Copyright © 2003 by the Genetics Society of America

Natural Selection for Polymorphism in the Disease Resistance Gene *Rps2* of *Arabidopsis thaliana*

Rodney Mauricio,*,†,1 Eli A. Stahl,*,‡,1 Tonia Korves,* Dacheng Tian,* Martin Kreitman*,‡ and Joy Bergelson*,‡,2

**Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637,* †*Department of Genetics, University of Georgia, Athens, Georgia 30602-7223 and* ‡*Committee on Genetics, University of Chicago, Chicago, Illinois 60637*

Manuscript received August 7, 2002
Accepted for publication November 11, 2002

ABSTRACT

Pathogen resistance is an ecologically important phenotype increasingly well understood at the molecular genetic level. In this article, we examine levels of *avrRpt2*-dependent resistance and *Rps2* locus DNA sequence variability in a worldwide sample of 27 accessions of *Arabidopsis thaliana*. The rooted parsimony tree of *Rps2* sequences drawn from a diverse set of ecotypes includes a deep bifurcation separating major resistance and susceptibility clades of alleles. We find evidence for selection maintaining these alleles and identify the N-terminal part of the leucine-rich repeat region as a probable target of selection. Additional protein variants are found within the two major clades and correlate well with measurable differences among ecotypes in resistance to the avirulence gene *avrRpt2* of the pathogen *Pseudomonas syringae*. Long-lived polymorphisms have been observed for other resistance genes of *A. thaliana*; the *Rps2* data suggest that the long-term maintenance of phenotypic variation in resistance genes may be a general phenomenon and are consistent with diversifying selection acting in concert with selection to maintain variation.

PLANTS are attacked by a multitude of pathogens and can respond to a subset of these attacks by mounting an induced defense response (Burdon 1987). The initial step in the induction of a defense response involves a genetic interaction between a specific allele of a disease resistance (R) gene and a complementary pathogen avirulence (avr) gene, the so-called gene-for-gene interaction (Flor 1956, 1971; Staskawicz *et al.* 1995). In *Arabidopsis thaliana*, the *Rps2* resistance gene confers resistance to pathogens with the avirulence gene *avrRpt2* in the pathogen *Pseudomonas syringae* (Dong *et al.* 1991; Whalen *et al.* 1991; Kunkel *et al.* 1993; Yu *et al.* 1993; Bent *et al.* 1994; Mindrinos *et al.* 1994). Recently, *P. syringae* strains have been found to infect *A. thaliana* in natural populations (Jakob *et al.* 2002).

The *RPS2* protein contains a nucleotide-binding site (NBS) and a leucine-rich repeat (LRR) region, two characteristics of a large family of plant R genes (*e.g.*, Salmeron *et al.* 1996; Thomas *et al.* 1997; McDowell *et al.* 1998; Meyers *et al.* 1998; Ellis *et al.* 1999; Noel *et al.* 1999; Bittner-Eddy *et al.* 2000; Luck *et al.* 2000). The LRR region is thought to function in pathogen recognition and thereby determine resistance specificity (Jones and Jones 1997; Leister and Katagiri 2000; Tao *et al.* 2000; Axtell *et al.* 2001). Within the LRR, solvent-exposed amino acid residues framed by conserved aliphatic residues are predicted to make direct contacts with the avirulence gene product or avr gene-dependent factor(s) (Jones and Jones 1997). Evolutionary analyses point to the framed, solvent-exposed residues as exhibiting very fast substitution rates due to positive Darwinian selection (Parniske *et al.* 1997; Meyers *et al.* 1998; Bittner-Eddy *et al.* 2000; Bergelson *et al.* 2001; Mondragón-Palomino *et al.* 2002), consistent with their direct role in pathogen (*i.e.*, avirulence gene) recognition. Other regions may also determine recognition (Ellis *et al.* 1999; Luck *et al.* 2000), however, and R gene-mediated resistance levels can also depend on other host factors (Banerjee *et al.* 2001).

Disease resistance genes are often polymorphic for resistance and susceptibility alleles (Kunkel 1996; Stahl *et al.* 1999; Ellis *et al.* 2000; Bergelson *et al.* 2001; Holub 2001). Caicedo *et al.* (1999) examined patterns of polymorphism among eight independent alleles of *Rps2* and found evidence of two divergent classes. Statistical tests of the data failed to detect evidence for natural selection, but several features of the data led the authors to suggest that selection, nonetheless, might be important at *Rps2*. First, the locus contained a high level of nucleotide polymorphism, with almost half of the polymorphisms resulting in amino acid changes. Second, the unrooted gene tree structure included one long branch separating a susceptibility allele (present in accessions Wu-0 and Zu-0-1) from a cluster of more closely related resistance and susceptibility alleles, a structure consis-

Sequence data from this article have been deposited with the EMBL/Genbank Data Libraries under accession nos. AF487796–AF487823.

[1]These authors contributed equally to this work.

[2]*Corresponding author:* Department of Ecology and Evolution, 1101 E. 57th St., Chicago, IL 60637. E-mail: jbergels@midway.uchicago.edu

tent with balancing selection maintaining *Rps2* polymorphism. Finally, the tree indicated a preponderance of amino acid changes between more closely related alleles, suggesting that diversifying selection may have generated *Rps2* sequence variation.

Here we extend the results of Caicedo *et al.* (1999) by carrying out statistical tests of selective neutrality and balancing selection at the *Rps2* locus with a larger sample of *A. thaliana* accessions and a sequence from the closely related congener, *A. lyrata*. We relate quantitative resistance phenotypes to the evolutionary history of the alleles and identify *RPS2* mutations that may confer phenotypic variation. We also test for associations of *Rps2* sequence variation and the geographic origin of alleles. The data are discussed in reference to the evolutionary processes thought to underlie plant disease resistance polymorphism.

MATERIALS AND METHODS

Plant materials: Twenty-seven accessions of *A. thaliana* were chosen to create a worldwide sample for *Rps2* sequencing representing the major geographic regions in the species' distribution (Table 1). Twelve of these accessions were taken from collections of J. Bergelson and R. Mauricio. Fifteen were selected from those at the *Arabidopsis* Biological Resource Center (ABRC) at random, except avoiding an excess of accessions from any one country. These accessions were obtained from the ABRC, and seeds from single individuals were harvested to create single-seed stocks for producing the plant materials used in the study. Two individuals of *A. lyrata* from Indiana (collected by R. Mauricio and D. Jacobson) were used to determine a consensus sequence of the locus for this species.

Phenotype assessment: Resistance phenotypes to the *P. syringae* avirulence gene, *avrRpt2*, were determined in all but six of the sequenced accessions, as well as the "Columbia" accession and the mutant, *rps2-101C* (in a Columbia background). Plants were grown from seed in Promix soil with a 12-hr day length at 20°. When the plants were 3–4 weeks old, one entire new leaf was infiltrated with *P. syringae* pv. *tomato* strain DC3000 at OD of 0.0002 using a blunt 1-cc syringe. The pathogen strain used in these infections contained a plasmid: either pLAFR3 or pLABL18. The pLABL18 plasmid is identical to the pLAFR3 plasmid, but contains an additional 3.6-kb fragment containing the *avrRpt2* gene (Whalen *et al.* 1991). Three days postinfection, bacterial levels were measured by grinding standard hole-punch-size leaf punctures in 10 mm $MgSO_4$ and plating dilutions on King's medium B with 40 mg/ml tetracycline. Five to eight replicate plants of each accession were infected with each bacterial strain per experiment. Phenotyping was replicated in at least two experiments for each sequenced accession, except for Po-1 and Mt-0, which were tested in only one experiment.

Plants were identified as resistant or susceptible by comparing the growth (colony-forming units per leaf punch, log-transformed) of the two pathogen strains in an analysis of variance (ANOVA) that included experimental day for accessions evaluated on multiple days. Accessions in which the pathogen strain with pLABL18 grew significantly less than the strain with pLAFR3 (Table 1, column 3) were designated "resistant." Other accessions were designated "susceptible."

For those accessions designated resistant, resistance was quantified by comparing pathogen growth in each focal accession with growth in the Columbia accession. Resistance relative to Columbia (Table 1, column 4) was calculated by dividing the difference in the growth of the strains pLABL18 and pLAFR3 in the focal accession by the difference in the growth of the strains in Columbia, as assessed on the same experimental days. Gaps in the distribution of relative resistance values were used to categorize accessions by degree of resistance. Accessions in the group with lower resistance than that of Columbia were labeled "mildly resistant" (mR), and those in the group with higher values than that of Columbia were labeled "strongly resistant" (sR). To determine the significance level for the degree of resistance relative to Columbia, we evaluated the significance of the interaction between accession (Columbia *vs.* the focal accession) and the pathogen strain (containing pLAFR3 or pLAB18) using ANOVA.

DNA sequence determination: For each accession, DNA was extracted from young rosette leaves using protocols described previously (Bergelson *et al.* 1998). The region encompassing *Rps2* was amplified in three overlapping amplicons, using primers GTTAGTTGGGTGGCGGGAGAG and GGCACAACCGAAACAACTGAGG, AACGGAGACTAAAACAGCCC and GACATGCATCTTCACC, and GTGGATCCATGCTAGTCACATTGAAGTTC and GACCTTTTTATTCCTTTTTCCG, in standard PCR protocols. Both strands were sequenced throughout the region using internal primers (sequences available from the authors) and ABI (Applied Biosystems, Foster City, CA) dye terminator sequencing chemistry. Sequences for each accession were compiled and aligned using Sequencher 3.0 (Genecodes, Ann Arbor, MI). A single consensus sequence for *A. lyrata* was generated from partial sequences of the two *A. lyrata* individuals. A small number of sites in our *A. lyrata* sequences were polymorphic; in each case one of the two alleles included the base found in *A. thaliana*, and we assigned the *A. thaliana* base to the consensus *A. lyrata* sequence for analyses. Multiple large insertions and deletions between the *A. thaliana* and *A. lyrata* sequences in the 5′ and 3′ noncoding regions substantially decreased the number of sites at which between-species comparisons could be made. As a consequence, some polymorphism analyses were conducted without considering the outgroup sequence.

We found differences between our sequences from accessions Wu-0 and Zu-0 and those reported by Caicedo *et al.* (1999) for the same accessions. In particular, Caicedo *et al.* (1999) did not detect mutations at positions 1279, 2554, and 3085 (Figure 1), and they found that variants at positions 3396 and 3502 were shared by Wu-0 and Zu-0 (Figure 1). Variation within accessions has been noted (Breyne *et al.* 1999; Caicedo *et al.* 1999) and may reflect ecotype seed collection from multiple (or heterozygous) individuals in the field.

Population genetic analyses: Silent (noncoding and synonymous) and amino acid replacement (nonsynonymous) polymorphism and divergence (Jukes-Cantor corrected) calculations were conducted using DnaSP (Rozas and Rozas 1997). Genealogy estimation was conducted by parsimony using PAUP (Swofford 1996), with 500 bootstrap replicates. Standard tests of a panmictic population, neutral mutation model utilized coalescent simulations with a fixed number of segregating sites (Hudson 1993), with programs available from R. Hudson (http://home.uchicago.edu/~rhudson1). Analyses testing for heterogeneity of polymorphism to divergence ratios were conducted using DNASlider (McDonald 1998); sliding window average *G*-statistics were analyzed for scaled population recombination rates, from $R_{Slider} = 0$ to 100 ($R_{Slider} = 4N_e rL$, where N_e is the effective population size, r is the recombination rate per base pair per generation and L is the length of the analyzed region), with the most conservative P values obtained for $R_{Slider} = 7$ for the entire sequenced region and $R_{Slider} = 6$ for the coding sequence. Linkage disequilibrium was tested for in 2×2 contingency tables by Fisher's exact

tests, using shareware available from W. Engels. Differentiation among "populations" (groups of sequences defined by phenotype or geographic origin) was calculated as $F_{ST} = 1 - \pi_W/\pi_T$, where π_W is average pairwise nucleotide difference within populations and π_T is that in total (Hudson *et al.* 1992b), and was tested by resampling with sequences permuted across groups (following Hudson *et al.* 1992a; Holsinger and Mason-Gamer 1996; Bergelson *et al.* 1998), using programs written by E. A. Stahl.

We analyzed a coalescent model with selection and recombination as described in Tian *et al.* (2002). Selection was assumed to maintain two alleles at fixed frequency 0.81 (and 0.19), acting at the beginning of the LRR (site 2654 in our alignment). The scaled per-base-pair recombination rate $R = 4N_e(1 - s)r = 0.00057$ uses published estimates of effective population size and selfing rate for *A. thaliana* (see Tian *et al.* 2002) and a recombination rate per meiosis estimated from regression of genetic and physical positions of markers near *Rps2* (2.71 cM/Mb, $r^2 = 0.96$; data from the Arabidopsis Genome Resource, http://ukcrop.net/agr, markers *mi475*, *SEP2B*, *m600*, *PG11*, *DD1*, *mi123*, *RLK5*, *mi232*, *prha*, *g8300*, and *mi431*). The scaled mutation rate between selected alleles (0.0125) was adjusted to fit the observed data near the selected site.

RESULTS

***avrRpt2*-dependent resistance phenotypes:** For each of 21 accessions, we compared the growth of *P. syringae* strain DC3000 with *avrRpt2* and DC3000 without *avrRpt2*. If an accession is resistant, the growth of the strain with *avrRpt2* should be significantly less than the growth of the strain without *avrRpt2*. The log of growth of the pathogen without *avrRpt2* minus that of the pathogen with *avrRpt2* is listed in Table 1; this measure of resistance is unitless since it is equivalent to the log of the ratio of growth for the two pathogen strains. The results of our ANOVAs indicate that 17 of the 21 accessions tested were resistant. Accessions BG-4, Po-1, Zu-0, and Knox-2 and the Columbia *rps2* mutant showed no indication of resistance. Statistical designations of resistance and susceptibility were consistent with observed disease symptoms.

We determined whether resistant accessions inhibited bacterial growth of DC3000 with *avrRpt2* to different extents by comparing bacterial growth in each line relative to this same measure in a common paired control line, Columbia. Relative resistance values (Table 1) ranged from 0.285 to 1.61. Gaps in the distribution of relative resistance values, between 0.67 and 0.945 and between 1.14 and 1.39, allowed us to group alleles into three operational subclasses of resistance, mild (mR), intermediate (R), and strong (sR). We used relative resistance values rather than ANOVA *P* values (Table 1) to categorize accessions because the power to detect differences from Columbia varied among accessions. The mR group included accessions AB-7, GR-6, Wu-0, Yo-0, and Cvi-0, and the sR group included Pog-0, RLD-1, Co-1, and Tsu-0.

Low growth of DC3000 without *avrRpt2* in Pu-8 suggested partial resistance to the DC3000 background; additional resistance in the presence of *avrRpt2* indicated that Pu-8 is resistant, but we were unable to measure its relative resistance. Also, Wu-0 has been reported previously as susceptible (Caicedo *et al.* 1999) although it exhibits growth and symptoms consistent with an intermediate phenotype (Kunkel *et al.* 1993; this study). It is possible that Caicedo *et al.* (1999) studied a different genotype within Wu-0 (see materials and methods).

Molecular variation at *Rps2*: We surveyed DNA sequence variability in 27 accessions from throughout the species range, including the accessions whose resistance phenotypes we determined (Table 1), and from the closely related species *A. lyrata*. The sequenced region spans 4248 base pairs (bp) in *A. thaliana* accession Columbia (GenBank accession no. AL049483), from 1003 bp upstream of the *Rps2* start codon to 521 bp downstream of its stop codon. Our survey yielded a 4461-bp alignment including the *A. lyrata* sequence, with 3755 sites at which polymorphism and divergence were ascertained (Table 2).

The data including the outgroup sequence revealed a total of 197 nucleotide differences fixed between *A. lyrata* and all *A. thaliana* sequences and 58 single nucleotide polymorphisms distinguishing 18 haplotypes in the 27 *A. thaliana* alleles (Figure 1; Table 2). Within the *Rps2* coding sequence, we detected 55 nonsynonymous (amino acid changing) differences between species and 20 nonsynonymous polymorphisms. The *Rps2* coding sequence reading frame is intact in all individuals, despite two one-codon insertions in *A. lyrata* relative to *A. thaliana* at Columbia residues 741 and 771 (both in the LRR region) and a four-codon deletion at 877 (near the *RPS2* C terminus). We also introduced one-codon insertion/deletions (indels; in both *A. lyrata* and *A. thaliana*) at Columbia residues 86 and 737, where the two species differ at all three nucleotide positions; these three-base differences were not included in polymorphism analysis. We found numerous indels between species in noncoding regions and five indel polymorphisms all outside of the coding sequence. A homonucleotide run at 821 varied between two *A. lyrata* individuals, but in *A. thaliana* no microsatellites were detected. No heterozygous sites were detected in *A. thaliana* individuals. Overall levels of polymorphism and divergence at *Rps2* (Table 2) fall within the range seen at other loci in *A. thaliana* and *A. lyrata* (Kawabe and Miyashita 1999; Purugganan and Suddith 1999; Aguadé 2001).

In Table 3, levels of polymorphism and divergence in the LRR region are presented. Within this region, the β-pleated sheet structural motif consensus sequence (Jones and Jones 1997) allows framed solvent-exposed amino acid residues, specific candidates for positive selection, to be analyzed and compared with conserved structural residues and nonconserved residues between frames. Significantly greater K_a than K_s between R gene paralogs at framed exposed residues (Meyers *et al.* 1998; Bergelson *et al.* 2001) has provided strong evidence

TABLE 1

***Arabidopsis thaliana* accessions and their *avrRpt2*-dependent resistance phenotypes**

Accession	Geographic origin	Resistance[a]	Relative resistance[b]	Phenotypic category[c]
Po-1	Germany	−0.45		S
Knox-2	Indiana	−0.26		S
rps2-101C	Columbia mutant	−0.14		S
BG-4	Seattle	0.08		S
Zu-0	Switzerland	0.1		S
Yo-0	California	0.68*	0.285***	mR
AB-7	Indiana	0.77***	0.550*	mR
Pu-8	Czech Republic	1.09**		R[d]
Wu-0	Germany	1.46***	0.627**	mR
GR-6	Michigan	1.56***	0.657†	mR
Cvi-0	Cape Verde Islands	1.45***	0.670	mR
Ang-0	Belgium	1.28***	0.945	R
Col-0	Columbia, SC	1.71***	1	R
Kas-1	Kashmir, India	1.91***	1.010	R
Ct-1	Catania, Italy	1.10***	1.072	R
Bur-0	Burren, Ireland	2.12***	1.099	R
Mt-0	Libya	1.94***	1.099	R
Bla-2	Spain	2.21***	1.104	R
Gott-20	Germany	2.07***	1.143	R
Pog-0	B. C., Canada	2.52***	1.392†	sR
RLD-1	Russia	2.56***	1.547**	sR
Co-1	Portugal	2.97***	1.556†	sR
Tsu-0	Japan	3.77***	1.611***	sR
FM-17	Ithaca, NY	ND		
HS-12	Massachusetts	ND		
C2-1	N. Carolina	ND		
Tamm-17	Finland	ND		
G2-1	N. Carolina	ND		
D2-9	N. Carolina	ND		

ND, not determined.

[a] Resistance is log of growth for the strain without *avrRpt2* minus that for the strain with *avrRpt2*. ANOVA significance values determine resistance or susceptibility designation. †$P < 0.10$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

[b] Relative resistance is the resistance value for the accession divided by that for Columbia in the same experiments. ANOVA P values are indicated as for resistance.

[c] *avrRpt2*-dependent resistance phenotypic categories are S, susceptible; mR, mildly resistant; R, resistant; and sR, strongly resistant. Assignment of categories was determined by the relative resistance values.

[d] Pu-8 is significantly *Rps2*-dependent resistant, but its relative resistance level could not be accurately determined.

for positive selection on plant R genes. In contrast, synonymous and nonsynonymous divergence reveals no evidence for positive selection on *Rps2* (framed exposed residues, $K_a = 0.033$, $K_s = 0.12$); functional constraint is evident for all categories of LRR region amino acid residues (Table 3). Contingency tables comparing synonymous and replacement polymorphism and divergence (McDonald and Kreitman 1991) also provide no indication of selectively driven protein evolution (Table 3).

Evidence for balancing selection at *Rps2*: Figure 2 shows a parsimony tree inferred from silent and nonsynonymous polymorphism and divergence, with accession name and *avrRpt2*-dependent resistance phenotype shown for each allele. The *Rps2* gene tree reveals the presence of two highly supported major clades. This haplotype structure is evident for synonymous as well as amino acid replacement polymorphisms, but only for polymorphisms falling in the middle of the coding sequence. Tests for nonrandom associations between all pairs of nonsingleton polymorphisms (Figure 3) reveal that linkage disequilibrium is clustered within a central segment of the *Rps2* coding sequence. Indeed only polymorphisms in this segment show significant linkage disequilibrium after correction for multiple tests of association. Outside of this central segment of the *Rps2* coding sequence, the data reveal little haplotype structure (Figures 1 and 3).

Sliding window analysis of nucleotide diversity between the two major clades (Figure 4) shows a peak of

TABLE 2

Levels of variability across the *Rps2* locus and among *RPS2* functional domains

	Position in alignment	No. sites			Divergence[b]		Polymorphism[b]	
		Total	Silent[a]	Replacement[a]	Silent	Replacement	Silent	Replacement
Entire region	1–4461	3755	1651.9	2097.1	142 (0.0904)	55 (0.0280)	33 (0.0047)	20 (0.0023)
Excluding *A. lyrata*[c]		4216	2102.9	2113.1	—	—	38 (0.0042)	20 (0.0023)
5′ noncoding	1–1137	802	802	0	62 (0.0783)	—	7 (0.0017)	—
Excluding *A. lyrata*[c]		975	975	0	—	—	8 (0.0015)	—
Coding	1138–3879	2709	610.9	2098.1	71 (0.1266)	55 (0.0280)	20 (0.0092)	20 (0.0023)
Excluding *A. lyrata*[c]		2727	613.9	2113.1	—	—	20 (0.0091)	20 (0.0023)
LRR	2654–3760	1095	248.9	846.1	26 (0.1203)	26 (0.0333)	8 (0.0087)	14 (0.0041)
Other structural[d]		225	57.0	168.0	5 (0.0877)	1 (0.0064)	0 (0.0)	1 (0.0009)
Nonstructural[d]		1389	305.0	1084.0	40 (0.1390)	28 (0.0272)	12 (0.011)	5 (0.0012)
3′ noncoding	3880–4461	241	241	0	9 (0.0392)	—	6 (0.0035)	—
Excluding *A. lyrata*[c]		514	514	0	—	—	10 (0.0036)	—

[a] Silent sites include noncoding sites as well as synonymous sites in the coding sequence.

[b] Divergence is shown as number of fixed differences (per site average difference); polymorphism is number of segregating sites (per site nucleotide diversity π).

[c] Results are shown for an alignment not including the consensus *A. lyrata* sequence, allowing polymorphism to be ascertained at a greater number of sites, for 5′ noncoding, coding, and 3′ noncoding regions.

[d] Other structural regions defined in Mindrinos *et al.* 1994 (and their positions in alignment) include leucine zipper (1228–1293), nucleotide-binding site (1684–1710, 1912–1926, 2128–2208), transmembrane domain (2158–2208), and hydrophobic N terminus (1156–1203). Nonstructural regions are made up of the coding sequence excluding the leucine-rich repeat region and other structural regions.

```
Position                1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4
            1 1 2 5 7 7 0 1 2 4 5 6 2 3 3 3 4 4 4 4 4 5 5 5 5 5 6 6 6 7 7 7 7 8 9 0 0 0 0 0 0 2 2 3 4 5 6 8 8 9 9 9 9 0 0 2 2 3
            2 8 6 5 4 9 4 0 7 5 6 0 3 7 8 9 5 5 6 6 9 1 5 7 8 9 8 8 8 0 2 7 7 3 5 5 6 6 7 8 8 9 3 4 9 8 0 4 9 9 0 5 9 2 2 3 7 3
            4 9 2 6 7 8 8 5 9 1 6 1 2 3 5 5 1 5 0 6 9 4 4 8 0 8 3 4 8 9 0 2 4 8 5 1 3 6 1 5 6 8 7 9 6 3 2 7 6 8 4 1 5 0 6 3 2 9
A. lyrata   A T G A T - A A A G A C C C G T C C C C T T C C G G G T T C C C C T G C G T A G G T A T G T C A G G T T A - - - A G
A. thaliana accessions
AB-7        G . . . . A . . . A . . A . A . . . . . . . . . T . . . . C . A . . G A . C . . . . C . A . C . G . . . . T A A G . .
FM-17       G . . . . A . . . A . . A . A . . . . . . . . . T . . . . C . A . . G A . C . . . . C . A . C . G . . . . T A A G . .
GR-6        G . . . . A . . . A . . A . A . . . . . . . . . T . . . . C . A . . G A . C . . . . C . A . C . G . . . . T A A G . .
HS-12       G . . . . A . . . A . . A . A . . . . . . . . . T . . . . C . A . . G A . C . . . . C . A . C . G . . . . T A A G . .
Ang-0       G . . . . A . . . A . . A . A . . . . . . . . . T . . T . C . . . . G A . C . . . . C . . . C . G . C . . T G A A . .
Mt-0        G . . . . A . . . A . . A . A . . . . . . . . . T . . T . C . . . . G A . C . . . . C . . . C . G . C . . T G A A . .
G2-1        G . . . . A . . . A . . A . A . . . . T . . . . T . . . . C . . . . G A . C . . . . C . . . G . C . . . . T A A G . .
D2-9        . . . . . A . . . A . . A . A . . . . T . . . . T . . . . C . . . . G A . C . . . . C . . . G . C . . . . T G A G T .
Tsu-0       . . . . . A . . . A . . A . A . . . . T . . . . T . . . . C . . . . G A . C . . . . C . . . G . C . . . . T G A G T .
RLD-1       . . . . G A . . . A . . A . A . . . . . . . . . T . . T . C . . . . G A . C . . . . C . . . C . G . C . . T G A A . .
Bla-2       . . . . . A . . . A . . A . A . . . . . . . . . T . . . . C . . . . G A . C . . . . C . A . C . G . . . . T A A G . .
C2-1        . . . . . A . . . A . . A . A . . . . . . . . . T . . . . C . . . . G A . C . . . . C . A . C . G . . . . T A A G . .
Cvi-0       . . . . . A . . . A . . A . A . . . . . . . . . T . . . . C . . . . G A . C . . . . C . A . C . C . . . . T G A G . .
Gott-20     . . . . . A . . . A . . A . A . . . . . . . . . T . . . . C . . . . G A . C . . . . C . A . C . G . . . . T A A G . .
Bur-0       . . . . . A . . . A . . A . A . . . . . . . . . T . . . . C . . . . G A . C . . . . C . . . C . G . . A . T A A G . .
Tamm-17     . . . . . A . . . A . . A . A . . . . . . . . . T . . . . C . . . . G A . C . . . . C . . . C . G . . A . T A A G . .
Co-1        . . . . . A . . . A . . A . A . . . T . . . . . T . . . . C . . . . G A . C . . . . C . . . C . G A . . . T A A G . .
Kas-1       . . . . . T . . . A . . A . A . . . . . . . . . T . . . . C . . . . G A . C . . . . C . . . C . C . . . . T G A G T .
Pu-8        . . . . . A . . . A . . A . A . . . . . . . . . T . . . . C . . . . G A . C . . . . C . . T C . C . . . A . G G G . .
Ct-1        . . . . G A . . . A . . A . A . . . . . . . . . T . . . . C . . . . G A . C . . . . C . . T C . C . . . A . G G G . .
Pog-0       . . . . . A . T . A . A A . A . . . . . . . . . T . . . . C . . . . G A . C . . . . C . . . C . G . . . . T A A G . C
Yo-0        . . . C . A . T . A . . A . A . . . . . . . . . T . . . C C . . . . G A . C . . . . C C . . C . G . . . . T A A G . .
Knox-2      . . . . . A . . . . C . T T T C T A . . C C T . A A . . . T . . A . . . . A G C A . . . . C . C . . . . T G A G . .
Wu-0        . . . . . A . . . . C . T T T C T A . . C C . . A A . . . T . A A . . T . A G . A . . . T C G C . . . . T G A G . .
BG-4        . G A . . A C . . . C . T T T C T A . . C C T . A A . . . T . . A . . . . A G C A . . . . C . C . . . . T G A G . .
Zu-0        . G A . . A C . G . C . T T T C T A . . C C T . A A . . . T . . A . . . . A G C A . . . . C . C . . . . T G A G . .
Po-1        . G A . . A C . G . C . T T T C T A . . C C T . A A . . . T . . A . . . . A G C A . . . . C . C . . . . T G A G . .

Amino acid polymorphisms
                        I C   S       S H         H     V V     P S S         E G G A E E   N L R
                        4 1   1       4 4         4     5 5     5 5 5         6 6 6 6 6 7   7 7 8
                        8 0   5       1 3         7     1 1     2 4 4         4 4 4 5 9 0   7 8 3
                        V 4   4       9 9         2     5 5     7 4 5         4 9 9 3 9 3   8 5 3
                          Y   Y       P N         Y     L A     H R Y         G Q E V D D   K V T
```

FIGURE 1.—*Rps2* region polymorphic sites. Shown are positions in the alignment and the bases of *A. lyrata* (consensus) and *A. thaliana* sequences. Periods represent ancestral bases inferred from *A. lyrata*, and bases indicate derived polymorphic mutations. Amino acid replacement mutations are indicated relative to the wild-type Columbia accession (GenBank accession no. AL049483), which is identical to Bla-2, C2-1, and Gott-20. The region encoding the *RPS2* leucine-rich repeat (LRR) region is indicated by the line above the amino acid replacement mutations, with the thicker line indicating its 5′ half.

TABLE 3

Polymorphism and divergence within the *RPS2* leucine-rich repeat (LRR) region

	Synonymous sites					Nonsynonymous sites				
	N	*S*	π	*D*	K_s	*N*	*S*	π	*D*	K_a
LRR	248.94	8	0.00867	26	0.1203	846.06	14	0.00405	26	0.03331
Exposed[a]	43.71	2	0.00886	5	0.1313	166.29	5	0.00644	6	0.03786
Structural[a]	103.69	4	0.01105	10	0.1089	274.31	2	0.00123	5	0.01890
Between[a]	101.54	2	0.00617	11	0.1244	405.46	7	0.00499	15	0.04120

N, no. of sites; *S*, no. of polymorphic sites; *D*, no. of fixed differences (see Table 2).

[a] LRR residue functional categories are Exposed, hypervariable framed residues; Structural, conserved residues throughout the LRR; and Between, nonframed nonconserved residues (Jones and Jones 1997; Bergelson *et al.* 2001).

silent polymorphism in the center of the coding sequence—the 300 bp 5′ of the region encoding the *RPS2* LRR region and in the 5′ half of the region encoding the *RPS2* LRR region itself (hereafter referred to as the 5′ LRR region)—corresponding to the region containing the cluster of polymorphic sites in linkage disequilibrium. Peak nucleotide diversity between the two major clades reaches $\pi_b = 0.086$ in the *Rps2* 5′ LRR region, a value approaching estimates of silent divergence between species. Clustering of silent polymorphism within this segment of the *Rps2* coding sequence results in significant heterogeneity in the ratio of poly-

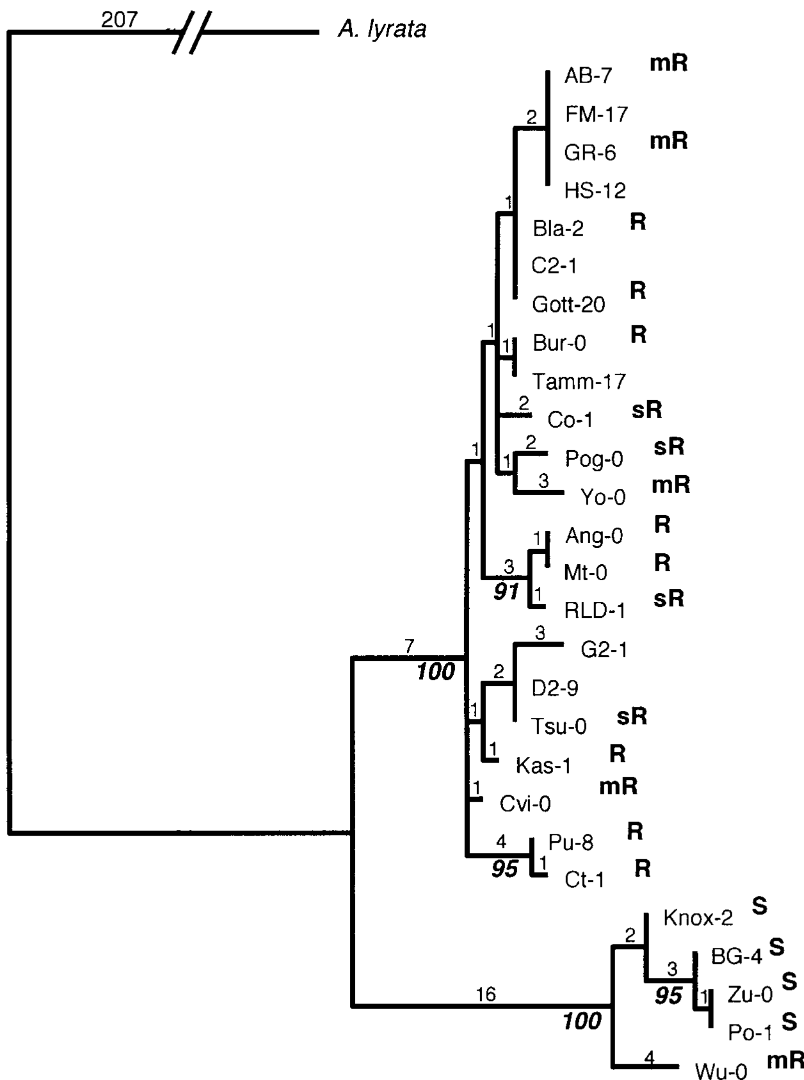

Figure 2.—Phylogeny of *Rps2* sequences based on parsimony analysis of silent, synonymous, and amino acid replacement variability. Accession names are indicated for each *Rps2* sequence with *avrRpt2*-dependent phenotype in boldface type (R, resistant; mR, mildly resistant; sR, strongly resistant; S, susceptible). The tree shown is one of three most parsimonious trees (length 265, consistency index 0.974) that differ only in the resistance (upper) clade. Numbers of mutations are shown above branches, with proportional branch length. Bootstrap values >90% are shown below branches (boldface italics).

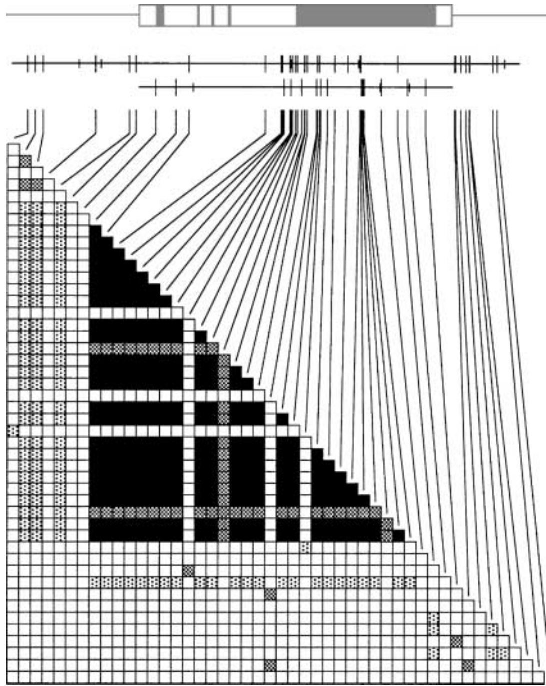

Figure 3.—Linkage disequilibrium between polymorphisms in the *Rps2* region. The *Rps2* region diagram shows the coding sequence (box) with *RPS2* functional regions (LZ, leucine zipper; NBS, nucleotide-binding site; LRR, leucine-rich repeat). On the horizontal lines below, singleton polymorphisms (small hash marks) and nonsingleton polymorphisms (sample frequency two or greater, large hash marks) are indicated for silent/synonymous polymorphisms (top line) and amino acid replacement polymorphisms in the coding sequence (bottom line). In the triangle at bottom, Fisher's exact test P values for each pair of nonsingleton polymorphic sites are indicated by shading, $P > 10^{-2}$ (white), $10^{-3} < P < 10^{-2}$ (stippled), $10^{-4} < P < 10^{-3}$ (shaded stipple), and $P < 10^{-4}$ (black). Only P values $<10^{-4}$ (black) remain significant after Bonferroni correction.

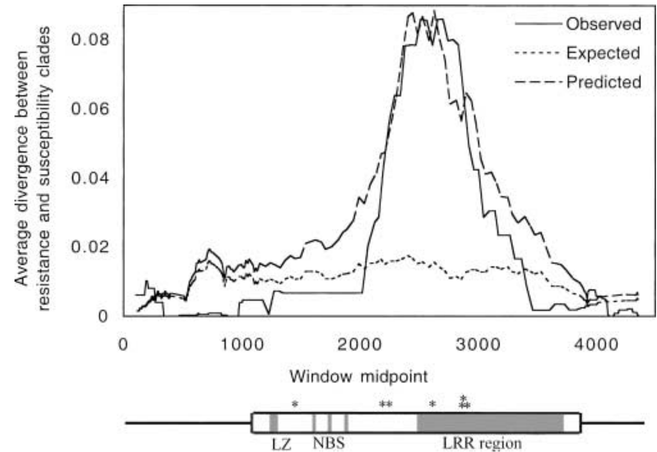

Figure 4.—Sliding window analysis of silent (noncoding and synonymous) divergence between resistance and susceptibility clades of *Rps2* alleles. Average numbers of pairwise differences per site within the window are shown with a solid line. Predicted levels under a coalescent model with selection and recombination (dashed line) assume that selection acts at the beginning of the LRR region (2654) and maintains *Rps2* polymorphism at frequency $22/27 = 0.81$, with independently estimated recombination rate 0.00057 and fitted mutation rate between selected alleles 0.0125 (see materials and methods). Expected levels under neutrality (dotted line) are calculated as divergence in the window times the ratio of averages across the region of polymorphism (π) and divergence, multiplied by the expected time to the most recent common ancestor for sample size 27 relative to expected average pairwise coalescence time, $2\Sigma_{i=2}^{27}(1/i(i-1))$. The window is 150 silent sites wide, slid by 10-site increments. Beneath the sliding window plot the corresponding functional regions of *RPS2* are shown, with amino acid differences between the clades indicated by asterisks (*).

morphism to divergence across the sequenced region (sliding window average G: entire region, $P \leq 0.004$; coding sequence, $P \leq 0.0014$; McDonald 1998). Variation at *Rps2*, therefore, is not compatible with an equilibrium model of selective neutrality in a panmictic population.

Alleles from resistant and susceptible accessions are not scattered throughout the *Rps2* gene tree, but are grouped together; therefore, we refer to the two major clades as the resistance (R) clade and susceptibility (S) clade. We tested for a significant association between *Rps2* sequence variation and *avrRpt2*-dependent resistance variation by analyzing differentiation (an F_{ST} estimator based on nucleotide diversities; Hudson *et al.* 1992b) between phenotypes. Overall differentiation between phenotypes is highly significant (S, mR, R, and sR; $F_{ST} = 0.52$, $P < 0.001$). Pairwise comparisons between phenotypes reveal significant differentiation between S and each of R, mR, and sR ($F_{ST} \geq 0.47$, $P \leq 0.019$), marginally significant differentiation between R and mR phenotypes ($F_{ST} = 0.12$, $P = 0.09$), and no significant differentiation for other comparisons between resistant phenotypes ($F_{ST} < 0.005$, $P > 0.3$). Thus, sequence variation at *Rps2* correlates with *avrRpt2*-dependent disease resistance, suggestive of causal links between the two (see discussion).

Geographic differentiation: In contrast to *avrRpt2*-dependent resistance, accessions from the same geographic region are scattered throughout the *Rps2* gene tree (Figure 2). We categorized accessions into five regions, (1) Eastern Europe, Asia, and Africa; (2) Central and Northern Europe; (3) Western and Southern Europe; (4) Eastern North America; and (5) Western North America, on the basis of the recent expansion of *A. thaliana* from Western Asia and Eastern Europe to its current worldwide distribution (Price *et al.* 1994; Sharbel *et al.* 2000). *Rps2* sequence variation reveals no differentiation among regions (overall $F_{ST} = 0.043$, $P = 0.3$; for all pairs of regions $F_{ST} \leq 0.14$, $P \geq 0.15$). In addition, *Rps2* variation does not differentiate North America from other continents ($F_{ST} = 0.043$, $P > 0.5$), revealing no evidence for a founder effect in the colonization of the Western hemisphere by *A. thaliana*. These

results are typical of studies of a single or few loci and moderate sample sizes in *A. thaliana* (Innan *et al.* 1996; Bergelson *et al.* 1998; Kawabe and Miyashita 1999).

DISCUSSION

Previously, Caicedo *et al.* (1999) found a high level of polymorphism at the *Rps2* locus and two highly divergent alleles, suggestive of balancing selection, but a statistical test (Tajima's D) could not reject selective neutrality. Here we find statistical evidence in support of the selection hypothesis and tentatively identify the *Rps2* 5′ LRR region as the target of selection. An *Rps2* sequence from sister species *A. lyrata* and a larger sampling of alleles allowed us to detect a clustering of polymorphism relative to divergence exceeding that possible under selective neutrality in a panmictic population. This result rules out the possibility that the region of high polymorphism is a mutational hotspot, since mutation rate heterogeneity would affect both polymorphism and divergence.

Our statistical confirmation of a peak of polymorphism should not be taken, in and of itself, as a strong refutation of neutral evolution. For example, Innan *et al.* (1996) identified a short highly diverged stretch in exon 4 of *Adh*, as well as in three adjacent sequence "blocks." While the authors argue in favor of balancing selection acting on exon 4 (owing to amino acid replacement differences between the alleles), they raise the possibility that population structure and history produced the "dimorphism" seen throughout the locus. Biallelic variation has also been found at several other loci in *A. thaliana* (Kawabe *et al.* 1997; Kawabe and Miyashita 1999; Stahl *et al.* 1999; Purugganan and Suddith 1999; Aguadé 2001; Hauser *et al.* 2001; Tian *et al.* 2002), adding to the appeal of a population structure hypothesis.

We favor balancing selection as an explanation for *Rps2* variation, on the basis of features of the data that distinguish our results from those of other studies that find biallelic variation but favor a population structure hypothesis. As indicated in the sliding window analysis (Figure 4), most of the variation is present in the coding segment of the *Rps2* gene and overlaps with the functional domain of the protein implicated in pathogen recognition. Seven amino acid replacement changes separate the R and S clades, four in the LRR region, and the suggestion that differences between *Rps2* allelic classes could be functional is consistent with a role of selection. Furthermore, accessions' *Rps2* allelic classes correspond closely with their resistance phenotypes. Since selection can act only if functionally distinct alleles exist, a correspondence between phenotype and genotype provides additional evidence in support of balancing selection. Others have also pointed to the importance of possible functional differences distinguishing diverged alleles. For example, Hauser *et al.* (2001) found two divergent alleles across part of the region in their analysis of polymorphism in *Glabrous1* (*Gl1*), a candidate gene for leaf trichome density variation; they argued against selection because the divergence was not in the coding region of the gene and variation in trichome density did not correlate with *Gl1* sequence variation. Kawabe *et al.* (2000) found divergent alleles of the cytosolic phosphoglucose isomerase (*PgiC*) and favored balancing selection because the alleles produced distinct allozymes (but note that phenotypic properties of the allozymes were not investigated).

Balancing selection is expected to lead to a signature in which neutral variation accumulates between the alleles surrounding the site(s) under selection. This signature is a simple manifestation of the genealogical correlation of tightly linked sites: as a balanced polymorphism becomes old, so too do the genealogical ages of sites tightly linked to it. In Hudson and Kaplan's (1988) coalescent treatment of balanced polymorphism, the physical scale of neutral polymorphism linked to the site under selection is, to a first approximation, determined by a balance between the origination of new neutral mutations (governed by the scaled neutral mutation rate, $4N_eu$, where N_e is the effective population size and u is the neutral mutation rate per site per generation) and the decay of the linkage disequilibrium between these mutations and the site under selection (governed by the scaled recombination rate, $4N_er$, where r is the per generation recombination rate between adjacent sites). Even for a highly self-fertilizing species, balancing selection can be expected to produce a relatively sharp peak of neutral polymorphism linked to a site under selection (Nordborg *et al.* 1996; Nordborg 1997). Based on available genetic and population genetic estimates of mutation and recombination rates in *A. thaliana* (Tian *et al.* 2002; see materials and methods and Figure 4), the peak of polymorphism seen at *Rps2* is compatible with theoretical predictions for a balanced polymorphism at the 5′ end of the region of the gene that encodes the *RPS2* LRR region (Figure 4).

We note that the balancing selection analysis is based on a constant-size panmictic population model and does not take into account departures from this model in the demographic history of *A. thaliana*. Nevertheless, given that the peak of polymorphism is restricted to within the *Rps2* coding sequence, that polymorphisms within the peak are not in linkage disequilibrium with polymorphisms outside it, and that significant linkage disequilibrium is rarely observed between loci in *A. thaliana* (Innan *et al.* 1997; Nordborg *et al.* 2002), we can identify the region of the peak, which includes the N-terminal half of the *RPS2* LRR region, as the target of natural selection.

In our balancing selection analysis, the best-fit mutation rate between allelic classes was found to be equal to 0.01, *i.e.*, on the order of one-hundredth the rate

```
                                        PK                  506
AENWRQ  A  |LVISLLD  | NR IQTL  PEK                 529
LI CPK  L  |TTLMLQQ  | NSSLKKI  PTGF                553
FMHMPV  L  |RVLDLSF  | TS ITEI  PLS                 576
IKYLVE  L  |YHLSMSG  | TK ISVL  PQE                 599
LGNLRK  L  |KHLDLQR  | TQFLQTI  PRDA                624
ICWLSK  L  |EVLNLYY  | SYAGWELQSFGEDEAE             654
LG FAD  L  |EYLENLT  | TLGITVLSLETLKTLFE            685
FGALHKH I  |QHLHVEE  | CNDLLYFNLPS                  711
LTNHGRN L  |RRLSIKS  | CHDLEYLVTPADFE               740
NDWLPS  L  |EVLTLHS  | LHNLTRVWGNSVS                767
QDCLRN  I  |RCINISH  | CNKLKNVSW                    791
VQKLPK  L  |EVIELFD  | CREIEELISEHESPSVED           823
PTLFPS  L  |KTLRTRD  | LPELNSI LPS                  847
RFSFQK  V  |ETLVITN  | CPRVKKL  PFQ                 871

a--L--  L  |E-L-L--  | C--L--L   P--
           |+        | T
```

FIGURE 5.—The *RPS2* LRR region, with polymorphic mutations. The amino acid sequence taken from JONES and JONES (1997) is shown, and codon number of the rightmost residue in each row is shown on the right. Residues matching the LRR consensus (at bottom) are shown in boldface type, and the vertical lines bracket the structural motif frame. Residues that differ between the resistance and susceptibility clades are indicated in red, and residues that differ in association with phenotypic changes within the clades are indicated in blue.

of neutral coalescence ($1/N_e$). Higher mutation rates between the selected alleles would lead to more recent common ancestry between them, and if large enough may not result in an observable peak of polymorphism even with balancing selection. Many kinds of mutations can cause loss of function; therefore the rate of origination of new susceptibility alleles might be expected to be quite high. An ancient balanced polymorphism between a resistance and a susceptibility allele would imply that selection favors one susceptibility allele over others and that the rate of origination of this particular susceptibility allele is low. Alternative resistance alleles, on the other hand, might be expected to have a low rate of origination. Thus, the observation of a signature of selection between the two major *Rps2* clades is consistent with the hypothesis that the two major allelic classes of *Rps2* contain functional resistance alleles. Indeed, BANERJEE *et al.* (2001) showed that the susceptibility allele of Po-1 is partially functional against *avrRpt2* when in the Col-0 genetic background. Note that the designation of resistance or susceptibility in this study is based only on the ability to recognize one specific avirulence gene, *avrRpt2*. We propose that the alleles represented by the *Rps2* resistance and susceptibility clades encode distinct specificities against natural pathogens in wild populations. The recent finding of infection by *P. syringae* in natural populations of *A. thaliana* (JAKOB *et al.* 2002) makes this a realistic possibility.

Rps2 exhibits marked sequence variability in association with phenotypic variation. Seven of the nine phenotypic changes that would be inferred by simply mapping phenotypes onto the *Rps2* gene tree are associated with amino acid polymorphisms, six with polymorphisms in the LRR region (Figures 1 and 2). Polymorphisms that distinguish the R and S clades are found upstream of the LRR region (not shown) and in nonconserved residues between LRR frames (Figure 5); these changes could confer phenotypic variation that is maintained by natural selection (ELLIS *et al.* 1999; LUCK *et al.* 2000). Polymorphisms associated with other phenotypic changes on the tree include framed solvent-exposed residues and conserved residues between frames (Figure 5). While we cannot rule out the possibility that changes at other loci contribute to phenotypic variation in these accessions, we suggest that these amino acid polymorphisms should be candidates for further study of *RPS2* function (AXTELL *et al.* 2001). Moreover, besides conferring phenotypic variation that is maintained by selection, hypervariability of amino acid residues N-terminal to the LRR region and in the N-terminal half of the LRR region may be consistent with diversifying selection on *Rps2*.

Previous studies have found evidence for rapid adaptive substitution rates in LRR region solvent-exposed residues among R gene paralogs (MEYERS *et al.* 1998; BERGELSON *et al.* 2001; HOLUB 2001; MONDRAGÓN-PALOMINO *et al.* 2002). In contrast, previous studies have not found evidence for positive selection at two Arabidopsis R genes that exhibit signatures of balancing selection, *Rpm1* (STAHL *et al.* 1999) and *Rps5* (TIAN *et al.* 2002). The possibility that the major alleles of *Rps2* represent a functional balanced polymorphism suggests that the maintenance of variation by natural selection may be a general feature of R gene evolutionary dynamics. At *Rps2*, we do not find evidence for adaptive protein evolution between species, but we do observe marked amino acid variability that could be consistent with diversifying selection. It remains to be seen whether even faster evolving R genes can also support balanced polymorphisms, as exemplified by genes of mammalian major histocompatibility complex (HUGHES and NEI 1988) and plant self-incompatibility loci (CLARK 1993).

We thank A. Berry, I. Cetl, M. Nachman, G. Robellen, O. Savolainen, and J. Winterer for collecting *A. thaliana* seeds, as well as the *Arabidopsis* Biological Resource Center at Ohio State University for providing seeds of *A. thaliana* accessions. F. Ausubel provided seed of *rps2-101C*. We acknowledge the assistance of M. Aguadé and the reviewers who provided careful and helpful reviews. This work was funded by a Sloan Foundation/National Science Foundation Fellowship in Molecular Evolution and University of Georgia Faculty Research grant to R.M., a Sloan Foundation/Department of Energy Fellowship in Computational Molecular Biology to E.A.S., and a Packard Fellowship and National Institutes of Health awards GM-57994 and GM-62504 to J.B.

LITERATURE CITED

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.

AXTELL, M. J., T. W. MCNELLIS, M. B. MUDGETT, C. S. HSU and B. J. STASKAWICZ, 2001 Mutational analysis of the Arabidopsis *RPS2* disease resistance gene and the corresponding *Pseudomonas syrin-*

gae avrRpt2 avirulence gene. Mol. Plant-Microbe Interact. **14:** 181–188.

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.

Bent, A. F., B. N. Kunkel, D. Dahlbeck, K. L. Brown, R. Schmidt *et al.*, 1994 *RPS2* of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. Science **265:** 1856–1860.

Bergelson, J., E. A. Stahl, S. Dudek and M. Kreitman, 1998 Genetic variation within and among populations. Genetics **148:** 1311–1323.

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

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.

Breyne, P., D. Rombaut, A. Van Gysel, M. Van Montagu and T. Gerats, 1999 AFLP analysis of genetic diversity within and between *Arabidopsis thaliana* ecotypes. Mol. Gen. Genet. **261:** 627–634.

Burdon, J. J., 1987 *Diseases and Plant Population Biology*. Cambridge University Press, Cambridge, UK.

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.

Clark, A. G., 1993 Evolutionary inferences from molecular characterization of self-incompatibility alleles, pp. 79–108 in *Mechanisms of Molecular Evolution*, edited by N. Takahata and A. G. Clark. Sinauer Associates, Sunderland, MA.

Dong, X., M. Mindrinos, K. R. Davis and F. M. Ausubel, 1991 Induction of Arabidopsis defense genes by virulent and avirulent *Pseudomonas syringae* strains and by a cloned avirulence gene. Plant Cell **3:** 61–72.

Ellis, J., P. Dodds and T. Pryor, 2000 Structure, function and evolution of plant disease resistance genes. Curr. Opin. Plant Biol. **3:** 278–284.

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.

Flor, H. H., 1956 The complementary genic systems in flax and flax rust. Adv. Genet. **8:** 29–54.

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

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.

Holsinger, K. E., and R. J. Mason-Gamer, 1996 Hierarchical analysis of nucleotide diversity in geographically structured populations. Genetics **142:** 629–639.

Holub, E. B., 2001 The arms race is ancient history in Arabidopsis, the wildflower. Nat. Rev. Genet. **2:** 516–527.

Hudson, R. R., 1993 The how and why of generating gene genealogies, pp. 23–36 in *Mechanisms of Molecular Evolution*, edited by N. Takahata and A. G. Clark. Sinauer Associates, Sunderland, MA.

Hudson, R. R., and N. L. Kaplan, 1988 The coalescent process in models with selection and recombination. Genetics **120:** 831–840.

Hudson, R. R., D. D. Boos and N. L. Kaplan, 1992a A statistical test to detect geographic subdivision. Mol. Biol. Evol **9:** 138–151.

Hudson, R. R., M. Slatkin and W. P. Madison, 1992b Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

Hughes, A. L., and M. Nei, 1988 Pattern of nucleotide substitution at major histocompatibility complex class-I loci reveals overdominant selection. Nature **335:** 167–170.

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.

Innan, H., R. Terauchi and N. T. Miyashita, 1997 Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. Genetics **146:** 1441–1452.

Jakob, K., E. M. Goss, H. Araki, T. Van, M. Kreitman *et al.*, 2002 *Pseudomonas viridiflava* and *P. syringae*—natural pathogens of *Arabidopsis thaliana*. Mol. Plant-Microbe Interact. **15:** 1195–1203.

Jones, D. A., and J. D. G. Jones, 1997 The role of leucine-rich repeat proteins in plant defenses. Adv. Bot. Res. **24:** 89–167.

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.

Kunkel, B. N., 1996 A useful weed put to work: genetic analysis of disease resistance in *Arabidopsis thaliana*. Trends Genet. **12:** 63–69.

Kunkel, B. N., A. F. Bent, D. Dahlbeck, R. W. Innes and B. J. Staskawicz, 1993 *RPS2*, an Arabidopsis disease resistance locus specifying recognition of *Pseudomonas syringae* strains expressing the avirulence gene *avrRpt2*. Plant Cell **5:** 865–875.

Leister, R. T., and F. Katagiri, 2000 A resistance gene product of the nucleotide binding site-leucine rich repeats class can form a complex with bacterial avirulence proteins *in vivo*. Plant J. **22:** 345–354.

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.

McDonald, J. H., 1998 Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. Mol. Biol. Evol **15:** 377–384.

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

McDowell, J. M., M. Dhandaydham, T. A. Long, M. G. M. Aarts, S. Goff *et al.*, 1998 Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of Arabidopsis. Plant Cell **10:** 1861–1874.

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

Mindrinos, M., F. Katagiri, G.-L. Yu and F. M. Ausubel, 1994 The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell **78:** 1089–1099.

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

Noel, L., T. L. Moores, E. A. van der Biezen, M. Parniske, M. J. Daniels *et al.*, 1999 Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of Arabidopsis. Plant Cell **11:** 2099–2112.

Nordborg, M., 1997 Structured coalescent processes on different time scales. Genetics **146:** 1501–1514.

Nordborg, M., B. Charlesworth and D. Charlesworth, 1996 Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. Proc. Roy. Soc. Lond. Ser. B **263:** 1033–1039.

Nordborg, M., J. O. Borevitz, J. Bergelson, C. C. Berry, J. Chory *et al.*, 2002 The extent of linkage disequilibrium in the highly selfing species *Arabidopsis thaliana*. Nat. Genet. **30:** 190–193.

Parniske, M., K. E. Hammond-Kosack, C. Golstein, C. M. Thomas, D. A. Jones *et al.*, 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.

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*, edited by E. M. Meyerowitz and C. R. Somerville. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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, 1997 DnaSP version 2.0: a novel software

package for extensive molecular population genetics analysis. Comput. Appl. Biosci. **13:** 307–311.

Salmeron, J. M., G. E. D. Oldroyd, C. M. T. Rommens, S. R. Scofield, H. S. Kim *et al.*, 1996 Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell **86:** 123–133.

Sharbel, T. F., B. Haubold and T. Mitchell-Olds, 2000 Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and post-glacial colonization of Europe. Mol. Ecol. **9:** 2109–2118.

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 **292:** 661–667.

Swofford, D., 1996 *PAUP: Phylogenetic Analysis Using Parsimony (and Other Methods)*, Version 4. Sinauer Associates, Sunderland, MA.

Tao, Y., F. Yuan, R. T. Leister, F. M. Ausubel and F. Katagiri, 2000 Mutational analysis of the *Arabidopsis* nucleotide binding site-leucine-rich repeat resistance gene *RPS2*. Plant Cell **12:** 2541–2554.

Thomas, C. M., D. A. Jones, M. Parniske, K. Harrison, P. J. Balint-Kurti *et al.*, 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.

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

Whalen, M. C., R. W. Innes, A. F. Bent and B. J. Staskawicz, 1991 Identification of *Pseudomonas syringae* pathogens of Arabidopsis and a bacterial locus determining avirulence on both Arabidopsis and soybean. Plant Cell **3:** 49–59.

Yu, G.-L., F. Katagiri and F. M. Ausubel, 1993 Arabidopsis mutations at the *RPS2* locus result in loss of resistance to *Pseudomonas syringae* strains expressing the avirulence gene *avrRpt2*. Mol. Plant Microbe Interact. **6:** 434–443.

Communicating editor: M. Aguadé