DOI: 10.1534/genetics.104.029371

Allelic and Haplotypic Diversity at the *Rp1* Rust Resistance Locus of Maize

Shavannor M. Smith,* Anthony J. Pryor† and Scot H. Hulbert*,[1]

**Department of Plant Pathology, Kansas State University, Manhattan, Kansas 66506 and †Commomwealth Scientific and Industrial Research Organization, Canberra, Australian Capital Territory 2601, Australia*

Manuscript received March 27, 2004
Accepted for publication May 4, 2004

ABSTRACT

The maize *Rp1* rust resistance locus is a complex consisting of a family of closely related resistance genes. The number of *Rp1* paralogs in different maize lines (haplotypes) varied from a single gene in some stocks of the inbred A188 to >50 genes in haplotypes carrying the *Rp1-A* and *Rp1-H* specificities. The sequences of paralogs in unrelated haplotypes differ, indicating that the genetic diversity of *Rp1*-related genes is extremely broad in maize. Two unrelated haplotypes with five or nine paralogs had identical resistance phenotypes (*Rp1-D*) encoded in genes that differed by three nucleotides resulting in a single amino acid substitution. Genes in some haplotypes are more similar to each other than to any of the genes in other haplotypes indicating that they are evolving in a concerted fashion.

PLANT genomes carry large numbers of genes that control active defenses to pathogens. The largest class of these disease resistance genes codes for proteins with nucleotide binding site (NBS) and leucine-rich repeat (LRR) domains (Meyers *et al.* 1999; Bai *et al.* 2002). Various NBS-LRR genes control resistances to fungi, bacteria, viruses, nematodes, and insects (Bent 1996; Baker *et al.* 1997; Hammond-Kosack and Jones 1997; Ellis *et al.* 2000; Hulbert *et al.* 2001).

The *Rp1* rust resistance complex of maize consists of a family of closely related NBS-LRR genes, which are all tightly linked at the *Rp1* locus (Collins *et al.* 1999). These genes mispair extensively in meiosis, and recombination events within the cluster while they are mispaired reassort the genes into new combinations or haplotypes. The best-characterized *Rp1* haplotype is *HRp1-D*, which carries the *Rp1-D* gene along with eight paralogs (Sun *et al.* 2001). Transcripts were detected for five of seven paralogs capable of coding for proteins similar to *RP1-D*, with 91–99% amino acid identity. An eighth paralog was predicted to code for a truncated protein with no LRR. Other than the *Rp1-D* gene, which confers race-specific rust resistance, no phenotype has been associated with the other genes in this haplotype.

Southern blot analysis of different maize lines has indicated that most lines carry different *Rp1* haplotypes. The number of different restriction fragments homologous to an *Rp1* probe varies from one or a few to more than a dozen. The partially conserved, highly duplicated nature of genes in complex loci like *Rp1* and *Rp3* (Sun *et al.* 2001; Webb *et al.* 2002) make it difficult to estimate gene number by gel blot analysis because of comigration of genes with similar restriction sites. Most haplotypes that carry known *Rp1* genes appear to have roughly as many or more *rp1* family members than the *HRp1-D* haplotype, making it very difficult to identify the genes controlling the resistance. The resistance reactions of some of the *Rp1* near-isogenic lines are indistinguishable with current collections of rust biotypes, indicating that they carry phenotypically identical *Rp1* genes, even though they were originally identified in different germplasm sources. In some cases these lines have been shown by DNA gel blot analysis to carry very different *Rp1* haplotypes, while in others the haplotypes appear identical (Collins *et al.* 1999).

This study was conducted to examine the differences between *Rp1* haplotypes, in both gene number and gene diversity. The number of *Rp1* paralogs in haplotypes with a range of numbers of *Rp1*-homologous restriction fragments was examined by sequencing PCR-amplified or directly cloned fragments of the genes. The haplotypes examined include pairs of haplotypes that appear unrelated by gel blot analysis but carry genes that are phenotypically indistinguishable. The feasibility of using such pairs of haplotypes to identify the genes controlling the resistance phenotype was examined.

MATERIALS AND METHODS

Amplification, cloning, and sequencing of *Rp1* genes: To sample as many *Rp1* genes as possible from different *Rp1* haplotypes, several PCR primers were designed from conserved regions in the genes. Three primer pairs were used to amplify nearly complete *Rp1* coding regions, as nonspecifically as possible (Table 1). The use of P6 with 4890R was used to amplify ~4-kb fragments that included nearly the whole coding re-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY309779–AY309783 and AY581258.

[1]*Corresponding author:* Department of Plant Pathology, 4024 Throckmorton Plant Sciences Center, Kansas State University, Manhattan, KS 66506-5502. E-mail: shulbrt@plantpath.ksu.edu

gion. In case this pair was unable to amplify certain sequences, two other primers pairs were also used. The primer pair P6 and P15 amplified an ~2-kb fragment corresponding to most of the 5′ half of the coding region of the genes. Primer pair 2290F and 4890R was used to amplify the 3′ half of the genes. Thirteen internal primers were used to sequence the coding region of selected genes (P6, P7, P8, K2, 2290F, P11, P15, P12, P18, 3990R, P19, 4450R, and P22; Table 1). To examine the number of different genes in a haplotype without sequencing the whole coding region, a highly polymorphic region (~835 bp) of the LRR was amplified using the primers 3535F and 4700R.

A variety of more specific primers were made to search for specific genes or recombinant genes in certain haplotypes (Table 1). Specific primers were designed by aligning the gene's sequence to all the genes from the *HRp1-D* haplotype and any other available full-length sequences and searching for unique polymorphisms among the aligned sequences. Gene-specific primers were typically paired with nonspecific primers in attempts to amplify specific genes.

PCR amplification of genomic DNA templates was performed using Enhanced DNA polymerase (Stratagene, La Jolla, CA) with ~1 min of extension time for every kilobase of fragment size. Optimal PCR parameters were followed as suggested by the manufacturer. Total RNA isolation was performed on fully expanded second leaves using GIBCO BRL (Rockville, MD) Trizol reagent as described by the manufacturer. RT-PCR was performed using a ProStar first-strand RT-PCR kit with optimal RT-PCR conditions suggested by the manufacturer (Stratagene). Following first-strand synthesis with an oligo(dT) primer, one primer pair (P12 and 3990R; Table 1) was used to amplify *Rp1* cDNA sequences. To distinguish between different *Rp1* genes, these primers were designed from regions that were conserved among the *Rp1-D* paralogs and flank a polymorphic region. These primers amplify an ~835-bp fragment. The PCR products were gel purified and cloned into the Invitrogen (Carlsbad, CA) TOPO TA cloning vector. Recombinant plasmids were prepared using the QIAGEN (Valencia, CA) miniprep purification system. Clones were digested with *Eco*RI (New England Biolabs, Beverly, MA) and analyzed by agarose gel electrophoresis prior to sequencing. M13 forward and reverse primers were used to fully sequence smaller fragments (835 bp) or sequence the ends of larger clones, and selected clones were sequenced using primers from conserved regions of the *Rp1* genes. All sequencing was performed at Kansas State University Sequencing Facility.

DNA sequence analysis: Sequence alignment of each clone (5′ or 3′ ends or entire coding regions) was done using the GCG Seqweb version 2 program or the BCM (//searchlauncher.bcm.tmc.edu/), Basic Local Alignment Search Tool (http://www.ncbi.nlm.nih.gov/blast/). Only single clones were examined for most of the genes from the *Rp1-A*, *Rp1-H*, and *Rp1-JF* haplotypes. Although high-fidelity polymerases were used to amplify the sequences, the potential for introduction of sequence errors during PCR amplification of the genes could not be ruled out. Therefore, polymorphic bases were scrutinized by determining if other *Rp1* genes shared the polymorphisms. Polymorphic nucleotides that were unique to a single gene were considered to be possible sequence errors. Alternatively, *informative* polymorphisms, where the polymorphic bases were shared by different genes, were considered to provide useful information for classifying the genes and estimating gene number. Only genes that differed for multiple informative nucleotide polymorphisms were considered to be unique genes.

Bioinformatics programs: Clustal X bioinformatics program was used for sequence alignment and bootstrap analysis (ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/) of the predicted protein sequences of 13 *rp1* genes and the LRR-coding region of 132 *Rp1* sequences. Neighbor-joining tree images were generated with the TreeView bioinformatics program (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). All parameters were set to default.

Isolation of an altered *Rp1-D* gene from the *Cuzco* haplotype: F_1 hybrids between maize lines homozygous for *HRp1-D-Cuzco* and *HRp1-I* were backcrossed by a susceptible *rp1* line (H95) and 26,000 progeny were screened for variant phenotypes. Ten-day-old seedlings were inoculated simultaneously with a mixture of rust spores (biotypes IN1 and IN2) and placed in high humidity for 12–15 hr. Both rust isolates are avirulent on lines carrying either *Rp1-D* or *Rp1-I*. Eight to 10 days following inoculation, seedlings were scored using a resistance reaction rating scale of 0–4. A high level of resistance with no sporulation was indicated by a 0 rating. A rating of 1 indicated a high level of resistance with one or a few uredinia per leaf. A rating of 2 indicates a large number of uredinia with necrotic or chlorotic spots and a visible hypersensitive response. One individual with a rating of 2 was self-fertilized to generate a homozygous line (*Rp1-D2I3*), which was screened with rust biotypes IN1, IN2, IN3, and HI1 separately to determine its race specificity and confirm the altered phenotype.

To identify the variant gene, genomic DNA isolated from the *Rp1-D2I3* line was PCR amplified with a forward specific primer designed to match the *Rp1-D* gene (500F) and nonspecific reverse primers (26-47-9-2-R, 2751R, and 3968R; Table 1). These primers amplified fragments up to ~4.4 kb that include the whole coding region and part of the promoter. The PCR products were purified, cloned, sequenced, and analyzed as described above.

RESULTS

Indistinguishable specificities in related and unrelated haplotypes: Three sources of *Rp1-D* were identified by Hooker and co-workers: the African cultivars *Kitale* and *Njoro* and the South American cultivar *Cuzco* (Wilkinson and Hooker 1968). When DNA from lines homozygous for each of the three haplotypes was examined by gel blot analysis, the lines derived from the African material appeared identical, but these appeared unrelated to the line with the *Cuzco* haplotype. The *HRp1-D-Cuzco* haplotype consistently had fewer fragments when compared to *HRp1-D*. The previously characterized *HRp1-D*, originally from the African cultivar *Kitale*, carried nine *rp1* paralogs (Sun *et al.* 2001). Gel blot hybridization with a 1.5-kb probe representing the 5′ half of the coding region typically detects fewer than nine restriction fragments because of comigration of similar size fragments. The enzyme *Nsi*I was exceptional in that it produced nine restriction fragments, corresponding to each of the nine genes in the haplotype. In contrast, only two fragments were detected in *Nsi*I digests of the *HRp1-D-Cuzco* haplotype. Similarly, only two fragments were observed in restriction digests with *Sph*I, *Xba*I, and *Eco*RV and three fragments were observed in *Nco*I and *Hin*dIII digests. Typically, some fragments hybridized more intensely than the others, indicating that there were probably more than two or three genes, but less than the previously characterized *HRp1-D*.

The resistances conferred by *Rp1-A* and *Rp1-F* are also identical in specificity and resistance phenotype (Hul-

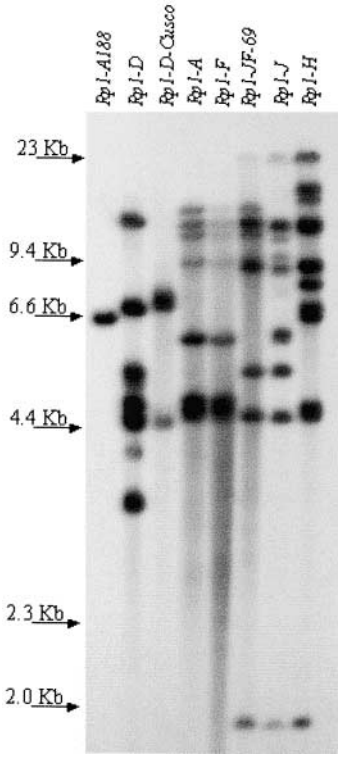

FIGURE 1.—Gel blot analysis of DNAs isolated from related and unrelated *Rp1* haplotypes. DNAs were digested with *Nco*I and probed with a probe corresponding to the 5′ half of the *Rp1-D* coding region. *Rp1* haplotypes with indistinguishable race specificities exhibited banding patterns that were either identical (*HRp1-A* and *HRp1-F*) or very different. Fragment sizes in kilobases are shown on the left.

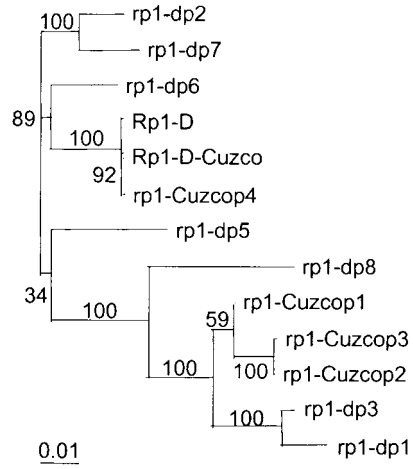

FIGURE 2.—A neighbor-joining tree of the coding region of *Rp1* genes from *HRp1-D-Cuzco* and *HRp1-D-Kitale*, generated using the predicted protein sequences of 13 *rp1* genes. Numbers at nodes represent the percentage (of 1000) of the time the genes cluster together in bootstrap analysis. The paralog designations -dp and -Cuscop indicate origins in the *HRp1-D-Kitale* and *HRp1-D-Cusco* haplotypes. Uppercase locus designations (*Rp1*) indicate a phenotypically detectable gene.

BERT *et al.* 1991). Two different sources of *Rp1-A* were originally identified from the cultivars Golden Glow and Golden King (HOOKER and LEROUX 1957). The source of *Rp1-F* is P.I. 172332, from Australia. This latter gene was originally given a different *Rp1* allelic designation because the resistance specificity appeared different, probably due to other factors in the genetic background in which the genes were identified (HOOKER and RUSSELL 1962). When compared, *HRp1-F* and the two haplotypes carrying *Rp1-A* all appeared identical by gel blot analysis (Figure 1). *HRp1-J* and *HRp1-H* have indistinguishable race specificities and resistance phenotypes (HULBERT *et al.* 1991) but appeared unrelated by gel blot analysis (Figure 1).

Isolation of genes from the *HRp1-D-Cuzco* haplotype: *Rp1* genes from *HRp1-D-Cuzco* were sequenced following PCR amplification and cloning. Twenty different clones carrying ~4-kb fragments from PCR products with the primers P6 and 4890R were partially sequenced. Comparison of these sequences indicated they corresponded to five different genes. Additional conserved primer pairs were used to amplify genes from *HRp1-D-Cuzco* because of the possibility that genes would be missed due to heterogeneity in the primer-annealing sequences. An additional 11 clones corresponding to the 5′ region of *Rp1* genes were amplified using the P12 and P15 primers and 9 clones of the 3′ half of the gene were isolated using the 2290F and 4890R primers. Partial sequence analysis of these clones showed that they corresponded to previously cloned sequences, indicating that the five genes identified account for all the *Rp1* genes in *HRp1-D-Cuzco*. At least 1 full-length clone was completely sequenced for each of these five genes.

The identification of only five different genes by sequence analysis agreed with the gel blot analysis of gene number. No more than four restriction fragments were observed in any enzyme digest. In digests such as *Nco*I and *Hin*dIII, three fragments were observed, but one hybridized more intensely than the others, suggesting that they probably represented two genes (Figure 2). The five *HRp1-D-Cuzco* genes were designated *rp1-Cuzcop1*–*rp1-Cuzcop4* (for paralogs 1–4) and *Rp1-D-Cuzco*. The latter gene was determined to code for a functional *Rp1-D* protein (see below).

Comparisons of the *HRp1-D* and *HRp1-D-Cuzco* haplotypes: In addition to differences in *Rp1* gene number, the *Rp1-D*-carrying haplotypes from *Kitale* and *Cuzco* differed in several other respects. The previously characterized *Kitale* haplotype carried nine *Rp1* paralogs, one of which was truncated before the LRR coding region. No truncated members were identified in *HRp1-D-Cuzco* by sequence analysis of PCR-amplified sequences or by hybridization to gel blots of genomic DNA with probes from different regions of the *Rp1-D* gene. Depending on the restriction enzyme used to cut the genomic DNA, probes from different regions of the gene hybridized to either the same fragments or similar numbers of fragments. The lack of a truncated family member is consis-

tent with the idea that truncated genes are not required for *Rp1*-mediated resistance (Sun *et al.* 2001).

Another novel aspect of *HRp1-D-Cuzco* is the fact that two of the five genes are very similar in sequence (Figure 2). The genes in the previously characterized *HRp1-D* from *Kitale* capable of coding for NBS-LRR genes coded for predicted amino acid sequences that ranged from 88 to 94% identity in amino acid sequence. While the least closely related genes in *HRp1-D-Cuzco* coded for predicted proteins only 89% identical to each other, two of the genes in this haplotype were predicted to code for proteins that differed by only two amino acids (>99.5% identity).

To determine the relationship and similarity between the genes in these two haplotypes, the DNAs of the coding regions and their predicted amino acid sequences were aligned. Only two of the genes from *HRp1-D-Cuzco* appeared similar to any of the genes in *HRp1-D*. The *Rp1-D* gene from *HRp1-D-Kitale* was identical to a gene from *HRp1-D-Cuzco* (*Rp1-D-Cuzco*) except for three nucleotides and different from the *rp1-Cuzcop4* gene at only five nucleotides. The three nucleotide differences between the *Rp1-D* and *Rp1-D-Cuzco* genes resulted in a single amino acid substitution. An aspartic acid in the *RP1-D* protein at position 1064 in the LRR region is changed to phenylalanine in the predicted protein coded by *Rp1-D-Cuzco*. Nonsynonymous substitutions have been observed at this site in two other *HRp1-D* genes (Sun *et al.* 2001). The other three genes in *HRp1-D-Cuzco* are only 86–89% identical to the *Rp1-D* gene but are more similar to each other. A neighbor-joining tree, generated using sequences representing all of the genes from *HRp1-D* and *HRp1-D-Cuzco*, showed that two genes from *HRp1-D-Cuzco* form a group with *Rp1-D*, while the three remaining genes form a second group (Figure 2).

Identification of the *Rp1-D* gene in *HRp1-D-Cuzco* and the recombinant *Rp1-D2I3* gene: The very high level of sequence identity between two *Rp1* genes from *HRp1-D-Cuzco* and the *Rp1-D* gene indicated that one or both of these genes were possible candidates for the gene conferring the resistance phenotype. Two approaches were used to determine which of the genes controls the phenotype. Previous analysis with *HRp1-D* from *Kitale* indicated some paralogs are not transcribed. To determine whether or not the *Rp1-D* candidates in the *Cuzco* haplotype are transcribed, cDNA isolated from *HRp1-D-Cuzco* seedlings was amplified using the P12 and 3990R primers (Table 1). These primers match sequences conserved among all the genes in this haplotype and flank polymorphic nucleotides. Sequence analysis of the resulting 33 RT-PCR clones found that 27 clones matched *rp1-Cuzcop2* and 6 clones matched *Rp1-D-Cuzco*. No RT-PCR clones were found to match the *rp1-Cuzcop1*, *rp1-Cuzcop3*, and *rp1-Cuzcop4* genes, indicating that they either are not transcribed or are transcribed at lower levels. Since transcripts were found for only one (*Rp1-D-Cuzco*) of the two genes with high sequence identity to the *Rp1-D* gene, the *Rp1-D-Cuzco* gene is the best candidate for the gene conferring the *Rp1-D* resistance phenotype in *HRp1-D-Cuzco*.

A second approach to identifying the *Rp1-D-Cuzco* gene with the *Rp1-D* phenotype was to identify a mutation or recombination event that altered the gene. A variant showing a reduced level of resistance was derived from a testcross population of 26,000 individuals from an *HRp1-I* × *HRp1-D-Cuzco* cross. This variant specified a reduced resistance reaction (reaction type 2) with biotypes IN1 and IN2 when compared to the parental lines, which were both highly resistant (reaction type 0 or 0–1) to both isolates. The variant specified a resistance reaction of 2 and 4 (fully susceptible) with biotypes IN3 and HI1, respectively, while the *Rp1-D* parental line specified a resistance reaction of 0 and 4 and the *Rp1-I* parental line specified a resistance reaction of 4 and 0 with biotypes IN3 and HI1, respectively. The variant line maintains the same resistance specificity as the *Rp1-D* parent, resistant to IN1, IN2, and IN3, but not HI1. The variant line has a reduced level of resistance to the biotypes it is resistant to, typically showing a dozen or more uredinia per seedling leaf while *Rp1-D*-carrying lines typically show none or a few. The putative modified *Rp1* gene was designated *Rp1-D2I3*. A similar reduced resistance variant (*Rp1-D*5*) was previously identified from *HRp1-D-Kitale* and was found to be derived from an intragenic recombination event involving the *Rp1-D* gene (Sun *et al.* 2001). The similar phenotype of the *HRp1-D2I3* variant suggested that the *Rp1-D* gene from *HRp1-D-Cuzco* was involved. Therefore, *HRp1-D2I3* genomic DNA was PCR amplified with primers designed from regions conserved in *Rp1* genes (P6-4890R), which include the entire coding region. Approximately 600 bp of the 5′ and 3′ ends of the resulting cloned PCR products were sequenced and aligned with *HRp1-D-Cuzco* paralogs for sequence comparisons. One of the *HRp1-D2I3* clones did not completely match any of the genes from *HRp1-D-Cuzco*. Sequences from this clone were identical to the *Rp1-D-Cuzco* gene on the 5′ end but did not match *Rp1-D-Cuzco* or any of the other genes for *HRp1-D-Cuzco* on the 3′ end. This indicated the 3′ end was contributed by an *HRp1-I* gene. To verify that this was actually a recombinant gene and not a previously unidentified gene from *HRp1-D-Cuzco*, genomic DNA from the putative recombinant and the two parental haplotypes, *HRp1-D-Cuzco* and *HRp1-I*, were PCR amplified using a nonspecific forward primer (P22) and a reverse gene-specific primer (26-47-9-2-R) designed to be specific to the 3′ region of the *Rp1-D2I3* putative recombinant gene (Table 1). The primer pair amplified a fragment of the expected size in the recombinant haplotype and *HRp1-I* but did not amplify from *HRp1-D-Cuzco* DNA, indicating that the putative recombinant was not an *HRp1-D-Cuzco* gene. Sequence analysis of the specific PCR-amplified fragments from *HRp1-I* and the recombinant haplotype verified the gene

TABLE 1

Primers used to amplify *Rp1* sequences

Primer name	Position[a]	Primer sequence	Specificity[b]
500F	−848 F	AGCGAGCAGTAGACCCTCAGC	Nonspecific
P6	67 F	AAGCTTCAGCTTACCTCAGT	Nonspecific
201F1	266 F	TTGAGGGCAAGGCCAAGAGCA	Specific
201F2	266 F	TTGAGGGCAAGGCCAAGAGCGA	Specific
P7	338 F	CCTCCACTGCAACTACTGTC	Nonspecific
P8	639 F	AAGCTCAGCTAAGTACTC	Nonspecific
1520R	641 R	GTTTGGCCATTGTTGGATTGG	Nonspecific
K2	940 F	CTCATAATGAGACAGAGTGGG	Nonspecific
2290F	1406 F	GGGTGGCAGAAGGATTTG	Nonspecific
P11	1860 F	TTGGTGAGCTGAAGCACCTC	Nonspecific
P12	2143 F	TTGAATGAGCTTGGTGGCAG	Nonspecific
P15	2200 R	GAGAATTGTCATTGGAAAGGAT	Nonspecific
P18	2592 F	ATCAATTGGTTGGTGCCCACTG	Nonspecific
JF-A-31F	2660 F	GATGTGGGAGGTGGATTCAGGAAAAG	Specific
2751R	2796 R	TTCCGGAGGAAAGAGAAGATAAA	Specific
JF-A-440F	3061 F	CAAAGCTTGACACGTTGGTTGTAAC	Specific
P19	3084 F	GATAGGTTGGTTGTAAGTG	Nonspecific
3990R	3116 R	GCTTACGTGCTGCTCCATCTC	Nonspecific
324R1	3404 R	ATGTTATGTTGGGGCAATGTT	Specific
324R2	3404 R	ATGTTATGTTGGGGCATCGTT	Specific
P22	3460 F	TGTCCAGGAATCGCTCACGG	Nonspecific
3535F	2685 F	GATGTGGGAGGTGGATTCAGG	Nonspecific
P22	3520 F	TGTCCAGGAATCGCTCACGG	Nonspecific
4700R	3544 R	CACATGCTAATGGCTGAAG	Nonspecific
4450R	3622 R	AGCCGTCAGTTTCATTTG	Nonspecific
GC-A1-704R	3765 R	CTTCAAGACGGGGCAATTTAC	Specific
392R2	3766 R	TCTTCAAGACGGGGCAACCT	Specific
458R	3840 R	AAGCTCTTCCAGCGAACGTGTAG	Specific
GC-A1-783R	3844 R	GGTATTAGCAGTTGATGTGCG	Specific
26-47-9-2-R	3860 R	TCATTTATTTCAAAGTGGTGCGG	Specific
522R	3897 R	AATTAGTAGAACCTGTACCTTCTCATTT	Specific
3968R	3995 R	CTGGAGCTTCAAGAGAAGAGGGG	Specific
636R	4007 R	CAAGAGGAGAGAGCAGTCGCA	Specific
4890R	4010 R	CTTGAAGCTCCAGAGTTCAGG	Nonspecific

Nonspecific primers were designed from conserved regions among the *Rp1-D* paralogs. Specific primers were designed from specific *Rp1* genes.

[a] Position of 5′ end of the homologous sequence of the *Rp1-D* gene where position 1 corresponds to the predicted start codon. F and R correspond to the sequences of the coding and noncoding strands of the *Rp1* gene from which the primers were designed.

[b] Specific primers were designed for amplification of specific *Rp1* genes while nonspecific primers were designed to amplify most or all *Rp1* genes.

amplified from the recombinant haplotype was a recombinant gene and provided partial sequence of the parental gene from *HRp1-I* (Figure 3). The recombinant gene was identical to the *Rp1-D-Cuzco* gene until base 3674. The sequence diverged at base 3675 where it became identical to the candidate *HRp1-I* gene. The crossover was determined to have occurred within a 63-bp region of perfect identity between the two parental genes.

The *rp1-Cuzcop4* gene is very similar to the *Rp1-D-Cuzco* gene involved in the recombination event. To determine if this gene is present in the recombinant haplotype, genomic DNA from the recombinant haplotype was PCR amplified with primer pair P12 and 2751R. This primer pair was designed to specifically amplify these two highly similar genes, but flank a polymorphic region to distinguish the two genes. Direct sequencing of the amplified product indicated that only the recombinant gene was present in the recombinant haplotype, not *Rp1-D-Cuzcop4*. This is consistent with our previous postulate, based on sequence similarity and expression, that the *Rp1-D-Cuzco* gene confers the *Rp1-D* resistance specificity in *HRp1-D-Cuzco*. This also indicates the reduced level of resistance conferred by the recombinant haplotype is due to the recombination event that replaces the last 173 nucleotides of the coding region, corresponding to the N-terminal 57 amino acids of the LRR region of the *RP1-D* protein. An alternative explanation is that the *rp1-Cuzcop4* contributes to the resistance con-

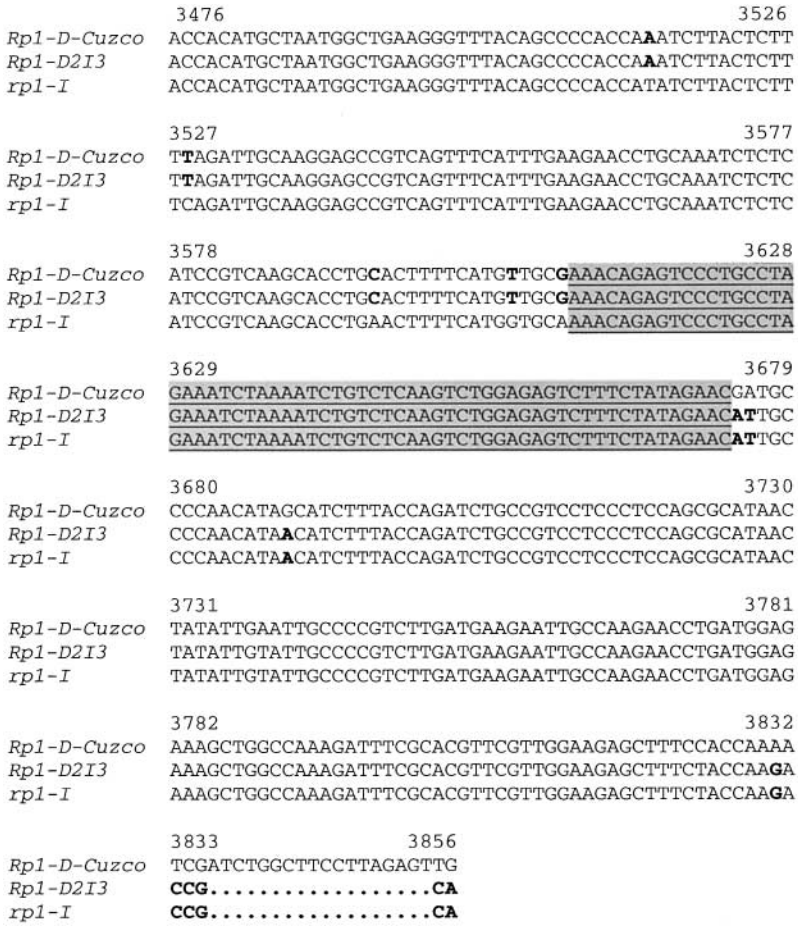

FIGURE 3.—DNA sequence alignment of the recombinant *Rp1-D2I3* gene with the two parental genes involved in the recombination event. Numbers correspond to the number of bases from the start codon in the *Rp1-D* gene. The exchange point occurred in the 63-nucleotide region of perfect identity (shaded) defined by the polymorphic bases between the parental genes. Bases that are polymorphic between the parental genes are in boldface type. The recombinant gene confers the *Rp1-D* specificity with a reduced level of resistance.

ferred by *Hrp1-D-Cuzco* and it is the loss of this gene in the recombinant haplotype that causes the reduced resistance. This explanation seems less likely since no transcripts were identified for this gene, although it is possible that very low levels of transcript are sufficient. It is also possible that the *rp1-Cuzcop4* gene does not contribute to resistance but is capable of coding for a functional *RP1-D* protein if it were transcribed.

Characterization of other *Rp1* haplotypes: Examination of several maize lines with no noticeable *Rp1*-mediated resistance identified one accession of the inbred A188 that had only a single restriction fragment in several enzyme digests (Figure 1). To determine the actual number of genes in this *Rp1-A188* haplotype we amplified and cloned fragments from genomic DNA of this line using primer pair P6 and 4890R, which amplifies an ~4-kb fragment (Table 1). The primer pair was used to amplify *rp1* sequences nonspecifically and was designed from the conserved coding region among the *Rp1-D* paralogs. The PCR-amplified fragment was cloned and used as template to sequence and characterize the clones. The 5′ and 3′ ends of 7 clones were sequenced, providing ~500 bp of the coding region for each end. No differences were found among the 7 clones, indicating that the *Rp1-A188* haplotype contains a single gene. Three additional *Rp1* haplotypes were examined by complete or partial sequence analysis of *Rp1* paralogs. The *Rp1-JF69* haplotype is a recombinant haplotype that carries both the *Rp1-F* gene and the *Rp1-J* gene (HULBERT *et al.* 1993). The *Rp1-JF69* haplotype was selected from 11 other recombinant haplotypes with both genes because it had the fewest *Rp1*-hybridizing restriction fragments (data not shown). The primers 2290F and 4890R were used to amplify and clone genomic fragments corresponding to part of the LRR-coding domain for sequencing. Of 65 clones sequenced from *HRp1-JF69*, 21 different sequences were observed (Figure 4). All 21 of the gene fragments had at least three informative differences between them. Genes in *HRp1-A* and *HRp1-H* were also partially sequenced. Forty-seven different sequences were observed among 77 clones sequenced from *HRp1-A*, and 52 were observed in 91 clones sequenced from *HRp1-H* (Figure 4). As with the genes from *HRp1-JF*, most of the sequences varied in many informative bases. The *Rp1-H* haplotype sequences all varied by at least four informative bases. Three of the *HRp1-A* sequences differed by only one informative base, but the remainder differed by at least three informative bases.

A neighbor-joining tree was generated using the sequences of the genes to visualize the relationships between the family members from the different *Rp1* haplotypes (Figure 4). The different *Rp1* haplotypes carry

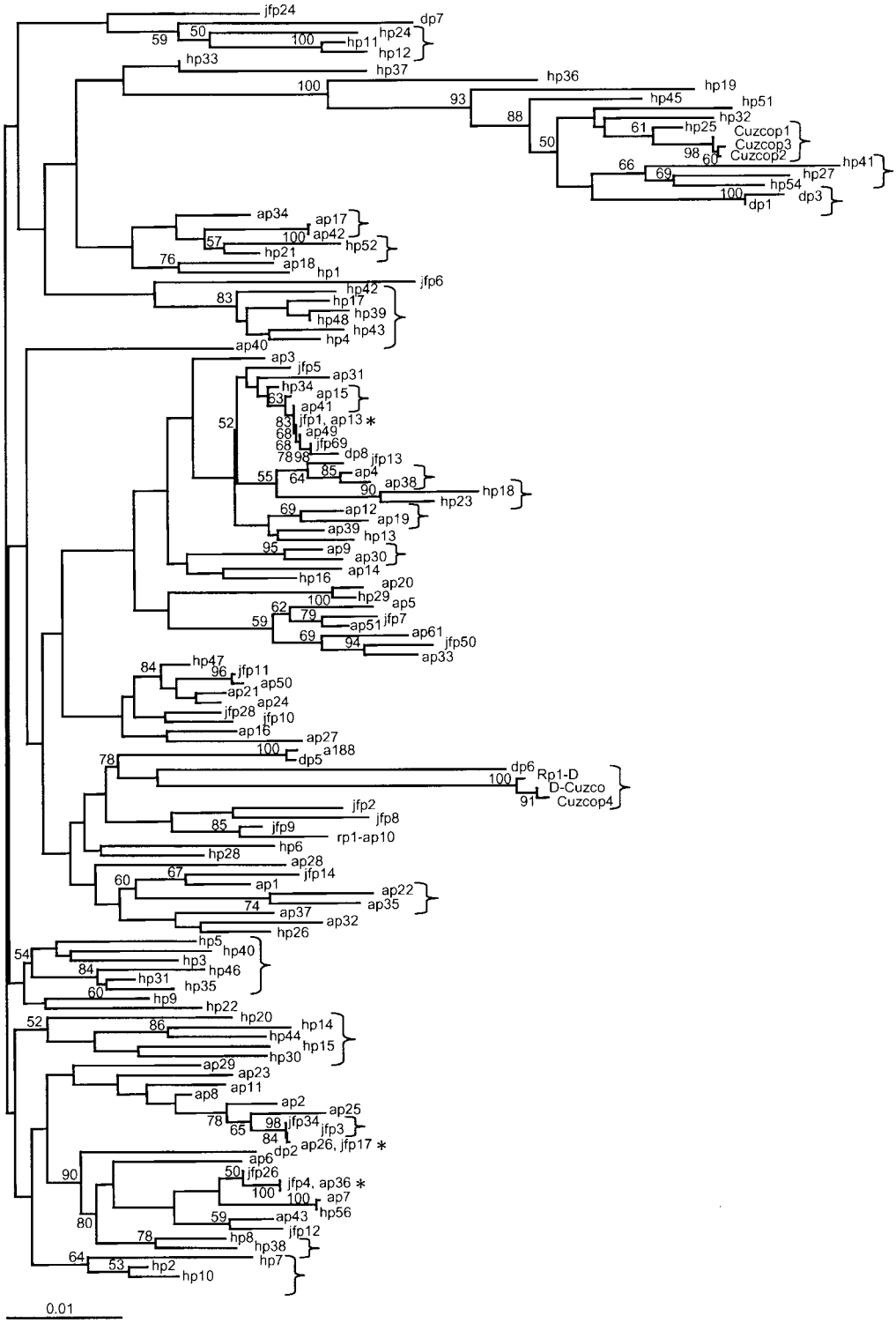

FIGURE 4.—A neighbor-joining tree of *Rp1* genes from *HRp1-D-Cuzco*, *HRp1-D-Kitale*, *HRp1-H*, *HRp1-A*, and *HRp1-JF*, generated using ~835 bases in the LRR-coding region of 132 *Rp1* sequences. Numbers at nodes indicate the frequency at which the genes cluster together in bootstrap analysis (percentage of 1000 trees generated) if >50%. Gene designations indicate haplotype (*a*, *HRp1-A*; *d*, *HRp1-D-Kitale*; *h*, *HRp1-H*; *jf*, *HRp1-JF69*; *a188*, *Rp1-A188*; and *Cuzco*, *HRp1-Cuzco*) and paralog number (p with number) corresponding to the specific paralog in that haplotype. Phenotypically detectable genes are indicated with an uppercase locus designation (*Rp1*). Braces indicate groups composed of genes from a single haplotype. Asterisks indicate identical genes in *HRp1-A* and *HRp1-JF69*. *Rp1* haplotypes contained from 1 to >50 different sequences.

different complements of *Rp1* genes and vary in their haplotypic diversity. Genes from unrelated haplotypes were always different in sequence. The genes clustered into a large number of different groups, as expected. Most of the groups include genes from more than one haplotype, but some of the groups were composed of genes from a single haplotype. This is similar to what was observed with genes in *HRp1-D-Cuzco* and demonstrates that genes in some *Rp1* haplotypes are evolving in a concerted manner (Figures 2 and 4).

If sufficient numbers of genes are sequenced from *HRp1-JF69*, *HRp1-A*, and *HRp1-H*, we reasoned it should be possible to identify genes that are identical or highly similar between the haplotypes. *HRp1-A* is very similar or identical to the *HRp1-F* parent of *HRp1-JF69*. Only three of the partially sequenced genes in each of the *HRp1-A*

and *HRp1-JF69* haplotypes appeared identical (Figure 4). It is possible one of the three identical genes identified in *HRp1-A* and *HRp1-JF69* controls the *Rp1-A/F* specificity but it is likely more identical genes would be identified if more genes were sequenced from these haplotypes. When >30 additional genes from *HRp1-A*, *HRp1-JF69*, and *HRp1-H* were sequenced, the majority of the sequences corresponded to previously identified genes with no more than eight sequences representing unidentified genes, suggesting that the majority of the genes in these haplotypes had been sampled efficiently. Consequently, it is likely these haplotypes carry more than the number of genes identified by sequencing. *HRp1-JF69* also carries the *Rp1-J* gene, which is identical in phenotype and race specificity to the *Rp1-H* gene in an unrelated haplotype. None of the sequences characterized were found in both haplotypes, so no candidates for the *Rp1-J/H* specificity were identified. This was again probably due to the large numbers of genes in these haplotypes, which were not exhaustively sampled by our cloning and sequencing efforts.

DISCUSSION

The second *Rp1-D*-carrying haplotype characterized shows some surprising contrasts to the first. Both haplotypes are phenotypically identical, carrying a gene conferring the *Rp1-D* specificity. The first haplotype characterized, from *Kitale*, carried *Rp1-D* (now designated *Rp1-D-Kitale*) and 7 other full-length genes capable of coding for proteins with amino acid identities ranging from 88 to 94% identical to the *RP1-D-Kitale* protein. This haplotype also carried a ninth gene that was transcribed, but truncated, coding for no LRR. Different pairs of genes in this haplotype sometimes had stretches of DNA sequence identity, as might be expected when family members recombine frequently, but none showed sequence affinities through the full coding region. In contrast, *HRp1-D-Cuzco* carried only 5 genes. None of these genes are closely related to the genes in *HRp1-D-Kitale*, other than the *Rp1-D-Kitale* gene. The two haplotypes are clearly not closely related or derived from each other by a small number of recombination events. Another interesting aspect of *HRp1-D-Cuzco* is that most of the genes are more similar to other members of the haplotype than to genes in the other haplotypes. The larger haplotypes we examined, like *HRp1-A* and *HRp1-H*, carried a great diversity of *rp1* genes, but also carried pairs or groups of highly similar genes. Similarly, when Ramakrishna *et al.* (2002) sequenced 4 of the estimated 15 genes from the B73 haplotype, 2 of the genes differed at only a single nucleotide. This is evidence that the genes are evolving in a concerted manner, as might be predicted for genes that mispair very frequently in meiosis (Hulbert *et al.* 2001). In most gene families, orthologs from different haplotypes are typically more similar in sequence than are paralogs in the same haplotype (Meyers *et al.* 1999). It is not clear why some haplotypes would show more evidence of within-haplotype homogenization than others. It is possible the haplotypes like *HRp1-D-Cuzco* evolved for an extended period in a population with limited variation at the locus.

The two *Rp1-D*-carrying haplotypes are unrelated, but both carry functional versions of the *Rp1-D* gene. These two versions have a very high level of sequence identity differing by only three nucleotides in the coding region, resulting in one nonsynonymous and two synonymous substitutions. Functional alleles at the *Rps2* locus of Arabidopsis were also found to be very similar. Five resistant accessions were found to have only one to four nucleotide substitutions (Caicedo *et al.* 1999) when compared to the *Rps2* allele from Colombia. The high levels of sequence identity among alleles conferring the same resistance specificities indicate these alleles may be fairly recent in their evolutionary origin. Genes conferring other *Rp1* specificities should be cloned from unrelated haplotypes (*e.g.*, *Rp1-J* and *Rp1-H*) to further examine whether these resistance specificities are recent or ancient in their evolutionary origins, but the large number of paralogs in many haplotypes will make their identification difficult.

Unrelated *Rp1* haplotypes typically have very different complements of *Rp1* genes. No identical genes were identified when genes from *HRp1-H*, *HRp1-A*, the two *Rp1-D*-carrying haplotypes, and the single gene from *A188* were compared. Typically, the different genes showed multiple informative polymorphic sites within the 835 bases compared. This, coupled with the fact that most maize lines appear to have different haplotypes when compared by gel blot analysis, indicates that maize germplasm carries many hundreds of different *Rp1* genes.

Different *Rp1* lines have different numbers of paralogs resulting in variability of haplotype size. Haplotype size at *Rp1* varies from 1 to >50 genes. It is highly likely there are more than the 47 and 52 genes identified in the *HRp1-A* and *HRp1-H* haplotypes, respectively, since the haplotypes were not sampled exhaustively and paralogs that were identical through the ~835-bp region sequenced would not be differentiated. Less dramatic differences in haplotype size have also been observed at the maize *Rp3* locus (Webb *et al.* 2002) and the soybean *Rsv1* locus (Hayes *et al.* 2004). Recently, genome sequence analysis of the *Rp1* complex in the maize line B73 has shed light on the nature of the intergenic regions (Ramakrishna *et al.* 2002). Regions between two pairs of *rp1* genes were sequenced and found to be ~38 and 68 kb in size. Considering these distances, the difference in size of a large *Rp1* haplotype with 50 or more genes and a haplotype with 1 or a few genes could be >2 Mb. These different size haplotypes could also vary considerably in the amount of *Rp1* transcript that is made. The bulk of the *Rp1* transcript is accounted for by 4 of the 9 genes in *HRp1-D-Kitale* (Sun *et al.* 2001) and 2 of the 5

genes in *HRp1-D-Cuzco*. Presumably the haplotypes with large numbers of genes have more genes that are transcribed, but this has not been demonstrated. It is often assumed that functional resistance genes are associated with a physiological cost to the plant in environments without the pathogen (Leonard 1997). There is now experimental evidence of this for the *Rpm1* gene of Arabidopsis, where lines engineered to carry the *Rpm1* gene were compared to isogenic lines without the gene (Tian *et al.* 2003). The *Rpm1* gene is unusual in that most accessions that do not carry the functional gene do not carry homologous sequences (Grant *et al.* 1998). Most R genes, like most other genes, have alternative alleles, which might be expected to exhibit similar physiological costs. A complex locus like *Rp1* would be more similar to a polymorphism like *Rpm1* in some respects, because different haplotypes would be expressing and maintaining different numbers of genes. A large haplotype probably has a higher physiological cost than a small one. Some recombinant *Rp1* haplotypes have been associated with adult-plant chlorotic spotting phenotypes in some genetic backgrounds that probably have a physiological cost (Hu *et al.* 1997; Hulbert and Drake 2000), but it is not known if this is associated with the number or expression of *Rp1* genes or due to specific genes or gene combinations.

In conclusion, characterization of different *Rp1* haplotypes found a wide range of haplotypic size and diversity. Unrelated haplotypes carry different genes and very different numbers of genes. Some of the genes in some haplotypes are more similar in sequence to each other than to genes in other haplotypes, suggesting that the genes in these haplotypes are evolving in a concerted manner. Comparisons of unrelated haplotypes that carry phenotypically identical genes could be a good method of identifying genes controlling the phenotype, but it may be necessary to sequence many genes to identify the common one.

This article is based upon work supported by the National Science Foundation under grant 0090883 and is publication 03361J from the Kansas Agricultural Experiment Station.

LITERATURE CITED

Bai, J., L. Pennill, J. Ning, S. W. Lee, J. Ramalingam *et al.*, 2002 Diversity of nucleotide binding site-leucine-rich repeat genes in cereals. Genome Res. **12:** 1871–1884.

Baker, B., P. Zambryski, B. Staskawicz and S. P. Dinesh-Kumar, 1997 Signaling in plant-microbe interactions. Science **276:** 726–733.

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

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.

Collins, N., J. Drake, M. Ayliffe, Q. Sun, J. Ellis *et al.*, 1999 Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants. Plant Cell **11:** 1365–1376.

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

Grant, M. R., J. M. McDowell, A. G. Sharpe, M. de Torres 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. G. Jones, 1997 Plant disease resistance genes. Annu. Rev. Plant Physiol. Plant Mol. Biol. **48:** 575–607.

Hayes, A. J., S. C. Jeonga, M. A. Gorea, Y. G. Yua, G. R. Bussa *et al.*, 2004 Recombination within a nucleotide-binding-site/leucine-rich-repeat gene cluster produces new variants conditioning resistance to soybean mosaic virus in soybeans. Genetics **166:** 493–503.

Hooker, A. L., and P. M. LeRoux, 1957 Sources of protoplasmic resistance to puccinia sorghi in corn. Phytopathology **47:** 113–187.

Hooker, A. L., and W. A. Russell, 1962 Inheritance of resistance to *Puccinia sorghi* in six corn inbred lines. Phytopathology **52:** 122–128.

Hu, G., C. A. Webb and S. H. Hulbert, 1997 Adult-plant phenotype of the *Rp1-DJ* compound rust resistance gene in maize. Phytopathology **87:** 236–241.

Hulbert, S. H., and J. A. Drake, 2000 Rust-resistant *sh2* sweet corn populations. Hortscience **35:** 145–146.

Hulbert, S. H., P. C. Lyons and J. L. Bennetzen, 1991 Reactions of maize lines carrying *Rp* resistance genes to isolates of the common rust pathogen, *Puccinia sorghi*. Plant Dis. **75:** 1130–1133.

Hulbert, S. H., M. A. Sudupak and K. S. Hong, 1993 Genetic relationships between alleles of the *Rp1* rust resistance locus of maize. Mol. Plant-Microbe Interact. **6:** 387–392.

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

Leonard, K. J., 1997 Modeling gene frequency dynamics, pp. 211–230 in *The Gene-for-Gene Relationship in Plant-Parasite Interactions*, edited by I. R. Crute, E. B. Holub and J. J. Burdon. CAB International, London.

Meyers, B. C., A. W. Dickerman, R. W. Michelmore, R. M. Pecherer, S. Sivaramakrishnan *et al.*, 1999 Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J. **20:** 317–332.

Ramakrishna, R., J. Emberton, M. Ogden, P. SanMigual and J. Bennetzen, 2002 Structural analysis of the maize Rp1 complex reveals numerous sites and unexpected mechanisms of local rearrangement. Plant Cell **14:** 3213–3223.

Sun, Q., N. C. Collins, M. Ayliffe, S. M. Smith, J. Drake *et al.*, 2001 Recombination between paralogues at the *rp1* rust resistance locus in maize. Genetics **158:** 423–438.

Tian, D., M. B. Traw, J. Q. Chen, M. Kreitman and J. Bergelson, 2003 Fitness costs of R-gene-mediated resistance in Arabidopsis thaliana. Nature **423:** 74–77.

Webb, C. A., T. E. Richter, N. C. Collins, M. Nicolas, H. N. Trick *et al.*, 2002 Genetic and molecular characterization of the maize *rp3* rust resistance gene. Genetics **162:** 381–394.

Wilkinson, D. R., and A. L. Hooker, 1968 Genetics of reaction to *Puccinia sorghi* in ten corn inbreds from Africa and Europe. Phytopathology **58:** 605.

Communicating editor: J. A. Birchler