Allelic Variation in the Effector Genes of the Tomato Pathogen *Cladosporium fulvum* Reveals Different Modes of Adaptive Evolution

Ioannis Stergiopoulos,¹ Maarten J. D. De Kock,^{1,2} Pim Lindhout,² and Pierre J. G. M. De Wit¹

¹Laboratory of Phytopathology, Wageningen University & Research Centre, Binnenhaven 5, 6709 PD Wageningen, The Netherlands; ²Laboratory of Plant Breeding, Wageningen University & Research Centre, Droeendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

Submitted 27 February 2007. Accepted 14 June 2007.

The allelic variation in four avirulence (*Avr*) and four extracellular protein (*Ecp*)–encoding genes of the tomato pathogen *Cladosporium fulvum* was analyzed for a world-wide collection of strains. The majority of polymorphisms observed in the *Avr* genes are deletions, point mutations, or insertions of transposon-like elements that are associated with transitions from avirulence to virulence, indicating adaptive evolution of the *Avr* genes to the cognate *C. fulvum* resistance genes that are deployed in commercial tomato lines. Large differences in types of polymorphisms between the *Avr* genes were observed, especially between *Avr2* (indels) and *Avr4* (amino-acid substitutions), indicating that selection pressure favors different types of adaptation. In contrast, only a limited number of polymorphisms were observed in the *Ecp* genes, which mostly involved synonymous modifications. A haplotype network based on the polymorphisms observed in the effector genes revealed a complex pattern of evolution marked by reticulations that suggests the occurrence of genetic recombination in this presumed asexual fungus. This, as well as the identification of strains with identical polymorphisms in *Avr* and *Ecp* genes but with opposite mating-type genes, suggests that development of complex races can be the combined result of positive selection and genetic recombination.

Additional keywords: haplotype variation, single nucleotide polymorphisms (SNP).

Cladosporium fulvum [syn. *Passalora fulva* [Braun et al. [2003]] is a nonobligate biotrophic fungus that causes leaf mold on tomato (*Lycopersicon esculentum*, Miller) (Thomma et al. 2005). During infection, the fungus colonizes the apoplastic space of tomato leaves, in which it secretes several

effector proteins that are thought to facilitate fungal growth in the absence of cognate *C. fulvum* resistance genes but act as elicitors of defense in the presence of *C. fulvum* genes. This concept is known as the gene-for-gene hypothesis, and over the years, the *C. fulvum*-tomato pathosystem has served as a model to study gene-for-gene-based resistance (De Wit 1992; Joosten and De Wit 1999). So far, eight effector genes have been cloned and characterized from *C. fulvum* that all code for small secreted proteins with an even number of cysteine residues that contribute to protein stability through the formation of disulfide bridges. Historically, *C. fulvum* effectors have been grouped into avirulence (*Avr*) and extracellular proteins (*Ecp*) that are, however, structurally and functionally related (De Wit 2002; Rivas and Thomas 2005; Thomma et al. 2005; Westerink et al. 2004a). The division is only based on the fact that, for *Avr* genes, the cognate *C. fulvum* genes were already available in near-isogenic tomato lines when they were cloned, whereas this was not the case for *Ecp* genes.

Four *Avr* genes (*Avr2*, *Avr4*, *Avr4E*, and *Avr9*) have been cloned from *C. fulvum*, and their encoded proteins trigger a hypersensitive response (HR) in tomato plants that carry the cognate *Cf-2*, *Cf-4*, *Hcr9-4E* (also known as *Cf-4E*), and *Cf-9* genes, respectively (Joosten et al. 1994; Luderer et al. 2002; Van Kan et al. 1991; Westerink et al. 2004b). However, several strains of *C. fulvum* exist in natural populations of the fungus that are able to overcome as many as five different *C. fulvum* genes that are present in commercial tomato lines (Lindhout et al. 1989). Such strains with complex virulence spectra are thought to have emerged as a result of strong selection pressure imposed on the fungus after the introduction of the different *C. fulvum* genes into tomato cultivars since the 1930s (Lindhout et al. 1989; Rivas and Thomas 2005). The transition from avirulence to virulence is associated with DNA modifications in the *Avr* genes that include deletions, point mutations, or insertions of transposon-like elements, resulting either in complete loss of the *Avr* protein or in the production of altered forms that no longer trigger HR in tomato plants carrying the cognate *C. fulvum* genes (Westerink et al. 2004a). In most cases, loss or mutation of *Avr* proteins does not significantly affect the fitness of the fungus on cultivated tomatoes, suggesting that *Avrs* mostly represent virulence factors with minor effects (De Wit 2002; Westerink et al. 2004a).

Four *Ecp* genes (*Ecp1*, *Ecp2*, *Ecp4*, and *Ecp5*) have been cloned from *C. fulvum* (Laugé et al. 2000; Van den Ackerveken et al. 1993a). The encoded *Ecp* proteins are abundantly produced during infection by all known strains of *C. fulvum* and trigger HR in some accessions of wild *Lycopersicon* species

Corresponding author: P. J. G. M. De Wit; E-mail: Pierre.deWit@wur.nl; Telephone: +31 317483130; Fax: +31 317483412.

Current address for M. J. D. De Kock: bioMérieux, Molecular Diagnostics R&D, Boseind 15, PO Box 84, 8280 AB Boxtel, The Netherlands.

Current address for P. Lindhout: De Ruiter Seeds, PO Box 1050, 2660 BB Bergschenhoek, The Netherlands.

The genomic sequence data for *Ecp4* and *Ecp5* are available in the NCBI GenBank database under the accession numbers EF104526 and EF104527, respectively.

*The e-Xtra logo stands for “electronic extra” and indicates that two supplementary figures appear online.

that carry cognate *C. fulvum*-*Ecp* genes, such as *L. pimpinellifolium* (Laugé et al. 1998, 2000). However, no DNA modifications have been reported so far in the *Ecp* genes of the fungus (Westerink et al. 2004a). This is likely due to the lack of selection pressure imposed on the pathogen to overcome *C. fulvum*-*Ecp*-mediated resistance, since these *R* genes have not been deployed yet in commercial tomato lines. Alternatively, *Ecp* proteins might not be dispensable for the fungus, since it has been shown that disruption or modification of *Ecp1*, *Ecp2*, or both reduces virulence of the pathogen (Laugé et al. 1997; Wubben et al. 1994).

Although several mutations causing a transition from avirulence to virulence have been described in the *Avr* genes of some *C. fulvum* strains, no extensive survey of allelic variation in *Avr* and *Ecp* genes in field and greenhouse populations of the fungus has ever been performed. Here, we analyze the allelic variation in the *Avr* and *Ecp* effector genes of *C. fulvum* strains from a world-wide collection that was sampled over a period of 50 years from greenhouse- and field-grown tomatoes. In many cases, the tomato cultivars and their resistance genes from which the strains were isolated were not known, but strains were usually collected from tomato plants that were originally resistant to the fungus. Several novel polymorphisms were identified in the *Avr* genes, and for the first time, also in the *Ecp* genes of *C. fulvum*. The vast majority of DNA polymorphisms in the *Avr* genes are either nonsynonymous modifications (*Avr2*, *Avr4*), complete gene deletions (*Avr4E*, *Avr9*), or both that are generally associated with transitions from avirulence to virulence. In contrast, polymorphisms in the *Ecp* genes were considerably less frequently observed and mostly involve synonymous modifications. A haplotype network based on the polymorphisms observed in the *Avr* and *Ecp* genes of all strains in the collection revealed a complex pattern of evolution marked by multiple reticulations that could have resulted from genetic recombination in *C. fulvum*. These findings suggest that novel virulence spectra in strains of the presumed asexual fungus originate from both positive selection imposed on the *Avrs* by the cognate *C. fulvum* genes and, possibly, genetic recombination.

RESULTS

DNA polymorphisms in the *Avr* genes.

In total, 26 DNA modifications were identified in the different *Avr* genes, ten of which had never been described before (Tables 1 and 2; Figs. 1 and S1). When examining each *Avr* gene separately, large differences in the types and number of polymorphisms were observed. Such differences are clearly reflected in the number of segregating sites (θ_w) and in the esti-

mates of nucleotide diversity (π) for each individual *Avr* gene (Table 3).

Eleven DNA polymorphisms were identified in *Avr2* (Table 1; Fig. 1), eight of which had been described before, and all but one were shown to confer virulence on *Cf-2* tomato plants to strains of *C. fulvum* (Table 2) (Luderer et al. 2002). The effect on *Cf-2*-mediated resistance of the three additional mutations identified in *Avr2* has not been analyzed yet, but two mutations (c.2C>T and c.158+12_*60del179) are predicted to abolish the *Avr2* protein and, thus, are no longer expected to induce *Cf-2*-mediated resistance. Nonsynonymous polymorphisms and other modifications altering the amino-acid sequence of the *Avr2* protein accounted for ten of the 11 DNA modifications identified in *Avr2*, suggesting positive diversifying selection. These caused primarily the production of truncated *Avr2* proteins (9 of 10) rather than an amino-acid substitution (1 of 10) in *Avr2* (Table 2; Fig. 1). Based on the allelic variation observed for *Avr2*, 14 different haplotypes (haplotypes A to M) could be inferred from the collection. A haplotype network assuming a minimal number of mutational events reveals that most haplotypes are distinguished from the wild-type *Avr2* allele by only a single polymorphism (Fig. 2). Although haplotypes that can or are predicted to avoid *Cf-2*-mediated resistance (race 2 haplotypes) accounted for 12 of the 14 haplotypes identified, they represent only 32 (39.5%) of the strains in the collection. Indeed, 11 of the race 2 haplotypes were sampled only one to six times and are fairly unique for one geographic region. In addition, most of the haplotypes did not share any polymorphisms with other haplotypes, indicating that most polymorphisms surfaced independently on the background of the wild-type *Avr2* allele. However, two haplotypes were identified (haplotypes M and N) that each shared polymorphisms present in other haplotypes, suggesting occurrence of genetic recombination or recurrent mutations (Fig. 2).

For *Avr4*, nine DNA polymorphisms were identified (Table 1; Fig. 1), six of which had been described before (Joosten et al. 1994, 1997) and were shown to cause a transition from avirulence to virulence on *Cf-4* tomato plants (Table 2). The effect on the *Cf-4*-mediated resistance of the three additional modifications identified has not yet been determined. However, one (c.257G>T) is predicted to result in the amino-acid substitution Cys86-to-Phe in *Avr4*, and since Cys86 is involved in one of the three disulfide bridges that are essential for stability of the protein (Van den Burg et al. 2003), the modification is expected to cause a transition from avirulence to virulence. An excess of nonsynonymous modifications (8 of 9) was observed, seven of which resulted in an amino-acid substitution and one in a truncation of the *Avr4* protein (Table 2; Fig. 1). This is in contrast to *Avr2*, in which almost all mutations

Table 1. Number and types of polymorphisms observed in the avirulence (*Avr*) and extracellular protein (*Ecp*) effector genes of *Cladosporium fulvum*

Genes	Number of polymorphisms in types of sequences				Types of polymorphisms in protein-coding sequences			
	Total (bp)	Coding (bp)	Introns (bp)	Noncoding (5' and 3') (bp)	SNPs	Indels (bp)	Nonsynon. ^a	Synon. ^b
<i>Avr2</i>	11 ^c (570)	9 ^c (237)	2 ^c (54)	0 (279)	3	5 ^d (5)	3	0
<i>Avr4</i>	9 (806)	8 (408)	– (–)	1 (398)	7	1 (1)	7	0
<i>Avr4E^c</i>	3 (641)	2 (366)	– (–)	1 (275)	2	0 (–)	2	0
<i>Avr9^e</i>	3 (712)	1 (192)	0 (59)	2 (461)	1	0 (–)	1	0
<i>Ecp1</i>	1 (778)	0 (291)	1 (154)	0 (333)	0	0 (–)	0	0
<i>Ecp2</i>	10 (799)	4 (498)	0 (56)	6 (245)	4	0 (–)	2	2
<i>Ecp4</i>	1 (508)	1 (360)	0 (57)	0 (91)	1	0 (–)	1	0
<i>Ecp5</i>	1 (882)	0 (348)	1 (397)	0 (137)	0	0 (–)	0	0

^a Nonsynonymous are polymorphisms that are predicted to result in a mutation in the encoded proteins. They only refer to single nucleotide polymorphisms (SNP).

^b Synonymous are polymorphisms in protein-coding sequences that are not predicted to result in a mutation in the encoded proteins.

^c Including a large deletion of 179 bp and the insertion of a 5.175 kb LINE-like element (Luderer et al. 2002) in the sequence of *Avr2*.

^d Excluding the insertion of a 5.175-kb LINE-like element that has been observed in the coding region of this gene (Luderer et al. 2002).

^e For *Avr4E* and *Avr9*, the entire gene was deleted in 14 (17%) and seven (9%) strains of the collection, respectively.

lead to truncated Avr2 proteins, suggesting that selection is favoring different types of adaptive mutations in these two proteins. Furthermore, seven of the polymorphisms in *Avr4* were present in parts of the gene that code for the mature Avr4 protein, and four result in the substitution of a Cys residue involved in a disulfide-bridge (Van den Burg et al. 2003). This excess of nonsynonymous substitutions, as compared with synonymous ones in the protein-coding sequences of *Avr4*, suggests positive diversifying selection (Table 3). Based on the polymorphisms present in *Avr4*, a haplotype network with ten different haplotypes could be inferred that is dominated by two haplotypes that can not overcome *Cf-4*-mediated resistance (haplotype with wild-type *Avr4* allele and haplotype A) and from which all other predominantly race 4 haplotypes (avoiding *Cf-4*-mediated resistance) are derived by single independent mutations (Fig. 2). The seven race 4 haplotypes, as well as the haplotype for which we have not yet tested whether it overcomes *Cf-4*-mediated resistance, accounted for only 15 strains in the collection. The most common race 4 haplotype (haplotype H) was present in ten European strains, while the two non-race 4 haplotypes (haplotypes with wild-type *Avr4* allele and haplotype A) were present world wide.

Two nonsynonymous single nucleotide polymorphisms (SNP) had been described before in *Avr4E*, as well as the absence of this gene in the genome of some strains. Both mutations cause a transition from avirulence to virulence on *Hcr9-4E* tomato plants (Table 2) (Westerink et al. 2004b). Sequence analysis of all strains in the collection revealed the presence of only one additional SNP in the 3' flanking region of *Avr4E* (c.*36C>T) (Fig. 1). All three identified SNP were always present in the same open reading frames, thus giving rise to only one isoform of the Avr4E protein. Based on the observed polymorphisms in *Avr4E*, only three haplotypes could be distinguished (Fig. 2). The haplotype corresponding to the wild-type *Avr4E* allele was the most predominant one (present in 44 strains), followed by the haplotype corresponding to the mutant *Avr4E* allele (23 strains) and the haplotype that lacked the *Avr4E* gene (14 strains).

Loss of *Avr9* from the genome of some strains of *C. fulvum* is the only mechanism described so far to avoid *Cf-9*-mediated resistance (Table 2) (Van Kan et al. 1991). This study revealed the presence of two nucleotide substitutions in the 5' and 3' flanking regions of *Avr9*, respectively, in addition to one nonsynonymous substitution present in the coding sequence of *Avr9* (Table 1; Fig. 1). The effect of the nonsynonymous substitution in *Avr9* on *Cf-9*-mediated resistance is not known, but as it is located within the signal peptide of the precursor protein (Van den Ackerveken et al. 1993b), it is not expected to significantly affect Avr9 activity. All three SNP identified in *Avr9* were observed in the same five Japanese strains, while absence of *Avr9* was detected in seven mainly European strains. Thus, the majority of strains in the collection cannot overcome *Cf-9*-mediated resistance (74 out of the 81 strains) (Fig. 2).

DNA polymorphisms in the *Ecp* genes.

For the first time, we report the presence of sequence variation in the *Ecp* genes of *C. fulvum*. In total, 13 DNA polymorphisms involving 12 SNP and one indel were identified (Tables 1 and 2; Fig. 1). In contrast to the *Avr* genes, polymorphisms in protein-coding sequences ($n = 5$) of the *Ecp* genes were less frequently observed than polymorphisms in non-protein coding regions ($n = 8$), whereas no deletions of an entire gene or other large deletions or insertions were observed. Indices of nucleotide diversity as estimated by θ_w and π were considerably lower for the individual *Ecp* genes as compared with the *Avr* genes, reflecting the limited sequence variation in the *Ecp*

genes (Table 3). Ten polymorphisms were identified in *Ecp2* (Table 1; Fig. 1), four of which were unique to one strain (haplotype A), one unique to another strain (haplotype B), and four were common to these two strains (Fig. 2). Only one SNP was found in each of the other *Ecp* genes examined; in *Ecp5* a single SNP was found in one strain, whereas the single SNP found in *Ecp1* and *Ecp4* were observed in 20 and 28 strains, respectively (Table 1; Figs. 1 and S1). Consequently, although diversity based on the number of segregating sites was the highest for *Ecp2* ($\theta_w = 1.81$), nucleotide diversity based on the

Table 2. Allelic variation in the avirulence (*Avr*) and extracellular protein (*Ecp*) effector genes of *Cladosporium fulvum*^a

Gene	Polymorphisms at DNA level ^a	Mutation in protein	Effect ^b
<i>Avr2</i>	c.2T>C	p.Met1?	Yes ^c
	c.(64_69)delA	p.Lys22AsnfsX11	Yes ^d
	c.(64_69)insA	p.Leu24ThrfsX18	Yes ^d
	c.70delC	p.Leu24TyrfsX10	Yes ^d
	c.(117_119)insA	p.Tyr41ValfsX1	Yes ^d
	c.127T>G	p.Leu43Val	N.d. ^e
	c.158+12_*60del179	p.Asp54_Gly78del	Yes
	c.158+26_158+28delTGA	–	No ^d
	c.196C>T	–	–
	p.Gln67X	Yes ^c	–
	c.211delT	p.Cys72AlaX?	Yes ^d
	c.57_58ins5kb	p.Ala20ins5kb	Yes ^d
	c.-57G>A	–	No
	c.124delC	p.Gln42LysfsX63	Yes ^f
<i>Avr4</i>	c.191G>A	p.Cys64Tyr	Yes ^f
	c.197C>T	p.Thr66Ile	Yes ^f
	c.199T>C	p.Tyr67His	Yes ^f
	c.209G>A	p.Cys70Tyr	Yes ^f
	c.257G>T	p.Cys86Phe	Yes ^c
	c.265G>A	p.Gly89Arg	N.d. ^e
	c.326G>A	p.Cys109Tyr	Yes ^c
	c.454T>C	p.Phe82Leu	Yes ^g
	c.488T>C	p.Met93Tyr	Yes ^g
	c.*36C>T	–	No
<i>Avr4E</i>	Gene deletion	–	Yes ^g
	c.-103A>G	–	No ^c
	c.23T>C	p.Val8Ala	No ^c
	c.*47A>C	–	No ^c
<i>Avr9</i>	Gene deletion	–	Yes ^h
<i>Ecp1</i>	c.171+33T>C	–	–
<i>Ecp2</i>	c.-126C>T	–	No ^c
	c.-95T>A	–	No ^c
	c.-94T>C	–	No ⁱ
	c.-50C>T	–	No ⁱ
	c.-25_-26insC	–	No ⁱ
	c.-14C>A	–	No ^c
	c.39T>C	–	N.d.
	c.87C>A	–	No ⁱ
	c.191G>T	p.Arg64Ile	No ⁱ
	c.196T>G	p.Ser66Ala	N.d.
<i>Ecp4</i>	c.323G>T	p.Gly90Val	No ^j
<i>Ecp5</i>	c.86+146G>A	–	–

^a The predicted effect of the DNA modifications on the produced Avr and Ecp proteins, respectively, and on the *Cf*-mediated resistance (when known) is also indicated. Polymorphisms at the DNA and protein level are described as recommended by the Human Genome Variation Society (den Dunnen and Antonarakis 2000).

^b Effect on overcoming *Cf*-mediated resistance

^c Not experimentally tested but deduced from the positioning of the polymorphism outside the mature protein coding sequence of the genes or based on their effect on stability of the protein.

^d Luderer et al. 2002.

^e N.d. = Not determined.

^f Joosten et al. 1994; Joosten et al. 1997.

^g Westerink et al. 2004a.

^h Van Kan et al. 1991.

ⁱ as deduced by Laugé et al (1998).

^j as deduced by Laugé et al (2000).

average number of nucleotide differences between two random sequences was highest for *Ecp4* ($\pi = 0.0009$) when compared with the other *Ecp* genes (Table 3). Polymorphisms in protein-coding sequences of the *Ecp* genes were, in general, rare. Two nonsynonymous and two synonymous substitutions were present in *Ecp2*, and only a single nonsynonymous modification was present in *Ecp4*. In contrast, no polymorphisms were observed in the protein-coding sequences of the *Ecp1* and *Ecp5* genes. This absence of polymorphisms could imply that selective constraints for maintaining the protein sequence are acting on these genes or, alternatively, could be indicative of a recent common ancestor for the strains. Sequencing of the *Ecp4*- and *Ecp5*-specific polymerase chain reaction (PCR) products also

revealed the presence of a 57- and a 397-bp intron in these two genes, respectively (Fig. 1), that had not been reported before, since *Ecp4* and *Ecp5* were cloned from a fungal cDNA library and no genomic clones had been isolated so far (Laugé et al. 2000).

Haplotype variation.

Among the 81 strains of the *C. fulvum* collection, 42 different multilocus haplotypes (N_{hapl}) could be distinguished, based on all types of polymorphisms (including gene deletions or other large deletions or insertions) in the *Avr* and *Ecp* effector genes. Most haplotypes were present only once ($N_{\text{hapl}} = 26$) or twice ($N_{\text{hapl}} = 8$), and only eight haplotypes were represented

Fig. 1. Sequence variation at the nucleotide level in the *Cladosporium fulvum* avirulence (*Avr*) and extracellular protein (*Ecp*) effector genes. Open reading frames are shown as white boxes and introns as black boxes. Parts of the effector gene-encoded proteins that are processed at the amino and carboxy terminus are shown as gray boxes. DNA modifications leading to a mutation in the produced protein are indicated by black flags when they cause a frameshift or by white flags when they cause a single amino-acid substitution in the protein. In the case of *Avr2*, two of the frameshift mutations are caused by nucleotide substitutions instead of indels and these are also indicated by black flags. DNA modifications that do not result in a mutation in the protein are indicated by a T. Double asterisks (**) indicate the newly identified allelic variations reported for the first time in this study. For *Ecp1*, the open circles indicate sequence errors in the previously published sequence of this gene (Van den Ackerveken et al. 1993a). For *Avr4E* and *Avr9*, absence of the entire gene was also observed. Figure is drawn at scale.

by more than three strains (Figs. 3 and S1). This suggests a large number of rare polymorphisms in the *Avr* and *Ecp* genes, but it could also suggest that polymorphisms are arranged within haplotypes in different complex combinations rather than in a few clonal lineages as previously suggested. All but two of the haplotypes were specific to one of the geographically-based subcollections but, as most haplotypes were sampled at very low frequencies, it reduces the possibilities of encountering each haplotype in every subcollection. The haplotype network constructed based on the polymorphisms present in the *Avr* and *Ecp* effector genes shows no clustering of haplotypes within specific geographic regions (Fig. 3). In contrast, a number of polymorphisms in some of the effector genes are shared among strains of different geographic origin. Indeed, of the 14 polymorphisms that could be characterized as nonrare variants (present in more than four strains, which is 5% of the collection) one (*Avr*4: c.326G>A) was found to be unique to the European subcollection and three (*Avr*9: c.-103A>G, c.23T>G, c.*47A>C) to the Japanese one, while all the others were shared by at least two different subcollections. The haplotype network also revealed a complex pattern of evolution marked by multiple reticulations and a number of intermediate hypothetical haplotypes (Fig. 3). This suggests that the virulence spectra of some of the haplotypes did not only evolve by sequential accumulation of mutations in *Avr* and *Ecp* genes but could also be the result of recombination between different haplotypes. Alternatively, recurrent mutation under a scheme of positive selection could lead to a similar outcome. The most complex haplotypes showed nine and eight polymorphisms in the *Avr* and *Ecp* effector genes, while 90% of the haplotypes showed between two and five polymorphisms (Fig. S2). One haplotype (haplotype 9) could be described as race 2.4.4E.9, as it overcomes the cognate *C. fulvum* genes *Cf*-2, *Cf*-4, *Cf*-4E (*Hcr*9-4E), and *Cf*-9. This haplotype was present in four strains, two of which were mating-type MAT1-1 and

two mating-type MAT1-2 (Fig. S1) (Stergiopoulos et al. 2007). In addition, two additional haplotypes (haplotypes 1 and 23) were observed in strains with fewer identical polymorphisms in their *Avr* or *Ecp* effector genes, or both, but with an opposite mating-type (Fig. S1). Other races able to overcome cognate *C. fulvum* genes included race 2 ($N_{\text{hapl}} = 11$), race 4E ($N_{\text{hapl}} = 5$), race 2.4 ($N_{\text{hapl}} = 1$), race 2.4E ($N_{\text{hapl}} = 2$), race 2.9 ($N_{\text{hapl}} = 1$), race 4.4E ($N_{\text{hapl}} = 5$), race 4E.9 ($N_{\text{hapl}} = 1$), race 2.4.4E ($N_{\text{hapl}} = 6$), and finally, race 2.4E.9 ($N_{\text{hapl}} = 1$). Eight haplotypes were also detected that are supposedly not able to overcome any *C. fulvum* gene and thus could be designated as race 0. However, it is still possible that some of the haplotypes might contain mutations in novel *Avr* genes that have not yet been cloned, including *Avr*5, *Avr*8, and *Avr*11. As even some of the most complex races are represented by more than one haplotype, it is clear that races with similar virulence spectra are not always defined by the presence of identical mutations in their *Avr* or *Ecp* genes.

DISCUSSION

Race evolution and alleged genetic recombination in *C. fulvum*.

In early breeding programs dating as far back as 1933 (Butler and Jones 1949; Langford 1937), *C. fulvum* genes were incorporated sequentially in commercial tomato lines from several wild *Lycopersicon* species (Rivas and Thomas 2005). Yet, in many cases, soon after the deployment of resistant varieties carrying novel *C. fulvum* genes, resistance was overcome by the appearance of new virulent races of the fungus (Bailey 1950; Boukema 1981; Day 1957; Higgins and Hollands 1987; Kishi and Abiko 1976; Langford 1937; Laterrot 1986; Leski 1970; Lindhout et al. 1989). Based on the present and previous studies, overcoming *C. fulvum*-mediated resistance can now be attributed to DNA modifications in the *Avr* genes of these

Table 3. Calculated indices of molecular diversity based on the sequence variation in the avirulence (*Avr*) and extracellular protein (*Ecp*) effector genes of *Cladosporium fulvum*

Genes	N^a	Nucleotide diversity						Haplotype diversity ^d		Codon usage ^e		
		π_n^b	θ_w^c	π_n^b	θ_w^c	π_{syn}^b	θ_{syn}^c	N_{hapl}	Hd	ENC	CBI	G_{syn}
<i>Avr</i> 2	78	0.0039 (± 0.0024) ^{s.d.}	0.609 (± 0.373) ^{s.d.}	0.0021	0.0035	0	0	14	0.83	60.59	0.38	0.49
<i>Avr</i> 4	81	0.0011 (± 0.0009)	1.611 (± 0.682)	0.0015	0.0046	0	0	10	0.691	49.74	0.38	0.59
<i>Avr</i> 4E	67	0.0021 (± 0.0015)	0.628 (± 0.386)	0.0033	0.0015	0	0	3	0.602	50.17	0.43	0.62
<i>Avr</i> 9	74	0.0005 (± 0.0006)	0.616 (± 0.378)	0.0009	0.0014	0	0	3	0.266	40.15	0.56	0.44
<i>Ecp</i> 1	81	0.0005 (± 0.0005)	0.201 (± 0.201)	0	0	0	0	2	0.376	41.97	0.55	0.69
<i>Ecp</i> 2	81	0.0006 (± 0.0006)	1.813 (± 0.739)	0.0003	0.0011	0.0008	0.0034	3	0.073	59.24	0.45	0.61
<i>Ecp</i> 4	81	0.0009 (± 0.0009)	0.201 (± 0.201)	0.0017	0.0007	0	0	2	0.450	41.75	0.60	0.76
<i>Ecp</i> 5	81	0.00003 (± 0.0001)	0.201 (± 0.201)	0	0	0	0	2	0.0247	51.89	0.37	0.55

^a Number of strains used for the analysis. Strains where gene deletion or other large deletions and insertions have occurred are excluded from calculations of nucleotide diversity and the estimation of the ratio nonsynonymous site and the number of synonymous substitutions per synonymous site (Ka/Ks). In contrast, haplotypic diversity and codon usage were calculated for all genes in the complete collection of 81 strains.

^b Nucleotide diversity based on the average number of nucleotide differences between two random sequences (π ; Nei 1987), estimated for the entire length of the sequences analyzed (π_n), for only nonsynonymous substitutions (π_n), and synonymous ones (π_{syn}).

^c The statistic of the number of segregating sites (θ_w ; Watterson 1975), estimated for the entire length of the sequences analyzed (θ_w), for nonsynonymous (θ_w) and synonymous (θ_{syn}) substitutions.

^d Number of haplotypes (N_{hapl}) and Nei's diversity index (Nei 1987) corrected for sample size (Hd), calculated based on polymorphisms in the *Avr* and *Ecp* genes for the complete collection of 81 strains.

^e Parameters related to codon usage, such as the effective number of codons (ENC; Wright 1990), codon bias index (CBI; Morton 1993), and the G+C content at synonymous (third) codon positions (G_{syn}).

Fig. 2. Median-joining network of haplotypes inferred based on polymorphisms observed in individual avirulence (*Avr*) and extracellular protein (*Ecp*) effector genes. Haplotypes are depicted as circles the radius of which is proportional to their allele frequency. Black circles indicate haplotypes that overcome cognate *Cladosporium fulvum*-mediated resistance, gray ones haplotypes that are predicted to overcome cognate *C. fulvum*-mediated resistance but have not been tested experimentally, whereas white circles represent haplotypes that still trigger hypersensitive response in the presence of cognate *C. fulvum* genes. Branch length is proportional to the number of nucleotide changes. Disrupted branches indicate complete deletion of an effector gene or other extensive deletions or insertions. Predicted intermediate (median) vectors are shown as squares. Networks were drawn only for the individual *Avr* genes and *Ecp2*, as only one polymorphism was found in each of the other *Ecp* genes examined.

ances that range from complete loss of an *Avr* gene to single amino-acid substitutions and frameshifts in *Avr* proteins that are no longer recognized by the cognate *C. fulvum* protein. Mutations in *Avr* genes leading to gain of virulence on previ-

ously resistant tomato cultivars have emerged independently several times in tomato-growing areas, since different alleles overcoming the same cognate *C. fulvum* genes have been observed. This is also clear from the minimum mutation net-

Fig. 3. Median-joining network of haplotypes inferred based on combined polymorphisms observed in avirulence (*Avr*) and extracellular protein (*Ecp*) effector genes. Haplotypes are depicted as circles, the radii of which are proportional to their sampled frequency. Intensity of the circles indicates the geographic region from which haplotypes were sampled, which is also indicated next to the haplotype's name. Black circles represent Europe (Eu), dark gray Japan (Jap), and light gray America (Am), while the substantially smaller collections are shown in various other shades of gray. Predicted intermediate (median) vectors are shown as square. The virulence pattern (race) of the haplotypes, based on mutations in their *Avr* genes, is also indicated. Branch length is proportional to the number of nucleotide changes. These numbers are also indicated next to branches. Disrupted branches indicate complete gene deletion of an effector gene or other extensive deletions or insertions.

works constructed for each individual *Avr* gene, in which most of the haplotypes corresponding to a specific allelic variant differ from the wild-type *Avr* allele by only a single mutation. The extremely low levels of synonymous substitutions observed in the eight effector genes of *C. fulvum* as well as in house-keeping (0% in internal transcribed spacer sequences in 44 strains examined so far; data not shown) and mating-type genes (Stergiopoulos et al. 2007) suggests that the examined races of the fungus could share a recent common ancestor that followed the worldwide movement of cultivated tomatoes by breeders and growers. In addition, haplotypes with complex virulence spectra have been identified that can overcome as many as four cognate *C. fulvum* genes present in commercial tomato lines. The haplotype network constructed based on the overall allelic variation observed in the effector genes suggests that the complex virulence spectra of some races can be explained by a stepwise accumulation of mutations in multiple *Avr* genes, as expected for a clonal population under selection pressure imposed by the sequential pyramiding of cognate *C. fulvum* genes in commercial tomato lines (Rivas and Thomas 2005). However, reticulate relationships among several multi-locus haplotypes are also observed, suggesting that genetic recombination could also create novel combinations of mutated *Avr* alleles. Indeed, based on the pattern of polymorphisms present in the *Avr* and *Ecp* genes, a minimum number of three recombination events ($R_m = 3$) (Hudson and Kaplan 1985) are suggested to have occurred in the history of the collection. Alternatively, recurrent mutation could explain the apparent homoplasy, but as most of the polymorphisms occur only in low frequencies, it is unlikely that, in all cases, the homoplasious sites are targets of recurrent mutation. Furthermore, the recent cloning of two opposite mating-type genes (*MAT1-1-1* and *MAT1-2-1*) from *C. fulvum*, combined with amplification fragment length polymorphism analysis, revealed high levels of haplotypic diversity that might be the result of genetic recombination (Stergiopoulos et al. 2007). Therefore, it is proposed that complex virulence spectra observed in strains of *C. fulvum* do not only arise from a stepwise accumulation of mutations in *Avr* genes, as previously suggested (Joosten and De Wit 1999; Westerink et al. 2004b), but are most probably the result of a combination of mutation, positive selection, and genetic recombination.

Allelic variation in *Avr* genes.

Most allelic variations in the individual *Avr* genes of *C. fulvum* identified so far are present in protein-coding sequences and cause transitions from avirulence to virulence on tomato lines carrying the cognate *C. fulvum* genes (Luderer et al. 2002; Westerink et al. 2004a). Therefore, it is suggested that the majority of allelic variation observed in *Avr* loci is a consequence of the selection pressure imposed on the pathogen by the deployment of the cognate *C. fulvum* genes in commercial tomato lines. However, large differences in the types of mutations were observed in the different *Avr* genes, indicating that selection pressure favors different types of adaptation, which could reflect differences in effector functions and point towards direct or indirect interactions of the effectors with their cognate *C. fulvum* proteins (Jones and Dangl 2006).

Most polymorphisms observed in *Avr4* are SNP that cause amino-acid substitutions of mainly Cys residues in the produced protein, which, in combination with the complete absence of synonymous substitutions, points towards positive diversifying selection. Structural studies revealed four disulfide-bonds in *Avr4*, three of which have been shown to destabilize the protein when any of these Cys residues were substituted artificially by alanine (Van den Burg et al. 2003). It is demonstrated that *Avr4* can bind to chitin and protects the fungus against plant

chitinases during pathogenesis (Van den Burg et al. 2006; Van Esse et al. in press), while *Avr4* proteins with substituted Cys-to Ala residues are unstable in tomato apoplastic fluids and no longer induce *Cf-4*-mediated resistance. Yet, these mutants are still able to bind chitin and protect *C. fulvum* against deleterious effects of plant chitinases (Van den Burg et al. 2006). Therefore, it is suggested that *Avr4* is an essential virulence factor, when selection of mutants favors variants that abolish recognition by *Cf-4* but still retain their ability to bind chitin. Sequence diversification instead of gene loss as a means of overcoming host resistance might further suggest a direct interaction between the wild-type *Avr4* protein and *Cf-4*, as has been observed before for cases in which effectors directly interact with plant resistant proteins (Allen et al. 2004; Dodds et al. 2006; Rehmany et al. 2005).

In contrast to *Avr4*, almost all polymorphisms identified in *Avr2* are nonsynonymous modifications that involve indels leading to truncated nonfunctional *Avr2* proteins. This suggests that *Avr2* is under positive diversifying that favors mutations that abolish the *Avr2* gene product. The abolishment of the *Avr2* gene product is functionally similar to entire gene deletions observed in the cases of *Avr4E* and *Avr9*. Unlike sequence diversification, circumvention of host recognition by jettison of an effector gene product can lead to long-term stable resistance and has been frequently observed with fungal pathogens (Couch et al. 2005; Gout et al. 2006; Orbach et al. 2000; Schurch et al. 2004). This type of adaptive evolution is often witnessed with effectors that modify host targets and are thus indirectly recognized by plant resistance proteins that detect changes in virulence targets (Jones and Dangl 2006; Rohmer et al. 2004). This also seems to be the case for *Avr2*, which does not directly interact with *Cf-2* but is perceived through its interaction with the tomato cysteine protease *Rcr3* (Krüger et al. 2002; Rooney et al. 2005). Moreover, similar to *Avr2*, cases of adaptive loss of gene function have been previously observed in mammalian and microbial systems and are thought to be one of the driving forces in molecular evolution (Olson et al. 1999). In that respect, the conditionally beneficial mutated alleles of *Avr2* that sweep to high frequency in the genome of *C. fulvum* might serve as a reservoir, available for reversion once selection pressure by *Cf-2* is lifted (Olson et al. 1999). Frameshift mutations, in particular, are often reverted by the occurrence of a second complementary frameshift that could restore the function of the protein.

Although an attractive way for avoiding host recognition, jettison of effector genes assumes a minimal fitness penalty or compensation of the effector's function by other effectors. Unfortunately, the intrinsic functions of *Avr4E* and *Avr9* are not yet known. It is again possible that the conditional benefits from losing a gene under strong selection pressure might be higher than the costs for the loss. As fitness in an ancestral niche does not necessarily imply fitness in a new shifting environment, most genetic traits will be neutral under a change in the host environment, but others, such as the ones encoded by avirulence genes, may negatively affect fitness by triggering host defense responses (Arnold et al. 2007). The relative frequency of gene deletions could reflect the strength of the positive selection pressure imposed on these genes by the cognate *C. fulvum* genes. Finally, the chromosomal location of the genes could also have affected the type of polymorphisms observed, as is the case for the telomeric sequences that are more prone to be lost from the genome. For *Avr4E*, in addition to gene loss, point mutations that result in amino-acid substitutions have also been observed. It is possible that these mutations represent compensatory mutations, which may have evolved recently and restored a possible loss of fitness caused by the *Avr4E* gene deletions.

It has to be noted that inferences on positive selection acting on the Avr proteins are mostly based on i) phenetic observations that the majority of mutations are associated with transitions from avirulence to virulence and ii) presence of a number of nonsynonymous substitutions in the coding sequences of the Avr genes in contrast to the complete absence of synonymous ones. In most cases however, presence of indels, the low frequency of individual polymorphisms present in the Avr genes, and, in most cases, absence of synonymous substitutions in protein-coding sequences of the genes did not permit a rigorous statistical analysis on positive selection using common approaches, such as tests based on comparisons of nonsynonymous (Ka) to synonymous (Ks) substitution rates (Nei and Gojobori 1986) and likelihood ratio tests (LRT) implemented in programs like DnaSP (Rozas et al. 2003) and CODEML (Yang 1997), respectively. Due to these limitations, for Avr2, Avr4E, and Avr9, LRT failed to reject the null selection models in favor of positive selection (data not shown). Conversely, positive selection could, indeed, be inferred for Avr4, while a Bayes empirical Bayes approach (Yang et al. 2005) indicated several sites to evolve under positive selection, although, in the absence of synonymous substitutions in Avr4, extremely inflated values of ω ($=K_a/K_s$) were obtained (Model M2a [positive selection]: $P_2 = 1.000$, $\omega_2 = 951.431$, M8 (Beta ω) $P_1 = 1.000$, $\omega = 951.431$). Extensive computer simulations performed by Anisimova and associates (2001, 2002) showed that LRT become too conservative to detect adaptive evolution when sequences are very similar and lineages are small. In our set of data, polymorphisms in Avr2 are mostly indels, while nucleotide variation in the coding sequences of Avr9 and Avr4E is very limited. Thus, failure of the LRT to confirm positive selection by statistical analysis in these cases is not really surprising. The same holds also for analysis of the Ecp genes in which even fewer polymorphisms are present in the coding sequences.

Allelic variation in the Ecp genes.

The present study represents the first survey on allelic variation in the Ecp genes of *C. fulvum*. However, in contrast to the large nucleotide diversity observed in the Avr genes, nucleotide diversity in the Ecp genes is considerably less and mostly involves silent substitutions or other modifications in non-protein coding sequences. Such modifications can be attributed to neutral natural variation rather than *C. fulvum*-Ecp-driven mutations, as presumably no selection pressure has been imposed on the Ecp genes (Rivas and Thomas 2005). Indeed, as far as we know, *C. fulvum*-Ecp genes have only been identified in wild accessions of tomato and have not yet been introduced into commercially grown tomatoes (Laugé et al. 1998, 2000). In addition, a relatively strong bias on preserving the sequence of the produced proteins was observed for the Ecp genes (Table 3), which supports the hypothesis that Ecp proteins might be essential for full virulence of the fungus. Indeed, a high in planta expression level of Ecp genes has been observed, suggesting that they play a role in virulence (Marmeisse et al. 1994; Wubben et al. 1994), while disruption mutants of Ecp1 and Ecp2 showed reduced mycelial growth and reduced conidiation on tomato seedlings (Laugé et al. 1997). Overall, the limited number of synonymous modifications in the Ecp genes suggests that most of the Ecp proteins are under purifying selection, although the generally low levels of polymorphisms in the Ecp genes could also be indicative of a recent common ancestor for most of the strains.

The high frequency (35% of the strains) of the one and only nonsynonymous modification observed in Ecp4 could be a result of directional positive selection or alternative mechanisms, such as random genetic drift, hitchhiking, or a bias in codon-usage (Fay and Wu 2001, 2003). However, as ten of the 29 strains carrying this substitution are fully virulent on sus-

ceptible tomato plants that lack Cf-Ecp4 (Laugé et al. 1998), the predicted amino-acid substitution in the Ecp4 protein does not affect virulence. Therefore, random genetic drift, which can strongly influence the frequency of neutral alleles in a population leading to either their gradual elimination or total fixation (Fay and Wu 2003; McDonald and Linde 2002), might account for the high frequency of the seemingly neutral Ecp4 allele in the collection.

Conclusion.

Overall, our analysis on allelic variation in the effector genes of *C. fulvum* shows that different selective forces have been and are still acting on these genes and that races of the fungus with complex virulence spectra could be the combined result of positive selection imposed on the Avr genes after introduction of *C. fulvum* genes into commercial tomato lines and genetic recombination. Future studies will focus on determining allelic variation in effector genes in freshly world-wide sampled strains isolated from both cultivated and wild tomatoes.

MATERIALS AND METHODS

Fungal material and culture conditions.

A total of 81 strains of *C. fulvum*, collected over a period of 50 years from different geographic regions around the world, were used in this study. The collection has been previously described by Stergiopoulos and associates (2007) and is stored at -80°C at the Laboratory of Phytopathology, Wageningen University, The Netherlands. Briefly, strains were collected from different geographic regions and were grouped according to the continent from which they were collected. The subcollection from Europe contained 49 strains originating from The Netherlands ($n = 22$), France ($n = 13$), Belgium ($n = 4$), Bulgaria ($n = 1$), the U. K. ($n = 5$), Italy ($n = 1$), and Poland ($n = 3$). The subcollection from the Americas contained 11 strains originating from Canada ($n = 7$), the U. S. A. ($n = 2$), Argentina ($n = 1$), and Brazil ($n = 1$). Additional subcollections included Japan ($n = 12$) and the substantially smaller collections from the former U. S. S. R. ($n = 2$), Turkey ($n = 4$), New Zealand ($n = 2$), and the African country of Zimbabwe ($n = 1$) (Fig. S1). Most strains were isolated from tomato plants grown in glasshouses, while a few were collected from outdoor-grown tomatoes. It should be taken into account that our collection, although derived from naturally occurring populations, was built by strains usually collected from cultivars that were previously resistant to these strains, therefore not representing a random population sampled from susceptible plants without *C. fulvum* genes. Thus, individuals with divergent phenotypes were explicitly selected for analysis. In addition, as the strains in this study were collected from distinct geographic areas and over a period of several years, the collection is not suited for proper population genetic analysis. Unfortunately, the tomato cultivars and their resistance genes from which the strains were collected are mostly not known. However, since we are interested in characterizing allelic variation in both Avr and Ecp effector genes, sampling from resistant host plants is not in that respect a drawback. Strains were cultured on half potato dextrose agar (19.5 g per liter, Agar Technical 15 g per liter) (Oxoid Ltd., Hampshire, England) at 22°C . Conidia were harvested from 15-day-old cultures and were freeze-dried prior to DNA extraction. Genomic DNA isolations were performed using the DNeasy plant mini kit (Qiagen Benelux bv, Venlo, The Netherlands) according to the manufacturer's instructions.

EcoTILLING assays.

Pre-amplification and formation of heteroduplexes. In all cases a race 0 strain (strain 1) was used as the wild-type refer-

ence strain to which all other strains were compared for the presence of allelic variation using the EcoTILLING technology (Comai et al. 2004; Henikoff et al. 2004). Gene-specific primers were designed based on the published sequences of *Avr2* (AJ421628), *Avr4* (Y08356), *Avr4E* (AY546101), *Avr9* (X60284), *Ecp1* (Z14023), *Ecp2* (Z14024), *Ecp4* (AJ271890), and *Ecp5* (AJ271891). All primers, except for the *Avr4*-specific primers, contained a M13-F or M13-R universal primer extension, which in some cases also served as a sequencing primer (Table 4).

Amplification of an *Avr* gene from an examined strain and the reference wild-type strain was performed in the same PCR reaction and was immediately followed by formation of the heteroduplexes between the two products generated. Both forward and reverse gene-specific primers were labeled with the fluorescent dyes IRD800 (795 nm) and IRD700 (685 nm), respectively. However, for the *Ecp* genes a two-step PCR approach was followed. In the first step, PCR assays were performed separately on the examined and the reference wild-type strain, using unlabeled *Ecp*-specific primers to which a M13-F or M13-R universal primer extension was attached. This preamplification step enables the fluorescent labeling of the two PCR products in a second nested PCR, using only one set of IRD800-M13F- and IRD700-M13R-labeled primers, thus reducing the running costs of this method. In all cases, PCR reaction mixes contained 2.5 µl of 10× PCR reaction buffer (Roche Diagnostics Nederland bv, Almere, The Netherlands), 1.8 µl of MgCl₂ (100 µM stock) (Roche Diagnostics Nederland bv), 0.6 µl of dNTPs (5 mM stock) (Promega Benelux bv, Leiden The Netherlands), 0.9 µl of each primer (10 µM stocks) (Biolegio bv, Nijmegen, The Netherlands), 0.075 µl of AmpliTaq DNA polymerase (5 units per µl) (Roche Diagnostics Nederland bv) and 1.2 µl of genomic DNA (25 ng per microliter of stocks) from the examined and the reference wild-type strain, respectively. The final volume of the reaction was adjusted to 15 µl with H₂O. In the case of the *Avr* genes, a 40-cycle PCR reaction consisting of 5 min of denaturation at 94°C, followed by an additional step of 30 s at the same temperature, 30 s of annealing at 50.4, 65, 53.5, or 62.5°C for *Avr2*, *Avr4*, *Avr4E*, and *Avr9*, respectively, and 1 min of extension at 72°C, was performed. A final extension step of 7 min at 72°C completed this part of the reaction, which then immediately entered into the heteroduplex formation program. This consisted of a denaturing step of 10 min at 95°C and linearly cooling to 4°C in the course of 10 min. The preamplification step employed in the case of the *Ecp* genes included a 25-cycle PCR reaction consisting of 5 min of denaturation at 94°C, fol-

lowed by an additional step of 30 s at the same temperature, 30 s of annealing at 54 or 49°C for *Ecp1* and *Ecp2* and *Ecp4* and *Ecp5*, respectively, and 1 min of extension at 72°C. A final extension step of 7 min at 72°C completed the reaction. The second nested PCR using the fluorescent-labeled M13 primers was carried out by mixing 30 ng of the preamplified DNA from both the examined and the reference wild-type strain in a PCR reaction mix that also contained 2.5 µl of 10× SuperTaq PCR reaction buffer (Roche Diagnostics Nederland bv), 1.0 µl of dNTPs (5 mM stock) (Promega Benelux bv), 1.0 µl of each of the IRD-labeled M13 primers (10 µM stocks) (Biolegio bv), and 0.1 µl of SuperTaq DNA polymerase (5 units per µl) (Roche Diagnostics Nederland bv). A PCR program similar to the one applied for the *Avr*-amplification and heteroduplex formation was then followed using an annealing temperature of 49°C. Finally, amplification products and heteroduplexes were visualized on a 1% agarose gel for quality and quantity control before proceeding to the enzymatic digestion of the heteroduplexes.

DNA-mismatch cleavage assay and analysis of products. Identification of mismatches between the examined and the reference wild-type strain was performed by the enzymatic cleavage of the produced heteroduplexes using the Surveyor nuclease provided by Transgenomic (Transgenomic Ltd, Cheshire, U.K.), according to the manufacturer's instructions. The Surveyor nuclease is a member of the CEL nuclease family of mismatch-specific nucleases derived from celery, which recognizes all insertions, deletions, and base substitutions, cleaving mismatches on the 3' side of the distortion (Qiu et al. 2004). Enzyme concentrations in these assays were adjusted according to the amount of DNA template digested in the reaction. Following treatment with the Surveyor nuclease, cleavage products were purified from the rest of the reagents using the GFX PCR DNA and gel band purification kit (Amersham Biosciences, Buckinghamshire, U.K.) before further analysis. This step significantly reduces the background noise on the images generated by the LI-COR system. The cleaved fragments were size-separated and analyzed by fluorescent thin-gel electrophoresis on a LI-COR 4300 DNA analysis system (LI-COR Biosciences, Lincoln, NB, U.S.A.), essentially following the method of Myburg and associates (2001).

Sequence analysis. Potential polymorphisms identified by EcoTILLING were confirmed by direct sequencing of the PCR products. All PCR products used for sequencing were generated from the examined and the reference wild-type strain using unlabeled analogs of the gene-specific primers. In order to evaluate the validity of the EcoTILLING method, a number of

Table 4. EcoTILLING and sequencing primers used in this study

Gene	Gene-specific primers (5'-3') used in EcoTILLING	Sequencing primers (5'-3')
<i>Avr2</i>	F: M13-F ^a + CATCAGCATATCCTCTTCCATCC R: M13-R ^b + CAGTACGTTCAAAAGCAGATAAGG	GGATAACAATTTACACAGG
<i>Avr4</i>	F: ACGGTAGGTCTGTACACGAGCC R: AACGAACTGGGTCATGGAATG	TCAGGAAGAGATTTCAGGAC
<i>Avr4E</i>	F: M13-F + GCCCGGTATATCGCTGTGC R: M13-R + CGGAACCCCTGGCTGAGA	GCCCGGTATATCGCTGTGC
<i>Avr9</i>	F: M13-F + AATACAACCTTGAAACAGCTAGG R: M13-R + GGACTCTACGGGGCTTGG	GGACTCTACGGGGCTTGG
<i>Ecp1</i>	F: M13-F + ACCAGGATAAGTTTCGCAGTC R: M13-R + GCCTCCGCGATGCTTCCT	TACAATGAGTTGCGGAAC
<i>Ecp2</i>	F: M13-F + GTCCAAGGATCGTGTCTCAAG R: M13-R + TTCTAGCAAACCCGTCTGA	ACCAGAAATGACCTCGAG
<i>Ecp4</i>	F: M13-F + AACCTAACATCACAATCTTCAA R: M13-R + TTGCTCAAGCCAGACAG	GGATAACAATTTACACAGG
<i>Ecp5</i>	F: M13-F + TATGAACACTTTTACTCTACTC R: M13-R + CGCGTCGCCTGATAGAT	TTTCCCAGTCACGACGTTG GGATAACAATTTACACAGG

^a M13-F: TTTCCCAGTCACGACGTTG.

^b M13-R: GGATAACAATTTCA.

PCR amplicons from the examined strains that did not show any polymorphisms compared with the reference strain for a particular gene were also sequenced. In that respect, we sequenced all the *Avr* genes from all 81 strains in our collection and 42, 46, 49, and 19 strains for *Ecp1*, *Ecp2*, *Ecp4*, and *Ecp5*, respectively. Indeed, based on the sequencing results, EcoTILLING proved a reliable method for the detection of allelic variation in the effector genes of *C. fulvum*, and the overall success of the method ranged between 84 to 100% (data not shown). Sequencing was performed at MACROGEN Inc. (Seoul, South Korea) directly on the purified PCR products and the generated chromatographs were analyzed using Vector NTI Suite 8 (InforMax Inc, Oxford).

Data analysis.

Polymorphisms were identified based on nucleotide sequence alignments performed using MEGA 3.1 (Kumar et al. 2004). The description of the polymorphisms was based on the nomenclature as recommended by the Human Genome Variation Society (den Dunnen and Antonarakis 2000). Standard molecular diversity indices, such as the average number of nucleotide differences between two random sequences (π ; Nei 1987) and the statistic of the number of segregating sites (θ_w ; Watterson 1975), were calculated using ARLEQUIN 3.01 (Excoffier et al. 2005).

We attempted to study the evolutionary forces acting on the individual effectors using commonly applied statistical tests, based on comparisons of nonsynonymous to synonymous substitutions rates (Nei and Gojobori 1986) and LRT. DNAsp 4.10.8 (Rozas et al. 2003) was used to calculate the ratio between the number of nonsynonymous substitutions per nonsynonymous site (K_a) and the number of synonymous substitutions per synonymous site (K_s), as estimated for only the protein-coding sequences of each individual effector gene. Indels as well as other large deletions and insertions in the sequence of the genes were excluded from the analysis. DNAsp uses the method of Nei and Gojobori (1986) to identify synonymous and nonsynonymous substitutions and a Jukes and Cantor correction for multiple hits (Jukes and Cantor 1969). In a second approach, we used a maximum likelihood method based on codon-substitution models to identify adaptive evolution (Nielsen and Yang 1998; Yang et al. 2000). Several likelihood models that allow for variation in the selective pressure among sites are implemented in the CODEML program of the PAML package v.3.15 (Yang 1997). Five site models [M0 [one ratio], M1a [nearly neutral], M2a [positive selection], M7 [beta], and M8 [beta& ω] that allow for variable ω (equivalent to nonsynonymous [d_N]/synonymous [d_S] substitution ratios) among codon sites were compared in a LRT of neutrality against positive selection (Anisimova et al. 2001; Yang et al. 2000). However, the low frequency of individual polymorphisms present in the *Avr* genes, the presence of indels, and the total absence, in most cases, of synonymous substitutions present in protein-coding sequences of the genes did not permit a meaningful statistical analysis, and therefore, data are not presented.

DNAsp was also used to examine several parameters related to codon usage bias, such as the effective number of codons (ENC) (Wright 1990), the codon bias index (CBI) (Morton 1993), and the G+C content at synonymous (third) codon positions (GC_{syn}) that are presented in Table 3. ENC and GC_{syn} are known to be negatively and positively correlated with codon bias, respectively, while CBI is a measure of deviation from the equal use of codons (Morton 1993) ranging in values from 1 (maximum codon bias) to 0 (equal use of codons). Strains that showed the complete deletion of one of the *Avr* genes or the presence of other large inserts or deletions were excluded from all the analyses described above, as their presence in the

sequence alignments results in an underestimation of the number of segregating sites (S). Instead, the complete set of 81 strains was used for haplotype inferences, while indices of haplotypic diversity (Hd) were calculated using Nei's (1987) diversity index corrected for sample size. Haplotypes were inferred based on the allelic variation observed in i) individual *Avr* and *Ecp* effector genes and ii) the combined *Avr* and *Ecp* effector genes. The NETWORK 4.2.0.1 software package was used to construct a minimal mutation network, which reflects the mutational relationships among the haplotypes by means of the median-joining algorithm (Bandelt et al. 1999).

Finally, DNAsp was also used for calculation of the minimum R_m number (Hudson and Kaplan 1985) in the history of the collection of strains, based on the pattern of the polymorphisms present. The method utilizes a four-gamete test for the estimation of R_m . For any pair of nucleotide sites, only three combinations are possible (in binary mode: 00, 01, 10). Therefore, producing all four gamete combinations requires either recombination to take place or, alternatively, it implies a second mutation in one of the nucleotide sites (Hudson and Kaplan 1985).

ACKNOWLEDGMENTS

We kindly acknowledge R. Dupont for technical assistance with the EcoTILLING assays. We are thankful to M. H. A. J. Joosten, B. P. H. J. Thomma, F. T. Bakker, and R. Jiang for critical reading of the manuscript and useful suggestions. I. Stergiopoulos was financially supported by a Research Training Network (RTN) Grant of the European Commission (ACE; Contact No. HPRN-CT-2002-00249). Since 2007, I. Stergiopoulos was supported by an international ERA-Plant Genomics network, which is financed by the Dutch Science Foundation for Earth and Life Sciences.

LITERATURE CITED

- Allen, R. L., Bittner-Eddy, P. D., Grenville-Briggs, L. J., Meitz, J. C., Rehmany, A. P., Rose, L. E., and Beynon, J. L. 2004. Host-parasite co-evolutionary conflict between *Arabidopsis* and downy mildew. *Science* 306:1957-1960.
- Anisimova, M., Bielawski, J. P., and Yang, Z. 2001. The accuracy and power of likelihood ratio tests to detect positive selection at amino acid sites. *Mol. Biol. Evol.* 18:1585-1592.
- Anisimova, M., Bielawski, J. P., and Yang, Z. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Mol. Biol. Evol.* 19:950-958.
- Arnold, D. L., Jackson, R. W., Waterfield, N. R., and Mansfield, J. W. 2007. Evolution of microbial virulence: The benefits of stress. *Trends Genet.* 23:293-300.
- Bailey, D. L. 1950. Studies in racial trends and constancy in *Cladosporium fulvum*. *Can. J. Res.* 28:535.
- Bandelt, H.-J., Forster, P., and Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16:37-48.
- Boukema, I. W. 1981. Races of *Cladosporium fulvum* Cooke. (*Fulvia fulva*) and genes for resistance in the tomato (*Lycopersicon* Mill.). Pages 287-292 in: Genetics and breeding of tomato: Proceedings of the meeting of the Eucarpia Tomato Working Group, Avignon, France.
- Braun, U., Crous, P. W., Dugan, F., Groenewald, J., and Sybren De Hoog, G. 2003. Phylogeny and taxonomy of *Cladosporium*-like hyphomycetes, including *Davidiella*, the teleomorph of *Cladosporium*. *Mycol. Prog.* 2:3-18.
- Butler, E. J., and Jones, S. G. 1949. Tomato leaf mold *Cladosporium fulvum* (Cooke). Page 671 in: Plant Pathology. E. J. Butler and S. G. Jones, eds. McMillen and Co. Ltd. London.
- Comai, L., Young, K., Till, B. J., Reynolds, S. H., Greene, E. A., Codomo, C. A., Enns, L. C., Johnson, J. E., Burtner, C., Odden, A. R., and Henikoff, S. 2004. Efficient discovery of DNA polymorphisms in natural populations by ecotilling. *Plant J.* 37:778-786.
- Couch, B. C., Fudal, I., Lebrun, M.-H., Tharreau, D., Valent, B., van Kim, P., Nottéghem, J.-L., and Kohn, L. M. 2005. Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics* 170:613-630.
- Day, P. R. 1957. Mutation to virulence in *Cladosporium fulvum*. *Nature* 179:1141.

- den Dunnen, J. T., and Antonarakis, S. E. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. *Hum. Mutat.* 15:7-12.
- De Wit, P. J. G. M. 1992. Molecular characterization of gene-for-gene systems in plant-fungus interactions and the application of avirulence genes in control of plant-pathogens. *Annu. Rev. Phytopathol.* 30:391-418.
- De Wit, P. J. G. M. 2002. Plant biology—On guard. *Nature* 416:801-803.
- Dodds, P. N., Lawrence, G. J., Catanzariti, A.-M., Theh, T., Wang, C.-I. A., Ayliffe, M. A., Kobe, B., and Ellis, J. G. 2006. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc. Natl. Acad. Sci. U.S.A.* 103:8888-8893.
- Excoffier, L., Laval, G., and Schneider, S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1:47-50.
- Fay, J. C., and Wu, C. I. 2001. The neutral theory in the genomic era. *Curr. Opin. Genet. Dev.* 11:642-646.
- Fay, J. C., and Wu, C. I. 2003. Sequence divergence, functional constraint, and selection in protein evolution. *Annu. Rev. Genomics Hum. Genet.* 4:213-235.
- Gout, L., Fudal, I., Kuhn, M. L., Blaise, F., Eckert, M., Cattolico, L., Balesdent, M. H., and Rouxel, T. 2006. Lost in the middle of nowhere: The *AvrLm1* avirulence gene of the Dothideomycete *Leptosphaeria maculans*. *Mol. Microbiol.* 60:67-80.
- Henikoff, S., Till, B. J., and Comai, L. 2004. TILLING. Traditional mutagenesis meets functional genomics. *Plant Physiol.* 135:630-636.
- Higgins, V. J., and Hollands, J. 1987. Prevalent races of *Cladosporium fulvum* in southern Ontario and their benomyl sensitivity. *Can. J. Plant Pathol.* 9:32-35.
- Hudson, R. R., and Kaplan, N. L. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147-164.
- Jones, J. D. G., and Dangl, J. L. 2006. The plant immune system. *Nature* 444:323-329.
- Joosten, M. H. A. J., and De Wit, P. J. G. M. 1999. The tomato-*Cladosporium fulvum* interaction: A versatile experimental system to study plant-pathogen interactions. *Annu. Rev. Phytopathol.* 37:335-367.
- Joosten, M. H. A. J., Cozijnsen, T. J., and 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., and De Wit, P. J. G. M. 1997. The biotrophic fungus *Cladosporium fulvum* circumvents *Cf-4*-mediated resistance by producing unstable AVR4 elicitors. *Plant Cell* 9:367-379.
- Jukes, T., and Cantor, C. 1969. Evolution of protein molecules. In: *Mammalian Protein Metabolism*. Academic Press, New York.
- Kishi, K., and Abiko, K. 1976. Studies on the physiological specialization of *Cladosporium fulvum* Cooke. Pathogenicity to Japanese resistant varieties of tomato and racial identification of isolates collected from 11 prefectures in Japan from 1971 to 1973. *Ann. Phytopathol. Soc. Jap.* 42:497-499.
- Krüger, J., Thomas, C. M., Golstein, C., Dixon, M. S., Smoker, M., Tang, S. K., Mulder, L., and Jones, J. D. G. 2002. A tomato cysteine protease required for Cf-2-dependent disease resistance and suppression of auto-necrosis. *Science* 296:744-747.
- Kumar, S., Tamura, K., and Nei, M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* 5:150-163.
- Langford, A. N. 1937. The parasitism of *Cladosporium fulvum* and the genetics of resistance to it. *Can. J. Res.* 15:35.
- Laterrot, H. 1986. Race 2.5.9, a new race of *Cladosporium fulvum* (*Fulvia fulva*) and sources of resistance in tomato. *Neth. J. Plant Pathol.* 92:305-307.
- Laugé, R., Joosten, M. H. A. J., Van den Ackerveken, G. F. J. M., Van den Broek, H. W. J., and De Wit, P. J. G. M. 1997. The in planta-produced extracellular proteins ECP1 and ECP2 of *Cladosporium fulvum* are virulence factors. *Mol. Plant-Microbe Interact.* 10:725-734.
- Laugé, R., Joosten, M. H. A. J., Haanstra, J. P. W., Goodwin, P. H., Lindhout, P., and De Wit, P. J. G. M. 1998. Successful search for a resistance gene in tomato targeted against a virulence factor of a fungal pathogen. *Proc. Natl. Acad. Sci. U.S.A.* 95:9014-9018.
- Laugé, R., Goodwin, P. H., De Wit, P. J. G. M., and Joosten, M. H. A. J. 2000. Specific HR-associated recognition of secreted proteins from *Cladosporium fulvum* occurs in both host and non-host plants. *Plant J.* 23:735-745.
- Leski, B. 1970. Physiological specialization of the fungus *Cladosporium fulvum* Cooke on tomatoes in Poland. *Acta Agrobot.* 23:133-149.
- Lindhout, P., Korta, W., Cislík, M., Vos, I., and Gerlagh, T. 1989. Further Identification of races of *Cladosporium fulvum* syn. *Fulvia fulva* on tomato originating from the Netherlands, France and Poland. *Neth. J. Plant Pathol.* 95:143-148.
- Luderer, R., Takken, F. L. W., De Wit, P. J. G. M., and Joosten, M. H. A. J. 2002. *Cladosporium fulvum* overcomes Cf-2-mediated resistance by producing truncated AVR2 elicitor proteins. *Mol. Microbiol.* 45:875-884.
- Marmeisse, R., Van den Ackerveken, G. F. J. M., Goosen, T., De Wit, P. J. G. M., and Van den Broek, H. W. J. 1994. The *in-planta* induced *Ecp2* gene of the tomato pathogen *Cladosporium fulvum* is not essential for pathogenicity. *Curr. Genet.* 26:245-250.
- McDonald, B. A., and Linde, C. 2002. Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* 40:349-379.
- Morton, B. R. 1993. Chloroplast DNA Codon Use—Evidence for selection at the *Psb-a* locus based on transfer-RNA availability. *J. Mol. Evol.* 37:273-280.
- Myburg, A. A., Remington, D. L., O'Malley, D. M., Sederoff, R. R., and Whetten, R. W. 2001. High-throughput AFLP analysis using infrared dye-labeled primers and an automated DNA sequencer. *Biotechniques* 30:348-357.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nei, M., and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3:418-426.
- Nielsen, R., and Yang, Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929-936.
- Olson, M. V. 1999. When less is more: Gene loss as an evolutionary change. *A. J. Hum. Genet.* 64:18-23.
- Orbach, M. J., Farall, L., Sweigard, J. A., Chumley, F. G., and Valent, B. 2000. A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta*. *Plant Cell* 12:2019-2032.
- Qiu, P., Shandilya, H., D'Alessio, J. M., O'Connor, K., Durocher, J., and Gerard, G. F. 2004. Mutation detection using Surveyor nuclease. *Biotechniques* 36:702-707.
- Rehmany, A. P., Gordon, A., Rose, L. E., Allen, R. L., Armstrong, M. R., Whisson, S. C., Kamoun, S., Tyler, B. M., Birch, P. R. J., and Beynon, J. L. 2005. Differential recognition of highly divergent downy mildew avirulence gene alleles by *RPPI* resistance genes from two Arabidopsis lines. *Plant Cell* 17:1839-1850.
- Rivas, S., and Thomas, C. M. 2005. Molecular interactions between tomato and the leaf mold pathogen *Cladosporium fulvum*. *Annu. Rev. Phytopathol.* 43:395-436.
- Rohmer, L., Guttman, D. S., Dangl, J. L. 2004. Diverse evolutionary mechanisms shape the type III effector virulence factor repertoire in the plant pathogen *Pseudomonas syringae*. *Genetics* 167:1341-1360.
- Rooney, H. C. E., van't Klooster, J. W., van der Hoorn, R. A. L., Joosten, M. H. A. J., Jones, J. D. G., and De Wit, P. J. G. M. 2005. *Cladosporium Avr2* inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. *Science* 308:1783-1786.
- Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X., and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496-2497.
- Schurch, S., Linde, C. C., Knogge, W., Jackson, L. F., and McDonald, B. A. 2004. Molecular population genetic analysis differentiates two virulence mechanisms of the fungal avirulence gene *NIP1*. *Mol. Plant-Microbe Interact.* 17:1114-1125.
- Stergiopoulos, I., Groenewald, M., Staats, M., Lindhout, P., Crous, P. W., and De Wit, P. J. G. M. 2007. Mating-type genes and the genetic structure of a world-wide collection of the tomato pathogen *Cladosporium fulvum*. *Fungal Genet. Biol.* 44:415-429.
- Thomma, B. P. H. J., Van Esse, H. P., Crous, P. W., and De Wit, P. J. G. M. 2005. *Cladosporium fulvum* (syn. *Passalora fulva*), a highly specialized plant pathogen as a model for functional studies on plant pathogenic Mycosphaerellaceae. *Mol. Plant Pathol.* 6:379-393.
- Van den Ackerveken, G. F. J. M., Van Kan, J. A. L., Joosten, M. H. A. J., Muijsers, J. M., Verbakel, H. M., and De Wit, P. J. G. M. 1993a. Characterization of two putative pathogenicity genes of the fungal tomato pathogen *Cladosporium fulvum*. *Mol. Plant-Microbe Interact.* 6:210-215.
- Van den Ackerveken, G. F. J. M., Vossen, P., and De Wit, P. J. G. M. 1993b. The Avr9 race-specific elicitor of *Cladosporium fulvum* is processed by endogenous and plant proteases. *Plant Physiol.* 103:91-96.
- Van den Burg, H. A., Westerink, N., Francoijs, K. J., Roth, R., Woestenenk, E., Boeren, S., De Wit, P. J. G. M., Joosten, M. H. A. J., and Vervoort, J. 2003. Natural disulfide bond-disrupted mutants of AVR4 of the tomato pathogen *Cladosporium fulvum* are sensitive to proteolysis, circumvent *Cf-4*-mediated resistance, but retain their chitin binding ability. *J. Biol. Chem.* 278:27340-27346.

- van den Burg, H. A., Harrison, S. J., Joosten, M. H. A. J., Vervoort, J., and de Wit, P. J. G. M. 2006. *Cladosporium fulvum* Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. *Mol. Plant-Microbe Interact.* 19:1420-1430.
- van Esse, H. P., Bolton, M. D., Stergiopoulos, I., De Wit, P. J. G. M., and Thomma B. P. H. J. 2007. The chitin-binding *Cladosporium fulvum* effector protein Avr4 is a virulence factor. *Mol. Plant-Microbe Interact.* 20:1092-1101.
- Van Kan, J. A. L., Van den Ackerveken, G. F. J. M., and De Wit, P. J. G. M. 1991. Cloning and characterization of cDNA of avirulence gene *Avr9* of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. *Mol. Plant-Microbe Interact.* 4:52-59.
- Watterson, G. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7:256-276.
- Westerink, N., Joosten, M. H. A. J., and De Wit, P. J. G. M. 2004a. Fungal (a)virulence factors at the crossroads of disease susceptibility and resistance. Pages 93-137 in: *Fungal Disease Resistance in Plants: Biochemistry, Molecular Biology, and Genetic Engineering*. Z. K. Punja, ed. Food Products Press, Binghamton, U.S.A.
- Westerink, N., Brandwagt, B. F., De Wit, P. J. G. M., and Joosten, M. H. A. J. 2004b. *Cladosporium fulvum* circumvents the second functional resistance gene homologue at the *Cf-4* locus (*Hcr9-4E*) by secretion of a stable *avr4E* isoform. *Mol. Microbiol.* 54:533-545.
- Wright, F. 1990. The 'effective number of codons' used in a gene. *Gene* 87:23-29.
- Wubben, J. P., Joosten, M. H. A. J., and De Wit, P. J. G. M. 1994. Expression and localization of two *in planta* induced Extracellular proteins of the fungal tomato pathogen *Cladosporium fulvum*. *Mol. Plant-Microbe Interact.* 7:516-524.
- Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13:555-556.
- Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A. M. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431-449.
- Yang Z, Wong, W. S., and Nielsen, R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22:1107-1118.

AUTHOR-RECOMMENDED INTERNET RESOURCES

- ARLEQUIN software: lgb.unige.ch/arlequin/
 DNA Sequence Polymorphism (CnaSP) website: www.ub.es/dnasp/
 Fluxus-engineering NETWORK 4.2.0.1 software: www.fluxus-technology.com/
 Human Genome Variation Society website: www.hgvs.org/mutnomen/
 Molecular Evolutionary Genetics Analysis MEGA webpage: www.megasoftware.net/
 Phylogenetic analysis by maximum likelihood (PAML) website: abacus.gene.ucl.ac.uk/software/paml.html