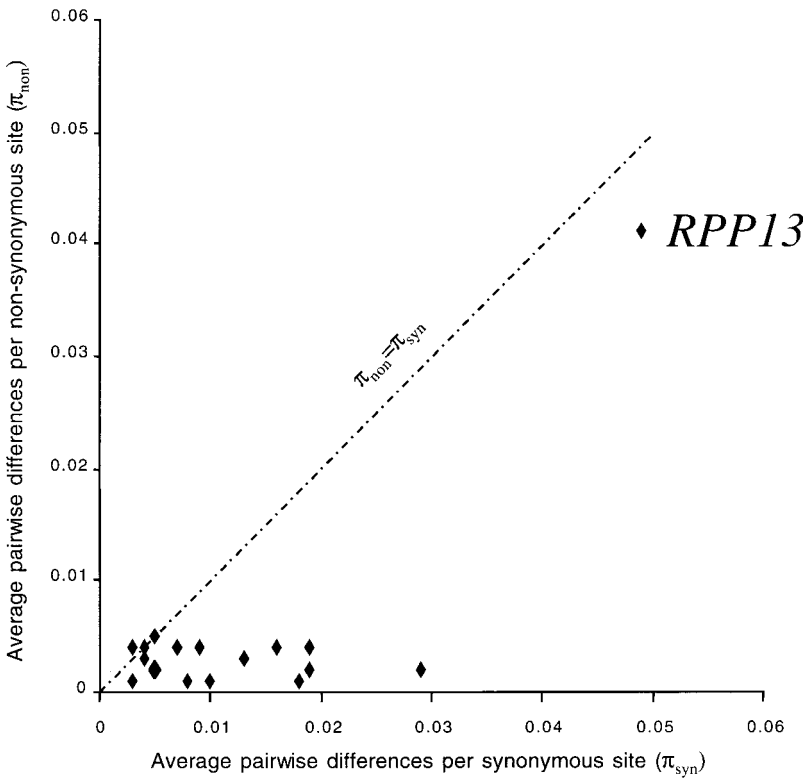

FIGURE 2.—The ratio of π_{non} to π_{syn} across 18 genes of *A. thaliana*. The sequence data for the interlocus comparisons were obtained from GenBank and the literature (HANFSTINGL *et al.* 1994; KAWABE *et al.* 1997, 2000; HENIKOFF and COMAI 1998; PURUGGANAN and SUDDITH 1998, 1999; CAICEDO *et al.* 1999; KAWABE and MIYASHITA 1999; STAHL *et al.* 1999; KAMIYA *et al.* 2000; KUITTINEN and AGUADÉ 2000; AGUADÉ 2001; HAUSER *et al.* 2001; KLIEBENSTEIN *et al.* 2001).

In contrast, 124/282 (43%) of the codons in the LRR exhibited nonsynonymous polymorphisms. Not only was the level of polymorphism higher, but also 55% of the polymorphic codons encoded three or more amino acids and more than one-quarter encoded four or more amino acids. These codons with greater than two amino acids segregating are concentrated in the junctions between the β-strand, β-turn motif and the connecting β-α-loop of the individual LRRs (see supplemental Figure 1 at http://www.genetics.org/supplemental/).

Frequency spectrum of variation: The Tajima's *D* (Tajima 1989) value was negative, but did not differ significantly from zero. Partitioning the gene into two regions (exclusively the LRR *vs.* the rest of the gene) to test if the pattern of evolution differed between these two functionally differentiated parts did not result in a significant deviation from zero (Table 2). The frequency of singletons (polymorphic sites where the rarest variant is present only once in the sample) in this sample is 27%. An excess of singleton mutations relative to the neutral expectation has been observed at multiple *A. thaliana* loci including *Adh1*, *AP3*, *CAL*, *CHI*, *ChiA*, and *PI* (Innan *et al.* 1996; Kawabe *et al.* 1997; Purugganan and Suddith 1998, 1999; Kuittinen and Aguadé 2000). To examine the difference between *RPP13* and samples from other loci, the frequency spectra of polymorphism were compared across 14 loci (see supplemental Figure 2 at http://www.genetics.org/supplemental/). In this comparison, over half of the non-*RPP13* genes show a high proportion of singletons (>50%). Furthermore, singletons make up a large proportion (60%) of the polymorphic amino acids at these other loci, while singletons make up a smaller proportion (26%) of the polymorphic amino acids at the *RPP13* locus.

Haplotype structure: Parsimony and neighbor-joining trees were inferred on the basis of the nucleotide sequences of the *RPP13* alleles from *A. thaliana* and *A. arenosa* (Figure 3; see also supplemental Figure 3 at http://www.genetics.org/supplemental/). High bootstrap values support the monophyly of the clade composed of the *RPP13* alleles from *A. thaliana*, as well as the larger clade containing the *RPP13* alleles from *A. thaliana* plus the ortholog from *A. arenosa*. Nineteen different *RPP13* alleles were detected in the 24 accessions. One allele was found in four accessions (*i.e.*, Hil-1, Duc-1, Leg-1, and Lha-1) and two alleles were found in two accessions (Ci-1 and Ti; Crl-1 and Bra-1). While multiple clades within *A. thaliana* were well supported in both analyses, the alleles in clade A and clade B are the most differentiated; alleles from these two clades show 66 fixed differences distributed across the entire *RPP13* coding region. Clade A shows a low level of within-clade variation ($\pi = 0.0019$). Among the five alleles in this clade, 14 of 15 total polymorphisms are singletons and nearly all of these result in an amino acid difference. Variation in clade B is much greater

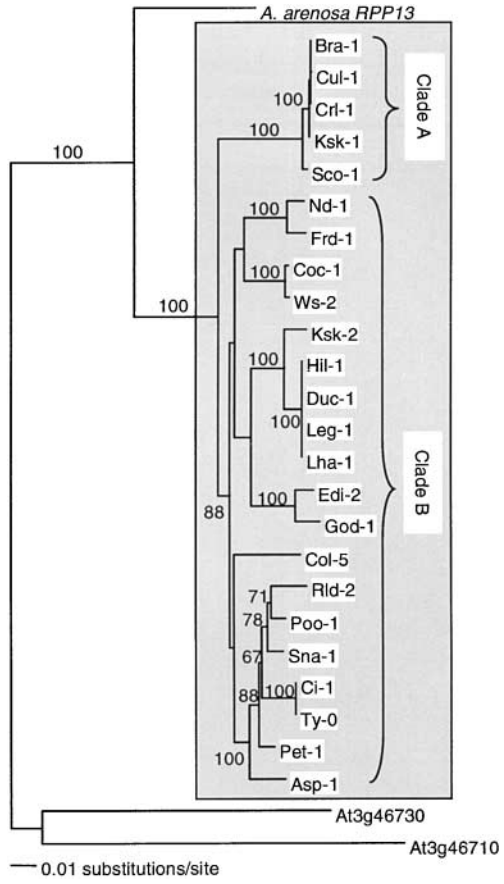

Figure 3.—Neighbor-joining tree of alleles of *RPP13*, ortholog from *A. arenosa*, and paralogs from *A. thaliana*. A neighbor-joining tree was inferred on the basis of the nucleotide sequences of the *RPP13* alleles from *A. thaliana* and *A. arenosa* using PAUP*4.0b8 (Swofford 1999). The tree was rooted using the sequence of two paralogs from the accession Columbia (GenBank accession nos. At3g46730 and At3g46710). The HKY85 substitution model was used. Bootstrap proportions of 100 bootstrap replicates >50 are indicated on the branches. Shaded area highlights the cluster of *A. thaliana RPP13* alleles.

($\pi = 0.04$) and the overall proportion of singletons is much lower within this haplotype (25.7%). Within clade B, there is some evidence for further divisions among alleles. The three pairs of alleles (Nd-1 and Frd-1, Ws-2 and Coc-1, and God-1 and Edi-2) each emerge in both neighbor-joining and parsimony analyses with high bootstrap support. These three allele pairs are diverged relative to all others in their sequences, despite evidence for recombination at this locus (see below). Another cluster of seven alleles (Rld-2, Poo-1, Sna-1, Ci-1, Ty-0, Pet-1, and Asp-1) also share distinct substitutions with one another relative to all other alleles. However, a straightforward haplotype tree could not be constructed due to recombination.

Evidence for recombination: Although estimates of outcrossing rates in *A. thaliana* are very low (Abbott and Gomes 1989), there was abundant evidence for recombination at the *RPP13* locus. Given such high

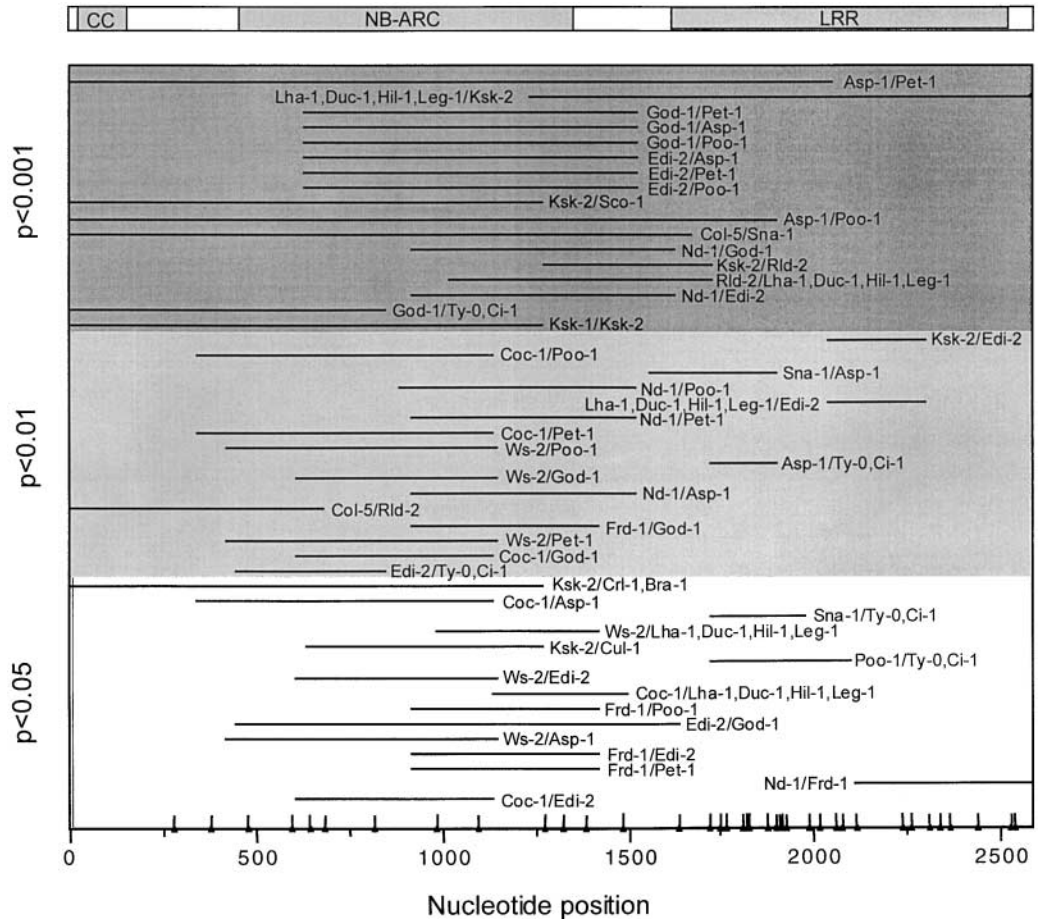

FIGURE 4.—Length and position of putative gene conversion/recombination tracks ordered by decreasing *P* value. The *P* values are Karlin-Altschul *P* values, based on the BLAST method for finding sequence matches in DNA or protein databases (ALTSCHUL *et al.* 1990) and are Bonferroni corrected. The names of the alleles involved in these gene conversion/recombination events are separated by a slash. Three sets of alleles are identical in sequence and are separated by a comma. The tick marks along the *x*-axis are the midpoints of the 38 minimum recombination intervals based on Hudson's four-gamete test. The structure of the *RPP13* gene is indicated at the top of the figure.

levels of polymorphism at *RPP13*, there was even greater power to detect recombinants at this locus than at other loci in *A. thaliana*. The four-gamete test (HUDSON and KAPLAN 1985) indicated that a minimum of 38 recombination events are necessary to explain the pattern of variation at this locus (Figure 4). The program Geneconv (SAWYER 1999) was used to infer recombination or gene conversion events in the genealogy of the alleles. Over 50 putative gene conversion/recombination tracks were identified in this sample, reinforcing the observation of extensive recombination at this locus (Figure 4). In some places the two analyses identified the same endpoints of recombinational events, while in other regions, the two analyses were not in agreement (Figure 4).

Only one putative recombinant between haplotype clades A and B was detected. This allele, Ksk-2, shared 82 unique polymorphisms and only two differences with Sco-1 between sites 1 and 1286; whereas Ksk-2, Lha-1, Leg-1, Duc-1, and Hil-1 shared 300 polymorphisms and no differences between sites 1243 and 2652. The Ksk-2 allele may have originated fairly recently because only one event is needed to infer the origin of this allele and the potential donor sequences found among other alleles in this study. The recent origin of this allele is further supported by the high sequence identity shared between each "recombinant" portion of the Ksk-2 allele with the inferred donor sequences.

The program Geneconv was also used to evaluate whether recombination or gene conversion events involving alleles of RPP13 and RPP13 paralogs had occurred. No evidence of conversion or recombination was found between these alleles and the two paralog sequences available from the Columbia ecotype.

Interspecific comparisons: Joint analyses of within- and between-species divergence can increase the statistical power to detect deviations from neutral evolution (AKASHI 1999). To identify orthologous sequences from close relatives of *A. thaliana*, sequences were PCR amplified with *RPP13*-specific *A. thaliana* primers. Phylogenetic analyses revealed that one of these amplified sequences, Aren1 from *A. arenosa*, was sister to the clade of *RPP13* alleles from *A. thaliana* (Figure 3; see also supplemental Figure 3 at http://www.genetics.org/supplemental/). Many other sequences amplified from *A. arenosa* and *A. lyrata* cluster with either one of the two previously described paralogs of *A. thaliana* (*e.g.*, Lyr3, Aren3, and Aren4). A number of studies use *A. lyrata* for interspecific sequence comparisons to *A. thaliana* (*e.g.*, WRIGHT *et al.* 2002; BARRIER *et al.* 2003). However, none of our amplification products from *A. lyrata* appeared to be orthologous to *RPP13*. Two sequences, Lyr1 and Lyr2, possess the greatest nucleotide identity to *RPP13* alleles among the collection of *A. lyrata* sequences. However, frameshift mutations within both of

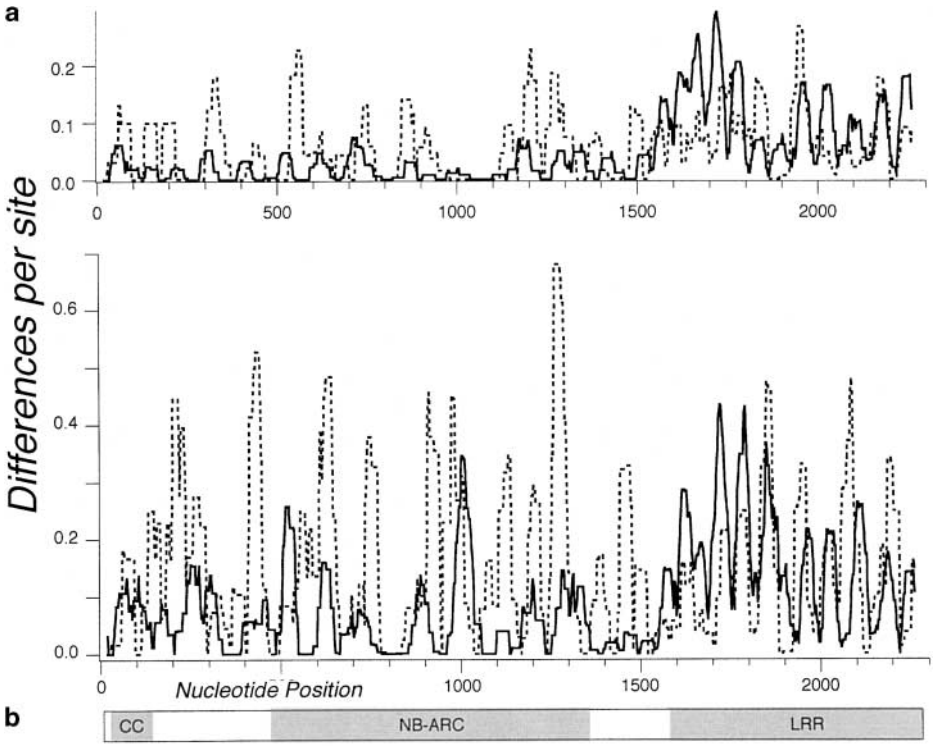

FIGURE 5.—Sliding window analyses: (a) average number of differences per site between *RPP13* alleles within *A. thaliana* and (b) between *A. thaliana* and *A. arenosa*. Dashed lines are synonymous variation, while solid lines are nonsynonymous. The sliding window analysis was conducted as in AGUADÉ *et al.* (1992) on the basis of the estimation of silent and replacement substitutions proposed by NEI and GOJOBORI (1986). Values are midpoints of 36-bp windows. The positions of the coiled coil (CC), the nucleotide-binding and ARC domains (NB-ARC), and the leucine-rich repeats (LRR) are indicated below the plots.

these sequences prevent their use in population genetic analyses (*i.e.*, calculations of K_s and K_a).

Average divergence at synonymous sites between Aren1 and the *RPP13* alleles from *A. thaliana* is 0.15. This value is close to the estimated divergence at synonymous sites across several loci between *A. thaliana* and *A. lyrata*: WRIGHT *et al.* (2002) report an average K_s = 0.126 (range 0.027–0.23) for 24 loci; BARRIER *et al.* (2003) report an average K_s = 0.119 (range 0.0–0.55) for 304 expressed sequence tags. However, nonsynonymous divergence (K_a) between *A. thaliana* and *A. arenosa* at *RPP13* (K_a = 0.089) exceeds the estimates reported for the *A. thaliana/A. lyrata* comparisons: K_a = 0.0211 (range 0.0–0.092) from WRIGHT *et al.* (2002) and K_a = 0.025 (range 0.0–0.16) from BARRIER *et al.* (2003). Amino acid divergence reaches 16% in the LRR, at least six times greater than the average amino acid divergence in the *A. thaliana/A. lyrata* gene comparisons.

Distribution of variation across the gene: Sliding window analyses were used to characterize the pattern of polymorphism and divergence across the *RPP13* gene (Figure 5). The pattern of nonsynonymous and synonymous polymorphism and divergence differs across the gene with nonsynonymous polymorphism and divergence peaking in the LRR. For the interspecific comparison of *RPP13* in *A. thaliana* and *A. arenosa*, the level of synonymous divergence (K_s) exceeded nonsynonymous divergence (K_a) in the first two-thirds of the gene (Figure 5). The K_a/K_s ratio of 0.389 in this region is comparable to π_{non}/π_{syn} = 0.44 for the intraspecific comparison. The pattern of sequence divergence also mirrors polymorphism in the LRR region, showing greater amino acid divergence relative to silent divergence. K_a/K_s is ~1 in the LRR and exceeds 1 at the junctions between β-strand, β-turn motif and the connecting β-α-loop of the individual LRRs. Both the interspecific and the intraspecific comparisons indicate that *RPP13* has a rate of amino acid evolution higher than that of any gene studied in *A. thaliana*, especially in the LRR.

The McDonald-Kreitman test was also used to determine if the level of nonsynonymous polymorphism observed at the *RPP13* locus exceeded that expected under neutrality. Under neutrality, the levels of intraspecific polymorphism and interspecific divergence are expected to be correlated (MCDONALD and KREITMAN 1991). At the *RPP13* locus, we detected a significant departure from the null expectation (*i.e.*, the ratios of polymorphism and divergence differed between synonymous and nonsynonymous mutations, $P < 0.0005$; Table 3). Since studies of other *R* genes revealed a pattern of diversifying selection acting on the LRR region and functional studies indicate that pathogen-recognition specificity is encoded by this region, we tested whether this region alone was responsible for the deviation from the neutral expectation. Indeed, an analysis of the first half of the gene, which does not contain the LRR, does not show a deviation from neutrality, while the analysis of the LRR region alone does (Table 3). These results indicate an excess of amino acid polymorphisms segregating in the LRR region relative to the neutral expectation.

Correlation with resistance phenotype: The resistance responses to three isolates of *P. parasitica*, Maks9, Emco5, and Wela3, were determined for the 24 accessions of *A. thaliana* (Table 1). The genetic basis of resistance to these isolates of *P. parasitica* was investigated

TABLE 3

Summary of McDonald-Kreitman tests

	Polymorphic sites within *A. thaliana*	Fixed differences between *A. thaliana* and *A. arenosa*	Fisher's exact test (two-tailed)
Entire gene			
Synonymous	79	49	
Nonsynonymous	240	66	$P < 0.0005$
First part of gene (excludes LRR)			
Synonymous	43	33	
Nonsynonymous	76	40	$P = 0.227$ (NS)
Second part of gene (includes LRR)			
Synonymous	36	16	
Nonsynonymous	164	26	$P = 0.0067$

NS, not significant.

in greater depth by transgenic complementation (Table 1; Bittner-Eddy *et al.* 2000). The Nd-*RPP13* allele encodes specific recognition to both the Maks9 and the Emco5 isolates of *P. parasitica*, while the Col and Rld alleles do not encode specific resistance to these two isolates (Bittner-Eddy *et al.* 2000). Furthermore, the Rld-*RPP13* gene encodes resistance to a third isolate, Wela3, while Col and Nd alleles do not encode resistance to this isolate (Bittner-Eddy *et al.* 2000). Recently a third allele, Frd, has been transformed into the susceptible Col ecotype; this allele encodes resistance to Emco5, but not to Maks9 (P. Bittner-Eddy, unpublished data). Although the *RPP13* allele isolated from the Col individual does not confer resistance to any of the three *P. parasitica* isolates used in this study, the Col-*RPP13* allele does not contain any frameshift mutations nor are there any other indications that this allele is nonfunctional. Downstream genes in the *RPP13* pathway of Col are not compromised, as this genotype showed isolate-specific resistance when transformed with each of the Rld, Frd, or Nd alleles.

The Nd, Frd, Rld, and Col alleles differ from each other by a large number of amino acids, so identifying residues that may be involved in pathogen recognition was not possible through simple pairwise comparisons. However, all of the accessions in this study were phenotyped for resistance to Maks9 and Emco5. Since resistance due specifically to the *RPP13* gene has been demonstrated only for the Nd and Frd alleles, resistance in other plants may not be encoded by the *RPP13* locus. The Rld accession is resistant to Maks9 and Emco5 but this resistance maps to other position(s) in the genome (Bittner-Eddy *et al.* 2000). Furthermore, some accessions had the same *RPP13* allele but differed in their reaction to *P. parasitica* (*i.e.*, Bra and Crl). Therefore the only informative comparison to identify amino acid differences that are associated with resistance or susceptibility was between the alleles from the susceptible accessions and those that have been shown by transformation to confer resistance.

Over half of the accessions studied were susceptible to one of the two isolates. The subset of 10 accessions that were resistant to the Maks9 isolate was different from the subset of 11 accessions that were resistant to Emco5. Nine were susceptible to both isolates. The "susceptible" alleles to each pathogen isolate are found in all the major clusters of the gene tree, indicating that a large diversity in amino acid sequence was found among alleles from susceptible plants; *i.e.*, susceptibility alleles are not more similar in sequence to one another than they are to alleles from resistant plants.

Alleles of *RPP13* from the Nd and Frd accessions conferred resistance to Emco5. There are 13 amino acid positions in which Nd and Frd have the same amino acid, but all of the susceptible alleles have a different amino acid (see supplemental Table 1 at http://www.genetics.org/supplemental/). Twelve of 13 of these polymorphisms are located in the LRR. Since the amino acid residues associated with resistance to Emco5 reside predominantly in the LRR, and the LRR regions of the Nd and Frd alleles are substantially differentiated from other alleles in the sample, it is likely that at least some of the recognition determinants of Emco5 are localized to the LRR region. The Nd allele has been demonstrated to confer resistance to Maks9 (Bittner-Eddy *et al.* 2000). A comparison of the Nd sequence to the alleles from accessions susceptible to Maks9 revealed that five amino acid residues were unique to Nd (see supplemental Table 2 at http://www.genetics.org/supplemental/). These five amino acids are located between sites 167 and 313 of the amino acid alignment and were restricted to the NBS. One difference at site 167 is found within the conserved motif 1 as described in van der Biezen and Jones (1998). None of the other four differences are located in conserved regions or regions for which functions have been ascribed such as kinase domains of the NBS.

DISCUSSION

A. thaliana is naturally infected by *P. parasitica* and genetic material of both organisms used in this study was

collected from natural populations in northern Europe, predominantly in the United Kingdom. The resistance gene, *RPP13*, shows both extreme sequence diversity and functional diversity in pathogen recognition. The pattern of sequence variation at *RPP13* suggests a coevolutionary interaction between host and parasite that is still very active. The presence of extreme polymorphism at this locus is consistent with the prediction that genes involved in pathogen recognition and defense should show elevated levels of polymorphism (Haldane 1949).

Such extreme intraspecific amino acid polymorphism has not been described at other *R*-gene loci in *A. thaliana* (Bergelson *et al.* 2001). The *RPS5* and *RPM1* loci segregate only two haplotypes each: one haplotype with a functional version of the gene and one having a null (deleted) allele. Purifying selection resulting in the conservation of the amino acid sequences of *RPS5* and *RPM1* alleles may be responsible for the low level of amino acid polymorphism at these loci. A third well-described locus, *RPS2*, also does not show the levels of nonsynonymous and synonymous polymorphism found at *RPP13* (Caicedo *et al.* 1999; Mauricio *et al.* 2003). Unlike the *RPP13* locus, only one pathogen-recognition specificity has so far been described at the *RPS2* locus. The disparity in amino acid polymorphism found between these loci may be related to the maintenance of multiple recognition specificities at the *RPP13* locus due to selection by naturally occurring pathogens. *P. parasitica* commonly infects *A. thaliana* in the wild, and variation in host resistance and pathogen virulence has been shown (Holub *et al.* 1994). Although alleles of *RPS2* from *A. thaliana* encode recognition to the *Pst* pathogen, little is known about the natural infection of *A. thaliana* by *Pst*, a tomato pathogen.

The pattern of sequence variation and segregation of multiple functionally distinct alleles at the *RPP13* locus most closely resembles the observations of the allelic variation at the *L* locus in flax. Thirteen alleles of the *L* locus have been described and each confers a different rust-resistance specificity (Ellis *et al.* 1999). The levels of both silent and amino acid polymorphism are high at this locus: π_{non} is 0.017 in the nonLRR region and reaches 0.051 in the LRR region. As observed at the *RPP13* locus, π_{non} exceeds π_{syn} in the LRR region, but not in the regions excluding the LRR. However, the sample of alleles from the *L* locus is not random; these alleles were specifically selected because they conferred different rust-resistance specificities. Our Arabidopsis sample was derived from naturally occurring populations from across Europe and was not selected on the basis of *a priori* phenotypic observations. In light of the random sampling undertaken in our study, the polymorphism at the *RPP13* locus is perhaps even more extraordinary because individuals with divergent phenotypes were not explicitly selected for analysis.

The observation that sequence variation is highest in the LRR portion of the gene is consistent with other studies of *R* genes encoding LRR domains (*e.g.*, Mondragon-Palomino *et al.* 2002) and the McDonald-Kreitman test indicates that this region has more amino acid polymorphism than expected under neutrality. The LRR regions of resistance proteins may be involved in determining specificity of gene-for-gene interactions found in plants (Staskawicz *et al.* 1995). This has been supported by domain swaps and mutational analyses of *R* genes, although physical interaction between a pathogen avr protein and the LRR region of a resistance protein has been demonstrated only between the AvrPita molecule expressed by the fungus *M. grisea* and the Pi-ta resistance protein from rice (Jia *et al.* 2000). In addition, studies on the *L* locus demonstrated that other domains may also be involved in determining specificity (Ellis *et al.* 1999; Luck *et al.* 2000). However, in all *R*-gene studies, protein sequence variation in the LRRs of *R* genes is correlated with different pathogen-recognition specificities (Wang *et al.* 1998; Bryan *et al.* 2000; Hwang *et al.* 2000; Banerjee *et al.* 2001; Dodds *et al.* 2001; Van der Hoorn *et al.* 2001; Wulff *et al.* 2001). At the *RPP13* locus, the excess amino acid polymorphism relative to silent polymorphism is consistent with the hypothesis that this region is experiencing diversifying selection. Furthermore, 12 of 13 amino acid differences between two alleles that confer recognition to the *P. parasitica* isolate Emco5 and the alleles from susceptible individuals occurred in the LRR portion of the protein. This observation is consistent with the LRR playing a central role in pathogen recognition.

Is it possible that the amount of amino acid polymorphism observed at this locus is due to relaxed selection pressure at this locus? Two factors could result in relaxed selection:

1. The pathogen is not a consistent selective agent; *i.e.*, allelic variation accumulates during episodes when this host is not exposed to the pathogen.
2. Host demographic factors, such as the predominantly selfing nature of the species and population dynamics dominated by rounds of colonization and extinction, result in a reduction in effective population size, which has been shown to affect the efficacy of selection.

At other loci, the prevalence of amino acid replacements occurring as singletons has been interpreted as evidence that selection against slightly deleterious mutations has not been as effective in *A. thaliana* as in other organisms (Sawyer *et al.* 1987; Purugganan and Suddith 1999). However, several lines of evidence argue against the accumulation of amino acid polymorphisms due to strictly neutral processes. If the gene were evolving neutrally (perhaps because the pathogen was not a consistent selective agent) we might expect that some individuals would lose the gene through deletion, or the gene might accumulate frameshift or nonsense mutations. To the contrary, none of the 24 randomly sampled individuals possessed null alleles. All genes were full length

(none encode obviously truncated proteins) and none have frameshift mutations.

Amino acid variation is associated with functional differentiation; that is, amino acid-differentiated alleles encode recognition to different pathogen isolates, indicating that at least some of the amino acid differences in these alleles contribute to functional differentiation. In the case of resistance to the pathogen Emco5, 12 of 13 of the amino acid residues shared among resistant alleles were found in the LRR, a region shown to affect pathogen recognition in other *R* genes. It would be unlikely to observe such an association between protein function and protein sequence if the gene were evolving neutrally.

Multiple analyses suggest the nonneutral evolution of the LRR; the amino acid polymorphism in this region exceeds that of the neutral expectation. While not all of the segregating variation necessarily has functional consequences, it is likely that at least some of these predominantly nonconservative amino acid changes concentrated in the putatively exposed residues affect pathogen recognition. Experiments involving domain swaps between alleles and site-directed mutational analyses will help to resolve precisely which of the many amino acid differences are functionally important.

All of these lines of evidence point to the selective maintenance of sequence variation at this locus, driven by a variable pathogen species. Furthermore, given the demography and mating system of *A. thaliana*, the allelic polymorphism has most likely been maintained through negative frequency-dependent selection and not overdominance. The long-term maintenance of many differentiated alleles is clearly inconsistent with recurrent selective sweeps operating at this locus over large geographic scales. The presence of several recombinant *RPP13* alleles indicates that heterozygotic individuals must have been present multiple times in the past. This indicates some, albeit potentially infrequent, outcrossing and segregation of differentiated alleles that affect disease resistance within *A. thaliana* populations. The characterization of allelic variation at the *RPP13* locus and observation of recombinant alleles provide the necessary materials for future investigations of the role of recombination in generating novel recognition specificities in a natural host-parasite interaction.

We thank H. Akashi, J. Parsch, and two anonymous reviewers for their helpful comments. We are grateful to A. Kawabe, K. Olsen, and E. Stahl for sharing their alignments of some of the genes used for the interlocus comparisons. This work was supported by grants from the U.S. National Science Foundation (to L.E.R., R.W.M., and C.H.L.) and the Biotechnology and Biological Sciences Research Council (to P.B.-E., E.B.H., and J.L.B.).

LITERATURE CITED

Abbott, R. J., and M. F. Gomes, 1989 Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. Heredity **62:** 411–418.

Aguadé, M., 2001 Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana*. Mol. Biol. Evol. **18:** 1–9.

Aguadé, M., N. Miyashita and C. H. Langley, 1992 Polymorphism and divergence in the *Mst26a* male accessory gland gene region in Drosophila. Genetics **132:** 755–770.

Akashi, H., 1999 Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. Genetics **151:** 221–238.

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

Banerjee, D., X. Zhang and A. F. Bent, 2001 The leucine-rich repeat domain can determine effective interaction between RPS2 and other host factors in Arabidopsis *RPS2*-mediated disease resistance. Genetics **158:** 439–450.

Barrier, M., C. D. Bustamante, J. Yu and M. D. Purugganan, 2003 Selection on rapidly evolving proteins in the Arabidopsis genome. Genetics **163:** 723–733.

Bergelson, J., M. Kreitman, E. A. Stahl and D. Tian, 2001 Evolutionary dynamics of plant R-genes. Science **292:** 2281–2285.

Bittner-Eddy, P., C. Can, N. Gunn, M. Pinel, M. Tor *et al.*, 1999 Genetic and physical mapping of the *RPP13* locus, in Arabidopsis, responsible for specific recognition of several *Peronospora parasitica* (downy mildew) isolates. Mol. Plant-Microbe Interact. **12:** 792–802.

Bittner-Eddy, P. D., and J. L. Beynon, 2001 The Arabidopsis downy mildew resistance gene, *RPP13*-Nd, functions independently of *NDR1* and *EDS1* and does not require the accumulation of salicylic acid. Mol. Plant-Microbe Interact. **14:** 416–421.

Bittner-Eddy, P. D., I. R. Crute, E. B. Holub and J. L. Beynon, 2000 *RPP13* is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. Plant J. **21:** 177–188.

Bryan, G. T., K.-S. Wu, L. Farrall, Y. Jia, H. P. Hershey *et al.*, 2000 ta single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell **12:** 2033–2045.

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

Dodds, P. N., G. J. Lawrence and J. G. Ellis, 2001 Six amino acid changes confined to the leucine-rich repeat beta-strand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax. Plant Cell **13:** 163–178.

Doyle, J. J., and J. L. Doyle, 1987 A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. **19:** 11–15.

Ellis, J. G., G. J. Lawrence, J. E. Luck and P. N. Dodds, 1999 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., L. Godiard, E. Straube, T. Ashfield, J. Lewald *et al.*, 1995 Structure of the Arabidopsis *RPM1* gene enabling dual specificity disease resistance. Science **269:** 843–846.

Grant, M. R., J. M. McDowell, A. G. Sharpe, T. Z. M. De, D. J. Lydiate *et al.*, 1998 Independent deletions of a pathogen-resistance gene in Brassica and Arabidopsis. Proc. Natl. Acad. Sci. USA **95:** 15843–15848.

Haldane, J. B. S., 1949 Disease and evolution. Ric. Sci. Suppl. **19:** 1–11.

Hanfstingl, U., A. Berry, E. A. Kellogg, J. T. I. Costai, W. Ruediger *et al.*, 1994 Haplotypic divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: Roles for both balancing and directional selection? Genetics **138:** 811–828.

Hauser, M.-T., B. Harr and C. Schlotterer, 2001 Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: molecular analysis of the candidate gene *GLABROUS1*. Mol. Biol. Evol. **18:** 1754–1763.

Henikoff, S., and L. Comai, 1998 A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in Arabidopsis. Genetics **149:** 307–318.

Holub, E. B., J. L. Beynon and I. R. Crute, 1994 Phenotypic and genotypic characterization of interactions between isolates of *Pe-*

ronospora parasitica and accessions of *Arabidopsis thaliana*. Mol. Plant-Microbe Interact. **7:** 223–239.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

Hulbert, S. H., C. A. Webb, S. M. Smith and Q. Sun, 2001 Resistance gene complexes: evolution and utilization. Annu. Rev. Phytopathol. **39:** 285–312.

Hwang, C.-F., A. V. Bhakta, G. M. Truesdell, W. M. Pudlo and V. M. Williamson, 2000 Evidence for a role of the N terminus and leucine-rich repeat region of the Mi gene product in regulation of localized cell death. Plant Cell **12:** 1319–1329.

Innan, H., F. Tajima, R. Terauchi and N. T. Miyashita, 1996 Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana*. Genetics **143:** 1761–1770.

Jia, Y., S. A. McAdams, G. T. Bryan, H. P. Hershey and B. Valent, 2000 Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. EMBO J. **19:** 4004–4014.

Kamiya, T., A. Kawabe and N. T. Miyashita, 2000 Nucleotide polymorphism at the *Atmyb2* locus of the wild plant *Arabidopsis thaliana*. Genes Genet. Syst. **75:** 409.

Kawabe, A., and N. T. Miyashita, 1999 DNA variation in the basic chitinase locus (*ChiB*) region of the wild plant *Arabidopsis thaliana*. Genetics **153:** 1445–1453.

Kawabe, A., H. Innan, R. Terauchi and N. T. Miyashita, 1997 Nucleotide polymorphism in the acidic chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. Mol. Biol. Evol. **14:** 1303–1315.

Kawabe, A., K. Yamane and N. T. Miyashita, 2000 DNA polymorphism at the cytosolic phosphoglucose isomerase (*PgiC*) locus of the wild plant *Arabidopsis thaliana*. Genetics **156:** 1339–1347.

Kliebenstein, D. J., V. M. Lambrix, M. Reichelt, J. Gershenzon and T. Mitchell-Olds, 2001 Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. Plant Cell **13:** 681–693.

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

Kuittinen, H., and M. Aguadé, 2000 Nucleotide variation at the CHALCONE ISOMERASE locus in *Arabidopsis thaliana*. Genetics **155:** 863–872.

Maddison, D. R., and W. P. Maddison, 2000 *MacClade 4: Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Sunderland, MA.

Mauricio, R., E. A. Stahl, T. Korves, D. Tian, M. Kreitman *et al.*, 2003 Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*. Genetics **163:** 735–746.

May, R. M., and R. M. Anderson, 1983 Parasite-host coevolution, pp. 186–206 in *Coevolution*, edited by D. Futuyama and M. Slatkin. Sinauer Associates, Sunderland, MA.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in Drosophila. Nature **351:** 652–654.

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

Mondragon-Palomino, M., B. Meyers, R. Michelmore and B. Gaut, 2002 Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. Genome Res. **12:** 1305–1315.

Luck, J. E., G. J. Lawrence, P. N. Dodds, K. W. Shepherd and J. G. Ellis, 2000 Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. Plant Cell **12:** 1367–1377.

Nei, M., and T. Gojobori, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

Price, R. A., J. D. Palmer and I. A. Al-Shehbaz, 1994 Systematic relationships of Arabidopsis: a molecular and morphological perspective, pp. 7–19 in *Arabidopsis* (Cold Spring Harbor Monograph Series), edited by E. M. Meyerowitz and C. R. Somerville. Cold Spring Harbor Laboratory Press, Plainview, NY.

Purugganan, M. D., and J. I. Suddith, 1998 Molecular population genetics of the Arabidopsis *CAULIFLOWER* regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. Proc. Natl. Acad. Sci. USA **95:** 8130–8134.

Purugganan, M. D., and J. I. Suddith, 1999 Molecular population genetics of floral homeotic loci: departures from the equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana*. Genetics **151:** 839–848.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Sawyer, S. A., 1999 *GENECONV: A Computer Package for the Statistical Detection of Gene Conversion* (http://www.math.wustl.edu/~sawyer).

Sawyer, S. A., D. E. Dykhuizen and D. L. Hartl, 1987 Confidence interval for the number of selectively neutral amino acid polymorphisms. Proc. Natl. Acad. Sci. USA **84:** 6225–6228.

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

Staskawicz, B. J., F. M. Ausubel, B. J. Baker, J. G. Ellis and J. D. G. Jones, 1995 Molecular genetics of plant disease resistance. Science **268:** 661–667.

Swofford, D., 1999 *PAUP**, Version 4.0b10. Sinauer Associates, Sunderland, MA.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–596.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins, 1997 The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **24:** 4876–4882.

Tian, D., H. Araki, E. Stahl, J. Bergelson and M. Kreitman, 2002 Signature of balancing selection in Arabidopsis. Proc. Natl. Acad. Sci. USA **99:** 11525–11530.

van der Biezen, E. A., and J. D. G. Jones, 1998 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 Hoorn, R. A. L., R. Roth and P. J. G. M. De Wit, 2001 Identification of distinct specificity determinants in resistance protein Cf-4 allows construction of a Cf-9 mutant that confers recognition of avirulence protein AVR4. Plant Cell **13:** 273–285.

Wang, G.-L., D.-L. Ruan, W.-Y. Song, S. Sideris, L. Chen *et al.*, 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.

Wright, S. I., B. Lauga and D. Charlesworth, 2002 Rates and patterns of molecular evolution in inbred and outbred Arabidopsis. Mol. Biol. Evol. **19:** 1407–1420.

Wulff, B. B. H., C. M. Thomas, M. Smoker, M. Grant and J. D. G. Jones, 2001 Domain swapping and gene shuffling identify sequences required for induction of an Avr-dependent hypersensitive response by the tomato Cf-4 and Cf-9 proteins. Plant Cell **13:** 255–272.

Communicating editor: O. Savolainen