Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 5850–5855, May 1999
Plant Biology

Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family

MARTIN PARNISKE AND JONATHAN D. G. JONES*

The Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, United Kingdom

Edited by Frederick M. Ausubel, Harvard Medical School, Boston, MA, and approved February 10, 1999 (received for review December 21, 1998)

ABSTRACT **The tomato *Cf-4* and *Cf-9* genes are the founder members of a large gene family of homologues of *Cladosporium fulvum* resistance gene *Cf-9* (*Hcr9* genes), several of which confer resistance against *C. fulvum* through recognition of different pathogen-encoded avirulence determinants. Three loci of tandemly repeated *Hcr9* genes—*Southern Cross* (*SC*), *Milky Way* (*MW*), and *Northern Lights* (*NL*)—are located on the short arm of tomato chromosome 1. Comparisons between 2 *SC-Hcr9*s, 11 from *MW*, and 5 from *NL* implicated sequence exchange between gene family members in their evolution. The extent to which novel variants can be generated by recombination depends on the degree of sequence polymorphism available within the gene family. Here we show that physical separation of *Hcr9* genes can be associated with elevated sequence divergence. Two diverged subclasses of *Hcr9*s could be defined. These are physically separated from each other, with members of one class exclusively residing at *Northern Lights*. One exceptional *Hcr9* at *Northern Lights* carried sequence features specific for *Hcr9*s at other loci, suggesting a recent transfer of this gene by an interlocus recombination event. As members of diverged subclasses are brought into physical vicinity within a tandem repeat, a larger spectrum of sequence variants can potentially be generated by subsequent interhomologue sequence exchange.**

Biological surveillance systems for the detection of nonself molecules are essential for the defense of higher eukaryotes against pathogens. As pathogens evolve to evade these recognition systems, the generation of novel recognition specificities appears crucial for fitness and long-term survival of the host species. In plants, recognition of pathogens is mediated by resistance genes (*R* genes). The cloning of several different classes of *R* genes over the last few years enables investigation of their evolution. Certain *R* genes are members of multigene families (1–3), with distinct specificities encoded by individual family members. For example, the highly homologous L_6 and *M* genes of flax confer resistance toward races of flax rust carrying different avirulence genes (4). The tomato *Cf-4* and *Cf-9* genes for resistance to the leaf mould fungus *Cladosporium fulvum* belong to a large family of homologues of *C. fulvum* resistance gene *Cf-9* (*Hcr9*s) (5–7). This gene family is ideally suited for the analysis of evolutionary mechanisms, because at least five functional *R* genes have been identified encoding at least four distinct recognition specificities (5, 8–10). Furthermore, genetical and physical mapping as well as sequence data of multiple *Hcr9*s are available (5–7, 9, 11). Different recognition specificities within a gene family are conceptually based on amino acid sequence polymorphism. In *R* genes which consist of leucine-rich repeats (LRRs; refs. 11 and 12), the highest degree of amino acid variability is found in predicted solvent exposed residues of the LRR parallel β-sheet structure, which implicated this domain in the determination of recognition specificity (5, 13). In a previous study we observed that *Hcr9* genes located at the complex *Cf-4/9* locus consist of patchwork or mosaic genes, which suggested that novel sequence variants are generated by sequence exchange between family members (5). Conceptually, such a mechanism allows for the generation of novel specificities; however, excessive exchange would result in sequence homogenization within the gene family and the concomitant loss of specificities. Therefore, mechanisms must exist that allow for the conservation and coexistence of homologous *R* genes in tandem arrays. We suggested earlier that the unique sequence block composition of the *Hcr9* intergenic regions (IRs) at the *Cf-4/9* locus reduces the frequency of interhomologue sequence exchange, thereby limiting homogenization (5). Here, we analyze the ORFs and IRs of seven *Hcr9*s originating from additional gene clusters proximal and distal to the *Cf-4/9* locus. We show that physical separation of *Hcr9*s can be associated with increased sequence divergence. We describe the molecular traces of an interlocus recombination event which apparently transferred an entire *Hcr9* including flanking sequences into the center of another *Hcr9* cluster. This finding provides a mechanism by which sequence polymorphism that has built up during physical separation can be exploited. By occasional recruitment of diverged sequences, chimeras comprised of sequences derived from diverged family members can be generated.

MATERIALS AND METHODS

DNA sequences were established as described previously (5). Sequencing reads were assembled by using XBAP and GAP4 of the Staden package (14). Sequences were aligned and analyzed by using COMPARE, BESTFIT, GAP, and PILEUP of GCG (Wisconsin Package, Versions 8 and 9, Genetics Computer Group, Madison, WI). Alignments were optimized using the sequence editor of GCG9, and informative polymorphic sites (IPSs) were displayed with SEQUENCE OUTPUT (B. G. Spratt, University of Sussex, Brighton, U.K.).

RESULTS

High Sequence Similarity Between *Southern Cross* (*SC*) and *Milky Way* (*MW*). *Hcr9*s are organized in three loci comprising clusters of tandemly repeated genes (7), called *SC*, *MW*, and *Northern Lights* (*NL*) (Fig. 1). Comparisons between *SC* and *MW* revealed a stretch of 8.8 kb of near sequence identity between the *Lycopersicon esculentum* haplotype of *SC* and the *Lycopersicon hirsutum* haplotype of *MW*. This homologous region comprises the *SC0A* gene as well as its flanking regions.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

This paper was submitted directly (Track II) to the *Proceedings* office.
Abbreviations: LRR, leucine-rich-repeat; IR, intergenic region; IPS, informative polymorphic site.
Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. AF119040 (*Northern Lights*) and AF119041 (*Southern Cross*)].
†To whom reprint requests should be addressed. e-mail: jonathan.jones@bbsrc.ac.uk.

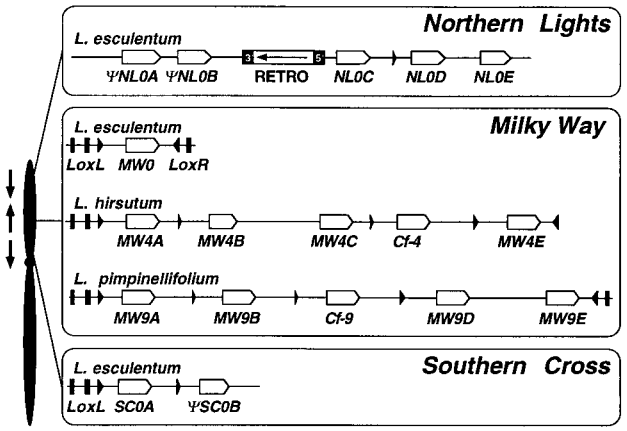

FIG. 1. Map position and physical structure of the *NL*, *MW*, and *SC* clusters of *Hcr9* genes. On the left is a schematic genetical map of tomato chromosome 1 showing the position of three *Hcr9* loci relative to each other. Arrows indicate the transcriptional polarity of *Hcr9*s at the different loci—e.g., *MW Hcr9*s are transcribed toward the telomere which is the opposite direction to *Hcr9*s at *SC* and *NL*. On the right, the physical organization of *Hcr9* clusters is shown. All of the displayed haplotypes have been entirely sequenced (5, 7). Open arrows indicate the position and transcriptional polarity of *Hcr9*s. The haplotypes at *MW* are all flanked by convergently orientated lipoxygenase genes (*LoxL* and *LoxR*). The exons of *Lox* are indicated by black boxes. The 3′ most exons are shown by black triangles, indicating the transcriptional polarity. Two types of fragments originating from the 3′ end of *LoxL* are interspersed between *Hcr9* genes at *MW*, *SC*, and *NL*, and are also represented by black triangles. *Hcr9* pseudogenes are labeled with a Ψ prefix. RETRO denotes a retrotransposon insertion in the *NL0* haplotype in which the terminal repeats are shown by black boxes. The transcriptional direction of the polyprotein gene is shown by an arrow.

It is delimited at the 5′ end by the extent of the sequenced region and at the 3′ end by the Ψ*SC0B* gene. All the sequenced haplotypes at *MW* carry a lipoxygenase sequence at their 5′ end (5) and a highly homologous sequence is present at *SC* (Fig. 1). Strikingly, the *SC0A* gene itself is more closely related to *MW4A* and *MW9A*—the most 5′ located *Hcr9* genes in the *Cf-4* and *Cf-9* clusters at *MW*—than to any other *Hcr9* gene. The nucleotide sequence of the IR between *SC0A* and Ψ*SC0B* is homologous to the IR between the *MW4A* and *MW4B* genes (Fig. 2) with which it also shares the same relative position within the cluster (Fig. 1). This sequence synteny between *SC* and *MW* is less clear for the Ψ*SC0B* gene for which no nearest evolutionary neighbor could be defined.Ψ*SC0B* carries two frameshift mutations and is, therefore, predicted to be a nonfunctional gene (7). However, comparisons of its nucleotide sequence with *Hcr9*s at *MW* showed extensive stretches of shared IPSs (Fig. 3), clearly indicating a close evolutionary relationship. The high degree of homology between *SC* and *MW* suggests these to be the products of a large duplication event encompassing tandemly arranged *Hcr9*s and flanking regions. This cluster duplication was either a very recent event, which is unlikely given the occurrence of both clusters in different *Lycopersicon* species, or sequences did not diverge because homogenizing sequence exchange still occurred after the physical separation of *MW* and *SC*. Because ectopic recombination involving reciprocal sequence exchange would lead to chromosomal rearrangements, these recombination events, if they occur, are more likely to involve gene conversions.

Specific Features of *Hcr9*s at *NL*. The overall structure of *Hcr9* genes at *NL* is similar to *Cf-9*, although variations do exist. *NL0E* carries a deletion of exactly 4 LRRs relative to Cf-9 starting after amino acid 14 of LRR 7. *NL0E* is the only *Hcr9* that does not encode the potential endoplasmic reticulum

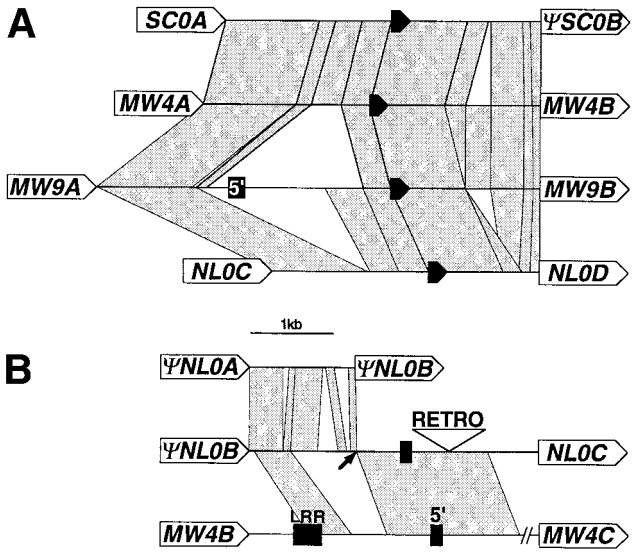

FIG. 2. Sequence affiliations between IRs. Stretches of high homology (>90%) are indicated by gray segments. The positions of *Hcr9* ORFs (or homologous regions in case of Ψ*NL0A*,Ψ*NL0B*, and Ψ*SC0B*) are shown by arrows (not to scale) labeled with the gene name. The positions of a truncated fragment of the *LoxL* gene are shown by filled arrows. 5′: *Hcr9* fragment with homology to a region flanking the transcriptional start site. LRR: sequence with homology to a region encoding LRRs 20–22 of *Cf-9*. (*A*) Stretches of near sequence identity between the IRs *SC0A*/Ψ*SC0B*, *MW4A*/*MW4B*, *MW9A*/*MW9B* and *NL0C*/*NL0D*. (*B*) The IR Ψ*NL0B*/*NL0C* schematically aligned to most closely related IRs Ψ*NL0A*/Ψ*NL0B* and *MW4B*/*MW4C*. Only the relevant region of the latter sequence is shown. The arrow indicates the position at which homology in the Ψ*NL0B*/*NL0D* IR switches between *NL* and *MW*. This point, therefore, marks a potential recombination breakpoint involved in the insertion of the *NL0C* gene into the *NL* cluster. The open triangle (RETRO) marks a retrotransposon insertion (not to scale). Homologies between IRs in *A* and *B* as well as some short insertions or deletions and imperfect repeats are not shown.

retrieval signal KKRY at the C terminus as the two lysines are replaced by glutamate residues (EERY).

Ψ*NL0A* carries a 2-bp deletion at position 501 and an in-frame stop codon at position 2269, whereas in Ψ*NL0B* the ORF is disrupted by insertion of 1 bp at position 452 and a deletion of 17 bp at position 1087. Therefore, we consider them to be pseudogenes. The 3′ half of Ψ*NL0A* is unique in that sequences homologous to the *Cf-9* region encoding LRRs 17–21 are imperfectly duplicated. The repetitive DNA structure of the region encoding LRRs might promote unequal crossing-over. Recombination between DNA repeats encoding different LRRs has been proposed to be responsible for the variability in LRR number and position in the *Hcr2* family (15) and for intragenic duplications in the *RPP5* (16) and L_6 (1) genes. The deletion of precisely four LRRs in *NL0E*, two LRRs in *Cf-4* (6), and the duplication of sequences encoding four LRRs in Ψ*NL0A* suggests that intragenic unequal recombination is also involved in the diversification of *Hcr9*s. However, it occurs to a lesser extent than in the evolution of *Hcr2*s, as most *Hcr9*s contain 27 LRRs. Intragenic unequal crossing over between *Hcr2*s is probably facilitated by the repetitive central part that encodes two types of highly conserved LRRs, whereas in *Cf-9* all the individual LRRs are distinct from each other (8, 9, 15, 17).

A Diverged Subclass of *Hcr9*s Resides at *NL*. The Ψ*NL0A*, Ψ*NL0B*, *NL0D*, and *NL0E* genes constitute a diverged subclass within the *Hcr9* gene family, the members of which all reside at *NL*. A quantitative analysis revealed that 16.7–26.8% of IPS are specific for *NL*; i.e., they are not found in *Hcr9*s at the *MW* or *SC* loci (Table 1). Likewise, 7.8–12% of the IPS in

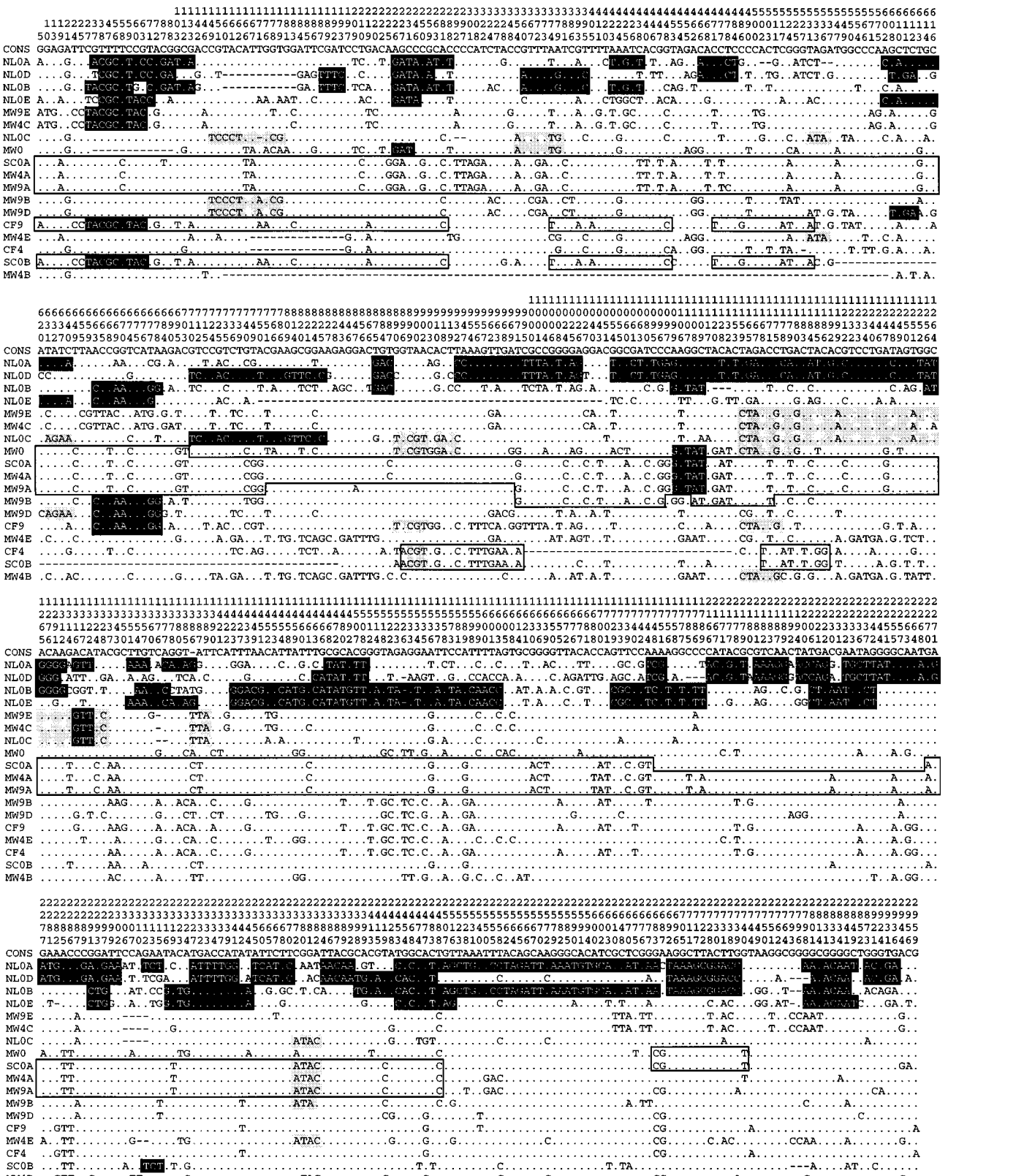

FIG. 3. Sequence relationships between *Hcr9* genes. Display of all 517 IPS (a polymorphic nucleotide which is shared between at least two *Hcr9* sequences within the alignment of 18 *Hcr9* genes). Dots indicate identity with the consensus; dashes indicate deletions. The position of each site relative to the first nucleotide of the ATG start codon within the consensus sequence is given by vertical numbers above each site. Highlighted are sequence patches of at least three consecutive IPS that deviate from the consensus sequence and are shared between at least two *Hcr9*s. Patches shared with Ψ*NL0A*, Ψ*NL0B*, *NL0D*, or *NL0E* are shown by white letters in black boxes. Patches shared with *NL0C*, but not with the other *NL Hcr9*s, are highlighted by gray boxes. Patches involving the *SC Hcr9*s *SC0A* and Ψ*SC0B* are surrounded by black lines. Sequence affiliations between *Hcr9*s at *MW* have been published elsewhere (5) and are only shown here when shared with genes in other clusters. The close relationship between *Hcr9*s at *SC* and *MW* is indicated by almost continous sequence affiliation between *SC0A*, *MW4A*, and *MW9A* and the shared patches between Ψ*SC0B* and *Cf-4* and *Cf-9*. In contrast, sequence affiliations of the *NL* genes Ψ*NL0A*, Ψ*NL0B*, *NL0D*, or *NL0E* are mostly confined to these four genes. Ψ*NL0A* carries an imperfect direct repeat of the sequence encoding LRRs 17–21 of which the first repeat unit was aligned.

Table 1. Polymorphic nucleotides in *Hcr9*s

Hcr9	Length, bp	Unique PS	IPS	Shared IPS* *MW/NL/SC*	Specific IPS† *NL*	Specific IPS† *MW/SC*
Ψ*NLOA*	2956	67	507	371	136	0
Ψ*NLOB*	2555	63	487	370	117	0
NLOD	2559	82	485	365	120	0
NLOE	2304	75	450	375	75	0
NLOC	2565	22	505	394	6†	105†
MW4A	2595	3	517	457	0	60
MW9A	2595	5	516	454	0	62
MW4B	2160	16	414	372	0	42
MW9B	2595	11	515	467	0	48
MW4C	2586	8	510	455	0	55
Cf-9	2589	11	512	464	0	48
Cf-4	2418	19	465	412	0	53
MW9D	2598	17	516	476	0	40
MW4E	2565	18	496	442	0	54
MW9E	2586	4	511	457	0	54
MWO	2535	26	503	460	0	43
SCOA	2595	7	516	456	0	60
Ψ*SCOB*	2289	27	454	420	0	34

Unique PS, number of polymorphic sites (PS) with nucleotides not found at the corresponding position in any other *Hcr9*. IPS, number of informative polymorphic sites (IPS) per sequence. This number differs between *Hcr9*s because of specific deletions and the coincidence of unique nucleotides at otherwise informative sites. Shared IPS, number of IPS carrying nucleotides found in sequences from other clusters. Specific IPS, number of IPS carrying nucleotides found only in sequences from *NL* or *MW*. No sites with IPS specific for *SC* exist.
*For *NLOC* the number of IPS shared with both *NL* and *MW* is given.
†For *NLOC* the number of sites shared only with *NL* (specific IPS NL) or only with *MW* (specific IPS *MW*) is given.

*Hcr9*s at *MW* are specific for *MW* and *SC*. In contrast, not a single IPS was found to be specific for *SC Hcr9*s, reflecting the homogeneity with *Hcr9*s from *MW* (Table 1 and Fig. 3). The members of the *NL* subclass are more related to each other than to the remainder of the gene family. They share extensive patches of IPS sequence, and the most closely related genes vary with the position in the alignment (Fig. 3). For example, Ψ*NL0A* and *NL0D* share IPS sequence from position 1054 to 1291, but Ψ*NL0A* is most closely related toΨ*NL0B* from position 2483 to 2683. This patchwork pattern indicates the occurrence of sequence exchange between Ψ*NL0A*, Ψ*NL0B*, *NL0D*, and *NL0E*, possibly as a result of unequal crossing-over or gene conversion.

Genetic evidence for unequal crossing-over near or within complex *R* gene loci has been obtained for the *rp1* complex in maize (18) and the *M* locus in flax (4). Accumulating evidence suggests unequal recombination is a major mechanism diversifying *R* gene sequences (3–6, 15, 19, 20). The finding of intergenic recombination within the *NL Hcr9* subclass is in full agreement with these results.

In contrast, only limited sequence information appears to be exchanged between the *NL* subclass and *Hcr9*s at other loci. Only three stretches longer than 5 IPSs were shared between the *NL* subclass and *Hcr9*s from *SC* or *MW*, all located 5′ of position 1109 of the consensus alignment. A single patch of three IPSs is shared between Ψ*NL0A* and Ψ*SC0B* from position 2312 to 2315 (Fig. 3). The divergence of the *NL* subclass is probably a consequence of its genetic isolation. An ancient duplication in combination with a low frequency of interlocus recombination could have provided the environment for such an independent evolution.

Under the assumption of stable mutation rate and absence of recombination, the number of acquired point mutations leading to polymorphisms could be taken as a measure for the relative age of a gene family. All *Hcr9* sequences at *MW* carry a low number of unique polymorphic nucleotides, suggesting a relatively young age. However, the finding of more diverged *Hcr9*s at *NL* shows that the gene family is older than suggested by the inspection of genes at *MW* alone. The homogeneity of sequences at *MW* contrasts with the age of the family and suggests that acquired point mutations were rapidly distributed at this locus. This independently supports the concept of sequence exchange between *Hcr9*s, which we concluded earlier from the patchwork pattern of sequence identities (5).

In the absence of recombination, the vast majority of polymorphic nucleotides are expected to be unique for a given sequence. Recombination between *Hcr9*s will lead to sequence exchange that influences the distribution of acquired point mutations in the gene family. The higher the frequency of sequence exchange, the higher the ratio of shared versus unique polymorphic nucleotides in each sequence. To obtain an independent measure for the degree of homogenization, we determined the ratio of unique polymorphic sites versus shared IPS in each *Hcr9*. The vast majority (>94%) of polymorphic nucleotides in an *Hcr9* sequence from *MW* or *SC* are shared with at least one other *Hcr9*, suggesting a high rate of sequence exchange relative to the mutation rate. In contrast, in Ψ*NL0A*, Ψ*NL0B*, *NL0D*, and *NL0E* a higher percentage of the polymorphic sites are unique, suggesting that homogenization at *NL* does not occur as rapidly as at *MW* (Table 1).

The analysis of recombination events within the *MW* locus has implicated *Hcr9* IRs in determining the position and frequency of unequal crossing-over events (5). IRs within the *NL0* cluster exhibit a high degree of polymorphism. Their length varies between 1253 bp (betweenΨ*NL0A and* Ψ*NL0B*) and 9542 bp (between Ψ*NL0B* and *NL0C*, including a retrotransposon). Sequence homologies between IRs at *NL0* are generally restricted to interrupted segments shorter than 600 bp and with the overall level of homology usually as low as 80% (data not shown). Only two longer stretches of homology were detected. The first 1243 bp of the sequence immediately 3′ of Ψ*NL0A* and the sequence immediately 3′ of Ψ*NL0B* (Fig. 2) are the most closely related IRs within the *NL* cluster, with an overall level of identity of 96%. A 1.1-kb stretch of only 81% identity interrupted by multiple insertions or deletions is shared between the sequences immediately flanking the 3′ ends of *NL0C* and *NL0D*.

***NL0C*: An *Hcr9* of the *MW* Subclass at *NL*.** Although physically positioned in the center of the *NL* cluster, the *NL0C* gene is clearly different from the other genes at *NL*. Strikingly, *NL0C* exhibits the sequence signature of *Hcr9*s at *MW*. A strong affiliation with *Hcr9*s at *MW* and *SC* is indicated by an extensive sharing of polymorphic nucleotides (Table 1). *NL0C* shares only a single patch of IPSs with *NL0D* but nine patches with *Hcr9*s at *MW* or *SC* (Fig. 3). The inconsistent sequence signature of *Hcr9* genes at *NL* suggests *NL0C* was translocated from another locus. This idea is also supported from analysis of flanking IRs that comprise extensive stretches of homology to IRs at *MW* and *SC* (Fig. 2), suggesting they could have arrived at *NL* together with *NL0C*. The sequence immediately 5′ of the ATG of *NL0C* has homology to the putative promoter regions of *Hcr9*s from different loci and is, therefore, not informative. Further 5′, a 1.5-kb region exhibits near sequence identity to the IR between the *MW4B* and *MW4C* genes (Fig. 2). This stretch of *MW* homology is flanked by a sequence with near identity outside multiple insertions or deletions to the IR between Ψ*NL0A* and Ψ*NL0B* (Fig. 2). The transition from *MW* to *NL* homology within the IR is possibly the result of an insertion of *MW* sequences into the *NL* cluster. The point of transition possibly defines a recombination breakpoint 5′ of the *NL0C* gene implicated in the insertion event (marked by an arrow in Fig. 2). The IR 3′ of the *NL0C* gene exhibits near sequence identity over almost the entire length with the IR between *MW9A* and *MW9B* (Fig. 2), encompassing a fragment with homology to the lipoxygenase sequence that flanks the *MW* and *SC* clusters at their 5′ ends (Fig. 1). This further

supports the idea of translocation, because lipoxygenase sequences are otherwise not detectable at *NL* (Fig. 1 and 2). No *NL* specific sequences could be identified in this IR, but the *NL0D* gene itself exhibits features of the *NL* subclass. We hypothesize that the ≈2-kb putative promoter region that shares homology with *Hcr9*s from all three clusters might have served as a recombination template which promoted the integration of the DNA segment carrying the *NL0C* gene.

DISCUSSION

Ectopic Recombination Between Diverged *R* Gene Clusters. The characterization of three clusters of *Hcr9* genes at the genetic, physical and sequence level allows us to correlate sequence relationships between *Hcr9*s with their physical location. The analysis of IPS led to the definition of two subclasses of *Hcr9*s. The members of both subclasses of *Hcr9*s tend to be separated from each other in individual clusters. The *NL* subclass is physically confined to *NL*. Restriction-fragment-length polymorphism analysis of *NL* haplotypes allowed the identification of polymorphism between different *Lycopersicon* species or *L. esculentum* cultivars (7). Among the cultivars analyzed, the *L. esculentum Verticillium* wilt, *Fusarium* wilt, root knot nematode, tobacco mosaic virus (VFNT) Cherry haplotype of *NL* resembled most closely the *L. esculentum* Moneymaker (Cf0) haplotype analyzed in the present study. Restriction-fragment-length polymorphism and partial sequence analysis suggested the presence of four *Hcr9*s at *NLV* that are orthologous to Ψ*NL0A*, Ψ*NL0B*, *NL0D*, and *NL0E*. However, a gene orthologous to *NL0C* was not detectable at *NL* in VFNT Cherry (7). Only the *NL* class of *Hcr9*s are found within the VFNT Cherry haplotype of *NL*. Likewise at *MW*, only the *MW* subclass is present in haplotypes originating from three species. Given this ordered occurrence, the presence of an *Hcr9* of the *MW* subclass in the *NL0* cluster is inconsistent and most likely the result of an ectopic recombination event (Fig. 4). In this scenario, the VFNT Cherry haplotype resembles a hypothetical progenitor locus of Cf0 prior to the insertion of the *NL0C* gene.

Due to the alternating orientation of the three loci each cluster is inverted with respect to its nearest cluster (Fig. 1). This arrangement results in deletions or duplications of intervening chromosomal segments upon recombination between loci, whereas a direct orientation would create dicentric bridges and acentric fragments. There is also the possibility of intrachromosomal exchange which would lead to inversion of the intervening chromosomal segment from recombination between inverted loci but a deletion in the direct orientation. The inverted orientation of the loci might be conducive to pairing, which allows rare interlocus gene conversion or unequal exchange while minimizing the risk of gross chromosomal rearrangement.

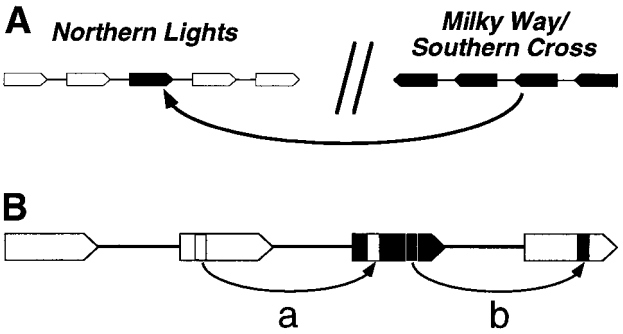

FIG. 4. Interactions between *Hcr9* genes and clusters. (*A*) Genetic isolation between *Hcr9* clusters allows sequence polymorphism to accumulate (members of the two diverged subclasses of *Hcr9*s are represented by black and white pentagons). The arrow symbolizes a hypothetical intercluster recombination event transferring a member of the *SC*/*MW* subclass of *Hcr9* genes to the *NL* locus. Because *NL* otherwise only comprises *Hcr9*s of the *NL* subclass, such recombination events occurred at a frequency too low to homogenize the sequences between distant *Hcr9* clusters. (*B*) By recombination between diverged clusters the receiving cluster (white *Hcr9*s) acquires sequence novelty (black *Hcr9*). By subsequent intracluster shuffling of sequence stretches (a and b) more diverse *Hcr9* variants can be generated than without the input of diverged family members.

Ectopic recombination events have been implicated in the evolution of *R* gene homologues in cereal genomes. In barley, rice, and foxtail millet, homologues of the nucleotide binding site LRR type of *R* genes (21) are mostly organized in genetically linked clusters comprising diverged members (22). The composition and size of these clusters exhibited strong intra- and interspecific variation, suggesting rapid reorganization (22). Similar ectopic events might have been involved in the generation of the *Dm3* complex of lettuce, in which diverged *R* gene homologues are dispersed over several Mb (13, 23). The *L* and *M* loci in flax are composed of members of the same gene family which suggests a common progenitor; however, the loci reside on different chromosomes (4). Comparative analyses of sequences of members at the *M* and *L* loci might reveal whether sequence exchange between the loci still occurred after the duplication event that separated the loci physically.

***R* Gene Evolution: A Balance Between Diversification and Conservation.** Conceptually, two opposing forces determine the evolutionary dynamics in an *R* gene family. The need to generate sequence novelty is imposed by the evolution of the pathogen. Positive selection for diversification of the putative recognition domain of *R* genes has been revealed by the analysis of synonymous and nonsynonymous substitution rates (5, 13, 19, 20, 24).

The chances of generating a gain of function allele by random mutation alone are extremely low. A strategy that also involves sequence exchange allows successful DNA fragments that have been selected for in one homologue to be tested in different *R* gene backgrounds. The validity of this concept has recently been demonstrated *in vitro*. By sequence shuffling between cephalosporinase genes of different bacterial species, alleles 270–540 times more active than any of the progenitor genes have been obtained (25). The higher the existing sequence diversity between functional genes, the larger the sequence repertoire that can be generated by DNA shuffling.

On the other hand, functional genes have to be conserved and protected against the eroding force of homogenization that is invariably associated with sequence exchange. Sequence homogenization of gene families is a well-recognized phenomenon termed "concerted evolution" (26). Selection by a specific pathogen is only imposed on a plant population at intervals and many generations may set seed without experiencing a specific avirulence determinant. Therefore, successful alleles have to be maintained in the absence of the pathogen during phases in which selection is absent. The generation and maintenance of sequence diversity, therefore, requires a reduction of the frequency at which homogenizing recombination occurs. Our analysis of *Hcr9* clusters indicates that their genetic map position, the degree of polymorphism within a cluster and the physical separation of clusters all affect the degree of homogenization.

It is well documented that recombination frequencies can vary greatly along plant chromosomes. Within the *NL0* cluster, we observe a significantly higher degree of *Hcr9* sequence divergence than at *MW* and *SC*. Although IRs at both *MW* and *NL* are polymorphic, their structure suggests IR sequence homogenization occurs to a lesser degree at *NL*. IRs at MW are composed of segments of high homology between multiple IRs and polymorphism is mainly due to the distinct segment composition of each IR (5). In contrast, IRs within *NL0* exhibit

a generally low level of homology. Therefore, it is possible that *NL* is located in a less recombinogenic region of tomato chromosome 1 than *MW* and *SC*.

Likewise, the frequency at which mispairing followed by unequal crossing-over occurs, is affected by the degree of sequence homology between the substrates. It is possible that polymorphic IRs reduce the rate of sequence exchange at *NL* both between IRs and the adjacent *Hcr9* coding regions.

The generally localized occurrence of the two subclasses of *Hcr9*s at *NL* and *MW/SC* also shows that physical separation of clusters can provide genetic niches within which sequences can diverge from the remainder of the gene family. The degree of divergence between the two subclasses of *Hcr9* genes is probably not sufficient to abolish sequence exchange between the classes, as suggested by few patches of shared IPS sequence (Fig. 3 and 4).

Genetic isolation of parts of a gene family might ultimately result in a degree of divergence that will split the family into nonrecombining classes. Members of the *Xa21 R* gene family of rice belong to two classes that are physically linked in a cluster of tandemly repeated genes. Within each class more than 95% of DNA sequence identity was observed on average, but the members of the two classes were typically only 63.5% identical (3). Recombination between both diverged classes only occurred in a short stretch of high homology near the 5′ end (3). The similar overall structure and the homology in the C-terminal portion of the Hcr2 and Hcr9 proteins together with the similar exon–intron structure of their genes suggests the *Hcr2* and *Hcr9* genes originated from a common ancestor. The *Hcr* genes as a whole are an extreme example of sequence diversification that illustrates how genetic isolation can ultimately lead to the generation of independent gene families.

The balance between diversification and conservation within the *Hcr9* gene family is accommodated by the physical distribution of clustered sequences and exploitation of the standard recombination machinery. Sequence exchange occurs in tandem arrays of *Hcr9* genes and the frequency appears to be adjusted by the polymorphism of the IRs and the position on the chromosome. The spatial organization in three clusters provides the environment for sequence polymorphism to build up and coexist. The discovery of intercluster gene transfer suggests that the machinery generating variability is fuelled by slow acquisition of this polymorphism (Fig. 4).

We thank Brande B. H. Wulff, David A. Jones, and Colwyn M. Thomas for help and useful suggestions, Erik van der Biezen for critical reading of the manuscript, and Brian G. Spratt for providing the SEQUENCE OUTPUT software. M.P. was supported by fellowships of the European Molecular Biology Organization (EMBO) and the European Community. Research at the Sainsbury Laboratory is funded by the Gatsby Charitable Foundation.

1. Anderson, P. A., Lawrence, G. J., Morrish, B. C., Ayliffe, M. A., Finnegan, E. J. & Ellis, J. G. (1997) *Plant Cell* **9,** 641–651.
2. Salmeron, J. M., Oldroyd, G., Rommens, C., Scofield, S. R., Kim, H. S., Lavelle, D. T., Dahlbeck, D. & Staskawicz, B. J. (1996) *Cell* **86,** 123–133.
3. Song, W. Y., Pi, L. Y., Wang, G. L., Gardner, J., Holsten, T. & Ronald, P. C. (1997) *Plant Cell* **9,** 1279–1287.
4. Ellis, J., Lawrence, G., Ayliffe, M., Anderson, P., Collins, N., Finnegan, J., Frost, D., Luck, J. & Pryor, T. (1997) *Annu. Rev. Phytopathol.* **35,** 271–291.
5. 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) *Cell* **91,** 821–832.
6. Thomas, C. M., Jones, D. A., Parniske, M., Harrison, K., Balint-Kurti, P. J., Hatzixanthis, K. & Jones, J. D. G. (1997) *Plant Cell* **9,** 2209–2224.
7. Parniske, M., Wulff, B. B. H., Bonnema, G., Thomas, C. M., Jones, D. A. & Jones, J. D. G. (1999) *Mol. Plant Microbe Interact.* **12,** 93–102.
8. Jones, D. A., Thomas, C. M., Hammond-Kosack, K. E., Balint-Kurti, P. J. & Jones, J. D. G. (1994) *Science* **266,** 789–793.
9. Thomas, C. M., Dixon, M. S., Parniske, M., Golstein, C. & Jones, J. D. G. (1998) *Phil. Trans. R. Soc. London B* **353,** 1413–1424.
10. Takken, F. L. W., Schipper, D., Nijkamp, H. J. J. & Hille, J. (1998) *Plant J.* **14,** 401–411.
11. Jones, D. A. & Jones, J. D. G. (1997) *Advances In Botanical Research Incorporating Advances In Plant Pathology* **24,** 89–167.
12. Kobe, B. & Deisenhofer, J. (1994) *Trends Biochem. Sci.* **19,** 415–421.
13. Meyers, B. C., Shen, K. A., Rohani, P., Gaut, B. S. & Michelmore, R. W. (1998) *Plant Cell* **10,** 1833–1846.
14. Staden, R. (1996) *Mol. Biotech.* **5,** 233–241.
15. Dixon, M., Hatzixanthis, K., Jones, D. A., Harrison, K. & Jones, J. D. G. (1998) *Plant Cell* **10,** 1915–1926.
16. Parker, J. E., Coleman, M. J., Szabo, V., Frost, L. N., Schmidt, R., van der Biezen, E. A., Moores, T., Dean, C., Daniels, M. J. & Jones, J. D. G. (1997) *Plant Cell* **9,** 879–894.
17. Dixon, M. S., Jones, D. A., Keddie, J. S., Thomas, C. M., Harrison, K. & Jones, J. D. G. (1996) *Cell* **84,** 451–459.
18. Hulbert, S. H. (1997) *Annu. Rev. Phytopathol.* **35,** 293–310.
19. Botella, M. A., Parker, J. E., Frost, L. N., Bittner-Eddy, P. D., Beynon, J. L., Daniels, M. J., Holub, E. B. & Jones, J. D. (1998) *Plant Cell* **10,** 1847–1860.
20. McDowell, J. M., Dhandaydham, M., Long, T. A., Aarts, M. G., Goff, S., Holub, E. B. & Dangl, J. L. (1998) *Plant Cell* **10,** 1861–1874.
21. Van der Biezen, E. A. & Jones, J. D. G. (1998) *Curr. Biol.* **8,** R226–R227.
22. Leister, D., Kurth, J., Laurie, D. A., Yano, M., Sasaki, T., Devos, K., Graner, A. & Schulze-Lefert, P. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 370–375.
23. Meyers, B. C., Chin, D. B., Shen, K. A., Sivaramakrishnan, S., Lavelle, D. O., Zhang, Z. & Michelmore, R. W. (1998) *Plant Cell* **10,** 1817–1832.
24. Wang, G. L., Ruan, D. L., Song, W. Y., Sideris, S., Chen, L. L., Pi, L. Y., Zhang, S. P., Zhang, Z., Fauquet, C., Gaut, B. S., *et al*. (1998) *Plant Cell* **10,** 765–779.
25. Crameri, A., Raillard, S.-A., Bermudez, E. & Stemmer, W. P. C. (1998) *Nature (London)* **391,** 288–291.
26. Dover, G. A., Linares, A. R., Bowen, T. & Hancock, J. M. (1993) *Methods Enzymol.* **224,** 525–541.