

# Diverse Evolutionary Mechanisms Shape the Type III Effector Virulence Factor Repertoire in the Plant Pathogen *Pseudomonas syringae*

Laurence Rohmer,<sup>\*,1</sup> David S. Guttman<sup>†</sup> and Jeffery L. Dangl<sup>\*,‡,2</sup>

<sup>\*</sup>Department of Biology and <sup>†</sup>Curriculum in Genetics, Department of Microbiology and Immunology and Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, North Carolina 27599 and <sup>‡</sup>Department of Botany, University of Toronto, Toronto, Ontario M5S 3B2, Canada

Manuscript received July 4, 2003

Accepted for publication March 26, 2004

## ABSTRACT

Many gram-negative pathogenic bacteria directly translocate effector proteins into eukaryotic host cells via type III delivery systems. Type III effector proteins are determinants of virulence on susceptible plant hosts; they are also the proteins that trigger specific disease resistance in resistant plant hosts. Evolution of type III effectors is dominated by competing forces: the likely requirement for conservation of virulence function, the avoidance of host defenses, and possible adaptation to new hosts. To understand the evolutionary history of type III effectors in *Pseudomonas syringae*, we searched for homologs to 44 known or candidate *P. syringae* type III effectors and two effector chaperones. We examined 24 gene families for distribution among bacterial species, amino acid sequence diversity, and features indicative of horizontal transfer. We assessed the role of diversifying and purifying selection in the evolution of these gene families. While some *P. syringae* type III effectors were acquired recently, others have evolved predominantly by descent. The majority of codons in most of these genes were subjected to purifying selection, suggesting selective pressure to maintain presumed virulence function. However, members of 7 families had domains subject to diversifying selection.

*PSEUDOMONAS syringae* is associated with numerous important plant diseases including bacterial speck of tomato and halo blight of beans. The species is subdivided into ~50 pathogenic varieties [pathovars (pv.)] on the basis of the original plant host of isolation (RUDOLPH 1995). The evolutionarily conserved type III secretion system was acquired by *P. syringae* prior to pathovar differentiation and represents a common strategy of infection for all the bacterial species that utilize it (SAWADA *et al.* 1999; FOUTS *et al.* 2003; JIN *et al.* 2003). The type III pilus is a conduit for delivery of type III effector proteins into the plant intercellular space (the apoplast; HE *et al.* 1993; ALFANO and COLLMER 1997; JIN and HE 2001). Inactivation of the type III secretion system in *P. syringae* results in a total loss of pathogenesis (LINDGREN *et al.* 1986, 1988). This indicates that the proteins secreted by the system are required for bacterial virulence.

Recent attempts to identify type III effector genes in *P. syringae* used newly available genome sequences (*P. syringae* pv. *tomato* DC3000, *P. syringae* pv. *syringae* B728a, and *P. syringae* pv. *phaseolicola* race 6), *in vivo* and *in vitro* expression assays, and secretion assays (BOCH *et al.* 2002; GUTTMAN *et al.* 2002; PETNICKI-OCWIEJA *et al.*

2002; ZWIESLER-VOLLIK *et al.* 2002; FOUTS *et al.* 2003). According to these studies, *P. syringae* pv. *tomato* DC3000 (*Pst* DC3000) may secrete as many as 50 effectors through its type III secretion system (for review, see COLLMER *et al.* 2002; GREENBERG and VINATZER 2003). Homologs of many of these type III effectors are distributed across *P. syringae* pathovars, suggesting that divergent strains carry at least an overlapping set of type III effectors that might provide a common core of virulence functions. Some *P. syringae* type III effectors are also shared with other phytopathogenic species (COLLMER *et al.* 2002; GUTTMAN *et al.* 2002; GREENBERG and VINATZER 2003), which may indicate similar infection strategies. Many type III effector genes are located in pathogenicity islands or are associated with remnants of mobile elements (KIM *et al.* 1998). Their distribution among strains of *P. syringae* is highly variable (GUTTMAN *et al.* 2002; D. S. GUTTMAN, unpublished data). These data support an important role for horizontal gene transfer in the evolution of type III effectors and pathogenesis.

A limited number of type III effectors from plant pathogenic bacteria have been assigned proven or probable biochemical functions (NIMCHUK *et al.* 2001; COLLMER *et al.* 2002; CHANG *et al.* 2004). For example, AvrBs2 is similar to phosphodiesterases, while AvrPpiG1, AvrRxv, AvrBst, and AvrXv4 and AvrRpt2 share key residues with cysteine proteases (ORTH 2002). Functions inside the host cell of some plant pathogen type III effectors have been elucidated. AvrBs3 family members alter plant gene expression (YANG *et al.* 2000a; SZUREK *et al.* 2001; MAROIS

<sup>1</sup>Present address: Medical Genetics, University of Washington, Seattle, WA 98195.

<sup>2</sup>Corresponding author: Department of Biology, Coker Hall, Room 108, University of North Carolina, Chapel Hill, NC 27599.  
E-mail: dangl@email.unc.edu

*et al.* 2002). AvrRpm1 (*P. syringae* pv. *maculicola*), AvrB (*P. syringae* pv. *glycinea*), and AvrRpt2 (*P. syringae* pv. *tomato*) all target the Arabidopsis RIN4 protein (MACKEY *et al.* 2002, 2003; AXTELL and STASKAWICZ 2003). In *Erwinia amylovora*, DspA, a member of the *P. syringae* AvrE family, triggers the release of reactive oxygen species by the host (pear) cell, which is necessary for the colonization by the pathogen (VENISSE *et al.* 2003).

To defend themselves against pathogen attack, plants have evolved a surveillance system to detect bacterial invasion. The surveillance proteins in plants are encoded by disease resistance (*R*) genes. These are believed to “guard” host targets against virulence proteins delivered into the plant cell by the bacteria (DANGL and JONES 2001; SCHNEIDER 2002). According to the guard hypothesis, the host defense response is triggered when the *R* protein detects the action of a virulence factor. In one example from tomato, the direct recognition of the *P. syringae* AvrPto virulence protein by the plant Pto resistance protein (SHAN *et al.* 2000) leads to subsequent signaling through the *R* protein Prf. In other cases, the virulence factor alters the activity of its host target, and this change triggers *R* protein action. The *R* protein RPS2, for example, triggers defense reactions in response to the disappearance of the AvrRtp2-mediated disappearance of RIN4 protein (AXTELL and STASKAWICZ 2003; MACKEY *et al.* 2003), and activation of the RPS5 *R* protein by AvrPphB requires cleavage of the host PBS1 kinase by that type III effector cysteine protease (SHAO *et al.* 2003).

There is a never-ending battle between bacterial virulence systems and plant surveillance and resistance systems. The pathogen is under strong selection to avoid or suppress recognition by the host. Virulence proteins that trigger the plant defense response will be strongly selected against and will therefore either be lost or diverge in sequence so that they are no longer recognized by their host. This is potentially problematic for the pathogen, if the biochemical function of the virulence factor required for its activity is the same as that sensed by the host to trigger a successful host defense response (SHAO *et al.* 2003). As a possible consequence, many virulence type III effectors may act to suppress the host defense response (STASKAWICZ *et al.* 2001; ORTH 2002). Plants can recognize conserved molecular features of plant pathogens, leading to induction of a basal defense response. The suite of type III effectors carried by a given bacterial pathogen acts, at least in part, to block or dampen this defense response (JAKOBEK *et al.* 1993; HAUCK *et al.* 2003). The type III effectors, AvrPphC and AvrPphF in *P. syringae* pv. *phaseolicola* (JACKSON *et al.* 1999; TSAMIS *et al.* 2000) and AvrPtoB in *P. syringae* pv. *tomato* (ABRAMOVITCH *et al.* 2003), are believed to play this role. Some evidence suggests that these virulence-associated proteins have been acquired through horizontal transfer (JACKSON *et al.* 1999, 2002). Thus, to counteract the plant surveillance system without sacrificing a poten-

tially useful virulence strategy, these bacteria appear to have acquired new virulence factors that inhibit or delay host defense response. In reaction, some genotypes of the host have evolved *R* genes that in fact can detect these “defense inhibiting” virulence factors, thus triggering the defense response. It is likely that host interactions with pathogens drive the diversification of plant *R* genes, to detect the action of new or divergent virulence factors on their nominal host targets (OHTA 1991; APANIUS *et al.* 1997; MICHELMORE and MEYERS 1998; YEAGER and HUGHES 1999; DANGL and JONES 2001).

The number of available sequences for confirmed or suspected type III effectors from *P. syringae* has increased dramatically. These data provided us with the opportunity to investigate the evolutionary history of type III effectors in *P. syringae* and the role played in this evolution by interaction with the host. Using 46 type III effector proteins as query sequences, we identified 24 families of type III effector genes on the basis of similarity. We analyzed the genes from these families for their distribution and relatedness and for features indicating horizontal transfer. We also assessed the role played by diversifying and purifying selection in the evolution of these type III effector gene families.

## MATERIALS AND METHODS

**Database mining:** The National Center for Biotechnology Information (NCBI) databases were explored using BLASTP (ALTSCHUL *et al.* 1990). Additional sequences were retrieved from the unfinished genome of *P. syringae* pv. *syringae* B728a, sequenced at a coverage of approximately eight times ([http://www.jgi.doe.gov/JGI\\_microbial/html/pseudomonas\\_syr/pseudo\\_syr\\_homepage.html](http://www.jgi.doe.gov/JGI_microbial/html/pseudomonas_syr/pseudo_syr_homepage.html)), and from the then-unfinished genome of *P. syringae* pv. *tomato* DC3000 (<http://www.tigr.org/tdb/mdb/mdbinprogress.html>). The maximum threshold of BLASTP expected (*E*) values was 0.005. For phylogenetic analyses, the similarities of the sequence groups were examined by global alignment (NEEDLEMAN and WUNSCH 1970), using the program “Needle” from the EMBOSS package. Homologous sequences were selected if they had a similarity (bit) score larger than a quarter of the score resulting from the comparison of the type III effector sequence to itself (ENDO *et al.* 1996).

**Phylogenetic analyses:** Protein alignments were generated using ClustalW (THOMPSON *et al.* 1994). The alignments were used to generate phylogenetic neighbor-joining trees using the Poisson correction for multiple substitution events with MEGA v2.0 (KUMAR *et al.* 2001). Bootstrap confidence levels were determined by randomly resampling of the sequence data 1000 times. Phylogenetic analyses were also performed using PHYLIP. The program Prodist was used to calculate protein distances, using the Dayhoff PAM substitution model, and neighbor-joining trees were constructed with the program Neighbor. DNA sequences were aligned using DIALIGN2 (MORGENSTERN 1999). Some sequences were trimmed at the 5′ or the 3′ end when the alignments in these regions were not reliable. The distance between sequences was assessed with the software Dnadist of the Phylip package based on the Kimura two-parameter model (KIMURA 1980). Neighbor-joining phylogenetic trees were then built with the software Neighbor, on the basis of these distances.

**GC content and codon usage:** The overall CG content and

the GC content at the third codon position (GC3) of type III effector genes was analyzed using a custom PERL script. Additionally, the open reading frames (ORFs) of available genomes were retrieved from TIGR ([http://www.tigr.org/tigr-scripts/CMR2/batch\\_download.dbi](http://www.tigr.org/tigr-scripts/CMR2/batch_download.dbi)) and were used to calculate the GC content and the GC3 content of the ORFs in these genomes. These organisms are: *Xanthomonas campestris* pv. *campestris*, *X. axonopodis* pv. *citri*, *Ralstonia solanacearum*, *Bacillus subtilis*, *P. syringae* pv. *tomato* DC3000, and *Streptomyces coelicolor*. Since the GC content of closely related organisms is very similar, we used the value for *P. syringae* pv. *tomato* DC3000 (58.4% G + C) to represent all *P. syringae* strains (LAWRENCE and OCHMAN 1997; note that the GC content of *P. syringae* pv. *syringae* B782a is 59.2%).

To assess whether type III effector gene GC and GC3 contents match the GC and GC3 content of their respective genomes, we compared the content of type III effector genes to the average content of the genome's ORFs. We excluded from this analysis disrupted ORFs. The cutoff (two standard deviations) is based on the assumption that the distribution of the ORF GC and GC3 content in a genome is normal. This assumption is supported by the result of normal quantile plots: most of the GC and GC3 values for each genome lie close to a straight line, indicating that a normal model fits well. According to the standard normal distribution, if the GC content of a gene is below or above this cutoff (twofold the standard deviation above or below the mean), this value is different from the rest of the genome with a probability of 97.7%. The overall GC content of *E. carotovora* pv. *atroseptica* (~50%) is from <http://bitrws400.scri.sari.ac.uk/TIPP/Erwinia.htm>, of *P. aeruginosa* (66.6%) is from STOVER *et al.* (2000), and of *P. fluorescens* (60.6%) is from [http://www.jgi.doe.gov/JGI\\_microbial/html/pseudomonas/pseudo\\_homepage.html](http://www.jgi.doe.gov/JGI_microbial/html/pseudomonas/pseudo_homepage.html).

Codon usage was assessed for organisms from which the ORFs were available using the program CodonW (downloaded from <http://www.molbiol.ox.ac.uk/cu/>) to calculate the codon adaptation index (CAI; SHARP and LI 1987). To calculate CAI for the ORFs of these organisms (including the type III effector homologs we analyzed), we used as a reference pool every ORF identified in the genomes, since it is impossible for us to determine which genes are highly expressed.

It was not possible to obtain the complete set of predicted ORF sequences for all relevant genomes and hence not possible to compare the average GC3, GC content, and CAI for these genomes to their respective type III effector genes. In these cases, we compared the overall GC value of the organism, as given by the sources in MATERIALS AND METHODS, to the overall GC content of the respective effector genes. The cutoff used in these cases is two standard deviations calculated for *Pst* DC3000.

**Positive selection assessment:** Nucleotide alignments were made with DIALIGN2 on the basis of the translation of nucleotide diagonals into peptide diagonals. The alignments and derived phylogenetic trees were used in the program CODEML from the PAML package (YANG 1997) to calculate the  $\omega$ -ratio (of nonsynonymous to synonymous changes;  $d_n/d_s$ ) for each site. Different evolutionary models were tested (YANG *et al.* 2000b): model M0 assumed a constant  $\omega$ -ratio; models M1 and M7 assumed that amino acid sites substitutions are either neutral ( $\omega \sim 1$ ) or conservative ( $\omega \sim 0$ ); and models M3 and M8 allow the occurrence of positively selected sites ( $\omega > 1$ ). M7 and M8 assume a  $\beta$ -distribution for the  $\omega$ -value between 0 and 1. For each codon, the probability of observing conservative, neutral, or positive selection was computed using the proportion of sites belonging to these categories. The log likelihood is the sum of these probabilities over all codons in the sequence. The likelihood ratio of two models is compared (M3 vs. M1 or M0 and M8 vs. M7) to test which model fits the

data significantly better: twice the difference in log likelihood between the two models is compared with a chi-square distribution with  $n$  d.f.,  $n$  being the difference between the numbers of parameters of the two models. The cutoff chosen was  $P = 0.1$ , which is acceptable because this likelihood ratio test is very conservative (ANISIMOVA *et al.* 2001). An empirical Bayesian approach implemented in CODEML was used to infer to which category (defined by a  $\omega$ -ratio estimated by the program) each amino acid most likely belongs.

## RESULTS

**Distribution of type III effector homologs among bacteria:** Forty-four known and predicted *P. syringae* type III effectors (see supplemental Table 1 at <http://www.genetics.org/supplemental/>) retrieved from the NCBI databases were used to search the same databases for homologs using the BlastP algorithm (ALTSCHUL *et al.* 1990). Two suspected chaperones (AvrF and AvrPphF-ORF1) were also considered, since they are associated in operons with specific type III effectors. Ancient homologs detected using BlastP (see supplemental Table 1 at <http://www.genetics.org/supplemental/>) were not considered later in this study, since phylogenetic analyses have little power when performed on highly divergent sequences (see MATERIALS AND METHODS). High rates of gene acquisition and loss and horizontal gene transfer make the discrimination between orthologs and paralogs extremely difficult in this data set. Technically, many of these sequences are more accurately described as xenologs, which is a homology relationship brought about by horizontal transfer. We largely avoid the issue by simply referring to all related sequences as homologs, without making further distinctions.

Eleven effectors had no homologs in the databases (see supplemental Table 1 at <http://www.genetics.org/supplemental/>). Eight of these (HolPtoS, HolPtoT, HolPtoU, HolPtoU2, HolPtoV, HolPtoY, HolPtoZ, and HopPtoB) are confirmed or predicted type III effectors in *P. syringae* pv. *tomato* DC3000 (*Pst* DC3000). Eleven other type III effectors had only one homolog; six of these were found in one of the two sequenced genomes, *Pst* DC3000 and *P. syringae* pv. *syringae* B728a (*Psy* B728a). The other five were found in only one of these two sequenced genomes, so they were either acquired through horizontal transfer in the respective strains or deleted from one strain following divergence. These 22 genes were not considered further in this study.

Twenty-four of the 46 type III effectors and chaperones originally selected had three or more homologs (see supplemental Table 1 at <http://www.genetics.org/supplemental/>). The number of protein sequences in the families ranged from 3 to 16. Seventeen of these 24 families contained a sequence from *Pst* DC3000, 14 from *Psy* B728a, and 9 from *P.s.* pv. *maculicola* ES4326 (*Pma* ES4326). In contrast, 4 of the families were absent from the two sequenced strains, including AvrA, isolated from *P. s.* pv. *glycinea*, and AvrPpiG1, isolated from *P. s.* pv.

*psi*, HopPmaB from *Pma* ES4326, and AvrD (present in many *P. syringae* strains). These four genes may have been acquired through horizontal transfer after strain divergence. BLAST results comparing the two sequenced *P. syringae* genomes indicated that some effectors are unique to one or the other genome. Homologs to AvrB and AvrPpiA1 were found only in the *Psy* B728a, whereas homologs of AvrPphD and AvrPpiB were found only in *Pst* DC3000. Additionally, homologs of the putative type III effectors HolPtoQ, HolPtoR, and HolPtoW identified in *Pst* DC3000 were not detected in *Psy* B728a.

Eighteen of the 24 type III gene families have homologs in distantly related phytopathogenic species. Twelve families contain one to four homologs from *Xanthomonas* species and 10 families contain one homolog from *R. solanacearum*. In addition, 6 of 24 effector families have homologs found in nonphytopathogenic species such as *P. aeruginosa* or *P. fluorescens* (see supplemental Table 1 at <http://www.genetics.org/supplemental/>). This suggests that at least a set of *P. syringae* type III effector genes that occur throughout the species is also dispersed broadly beyond it (see supplemental Table 2 at <http://www.genetics.org/supplemental/>).

**Phylogenetic relationships among homologs from type III effector gene families:** We used the amino acid sequences for phylogenetic analyses of the 24 type III effector families that contained three or more homologs. These sequences were aligned using ClustalW (see MATERIALS AND METHODS; THOMPSON *et al.* 1994) and analyzed by neighbor-joining trees using 1000 bootstrapping replicates (Figure 1). The topologies of the type III effector trees were compared to the topologies of 16S rDNA species trees or intraspecific data from multiple housekeeping genes (EISEN 1995; SAWADA *et al.* 1999; DALE *et al.* 2002; D. S. GUTTMAN, unpublished data). Topological comparisons were based on the presence of common, well-supported clades (bootstrap scores >70). Topologies of only 3 of 24 trees (AvrB, AvrPpiG, and HopPmaG) were clearly incongruent with the topologies of species or subspecies trees (Figure 1). Interestingly, AvrB and AvrPpiG are both encoded by genes that appear to be frequently transferred horizontally, on the basis of GC composition and the presence of mobile elements (Table 1). The incongruence of the tree supports the idea of a recent acquisition of these two genes (DAUBIN *et al.* 2003b). HopPmaG, on the other hand, is nearly ubiquitous among *P. syringae* strains and in many other bacterial species. Given the extensive similarity between HopPmaG from *Psy* B728A and its homolog from *P. aeruginosa* and the dissimilarity among *P. syringae* HopPmaG sequences, it appears that this locus has moved into *P. syringae* multiple times.

There are a number of other cases of probable horizontal transfer between *P. syringae* strains. Unfortunately, the current sampling of effector sequences makes it impossible to rule out alternative scenarios such as phylogenetic

incongruence due to extensive gene duplication and loss during the course of strain diversification.

**Evidence of horizontal transfer:** Most genes in a given genome have a similar GC content. Deviations from these values suggest a recent acquisition in the genome, most probably via a mobile element (GALTIER and LOBRY 1997; HACKER and KAPER 1999). Thus, the GC content of a particular gene can be compared to that of the genome to determine whether it was likely acquired through horizontal transfer. Eventually, the GC content of an acquired gene may coalesce with the host genome (LAWRENCE and OCHMAN 1997), and the sequences necessary for its mobilization may be eliminated (HACKER and KAPER 2000). The time necessary for sequence homogenization to the GC content of the genome is unknown and probably depends on how different these features were at the time of transfer, as well as the weak selective pressure imposed by the transcriptional and replication apparatus of the new host.

Genomic GC content influences the frequency of alternative synonymous codons. Synonymous codons differ from each other largely at the third base. Hence, the GC3 content is less likely to be influenced by selection and represents a more objective measurement of a genome's GC content than its overall GC content (LAWRENCE and OCHMAN 1997). We measured GC3 for the type III effector genes in the 24 families (MATERIALS AND METHODS). These values were graphed atop the distribution of GC3 contents from the respective genomes or closely related genomes where possible (Figure 2; see MATERIALS AND METHODS). We also used the overall GC content, as well as the codon usage (CAI value), for similar comparisons (Figure 2). The overall GC content and CAI comparisons were very similar to the GC3 analysis, as observed for other bacterial species (BELLGARD and GOJOBORI 1999). Additionally, we searched 20 kb surrounding each type III gene in our study, where the DNA sequences were available, for mobile elements or remnants of such. Their presence could indicate whether the region where the type III effector resides is likely to have transferred recently.

Type III effector genes with values that deviate from the genome values tend to be located either on plasmids or near mobile elements and remnants thereof, while type III effector genes with values close to the respective genome mean do not. Consistently deviant GC3, GC, and CAI values, the presence of mobile elements in the surrounding sequences, and the phylogenetic analysis (Figure 1) suggest that a fraction of type III effector genes have been recently acquired by the genomes in which they reside, whereas others have been in the genome long enough to express the GC content and codon usage of that genome and to have been selected for the elimination of the surrounding mobile element sequences.

**Recently acquired gene families define the "variable" type III effector suite:** In *P. syringae* genomes, we identi-

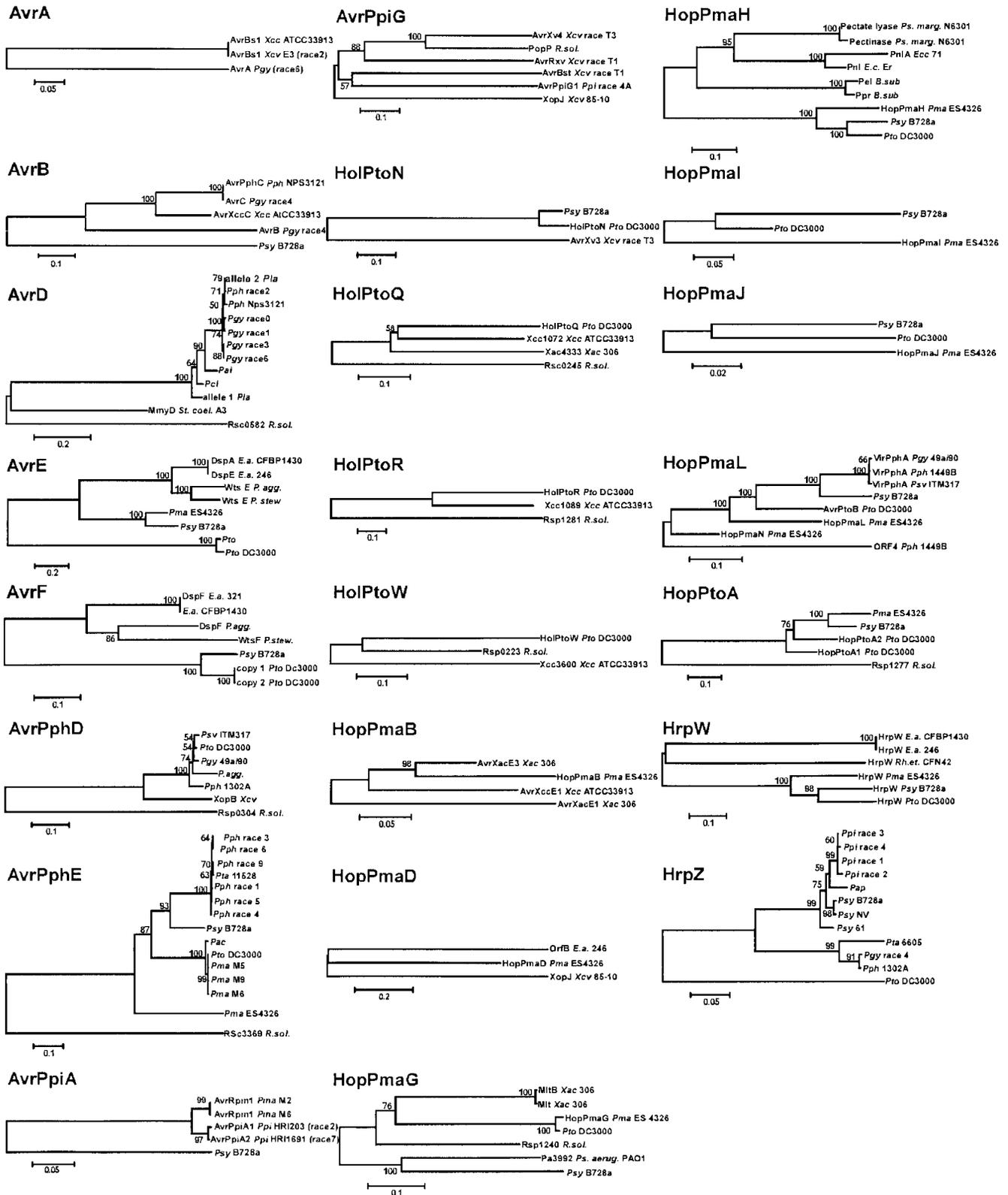


FIGURE 1.—Phylogenetic analyses of *P. syringae* type III effectors. Gene trees were inferred using the neighbor-joining method by the program MEGA based on the Poisson-correction distance model with protein sequence alignments. The name of each protein is given, when it exists. Otherwise, only the strain is given in the tree. The horizontal length of the branches is proportional to the estimated number of substitutions. The numbers above or below the internal branches show the local bootstrap probability (percentage) obtained for 1000 repetitions.

TABLE 1  
Genes in the 24 type III effector gene families were analyzed for features suggesting horizontal transfer: discrepancy in GC3 content and codon usage and genomic association with mobile elements

Family	No. of sequences	<i>P. syringae</i>				<i>Xanthomonas spp.</i>				<i>R. solanacearum</i>				Other			
		GC3 content <sup>a</sup>	Codon usage <sup>a</sup>	Mobile element <sup>a</sup>	Group	GC3 content <sup>a</sup>	Codon usage <sup>a</sup>	Mobile element <sup>a</sup>	Group	GC3 content <sup>a</sup>	Codon usage <sup>a</sup>	Mobile element <sup>a</sup>	Group	GC3 content <sup>a</sup>	Codon usage <sup>a</sup>	Mobile element <sup>a</sup>	Group
<i>avrA</i>	3	1/1	1/1	1/1	Recent	2/2	2/2	2/2	Recent	—	—	—	—	—	—	—	—
<i>avrB</i>	5	4/4	4/4	4/4	Recent	1/1	1/1	1/1	Recent	—	—	—	—	—	—	—	—
<i>avrD</i>	12	10/10	10/10	10/10	Recent	—	—	—	—	1/1	1/1	1/1	ND	0/1	ND	1/1	ND
<i>avrE</i>	8	0/2	1/2	0/2	Ancient	—	—	—	—	—	—	—	—	—	ND	0/5	Ancient
<i>avrF</i>	6	0/2	0/2	0/2	Ancient	—	—	—	—	—	—	—	—	—	ND	0/4	Ancient
<i>avrPphD</i>	7	0/4	0/4	1/2	Ancient	0/1	0/1	ND	Ancient	0/1	0/1	1/1	Ancient	ND	ND	1/1	ND
<i>avrPphE</i>	15	5/14	5/14	4/12	Intermediate	—	—	—	—	0/1	0/1	1/1	ND	—	—	—	—
<i>avrPpiA1</i>	5	5/5	5/5	5/5	Recent	—	—	—	—	—	—	—	—	—	—	—	—
<i>avrPpiB</i>	3	3/3	3/3	3/3	Recent	—	—	—	—	—	—	—	—	—	—	—	—
<i>avrPpiG1</i>	6	1/1	1/1	1/1	Recent	2/4	2/4	1/1	Intermediate	0/1	0/1	1/1	ND	—	—	—	—
<i>holPtoN</i>	3	2/2	2/2	2/2	Recent	1/1	0/1	ND	ND	—	—	—	—	—	—	—	—
<i>holPtoQ</i>	4	1/1	0/1	1/1	Recent	0/2	0/2	1/2	Ancient	0/1	0/1	0/1	Ancient	—	—	—	—
<i>holPtoR</i>	3	0/1	0/1	0/1	Ancient	0/1	0/1	0/1	Ancient	0/1	0/1	0/1	Ancient	—	—	—	—
<i>holPtoW</i>	3	1/1	1/1	1/1	Recent	0/1	0/1	0/1	Ancient	0/1	0/1	0/1	Ancient	—	—	—	—
<i>hopPmaB</i>	4	0/1	0/1	0/1	Ancient	0/3	0/3	2/3	Ancient	—	—	—	—	—	—	—	—
<i>hopPmaD</i>	3	0/1	0/1	ND	Ancient	0/1	0/1	ND	Ancient	—	—	—	—	ND	ND	0/1	ND
<i>hopPmaG</i>	7	0/3	0/3	0/3	Ancient	0/2	0/2	1/1	Ancient	0/1	0/1	0/1	Ancient	0/1	ND	1/1	ND
<i>hopPmaH</i>	9	0/3	0/3	0/3	Ancient	—	—	—	—	—	—	—	—	—	ND	ND	ND
<i>hopPmaI</i>	3	0/3	0/3	1/3	Ancient	—	—	—	—	—	—	—	—	—	—	—	—
<i>hopPmaJ</i>	3	0/3	0/3	1/3	Ancient	—	—	—	—	—	—	—	—	—	—	—	—
<i>hopPmaL</i>	8	3/8	5/8	5/6	Intermediate	—	—	—	—	—	—	—	—	—	—	—	—
<i>hopPtoA</i>	5	0/4	0/4	0/4	Ancient	—	—	—	—	0/1	0/1	0/1	Ancient	—	—	—	—
<i>hrpW</i>	5	0/3	0/3	0/3	Ancient	—	—	—	—	0/1	1/1	0/1	ND	—	—	—	—
<i>hrpZ</i>	12	0/12	0/12	0/6	Ancient	—	—	—	—	—	—	—	—	—	—	—	—

Results are given by species or genus. ND, not done.

<sup>a</sup> The proportion of genes for which the discrepancy or linkage to mobile elements (when available) is indicated.

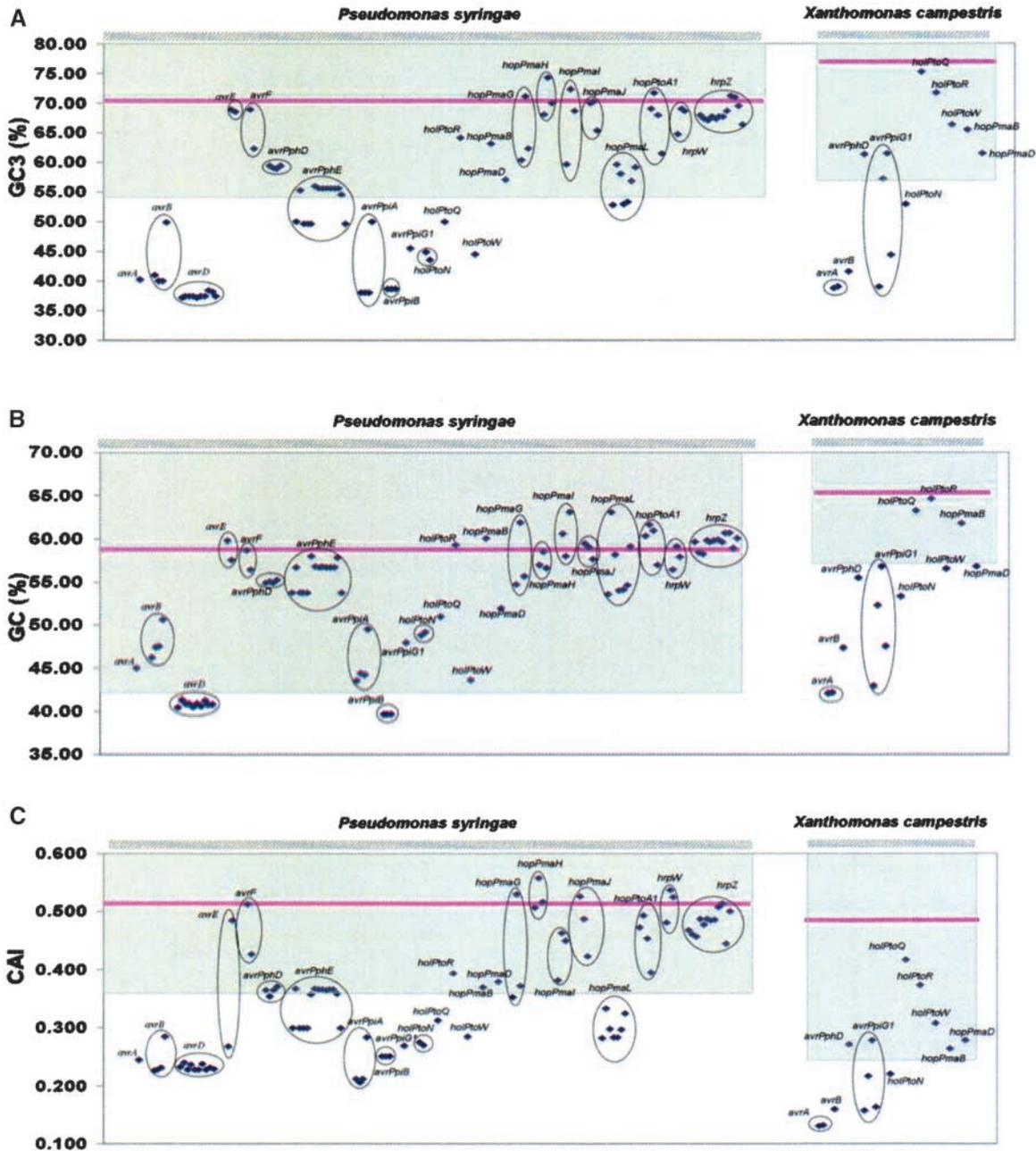


FIGURE 2.—Inference of *P. syringae* type III effector horizontal transfer. (A) GC3 content for *P. syringae* type III effector and chaperone xenologs (grouped into families) compared to the average GC3 content of the *P. syringae* pv. *tomato* DC3000 ORFs (left). Each diamond indicates a value for a specific type III effector gene. The purple horizontal line indicates the mean GC3 content of *Pst* DC3000 (70.25%) and the mean GC3 content of *X. campestris* pv. *campestris* (*Xcc*, 76.81%), respectively. The shaded box indicates two standard deviations from either side of the mean for *Pto* DC3000 ( $\pm 16.29$  from the mean) and *Xcc* ( $\pm 20.09$ ), respectively. Among the other complete genome sequences available, only the xenologs found in *Xcc* showed significantly different values compared to the average GC3 content of ORFs in the genome (right). (B) GC content for *P. syringae* type III effector and chaperone xenologs (grouped into families) compared to the average GC content of the *P. syringae* pv. *tomato* DC3000 ORFs (left) and GC content for *Xcc* xenologs compared to the average GC content of the *Xcc* ORFs (right). Each diamond indicates a value for a specific type III effector gene. The purple horizontal line indicates the mean GC content of *Pst* DC3000 (58.72%) and the mean GC content of *Xcc* (65.27%), respectively. The shaded box indicates two standard deviations from either side of the mean for *Pto* DC3000 ( $\pm 16.57$ ) and *Xcc* ( $\pm 8.27$ ), respectively. (C) Codon usage represented by the CAI for *P. syringae* type III effector and chaperone xenologs compared to the average CAI of the *P. syringae* pv. *tomato* DC3000 ORFs (left) and CAI of the *Xcc* xenologs compared to the average CAI of the *Xcc* ORFs (right). Each diamond indicates a value for a specific type III effector gene. The purple horizontal line indicates the mean CAI of *Pst* DC3000 (0.512) and the mean CAI of *Xcc* (0.48), respectively. The shaded box within the graph indicates two standard deviations from either side of the mean for *Pto* DC3000 ( $\pm 0.203$ ) and *Xcc* ( $\pm 0.31$ ), respectively.

fied nine families of genes where all members were probably acquired recently (*avrA*, *avrB*, *avrD*, *avrPpiA1*, *avrPpiB*, *avrPpiG1*, *holPtoN*, *holPtoQ*, and *holPtoW*). Each of these families exhibits corroborative evidence of recent transfer, including association with a plasmid or remnants of mobile elements (Table 1). For example, the *avrD* family is a large family, with homologs in 26 *P. syringae* strains from 12 pathovars (YUCEL *et al.* 1994). Homologs are also found in *R. solanacearum* and *S. coelicolor*. Every member of this family, except the member from *R. solanacearum*, is carried on a plasmid and therefore likely to be horizontally transferred (HANEKAMP *et al.* 1997).

Interestingly, three of these nine type III effector gene families contain some members recently acquired by *P. syringae*, but other members that seem ancient in other plant pathogen genomes. Xenologs to *holPtoQ* and *holPtoW* are found in both *R. solanacearum* and *Xanthomonas* species, where they were characterized as ancient genes (Table 1). It is, however, impossible to determine the source from which *P. syringae* obtained these genes. Similarly, *avrPpiG1* appears to have been acquired recently in the genome of *P. syringae* pv. *ptisi*. In contrast, two of four *avrPpiG1* xenologs found in *Xanthomonas* seem to be ancient while the other two seem to have been acquired recently, in their respective genomes. Thus, the *avrPpiG1* family consists of members that were acquired at different stages of *Xanthomonas* evolution.

The members of *avrPphE* and *hopPmaL* type III effector families also seem to match this evolutionary profile in *P. syringae*. Only 5 of 14 *avrPphE* family genes show all the features associated with horizontal transfer, whereas the 9 other genes in this family do not (Table 1; Figure 2; also see supplemental Table 2 at <http://www.genetics.org/supplemental/>). For *hopPmaL*, two genes were discovered in *P. syringae* pv. *phaseolicola*. These arose either from gene duplication or via separate horizontal transfers. If these genes arose through gene duplication, the high level of nucleotide diversity (data not shown) between the two sequences would imply very rapid evolution. These data indicate that the members of the *avrPphE* and *hopPmaL* gene families have different origins and could have been acquired at different steps in the evolution of these strains.

**A suite of “core” type II effector gene families:** The type III effector gene families exhibiting no difference in GC3, GC, or CAI values from the rest of the genome (in any phytopathogenic species or genera) are *avrE*, *avrF*, *avrPphD*, *holPtoR*, *hopPmaB*, *hopPmaD*, *hopPmaG*, *hopPmaH*, *hopPmaI*, *hopPmaJ*, *hopPtoA*, *hrpW*, and *hrpZ* (Table 1). These are likely to represent the ancient type III effector gene core set, acquired by *P. syringae* before diversification of the various pathovars.

The *hrp/hrc* pathogenicity island (PAI) exhibits an overall GC content of 58.7% in *Pto* DC3000 and *Psy* B278a, consistent with the average for those two genomes (58.4% for the whole genome of *Pto* DC3000

and an estimated 59.2% from the published draft of the genome of *Psy* B278a) and suggesting that it has not been recently acquired (SAWADA *et al.* 1999). There are two PAIs that flank the *hrp/hrc* cluster and that can contain type III effector genes (ALFANO *et al.* 2000). From our data, we suggest that at least some of the type III effector genes from these *hrp/hrc* linked families have not experienced recent horizontal gene transfer, but have probably been stable and evolved along with their respective genome. Some, like *avrF*, *hrpW*, and *hrpZ*, are linked to the *hrp/hrc* PAI, which encodes the structural and regulatory factors for the type III secretion pilus. Similarly, *hopPtoA1* family members are also *hrp/hrc* linked, except for the *hopPtoA2* homolog, found unlinked to *hopPtoA1* in the *Pst* DC3000 genome. Consistent with our proposal, no linkage to mobile elements could be detected for the members of the families of *hopPmaI* and *hopPmaJ*. These conclusions are also supported by DNA blot analyses showing that *hopPmaI*, *hopPmaJ*, *hopPmaG*, and *hrpW* are found almost universally among *P. syringae* strains (D. S. GUTTMAN, unpublished data).

Two families (*hopPmaB* and *hopPmaG*) contain members associated with a plasmid (*avrXacE3* and *mlt* of *X. axonopodis* pv. *citri* 306, homologous to *hopPmaB* and *hopPmaG*, respectively). *AvrXacE3* and *avrXacE1* in *X. axonopodis* pv. *citri* 306 are homologous and appear to have arisen by duplication. The presence of *avrXacE3* on a plasmid could facilitate the spread of this gene to other species, including *P. syringae*.

To summarize, the type III effector gene families analyzed can be classified into three distinct groups on the basis of their phylogeny, GC content, and genomic location (Figures 1 and 2; Table 1). The first group includes xenolog families in which all genes show evidence of horizontal transfer within *P. syringae* (*avrA*, *avrB*, *avrD*, *avrPpiA1*, *avrPpiB*, *avrPpiG1*, *holPtoN*, *holPtoQ*, and *holPtoW*) and between *P. syringae* and other species (*avrA*, *avrB*, and *avrPpiG1* xenologs). The second group contains type III effector gene families in which some members have probably been horizontally transferred, but at different times in the evolution of the species (*avrPphE* and *hopPmaL*). The last group contains the families that apparently have constituted an ancient or stable suite of virulence factors in *P. syringae* and, in some cases, other phytopathogenic bacteria (*avrE*, *avrF*, *avrPphD*, *holPtoR*, *hopPmaB*, *hopPmaD*, *hopPmaG*, *hopPmaH*, *hopPmaI*, *hopPmaJ*, *hopPtoA*, *hrpW*, and *hrpZ*).

**A role for diversifying selection in the evolution of some type III effector families:** The interaction between *P. syringae* pathovars and their plant hosts is strongly influenced by the evolution of both type III effector genes and the corresponding plant *R* genes that might detect their presence in the host. The extent of diversifying selection acting on type III effector genes might be constrained by a requirement to maintain their virulence function. If, as the “guard hypothesis” suggests,

the virulence function of a given type III effector initiates plant *R* action, then this constraint may be quite difficult to overcome. This perhaps explains the common, but counterintuitive, finding that type III effectors can be found as presence/absence alleles within a pathovar. Diversifying selection may, then, act on type III effector genes to facilitate (1) escape from host recognition, (2) adaptation to new alleles of their original host target (in their role as virulence factors), or (3) adaptation to additional host targets while maintaining their core virulence function.

To test the influence of diversifying selection on the evolution of type III effector genes, we used the program CODEML from the PAML package (YANG 1997). This program uses maximum likelihood to estimate the ratio ( $\omega$ ) of nonsynonymous ( $d_N$ ) to synonymous ( $d_S$ ) substitution rates for each codon position of a nucleotide alignment. It automatically performs corrections for multiple substitutions, which are likely in distantly related sequences (as illustrated by the tree lengths in Figure 1 and Table 2).

Using the M0 model of this program (YANG 1997; see MATERIALS AND METHODS), the  $\omega$ -value is constrained to be constant at each codon position, as if there were no variation in the selection acting across the sequence. With the M1 model, the program considers that some positions are under strong purifying selection ( $\omega$  close to 0) and the others are evolving neutrally ( $\omega$  close to 1). With models M3, M7, and M8, the program calculates the  $\omega$ -ratio for each codon and assigns the codons to a "site class," for which it estimates a  $\omega$ -ratio. Three discrete site classes ( $\omega_0$ ,  $\omega_1$ , and  $\omega_2$  as represented in Table 2) were considered for the M3 model, and 10 classes were considered for the M7 and M8 models. Models M7 and M8 use a  $\beta$ -distribution for the  $\omega$ -values. The 10 classes approximate as closely as possible the  $\beta$ -distribution of the  $\omega$ -values ranging between 0 and 1 (represented as parameters  $p$  and  $q$  in Table 2 for the M8 model; YANG *et al.* 2000b). M3 and M8 models also allow  $\omega$ -values above 1 (indicative of positive selection), while M7 does not. Since many of these tests are nested, it is possible to identify which model best fits the data using a likelihood ratio test (LRT), comparing one model against another and indicating the probability that one model fits the data better than the other.

We first compared the results obtained from the M3 and M0 models, using an LRT (ANISIMOVA *et al.* 2001). If the M3 model fits the data set better than the M0 model, then evolutionary pressures are not equal for all codons. The next comparison, between models M1 and M3, indicates whether the evolution of the gene is neutral or if there are any positively selected sites ( $\omega > 1$ ). The last comparison, between models M7 and M8, tests a model allowing positive selection against a neutral model. The comparison between M7 and M8 is more conservative than the comparison between M1 and M3 and may miss some genes undergoing weak positive

selection (ANISIMOVA *et al.* 2002). However, with our data set, the two models consistently suggested positive selection for the same gene families. The results for these tests are presented in Table 2. They suggest positive selection ( $\omega > 1$ ) during the evolution of 7/19 analyzed type III effector gene families (*avrD*, *holPtoN*, *holPtoQ*, *hopPmaB*, *hopPmaI*, *hopPmaL*, and *hrpW*). For the 12 other families, the models allowing positive selection (M3 and M8) did not detect any codon with  $\omega > 1$ .

Among the seven families potentially subjected to positive selection, four (*holPtoN*, *holPtoQ*, *hopPmaL*, and *hrpW*) have LRT  $P < 0.1$  (Table 2 and MATERIALS AND METHODS). In three cases (*avrD*, *hopPmaB*, and *hopPmaI*), the overall LRT  $P$  values are not significant, even though the M3 and M8 models suggest amino acid sites with  $\omega > 1$ . The M0 model has the best overall fit to the data for the *avrD* family ( $\omega = 0.5905$ ). This could be due to the majority of the protein undergoing constant purifying selection, while a few sites undergo positive selection (ANISIMOVA *et al.* 2001). Second, in the *hopPmaB* family, the strength of support for model M8 is not significantly better than that for the M7 model (Table 2). This could be due to the large number of sites (20%) that belong to classes with  $0.9 < \omega < 1$ , indicative of neutral evolution. For this data set, the LRT lacks power to differentiate between the M7 and M8 models (ANISIMOVA *et al.* 2001). Third, the *hopPmaI* family contains three moderately divergent genes (the tree length is 2.33). The M3 model fits the data set significantly better than the M1 and M0 models do. The M3 model predicts that 8.5% of sites are under positive selection ( $\omega = 2.2$ ). The more conservative model M8 also suggests that 5% of sites are under diversifying selection ( $\omega = 2.5$ ). The *hopPmaJ* and *avrPphE* gene families show  $\omega_2 > 1$  for the M3 model. However, none of the sites in the genes fell into the class  $\omega_2$  according to this model. Additionally, the  $\omega_2$  values were close to one, suggesting neutral, rather than positive, selection.

**Analysis of positively selected sites in type III effector gene families:** We sought to identify which sites in particular type III effector genes were subjected to positive selection for members of the *avrD*, *holPtoN*, *holPtoQ*, *hopPmaB*, *hopPmaI*, *hopPmaL*, and *hrpW* type III effector gene families. We employed the Bayesian calculation of posterior probabilities to identify which sites may be under positive selection, according to the site class to which they belong (NIELSEN and YANG 1998; YANG *et al.* 2000b). The probability for each site belonging to the class of positively selected sites is represented in Figure 3. We next correlated the location of these sites with putative functional domains assigned to the proteins from the following five families (for which information was available), on the basis of homologies or experimental data.

The *holPtoQ* family members were detected by bioinformatic analyses as putative type III effectors (GUTTMAN *et al.* 2002). We used the protein consensus sequence of

TABLE 2  
Overview of the results modeling selection on type III effector families obtained with CODEML

Family	No. of sequences in the analysis	Average $d_N/d_S$	Tree length (CODEML)	P-value M3:M0	P-value M3:M1	P-value M8:M7	Proportion of sites under positive selection detected		Model best fitting the data	ln L for the model fitting the best	Parameters in the $\omega$ -distribution	
							With M3 (%)	With M8 (%)			M3 model	M8 model
<i>avrA</i>	3	0.2348	3.33406	1.09E-05	4.2E-06	7.94E-01	NP	NP	M3	-2749.29	$\beta_0 = 0.34653$	$\beta_0 = 0.59405$
											$\beta_1 = 0.23706$	$\beta = 6.07760$
											$\beta_2 = 0.41641$	$q = 99.00000$
											$\omega_0 = 0.05365$	$(\beta_1 = 0.40595)$
											$\omega_1 = 0.05365$	$\omega = 0.49497$
<i>avrB</i>	5	0.1758	29.93223	1.62E-23	3.6E-44	4.49E-02	NP	NP	M8	-3573.96	$\beta_0 = 0.31227$	$\beta_0 = 0.69727$
											$\beta_1 = 0.39459$	$\beta = 0.32013$
											$\beta_2 = 0.29314$	$q = 0.95376$
											$\omega_0 = 0.00834$	$(\beta_1 = 0.30273)$
											$\omega_1 = 0.06173$	$\omega = 0.03425$
<i>avrD</i>	10	0.5905	0.42266	NA	9.4E-01	7.43E-01	0.40	0.40	M0	-2109.60	$\beta_0 = 0.25498$	$\beta_0 = 0.99594$
											$\beta_1 = 0.74069$	$\beta = 0.42952$
											$\beta_2 = 0.00432$	$q = 0.28467$
											$\omega_0 = 0.00001$	$(\beta_1 = 0.00406)$
											$\omega_1 = 0.80099$	$\omega = 20.39960$
<i>avrF</i>	7	0.1617	9.1427	6.64E-10	7.0E-21	6.34E-01	NP	NP	M3	-1696.30	$\beta_0 = 0.34049$	$\beta_0 = 0.65160$
											$\beta_1 = 0.54809$	$\beta = 0.31952$
											$\beta_2 = 0.11142$	$q = 1.61894$
											$\omega_0 = 0.01238$	$(\beta_1 = 0.34840)$
											$\omega_1 = 0.16599$	$\omega = 0.15917$
<i>avrPphD</i>	6	0.3591	1.69737	1.29E-18	2.5E-02	5.96E-01	NP	NP	M3	-5494.22	$\beta_0 = 0.59723$	$\beta_0 = 0.67219$
											$\beta_1 = 0.3637$	$\beta = 6.42498$
											$\beta_2 = 0.32954$	$q = 99.00000$
											$\omega_0 = 0.05915$	$(\beta_1 = 0.32781)$
											$\omega_1 = 0.05915$	$\omega = 0.97081$
<i>avrPphE</i>	15	0.2248	3.60038	3.02E-11	3.8E-20	7.37E-01	NP	NP	M3	-3453.96	$\beta_0 = 0.96926$	$\beta_0 = 0.44450$
											$\beta_1 = 0.31937$	$\beta = 0.01294$
											$\beta_2 = 0.59722$	$q = 0.03997$
											$\beta_2 = 0.08341$	$(\beta_1 = 0.55550)$
											$\omega_0 = 0.00001$	$\omega = 0.21628$
				$\omega_1 = 0.23083$								
				$\omega_2 = 1.04236$								

(continued)

TABLE 2  
(Continued)

Family	No. of sequences in the analysis	Average $d_N/d_S$	Tree length (CODEML)	$P$ -value M3:M0	$P$ -value M3:MI	$P$ -value M8:M7	Proportion of sites under positive selection detected		Model best fitting the data	$\ln L$ for the model fitting the best	Parameters in the $\omega$ -distribution	
							With M3 (%)	With M8 (%)			M3 model	M8 model
<i>avrPpiA</i>	4	0.342	3.12641	5.04E-11	8.0E-01	9.08E-01	NP	NP	M3	-1590.64	$\beta_0 = 0.28033$	$\beta_0 = 0.56549$
											$\beta_1 = 0.28447$	$\beta_1 = 1.07220$
<i>avrPpiG</i>	6	0.0311	75.40503	2.48E-58	NA	9.9E-01	NP	NP	M3	-7405.67	$(\beta_2 = 0.43521)$	$\beta_0 = 0.98934$
											$\omega_0 = 0.00993$	$\beta_1 = 0.40989$
<i>holPtoN</i>	3	0.6109	50.62056	1.22E-76	8.3E-84	4.1E-78	4.50	3.50	M8	-2347.01	$(\beta_2 = 0.37419)$	$\beta_0 = 0.96412$
											$\omega_0 = 0.00259$	$\beta_1 = 0.6133$
<i>holPtoQ</i>	4	0.2741	25.89776	1.38E-36	1.2E-50	0.07414972	11.10	9.30	M3	-4716.92	$(\beta_2 = 0.04534)$	$\beta_0 = 0.42576$
											$\omega_1 = 0.01089$	$\beta_1 = 0.46246$
<i>holPtoW</i>	3	0.1343	10.5815	1.85E-21	NA	NA	NP	NP	M3	-4886.02	$(\beta_2 = 0.11178)$	$\beta_0 = 0.40787$
											$\omega_1 = 0.05947$	$\beta_1 = 0.61712$
<i>hopPmaB</i>	4	0.4784	10.95719	3.13E-14	9.3E-02	4.1E-01	8	3.50	M3	-3354.29	$(\beta_2 = 0.07325)$	$\beta_0 = 0.96484$
											$\omega_1 = 0.10674$	$\beta_1 = 0.5168$
											$\omega_2 = 0.92158$	$\beta_0 = 0.24336$
											$\omega_2 = 2.18535$	$\beta_1 = 0.05149$
											$\omega_2 = 0.30963$	$q = 0.22848$
											$\omega_2 = 0.92158$	$(\beta_1 = 0.59213)$
											$\omega_2 = 0.92158$	$\omega = 0.09956$
											$\omega_2 = 0.39762$	$\beta_0 = 0.96484$
											$\omega_2 = 0.92158$	$\beta_1 = 0.24336$
											$(\beta_2 = 0.08558)$	$q = 0.38236$
											$\omega_0 = 0.00001$	$(\beta_1 = 0.03516)$
											$\omega_1 = 0.56004$	$\omega = 3.04737$
											$\omega_2 = 2.20767$	

(continued)

TABLE 2  
(Continued)

Family	No. of sequences in the analysis	Average $d_S/d_S$	Tree length (CODEML)	P-value M3:M0	P-value M3:M1	P-value M8:M7	Proportion of sites under positive selection detected		Model best fitting the data	ln $L$ for the model fitting the best	Parameters in the $\omega$ -distribution	
							With M3 (%)	With M8 (%)			M3 model	M8 model
<i>hopPmaG</i>	6	0.1338	8.21425	2.21E-45	7.3E-73	1.1E-01	NP	NP	M3	-5526.22	$\rho_0 = 0.2444$ $\rho_1 = 0.34908$ ( $\rho_2 = 0.40652$ ) $\omega_0 = 0.00001$ $\omega_1 = 0.06712$ $\omega_2 = 0.2715$	$\rho_0 = 0.74314$ $p = 0.42775$ $q = 3.85984$ ( $\rho_1 = 0.25686$ ) $\omega = 0.24394$
<i>hopPmaI</i>	3	0.4228	2.3327	9.07E-12	8.5E-04	7.4E-01	8.50	5	M3	-3704.37	$\rho_0 = 0.35997$ $\rho_1 = 0.55466$ ( $\rho_2 = 0.08537$ ) $\omega_0 = 0.00001$ $\omega_1 = 0.41804$ $\omega_2 = 2.23638$	$\rho_0 = 0.94869$ $p = 0.40628$ $q = 0.92608$ ( $\rho_1 = 0.05131$ ) $\omega = 2.51152$
<i>hopPmaJ</i>	3	0.1659	0.80391	0.99903	1.6E-03	9.7E-01	NP	NP	M3	-1002.55	$\rho_0 = 0.0513$ $\rho_1 = 0.93801$ ( $\rho_2 = 0.01069$ ) $\omega_0 = 0.15336$ $\omega_1 = 0.15336$ $\omega_2 = 1.3276$	$\rho_0 = 0.01487$ $p = 2.54205$ $q = 0.00100$ ( $\rho_1 = 0.98513$ ) $\omega = 0.15215$
<i>hopPmaL</i>	8	0.9709	3.62074	3.55E-17	4.0E-04	1.8E-03	6.80	8.20	M3	-6003.88	$\rho_0 = 0.39724$ $\rho_1 = 0.54094$ ( $\rho_2 = 0.06182$ ) $\omega_0 = 0.13304$ $\omega_1 = 0.91632$ $\omega_2 = 6.83261$	$\rho_0 = 0.93207$ $p = 0.45587$ $q = 0.34351$ ( $\rho_1 = 0.06793$ ) $\omega = 6.26811$
<i>hopPtoA</i>	4	0.1998	2.60053	8.39E-18	7.0E-22	NA	NP	NP	M3	-4736.63	$\rho_0 = 0.71581$ $\rho_1 = 0.00000$ ( $\rho_2 = 0.28419$ ) $\omega_0 = 0.05406$ $\omega_1 = 0.35487$ $\omega_2 = 0.56702$	$\rho_0 = 1.00000$ $p = 0.37860$ $q = 1.50673$ ( $\rho_1 = 0.00000$ ) $\omega = 0.35156$
<i>hrpW</i>	5	0.2873	16.52812	1.46E-37	1.7E-39	5.6E-02	12.50	1.50	M3	-5321.77	$\rho_0 = 0.28823$ $\rho_1 = 0.58613$ ( $\rho_2 = 0.12565$ ) $\omega_0 = 0.00489$ $\omega_1 = 0.186$ $\omega_2 = 1.4079$	$\rho_0 = 0.98450$ $p = 0.43715$ $q = 1.53500$ ( $\rho_1 = 0.01550$ ) $\omega = 9.53042$

(continued)

TABLE 2  
(Continued)

Family	No. of sequences in the analysis	Average $d_n/d_s$	Tree length (CODEML)	$P$ -value M3:M0	$P$ -value M3:M1	$P$ -value M8:M7	Proportion of sites under positive selection detected		Model best fitting the data	ln $L$ for the model fitting the best	Parameters in the $\omega$ -distribution	
							With M3 (%)	With M8 (%)			M3 model	M8 model
<i>hnpZ</i>	11	0.2105	2.87881	6.9E-09	1.1E-19	4.2E-01	NP	NP	M3	-3538.67	$\rho_0 = 0.49306$ $\rho_1 = 0.20477$ ( $\rho_2 = 0.30217$ ) $\omega_0 = 0.06795$ $\omega_1 = 0.06796$ $\omega_2 = 0.53965$	

NP, no positive selection could be detected using this test; NA, the  $P$ -value could not be calculated.

the HolPtoQ homologs to search for domain homology. Positions 55–282 of the consensus are homologous to the inosine-uridine nucleoside N-ribohydrolase domain (URH1,  $E$ -value =  $2 \times 10^{-4}$ , alignment of 70% of the sequence) from the Conserved Domain Database (CDD; MARCHLER-BAUER *et al.* 2002, 2003) and the Structural Classification Of Proteins database (SCOP; Cambridge, UK). Catalytic residues of some hydrolases from *Trypanosoma brucei brucei* and *Crithidia fasciculata* have been identified by directed mutagenesis (DEGANO *et al.* 1996; GOPAUL *et al.* 1996; PELLE *et al.* 1998). Oddly, these were not the same in both species, and thus it is perhaps unsurprising that none of them were conserved in the HolPtoQ proteins. However, other domains highly conserved between every purine hydrolase (KURTZ *et al.* 2002) are also conserved in the HolPtoQ proteins (Figure 3). The domain from codon 59 to 68, identified in some purine hydrolases and in HolPtoQ proteins, is part of a metal ligand-binding pocket (PELLE *et al.* 1998). While no diversification was detected in this motif, positive selection was detected in the substrate-binding pocket domain encoded by codons 190–236 (PELLE *et al.* 1998). If these putative type III effectors have a virulence function, then variation in the substrate pocket could be driven by the necessity to adapt to new substrates while escaping the host surveillance system or to expand the range of possible virulence targets. Additional residues carrying signs of potential positive selection are located on the probable exposed surface of the protein (PELLE *et al.* 1998), the most significant being codons 4, 7, 9, 27, 317, and 320. These may also be important in evasion of host recognition.

Genes in the *hopPmaB* family contain at least three positions that are subjected to diversifying selection: at codons 23, 141, and 193. It is possible that codons 27 and 60 are diversifying as well; they were detected using model M3, but not detected when using model M8. These proteins are putative cysteine proteases, characterized by a catalytic triad, involving a cysteine, a histidine, and a glutamic acid or aspartic acid (positions 164, 47, and 120 in the consensus protein sequence). The N terminus of the consensus sequence is homologous to PD563556 of the PRODOM database, which is found in peptidase C55. Two HopPmaB domains were identified as being part of the catalytic domain. The domain GAGNCDXNAAI (positions 160–171) contains the cysteine residue of the catalytic triad and was detected in the CDD by PSI-blast. The positively selected codon 141 is 23 residues upstream of the cysteine residue. The Prositecan search program detected the PS00639 domain in the consensus HopPmaB family sequence (positions 45–57), containing the histidine residue of the catalytic triad, SIHGLXALG<sub>s</sub>XX. Position 56 may be positively selected in the family ( $P = 0.596$ ,  $\omega = 1.53$ ). The histidine residue from the consensus has been replaced in HopPmaB by a glutamine residue. The last active residue of the triad is likely an aspartic acid in the domain NVDSLRLSNG

(positions 116–127). This aspartic acid is conservatively substituted with a glutamic acid in HopPmaB. This region is under strong purifying selection ( $P > 0.95$ ), suggesting the importance of this domain for the function of the protein. Overall, the catalytic triad is strictly conserved in the *Xanthomonas* species, but not in *Pma* ES4326. Furthermore, diversifying selection near conserved domains was detected, suggesting that despite the conservation of the function in the *Xanthomonas* species, the activity or substrate preference of these proteins could vary.

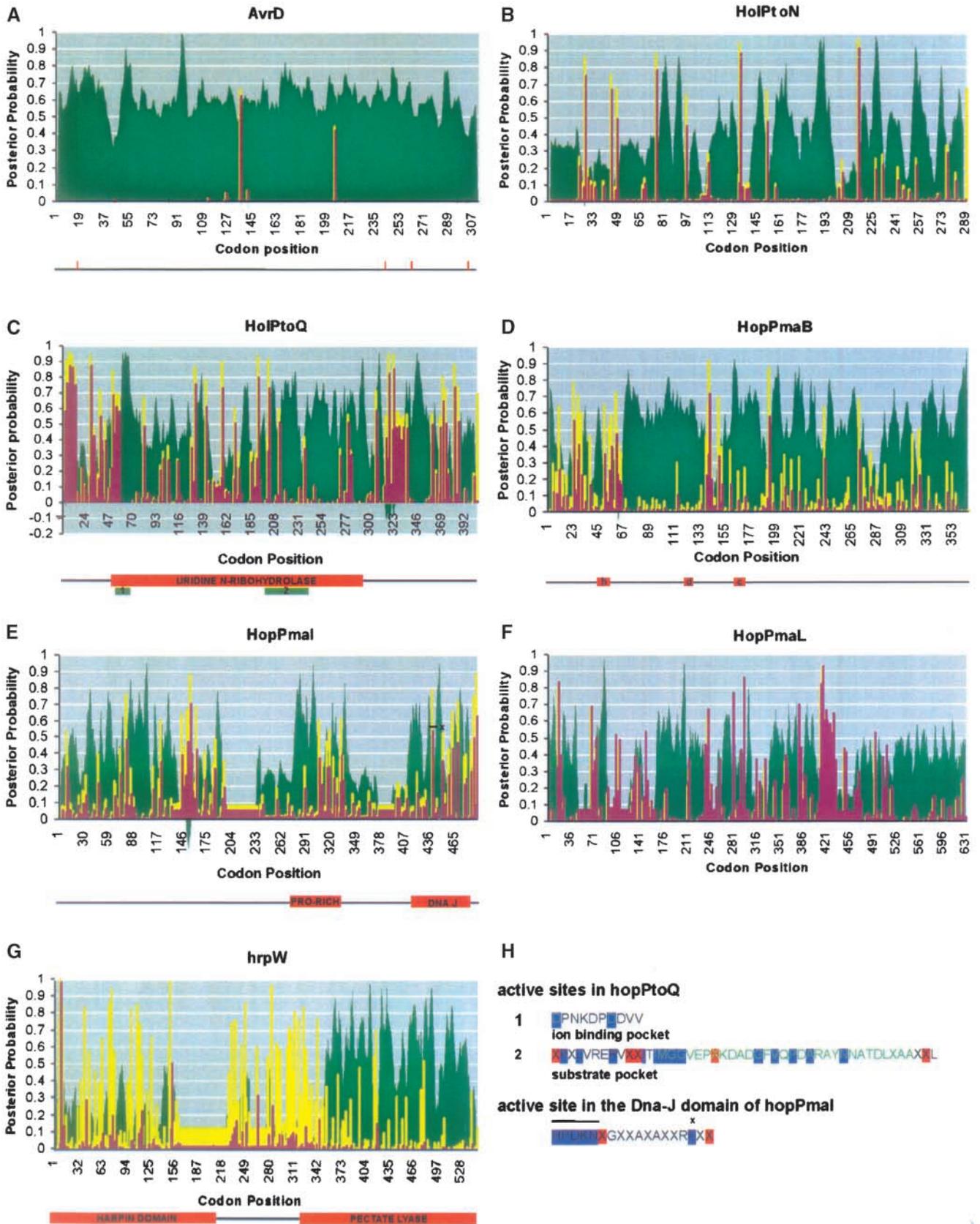
The proteins encoded by the genes of the *hopPmaI* family contain a J domain near their deduced carboxyl terminus. This domain is involved in protein-protein interactions. It is found in cochaperones in eukaryotes and prokaryotes (for review see KELLEY 1998). Most proteins containing the J domain are involved in signaling pathways, such as those required for apoptosis (plant and human viruses; SULLIVAN and PIPAS 2002) and heat shock. In Hsp40, a cochaperone of Hsp70, the J domain orchestrates interaction with the DnaK ATPase domain. The Hsp40 cochaperones specifically help stimulate ATP hydrolysis and deliver substrates to Hsp70. Active residues of this domain were found by directed mutagenesis and the structure of the J domain of Hsp40 (GENEVAUX *et al.* 2002). The *hopPmaI* codons subjected to positive selection are directly next to the sites determining J-domain activity (positions 445–463). The positively selected site 450 is next to the conserved domain HPDKN in the J domain (445–449). The positively selected site 462 is very near the conserved phenylalanine 460. The positively selected site 438 corresponds to the site 26 of Hsp40. Mutation of K26 in Hsp40 reduces its activity (GENEVAUX *et al.* 2002). As the putative function of J domain is to bridge proteins and substrates, a diversification in this domain near key positions could preserve the activity of the protein but allow diversification of host target binding. A region matching a proline-rich domain is detected from positions 282 to 335. This domain may be involved in protein-protein interaction. Several serines are conserved in this region. Three sites putatively subjected to diversifying selection are also detected in this region, at positions 305, 317, and 331.

The genes in the *hrpW* family encode pectate-binding

proteins (CHARKOWSKI *et al.* 1998; KIM and BEER 1998). The N terminus of HrpW from *Pst* DC3000 and *E. amylovora* Ea321 is sufficient to trigger the hypersensitive response (HR) on nonhost plants, while the C terminus of the protein is not involved in this process (CHARKOWSKI *et al.* 1998; KIM and BEER 1998). This correlates with our finding that the N terminus is highly variable between these genes and contains many amino acid insertions and deletions. The putatively positively selected sites are concentrated in the region that triggers the host defense response. According to model M3, three sites are positively selected with a  $P > 95\%$ , residue 8 ( $\omega = 1.4$ ), residue 151 ( $\omega = 1.4$ ), and residue 282 ( $\omega = 1.36$ ). The C-terminal domain, starting at codon position 350 in the consensus sequence, encodes the pectate lyase activity (pfam03211.5, pectate\_lyase). The 3' end is highly conserved, with particularly low  $\omega$ -values. According to M3, M7, and M8, 75% of all sites in this domain are subjected to strong purifying selection. These results strongly suggest that either host recognition or adaptation to a new virulence target drove the evolution of the 5' end of these genes. The 3' end of the gene is strongly conserved, indicative of conservation of pectate lyase function. This result further suggests that pectate lyase function is not selected against during the interaction with the host.

Virulence functions for members of the HopPmaL family (VirPphA<sub>Pph</sub>, VirPphA<sub>Psv</sub>, VirPphA<sub>Pgy</sub>, and AvrPtoB) were recently described (JACKSON *et al.* 2002; ABRAMOVITCH *et al.* 2003), and these proteins can complement each other's function (JACKSON *et al.* 2002). An HR defense response in tomato is triggered by the direct recognition of the AvrPtoB N terminus by the host Pto protein (KIM *et al.* 2002). However, the C terminus of AvrPtoB can inhibit the HR triggered by recognition of other type III effectors (ABRAMOVITCH *et al.* 2003). Positive selection is acting on AvrPtoB codons 16 and 295, in the region involved in recognition. Interestingly, when we restricted our alignment to only the family members experimentally proven to encode type III effectors (VirPphA<sub>Pph</sub>, VirPphA<sub>Psv</sub>, VirPphA<sub>Pgy</sub>, HopPmaL, and AvrPtoB), the number of positively selected sites and the probabilities for positive selection were increased. Additional positively selected sites, with  $P > 95\%$ , were found at codons 65, 70, 237, 357, 411, and

FIGURE 3.—Positive selection of *P. syringae* type III effectors revealed by PAML. For each of the type III effector gene families, the posterior probabilities of the positively selected class ( $\omega > 1$ ,  $\omega = d_n/d_s$ ) are plotted for all codon positions of the alignment (A–G). Bars in yellow represent the probabilities calculated with model M3, which assumes 3 classes of sites (2 classes with  $0 \leq \omega \leq 1$ ; 1 class with  $\omega > 1$ ). Bars in violet represent the probabilities calculated with model M8, which assumes 10 classes of sites (all but 1 class  $0 \leq \omega \leq 1$ ). The green area represents the degree of similarity between sequences in the alignment for each position measured by the program plotcon from the EMBOSS package (the maximum for this similarity is 1). Below the plot, a schematic of the proteins encoded by the gene family depicts domains identified for each of them. More detailed analyses of active sites for HoiPtoQ and HopPmaI are shown in H. The residues in blue are conserved among proteins containing this domain. The residues in red are potentially positively selected. The residues in green (for HoiPtoQ) are the catalytic domain identified in the hydrolase of *T. brucei brucei*. The position of the conserved residues of the DNA-J domain is indicated by a line and the x above the sequence and on the graph (E).



412. These residues are located in the region containing the Pto-binding domain. Positive selection was also seen acting on codon 510, for which no function is defined. This result might signify that evolution in the AvrPtoB gene family is ongoing and leading to functional divergence at the margins of the family. The three homologs left out of the more restricted analysis may be degenerating and no longer subjected to positive selection.

## DISCUSSION

Xenologs were identified for 36 putative type III effectors or chaperones among 46 genes used to search the databases using BlastP. A large majority of the identified xenologs were found in other plant pathogen genomes, although many more animal bacterial pathogen genome sequences have been published. This may be explained by the fact that the plant pathogenic bacteria share the same niche, which favors horizontal transfer between them. Additionally, virulence factors identified in phytopathogens may target factors specific to plant cells. The type III secretion system in plant pathogenic bacteria for instance is different from its animal pathogen counterparts since it must cross the thick plant cell wall.

We provide evidence for recent horizontal acquisition by *P. syringae* in 11/24 type III effector gene families (nine "recent" and two "intermediate" families). We corroborated probable horizontal transfer of type III effector genes through phylogenetic incongruence, atypical GC3 and GC content and genomic location. The variable distribution of many of these genes among *P. syringae* pathovars also suggests their recent acquisition (D. S. GUTTMAN, unpublished data; J. H. CHANG, unpublished data). This confirms what had been previously suggested for several known type III effector genes (HANEKAMP *et al.* 1997; KIM *et al.* 1998; VIVIAN *et al.* 2001), on the basis of only their physical linkage to remnants of transposons. Our data, however, are based on several independent criteria collectively associated with horizontal transfer.

These 11 families can be further subdivided into those in which all members showed evidence for recent horizontal transfer and those in which only some members were found to be recently acquired. It is interesting that in at least three cases (*avrPpiG1*, *holPtoQ*, and *holPtoW*), type III effector genes seem to have been acquired recently in the genome of *P. syringae*, but not in the genome of other phytopathogenic species, such as *Xanthomonas* and *Ralstonia*. We cannot, however, postulate a source species for these genes. A low GC content, for example, does not mean that it was transferred from an organism exhibiting a low GC content (DAUBIN *et al.* 2003a). In addition, we identified a third class of *P. syringae* type III effector/chaperone families with ancient genes that show no evidence for recent horizontal transfer.

Our analysis of the distribution of type III effector gene families has established that virulence factors are exchanged not only between pathovars of the same species, but also sometimes between different phytopathogenic species. For example, members of the *hopPmaB* family may have been transferred from *Xanthomonas* to *P. syringae*. Some *P. syringae* type III effectors can be secreted through type III secretion systems of other pathogens, such as *E. amylovora*, *Xanthomonas*, or *Yersinia pestis*, demonstrating the conservation of the mechanism of secretion between bacteria genera (HAM *et al.* 1998; ANDERSON *et al.* 1999; CORNELIS and VAN GIJSEGEM 2000). This is consistent with data suggesting that the first ~50–75 amino acids of type III effector proteins may encode an amphipathic NH<sub>2</sub> terminus potentially required for delivery through the type III secretion system (MUDGETT and STASKAWICZ 1999; LLOYD *et al.* 2001; GUTTMAN *et al.* 2002). Since type III effector proteins can be secreted by heterologous type III systems, the transfer of these genes might allow for rapid movement of strains into new and perhaps unexplored niches. However, the actual situation is certain to be complicated by different regulatory controls operating between different bacterial species.

The acquisition of type III effector/chaperone genes occurred in different temporal frames: some were present in a strain ancestral to all the pathovars analyzed, while others were acquired after pathovar differentiation. The ancient genes, present in nearly all *P. syringae* strains analyzed to date, could provide virulence functions of broad utility and may encode proteins whose host targets are not monitored by the plant surveillance system. More recently acquired type III effector genes could contribute to strategies specific to particular pathogen-host interactions. The evolution of type III effector genes is driven largely by the necessity to escape recognition by host surveillance proteins and probably modulated by evolving host virulence targets. This type of evolutionary change is influenced predominantly by positive evolution, in which strains that carry beneficial alleles increase in frequency in the population.

Numerous virulence genes in pathogenic bacteria and viruses have been shown to be under positive selection (MCGRAW *et al.* 1999; REID *et al.* 1999, 2000; MOURY *et al.* 2002; TARR and WHITTAM 2002). We were interested in understanding the selective pressures that act on *P. syringae* type III effector genes and whether there was an obvious difference between the pressures that act on ancient *vs.* recent genes. We believed that ancient type III effector genes would diversify to adapt to specific host targets or to avoid host-plant recognition. It was less apparent whether recently acquired type III effector genes would be under positive selection. Some of these may have been passively transferred to new strains, while others may be sweeping through a local population due to a transient selective advantage.

We identified 13 probable ancient type III effector/

chaperones gene families in this study (*avrE*, *avrF*, *avrPphD*, *holPtoR*, *hopPmaB*, *hopPmaD*, *hopPmaG*, *hopPmaH*, *hopPmaI*, *hopPmaJ*, *hopPtoA*, *hrpW*, and *hrpZ*). We were able to identify clear evidence of positive selection in only 3 of them (*hopPmaB*, *hopPmaI*, and *hrpW*). Why do we not see positive selection in a larger number of these gene families? The answer may be that the functions of these genes are so important that purifying selection dominates their evolution. If this were the case, then the products of these highly constrained genes would be ideal targets for pathogen surveillance systems and we would expect to find that many of these effector proteins are recognized by host resistance proteins. Contrary to this argument, very few of these type III effector genes correspond to defined host disease resistance genes. A possible explanation is that our sampling may be simply not deep enough to identify positive selection when it is occurring, given the small size and extensive diversity found in some of these families. Alternatively, *P. syringae* may use additional type III effector genes to suppress the HR-type resistance triggered by the effector recognition in the host cell, such as AvrPphC suppressing the response triggered by the recognition of AvrPphF from *P. syringae* pv. *phaseolicola* (TSIAMIS *et al.* 2000).

Positive selection is also found in type III effector genes that have undergone recent horizontal transfer (*avrD*, *hopPmaL*, *holPtoN*, and *holPtoQ*), although it is not clear whether the selection occurred prior to or after their transfer. Additionally, it is equally unclear whether diversification of these type III effector genes led to changes in host specificity or diversification of function. For *HolPtoQ*, the presence of positively selected sites in the probable substrate-binding pocket suggests that the substrate may vary between the homologs, perhaps because of variation in the target sequence between host plants. Alternatively, the genes of one type III effector family might be used by different pathovars for slightly different functions on different host targets. It has already been shown that similar genes can accomplish different virulence functions. In *X. oryzae* pv. *oryzae* strain PXO86, seven genes of the *avrBs3* family encode putative transcriptional regulators (BAI *et al.* 2000). Six of them contribute to pathogenicity. They do not complement each other, showing that they evolved different functions that are all more or less important for the virulence strategy (BAI *et al.* 2000).

We can also partition the set of positively selected type III effector genes into those in which the positive selection is acting on the probable host recognition domain and those in which the selection is acting on the putative virulence domain. *HrpW* and *hopPmaL* are in the former group and appear under selection to avoid host recognition. *HolPtoQ*, *hopPmaB*, and *hopPmaI* are in the latter group and are presumably tracking host virulence targets. It is possible that the recent acquisition and positive selection on *holPtoQ* enables strains to change host specificity or adapt to new host targets. In

sum, our data suggest that these cases of positive selection are most consistent with a model in which type III effectors evolve to maintain a core virulence function while potentially expanding the repertoire of host targets they can manipulate.

A better understanding of the acquisition of type III effector genes will provide insight into how any given type III effector fits into the overall virulence strategy of a pathogen. Tracking selectively diversified or constrained regions will help identify domains important for the type III effector gene's function or its interaction with the host recognition machinery. The ongoing search for type III effector genes in a variety of *P. syringae* pathovars and further elucidation of the forces driving their evolution will shed light on the critical and central role these proteins play in pathogen-host interactions.

The temporary sequence for the genome of *P. syringae* pv. *phaseolicola* generated by random shotgun sequencing was made available in February 2004 by TIGR (<http://tigrblast.tigr.org/ufmg/>), during the process of editing this article, although the sequence contains gaps and no ORFs have been predicted. This sequence was searched for homologs with the TBLASTN algorithm. The query sequences were the protein sequences we used for the original search that gave rise to xenolog families in this article (see supplemental Table 1 at <http://www.genetics.org/supplemental/>). The results of the search are shown in supplemental Table 3 at <http://www.genetics.org/supplemental/>. Briefly, of 13 families of ancient genes, 11 were detected in the genome of *P. syringae* pv. *phaseolicola*. Ancient genes that are present are located on the chromosome except for *avrPphD*. The two xenologs that were not detected are *hopPmaB* and *hopPmaD*. These two genes were labeled as ancient. At present, it is not possible for us to determine whether the missing genes are present in the genome but not in the unfinished sequence released by TIGR, whether the genes have been deleted from the genome throughout evolution of the strain due to selective pressures, or whether the genes are not ancient, as we assumed on the basis of sequence analysis. As expected, only five of nine genes labeled as recent were found in the genome of *P. s.* pv. *phaseolicola*: *avrB*, *avrD*, and *holPtoQ* on a plasmid and *holPtoN* and *holPtoW* on the chromosome. These preliminary results corroborate, for the most part, our prediction of chronology for the acquisition of putative type III effectors or chaperones in *P. syringae*.

The authors thank Jeff Chang for the careful and critical reading of the manuscript and Todd Vision and Jason Phillips for providing important analysis tools and for their advice during the course of this work. We are very grateful to The Institute for Genome Research (TIGR) and the U.S. Department of Energy-Joint Genome Initiative for making the *Pst* DC3000 and *Psy* B278a genome sequences, respectively, available publicly before finishing. This work was supported by the DOE Office of Basic Energy Biosciences grant DE-FG05-95ER20187 and National Institutes of Health grant RO1 GM066025 to J.L.D. D.S.G. was supported by a grant from the National Science and Engineering Research Council of Canada and the Canadian Foundation for Innovation.

## LITERATURE CITED

- ABRAMOVITICH, R. B., Y. J. KIM, S. CHEN, M. B. DICKMAN and G. B. MARTIN, 2003 Pseudomonas type III effector AvrPtoB induces plant disease susceptibility by inhibition of host programmed cell death. *EMBO J.* **22**: 60–69.
- ALFANO, J. R., and A. COLLMER, 1997 The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death. *J. Bacteriol.* **179**: 5655–5662.
- ALFANO, J. R., A. O. CHARKOWSKI, W. L. DENG, J. L. BADEL, T. PETNICKI-OCWIEJA *et al.*, 2000 The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. *Proc. Natl. Acad. Sci. USA* **97**: 4856–4861.
- ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- ANDERSON, D. M., D. E. FOUTS, A. COLLMER and O. SCHNEEWIND, 1999 Reciprocal secretion of proteins by the bacterial type III machines of plant and animal pathogens suggests universal recognition of mRNA targeting signals. *Proc. Natl. Acad. Sci. USA* **96**: 12839–12843.
- ANISIMOVA, M., J. P. BIELAWSKI and Z. YANG, 2001 Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**: 1585–1592.
- ANISIMOVA, M., J. P. BIELAWSKI and Z. YANG, 2002 Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Mol. Biol. Evol.* **19**: 950–958.
- APANUS, V., D. PENN, P. R. SLEV, L. R. RUFF and W. K. POTTS, 1997 The nature of selection on the major histocompatibility complex. *Crit. Rev. Immunol.* **17**: 179–224.
- AXTELL, M. J., and B. J. STASKAWICZ, 2003 Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4. *Cell* **112**: 369–377.
- BAI, J., S. H. CHOI, G. PONCIANO, H. LEUNG and J. E. LEACH, 2000 *Xanthomonas oryzae* pv. *oryzae* avirulence genes contribute differently and specifically to pathogen aggressiveness. *Mol. Plant-Microbe Interact.* **13**: 1322–1329.
- BELLGARD, M. I., and T. GOJOBORI, 1999 Significant differences between the G+C content of synonymous codons in orthologous genes and the genomic G+C content. *Gene* **238**: 33–37.
- BOCH, J., V. JOARDAR, L. GAO, T. L. ROBERTSON, M. LIM *et al.*, 2002 Identification of Pseudomonas syringae pv. tomato genes induced during infection of Arabidopsis thaliana. *Mol. Microbiol.* **44**: 73–88.
- CHANG, J. H., A. K. GOEL, S. R. GRANT and J. L. DANGL, 2004 Wake of the flood: ascribing functions to the wave of type III effector proteins of phytopathogenic bacteria. *Curr. Opin. Microbiol.* **7**: 11–18.
- CHARKOWSKI, A. O., J. R. ALFANO, G. PRESTON, J. YUAN, S. Y. HE *et al.*, 1998 The Pseudomonas syringae pv. tomato HrpW protein has domains similar to harpins and pectate lyases and can elicit the plant hypersensitive response and bind to pectate. *J. Bacteriol.* **180**: 5211–5217.
- COLLMER, A., M. LINDBERG, T. PETNICKI-OCWIEJA, D. J. SCHNEIDER and J. R. ALFANO, 2002 Genomic mining type III secretion system effectors in Pseudomonas syringae yields new picks for all TTSS prospectors. *Trends Microbiol.* **10**: 462–469.
- CORNELIS, G. R., and F. VAN GJSEGEM, 2000 Assembly and function of type III secretory systems. *Annu. Rev. Microbiol.* **54**: 735–774.
- DALE, C., G. R. PLAGUE, B. WANG, H. OCHMAN and N. A. MORAN, 2002 Type III secretion systems and the evolution of mutualistic endosymbiosis. *Proc. Natl. Acad. Sci. USA* **99**: 12397–12402.
- DANGL, J. L., and J. D. JONES, 2001 Plant pathogens and integrated defence responses to infection. *Nature* **411**: 826–833.
- DAUBIN, V., E. LERAT and G. PERRIÈRE, 2003a The source of laterally transferred genes in bacterial genomes. *Genome Biol.* **4**: R57.
- DAUBIN, V., N. A. MORAN and H. OCHMAN, 2003b Phylogenetics and the cohesion of bacterial genomes. *Science* **301**: 829–832.
- DEGANO, M., D. N. GOPAUL, G. SCAPIN, V. L. SCHRAMM and J. C. SACCHETTINI, 1996 Three-dimensional structure of the inosine uridine nucleoside N-ribohydrolase from Crithidia fasciculata. *Biochemistry* **35**: 5971–5981.
- EISEN, J. A., 1995 The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J. Mol. Evol.* **41**: 1105–1123.
- ENDO, T., K. IKEO and T. GOJOBORI, 1996 Large-scale search for genes on which positive selection may operate. *Mol. Biol. Evol.* **13**: 685–690.
- FOUTS, D. E., J. L. BADEL, A. R. RAMOS, R. A. RAPP and A. COLLMER, 2003 A pseudomonas syringae pv. tomato DC3000 Hrp (type III secretion) deletion mutant expressing the Hrp system of bean pathogen P. syringae pv. syringae 61 retains normal host specificity for tomato. *Mol. Plant-Microbe Interact.* **16**: 43–52.
- GALTIER, N., and J. R. LOBRY, 1997 Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J. Mol. Evol.* **44**: 632–636.
- GENEVAUX, P., F. SCHWAGER, C. GEORGIOPOULOS and W. L. KELLEY, 2002 Scanning mutagenesis identifies amino acid residues essential for the *in vivo* activity of the *Escherichia coli* DnaJ (Hsp40) J-domain. *Genetics* **162**: 1045–1053.
- GOPAUL, D. N., S. L. MEYER, M. DEGANO, J. C. SACCHETTINI and V. L. SCHRAMM, 1996 Inosine-uridine nucleoside hydrolase from Crithidia fasciculata. Genetic characterization, crystallization, and identification of histidine 241 as a catalytic site residue. *Biochemistry* **35**: 5963–5970.
- GREENBERG, J. T., and B. A. VINATZER, 2003 Identifying type III effectors of plant pathogens and analyzing their interaction with plant cells. *Curr. Opin. Microbiol.* **6**: 20–28.
- GUTTMAN, D. S., B. A. VINATZER, S. F. SARKAR, M. V. RANALL, G. KETTLER *et al.*, 2002 A functional screen for the type III (Hrp) secretome of the plant pathogen Pseudomonas syringae. *Science* **295**: 1722–1726.
- HACKER, J., and J. B. KAPER, 1999 The concept of pathogenicity islands, pp. 1–12 in *Pathogenicity Islands and Other Mobile Virulence Elements*, edited by J. HACKER and J. B. KAPER. ASM Press, Washington, DC.
- HACKER, J., and J. B. KAPER, 2000 Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* **54**: 641–679.
- HAM, J. H., D. W. BAUER, D. E. FOUTS and A. COLLMER, 1998 A cloned Erwinia chrysanthemi Hrp (type III protein secretion) system functions in Escherichia coli to deliver Pseudomonas syringae Avr signals to plant cells and to secrete Avr proteins in culture. *Proc. Natl. Acad. Sci. USA* **95**: 10206–10211.
- HANEKAMP, T., D. KOBAYASHI, S. HAYES and M. M. STAYTON, 1997 Avirulence gene D of Pseudomonas syringae pv. tomato may have undergone horizontal gene transfer. *FEBS Lett.* **415**: 40–44.
- HAUCK, P., R. THILMONY and S. Y. HE, 2003 A Pseudomonas syringae type III effector suppresses cell wall-based extracellular defense in susceptible Arabidopsis plants. *Proc. Natl. Acad. Sci. USA* **100**: 8577–8582.
- HE, S. Y., H. C. HUANG and A. COLLMER, 1993 Pseudomonas syringae pv. syringae harpinPss: a protein that is secreted via the Hrp pathway and elicits the hypersensitive response in plants. *Cell* **73**: 1255–1266.
- JACKSON, R. W., E. ATHANASSOPOULOS, G. TSIAMIS, J. W. MANSFIELD, A. SESMA *et al.*, 1999 Identification of a pathogenicity island, which contains genes for virulence and avirulence, on a large native plasmid in the bean pathogen Pseudomonas syringae pathovar phaseolicola. *Proc. Natl. Acad. Sci. USA* **96**: 10875–10880.
- JACKSON, R. W., J. MANSFIELD, H. AMMOUNEH, L. C. DUTTON, B. WHARTON *et al.*, 2002 Location and activity of members of a family of virPphA homologues in pathovars of Pseudomonas syringae and P. savastanoi. *Mol. Plant Pathol.* **3**: 205–215.
- JAKOBEK, J. L., J. A. SMITH and P. B. LINDGREN, 1993 Suppression of bean defense responses by Pseudomonas syringae. *Plant Cell* **5**: 57–63.
- JIN, Q., and S. Y. HE, 2001 Role of the Hrp pilus in type III protein secretion in Pseudomonas syringae. *Science* **294**: 2556–2558.
- JIN, Q., R. THILMONY, J. ZWIESLER-VOLLICK and S. Y. HE, 2003 Type III protein secretion in Pseudomonas syringae. *Microbes Infect.* **5**: 301–310.
- KELLEY, W. L., 1998 The J-domain family and the recruitment of chaperone power. *Trends Biochem. Sci.* **23**: 222–227.
- KIM, J. F., and S. V. BEER, 1998 HrpW of Erwinia amylovora, a new harpin that contains a domain homologous to pectate lyases of a distinct class. *J. Bacteriol.* **180**: 5203–5210.
- KIM, J. F., A. O. CHARKOWSKI, J. R. ALFANO, A. COLLMER and S. V. BEER, 1998 Sequences related to transposable elements and

- bacteriophages flank avirulence genes of *Pseudomonas syringae*. *Mol. Plant-Microbe Interact.* **11**: 1247–1252.
- KIM, Y. J., N. C. LIN and G. B. MARTIN, 2002 Two distinct *Pseudomonas* effector proteins interact with the Pto kinase and activate plant immunity. *Cell* **109**: 589–598.
- KIMURA, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Biol.* **16**: 111–120.
- KUMAR, S., K. TAMURA, I. B. JAKOBSEN and M. NEI, 2001 MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**: 1244–1245.
- KURTZ, J. E., F. EXINGER, P. ERBS and R. JUND, 2002 The URH1 uridine ribohydrolase of *Saccharomyces cerevisiae*. *Curr. Genet.* **41**: 132–141.
- LAWRENCE, J. G., and H. OCHMAN, 1997 Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44**: 383–397.
- LINDGREN, P. B., R. C. PEET and N. J. PANOPOULOS, 1986 Gene cluster of *Pseudomonas syringae* pv. "phaseolicola" controls pathogenicity of bean plants and hypersensitivity of nonhost plants. *J. Bacteriol.* **168**: 512–522.
- LINDGREN, P. B., N. J. PANOPOULOS, B. J. STASKAWICZ and D. DAHLBECK, 1988 Genes required for pathogenicity and hypersensitivity are conserved and interchangeable among pathovars of *Pseudomonas syringae*. *Mol. Gen. Genet.* **211**: 499–506.
- LLOYD, S. A., M. NORMAN, R. ROSQVIST and H. WOLF-WATZ, 2001 *Yersinia* YopE is targeted for type III secretion by N-terminal, not mRNA, signals. *Mol. Microbiol.* **39**: 520–531.
- MACKEY, D., B. F. HOLT, A. WIG and J. L. DANGL, 2002 RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* **108**: 743–754.
- MACKEY, D., Y. BELKHADIR, J. M. ALONSO, J. R. ECKER and J. L. DANGL, 2003 *Arabidopsis* RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell* **112**: 379–389.
- MARCHLER-BAUER, A., A. R. PANCHENKO, B. A. SHOEMAKER, P. A. THIESSEN, L. Y. GEER *et al.*, 2002 CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* **30**: 281–283.
- MARCHLER-BAUER, A., J. B. ANDERSON, C. DEWEES-SCOTT, N. D. FEDOROVA, L. Y. GEER *et al.*, 2003 CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res.* **31**: 383–387.
- MAROIS, E., G. VAN DEN ACKERKEKEN and U. BONAS, 2002 The xanthomonas type III effector protein AvrBs3 modulates plant gene expression and induces cell hypertrophy in the susceptible host. *Mol. Plant-Microbe Interact.* **15**: 637–646.
- MCGRAW, E. A., J. LI, R. K. SELANDER and T. S. WHITTAM, 1999 Molecular evolution and mosaic structure of alpha, beta, and gamma intimins of pathogenic *Escherichia coli*. *Mol. Biol. Evol.* **16**: 12–22.
- MICHELMORE, R. W., and B. C. MEYERS, 1998 Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* **8**: 1113–1130.
- MORGENSTERN, B., 1999 DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* **15**: 211–218.
- MOURY, B., C. MOREL, E. JOHANSEN and M. JACQUEMOND, 2002 Evidence for diversifying selection in potato virus Y and in the coat protein of other potyviruses. *J. Gen. Virol.* **83**: 2563–2573.
- MUDGETT, M. B., and B. J. STASKAWICZ, 1999 Characterization of the *Pseudomonas syringae* pv. tomato AvrRpt2 protein: demonstration of secretion and processing during bacterial pathogenesis. *Mol. Microbiol.* **32**: 927–941.
- NEEDLEMAN, S. B., and C. D. WUNSCH, 1970 A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**: 443–453.
- NIELSEN, R., and Z. YANG, 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **149**: 929–936.
- NIMCHUK, Z., L. ROHMER, J. H. CHANG and J. L. DANGL, 2001 Knowing the dancer from the dance: R-gene products and their interactions with other proteins from host and pathogen. *Curr. Opin. Plant Biol.* **4**: 288–294.
- OHTA, T., 1991 Role of diversifying selection and gene conversion in evolution of major histocompatibility complex loci. *Proc. Natl. Acad. Sci. USA* **88**: 6716–6720.
- ORTH, K., 2002 Function of the *Yersinia* effector YopJ. *Curr. Opin. Microbiol.* **5**: 38–43.
- PELLE, R., V. L. SCHRAMM and D. W. PARKIN, 1998 Molecular cloning and expression of a purine-specific N-ribohydrolase from *Trypanosoma brucei brucei*. Sequence, expression, and molecular analysis. *J. Biol. Chem.* **273**: 2118–2126.
- PETNICKI-OCWIEJA, T., D. J. SCHNEIDER, V. C. TAM, S. T. CHANCEY, L. SHAN *et al.*, 2002 Genomewide identification of proteins secreted by the Hrp type III protein secretion system of *Pseudomonas syringae* pv. tomato DC3000. *Proc. Natl. Acad. Sci. USA* **99**: 7652–7657.
- REID, S. D., R. K. SELANDER and T. S. WHITTAM, 1999 Sequence diversity of flagellin (fliC) alleles in pathogenic *Escherichia coli*. *J. Bacteriol.* **181**: 153–160.
- REID, S. D., C. J. HERBELIN, A. C. BUMBAUGH, R. K. SELANDER and T. S. WHITTAM, 2000 Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* **406**: 64–67.
- RUDOLPH, K., 1995 *Pseudomonas syringae* pathovars, pp. 47–138 in *Pathogenesis and Host Specificity in Plant Diseases*, edited by R. P. SINGH, U. S. SINGH and K. KOHMOTO. Elsevier Science, Oxford.
- SAWADA, H., F. SUZUKI, I. MATSUDA and N. SAITOU, 1999 Phylogenetic analysis of *Pseudomonas syringae* pathovars suggests the horizontal gene transfer of argK and the evolutionary stability of hrp gene cluster. *J. Mol. Evol.* **49**: 627–644.
- SCHNEIDER, D. S., 2002 Plant immunity and film noir: what gumshoe detectives can teach us about plant-pathogen interactions. *Cell* **109**: 537–540.
- SHAN, L., P. HE, J. M. ZHOU and X. TANG, 2000 A cluster of mutations disrupt the avirulence but not the virulence function of AvrPto. *Mol. Plant-Microbe Interact.* **13**: 592–598.
- SHAO, F., C. GOLSTEIN, J. ADE, M. STOUTEMYER, J. E. DIXON *et al.*, 2003 Cleavage of *Arabidopsis* PBS1 by a bacterial type III effector. *Science* **301**: 1230–1233.
- SHARP, P. M., and W. H. LI, 1987 The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**: 1281–1295.
- STASKAWICZ, B. J., M. B. MUDGETT, J. L. DANGL and J. E. GALAN, 2001 Common and contrasting themes of plant and animal diseases. *Science* **292**: 2285–2289.
- STOVER, C. K., X. Q. PHAM, A. L. ERWIN, S. D. MIZOGUCHI, P. WARRENER *et al.*, 2000 Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406**: 959–964.
- SULLIVAN, C. S., and J. M. PIPAS, 2002 T antigens of simian virus 40: molecular chaperones for viral replication and tumorigenesis. *Microbiol. Mol. Biol. Rev.* **66**: 179–202.
- SZUREK, B., E. MAROIS, U. BONAS and G. VAN DEN ACKERKEKEN, 2001 Eukaryotic features of the Xanthomonas type III effector AvrBs3: protein domains involved in transcriptional activation and the interaction with nuclear import receptors from pepper. *Plant J.* **26**: 523–534.
- TARR, C. L., and T. S. WHITTAM, 2002 Molecular evolution of the intimin gene in O111 clones of pathogenic *Escherichia coli*. *J. Bacteriol.* **184**: 479–487.
- THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- TSIAMIS, G., J. W. MANSFIELD, R. HOCKENHULL, R. W. JACKSON, A. SESMA *et al.*, 2000 Cultivar-specific avirulence and virulence functions assigned to avrPphF in *Pseudomonas syringae* pv. phaseolicola, the cause of bean halo-blight disease. *EMBO J.* **19**: 3204–3214.
- VENISSE, J. S., M. A. BARNY, J. P. PAULIN and M. N. BRISSET, 2003 Involvement of three pathogenicity factors of *Erwinia amylovora* in the oxidative stress associated with compatible interaction in pear. *FEBS Lett.* **537**: 198–202.
- VIVIAN, A., J. MURILLO and R. W. JACKSON, 2001 The roles of plasmids in phytopathogenic bacteria: Mobile arsenals? *Microbiology* **147**: 763–780.
- YANG, B., W. ZHU, L. B. JOHNSON and F. F. WHITE, 2000a The virulence factor AvrXa7 of *Xanthomonas oryzae* pv. *oryzae* is a type III secretion pathway-dependent nuclear-localized double-stranded DNA-binding protein. *Proc. Natl. Acad. Sci. USA* **97**: 9807–9812.

- YANG, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**: 555–556.
- YANG, Z., R. NIELSEN, A.-M. GOLDMAN and K. PEDERSEN, 2000b Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431–449.
- YEAGER, M., and A. L. HUGHES, 1999 Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunol. Rev.* **167**: 45–58.
- YUCEL, I., C. BOYD, Q. DEBNAM and N. T. KEEN, 1994 Two different classes of *avrD* alleles occur in pathovars of *Pseudomonas syringae*. *Mol. Plant-Microbe Interact.* **7**: 131–139.
- ZWIESLER-VOLLICK, J., A. E. PLOVANICH-JONES, K. NOMURA, S. BANDYOPADHYAY, V. JOARDAR *et al.*, 2002 Identification of novel *hrp*-regulated genes through functional genomic analysis of the *Pseudomonas syringae* pv. tomato DC3000 genome. *Mol. Microbiol.* **45**: 1207–1218.

Communicating editor: K. V. ANDERSON