The Tomato Cf-5 Disease Resistance Gene and Six Homologs Show Pronounced Allelic Variation in Leucine-Rich Repeat Copy Number

Mark S. Dixon,1,2 Kostas Hatzixanthis,1 David A. Jones,3 Kate Harrison,4 and Jonathan D. G. Jones5
Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich, NR4 7UH, United Kingdom

The tomato Cf-2 and Cf-5 genes confer resistance to Cladosporium fulvum and map to a complex locus on chromosome 6. The Cf-5 gene has been isolated and is predicted to encode a largely extracytoplasmic protein containing 32 leucine-rich repeats (LRRs), resembling the previously isolated Cf-2 gene, which has 38 LRRs. Three haplotypes of this locus from Lycopersicon esculentum, L. pimpinellifolium, and L. esculentum var cerasiforme were compared, and five additional homologs of Cf-5 were sequenced. All share extensive sequence identity, particularly within the C-terminal portions of the predicted proteins. In striking contrast to the Cf-9 gene family, six of seven homologs in the Cf-2/Cf-5 gene family vary in LRR copy number, ranging from 25 to 38 LRRs. Cf-5 and one adjacent homolog differ by only two LRRs. Recombination events that vary the LRR copy number in this region could provide a mechanism for the generation of new specificities for recognition of different ligands. A recombination breakpoint between the Cf-2 and Cf-5 loci was fully characterized and shown to be intragenic.

INTRODUCTION

Plants have evolved many mechanisms to protect themselves from pathogens. Some of these systems are passive, such as the production of preformed antimicrobial compounds (Osbourn, 1996), whereas others involve the induction of defense mechanisms when a specific pathogen is recognized (Hammond-Kosack and Jones, 1996). Dominant or semidominant resistance (R) genes in plants are thought to act as molecular receptors for pathogen-derived factors, which are the products of avirulence (Avr) genes (reviewed in Staskawicz et al., 1995). This specific interaction was first described for flax and flax rust and led to Flor’s classic gene-for-gene hypothesis (Flor, 1946).

The first cloned R gene involved in gene-for-gene recognition was Pto, which confers resistance to tomato to bacterial spot caused by Pseudomonas syringae strains that carry the AvrPto gene (Martin et al., 1993). Pto encodes a protein kinase that has been shown to interact directly with the Avr gene product AvrPto (Scofield et al., 1996; Tang et al., 1996). Several R genes have now been isolated, and these form separate classes based on their sequence identity (Bent, 1996; Hammond-Kosack and Jones, 1997). The most abundant class encodes cytoplasmically localized proteins containing a predicted nucleotide binding site and multiple leucine-rich repeats (LRRs) near the C terminus. This class is typified by the R genes N from tobacco (Whitham et al., 1994), L6 from flax (Lawrence et al., 1995), and RPS2, RPM1, and RPM5 from Arabidopsis (Bent et al., 1994; Mindrinos et al., 1994; Grant et al., 1995; Parker et al., 1997). Another distinct group of R genes comprises the tomato Cf-2, Cf-4, and Cf-9 genes, which confer race-specific resistance to the leaf mold pathogen Cladosporium fulvum (Jones et al., 1994; Dixon et al., 1996; Thomas et al., 1997). The Cf genes encode membrane-anchored proteins largely composed of extracytoplasmic LRRs. The Xa21 gene of rice contains components similar to both the Pto and Cf genes, encoding a transmembrane receptor-like protein kinase with 23 extracellular LRRs (Song et al., 1995).

Several different Cf genes confer resistance to specific races of C. fulvum (Stevens and Rick, 1988) and have been bred into cultivated tomato to generate near isogenic lines (see Methods). Cf-2 and Cf-9 were identified in Lycopersicon pimpinellifolium, and Cf-5 was identified in the land race L. esculentum var cerasiforme (Dickinson et al., 1993; Jones et al., 1993). The Cf-4 gene from L. hirsutum and the Cf-9 gene map to allelic complex loci on the short arm of chromosome 1 (Jones et al., 1994; Thomas et al., 1997).

1 These authors contributed equally to this work.
2 Current address: School of Biological Sciences, University of Southampton, Southampton, SO16 7PX, United Kingdom.
3 Current address: Plant Cell Biology, Research School of Biological Sciences, Australian National University, P.O. Box 475, Canberra ACT 2601, Australia.
4 Current address: Institute of Chemical and Physical Research (RIKEN) Frontier Research Program, Laboratory for Photoreception and Signal Transduction, Hirosawa 2-1, Wako, Saitama 351-01, Japan.
5 To whom correspondence should be addressed. E-mail jonathan.jones@bbsrc.ac.uk; fax 44-1603-250024.
A major biological question regarding R gene loci concerns whether the mechanisms by which allelic diversity is generated at these loci match pathogen evolution. Detailed comparisons with the allelic locus from susceptible L. esculentum have implicated multiple recombination and/or gene conversion events between the Hcr9 (for homologs of *Cladosporium* resistance gene Cf-9) genes in their evolution (Parniske et al., 1997). Cf-2 and Cf-5 map to a complex locus on chromosome 6 (Dickinson et al., 1993), and in the course of cloning Cf-2, a family of Hcr2 genes was isolated from both L. esculentum and L. pimpinellifolium (Dixon et al., 1995, 1996). Here, we report the isolation of the Cf-5 gene and characterization of the complex locus from three genotypes. By comparing gene sequences, we have identified a region likely to be involved in ligand recognition. This result suggests how unequal recombination could generate novel variation within a specific functional domain of a class of genes. In marked contrast to the Hcr9 gene family (Parniske et al., 1997), the members of the Hcr2 gene family differ dramatically in that each has a different number and/or combination of LRRs.

**RESULTS**

**Identification of Binary Cosmid Clones Carrying Cf-5**

A line containing Cf-2 (Cf2) was crossed to a line containing Cf-5 (Cf5) to generate a transheterozygous F1 plant carrying a single copy of each resistance locus. This F1 plant was crossed to a line lacking any detectable Cf genes (Cf0), and ~12,000 testcross progeny were screened for susceptible plants by inoculation with *C. fulvum*. All plants should carry either the Cf-2 or Cf-5 resistance genes and therefore be resistant, except when recombination has occurred between the two loci. This strategy led to the isolation of a single susceptible individual (V454) that was shown to carry a recombinant chromosome 6 by using restriction fragment length polymorphism markers closely linked to Cf-2 (Dixon et al., 1996). When we used a DNA probe hybridizing with part of Cf-2 as well as with an adjacent, related gene, DNA gel blot analysis revealed a pattern that had some but not all features of the pattern for both Cf-2 and Cf-5 genotypes. These data taken together suggested that the Cf-5 gene is not just closely linked to Cf-2 but is likely to be part of an allelic locus.

To clone the Cf-5 gene, we used genomic DNA from plants homozygous for Cf-4 and Cf-5 to construct a binary vector cosmid library (see Methods). A 3.8-kb XhoI fragment of Cf-2 was used to screen slot blots and gel blots of cosmid DNA from pools of the library. Several pools were identified and found to contain hybridizing cosmid clones, and individual positive clones were isolated by standard procedures. By a combination of restriction mapping, fingerprinting, and DNA hybridization, the cloned insert DNA in each cosmid was mapped and assembled into a single contig (Figure 1A). The assembled contig contained three main Cf-2-hybridizing regions (B1 to B6) and one small region of hybridization (B7). All major Cf-2-hybridizing bands on genomic DNA gel blots of Cf5 plants were represented within this Cf-5 cosmid contig.

**Complementation in Transgenic Plants Reveals Copy B2 to Be Cf-5**

Five independent cosmid clones (4/5-8, 4/5-9, 4/5-42, 4/5-52, and 4/5-96; Figure 1B) were mobilized into Agrobacterium; however, in each case except for cosmid 4/5-8, the cosmid insert DNA underwent recombination such that some DNA sequence was lost. Each of the cosmids (4/5-9, 4/5-42, 4/5-52, and 4/5-96) that underwent this recombination contained more than one region hybridizing with the Cf-2 DNA probe. Molecular analysis suggested that recombination between these regions of similarity was responsible for loss of sequence. Mobilization into Agrobacterium of a cosmid carrying, for example, more than one region of hybridization with Cf-2 resulted in the generation of a deleted cosmid containing a single chimeric region of homology to Cf-2. To overcome the generation of chimeric genes, we subcloned cosmid 4/5-42 to generate two new cosmids (KB2 and minHcr2-5D; Figure 1C), each containing a single complete region that hybridized with Cf-2.

Cosmids 4/5-8, KB2, and minHcr2-5D were mobilized into Agrobacterium and transformed into susceptible Cf0 plants. Transgenic plants were selected and inoculated with a race of *C. fulvum* that contains the β-glucuronidase gene (race 4 GUS). Three of six plants transformed with cosmid KB2 were resistant to *C. fulvum*, whereas none of the seven plants transformed with 4/5-8 and none of the 13 transformed with minHcr2-5D were resistant. One of the resistant KB2-transformed plants was polyplody as a consequence of the transformation process and produced no viable seed. Progeny from the remaining nonpolyplody KB2-transformed plants were screened with races of *C. fulvum* either containing or lacking *Avr5* (race 4 GUS compared with race 5). All progeny were susceptible to *C. fulvum* lacking *Avr5* (race 5), whereas ~75% of progeny from each transformant were resistant to *C. fulvum* carrying *Avr5* (race 4 GUS) (data not shown). These data demonstrate that *Avr5*-dependent resistance to *C. fulvum*, that is, Cf-5, must be located within the 13-kb BamHI fragment of cosmid KB2.

To further define the Cf-5 gene, we subcloned a 4.9-kb BamHI-EcoRI fragment containing the entire B2 region of homology to Cf-2 into the corresponding sites of pCLD04541 to create the plasmid minCf-5 (Figure 1C). Plasmid minCf-5 was mobilized into Agrobacterium and transformed into susceptible (Cf0) plants. Transgenic plants were selected and inoculated with *C. fulvum* race 4 GUS. Of nine transformants, four were fully resistant, whereas two were partially resistant. This result is entirely consistent with the semidominant nature of Cf genes and the implication that expression
levels affect the degree of resistance (Hammond-Kosack and Jones, 1994).

**Structure of the Cf-5 Gene**

The DNA sequence of the Cf-5 gene carried in plasmid minCf-5 was determined and contained a single major open reading frame encoding a polypeptide of 968 amino acids (Figure 2). Analysis using BESTFIT (Genetics Computer Group, Madison, WI) showed that Cf-5 is 90% identical and 93.3% similar to Cf-2. Like Cf-2, Cf-5 can be divided into seven domains. Domains B, C, and D show homology to the polygalacturonase inhibitor proteins and form the major portion of the predicted protein. Domain C consists entirely of 32 LRRs; several of these can be classified into two subgroups (types A and B). Subgroup A and B LRRs alternate in a pattern similar to that previously described for Cf-2 (Dixon et al., 1996). Compared with Cf-2, Cf-5 lacks a number of LRRs (Figure 2, dashes) and contains a number of other amino acid differences, which are shown in red (Figure 2). As in Cf-4 (Thomas et al., 1997) and the other Hcr9 proteins (Pamiske et al., 1997), amino acid differences are most frequent at positions flanking the leucine residues in the LxxLx-LxxN canonical LRR motif (where x can be any amino acid). This region is predicted to form a discriminatory parallel β sheet. Domain A is a predicted signal peptide of 26 amino acids. Domain F is a potential membrane-spanning region containing 24 uncharged amino acids. Domains E and G are rich in acidic and basic amino acids, respectively, which is consistent with a role in anchoring and orienting the protein within the cell membrane. Domains B to E contain 25 potential N-linked glycosylation sites.

Cf-5 is likely to be transcriptionally processed in a manner similar to Cf-2, because several well-characterized transcriptional control sequences are conserved. A putative TATA box, preceded by two CAAAT boxes, is found only 46 bases upstream of the first initiation codon. The next closest TATA box is 400 nucleotides upstream of the first and lacks any apparent CAAAT boxes. Forty-one bases downstream of the termination codon, there is conservation of sequences corresponding to an intron in Cf-2, with the intron-exon boundaries being completely conserved. Similarly, the polyadenylation signal 371 bases downstream of the termination codon is absolutely conserved.

**Comparison of the Cf0, Cf2, and Cf15 Haplotypes**

Previously, we reported the isolation and partial characterization of allelic sequences from Cf0 and Cf2 plants (Dixon et al., 1995, 1996). Extensive analysis of these cloned loci, using subcloned regions of the Cf-2 contig as hybridization probes, has shown that the cloned Cf-5 contig is colinear.
with these other loci (Figure 3). Figure 3 shows boxed regions representing sequences within these three loci that hybridize with the Cf-2 gene probe and that have been designated Hcr2 genes. The second digit denotes the genotype, and the final letter designation refers to the relative position within each haplotype. When this nomenclature is used, Cf-5 can also be referred to as Hcr2-5C.

DNA gel blot analysis revealed that all but one of the Hcr2 genes at the Cf-5 locus had been isolated. Hcr2-5A, corresponding to homologs Hcr2-0A and Hcr2-2A, was not cloned. A DNA probe upstream of Hcr2-2A hybridizes with the adjacent region of Hcr2-0A but not to regions adjacent to the other Hcr genes or Cf genes in the three contigs. This same probe detected sequences on DNA gel blots of genomic DNA from plants carrying Cf-5, indicating the presence of an Hcr2-5A (data not shown). It is likely that Hcr2-0A, Hcr2-2A, and Hcr2-5A are truly orthologs within the Cf gene family.

### Sequence of Other Hcr Genes

The individual Hcr2 genes from each cloned contig were subcloned and sequenced. For each, a major open reading frame, with homology to the Cf-2 and Cf-5 genes, was maintained (Figure 4). Only the small region of homology designated B4 in the Cf-5 contig (Figure 1) carried an incomplete homolog. B4 contained sequences corresponding only to the most N-terminal regions and lacks any homology to the remainder of the Cf genes (data not shown).

All of the Hcr2s sequenced (Hcr2-0A, Hcr2-0B, Hcr2-2A, Hcr2-5B, and Hcr2-5D) encode polypeptides that can be divided into the same seven general domains as Cf-2 and Cf-5 (A to G). The most notable differences between all of these predicted proteins are that they each contain different numbers and/or combinations of LRRs and that this variation occurs within the region of highly conserved alternating A/B-type

**Figure 2. Primary Structure of the Cf-5 Protein.**

The amino acid sequence predicted from the sequence of the Cf-5 gene is shown divided into seven domains (A to G), as described in the text. The sequence has been aligned with Cf-2; amino acid differences are shown in red, and dashes indicate the absence of amino acids in Cf-5. In domain C, the conserved L residues of the LRRs are often replaced by V, F, I, or M. The positions of these conserved residues are marked by asterisks at the top of domain C. Numbers to the right of domain C indicate the specific LRR number, and letters A and B to the right of the sequence indicate the type of LRR (see text). Underlined residues indicate potential N-linked glycosylation sites. Vertical lines bracket the region predicted to form the solvent-exposed β sheet. The Cf-5 gene sequence has GenBank accession number AF053993.

**Figure 3. Analysis of Allelic Regions of Chromosome 6.**

Shown is a schematic representation of allelic regions of chromosome 6 from the near-isogenic lines C0, C2, and C5 and the rare recombinant between Cf-2 and Cf-5 (V454). Boxed regions represent the open reading frames of the Cf genes and their homologs; arrows indicate the direction of transcription. Filled boxes show DNA derived from L. esculentum, open boxes show DNA derived from L. pimpinellifolium, and hatched regions indicate DNA identified by DNA gel blot analysis but not cloned. The diagonal line between Cf-2.2 and Hcr2-5B shows the approximate position of the recombination event between these two genes. The bar indicates 5 kb; however, distances for the uncloned region (broken lines) are not known and cannot be compared.
LRRs (Figure 4). Apart from the dramatic variation in the number of LRRs, several other amino acid differences exist between the Cf-2, Cf-5, and Hcr genes, although these tend to be distributed mainly throughout domain C. Significantly, Hcr2-5D is almost identical to Cf-5, only differing by a single amino acid substitution adjacent to two additional LRRs (Figure 5).

Transgenic plants containing Hcr2-5D do not confer resistance to C. fulvum, despite the near identity to Cf-5. To confirm this result, expression of Hcr2-5D was examined. The high degree of identity between Hcr2-5D and the Hcr2-0B homolog in the susceptible Cf0 line used for transformation precluded the use of RNA gel blot analysis for this purpose. Instead, reverse transcriptase–polymerase chain reaction (RT-PCR) experiments were performed with poly(A)⁺ RNA extracted from cotyledons of progeny from two Cf-5 and three Hcr2-5D transformants as well as from the original nontransgenic Cf5 line (Figures 6A and 6B). Primers CF5F2 and CF5R2, designed to anneal to a conserved 3’ portion of Cf5 genomic DNA but not to Hcr2-0B, amplified a product of ~1.6 kb from Cf5 genomic DNA but not from Cf0 genomic DNA, thus showing their specificity. These primers amplified a smaller product of ~1.4 kb from all RNA samples, indicating expression of the transgenes. The size difference between the products from genomic and cDNA confirms the presence of the predicted intron in the 3’ untranslated sequences and proves that the amplification is specific and not the result of contaminating genomic DNA (Figure 6A). Primers 2-5CF and 2-5CR were designed for sequences flanking the two additional LRRs in Hcr2-5D and therefore reveal characteristic size polymorphisms for the two genes and transcripts corresponding to Cf-5 and Hcr2-5D. These primers clearly reveal the expression of both transcripts in the original nontransgenic line Cf5 and the expression of individual transcripts in each of the five transgenic plants examined (Figure 6B).

Characterization of the Recombination Event in V454

Detailed DNA gel blot analysis of the recombinant V454 suggested that the crossover event between the Cf-2 and Cf-5 parental chromosomes was unequal and resulted in the loss
of several Hcr2 genes. The V454 recombinant contains Hcr2-5A and a chimeric gene consisting mainly of Hcr2-5B and a small part of Cf-2.2 (Figure 3). Using the sequence of these two genes, it was possible to design specific synthetic oligonucleotide primers for use in PCR that could amplify the site of recombination in V454. Primer V3F was designed to anneal to Hcr2-5B and no other sequenced homolog in Cf2 or Cf5 plants, whereas primer V3R was designed to fulfill the same specifications and anneal only to the Cf-2 genes (see Methods). Used in combination, these primers amplified a novel product in the recombinant plant V454 that did not amplify from Cf2 or Cf5 plant DNA. Sequencing of the novel product confirmed that V454 carried a chimeric sequence of Hcr2-5B and Cf-2.2 (Figure 7). The point of recombination can be located only to a region between the two closest flanking heterologous nucleotides, which in this case span a region of 107 identical nucleotides. This recombination event occurred within sequences encoding parts of domains A and B (Figure 7). In contrast, previously described recombination events between the Cf-9 and Cf-4 loci were in intergenic regions (Parniske et al., 1997).

DISCUSSION

Structure of the Cf-5 Gene Product

The Cf-5 gene closely resembles the Cf-2 genes because it encodes a membrane-anchored extracytoplasmic protein, as might be predicted for a receptor evolved to detect an extracellular pathogen. Like Cf-2, an extensive region of Cf-5 comprises extracytoplasmic LRRs (Dixon et al., 1996; Jones et al., 1996), 28 of which are exactly 24 amino acids in length. In addition, a region within domain C contains highly conserved LRRs, which can be classified into subgroups (A and B) based on their sequence. Type A repeats have the consensus EEIGL(R/S)LXNLGXXLNL(S/F)GSIP, whereas type B repeats have the consensus ASLGNLNL(S/F)XLXLYNN(Q/K)LSGSIP. Seven of each of these repeats form an alternating region within domain C. The uninterrupted region of LRRs is proposed to give the Cf proteins a highly ordered reiterated structure, although to date, the crystal structure of an extracellular-type LRR has not been determined. The Cf LRRs are predicted to form a stacked structure with the LxxLxLxx part of each LRR contributing to a solvent-exposed face composed of parallel β sheets (Jones et al., 1996). Indeed, modeling studies with a Cf-2 LRR strongly support this hypothesis (Kajava, 1998). Interestingly, within the LxxLxLxx region of the types A and B LRRs, 43% of the amino acids vary between Cf-5 and Cf-2, whereas outside this region but still within the A/B repeat region, only 9% differ. The leucine residues never vary, but the amino acids that are predicted to be solvent exposed in this region of parallel β sheet are hypervariable (Figure 2). Detailed analysis of Cf-9 and many Hcr9 proteins indicate that the corresponding region is hypervariable and potentially contributes to recognition specificity (Parniske et al., 1997; Thomas et al., 1997).

Structure of the Cf-0, Cf-2, and Cf-5 Allelic Loci

The allelic loci from the three genotypes Cf0, Cf2, and Cf5 each carry a different number of tandemly arranged Cf-like genes (two, three, and four, respectively). It is likely that these genes evolved from a common progenitor that existed before speciation and that duplication and unequal exchange events have resulted in the present gene arrangements. Phylogenetic tree analysis of Cf-2, Cf-5, and the Hcr proteins clearly shows that the predicted proteins fall into two classes: Hcr2-0A and Hcr2-2A form one group, and the Cf and remaining Hcr proteins form the second (Figure 8). Figure 4 demonstrates the relatedness of Hcr2-0A and Hcr2-2A and shows that they share the most similar combination of LRRs. BESTFIT analysis of the nucleotide sequences further clarifies this subclassification. Sequences 5’ of the translation initiation codon in Hcr2-0A and Hcr2-2A share >91% identity, whereas no significant homology can be found between these sequences and the upstream regions from any of the other Cf or Hcr genes. Hcr2-0A and Hcr2-2A originate from different Lycopersicon species, yet the similarity they share is greater than with the adjacent related genes in their respective species, suggesting that duplication and divergence must have occurred before speciation.
Similarly, the Xa21 gene family from rice comprises seven members that have been classified into two groups, suggesting an ancient gene duplication event (Song et al., 1997). In contrast to members of the Cf-2/Cf-5 locus, retrotransposon insertion and nucleotide changes disrupt the open reading frame in all but two of the Xa21 gene family members. Perhaps the Hcr2 genes arose relatively recently and underwent rapid evolution to create diversity, or maybe there is a selection pressure placed on them that requires maintaining the open reading frame in each homolog.

Recombination in V454

From the Cf-2/Cf-5 recombinant plant, V454, the DNA encompassing a rare crossover event was identified and sequenced. The plant V454 was susceptible because the recombinant chromosome carries no complete Cf resistance gene. Recombination had occurred between Cf-2.2 and Hcr2-5B in a region encoding domains A and B. The location of the crossover event was delimited to an interval of 107 bases between the two closest flanking polymorphic nucleotides. The chimeric gene in V454 encodes domain A of Cf-2.2 and domains B to G of Hcr2-5B. Because domain A is a predicted signal peptide and would be cleaved from any mature protein, the new chimeric protein would effectively be identical to Hcr2-5B. However, any differences in promoter activity would result in altered Hcr2-5B expression.

Interestingly, a comparison of the members of the Xa21 gene family indicates that past intragenic recombination has occurred within the very same region of the reading frame coding for the signal peptide and sequences just upstream of the LRRs (Song et al., 1997). This region in the Xa21 family members is GC rich and is seemingly significant in influencing specific recombination in the domain; however, no such analogous GC-rich region exists in the Hcr2 genes.

Contrasting Consequences of Evolution at the Hcr2 and Hcr9 Loci

Analysis of Cf-5 and the Hcr2 genes has identified a number of interesting features. Like the Hcr9 genes, the Hcr2 genes map to a locus carrying multiple tandemly arranged linked homologs. Detailed comparison of Cf-5, Hcr2-5B, and the Hcr9 genes provides evidence for recombination and gene conversion events that probably generated much of the gene diversity. This shuffling of segments of the Hcr9 genes has led to hypervariable amino acids at the presumed solvent-exposed residues of the parallel β sheet within the LRR.

Figure 7. Sequence of Recombination Product in V454.

Alignment of portions of Cf-2.2, Hcr2-5B, and the recombinant gene in V454 is shown. Vertical lines indicate identities, and gaps show mismatches. The amino acid sequence shown corresponds to that encoded by the recombinant gene within V454. Numbers to the left of the DNA sequences indicate the relative positions within each sequence. The upper alignment block corresponds to domain A, whereas the lower alignment block corresponds to the first portion of domain B. The sequence in lowercase letters indicates the region in which the crossover occurred.

Figure 8. Phylogenetic Relationship between Cf-2.1, Cf-5, and the Hcr2 Proteins.

The C-terminal portions of Cf-2.1, Cf-5, and the Hcr2s (covering the last 14 LRRs and domains D, E, and F) were aligned using the Clustal method (Higgins and Sharp, 1988) and plotted as a phylogenetic tree. The scale bar indicates the degree of amino acid dissimilarity. Hcr2-5D is not shown because it is indistinguishable from Cf-5 in this analysis.
domain. Exactly the same phenomenon of hypervariability is clearly apparent when Cf-5 is compared with Cf-2 (Figure 2).

In striking contrast to the Hcr9 proteins, the Hcr2 proteins all contain different numbers and/or combinations of types A and B LRRs (Figure 4). This is probably a consequence of the extremely regular sequence of the LRR motifs within the Hcr2 genes that provides increased scope for unequal exchange events or slippage during DNA replication. The LRRs of the Hcr9 proteins are much less regular in length and sequence, reducing the possibility of DNA misalignment before recombination. However, unequal crossover events must still be possible within the Hcr9 genes because Cf-4 encodes two fewer LRRs than does Cf-9 (Thomas et al., 1997).

A general feature of the Hcr2 proteins is that they have alternating types A and B LRRs. However, Hcr2-2A contains three consecutive A-type repeats, and Cf-2 contains four consecutive B-type repeats. This may reflect that either one or more ancestral Hcr2 proteins had stretches of multiple type A or B repeats or that on occasion, unequal recombination can occur when types A and B repeats align, perhaps at the nucleotides encoding the shared amino acids GSIP. It may be possible to examine this by selecting new recombination events within the Cf-2/Cf-5 locus. Alternatively, the presence of several consecu-
tive type A or B LRRs may support the idea of DNA slippage during replication. The mechanisms of recombination and slippage are not mutually exclusive, and both could contribute to the observed variation in LRR copy number.

It is probable that many recombination events between Cf-2 and Cf-5 were not detected during the selection of V454, because the screen was only for completely susceptible individuals. Any recombination events giving rise to resistant plants would have been missed. Recombination in plants can occur between small regions of identity. Analysis of 130 intragenic recombinants of the maize locus. Alternatively, the presence of several consecu-
tive type A or B LRRs may support the idea of DNA slippage during replication. The mechanisms of recombination and slippage are not mutually exclusive, and both could contribute to the observed variation in LRR copy number.

It is probable that many recombination events between Cf-2 and Cf-5 were not detected during the selection of V454, because the screen was only for completely susceptible individuals. Any recombination events giving rise to resistant plants would have been missed. Recombination in plants can occur between small regions of identity. Analysis of 130 intragenic recombinants of the maize bronze gene identified one recombination junction located at a 19-bp interval and six others at intervals of <30 bp (Dooner and Martínez-Férez, 1997).

Inferences about Function from Hcr2 Protein Structure

The Cf proteins are thought to work as molecular receptors that enable the plant to detect the presence of C. fulvum Avr gene products and then initiate the activation of specific defense responses. The C-terminal portions of the Cf proteins show greatest similarity and are proposed to have a role in signaling (Dixon et al., 1996; Thomas et al., 1997). In the absence of an extensive or highly conserved cytoplasmic region, any signaling role would most likely involve interaction with a protein that is integral to or closely associated with the plasma membrane. Central to this model for Cf protein function is the concept that the highly variable regions within the LRRs are responsible for recognition of pathogen-encoded avirulence determinants either directly or indirectly through some coreceptor. Thomas et al. (1997) have shown that the Cf-4 and Cf-9 proteins have identical C termini yet recognize distinct avirulence factors Avr4 and Avr9. Here, we describe Cf-5 and Hcr2-5D, which differ in only a small region of domain C yet have distinct specificities: Cf-5 confers resistance against C. fulvum races carrying the Avr5 gene, whereas Hcr2-5D does not. We propose that the presence of the two additional LRRs and/or the single amino acid change in Hcr2-5D disrupts a region involved in specific recognition of Avr5 (Figure 5), thus defining this region with unprecedented precision. The single amino acid change is immediately adjacent to the two additional LRRs; therefore, together they constitute only a single change at the nucleotide level. Based on the predictions of Parniske et al. (1997), the single residue change adjacent to the two additional LRRs in Hcr2-5D is unlikely to alter any recognition specificity because it falls outside the xxLxLxx domain hypothesized to form the discriminatory surface. In addition, the serine-to-alanine change is a conservative substitution. A more plausible explanation is that specific amino acids within LRRs both above and below the point of the two-LRR insertion in Hcr2-5D are required to provide Avr5 specificity in Cf-5. The insertion of two additional LRRs into the middle of any recognition domain, as in Hcr2-5D, would alter the spacing between amino acids on the discriminatory surface, thus disrupting any specific interaction/recognition event. The presence of very highly conserved types A and B LRRs within domain C increases the likelihood of unequal crossover events in a region of the Cf genes responsible for encoding recognition specificities. This would shuffle LRRs, mixing existing variability to generate new Hcr genes encoding novel combinations of LRRs and potentially new recognition capabilities.

Thomas et al. (1997) and Parniske et al. (1997) propose that the consecutive LRRs within the Cf proteins create a structure upon which the variant amino acids within each repeat confer recognition capacity. Cf-4 and Cf-9 either directly or indirectly recognize fungal avirulence determinants encoded by Avr4 and Avr9, respectively. These encode secreted proteins of 28 and 106 amino acids (van Kan et al., 1991; J oosten et al., 1994), of which Avr9 is known to form a small cystine-knot protein (Vervoort et al., 1997). If the interaction is direct, then these small peptides might only contact a few of the LRRs in the recognition domains of Cf-4 and Cf-9. Given the large number of consecutive LRRs in the Cf proteins, each Cf protein would probably have the potential to recognize many different ligands, thereby increasing the potential of each Cf gene as an R gene. The RPM1 gene of Arabidopsis recognizes two distinct pathogen avirulence genes, avrB and avrRPM1 (Bisgrove et al., 1994). Two Hcr genes of Cf-9 have the capacity to specify Avr9-independent resistance against C. fulvum in adult plants and therefore must contain additional recognitional specificities (Parniske et al., 1997).

Recombination and the Generation of New Resistance Specificities

Many of the features of the Hcr2 and Hcr9 loci are reflected in other characterized R gene loci, most notably local multi-
gene families, internal duplication of LRR-encoding sequences, and intralocus recombination.

Genetically linked multigene families are present within several other R gene loci, such as Pto (Martin et al., 1993), N (Whitham et al., 1994), Xa21 (Song et al., 1995, 1997), M (Anderson et al., 1996), and RPP5 (Parker et al., 1997). In addition, several R gene homologs have been mapped to clusters in Arabidopsis and potato (Leister et al., 1996; Botella et al., 1997). The flax M gene is part of a gene family comprising ~15 members, all of which map to a locus of <1 Mb (Anderson et al., 1996). However, not all R genes map to multigene loci. The flax L6 gene, which is a homolog of M, maps to a single gene locus that has multiple alleles (Lawrence et al., 1995; Anderson et al., 1996). RPS2 (Mindrinos et al., 1994) and RPM1 (Grant et al., 1995) map to single loci, and RPM1 has no related homolog at the allelic locus in susceptible lines.

Duplication of LRR-encoding domains has been observed within several R genes. In RPP5, four probable duplication events have been inferred, and an allele of the RPP5 resistance gene has been identified that contains an intragenic duplication of four LRRs (Parker et al., 1997). In L6 and M, there has been at least one large duplication within the LRR-encoding region. The L2 allele seems to have undergone a second duplication event so that it now contains four copies of this large-order repeat unit (Ellis et al., 1997). These observations demonstrate the existence of mechanisms that generate duplications and thus their potential, although definitive proof of their specific contribution is still lacking, to create novel resistance specificities.

The Rp1 locus of maize provides a good example of unequal recombination at a complex locus; lines carrying three Rp1 genes with different specificities can be generated (reviewed in Hulbert, 1997). Also, there is a correlation between unequal recombination, gene conversion, and the generation of new specificities, although the molecular basis for this has not been examined.

The presence of multiple genes within complex loci permits intergenic as well as intragenic unequal exchange events to generate new alleles. After any initial duplication event, one gene can be maintained to perform its normal role while variation accumulates in the other. As with the Hcr9 proteins, the solvent-exposed amino acids in the presumed β sheet of the LRR domain are hypervariable. Recombination and gene conversion probably contribute the variation at both Hcr2 and Hcr9 loci. However, because the Hcr2 proteins contain extremely regular repeating LRR motifs, there is increased scope for intragenic unequal exchange events. As a consequence, there exists dramatic variation in the number and combination of LRRs within the Hcr2 ligand recognition domain, even though the mechanism of generation of variation (recombination) is the same as for Hcr9 genes. It will be interesting to determine the extent of these contrasting modes of Hcr2 and Hcr9 gene evolution within other complex R gene loci.

**METHODS**

**Plant Stocks and Inoculation Procedures**

The near isogenic lines of Lycopersicon esculentum cv Moneymaker carrying either Cf-2 (Cf2) or Cf-5 (Cf5) and the original Moneymaker line carrying no detectable resistance genes to Cladosporium fulvum (Cf0) were obtained from R. Oliver (University of East Anglia, Norwich, UK). Assays for disease resistance were performed as described by Dickinson et al. (1993).

**Construction of Cosmid Library**

Genomic DNA from plants homozygous for Cf-5 and Cf-4 was extracted as described by Carroll et al. (1995), except that high molecular weight DNA was recovered by spooling rather than precipitation. All subsequent techniques were performed according to Sambrook et al. (1989), unless otherwise stated. Further purification of genomic DNA was performed on a cesium chloride-ethidium bromide gradient. This DNA was partially digested with MboI, dephosphorylated, and size fractionated twice on 10 to 40% (w/v) sucrose gradients. Fractionated insert DNA (1 µg) was ligated at 12°C for 18 hr in a total volume of 10 µL with 250 ng of BamHI-digested pCLD04541 binary cosmid vector. Ligated DNA was packaged using commercial extracts (Gigapak, Stratagene) according to the manufacturer’s instructions and transfected into SURE tetracycline-sensitive Escherichia coli (Stratagene).

**Tomato Transformation**

All DNA for transformation into plant cells was cloned into the binary cosmid vector pCLD04541 and mobilized into Agrobacterium tumefaciens LBA4404. Transformation of tomato cotyledons (variety Moneymaker cv) and plant regeneration was performed essentially as described by Fillatti et al. (1987).

**DNA Subcloning**

Cosmid KB2 was generated by subcloning a 13-kb BamHI fragment of cosmid 4/5-42 into the BamHI site of the binary cosmid vector pCLD04541. Sequence analysis and comparison with the published Cf-2 gene showed that the two BamHI sites in cosmid 4/5-42 correspond to a similar site downstream of the polyadenylation site in Cf-2. Transformed plants that were resistant to C. fulvum had previously been generated using Cf-2 constructs, which only contained sequences up to this conserved BamHI site (Dixon et al., 1996).

Sequences containing only the Hcr2-5D region or only the Cf-5 region were subcloned into the pCLD04541 vector as BamHI-EcoRI fragments to create the cosmids minHcr2-5D and minCf-5, respectively. Due to the presence of an internal EcoRI site in each region, these subclonings were achieved using partial digestion with EcoRI to obtain the desired fragments.

**DNA Sequencing**

All templates for sequencing were prepared from clones containing randomly sheared target DNA generated by nebulization. The ends
of sheared fragments were repaired using T4 DNA polymerase (Phar-
macia) and subsequently cloned into dephosphorylated HincII-
digested pUC119 or M13mp18 (Sambrook et al., 1989). Standard
sequencing reactions were performed using the PRISM Ready Reac-
tion Terminator Cycle Sequencing system (Perkin-Elmer). Double-
stranded plasmid DNA for sequencing was isolated according to in-
structions given in the PRISM sequencing protocol handbook (Perkin-
Elmer). Single-stranded DNA was prepared according to standard
procedures. The reaction products were run on Applied Biosystems
(Foster City, CA) 373 and 377 DNA sequencers, and data were as-
sembled using UNIX versions of the Staden programs package
(Roger Staden, MRC, Cambridge, UK), including TED and XBP.
Analysis of assembled sequences was performed using Genetics
Computer Group (Madison, WI) programs. Extended sequencing
reads of selected template DNA preparations were obtained using a
dye primer Thermo sequencing kit (Amersham Pharmacia, Little
Chalfont, UK) in conjunction with fluorescently labeled standard for-
dee. Reads of selected template DNA preparations were obtained using a
Li-Cor DNA sequencer model 4000L, and data were analyzed as previously described.

RNA Extraction and First-Strand cDNA Synthesis

Total RNA was extracted exactly as described by Dixon and Hammonds-
Kosack (1997). One milligram of oligo(dT)25 Dynabeads (Dynal, Oslo,
Norway) was used according to the manufacturer’s instructions to
select poly(A)+ RNA for each sample. Poly(A)+ RNA was recovered
using 10 μL of elution buffer, and half of each sample was used for the
synthesis of first-strand cDNA using Expand reverse tran-
scriptase (Boehringer Mannheim) according to the manufacturer’s instructions. Before polymerase chain reaction (PCR), cDNA was pu-
rified from oligo(dT) and other cDNA synthesis components using QIA-
quick spin columns (Qiagen, Hilden, Germany), according to the
manufacturer’s instructions.

PCR

PCR was performed using 35 cycles of amplification. Primers CF5F2
(5’-GTAATACGACTTCTTCCCTC-3’) and CF5R2 (5’-CTACTTCCAA-
ACTGAGAAGGAA-3’) were used at an annealing temperature of
55°C to amplify the 3′-conserved regions of Cf-5 and Hcr2-5D con-
taining the intron in the untranslated tails. The region spanning the
leucine-rich repeat (LRR) variability between Cf-5 and Hcr2-5D was am-
plified using the primers 2-5CF (5’-CATGTCATTGAGGATCCTAG-3’) and 2-5CR (5’-AGATGACATCGACCGAAAAATGTT-3’) at an
annealing temperature of 58°C. To achieve amplification of specific products from the recombinant plant V454, we used the primers V3F
(5’-CAGTGAGAGAAGACATTCG-3’) and V3R (5’-AAGATATTGCCCA-3’) at an annealing temperature of 49°C.

ACKNOWLEDGMENTS

We thank Sara Perkins, Margaret Shailer, and Jonathan Darby for plant
care and Matthew Smoker for assistance with plant tissue cul-
ture. We are also grateful to Martin Parniske and Colwyn Thomas for
constructive criticisms of this paper and to Patrick Bovill and Dave
Baker for running the ABI373 and ABI377 automated sequencers.

This work was supported by the Gatsby Foundation and also in part
by a European Union Human Capital and Mobility grant (No. 208/
A06586) awarded to K.H.

Received July 13, 1998; accepted August 21, 1998.

REFERENCES

Anderson, P.A., Lawrence, G.J., Morrish, B.C., Ayliffe, M.A.,
resistance gene M associated with loss of a repeated unit within
the leucine-rich repeat coding region. Plant Cell 9, 641–651.


Bent, A.F., Kunkef, B.N., Dahlbeck, D., Brown, K.L., Schmidt, R.,
Arabidopsis thaliana: A leucine-rich repeat class of plant disease

Bisgrove, S.R., Simonich, M.T., Smith, N.M., Sattler, A., and
specificity for two different pathogen avirulence genes. Plant Cell 6,
927–933.

Botella, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T.,
resistance gene

resistance gene M associated with loss of a repeated unit within
the leucine-rich repeat coding region. Plant Cell 9, 641–651.

Bisgrove, S.R., Simonich, M.T., Smith, N.M., Sattler, A., and
specificity for two different pathogen avirulence genes. Plant Cell 6,
927–933.

Botella, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T.,
resistance gene

resistance gene M associated with loss of a repeated unit within
the leucine-rich repeat coding region. Plant Cell 9, 641–651.

Bisgrove, S.R., Simonich, M.T., Smith, N.M., Sattler, A., and
specificity for two different pathogen avirulence genes. Plant Cell 6,
927–933.

Botella, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T.,
resistance gene

resistance gene M associated with loss of a repeated unit within
the leucine-rich repeat coding region. Plant Cell 9, 641–651.

Bisgrove, S.R., Simonich, M.T., Smith, N.M., Sattler, A., and
specificity for two different pathogen avirulence genes. Plant Cell 6,
927–933.

Botella, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T.,
resistance gene

resistance gene M associated with loss of a repeated unit within
the leucine-rich repeat coding region. Plant Cell 9, 641–651.

Bisgrove, S.R., Simonich, M.T., Smith, N.M., Sattler, A., and
specificity for two different pathogen avirulence genes. Plant Cell 6,
927–933.

Botella, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T.,
resistance gene

resistance gene M associated with loss of a repeated unit within
the leucine-rich repeat coding region. Plant Cell 9, 641–651.

Bisgrove, S.R., Simonich, M.T., Smith, N.M., Sattler, A., and
specificity for two different pathogen avirulence genes. Plant Cell 6,
927–933.

Botella, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T.,
resistance gene

resistance gene M associated with loss of a repeated unit within
the leucine-rich repeat coding region. Plant Cell 9, 641–651.

Bisgrove, S.R., Simonich, M.T., Smith, N.M., Sattler, A., and
specificity for two different pathogen avirulence genes. Plant Cell 6,
927–933.

Botella, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T.,
resistance gene

resistance gene M associated with loss of a repeated unit within
the leucine-rich repeat coding region. Plant Cell 9, 641–651.

Bisgrove, S.R., Simonich, M.T., Smith, N.M., Sattler, A., and
specificity for two different pathogen avirulence genes. Plant Cell 6,
927–933.

Botella, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T.,
resistance gene

resistance gene M associated with loss of a repeated unit within
the leucine-rich repeat coding region. Plant Cell 9, 641–651.

Bisgrove, S.R., Simonich, M.T., Smith, N.M., Sattler, A., and
specificity for two different pathogen avirulence genes. Plant Cell 6,
927–933.


