DOI: 10.1534/genetics.105.047290

Unique Evolutionary Mechanism in *R*-Genes Under the Presence/Absence Polymorphism in *Arabidopsis thaliana*

Jingdan Shen,*[,1] Hitoshi Araki,†[,1] Lingling Chen,* Jian-Qun Chen* and Dacheng Tian*[,2]

**State Key Laboratory of Pharmaceutical Biotechnology, Department of Biology, Nanjing University, Nanjing 210093, China and*
†Department of Zoology, Oregon State University, Corvallis, Oregon 97331

Manuscript received June 27, 2005
Accepted for publication November 1, 2005

ABSTRACT

While the presence/absence polymorphism is commonly observed in disease resistance (*R*-) genes in Arabidopsis, only a few *R*-genes under the presence/absence polymorphism (R-P/A) have been investigated. To understand the mechanism of the molecular evolution of R-P/A, we investigated genetic variation of nine R-P/A in *A. thaliana* from worldwide populations. The number of possessed *R*-genes varied widely among accessions (two to nine, on average 4.3 ± 1.6/accession). No pair of accessions shared the same haplotype, and no clear geographic differentiation was observed with respect to the pattern of presence/absence of the *R*-genes investigated. Presence allele frequencies also varied among loci (25–70%), and no linkage disequilibrium was detected among them. Although the LRR region in regular *R*-genes is known to be highly polymorphic and has a high K_a/K_s ratio in *A. thaliana*, nucleotide sequences of this region in the R-P/A showed a relatively low level of genetic variation ($\pi = 0.0002$–0.016) and low K_a/K_s (0.03–0.70, <1). In contrast, the nucleotide diversities around the deletion junction of R-P/A were constantly high between presence and absence accessions for the *R*-genes ($D_{xy} = 0.031$–0.103). Our results suggest that R-P/A loci evolved differently from other *R*-gene loci and that balancing selection plays an important role in molecular evolution of R-P/A.

MOST of the disease resistance (*R*) genes in plants share partial structures and genetic similarity, even though their target organisms of resistance are quite diverged (Belkhadir *et al.* 2004). However, genomic arrangements and distribution of the *R*-genes vary notably among species, and even within species (Hammond-Kosack and Jones 1997; Zhou *et al.* 2004). Many *R*-genes compose gene families, which are often physically clustered on a plant genome. *Rpp1*, *Rpp5*, *Rpp8*, and *Rps4* are in this category in Arabidopsis (Botella *et al.* 1998; McDowell *et al.* 1998; Gassmann *et al.* 1999; Noel *et al.* 1999), as well as *Cf9/4* in tomato (Parniske *et al.* 1997) and *RGC2* in lettuce (Kuang *et al.* 2004). The *R*-gene at a single locus is also common in plants, *e.g.*, *Rpp13* and *Rps2* in Arabidopsis and *L* locus in flax (Caicedo *et al.* 1999; Ellis *et al.* 1999; Bittner-Eddy *et al.* 2000). However, some *R*-genes are under presence/absence polymorphism (P/A), which are present in some ecotypes but absent in the others within species. *Rpm1* and *Rps5* are examples in this category (Grant *et al.* 1998; Henk *et al.* 1999).

Bergelson *et al.* (2001) showed that the leucine-rich repeat (LRR) regions in *R*-genes are subject to rapid evolution but also represent a high level of polymorphism, indicating a complex evolutionary mechanism of *R*-genes (see also Michelmore and Meyers 1998). For example, *Rpp5* in Arabidopsis compose a gene cluster that contains nine homologs in an ecotype Ler and seven in Col (Noel *et al.* 1999), and the ratio of evolutionary rate in nonsynonymous sites to that in synonymous sites, K_a/K_s, was 2.14, while the nucleotide diversity among these clusters was up to 0.23. The high level of polymorphism in *R*-gene clusters were also reported at *Rpp8* and *Rpp1* in Arabidopsis (Botella *et al.* 1998; McDowell *et al.* 1998), *Cf4/9* in tomato (Parniske *et al.* 1997), and *RGC2* in lettuce (Kuang *et al.* 2004). At a single *R*-gene locus, the high levels of polymorphism were observed in *Rpp13* and *Rps2* in Arabidopsis (Caicedo *et al.* 1999; Bittner-Eddy *et al.* 2000; Mauricio *et al.* 2003; Rose *et al.* 2004) and in *L* gene in flax (Ellis *et al.* 1999), while the low level of polymorphism was observed only in *Rps4* in Arabidopsis (Gassmann *et al.* 1999) and *pita* in rice (Jia *et al.* 2003).

The high level of polymorphism and fast rate of evolution in *R*-genes may be explained by natural selection adapting to rapidly changing pathogen populations under the arms race hypothesis with gene-for-gene interaction (Michelmore and Meyers 1998; Bergelson *et al.* 2001). The coevolutionary conflict between the

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. DQ220470–DQ220672.

[1] These authors contributed equally to this work.

[2] *Corresponding author:* Department of Biology, Nanjing University, 22 Hankou Rd., Nanjing 210093, China. E-mail: dtian@nju.edu.cn

host and pathogen indicates that the attempts to evade host resistance by pathogens are followed by the developments of new detection capabilities by their hosts. A typical example of the coevolution between a host and a pathogen is *Rpp13* in Arabidopsis and *ATR13* in the pathogen *Peronospora parasitica* (Allen *et al.* 2004).

Interestingly, however, population genetics studies on a few *R*-genes under the presence/absence polymorphism (R-P/A) in Arabidopsis exhibited a different pattern of evolution. Analyses on *Rpm1* and *Rps5* showed that the presence/absence polymorphism of these loci directly corresponded to the resistance phenotypes of individuals and that the origin of the presence/absence polymorphism was quite old (Stahl *et al.* 1999; Tian *et al.* 2002). These examples suggested that at least some R-P/A genes are under balancing selection, which contrasts to the expectation by the arms race hypothesis.

In this research, we studied the genetic variation of nine R-P/A loci in *Arabidopsis thaliana* to understand the general mechanism of molecular evolution of R-P/A. We identified seven R-P/A loci in *A. thaliana* and investigated their genetic variations, including two known R-P/A (*Rpm1* and *Rps5*). The allelic and nucleotide diversities of these R-P/A were surveyed in >33 accessions of *A. thaliana* from worldwide populations. We also investigate the nucleotide diversity of the flanking sequences in six R-P/A loci and discuss the origin of R-P/A. This is the first study that systematically evaluates the effect of balancing selection on many R-P/A in *A. thaliana.*

MATERIALS AND METHODS

Determination of R-P/A: Except for *Rpm1* and *Rps5*, the other seven *R*-genes as R-P/A in Arabidopsis were screened by two approaches. In the first approach, we used all single *R*-genes without closely a related paralog (<30% nucleotide diversity) in the Col genome (Meyers *et al.* 2003) for the BLAST search in the incompletely sequenced genome Ler (http://www.arabidopsis.org/cereon) to find the candidates that are absent in Ler. Then we randomly selected 10 *R*-genes from these candidates to perform PCR amplification in Ler (primers were designed on the basis of the sequence information of *R*-genes in Col) to check if these genes are absent in Ler. In the second approach, we selected 20 single *R*-genes of the Col genome (excluding the candidates identified in the first approach) to perform PCR amplification in 20 Arabidopsis accessions to check if these genes are absent in some accessions. After genotyping by PCR, we determined the junction region of insertion/deletion in these genes confirmed with the P/A by multiple PCRs. The primer pairs used in multiple PCRs were designed to be every 1 kb away from both flanking regions of the *R*-gene until a positive PCR product of the junction region was obtained in absent accessions. Finally, we sequenced the PCR product to confirm the breaking points of the insertion/deletion of these *R*-genes.

Genotyping and DNA sequencing: Thirty-six to 50 accessions of *A. thaliana* and three individual samples of *A. lyrata* were randomly chosen from worldwide populations for this study (DNA of Arabidopsis accessions was a gift from Joy Bergelson) to perform the PCR amplification for determination of their genotypes in nine R-P/A. For each of these *R*-genes, a three-primer PCR was used (see Figure 1; two primers are designed in 3′- and 5′-flanking regions of the breaking point and one in the insertion of the *R*-gene) to give an alternative product for the present or the absent genotype or a two-primer PCR was used to give a positive product for the present but a negative one for the absent genotypes. On the basis of the results of genotyping, 9–18 accessions were randomly selected from the present genotypes for sequencing of the *R*-gene, in which ~800 bp of the LRR region was sequenced for the evaluation of diversity. Meanwhile, ~10 absent and 10 present accessions were also randomly chosen to sequence their 3′- and/or 5′-flanking regions of breaking point (~600 bp) for estimation of the age of the insertions. All sequencing reactions were run on an ABI 3100-Avant automated sequencer.

Data analysis: Multiple alignments of the DNA sequences were performed by Sequencer 4.0 (Gene Codes, Ann Arbor, MI). The rooted tree of genealogy was determined by the neighbor-joining method (PAUP 4.0; Swofford 2000) in which either *A. lyrata*, if an ortholog could be identified, or the closest paralog in Arabidopsis was used as outgroup gene. Parameter estimation of population genetics (such as nucleotide divergence between population [d_{xy}], average nucleotide diversity [π], and K_a/K_s) was carried out using the program DnaSP 4.0 (Rozas and Rozas 1999).

RESULTS

Identification of R-P/A: In the first method for identifying R-P/A, 130 of 149 single *R*-genes in *A. thaliana* Col (Meyers *et al.* 2003) were surveyed by the BLAST search in an incompletely sequenced genome of *A. thaliana* Ler. Nineteen candidate loci were identified by this method, and 10 were selected for further identification. Two of 10 loci were recognized to have the *R*-gene in both accessions by the PCR method (*i.e.*, no R-P/A), but 8 loci remained as R-P/A candidates. The second approach provided two candidate loci of 20 single *R*-gene loci investigated. Including *Rpm1* and *Rps5* (Grant *et al.* 1998; Henk *et al.* 1999), we identified a total of 12 R-P/A candidate loci, and 9 of them were selected for further analysis (Table 1).

Of the nine R-P/A, five (At4g10780, At5g05400, At3g07040, At4g27220, and At1g12220) were coiled-coil motif (CC)–nucleotide binding site (NBS)–LRR and four (At5g49140, At5g45240, At5g18350, and At1g63870) were TIR-NBS-LRR genes (Meyers *et al.* 2003). The size of the CC-NBS-LRR genes was very similar (Table 1), ranging from 2625 to 2781 bp (2703 bp on average), but the size in four Toll/interleukin-1 receptor (TIR)–NBS–LRR genes varied from 4047 to 5384 bp (including introns; data not shown). All TIR types of *R*-genes contained introns but there was no intron in the CC type of *R*-genes.

To confirm that they were truly R-P/A, we sequenced the deletion junction of P/A (see materials and methods). We could identify six R-P/A loci, including the two known R-P/A (Figure 1). The efforts to determine the junction sequences in the other three genes failed due to the repeat sequences around these loci, which made the clear determination of the deletion junction difficult. Although the other three failed to identify the

TABLE 1

Summaries of insertions, *R*-genes, and their flanking sequences

Locus	R-P/A-1	-2	-3	-4	-5	-6	-7	*RPM1*	*Rps5*
Type	CC	TIR	CC	TIR	CC	TIR	TIR	CC	CC
Insertion									
Exon (bp)	2679	2943	2625	3582	2760	2439	3096	2781	2670
LRR sequenced (bp)	819	659	1011	1182	989	1032	1158	855	674
π in LRR[a]	0.00019	0.00018	0.0030	0.016	0.00035	0.00040	0.015	0.00069	0.0011
Weighted π[b]	0.00046	0.00028	0.0067	0.022	0.00092	0.0016	0.043	0.0010	0.0017
D_{xy} between species[c]	/	/	/	0.069	/	/	0.061	0.047	0.058
K_a/K_s	/	/	/	0.70	/	/	0.034	0.22	0.26
Flanking of junction									
Position to junction	−253, 341	−601, −1	−266, 636	−179, 272	/	/	/	−568, 534	−608, 697
Length (bp)	594	601	902	451	/	/	/	1102	1305
Total π[a]	0.02229	0.05523	0.01782	0.01436	/	/	/	0.02574	0.02097
π within P alleles, π_1	0.00208	0.00033	0.00097	0.00070	/	/	/	0.00235	0.00091
π within A alleles, π_2	0.01082	0.00338	0.00530	0.00000	/	/	/	0.00030	0.00411
$(\pi_1 + \pi_2)/\pi$	0.579	0.067	0.352	0.049	/	/	/	0.103	0.239
$E[(\pi_1 + \pi_2)/\pi]$[d]	0.700	0.731	0.694	0.763	/	/	/	0.745	0.722
D_{xy} between P/A[c]	0.037	0.103	0.031	0.043	/	/	/	0.048	0.038
Tajima's D[e]	2.23*	3.34**	1.58	2.28*	/	/	/	3.03**	2.47**

R-P/A-1–7 corresponds to loci At4g10780, At5g49140, At5g05400, At5g18350, At4g27220, At5g45240, and At1g63870, respectively. The relative position is labeled as − or + bp for 5′ and 3′ flanking regions, respectively, from the insertion/deletion junction (as 0). /, not analyzed because of the lack of data.

[a] π represents the average number of nucleotide differences (Nei 1987).

[b] Weighted π was calculated by weighting π by the proportion of presence alleles in the population (Table 2, Innan and Tajima 1997). This is an estimate of $\theta = 4N_e\mu$, where N_e is the effective population size and μ is the mutation rate.

[c] D_{xy} represents the average nucleotide diversity between two groups (x and y). We calculated D_{xy} between species (*A. thaliana* and *A. lyrata*) for LRR regions and between presence (P) and absence (A) alleles within species for flanking regions.

[d] Expected values of $(\pi_1 + \pi_2)/\pi$ were calculated on the basis of allele frequencies (Innan and Tajima 1997). Selective neutrality and no recombination were assumed.

[e] Tajima's D statistic (Tajima 1989) was obtained by DnaSP v. 4.0 (Rozas and Rozas 1999). * and ** represent significance levels of 5 and 1%, respectively.

sequences of the deletion junctions, they were likely to be R-P/A because the multiple PCR in our efforts among accessions for each gene gave very consistent results of the presence/absence genotype for each accession.

The architectures of the deletion junctions in R-P/A were very similar among loci (Figure 1). The insertions commonly contained a 4- to 7-kb-long fragment, including a 2.6-to 4.3-kb-long *R*-gene without any other

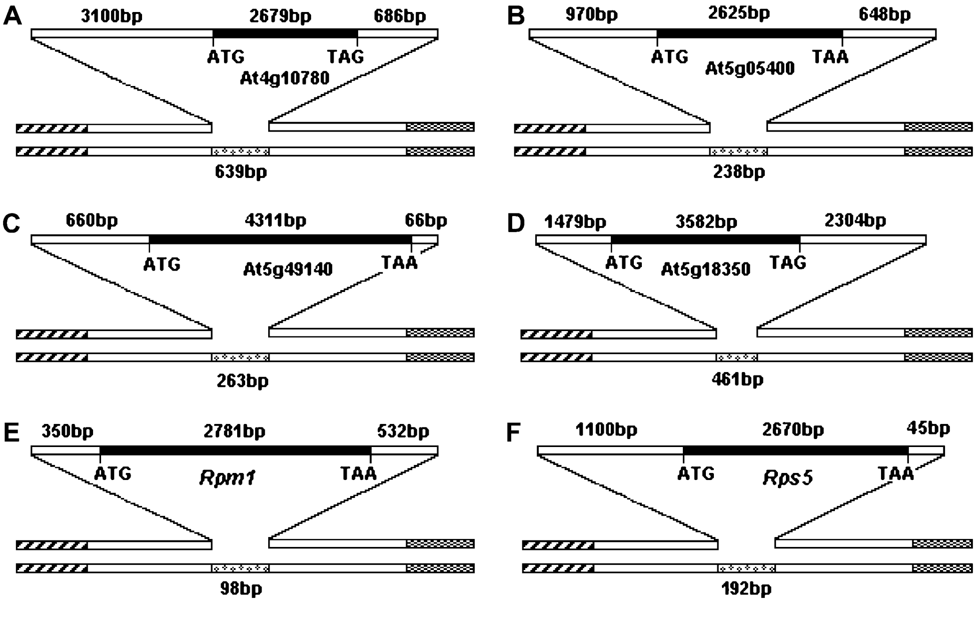

Figure 1.—Schematic comparison of *R*-gene presence and absence alleles. Insertions with the *R*-gene are represented for At4g10780 (A), At5g05400 (B), At5g49140 (C), At5g18350 (D), *Rpm1* (E), and *Rps5* (F). (E and F are drawn on the basis of reports by Henk *et al.* 1999 and Grant *et al.* 1998.) In all absence alleles examined, the *R*-genes were replaced by 98–~639-bp junk DNA of unknown origin.

TABLE 2

Distribution of allelic polymorphism in R-P/A among *A. thaliana* accessions

Accession	R-P/A-1	-2	-3	-4	-5	-6	-7	*Rpm1*	*Rps5*	sum
Col-0	+	+	+	+	+	+	+	+	+	9
Kas-0	+	+	+	+	+	+	+	−	−	7
Gr-24	+	+	+	+	−	−	−	+	+	6
Sorbo	+	+	−	+	−	−	+	+	+	6
Got-7	+	+	−	−	+	−	+	+	+	6
Lip-0	+	−	−	+	+	−	+	+	+	6
Pog-0	−	+	+	+	+	−	−	+	+	6
Bur-0	+	+	+	−	−	−	−	+	+	5
Yo-0	+	+	−	+	−	+	−	−	+	5
Ct-1	+	−	+	+	−	−	−	+	+	5
Gu-0	+	−	−	+	+	−	−	+	+	5
Sq-8	−	+	+	+	+	−	−	+	−	5
Wu-0	−	+	−	+	−	−	+	+	+	5
Zdr-6	−	−	+	−	+	−	+	+	+	5
Rf-4	+	+	−	+	−	+	−	−	−	4
Edi-0	+	+	−	+	−	−	−	+	−	4
Ang-0	+	+	−	−	−	−	+	−	+	4
Hr-5	−	+	+	−	−	−	−	+	+	4
Bla-2	−	+	−	+	−	−	+	−	+	4
Nd-1	−	−	+	+	−	−	+	−	+	4
Ms-0	−	−	+	+	−	−	−	+	+	4
Kz-9	−	−	−	+	−	−	+	+	+	4
Rrs-7	+	−	−	+	−	−	−	+	−	3
Ler-1	−	+	+	−	−	−	−	+	−	3
Up-14	−	+	−	+	−	+	−	−	−	3
Zu-0	−	+	−	−	+	−	−	−	+	3
Mt-0	−	−	+	+	−	−	+	−	−	3
Lov-1	−	−	+	+	−	−	−	+	−	3
Pu2-8	−	−	+	−	−	−	−	+	+	3
Kz-7	−	−	−	+	+	−	−	+	−	3
Nfc-5	−	+	−	+	−	−	−	−	−	2
Pu2-7	−	−	+	−	−	−	−	+	−	2
Ws-0	−	−	−	−	−	−	−	+	+	2
$F_{[\text{presence}]}$	15/36	30/46	20/45	28/40	19/50	12/48	13/37	29/43	28/44	
%	41.7	65.2	44.4	70.0	38.0	25.0	35.1	67.4	63.6	

Genotypes of these accessions are represented by + (presence) or − (absence). R-P/A-1–7 correspond to loci At4g10780, At5g49140, At5g05400, At5g18350, At4g27220, At5g45240, and At1g63870, respectively. Frequency and percentage of the presence allele in each locus are shown at the bottom.

ORF. The similar architectures of R-P/A indicate a common mechanism to create R-P/A. These R-P/A genes could be created by simple insertion/deletion or by transposable elements. The independent deletion events were proposed for the absence of *Rpm1* in Arabidopsis and in Brassica (Grant *et al.* 1998). An insertion event could be responsible for the presence of *Rps5* because a transposon (TAG2) target sequence and a perfect 17-bp reverse repeat were identified at 4 bp away from the deletion junction in the absence accessions. At exactly the same target site, *Rps5* and another 4-kb insert were uncovered in different accessions (Henk *et al.* 1999). Two independent inserts at the same site suggest that *Rps5* was created by a simple insertion event, probably by a transposable element. Except for *Rps5*, no transposon target sequence and reverse repeats could be identified around deletion junctions of the other R-P/A genes.

R-P/A allele frequency in population samples: Allelic diversities of the nine R-P/A were investigated in 36–50 accessions of *A. thaliana* (Table 2). The number of present *R*-genes per accession varied widely, ranging from two to nine (only in the reference accession Col). Of 36 accessions, 91.6% (33 accessions), which were completely genotyped, harbored an intermediate number of *R*-genes investigated (three to six genes of the nine loci). An average number of present *R*-genes among the 33 accessions was 4.3 (±1.6 SD). No single pair of accessions shared the same haplotype with respect to the nine R-P/A in these accessions, and no linkage disequilibrium between any pair of loci was detected (χ^2-test, $P > 0.075$). These results suggest no

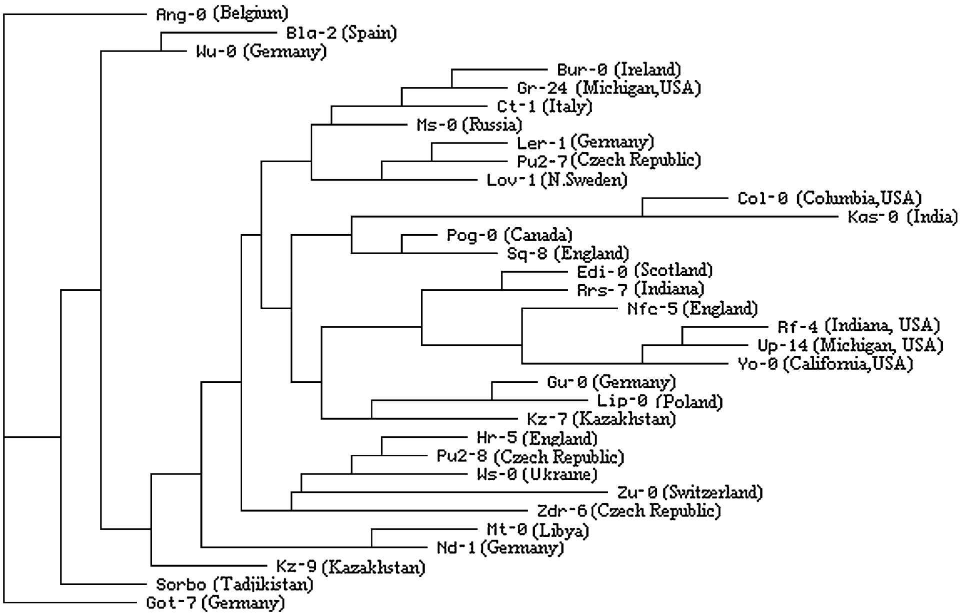

Figure 2.—Grouping of *A. thaliana* accessions with respect to the presence/absence of the nine R-P/A. This tree was constructed by the neighbor-joining method using PAUP*4.0 (Swofford 2000).

strong interaction among R-P/A loci investigated. Given allele frequencies of the nine R-P/A (Table 2), the expected number of identical haplotype pairs in 33 samples was calculated as 2.4, assuming no interaction and free recombination among R-P/A. The lack of identical haplotype pairs can be caused by diversifying selection among sets of R-P/A loci, while our observation of having no identical haplotype pairs did not significantly deviate from the neutral expectation ($P = 0.094$).

No clear geographic differentiation was observed in the pattern of presence/absence of the *R*-genes among the accessions (Figure 2). Although we had only two sets of accession pairs from the same population (Pu and Kz), these accessions showed no similarities within populations.

Allele frequencies in the nine R-P/A loci varied widely (Table 2). Within up to 50 accessions, R-P/A-6 (At5g45240) showed the lowest presence allele frequency (25.0%), while R-P/A-4 (At5g18350) showed the highest (70.0%). The average allele frequency of presence alleles was 50.1% (±16.6 SD). Further study is required for testing selective neutrality of R-P/A on the basis of allele frequency distribution. Such study must be conducted on randomly selected R-P/A loci with respect to allele frequency and on a higher number of loci that provide enough power to distinguish distributions predicted by different models (Wright 1931).

Nucleotide diversity of R-P/A: The LRR region in the *R*-genes, which is suggested to be involved in pathogen recognition, is known to be highly polymorphic (Michelmore and Meyers 1998; Bergelson *et al.* 2001). To evaluate the level of genetic variation of this region in R-P/A loci, we sequenced 659–1182 bp fragments in the LRR region of the nine R-P/A (Table 1). The observed nucleotide diversities (π, Nei 1987) were low in seven R-P/A loci (0.00018–0.00297), relative to those in the other two loci (0.0155 and 0.0152 in At5g18350 and At1g63870, respectively). Although nucleotide diversities in one allele class (*e.g.*, presence alleles) are expected to be lower than those in total (Innan and Tajima 1997), the average genetic diversity of the weighted π for the nine R-P/A (Table 1) was 0.011, which was still lower than that in the other *R*-genes in this species (π = 0.05–0.26, 0.088, >0.1, 0.08, and 0.013 in *Rpp5*, *Rpp1*, *Rpp1*3, *Rpp8*, and *Rps2*, respectively (Botella *et al.* 1998; Caicedo *et al.* 1999; Noel *et al.* 1999; Takahashi *et al.* 2002; Rose *et al.* 2004; see also Bergelson *et al.* 2001). In addition, the K_a/K_s ratio was <1 (0.03–0.70) in the LRR regions in four R-P/A for which homologs from an outgroup were available (Table 1), while many other *R*-genes have $K_a/K_s > 1$ (Bergelson *et al.* 2001). These results suggest that the evolutionary mechanism of R-P/A is not the same as that of other *R*-genes.

Nucleotide diversity around the deletion junctions: In *Rpm1* and *Rps5*, high levels of genetic diversity between presence and absence alleles were observed, as a consequence of balancing selection on these loci (Stahl *et al.* 1999; Tian *et al.* 2002). To examine whether this observation is general for R-P/A or not, we obtained 451- to 902-bp-long flanking sequences of the insert of the four R-P/A loci (Table 1). The 1102- to 1305-bp-long flanking sequences of *Rpm1* and *Rps5* were also obtained from GenBank (Stahl *et al.* 1999; Tian *et al.* 2002). A high level of total genetic diversities and those between presence and absence alleles were observed in all the six R-P/A (Table 1, average total π = 0.025, D_{xy} = 0.050), compared with their genetic background (<0.01, Nordborg *et al.* 2005). A high level of genetic diversity itself is not surprising because we intentionally selected regions linked to P/A loci. Interestingly, however, genetic diversities within presence and absence alleles

(π_1 and π_2, respectively) varied widely, and the observed ratio of the sum of them to total π, $(\pi_1 + \pi_2)/\pi$, was constantly and notably lower than the expected values under selective neutrality (6–83% of the neutral expectations, Table 1; see also Innan and Tajima 1997). Furthermore, Tajima's *D* statistic (Tajima 1989) was all positive (1.58–3.34, Table 1) and statistically significant for all but one of six loci. Positive *D* can be expected even under the neutrality in this case (because of the intentional selection of regions), but the statistical significance of this relatively conservative test, as well as low $(\pi_1 + \pi_2)/\pi$, suggests that balancing selection is responsible for the genetic variation around the R-P/A loci in general, as suggested in *Rpm1* and *Rps5* (Stahl *et al.* 1999; Tian *et al.* 2002).

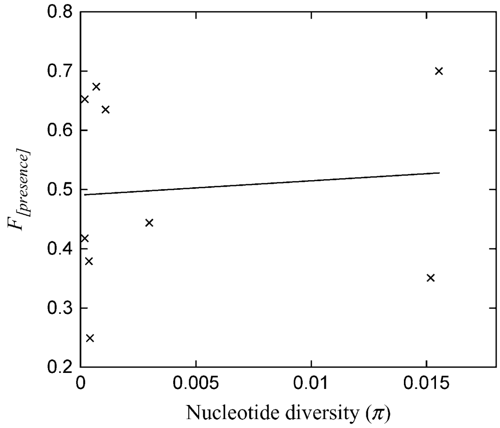

Figure 3.—Relationship between presence allele frequency ($F_{[presence]}$) and nucleotide diversity of the LRR region in the nine R-P/A. Correlation coefficient was 0.093.

DISCUSSION

Abundance of R-P/A in Arabidopsis genome: In our first approach toward identifying R-P/A, 19 candidate loci were identified by the BLAST search in the Ler genome, compared with 130 single *R*-gene loci in the Col genome in *A. thaliana* (Meyers *et al.* 2003). Of 10 candidate loci selected for PCR identification from the 19 candidates, 8 loci were confirmed as R-P/A. Therefore, the first approach can theoretically provide 15.2 R-P/A loci (= 19 × 0.8). In the second approach, we randomly selected 20 from 109 single *R*-genes (109 = 130 − 19 identified − *Rps5* − *Rpm1*) and found 2 R-P/A in Col. If we assume this proportion (2/20) as a proportion of R-P/A in the 109 single *R*-genes in *A. thaliana*, we predict another 10.9 R-P/A loci found by this approach (= 109 × 0.1). Including *Rpm1* and *Rps5* (Grant *et al.* 1998; Henk *et al.* 1999), we expect ~28 R-P/A in a genome, which account for 18.8% of the total 149 *R*-genes (Meyers *et al.* 2003). This result indicates that R-P/A is common phenomenon in *R*-genes in Arabidopsis. In fact, this number in *A. thaliana* can be even higher, because the estimate above does not take into account R-P/A that are absent in Col. The ongoing genome sequencing in the other accession of *A. thaliana* Ler will reveal that such R-P/A is present in Ler but absent in Col. The high proportion of R-P/A in plants was also observed in the rice genome (22.2%; J. Ding, J. Q. Chen, H. Araki, P. F. Zhang, J. Yang and D. Tian, unpublished results), suggesting that the R-P/A is common in plant genomes.

Evolutionary mechanism of R-P/A: *R*-genes are often arranged as tandem arrays as gene clusters. For example, there are 8–10 genes spanning 90 kb in the *Rpp5* gene cluster in Arabidopsis (Noel *et al.* 1999), 8 genes in 230 kb in the *Xa21* cluster in rice (Song *et al.* 1997), 8 genes in 261 kb in *Mla* in barley (Wei *et al.* 2002), 5 genes in 36 kb in *Cf4/Cf9* in tomato (Parniske *et al.* 1997), 1–>50 genes in *Rp1* in maize (Smith *et al.* 2004), and 12–32 genes in *RGC2* in lettuce (Kuang *et al.* 2004). The clustered genes are usually members of multigene families, and the sequence diversity in these genes is notably high. For instance, the diversity of genes is 0.05–0.26 in the *Rpp5* cluster, normally >0.1 in *Rp1*, *Mla*, and *RGC2*, and >0.05 in general in *Cf4/Cf9*. In these gene families, high K_a/K_s is common, and the genes in clusters are ready for the "birth-and-death process" to adapt higher divergent genes (Michelmore and Meyers 1998), similar to the vertebrate immune system (Nei *et al.* 1997).

In contrast, molecular evolution of R-P/A seems quite different from that of *R*-genes in the multigene family. The LRR region of R-P/A in *A. thaliana* showed a relatively low level of polymorphism and had low K_a/K_s. Why does R-P/A have a low level of polymorphism? One possible explanation is a suppressed recombination rate because of a lack of recombining pairs in the heterozygotes at the R-P/A loci. In this scenario, we expect a positive correlation between the presence allele frequency and the level of polymorphism, because the more presence alleles in a population, the more the alleles have chances to recombine in individuals. However, we did not find a clear correlation between these values in the nine R-P/A in this study (Figure 3, $r = 0.093$). This result indicates that a lower effective recombination rate alone may not explain the unique pattern of genetic variation in R-P/A, although it is noteworthy that we may not have enough power to detect the correlation (we have only nine plots available) and that our samples were collected from worldwide populations, not from a single natural population. In fact, observed values of $(\pi_1 + \pi_2)/\pi$ were much lower than selectively neutral expectations without recombination (Table 1), suggesting that some kind of balancing selection plays an important role in R-P/A.

Frequency-dependent selection, in which R-P/A is maintained in a spatio-temporal manner (Stahl *et al.* 1999), is one of the alternative explanations of balancing selection. In this scenario, a low level of genetic diversity

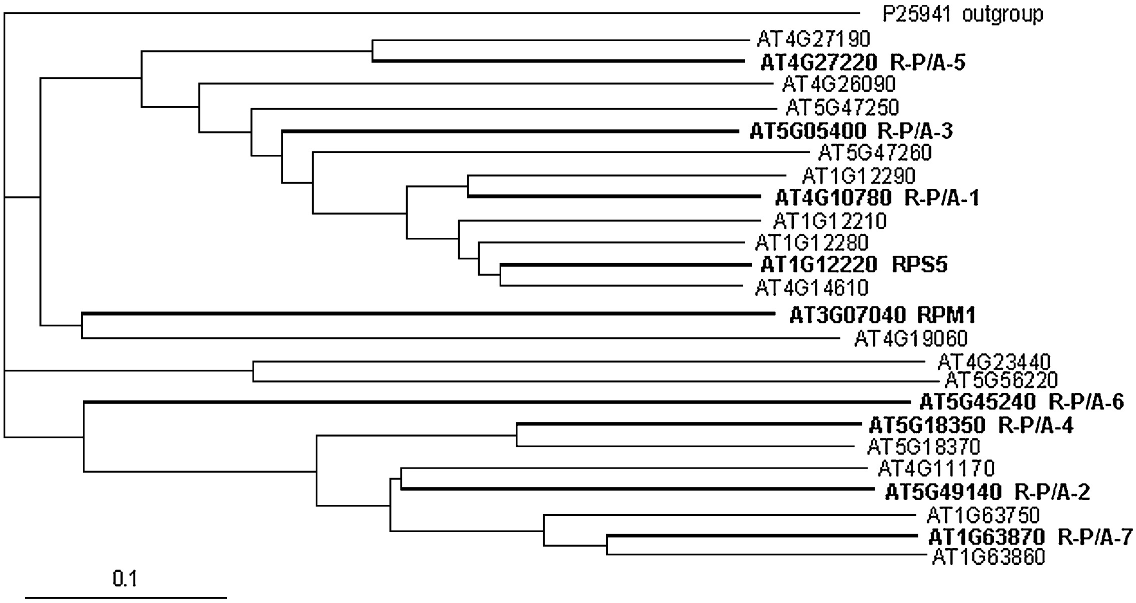

Figure 4.—Phylogenetic relationship of *R*-genes in the Col genome in *A. thaliana*. The tree was dissected from the one constructed by Meyers *et al.* (2003), using amino acid sequences of the NBS domain of *R*-genes by the neighbor-joining method. Nine R-P/A loci investigated in this study are indicated by boldface type. The Streptomyces sequence rooted the tree as the outgroup (Meyers *et al.* 2003).

in the presence allele is expected independently of allele frequencies because of the fluctuation of allele frequencies. Significantly positive Tajima's *D* (Tajima 1989) supports this scenario, and low K_a/K_s in R-P/A is also understandable if we assume selective constraint on presence *R*-genes under balancing selection between P and A.

The reason why the evolutionary mechanism of R-P/A differs from that of the other *R*-genes remains unknown. At the genome level, the number of *R*-genes varied largely among species, from 149 in Arabidopsis to 480 in rice (Meyers *et al.* 2003; Zhou *et al.* 2004), indicating that the maximum number of *R*-genes obtainable for individuals is different among species. The fitness cost of *R*-genes may limit the total number of *R*-genes in a genome (Tian *et al.* 2003), while it seems insufficient to confer resistance to the multitude of pathogens that a plant is likely to encounter (Dangl and Jones 2001). R-P/A may provide a potential for plants to hold the necessary and sufficient number of divergent *R*-genes without incurring too much cost. The unique haplotype of the set of nine R-P/A in each accession investigated (Table 2) is consistent with this scenario, but further study is required to address this question.

Origin of R-P/A: Because all NBS-LRR genes share similar sequences and structural motifs (Kanazin *et al.* 1996), they are believed to share a common ancestor. Nucleotide sequencing of R-P/A and its flanking region in a closely related species, *A. lyrata*, enabled us to identify four orthologous genes out of nine R-P/A (At5g18350 and At1g63870 and *Rpm1* and *Rps5* from Bergelson *et al.* 2001). Interestingly, the first two loci showed notably high nucleotide diversities in the LRR regions, supporting that they are the ancestral alleles in this species. The other five loci might be either under P/A in *A. lyrata* or acquired in *A. thaliana* after speciation. However, we found no similar paralogs for these five R-P/A on the basis of a conserved NBS region in *R*-genes in *A. thaliana* (Figure 4), indicating that duplication within *A. thaliana* after speciation is less likely. The possibility that the similar paralogs for these *R*-genes are missing in the Col genome by P/A of the paralogous loci remains, but the fact that all R-P/A found in Col shared a very old common ancestor (Figure 4) indicates that this possibility is low. Relatively high genetic variation of the flanking regions of R-P/A within absence alleles in R-P/A-1–R-P/A-3 (Table 1) may reflect that the absence alleles are the ancestral state in these loci, although this is not necessarily true if allele frequencies of these R-P/A fluctuated recently. In fact, we observed the same phenomenon (higher genetic variation of the flanking region of R-P/A among absence alleles than among presence alleles) in *Rps5*, where we know that the presence alleles are the ancestral state.

Regardless, our results suggest that these R-P/A have been maintained in *A. thaliana* for a long evolutionary time and that the balancing selection plays an important role in the molecular evolution of R-P/A. Further studies on the profiling of R-P/A in wild plants will reveal their disease resistance mechanisms in variable natural environments.

We thank Joy Bergelson for DNA samples of Arabidopsis accessions, Hideki Innan, the corresponding editor, and anonymous reviewers for much insightful advice. This work was funded by National Natural Science Foundation of China (30470924, 30570987, and 30470122), Specialized Research Fund for the Doctoral Program of Higher Education, and a SUR grant from IBM to D.T. or J.Q.C.

LITERATURE CITED

Allen, R. L., P. Bittner-Eddy, L. Grenville-Briggs, J. Meitz, A. P. Rehmany *et al.*, 2004 Host-parasite coevolutionary conflict between *Arabidopsis* and *Downy Mildew.* Science **306:** 1957–1960.

Belkhadir, Y., R. Subramaiam and J. L. Dangl, 2004 Plant disease resistance protein signaling: NBS-LRR proteins and their partners. Curr. Opin. Plant Biol. **7:** 391–399.

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

Bittner-Eddy, P. D., L. 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.

Botella, M. A., J. E. Parker, L. N. Forst, P. D. Bittner-Eddy, J. L. Beynon *et al.*, 1998 Three genes of the *Arabidopsis RPP1* complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. Plant Cell **10:** 1847–1860.

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.

Dangl, J. L., and J. D. Jones, 2001 Plant pathogens and integrated defence responses to infection. Nature **411:** 826–833.

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.

Gassmann, W., M. E. Hinsch and B. J. Staskawicz, 1999 The *Arabidopsis RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. Plant J. **20:** 265–277.

Grant, M. R., J. M. McDowell, A. G. Sharpe, M. deTorres Zabala, 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.

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

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

Innan, H., and F. Tajima, 1997 The amounts of nucleotide variation within and between allelic classes and the reconstruction of the common ancestral sequence in a population. Genetics **147:** 1431–1444.

Jia, Y., G. T. Bryan, L. Farrall and B. Valent, 2003 Natural variation at the *Pi-ta* rice blast resistance locus. Phytopathology **93:** 1452–1459.

Kanazin, V., L. F. Marek and R. C. Shoemaker, 1996 Resistance gene analogs are conserved and clustered in soybean. Proc. Natl. Acad. Sci. USA **93:** 11746–11750.

Kuang, H., S. S. Woo, B. C. Meyers, E. Nevo and R. W. Michelmor, 2004 Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. Plant Cell **16:** 2870–2894.

Mauricio, R., E. Stahl, T. Korves, D. Tian, M. Kreitman *et al.*, 2003 Natural selection for polymorphism in the disease resistance gene *RPS2* of Arabidopsis. Genetics **163:** 735–746.

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

Meyers, B. C., A. Kozik, A. Griego, H. Kuang and R. W. Michelmore, 2003 Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis.* Plant Cell **15:** 809–834.

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.

Nei, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

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

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:** 2099–2111.

Nordborg, M., T. T. Hu, Y. Ishino, J. Jhaveri and C. Toomajian, 2005 The pattern of polymorphism in *Arabidopsis thaliana.* PLoS Biol. **3:** e196.

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.

Rose, L. E., P. D. Bittner-Eddy, C. H. Langley, H. Charles, E. B. Holub *et al.*, 2004 Maintenance of extreme amino acid diversity at the disease resistance gene, *RPP13*, in *Arabidopsis thaliana.* Genetics **166:** 1517–1527.

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

Smith, S. M., A. J. Pryor and S. H. Hulbert, 2004 Allelic and haplotypic diversity at the *Rp1* rust resistance locus of maize. Genetics **167:** 1939–1947.

Song, W. Y., L. Y. Pi, G. L. Wang, J. Gardner, T. Holsten *et al.*, 1997 Evolution of the rice *Xa21* disease resistance gene family. Plant Cell **9:** 1279–1287.

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.

Swofford, D., 2000 *PAUP**: Phylogenetic Analysis Using Parsimony.* Sinauer Associates, Sunderland, MA.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Takahashi, M., F. Matsuda, N. Margetic and M. Lathrop, 2002 Automated identification of single nucleotide polymorphisms from sequencing data. Proc. IEEE Comput. Soc. Bioinform. Conf. **1:** 87–93.

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.

Tian, D., B. Traw, J. Chen, M. Kreitman and J. Bergelson, 2003 Fitness cost of *R*-gene mediated resistance in *Arabidopsis thaliana.* Nature **424:** 74–77.

Wei, F., R. A. Wing and R. P. Wise, 2002 Genome dynamics and evolution of the *Mla* resistance locus in barley. Plant Cell **14:** 1903–1917.

Wright, S., 1931 Evolution in Mendelian populations. Genetics **16:** 97–159.

Zhou, T., Y. Wang, J. Q. Chen, H. Araki, Z. Q. Jing *et al.*, 2004 Genome-wide identification of NBS genes in rice reveals significant expansion of divergent non-TIR NBS genes. Mol. Gen. Genet. **406:** 402–415.

Communicating editor: M. Nordborg