THE ROYAL SOCIETY

Genetic and molecular analysis of tomato *Cf* genes for resistance to *Cladosporium fulvum*

Colwyn M. Thomas, Mark S. Dixon, Martin Parniske, Catherine Golstein and Jonathan D. G. Jones*

Sainsbury Laboratory, John Innes Centre, Norwich Research Park, Colney Lane, Norwich NR4 7UH, UK

In many plant–pathogen interactions resistance to disease is controlled by the interaction of plant-encoded resistance (*R*) genes and pathogen-encoded avirulence (*Avr*) genes. The interaction between tomato and the leaf mould pathogen *Cladosporium fulvum* is an ideal system to study the molecular basis of pathogen perception by plants. A total of four tomato genes for resistance to *C. fulvum* (*Cf-2*, *Cf-4*, *Cf-5* and *Cf-9*) have been isolated from two genetically complex chromosomal loci. Their gene products recognize specific *C. fulvum*-encoded avirulence gene products (Avr2, Avr4, Avr5 and Avr9) by an unknown molecular mechanism. *Cf* genes encode extracellular membrane-anchored glycoproteins comprised predominantly of 24 amino acid leucine-rich repeats (LRRs). *Cf* genes from the same locus encode proteins which are more than 90% identical. Most of the amino-acid sequence differences correspond to the solvent-exposed residues within a β-strand/β-turn structural motif which is highly conserved in LRR proteins. Sequence variability within this motif is predicted to affect the specificity of ligand binding. Our analysis of *Cf* gene loci at the molecular level has shown they comprise tandemly duplicated homologous genes, and suggests a molecular mechanism for the generation of sequence diversity at these loci. Our analysis provides further insight into the molecular basis of pathogen perception by plants and the organization and evolution of *R* gene loci.

Keywords: tomato; disease resistance; *Cf* genes

1. INTRODUCTION

Plants are constantly subjected to attack by potentially pathogenic micro-organisms. In many plants, resistance to disease is controlled by the interaction of dominant plant-encoded resistance (*R*) genes and the products of pathogen-encoded avirulence (*Avr*) genes (the 'gene-for-gene' interaction). In many cases resistance is manifested in the form of a hypersensitive response, which results in plant cell death at the site of pathogen attack and arrest of pathogen ingress (Hammond-Kosack & Jones 1996). It has been postulated that *R* gene products act as receptors for Avr proteins which initiate a signal cascade resulting in activation of the plant defence response (reviewed in Staskawicz *et al*. 1995).

Most isolated *R* genes appear to encode cytoplasmic proteins which contain a central nucleotide binding site (NBS) domain and a C-terminal domain consisting of variable numbers of leucine-rich repeats (LRRs) (Bent 1996; Jones & Jones 1997). This class is further subdivided into genes encoding either an N-terminal leucine zipper (LZ) region (LZ–NBS–LRR), or an N-terminal region showing homology to the cytoplasmic signalling domain of the *Drosophila* Toll and mammalian interleukin-1 receptors (TIR), the TIR–NBS–LRR class (Bent 1996; Jones & Jones 1997). The homology between these latter proteins and resistance proteins led to the suggestion that they may function in a conserved pathway in eukaryotes, which is activated in response to pathogen challenge (Belvin & Anderson 1996). Recent studies have identified homologies between these NBS–LRR proteins and proteins regulating apoptosis in mammals and worms (Van der Biezen & Jones 1998), suggesting R proteins may function as components of cellular complexes which affect the plant defence response in an Avr-dependent manner.

The first *R* gene to be isolated of the gene-for-gene type was the tomato *Pto* gene, which encodes a cytoplasmic serine–threonine protein kinase and confers resistance to races of *Pseudomonas syringae* that express *AvrPto* (Martin *et al*. 1993). Pto interacts physically with AvrPto in the yeast two-hybrid system, consistent with the proposed receptor–ligand mechanism of pathogen perception *in planta* (Scofield *et al*. 1996; Tang *et al*. 1996). Yeast two-hybrid analysis demonstrated Pto interacts with a second serine–threonine protein kinase, Ptil (Zhou *et al*. 1995), and a number of transcription factors which may activate defence genes (Zhou *et al*. 1997). Mutational analysis has also identified the *Prf* gene encoding a distinct cytoplasmic NBS–LRR protein which is essential for Pto function (Salmeron *et al*. 1996). Therefore, the function of LRRs in Prf and other resistance proteins of the NBS–LRR type in the perception of Avr proteins remains unclear.

Another class of *R* genes includes *Cf-2*, *Cf-4*, *Cf-5* and *Cf-9* from tomato (Dixon *et al*. 1996, 1998; Jones *et al*.

*Author for correspondence (jonathan.jones@bbsrc.ac.uk).

1994; Thomas *et al*. 1997). These genes encode extracytoplasmic membrane-anchored glycoproteins consisting predominantly of LRRs. The rice *Xa-21* gene for resistance to *Xanthomonas oryzae* pv. *oryzae* encodes a similar type of protein but is distinguished from Cf proteins by an additional cytoplasmic serine–threonine protein kinase domain (Song *et al*. 1995). Genetic analysis of *Cf* gene loci (Balint-Kurti *et al*. 1994; Dickinson *et al*. 1993; Jones *et al*. 1993) and other *R* gene loci, including *L* and *M* in flax (reviewed in Ellis *et al*. 1997), *Rp1* in maize (reviewed in Hulbert 1997), *Dm* in lettuce (Witsenboer *et al*. 1995), *Mla* in barley (Jorgensen 1994), and various *RPP* loci in *Arabidopsis* (Holub 1997), have shown that these loci are genetically complex. A principal goal in plant pathology is to determine the molecular basis for pathogen perception by R proteins and the underlying molecular mechanisms which generate novel recognitional specificities.

Genes which encode R proteins of the NBS–LRR type have been isolated from a number of species and differ significantly in their primary sequence, overall length, and the numbers of their LRRs (Bent 1996; Jones & Jones 1997). Consequently, no meaningful conclusions could be made with respect to the molecular basis for pathogen perception. To specifically address this question we have recently isolated and characterized four *Cf* genes for resistance to *C. fulvum* from two complex chromosomal loci (Dixon *et al*. 1996, 1998; Jones *et al*. 1994; Thomas *et al*. 1997). In this paper we review our current knowledge of tomato *Cf* genes, their genetic organization and the possible molecular mechanisms for generating sequence diversity.

(a) *The interaction of* Cladosporium fulvum *with tomato*

Cladosporium fulvum is a biotrophic fungus which causes leaf mould in tomato, its only known host. In compatible interactions (i.e. disease-sensitive), fungal spores germinate on the abaxial surface of leaves and enter the leaf apoplast through stomata. Fungal hyphae grow intercellularly in the lower mesophyll cells and do not form haustoria. Conidiophores emerge through the stomata *ca*. 10–12 days post-infection (Hammond-Kosack & Jones 1994). In incompatible tomato–*C. fulvum* interactions fungal hyphae are arrested in their development soon after penetration of the sub-stomatal cavity, dependent on the particular *Cf* genotype (Hammond-Kosack & Jones 1994). With the possible exception of Cf-2, which can trigger guard cell death, cell death is not a prerequisite for arresting fungal growth.

C. fulvum physiological races differ in the spectrum of *Avr* genes which they express. Fungal races are defined according to their reaction on tomato lines containing specific *Cf* genes and named according to the resistance genes they can overcome: e.g. *C. fulvum* race 5 will cause disease on tomato plants expressing no *Cf* genes (Cf0), or on genotypes expressing only the *Cf-5* gene. Avirulence gene products secreted by the fungus can be isolated from the infected leaves of compatible interactions in apoplastic washing fluids (AF, De Wit & Spikman 1982). When injected into tomato leaves or cotyledons containing *Cf* genes, AF induces a range of responses varying from necrosis in *Cf-9* containing plants to mild chlorosis in cotyledons of *Cf-5*-containing plants (Hammond-Kosack & Jones 1994).

(b) Cladosporium fulvum*-encoded avirulence determinants*

Two genes for the *C. fulvum* encoded avirulence gene products Avr9 and Avr4 have been cloned (Joosten *et al*. 1994; Van den Ackerveken *et al*. 1992), which are recognized by *Cf-9* and *Cf-4*. Both *Avr* genes can be expressed in transgenic Cf0 tomatoes and secreted into the apoplastic space by fusing to an appropriate signal peptide sequence (Hammond-Kosack *et al*. 1994*a*; Thomas *et al*. 1997). When transgenic plants expressing either Avr9 or Avr4 are crossed to tomato lines containing the cognate *Cf* gene, the F_1 progeny undergo developmentally regulated seedling death. Avr9 and Avr4 can also be expressed in the form of recombinant potato virus × (PVX::Avr4 and PVX::Avr9). Inoculated tomato leaves containing the appropriate *Cf* gene exhibit systemic necrosis as a result of virus infection and subsequent activation of a plant cell death response (Hammond-Kosack *et al*. 1995; Thomas *et al*. 1997).

(c) *Tomato* Cf *genes*

In the breeding of cultivated tomato (*Lycopersicon esculentum*) a number of genes which confer resistance to *C. fulvum* have been introgressed from wild tomato species (Jones *et al*. 1993; Stevens & Rick 1988). The *Cf-9* (Stevens & Rick 1988) and *Cf-2* (Langford 1937) genes were introgressed from *L. pimpinellifolium*, *Cf-4* from *L. hirsutum* (Kerr & Bailey 1964) and *Cf-5* from *L. esculentum* var. *cerasiforme* (Dickinson *et al*. 1993). All four genes have been introgressed into the tomato cultivar Moneymaker (Cf0). Successive backcrossing was used to generate a series of near isogenic lines (Tigchelar 1984) containing single introgressed *Cf* genes (Cf2, Cf4, Cf5 and Cf9).

A combination of classical and restriction fragment length polymorphism (RFLP) mapping studies showed the *Cf* genes were located at two complex loci. *Cf-9* and *Cf-4* were mapped to identical locations on the short arm of chromosome 1 (Balint-Kurti *et al*. 1994; Jones *et al*. 1993) in a 5 cM interval between the RFLP markers TG236 (3 cM proximal to *Cf-4/Cf-9*) and CP46 (2 cM distal to *Cf-4/Cf-9*), in F_2 progeny of interspecific crosses between the Cf4 or Cf9 NILs and *L. pennellii* (Balint-Kurti *et al*. 1994). *Cf-2* and *Cf-5* were mapped at identical locations on the short arm of chromosome 6 (Dickinson *et al*. 1993; Jones *et al*. 1993) in a 4–5 cM interval between the RFLP markers CT119 (located proximal to *Cf-2/Cf-5*), and GP79 (located distal to *Cf-2/Cf-5*) in F_2 progeny of interspecific crosses between the Cf2 or Cf5 NILs with *L. pennellii* (Dickinson *et al*. 1993). Mapping relative to visible genetic markers located *Cf-2* and *Cf-5* between the morphological markers *yellow virescent* (*yv*) (2 cM proximal to *Cf-2/Cf-5*) and *thiamineless* (*tl*) (1 cM distal to C*f-2/Cf-5*) (Jones *et al*. 1993).

(d) *Isolation of* Cf-9 *and* Cf-2

Cf-9 was isolated by transposon tagging using an engineered version of the non-autonomous maize transposon *Dissociation* (*Ds*). A line containing *Cf-9* and a genetically linked T-DNA carrying *Ds* (located 3 cM proximal to *Cf-9*), was crossed to *Cf-9* plants containing an *Activator* (*Ac*) transposase source. Transposon tagged alleles of *Cf-9*

were identified as survivors after test-crossing to a Cf0 transgenic line expressing Avr9 (Jones *et al*. 1994).

Molecular analysis of recombinants in the *tl/yv* interval using cDNA and genomic probes was used to generate a high-resolution molecular map of the *Cf-2* locus (Dixon *et al*. 1995). *Cf-2* was shown to cosegregate with a cDNA probe MG112. This clone was isolated from a root-specific library, using as probes two YAC clones encompassing *Cf-2* and the closely linked *Mi* locus for resistance to root-knot nematodes (*Meloidogyne* spp.) (Dickinson *et al*. 1993). MG112 identified polymorphic sequences between the NILs Cf0, Cf2 and Cf5 (Dixon *et al*. 1995). There are two near-identical genes, *Cf-2.1* and *Cf-2.2*, which conferred Avr2-dependent resistance. These were identified by complementation experiments using binary vector cosmid clones from a tomato line containing *Cf-2* and *Cf-9* (Dixon *et al*. 1996).

(e) Cf-9 *and* Cf-2 *encode extracellular membrane-anchored LRR glycoproteins*

Cf-9 and the Cf-2 proteins encode membrane-anchored extracellular leucine-rich repeat glycoproteins (Dixon *et al*. 1996; Jones *et al*. 1994). This appears consistent with their proposed roles as extracellular receptors for pathogen-encoded signal molecules secreted into the leaf apoplast. *Cf-9* encodes a protein of 863 amino acids which contains several structural domains (Jones *et al*. 1994; Jones & Jones 1997). Most of the sequence comprises *ca*. 24 amino acid leucine-rich repeat units (27 in total) which show extensive homology to the consensus sequence for extracellular LRR proteins (LxxLxxLxxLxLxxNxLxGxIPxx, where x is any amino acid) (Kobe & Deisenhofer 1994). In Cf-9 the short cytoplasmic domain (domain G) is terminated by the amino-acid sequence KKRY. In mammalian and yeast proteins, the KKxx sequence motif at the C terminus of secreted proteins acts as a signal sequence for retrieval from the Golgi apparatus to the endoplasmic reticulum. It is not known if this sequence performs a similar role in plants. Cf-9 also contains a number of sequences which are potential targets for N-glycosylation (Jones *et al*. 1994).

The *Cf-2.1* and *Cf-2.2* genes encode proteins of 1112 amino acids which differ by only three amino acids near their C-termini (Dixon *et al*. 1996). Neither Cf-2 proteins contain the KKxx motif at their C-termini. The Cf-2 proteins are considerably larger than Cf-9 and contain 38 LRRs (Dixon *et al*. 1996) which, in contrast to Cf-9, are of a uniform 24 amino-acid length. Additionally, in domain Cl a sequence of 20 LRRs comprises, predominantly, two distinct alternating LRRs, type A with the consensus EEIGL(R–S)SLTxLxLGxNxL(N–S)GSIP, and type B with the consensus ASLGNLNNL(S–F)xLxLYNN(Q–K)LSGSIP (Dixon *et al*. 1996). Cf-2 proteins also contain a number of potential targets for N-glycosylation.

With the exception of LRR consensus residues, the N-terminal portions of the Cf-9 and Cf-2 proteins display little amino-acid sequence homology. Significant levels of amino-acid homology are observed in their C-termini in a region encompassing the last nine LRRs and domains D, E, F and G (Dixon *et al*. 1996; Jones & Jones 1997). This conserved region may perform a similar role in these functionally related proteins by interacting with a common or related signalling partner which activates the plant defence response upon ligand binding.

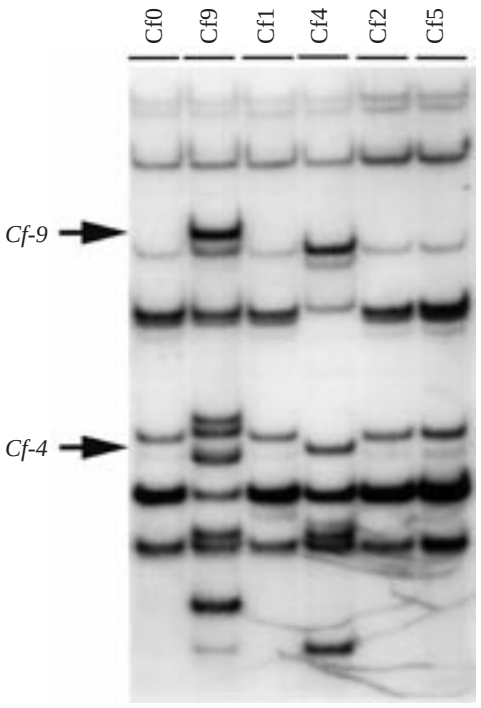

Figure 1. DNA from a line containing no known *Cf* genes (Cf0), NILs containing single introgressed *Cf* genes (Cf2, Cf4, Cf5 and Cf9), and *L. esculentum* cv Stirling Castle (Cf1) were digested with *Bgl*II and blotted. The filter was hybridized with a *Cf-9* probe (Jones *et al*. 1994). The 6.7 and 3.2 kb *Bgl*II fragments derived from *Cf-9* and *Cf-4* are indicated.

(f) Cf-9 *and* Cf-2 *are members of genetically linked multigene families*

Gel-blot analysis of DNA from several NILs using a probe to the 5′ end of *Cf-9* (Jones *et al*. 1994) revealed a polymorphic multigene family (see figure 1). This analysis also included *L. esculentum* cv Stirling Castle which contains *Cf-1* that also maps at the *Cf-4/Cf-9* locus (Kerr & Bailey 1964). A total of 10–15 hybridizing bands were detected. Several of the members of this family are polymorphic between Cf4 and Cf9 NILs (figure 1), demonstrating that some of these members are present on the introgressed DNA (Balint-Kurti *et al*. 1994) and are therefore genetically linked. Mapping studies have shown the non-polymorphic members of the multigene family detected with the *Cf-9* probe are located outside the region of introgressed DNA at several loci on the short arm of chromosome 1 (M. Parniske, unpublished data). The additional members of the *Cf-9*-homologous multigene family (*Hcr9*s) in the NILs Cf9 and Cf4 are designated *Hcr9-9* and *Hcr9-4* genes, respectively (for homologue of *Cladosporium* resistance gene *Cf-9*).

None of the members of the *Cf-2* or *Cf-5* multigene families were detected in this analysis (Thomas *et al*. 1997). Molecular analysis of the Cf0, Cf2 and Cf5 NILs using *Cf-2* as a probe has also revealed the presence of a genetically linked polymorphic multigene family at the *Cf-2/Cf-5* locus (Dixon *et al*. 1995, 1996). The additional members of the *Cf-2*-homologous multigene family in the NILs Cf2 and Cf5 have been designated *Hcr2-2* and

Hcr2-5 genes, respectively (for homologue of *Cladosporium* resistance gene *Cf-2*).

(g) *A genetic strategy to isolate* Cf-4 *and* Cf-5

The observation that *Cf-4* maps at an identical chromosomal location to *Cf-9*, and *Cf-5* at an identical location to *Cf-2* suggested *Cf-4* and *Cf-5* may be homologues of *Cf-9* and *Cf-2*, respectively. Because members of these multigene families are polymorphic, a genetic strategy was employed to identify candidate genes for *Cf-4* and *Cf-5* by generating recombinants between *Cf-4/Cf-9* and *Cf-2/Cf-5* which could be analysed for their *Hcr9* or *Hcr2* composition (Dixon *et al*. 1996; Thomas *et al*. 1997).

Two transheterozygous parents were generated (*Cf-4/Cf-9* and *Cf-2/Cf-5*) and test-crossed to Cf0 plants to identify recombinants. In the case of allelic genes, disease-sensitive recombinants could arise through rare intragenic recombination events. In the case of closely linked genes, recombinants possessing neither *R* gene (disease-sensitive), or both *R* genes in *cis* configuration, would occur through intergenic recombination. Plants containing both parental resistance specificities in *cis* arrangement, and plants lacking both genes could also arise through chromosomal mispairing and unequal crossing-over. Alternatively, disease-sensitive plants could arise through chromosomal deletions. Test-cross progeny of the *Cf-4/Cf-9* transheterozygous parent were inoculated with *C. fulvum* race 5 (which expresses *Avr4* and *Avr9*), and progeny of the *Cf-2/Cf-5* transheterozygous parent with *C. fulvum* race 4 (which expresses *Avr2* and *Avr5*) to identify disease-sensitive recombinants (Dixon *et al*. 1996, 1998; Thomas *et al*. 1997).

A total of five disease-sensitive individuals (V408, V512, V514, V516 and V517) were identified after screening 7400 *Cf-4/Cf-9* test-cross plants. RFLP analysis showed that all five plants contained the Cf0 and Cf9 alleles of CP46, suggesting *Cf-4* is located distal to *Cf-9*. In progeny of the *Cf-2/Cf-5* cross one disease-sensitive plant (V454) was identified in 12 000 test-cross progeny (Dixon *et al*. 1996). The five *Cf-4/Cf-9* disease-sensitive plants were screened by DNA gel-blot and AFLP analysis (Thomas *et al*. 1997). All disease-sensitive plants lacked restriction fragments from the 5′ and 3′ ends of *Cf-9* and several restriction fragments specific to the Cf4 parent. A total of three *Hcr9-4*s were identified as *Cf-4* candidates (Thomas *et al*. 1997). Analysis of the single *Cf-2/Cf-5* recombinant identified several *Hcr2-5* genes as candidates for *Cf-5* (Dixon *et al*. 1998). Binary vector cosmid clones containing the candidate genes were isolated and physical contigs of the *Cf-4/Cf-9* and *Cf-2/Cf-5* loci were constructed (Dixon *et al*. 1996, 1998; Parniske *et al*. 1997; Thomas *et al*. 1995, 1997).

(h) *Physical organization of the* Cf-4/Cf-9 *and* Cf-2/Cf-5 *loci*

A physical map of the *Cf-4* locus was compared with a cosmid contig of the *Cf-9* locus (Parniske *et al*. 1997; Thomas *et al*. 1997), and the corresponding locus from the disease-sensitive Cf0 NIL (Parniske *et al*. 1997). From DNA gel-blot analysis, using a *Cf-9* probe, it was shown that the *Cf-4* and *Cf-9* haplotypes each contain five *Hcr9* genes (*Hcr9-9A* to *E*, and *Hcr9-4A to E*) within a 36 kb physical interval (see figure 2). The tandemly duplicated units are flanked by convergently orientated lipoxygenase genes (*LoxL* and *LoxR*) (Parniske *et al*. 1997). Physical mapping and DNA sequence analysis of each contig has shown that the *Hcr9*s are tandemly duplicated transcription units (Parniske *et al*. 1997; Thomas *et al*. 1997). In contrast, the disease-sensitive Cf0 haplotype contains only a single *Hcr9* gene (*Hcr9-0A*, figure 2).

Physical contigs encompassing *Cf-2* (Dixon *et al*. 1996), and *Cf-5* (Dixon *et al*. 1998), and the corresponding locus in the Cf0 NIL (Dixon *et al*. 1995) have also been compared. The Cf2 line contains tandemly duplicated copies of *Cf-2* (*Cf-2.1* and *Cf-2.2*), and a third gene (*Hcr2-2A*) which encodes a protein of the same class as *Cf-2.1* and *Cf-2.2* (figure 2). The three genes are located within a 30-kb region (Dixon *et al*. 1996). Molecular analysis of cosmids encompassing *Cf-5* identified four tandemly orientated *Hcr2-5* genes (*Hcr2-5A* to *Hcr2-5D*) (Dixon *et al*. 1998).

(i) *Chromosomal mispairing and unequal crossing-over at the* Cf-4/Cf-9 *and* Cf-2/Cf-5 *loci*

There were two distinct classes of *Cf-4/Cf-9* recombinants which could be identified based on their *Hcr9* composition (Thomas *et al*. 1997). A total of four disease-sensitive plants (V408, V512, V516, and V517) contained *Hcr9-4A* and *Hcr9-4B* from the Cf4 parental line and *Hcr9-9E* from the Cf9 parent (figure 3) (Thomas *et al*. 1997) confirming these plants were generated by recombination at the *Cf-4/Cf-9* locus. Recombinant (V514) retained *Hcr9-9D* and *Hcr9-9E* from the Cf9 parental line and *Hcr9-4A* from the Cf4 parent (figure 3). From molecular analysis the approximate locations of recombination break-points in the five recombinants could be determined (figure 3). All recombinants were generated through chromosomal mispairing and unequal crossing-over, i.e. each recombinant chromosome retained three *Hcr9* genes at this locus compared with five in the parental lines. All five cross-overs were within the intergenic regions. The precise location of each recombination event was determined by DNA sequence analysis which confirmed their chimeric nature (Parniske *et al*. 1997). From this analysis it was shown that *Cf-4* was located distal to AFLP marker M2 (Thomas *et al*. 1997) and identified *Hcr9-4C*, *Hcr9-4D*, and *Hcr9-4E* as candidates for *Cf-4* (figure 2).

Analysis of the *Hcr2* composition of the *Cf-2/Cf-5* recombinant V454 using a *Cf-2* probe (Dixon *et al*. 1996, 1998) showed it lacked both *Cf-2.1* and *Cf-2.2* as predicted. Restriction fragments corresponding to two *Hcr2-5* genes (*Hcr2-5C* and *Hcr2-5D*) were also absent (figure 2) and some non-parental restriction fragments were also observed. DNA sequence analysis showed this was a consequence of intragenic recombination and located the region of recombination to a 107-bp interval encoding part of domains A and B in the *Cf-2.2* and *Hcr2-5B* genes (Dixon *et al*. 1998). Therefore, V454 was also generated through chromosomal mispairing and unequal crossing-over (figure 3) and this analysis identified *Hcr2-5B*, *Hcr2-5C* and *Hcr2-5D* as candidates for *Cf-5* (Dixon *et al*. 1998).

(j) *Identification and DNA sequence analysis of* Cf-4 *and* Cf-5

The identity of *Cf-4* (*Hcr9-4D*) was determined by transformation of overlapping cosmids containing various combinations of the three candidate *Cf-4* genes (Thomas *et al*. 1997). Transgenic Cf0 plants expressing *Hcr9-4D*

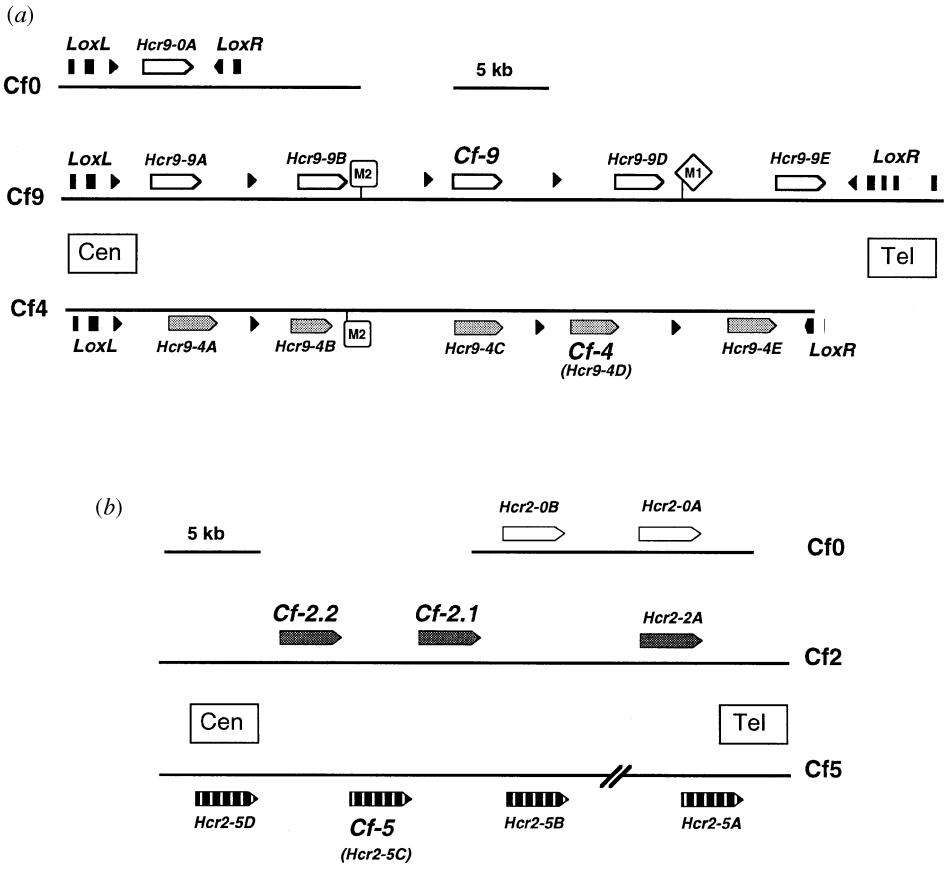

Figure 2. Physical organization of *Hcr9* and *Hcr2* genes in different *Lycopersicon* haplotypes. The transcriptional polarity of *Hcr9* and *Hcr2* genes is indicated by arrows. (*a*) Organization of *Hcr9* genes in Cf0, Cf9 and Cf4 NILs. *Hcr9*s in each genotype are flanked by lipoxygenase genes (*LoxL* and *LoxR*). Sequences derived from the 3′ exon of *LoxL* present in the *Hcr9* intergenic regions are also shown. M1 and M2 correspond to two AFLP markers (Thomas *et al*. 1995). The orientation of the *Cf-4/Cf-9* locus relative to the centromere (Cen) and telomere (Tel) was deduced as descibed previously (Thomas *et al*. 1997). (*b*) Organization of *Hcr2*s in Cf0, Cf2 and Cf5 NILs. The physical contig encompassing *Cf-5*, *Hcr2-5D* and *Hcr2-5B* does not include *Hcr2-5A* which has been detected by gel-blot analysis (Dixon *et al*. 1998), therefore a discontinuity is shown in the physical map. The chromosomal orientation of the *Cf-2/Cf-5* locus was deduced from RFLP analysis of recombinant V454, using CT119 and GP79 as probes. The orientations of these RFLP markers on the short arm of chromosome 6 have been determined previously (Van Wordragen *et al*. 1996).

under its own promoter, or CaMV 35S, were resistant to infection by *C. fulvum* race 5 (which expresses the Avr4 avirulence determinant) but not race 4 isolates which can overcome *Cf-4*-mediated resistance (Joosten *et al*. 1994). As predicted, transgenic plants expressing *Cf-4* also exhibited systemic necrosis after inoculation with PVX::Avr4 (Thomas *et al*. 1997).

Cf-4 encodes a polypeptide of 806 amino acids with 25 LRRs (figure 4) 91.5% identical with Cf-9 (863 amino acids). The reduction in length compared with Cf-9 is owing to a ten amino-acid deletion near the mature N-terminus in domain B (Jones *et al*. 1994) and a second deletion of 46 amino acids comprising two complete LRRs (figure 4). A total of 67 amino acids distinguish Cf-4 from Cf-9, of which six are located within the signal peptide sequence (Thomas *et al*. 1997).

Cf-5 was also identified by complementation analysis in Cf0 transgenic plants (Dixon *et al*. 1998). Transgenic Cf0 plants expressing *Cf-5* conferred resistance only on *C. fulvum* races expressing the Avr5 avirulence determinant. Cf-5 consists of 968 amino acids and contains a total of 32 LRRs, six fewer than in Cf-2 proteins. The overall level of sequence identity between Cf-5 and Cf-2 is 90% (Dixon *et al*. 1988); slightly lower than that observed between Cf-4 and Cf-9. In Cf-2 proteins, 20 of the 34 LRRs in domain C1 are of two distinct types; type A or type B. With the exception of a block of four consecutive type B LRRs in the middle of this domain, the repeats exhibit a strict alternating pattern (Dixon *et al*. 1996). In Cf-5, 14 consecutive LRRs in domain C1 exhibit a strictly alternating pattern of type A and type B repeat units (Dixon *et al*. 1998).

In both Cf-4 and Cf-5 the variable amino acids are not distributed at random. Amino acids which distinguish Cf-4 from Cf-9 are located in the N-terminal half of the protein delimiting a region which must determine the recognitional specificity of ligand binding in Cf-4 and Cf-9 (figure 4). A number of amino acids distinguishing Cf-5 from Cf-2 are found in the C-terminal half of the protein, but most are again located in the N-terminal half (figure 4). Part of the LRR consensus sequence (xxLxLxx, where L=leucine or any aliphatic amino acid and x is any amino acid) corresponds to a region which is highly conserved in members of the LRR superfamily of proteins (figure 4). It has been proposed that this sequence forms a β-strand/β-turn structural motif. The side chains of consensus aliphatic amino acids perform a structural role and project into the hydrophobic core of the molecule. The side chains of interstitial residues project into solvent, where they can interact with the cognate ligand (Kobe & Deisenhofer 1994). Sequence variability

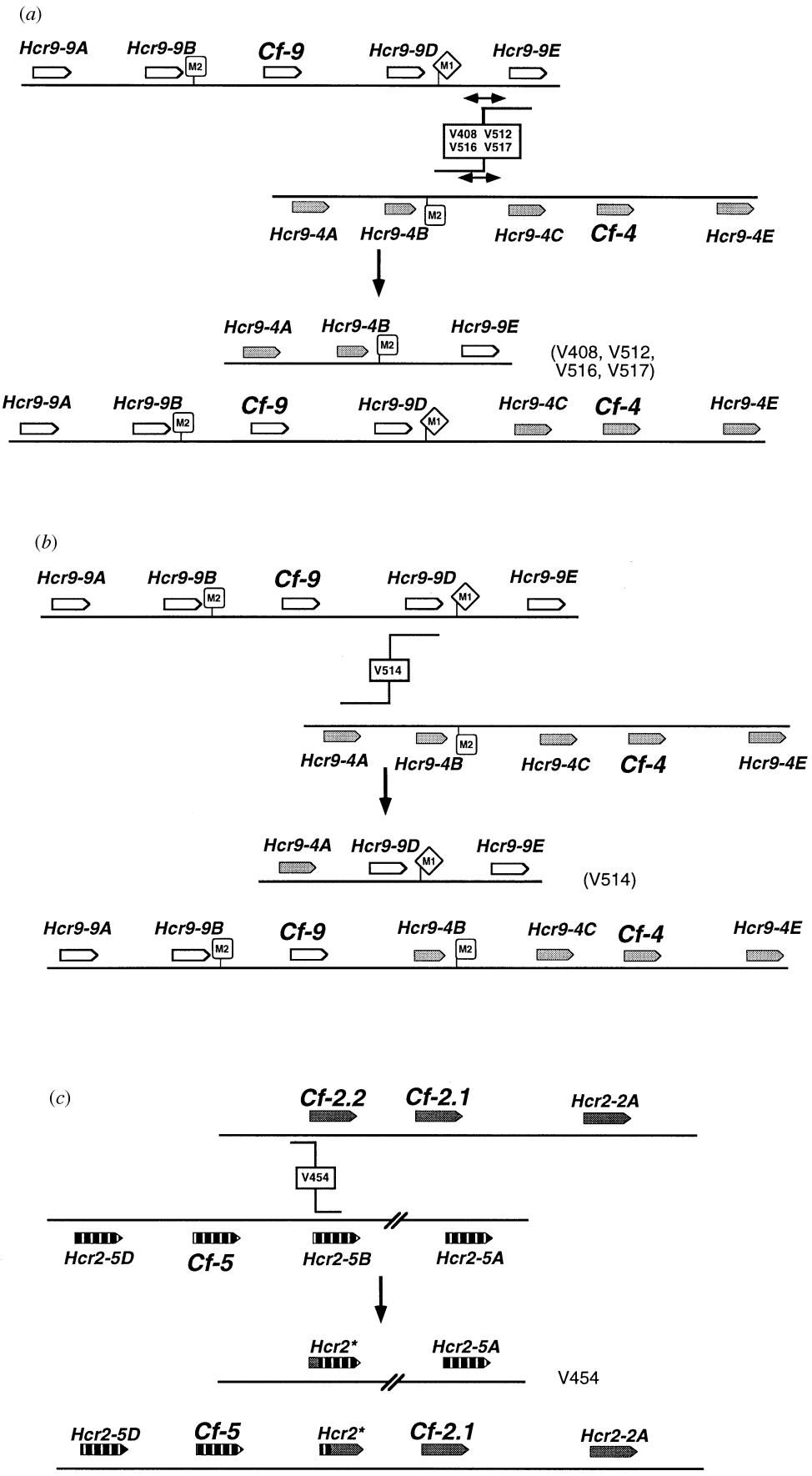

Figure 3. Chromosomal mispairing and unequal crossing-over at the *Cf-4/Cf-9* and *Cf-2/Cf-5* loci. Misalignment of Cf4 and Cf9 chromosomes generated two distinct disease-sensitive recombinant classes. (*a*) In four recombinants (V408, V512, V516 and V517) crossing-over occurred within a 3.0-kb interval upstream of *Hcr9-9E* in the Cf9 chromosome, and *Hcr9-4C* in the Cf4 chromosome. The *Hcr9* composition of disease-sensitive plants V408, V512, V516 and V517 is shown. The *Hcr9* composition of the reciprocal 'double resistant' recombinant chromosome is shown below to illustrate the subsequent variation in *Hcr9* copy number. (*b*) The recombination break-points in V514 were mapped to a different location. Recombinant V514 retains three *Hcr9*s distinct from the class described above. The *Hcr9* composition of the reciprocal 'double resistant' recombinant chromosome is shown below. (*c*) Chromosomal mispairing at the *Cf-2/Cf-5* locus as deduced from molecular analysis of recombinant V454. Crossing-over occurred within the coding sequences of *Cf-2.2* and *Hcr2-5B*. The *Hcr2* composition of recombinant V454 and the reciprocal 'double resistant' recombinant chromosome are shown. In both recombinant chromosomes novel *Hcr2* genes (*Hcr2**) were generated.

within the interstitial residues is therefore expected to affect the recognitional specificity of ligand binding (Kobe & Deisenhofer 1994). In Cf-4, 33 of the 57 amino acids in LRRs which distinguish it from Cf-9 are located within this motif, and 32 correspond to interstitial residues (figure 4). In Cf-5, 56 of the 88 amino acids within LRRs which distinguish it from Cf-2 correspond to interstitial residues (figure 4).

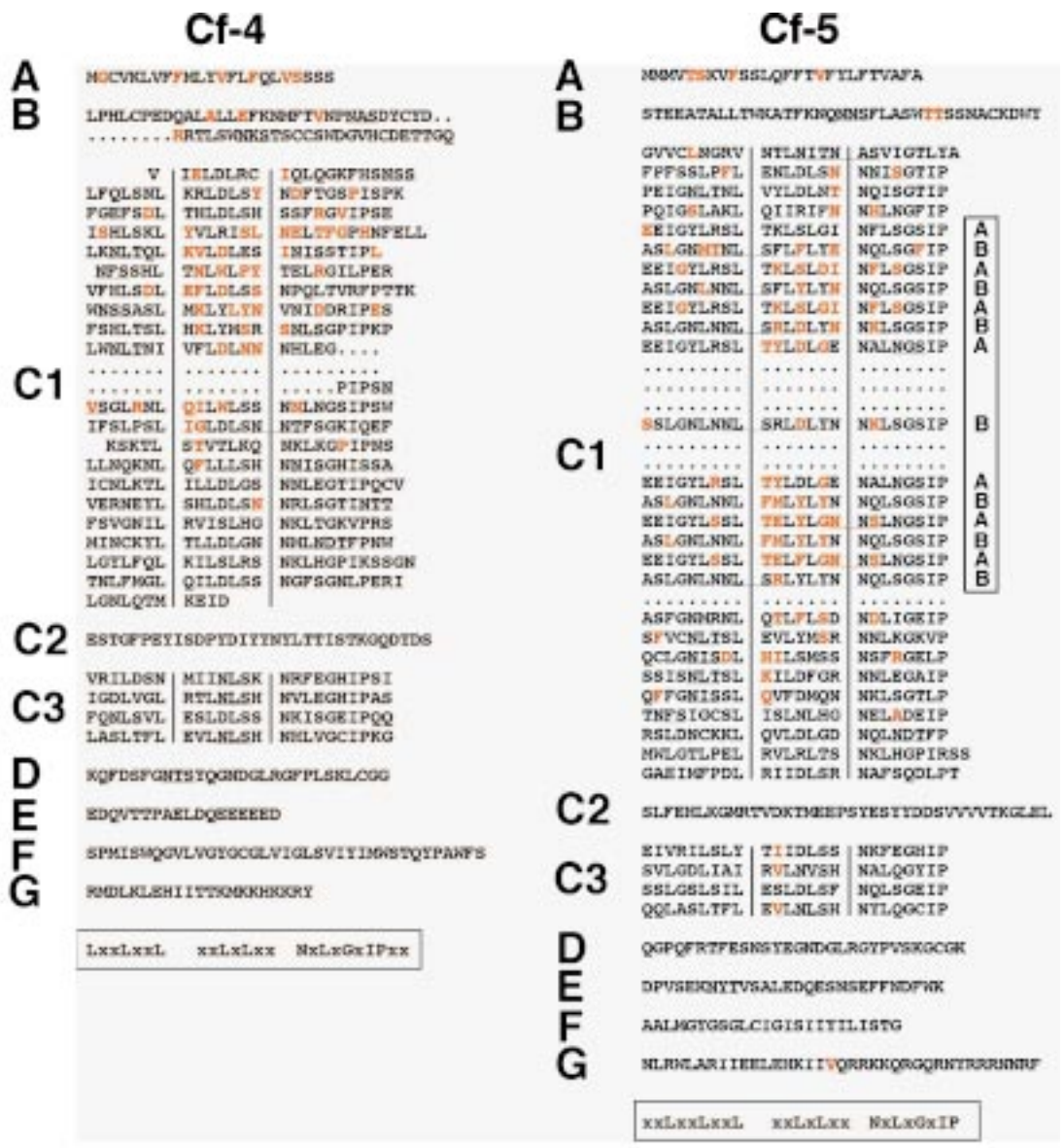

Figure 4. Amino-acid sequences of Cf-4 and Cf-5. Both sequences are divided into domains to emphasize structural similarities with Cf-9 and Cf-2. Domain A is a predicted signal peptide sequence, domain B the predicted mature N-terminus shows homology to the N-terminal sequences in a number of plant LRR receptor-like protein kinases, and to plant-encoded inhibitors of fungal polygalacturonases (Jones & Jones 1997). Domains C1 and C3 contain *ca*. 24 amino acid leucine-rich repeats which show homology to the consensus sequence, for extracellular LRR proteins shown boxed and aligned below each sequence. Domain C2 bisects domains C1 and C3 and shows poor homology to the LRR consensus sequence, as does domain D. Domains E, F and G anchor and orientate the protein within a membrane. Domain E contains a high proportion of negatively charged acidic residues. Domain F is rich in aliphatic amino acids and is predicted to form a membrane-spanning α-helix. Domain G comprises a short cytoplasmic domain rich in positively charged amino acids. Residues corresponding to the predicted β-strand/β-turn LRR structural motif are delimited by vertical lines in both sequences. Identical amino acids between Cf-4 and Cf-9, and between Cf-2 and Cf-5, are shown in black, and variant residues in red. Deletions in Cf-4 and Cf-5 relative to Cf-9 and Cf-2 are indicated by dots. The 14 alternating type A and type B LRRs in Cf-5 are also indicated. Potential NxS–T glycosylation sequences are underlined.

2. DISCUSSION

(a) Hcr9 *and* Hcr2 *gene organization*

Cf genes at the *Cf-4/Cf-9* and *Cf-2/Cf-5* loci are members of tandemly duplicated multigene families. The *Cf-4* and *Cf-9* haplotypes each comprise five *Hcr9* genes within a 36-kb interval (figure 2) while the corresponding locus in the Cf0 NIL has only one *Hcr9*. A total of three *Hcr2* genes are present in the *Cf-2* haplotype, whereas the *Cf-5* haplotype contains four—two *Hcr2*s are present at the corresponding locus in the Cf0 NIL (Dixon *et al*. 1996). Molecular analyses of other disease-resistance loci, such as *RPP5* in *Arabidopsis* (Parker *et al*. 1997), *N* in tobacco (Whitham *et al*. 1994), *Xa-21* in rice (Song *et al*. 1997), *M* in flax (Anderson *et al*. 1997), and the *Pto* and *I2* loci in tomato (Martin *et al*. 1993; Ori *et al*. 1997) have also revealed the presence of genetically linked multigene families. This type of organization was also predicted from genetic analyses of the *rp1 complex* in maize (Hulbert 1997), *M* in flax (Ellis *et al*. 1997) and

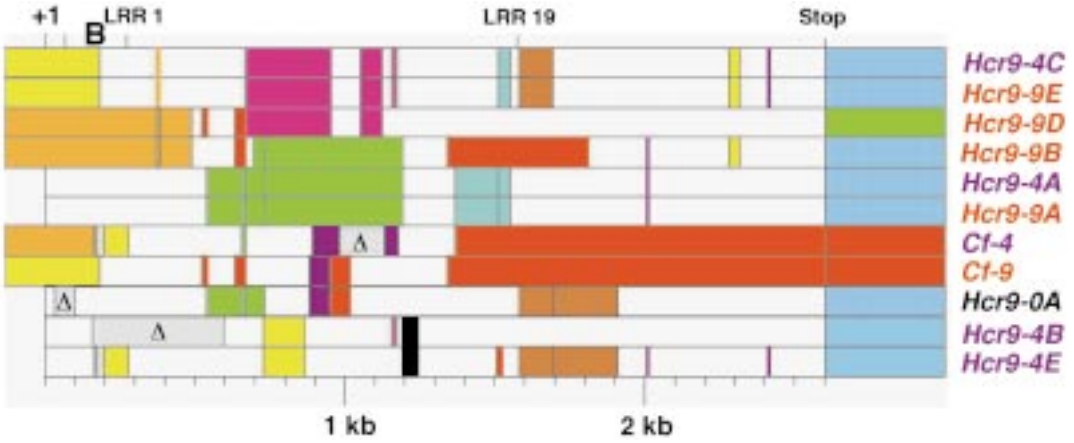

Figure 5. Schematic alignment of 11 *Hcr9* sequences highlighting the DNA sequence affiliations in their coding regions, upstream of the translation initiation site (+1), and downstream of the translation termination signal (Stop). The location of domain B (B), and the interval between LRR1 and LRR19 are also indicated. The identity of individual *Hcr9*s is indicated at the right of the panel, *Hcr9-9*s in red, *Hcr9-4*s in blue and *Hcr9-0A* in black. Homologous DNA sequence tracts shared between *Hcr9*s are shown in the same colour. Deletions in *Hcr9*s relative to a consensus sequence consisting of 27 LRRs, as in Cf-9, are indicated by the hatched boxes.

Mla in barley (Jorgensen 1994; Wise & Ellingboe 1985) to account for the generation of novel *R* gene variants through intergenic recombination between tandemly duplicated sequences.

All *Hcr9*s consist of an uninterrupted open reading frame with a single intron in the 3′ untranslated region (Parniske *et al*. 1997; Thomas *et al*. 1997). A single intron is present at a similar location in *Hcr2*s (Dixon *et al*. 1996, 1998) suggesting *Hcr9*s and *Hcr2*s are derived from a common ancestral gene. All *Hcr9*s appear to encode proteins with 27 LRRs, with the exception of *Cf-4* and *Hcr9-4B* which encode proteins with 25 and 23 LRRs, respectively (Parniske *et al*. 1997; Thomas *et al*. 1997). The number of LRRs in Hcr2s ranges from 25 in Hcr2-2A to 38 in the Cf-2 proteins (Dixon *et al*. 1998).

(b) *Evolution of novel* Cf *genes*

The presence of tandemly duplicated homologous sequences at the *Cf-4/Cf-9* and *Cf-2/Cf-5* loci could promote chromosomal mispairing, resulting in unequal crossing-over or gene conversion. Unequal crossing-over is the main mechanism accounting for the meiotic instability of *Rp1* alleles in maize, the generation of novel *Rp1* recognitional specificities, and *Rp1* disease-lesion mimic alleles (Hulbert 1997). Molecular analysis of the *Cf-4/Cf-9* and *Cf-2/Cf-5* loci demonstrated that unequal crossing-over and/or gene conversion have played an important role in their evolution (Parniske *et al*. 1997). The presence of two almost identical copies of *Cf-2* (Dixon *et al*. 1996) is most likely the result of a recent sequence duplication, most probably a result of unequal crossing-over. Also *Cf-9* and *Cf-4* differ by one nucleotide in 1057 bp at their 3′ ends and are identical for a further 5.2 kb downstream (Parniske *et al*. 1997; Thomas *et al*. 1997). This suggests one of these genes was also generated from a recent cross-over or gene conversion event.

DNA sequence analysis of the *Hcr9* intergenic regions revealed extensive blocks of sequence homology upstream of *Hcr9*s including sequences derived from the 3′ exon of *LoxL* (figure 2) (Parniske *et al*. 1997). An unequal cross-over in this region which occurred prior to speciation, and involved *LoxL* sequences, may have duplicated an *Hcr9* which was amplified by subsequent unequal cross-overs (Parniske *et al*. 1997). The composition and length of the *Hcr9* intergenic blocks varies between intergenic regions indicative of sequence rearrangements during their evolution. Within the *Cf-4* and *Cf-9* haplotypes the intergenic regions are largely distinct with respect to their sequence block composition. *Hcr9* intergenic regions containing similar, but non-identical sequence block patterns, are present in *Cf-4* and *Cf-9* haplotypes but are non-syntenic. Furthermore, DNA sequence analysis of 11 *Hcr9*s (Parniske *et al*. 1997) revealed that *Hcr9*s within the *Cf-4* and *Cf-9* haplotypes contained regions of extensive DNA sequence homology. *Hcr9*s appear to be patchworks of different *Hcr9* sequences, suggesting frequent intergenic sequence exchange has occurred (figure 5). Evidence for a similar mechanism acting on *Hcr2*s was demonstrated from sequence comparisons of seven *Hcr2*s (Dixon *et al*. 1998). *Hcr2*s differ primarily in the number of LRRs they encode, particularly those of the alternating type A and type B LRRs (Dixon *et al*. 1996, 1998). Most of the intergenic sequence exchange events in *Hcr2*s appear to have been mediated through the type A and type B LRR coding sequences.

One consequence of tandemly duplicated chromosomal sequences is meiotic instability, as observed at *Rp1* (Hulbert 1997), and sequence homogenization within a gene family through frequent intergenic sequence exchange. This is undesirable at *R* gene loci where novel variants are required to combat a constantly changing pathogen population. Genetic experiments to test the stability of *Cf-9* suggest this occurs infrequently at the *Cf-9* locus. In plants homozygous for *Cf-9*, despite the presence of four highly homologous *Hcr9*s and intergenic sequences, the level of *Hcr9* mispairing resulting in loss of *Cf-9* function is less than 1 in 22 000 (Parniske *et al*. 1997). It was proposed that the unique sequence block composition of *Hcr9* intergenic regions may suppress mispairing at meiosis, and gene conversion in sister chromatids, in plants homozygous for a particular haplotype, thereby preventing sequence homogenization at *Cf* loci (Parniske

et al. 1997). However, in transheterozygous plants the loss of *Cf-4* and *Cf-9* function is increased significantly. A similar phenomenon has been described in maize heterozygous for specific *Rp1* resistance alleles (Hulbert 1997). In plants containing two distinct haplotypes the non-syntenic nature of *Hcr9* intergenic sequence blocks increases the capacity for chromosomal mispairing, and may involve *Hcr9* coding sequences as is suggested by DNA sequence analysis (Parniske *et al*. 1997).

The model for sequence diversification within *Cf* genes (Parniske *et al*. 1997) was proposed on the basis of results obtained with haplotypes originating from different *Lycopersicon* species, *L. pimpinellifolium* (*Cf-9*) and *L. hirsutum* (*Cf-4*). Our analysis of the *Cf-4/Cf-9* and *Cf-2/Cf-5* loci has revealed their dynamic nature in a number of different haplotypes. If this mechanism contributes to the evolution of *Cf* genes in natural populations, it predicts significant levels of haplotype variation within *Lycopersicon* species. Evidence for this can be seen by comparing the *Cf* locus on chromosome 6 in Cf0 and Cf5 NILs, which are both from *L. esculentum* (figure 2). Consequently, any process which promotes outbreeding in wild tomato species, and the subsequent mixing of distinct *Hcr* haplotypes would result in increased sequence variation within *Hcr*s, and novel *Cf* gene variants through intergenic sequence exchange. In a number of wild *Lycopersicon* species self-pollination is limited by a gametophytic self-incompatability (SI) system (Mutschler & Liedl 1994). The *S* locus controlling SI is linked to the *Cf-4/Cf-9* locus on chromosome 1 (Tanksley & Loaiza-Figueroa 1985) and would promote the mixing of haplotypes in some wild *Lycopersicon* species.

(c) *Recombination at the* Cf-4/Cf-9 *and* Cf-2/Cf-5 *loci*

In the single *Cf-2/Cf-5* disease-sensitive recombinant unequal crossing-over occurred within the coding sequences of two genes, *Cf2.2* and *Hcr2-5B* (figure 3). All five *Cf-4/Cf-9* recombination events occurred in intergenic sequences (Parniske *et al*. 1997; Thomas *et al*. 1997). It is possible that adjacent gene conversion events were associated with these recombination events, causing subtle changes in *Hcr9* and *Hcr2* coding sequences which were not detected by hybridization analysis.

Only two classes of *Cf-4/Cf-9* recombinants were observed (figure 3). When the physical maps of the *Cf-4/Cf-9* locus are aligned (i.e. *Hcr9-9A* with *Hcr9-4A*), *Cf-4* is located immediately distal to *Cf-9*. In each of the mispairing events that generated disease-sensitive recombinants, *Cf-4* was displaced a distance of two homologues further distal to *Cf-9* (figure 3). In theory, an unequal cross-over event should not be necessary to generate a disease-sensitive recombinant. This could be achieved through a reciprocal cross-over located between *Hcr9-4C* and *Cf-4* on one chromosome and between *Cf-9* and *Hcr9-9D* on the other (figure 2). Also, in all five recombinants *Cf-9* was located proximal to *Cf-4*. Disease-sensitive recombinants could also occur through chromosomal mispairing with *Cf-9* located distal to *Cf-4* and crossing-over between *Hcr9-9B* and *Cf-9* on one chromosome and between *Cf-4* and *Hcr9-4E* on the other. Such a recombinant would retain *Hcr9-9A*, *Hcr9-9B*, and *Hcr9-4E*.

It was suggested that *Hcr9-9A*, *Hcr9-9B*, *Hcr9-4C* and *Hcr9-4E* may confer some Avr9- and Avr4-independent resistance in the plants screened for disease sensitivity (Thomas *et al*. 1997). *Hcr9-9A* and *Hcr9-9B* both confer resistance to *C. fulvum* infection in adult plants in an Avr9-independent manner (Parniske *et al*. 1997), which accounts for their absence in the disease-sensitive recombinants. This natural 'pyramiding' of *R* genes recognizing different Avr determinants could explain the apparent durability of *Cf-9* in commercial cultivars of glasshouse-grown tomatoes. Analysis of transgenic Cf0 plants expressing *Hcr9-4E* (figure 2), has shown that this gene confers resistance to *C. fulvum* race 5 at the seedling and adult plant stages (C. Golstein, C. M. Thomas and M. H. A. J. Joosten, unpublished data). This gene also confers resistance through recognition of a novel *C. fulvum* Avr protein (Thomas *et al*. 1997). Therefore, only a subset of *Cf-4/Cf-9* recombinants would be susceptible to *C. fulvum* infection, and recombinants containing *Hcr9-9A/Hcr9-9B* or *Hcr9-4E* would not have been detected.

(d) *Sequences which determine recognitional specificity in Cf proteins*

Amino acids within the N-terminal half of Cf proteins contain the critical sequence determinants which confer recognitional specificity (figure 4). Porcine ribonuclease inhibitor (PRI) is the only LRR protein for which a crystal structure has been determined and the molecular basis for the interaction with its ligand (RNase A) has also been reported (Kobe & Deisenhofer 1994, 1995). Part of the plant extracytoplasmic LRR consensus (xxLxLxx) (figure 4) corresponds to a structural motif in PRI, which forms a repeated β-strand/β-turn structure that generates a solvent-exposed surface, including a parallel β-sheet, that binds ribonuclease. In this structure the conserved leucines project into the hydrophobic core of the protein and perform a structural role, whereas the side chains of non-conserved interstitial residues are solvent-exposed and form an extensive ligand-binding surface (Kobe & Deisenhofer 1994). This motif is conserved in LRR proteins and may play an essential function in mediating protein–protein interactions (Buchanan & Gay 1996; Kobe & Deisenhofer 1994). Sequence variation within this motif will therefore affect the specificity of ligand binding.

Only seven amino acids within the 24-amino acid LRR consensus correspond to the conserved structural motif, but a disproportionate number of amino acids distinguishing Cf-4 from Cf-9 (33 of 57), and Cf-5 from Cf-2 (55 of 86), are located in this region and correspond to the solvent-exposed interstitial residues. Most of the sequences in Cf proteins, at least at their N-termini, appear to play a structural role and only a small proportion of critical sequence differences are required to generate a distinct recognitional specificity. Therefore, differences between the amino-acid composition of Cf proteins from the same locus within the conserved β-strand/β-turn motif in their N-terminal LRRs, together with variation in the number of their LRRs, provides an explanation for the molecular basis of ligand specificity.

Analysis of 11 Hcr9 sequences showed that most sequence variation was located within the conserved structural motif of the 16 N-terminal LRRs, and a

number of hypervariable amino-acid sequence positions were also identified (Parniske *et al*. 1997). Analysis of the corresponding *Hcr9* coding sequences in this region revealed a much higher level of non-synonymous nucleotide substitutions than in the remainder of the *Hcr9* coding sequence (Parniske *et al*. 1997). This demonstrated that there has been selection for sequence diversification in this region consistent with the proposal that it encodes a potentially variable ligand-recognition domain within Hcr9s.

Our analysis has shown that intragenic sequence exchange between *Hcr9*s can result in an increase or decrease in the number of the LRRs they encode, either as a consequence of gene conversion or unequal crossing-over. This is clearly seen in the variable numbers of LRRs in Hcr2s (Dixon *et al*. 1998). Intragenic sequence exchange may also result in the mixing of LRRs with unique interstitial residue composition, as exemplified by Hcr9s (figure 5). Together with the accumulation of point mutations in these sequences (Parniske *et al*. 1997), this results in sequence variability within *Hcr*s and potentially novel Cf recognitional specificities.

(e) *The C-terminal domain of Cf proteins*

Our analysis of Cf proteins (Dixon *et al*. 1996, 1998; Jones *et al*. 1994; Thomas *et al*. 1997) suggests that they comprise two functional domains: first, a variable N-terminal region, important in determining recognitional specificity, and second, a conserved C-terminal region which includes part of domain Cl, and domains C2, C3, D, E, F and G. In the absence of a substantial cytoplasmic domain in Cf proteins it has been proposed that this region interacts with another transmembrane protein to transduce a signal after ligand binding, which activates the plant defence response (Dixon *et al*. 1996; Jones & Jones 1997; Thomas *et al*. 1997). Analysis of 11 *Hcr9*s showed that most nucleotide substitutions in this region are synonymous, showing there has been selection for sequence conservation. Cf-4 and Cf-9 are identical in this region and may interact with the same specific partner protein. There are two EMS-induced mutant alleles of *Cf-9*, which completely abolish Cf-9 function (Hammond-Kosack *et al*. 1994), and which are owing to single base mutations resulting in amino acid changes in this C-terminal domain ($D_{508}N$ and $S_{676}L$) (C. M. Thomas, unpublished data). These amino acids, which are located in the conserved structural motif discussed above, are also present in the corresponding positions in the Cf-2 proteins and Cf-5.

(f) *The structure of Cf proteins*

In PRI, the LRRs comprise mainly alternating 28- and 29-amino-acid units that form α/β coil structures (Kobe & Deisenhofer 1994). Proteins with shorter LRRs, such as the mammalian thyrotropin and choriogonadotropin receptors, are predicted to adopt similar structures (Jiang *et al*. 1995; Kajava *et al*. 1995). Cf proteins are unlikely to adopt α/β coil structures because they contain residues which are not common in α-helices (Dixon *et al*. 1996; Jones & Jones 1997; Thomas *et al*. 1997). Alternative protein folds, such as the β-helical structure described for several proteins (Kobe & Deisenhofer 1994; Emsley *et al*. 1996) would facilitate parallel stacking of the β-strand/β-turn motif. This would result in the formation of a rod-like structure with an extensive and potentially variable ligand-binding surface.

Glycosylation may also play a role in determining Cf tertiary structure (Jones & Jones 1997). Glycosylation of the highly conserved asparagine residue in the β-turn region of the consensus structural motif (figure 4) would disrupt a regular repeat structure because the side chain of this amino acid projects into the hydrophobic core of the molecule (Buchanan & Gay 1996; Kobe & Deisenhofer 1994). Most of the potential glycosylation sites in Cf proteins lie outside this region (figure 4). Also, glycosylation of asparagine residues within the β-strand/β-turn motif in domain Cl may interfere with ligand binding. Only one potential glycosylation site is present within this motif in the N-terminal LRRs of Cf-4 and Cf-9 and four in the Cf-5 and Cf-2 proteins (figure 4). However, three potential N-glycosylation sites are present within this motif, in domain C3 of Cf-4 and Cf-9 (figure 4), and two in the corresponding domain of Cf-2 and Cf-5. The LRRs in domain C3 lie outside the region predicted to function in determining the recognitional specificity of ligand binding. One of the EMS-induced mutant alleles of *Cf-9* described here ($S_{676}L$) results in the loss of one of these potential glycosylation sequences.

(g) *Cf protein ligands*

Avr4 and Avr9 are the presumed ligands for Cf-4 and Cf-9. The molecular mechanism of avirulence protein perception remains unclear. No interaction was detected between Cf-4, Cf-9 and their cognate avirulence proteins by yeast two-hybrid analysis (C. M. Thomas, unpublished data). Experiments using purified plasma membrane preparations from a number of tomato genotypes did detect a high-affinity binding activity for ^{125}I-labelled Avr9 (Kooman-Gersmann *et al*. 1996), but no significant differences in binding activity were observed in tomato genotypes which contained or lacked Cf-9. A number of membrane proteins, such as a 'pathogenicity' target, may bind Avr9 and mask any Cf-9-specific binding.

Both Avr4 and Avr9 are small cysteine-rich peptides (Joosten *et al*. 1994; Van den Ackerveken *et al*. 1992). Avr9 (28 amino acids) comprises antiparallel β-strands cross-linked by disulphide bridges to form a cystine knot protein (Vervoort *et al*. 1997). Mutations in cysteine residues abolish elicitor function (Kooman-Gersmann *et al*. 1997). The tertiary structure of Avr4 has not been determined, but it may also be a cystine knot protein. Analysis of several *C. fulvum* race 4 isolates has identified two cysteine residues essential for stability of the secreted protein (Joosten *et al*. 1997). Cystine knot proteins act as signalling molecules, such as growth factors and glycoprotein hormones which interact with cell surface receptor molecules (Isaacs 1995; McDonald & Hendrickson 1993), and as proteinase inhibitors or ion-channel blockers (Kooman-Gersmann *et al*. 1997; Vervoort *et al*. 1997). The cystine knot structure may provide a stable structure in the leaf apoplast for the presentation of specific interacting residues to a target receptor. Cf-4 and Cf-9 may have also evolved the capacity to recognize these ligands, resulting in activation of a plant defence response.

Mutagenesis of Avr9 has identified a number of solvent-exposed residues in a short connecting loop between β-strands (F_{21} to L_{24}) and in a larger loop (T_7 to D_{11}) essential for elicitor function (Kooman-Gersmann *et al.* 1997). These residues are in close spatial proximity and may represent a face of the protein which interacts with a target receptor that triggers the plant defence response. Therefore, in the case of a direct interaction between Cf-9 and Avr9, the number of intermolecular contact points may be few, possibly spanning several LRRs. Domain swaps between Cf-4 and Cf-9 involving reciprocal exchange of defined numbers of LRRs within the variable N-terminal region (figure 4) have been constructed and tested for Avr4 and Avr9 recognition. None of the chimeric proteins tested induced plant cell death in response to Avr9 or Avr4 (C. M. Thomas, unpublished data). Avr9 and Avr4 could form homodimers, which is common in cystine knot proteins (Isaacs 1995; McDonald & Hendrickson 1993), or a heterodimer with a plant-encoded protein to form a ligand complex. In this case the Cf-9 and Cf-4 ligands may make multiple contact points spanning the N-terminal LRRs within domain Cl. It is also possible that Cf-4 and Cf-9 recognize Avr4- and Avr9-dependent conformational changes in other cell surface proteins, such as the binding activity described here.

3. CONCLUSIONS

Our understanding of the nature, genome organization and evolution of *R* gene loci, particularly *Cf* gene loci, has increased immensely in recent years. The next main challenge with Cf proteins is to determine the molecular mechanism of Avr perception, and dissection of the signal transduction pathway which activates plant defences. This may result in the rational design of Cf proteins that recognize defined pathogen proteins as a means to engineer plant disease resistance. In the tomato–*C. fulvum* interaction, genetic evidence for signalling partners has come from mutational screening and has identified *Rcr* genes (required for *Cladosporium* resistance) required for the function of either Cf-9 (Hammond-Kosack *et al.* 1994*b*) or Cf-2 (M. S. Dixon, unpublished data). The observation that Cf-9 functions in tobacco suspension culture cells in an Avr9-dependent manner to generate a rapid oxidative burst (Hammond-Kosack & Jones 1996; P. Piedras, unpublished data) should also lead to a better understanding of the biochemical basis of the Cf-dependent plant defence response.

REFERENCES

Anderson, P. A., Lawrence, G. J., Morrish, B. C., Ayliffe, M. A., Finnegan, E. J. & Ellis, J. G. 1997 Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the the leucine-rich repeat coding region. *Pl. Cell* **9**, 641–651.

Balint-Kurti, P. J., Dixon, M. S., Jones, D. A., Norcott, K. A. & Jones, J. D. G. 1994 RFLP linkage analysis of the *Cf-4* and *Cf-9* genes for resistance to *Cladosporium fulvum* in tomato. *Theor. Appl. Genet.* **88**, 691–700.

Belvin, M. P. & Anderson, K. V. 1996 A conserved signalling pathway: the *Drosophila* Toll-Dorsal pathway. *A. Rev. Cell Devl Biol.* **12**, 393–416.

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

Buchanan, S. G. S. T. C. & Gay, N. J. 1996 Structural and functional diversity in the leucine-rich repeat family of proteins. *Prog. Biophys. Molec. Biol.* **65**, 1–44.

De Wit, P. J. G. M. & Spikman, G. 1982 Evidence for the occurrence of race- and cultivar-specific elicitors of necrosis in intercellular fluids of compatible interactions between *Cladosporium fulvum* and tomato. *Pl. Pathol.* **21**, 1–11.

Dickinson, M. J., Jones, D. A. & Jones, J. D. G. 1993 Close linkage between the *Cf-2/Cf-5* and *Mi* resistance loci in tomato. *Molec. Plant–Microbe Interact.* **6**, 341–347.

Dixon, M. S., Jones, D. A., Hatzixanthis, K., Ganal, M. W., Tanksley, S. D. & Jones, J. D. G. 1995 High resolution mapping of the physical location of the tomato *Cf-2* gene. *Molec. Plant–Microbe Interact.* **8**, 200–206.

Dixon, M. S., Jones, D. A., Keddie, J. S., Thomas, C. M., Harrison, K. & Jones, J. D. G. 1996 The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* **84**, 451–459.

Dixon, M. S., Hatzixanthis, K., Jones, D. A., Harrison, K. & Jones, J. D. G. 1998 The tomato *Cf-5* resistance gene and six homologues; intragenic recombination and correlation of distinct specificities with pronounced allelic variation in leucine-rich repeat copy number. *Pl. Cell.* (Submitted.)

Ellis, J., Lawrence, G., Ayliffe, M., Anderson, P., Collins, N., Finnegan, J., Frost, D., Luck, J. & Pryor, T. 1997 Advances in the molecular genetic analysis of the flax–flax rust interaction. *A. Rev. Phytopathol.* **35**, 271–291.

Emsley, P., Charles, I. G., Fairweather, N. F. & Isaacs, N. W. 1996 Structure of bordetella-pertussis virulence factor P.69 pertactin. *Nature* **381**, 90–92.

Hammond-Kosack, K. E. & Jones, J. D. G. 1994 Incomplete dominance of tomato *Cf* genes for resistance to *Cladosporium fulvum*. *Molec. Plant–Microbe Interact.* **7**, 58–70.

Hammond-Kosack, K. E. & Jones, J. D. G. 1996 Resistance gene-dependent plant defense responses. *Pl. Cell* **8**, 1773–1791.

Hammond-Kosack, K. E., Harrison, K. & Jones, J. D. G. 1994*a* Developmentally regulated cell death on expression of the fungal avirulence gene *Avr9* in tomato seedlings carrying the disease-resistance gene *Cf-9*. *Proc. Natn. Acad. Sci. USA* **91**, 10 445–10 449.

Hammond-Kosack, K. E., Jones, D. A. & Jones, J. D. G. 1994*b* Identification of two genes required in tomato for full resistance to *Cladosporium fulvum*. *Pl. Cell* **6**, 361–374.

Hammond-Kosack, K. E., Staskawicz, B. J., Jones, J. D. G. & Baulcombe, D. C. 1995 Functional expression of a fungal avirulence gene from a modified Potato Virus X. *Molec. Plant–Microbe Interact.* **8**, 181–185.

Holub, E. B. 1997 Organization of resistance genes in *Arabidopsis*. In *The gene-for-gene relationship in plant–parasite interactions* (ed. I. R. Crute, E. B. Holub & J. J. Burdon), pp. 5–26. New York: CAB International.

Hulbert, S. H. 1997 Structure and evolution of the *rp1 complex* conferring rust resistance in maize. *A. Rev. Phytopathol.* **35**, 293–310.

Isaacs, N. W. 1995 Cystine knots. *Curr. Opin. Struct. Biol.* **5**, 391–395.

Jiang, X., Dreano, M., Buckler, D. R., Cheng, S., Ythier, A., Wu, H., Hendrickson, W. A. & Tayar, N. E. 1995 Structural predictions for the ligand-binding region of glycoprotein hormone receptors and the nature of hormone–receptor interactions. *Structure* **3**, 1341–1353.

Jones, D. A. & Jones, J. D. G. 1997 The role of leucine-rich repeat proteins in plant defences. *Adv. Bot. Res. Incorp. Adv. Pl. Pathol.* **24**, 89–167.

Jones, D. A., Dickinson, M. J., Balint-Kurti, P. J., Dixon, M. S. & Jones, J. D. G. 1993 Two complex resistance loci revealed in tomato by classical and RFLP mapping of the *Cf-2*, *Cf-4*, *Cf-5* and *Cf-9* genes for resistance to *Cladosporium fulvum*. *Molec. Plant–Microbe Interact.* **6**, 348–357.

Jones, D. A., Thomas, C. M., Hammond-Kosack, K. E., Balint-Kurti, P. J. & Jones, J. D. G. 1994 Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science* **266**, 789–793.

Joosten, M. H. A. J., Cozijnsen, T. J. & de Wit, P. J. G. M. 1994 Host resistance to a fungal tomato pathogen lost by a single base-pair change in an avirulence gene. *Nature* **367**, 384–386.

Joosten, M. H. A. J., Vogelsang, R., Cozijnsen, T. J., Verberne, M. C. & de Wit, P. J. G. M. 1997 The biotrophic fungus *Cladosporium fulvum* circumvents *Cf-4*-mediated resistance by producing unstable AVR4 elicitors. *Pl. Cell* **9**, 367–379.

Jorgensen, J. H. 1994 Genetics of powdery mildew resistance in barley. *Crit Rev. Pl. Sci.* **13**, 97–119.

Kajava, A. V., Vassart, G. & Wodak, S. J. 1995 Modeling of the three-dimensional structure of proteins with the typical leucine-rich repeats. *Structure* **3**, 867–877.

Kerr, E. A. & Bailey, D. L. 1964 Resistance to *Cladosporium fulvum* Cke obtained from wild species of tomato. *Can. J. Bot.* **42**, 1541–1553 .

Kobe, B. & Deisenhofer, J. 1994 The leucine-rich repeat: a versatile binding motif. *Trends Biochem. Sci.* **19**, 415–421.

Kobe, B. & Deisenhofer, J. 1995 A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* **374**, 183–186.

Kooman-Gersmann, M., Honee, G., Bonnema, G. & de Wit, P. J. G. M. 1996 A high-affinity binding site for the AVR9 peptide elicitor of *Cladosporium fulvum* is present on plasma membranes of tomato and other solanaceous plants. *Pl. Cell* **8**, 929–938.

Kooman-Gersmann, M., Vogelsang, R., Hoogendijk, E. C. M. & de Wit, P. J. G. M. 1997 Assignment of amino acid residues of the Avr9 peptide of *Cladosporium fulvum* that determine elicitor activity. *Molec. Plant–Microbe Interact.* **10**, 821–829.

Langford, A. N. 1937 The parasitism of *Cladosporium fulvum* Cooke and the genetics of resistance to it. *Can. J. Res.* C **15**, 108–128.

McDonald, N. Q. & Hendrickson, W. A. 1993 A structural superfamily of growth factors containing a cystine knot motif. *Cell* **73**, 421–424.

Martin, G. B., Brommonschenkel, S. H., Chunwongse, J., Frary, A., Ganal, M. W., Spivy, R., Wu, T., Earle, E. D. & Tanksley, S. D. 1993 Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**, 1432–1436.

Mutschler, M. A. & Liedl, B. E. 1994 *Genetic control of self-incompatibility and reproductive development in flowering plants* (ed. E. G. Williams, A. E. Clarke & R. B. Knox), pp. 164–188. Netherlands: Kluwer.

Ori, N., Eshed, Y., Paran, I., Presting, G., Aviv, D., Tanksley, S., Zamir, D. & Fluhr, R. 1997 The *I2C* family from the wilt disease resistance locus *I2* belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. *Pl. Cell* **9**, 521–532.

Parker, J. E., Coleman, M. J., Szabo, V., Frost, L. N., Schmidt, R., van der Biezen, E., Moores, T., Dean, C., Daniels, M. J. & Jones, J. D. G. 1997 The *Arabidopsis* downy mildew resistance gene *RPP5* shares similarity to the Toll and Interleukin-1 receptors with *N* and *L6*. *Pl. Cell* **9**, 879–894.

Parniske, M., Hammond-Kosack, K. E., Golstein, C., Thomas, C. M., Jones, D. A., Harrison, K., Wulff, B. B. H. & Jones, J. D. G. 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 .

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

Scofield, S. R., Tobias, C. M., Rathjen, J. P., Chang, J. H., Lavalle, D. T., Michelmore, R. W. & Staskawicz, B. J. 1996 Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. *Science* **274**, 2063–2065.

Song, W.-Y. (and 11 others) 1995 A receptor kinase-like protein encoded by the rice disease resistance gene *Xa-21*. *Science* **270**, 1804–1806.

Song, W.-Y., Wang, G.-L., Gardner, J., Holsten, T. & Ronald, P. 1997 Evolution of the rice *Xa21* disease resistance gene family. *Pl. Cell* **9**, 1279–1287.

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

Stevens, M. A. & Rick, C. M. 1988 Genetics and breeding. In *The tomato crop* (ed. J. G. Atherton & J. Rudich), pp. 35–109. London: Chapman & Hall.

Tang, X., Frederick, R. D., Zhou, J., Halterman, D. A., Jia, Y. & Martin, G. B. 1996 Initiation of plant disease resistance by physical interaction of AvrPto and Pto kinase. *Science* **274**, 2060–2063.

Tanksley, S. D. & Loaiza-Figueroa, F. 1985 Gametophytic self-incompatibility is controlled by a single major locus on chromosome 1 in *Lycopersicon peruvianum*. *Proc. Natn. Acad. Sci. USA* **82**, 5093–5096.

Thomas, C. M., Vos, P., Zabeau, M., Jones, D. A., Norcott, K. A., Chadwick, B. P. & Jones, J. D. G. 1995 Identification of amplified restriction fragment polymorphism (AFLP) markers tightly linked to the tomato *Cf-9* gene for resistance to *Cladosporium fulvum*. *Pl. J.* **8**, 785–794.

Thomas, C. M., Jones, D. A., Parniske, M., Harrison, K., Balint-Kurti, P. J., Hatzixanthis, K. & Jones, J. D. G. 1997 Characterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in Cf-4 and Cf-9. *Pl. Cell* **9**, 2209–2224.

Tigchelaar, E. C. 1984 Collections of isogenic tomato stocks. *Rep. Tomato Genet. Coop.* **34**, 55–57.

Van den Ackerveken, G. F. J. M, Van Kan, J. A. L. & de Wit, P. J. G. M. 1992 Molecular analysis of the avirulence gene *Avr9* of the fungal pathogen *Cladosporium fulvum* fully supports the gene-for-gene hypothesis. *Pl. J.* **2**, 359–366.

Van der Biezen, E. A. & Jones, J. D. G. 1998 The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Curr. Biol.* **8**, R226–R227.

Van Wordragen, M. F., Weide, R. L., Coppoolse, E., Koornneef, M. & Zabel, P. 1996 Tomato chromosome 6: a high resolution map of the long arm and construction of a composite integrated marker-order map. *Theor. Appl. Genet.* **92**, 1065–1072.

Vervoort, J., van den Hooven, H. W., Berg, A., Vossen, P., Vogelsang, R., Joosten, M. H. A. J. & de Wit, P. J. G. M. 1997 The race-specific elicitor AVR9 of the tomato pathogen *Cladosporium fulvum:* a cystine knot protein. Sequence-specific H-1 NMR assignments, secondary structure and global fold of the protein. *FEBS Lett.* **404**, 153–158.

Whitham, S., Dinesh-Kumar, S. P., Choi, D., Hehl, R., Corr, C. & Baker, B. 1994 The product of the tobacco mosaic virus resistance gene *N*: similarity to Toll and the Interleukin-1 receptor. *Cell* **78**, 1101–1115.

Wise, R. P. & Ellingboe, A. H. 1985 Fine structure and instability of the *ML-A* locus in barley. *Genetics* **111**, 113–130.

Witsenboer, H., Kesseli, R. V., Fortin, M. G., Stangellini, M. & Michelmore, R. W. 1995 Sources and genetic fine structure of a cluster of genes for resistance to three pathogens in lettuce. *Theor. Appl. Genet.* **91**, 178–188.

Zhou, J., Loh, Y.-T., Bressan, R. A. & Martin, G. B. 1995 The tomato *Pti1* gene encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitive response. *Cell* **83**, 925–935.

Zhou, J., Tang, X. & Martin, G. B. 1997 The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a *cis*-element of pathogenesis-related genes. *EMBO J.* **16**, 3207–3218.