4-2005

Proposed Guidelines for a Unified Nomenclature and Phylogenetic Analysis of Type III Hop Effector Proteins in the Plant Pathogen *Pseudomonas syringae*

Magdalen Lindeberg
Cornell University, ml16@cornell.edu

John Stavrinides
University of Toronto, john.stavrinides@uregina.ca

Jeffrey H. Chang
University of North Carolina

James R. Alfano
University of Nebraska-Lincoln, jalfano2@unl.edu

Alan Collmer
Cornell University, arc2@cornell.edu

See next page for additional authors

Authors

Magdalen Lindeberg, John Stavrinides, Jeffrey H. Chang, James R. Alfano, Alan Collmer, Jeffrey L. Dangl, Jean T. Greenberg, John W. Mansfield, and David S. Guttman

e-**X**tra*

Proposed Guidelines for a Unified Nomenclature and Phylogenetic Analysis of Type III Hop Effector Proteins in the Plant Pathogen *Pseudomonas syringae*

Magdalen Lindeberg,[1] John Stavrinides,[2] Jeffrey H. Chang,[3] James R. Alfano,[4] Alan Collmer,[1] Jeffery L. Dangl,[3] Jean T. Greenberg,[5] John W. Mansfield,[6] and David S. Guttman[2]

[1]Department of Plant Pathology, Cornell University, Ithaca, NY 14853, U.S.A.; [2]Department of Botany, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada; [3]Department of Biology and Curriculum in Genetics, University of North Carolina, Chapel Hill, 27599-3280, U.S.A.; [4]The Plant Science Initiative and the Department of Plant Pathology, University of Nebraska, Lincoln 68588-0660, U.S.A.; [5]The University of Chicago, Department of Molecular Genetics and Cell Biology, 1103 East 57th Street, Erman Biology Center, Chicago 60637 IL, U.S.A.; [6]Department of Agricultural Sciences, Imperial College London, Wye Campus, Wye, Ashford, Kent, TN25 5AH, U.K.

Submitted 6 December 2004. Accepted 10 December 2004.

Pathovars of *Pseudomonas syringae* interact with their plant hosts via the action of Hrp outer protein (Hop) effector proteins, injected into plant cells by the type III secretion system (TTSS). Recent availability of complete genome sequences for a number of *P. syringae* pathovars has led to a significant increase in the rate of effector discovery. However, lack of a systematic nomenclature has resulted in multiple names being assigned to the same Hop, unrelated Hops designated by the same alphabetic character, and failure of name choices to reflect consistent standards of experimental confirmation or phylogenetic relatedness. Therefore, specific experimental and bioinformatic criteria are proposed for proteins to be designated as Hops. A generic Hop name structure, HopXY#$_{pv \ strain}$, also is proposed, wherein family membership is indicated by the alphabetic characters, subgroup membership numerically, and source pathovar and strain in subscript. Guidelines are provided for phylogenetic characterization and name selection for Hops that are novel, related to previously characterized Hops, chimeras, pseudogenes, truncations, or nonexpressed alleles. Phylogenetic analyses of previously characterized Hops are described, the results of which have been used to guide their integration into the proposed nomenclature.

The bacterial plant pathogen *Pseudomonas syringae* comprises a large collection of pathovars and strains noted for their diverse and highly specific interactions with plant hosts. The bacterium–host interaction is governed in large part by effector proteins, which are injected into plant cells by the type III secretion system (TTSS) and are discussed in detail in recent

Corresponding author: D. S. Guttman; Telephone: 416-978-6865; Fax: 416-978-5878; E-mail: david.guttman@utoronto.ca

*The *e*-Xtra logo stands for "electronic extra" and indicates the HTML abstract available on-line contains three supplemental tables not included in the print edition.

reviews (Alfano and Collmer 2004; Chang et al. 2004; Collmer et al. 2002; Greenberg and Vinatzer 2003; Jin et al. 2003). TTSS substrates have been identified by various phenotypic characteristics over the last 20 years, but the recent availability of partial and completed genome sequences for a number of *P. syringae* pathovars has greatly increased the rate of effector discovery and the number of published, named effectors (Buell et al. 2003; Fouts et al. 1997; Greenberg and Vinatzer 2003; Guttman et al. 2002; Petnicki-Ocwieja et al. 2002; Zwiesler-Vollick et al. 2002).

The history of effector nomenclature generally reflects developments in our understanding of host–pathogen interactions in *P. syringae*. The first generation of effectors, identified by their ability to induce an avirulence reaction in plants carrying the corresponding resistance gene, was given the Avr designation. Many of these were named according to guidelines proposed by Vivian and Mansfield (1993). Following the identification of a TTSS in *P. syringae* and demonstration that virulence factors, including most of the Avr proteins, were substrates for this pathway, proteins targeted to the TTSS were given the Hrp outer protein (Hop) designation (Alfano and Collmer 1997), reflecting the Yersinia outer protein (Yop) precedence for the prototypical TTSS effectors (Cornelis and Van Gijsegem 2000).

However, the naming of individual Hop genes and proteins has not been coordinated among different research groups, leading to a situation where i) multiple names have been assigned to the same Hop, ii) unrelated Hops are designated by the same alphabetic character, iii) name choices do not reflect consistent standards of experimental determination or phylogenetic relatedness, and iv) named Hop "candidates" have proliferated, identified on the basis of conserved regulatory or targeting patterns, but lacking experimental evidence for expression or passage through the TTSS. With genome sequences already generated for three *P. syringae* pathovars and various strategies expected to generate more Hops and Hop candidates, the need for clear and consistent guidelines for Hop nomenclature has become apparent.

To address these problems, a set of standards for Hop nomenclature and name assignment are suggested. These guidelines i) specify criteria for a protein to become a named Hop, ii) codify a nomenclature system for Hops, Hop homologs, and

their chaperones, and iii) outline recommended procedures for assignment of new names. Phylogenetic analyses of previously published Hops were conducted as a first step in the implementation of the new nomenclature guidelines and are discussed here in detail. Evolutionary analyses of some of these sequences also have been recently published (Rohmer et al. 2004). The results have been used to assign Hops to homology families and subgroups, and to guide the assignment of new names.

Criteria for Hop name assignment.

The term "Hop" applies generically to expressed proteins that are secreted or translocated by the TTSS of *P. syringae* and related plant pathogens. A typical *hop* gene is preceded by an Hrp box promoter, is activated by the HrpL alternative sigma factor, and encodes a protein with N-terminal sequences that target it to the TTSS pathway for secretion in culture or translocation in planta, hereafter referred to as a TTSS targeting pattern. Because Hops are host range determinants whose presence can be specifically recognized by plants, it is likely that production of some members of a given Hop family will be disrupted in some *P. syringae* host-specific pathovars, races, or strains.

The majority of proteins that meet these criteria for expression and secretion or translocation are likely to be true effectors, with their primary function within the host cell (Cornelis and Van Gijsegem 2000). However, a subset of TTSS substrates may have a role as translocators or helpers assisting the delivery of true effectors across host cell barriers. The nomenclature system described here does not distinguish between effectors and translocators, and all proteins traveling the TTSS pathway will be considered Hops.

The best evidence for Hop name assignment is experimental verification of both expression and TTSS-dependent secretion or translocation; however, exclusive reliance on such stringent requirements would significantly delay name assignment to promising candidates. To ease the experimental burden required for Hop naming while minimizing the generation of artifactual Hops and Hop families, experimental evidence can be supplemented with computational or bioinformatic data, as described in criteria B and C (below). Homology with previously characterized Hops (criterion A) and evidence of function in planta (criterion D) also are accepted as evidence.

In order for newly identified proteins to receive a Hop designation, it is recommended that one or more of the following criteria be met.

A. Phylogenetic membership in an established Hop family and a consensus N-terminal targeting pattern. If >60% of a new protein's sequence can be significantly aligned ($e < 10^{-5}$) with one or more members of a Hop family previously characterized using criteria B, C, or D, the new protein can be given a Hop name reflecting this relationship. However, for a more accurate assessment of family and subgroup membership, the phylogenetic analyses described under "Implementation of Nomenclature Guidelines" are strongly recommended. The additional requirement for a TTSS targeting pattern (characteristics specified under criterion B) is included so that Hop names will be limited to those homologs that are plausible TTSS substrates.

If subsequent testing reveals that a protein named by this criterion is not expressed, its gene name should be modified by addition of the Greek letter psi (ψ) to reflect its status as a pseudogene. Likewise, if a protein is later found not to be secreted, an asterisk can be appended to the name, indicating that it is a nonsecreted allele (discussed in more detail in the section entitled "Name Structure and Selection").

B. Confirmed HrpL-dependence and a consensus TTSS targeting pattern. Proteins are eligible for a Hop designation if HrpL dependence or expression in bacteria during infection or in minimal media accepted as mimicking the plant apoplast is experimentally confirmed, and if an N-terminal TTSS targeting pattern is present. Although the consensus properties of this amino acid domain may change as more effectors are identified, the best information currently available indicates that most proteins targeted to the TTSS have N-terminal regions characterized by the following: i) ≥10% Ser in the first 50 amino acids, ii) Ile, Leu, Val, Ala, or Pro in the third or fourth position, and iii) no Asp or Glu residues in the first 12 amino acids (Greenberg and Vinatzer 2003; Guttman et al. 2002; Petnicki-Ocwieja et al. 2002). It is expected that effectors named using criterion B eventually will be tested for TTSS-dependent secretion or translocation. If they do not meet these criteria, their name should be changed accordingly.

C. Hrp-dependent (TTSS) secretion or translocation and evidence of expression. Proteins are eligible for a Hop designation if TTSS-dependent secretion or translocation is experimentally confirmed and there is bioinformatic or experimental evidence supporting the expression of at least one member of the Hop family. Although passage through the TTSS is the defining characteristic of Hop proteins, evidence of expression also is required to avoid the proliferation of Hop families that have been computationally identified and translocated under artificial conditions, but for which there is no evidence of actual expression. Evidence of expression should include assessment of the presence of an upstream Hrp box (Fouts et al. 2002; Innes et al. 1993; Shen and Keen 1993; Xiao and Hutcheson 1994). Accepted methods for demonstrating translocation include showing that a reporter such as ′AvrRpt2 or Cya can be translocated into the cell interior in an Hrp-dependent manner when fused to the full-length coding sequence of the candidate being tested (Casper-Lindley et al. 2002; Mudgett et al. 2000; Schechter et al. 2004).

D. Avirulence phenotype or hypersensitive response. Proteins that induce an avirulent or hypersensitive response when expressed in heterologous bacteria can be named as Hops. However, subsequent demonstration of secretion or translocation is encouraged, given that proteins such as AvrD, which direct the production of low molecular-weight elicitors (Yucel et al. 1994), are active in bacterial cells but may not be TTSS substrates.

Name structure and selection.

Generic name structure. The suggested Hop name structure, designed to reflect both phylogenetic membership in families and subgroups as well as information about the source pathovar and strain, is: HopXY#$_{pv\ strain}$.

Family. Individual Hops are grouped in homology families according to sequence relatedness. Family membership is indicated by the alphabetic character or character combination, shown in the generic name structure above as "XY".

Subgroup. Whereas orthology and paralogy relationships of Hops within a family generally are ignored due to the confounding effects of horizontal gene transfer, homology families can be further divided when the level of amino acid similarity is sufficiently high and the phylogenetic structure for these subgroups is strong. Although there is no way to determine from sequence if these subgroups are functionally distinct, they represent distinct and strongly supported evolutionary lineages, and researchers are encouraged to use phylogenetic means to assign new Hops to families and subgroups. Subgroup membership is indicated by the numerical designation following the alphabetic characters and can be determined using the guidelines listed. It is suggested that the subgroup designation "1" be assigned even in cases where the homology family is not subdivided, in order to facilitate possible future expansion of the family when additional members are identified. If a given strain is found to encode more than one member of a given family and subgroup, addition of a hyphenated

number (HopXY#-#$_{pv strain}$) will be used to distinguish between the copies. Instructions for determining family and subgroup membership are described in "Phylogenetic Characterization of Hop Protein Sequences".

Source. Pathovar and strain of the source bacterium are to be indicated in subscript. This format is generally consistent with American Society for Microbiology guidelines for gene naming, wherein genotypic designations are indicated by three-letter locus symbols, functionally related loci distinguished by capital letters, and subscripts used to distinguish genes from different organisms and strains. Pathovar abbreviations such as *Pma* and *Pto* follow the recommendations of Vivian and Mansfield (1993) and the update by D. S. Guttman published on the Pseudomonas–Plant Interaction website. Although strain designations have not been included routinely in past Avr and Hop name assignments, the anticipated sequencing of multiple strains from single pathovars necessitates their inclusion in the new nomenclature.

Nonhomologous Hops. If a Hop is confirmed by criteria B, C, or D but is not homologous to any previously identified Hop family, it will be assigned the next available alphabetic character and the subgroup designation "1", followed by a subscript indicating the pathovar and strain from which it was identified. Once an alphabetic character or character combination has been assigned, it cannot be applied to Hops outside that family. Once alphabetic characters A to Z have been assigned, naming shall proceed with AA, AB, AC…, BA, BB, BC…, and so on.

Homologs of previously named hops. If BLAST analyses reveal similarity with previously characterized Hops extending over 60% of the new protein's sequence, the new Hop can be assigned the alphabetic character of the family, followed by the appropriate pathovar and strain designation. However, researchers are strongly urged to perform the more detailed analyses described in "Phylogenetic Categorization of New Hops". In contrast to BLAST, a local alignment tool that groups sequences according to high levels of similarity in relatively small regions, the suggested procedures generate a more accurate alignment of prospective family members along their entire sequence. In addition, they provide guidelines for quantifying the degree of evolutionary similarity among family members for their assignment to subgroups. The subgroup model is especially useful when distinguishing among multiple homologs in the same strain.

Chimeras. As a general rule, Hops for which a contiguous stretch ≥40% of the coding region is unique or derived from an unrelated Hop should be considered chimeras. All chimeras thus far identified are composed of a region similar to a previously named Hop and a novel coding sequence. These have been assigned a unique alphabetic designation. If Hops are identified that appear to be chimeras composed of two previously characterized Hops, it is recommended that they be named using the alphabetic characters of both parents (e.g., a chimera Hop composed of regions from HopAB$_{pv strain}$ and HopXY$_{pv strain}$ would be named HopAB-XY$_{pv strain}$).

Pseudogenes. Nonexpressed homologs of characterized Hop families can be named according to the convention for bacterial pseudogenes wherein the gene name is preceded by the Greek letter ψ (e.g., ψhopXY1$_{pv strain}$).

Nonsecreted or translocated alleles. Homologs that are expressed but neither secreted nor translocated can be named after the Hop family and subgroup to which they are homologous followed by an asterisk (e.g., HopXY1 *$_{pv strain}$). Experimental confirmation of this phenotype is strongly encouraged, because not all Hops have an obvious N-terminal secretion or translocation domain.

Gene disruptions. Homologs of previously characterized *hop* genes that are truncated by a frame shift or premature stop codon can be indicated by addition of a single quotation mark to the *hop* gene name (i.e., *hopXY'$_{pv strain}$*). If a disrupted gene fails to be expressed, a ψ also can be added to reflect pseudogene status. Insertions by mobile genetic elements can be indicated with a double colon followed by the name of the inserted element (i.e., *hopXY*::IS*10*$_{pv strain}$).

Hop candidates. Although the range of criteria described above should eliminate the need to name Hop candidates, and understanding that naming of candidates can lead to downstream confusion, researchers who choose to name candidates are strongly urged to minimize name proliferation by depositing the sequences in GenBank under the provisional name "candidate Hop protein" and referring to them in print by their GenBank accession numbers rather than by provisional three-letter gene names. Interesting homologies or other properties can be indicated in the "Note" field in the GenBank submission. GenBank records of candidates or Hops named according to criterion C should be updated with secretion or translocation data and any name changes should be made in a timely fashion.

Hop chaperones. Three Hop chaperones have been experimentally confirmed and named with the designation "specific Hop chaperone" (Shc) followed by the letter of the first Hop for which they were shown to be a chaperone (Badel et al. 2003; van Dijk et al. 2002; Wehling et al. 2004). This designation was chosen to parallel the *Yersinia* spp. term "specific *Yersinia* chaperone" (Syc) (Wattiau and Cornelis 1993). Acceptable evidence for chaperone activity includes any method that directly demonstrates association of the chaperone and its cognate Hop, such as affinity chromatography, immunoblotting, yeast two-hybrid experiments, or crystal structure. In addition, a decrease in Hrp-dependent secretion of the chaperoned protein should be evident upon mutagenesis of the chaperone gene.

Phylogenetic analysis of previously characterized Hops.

Adoption of a new nomenclature system not only establishes suggested guidelines for naming new Hops but also provides an opportunity to rename previously identified Hops according to consistent phylogenetic standards.

Methodology.

To initiate phylogenetic analyses, homologous *P. syringae* sequences were collected from the NCBI GenBank database and the partially completed sequence of *P. syringae* pv. *phaseolicola* 1448A genome using tblastx to compare known TTSS effector protein sequences against the nr nucleotide database and a custom type III effector gene database. This search method eliminates any biases or errors due to incomplete or improper sequence annotation. Homologous protein sequences were aligned using MultiClustal (Yuan et al. 1999), which runs iterative ClustalW alignments in order to identify the optimal alignment parameters. In many cases, final multiple sequence alignments were adjusted manually. No assumptions concerning orthology or paralogy were imposed on the data due to the extensive horizontal transfer evident among the *hop* genes. Technically, most Hops probably should be considered xenologs, due to the apparent role of horizontal transfer in many, but not all, gene families (Rohmer et al. 2004).

Phylogenetic analyses were performed using neighbor-joining (NJ), maximum-likelihood (ML), and Bayesian techniques. NJ was performed in MEGA2 (ver. 2.1) (Kumar et al. 2001) using the gamma model with the gamma parameter set at 2.25, and PHYLIP 3.6a (Felsenstein 1993) using the Jones-Taylor-Thornton substitution model. Phylogenetic confidence was determined by bootstrapping, using 1,000 replicates. ML analyses were performed in PHYLIP, using the ProML module with the Jones-Taylor-Thornton model of amino acid change. Bootstrapping was performed using SeqBoot. Bayesian analyses
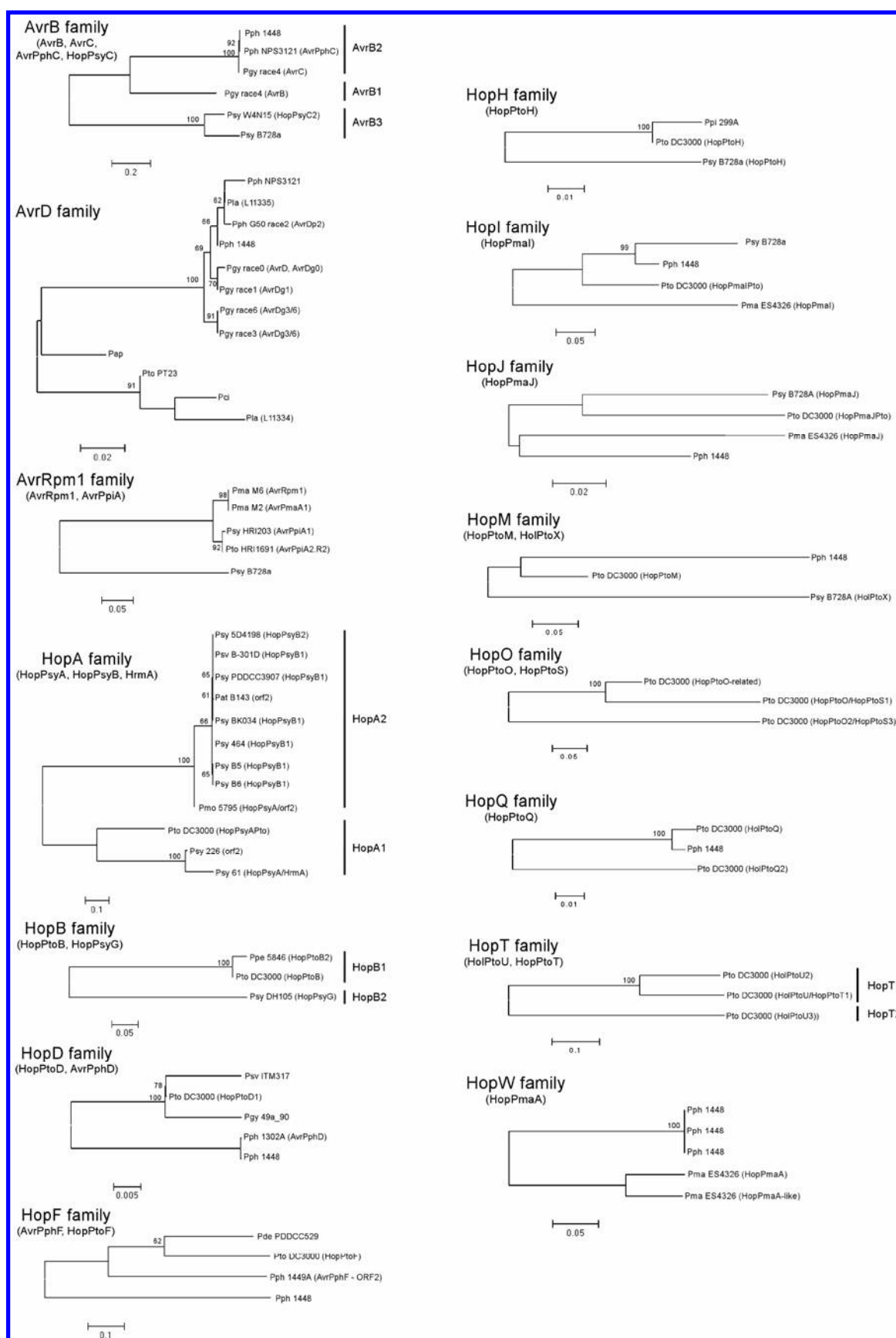
Fig. 1. Maximum likelihood phylogenetic trees for all Hrp outer protein (Hop) families with three or more members. Proposed subgroup names are demarcated by the vertical bar to the right of the phylogeny. The original name of the effector is in parentheses following the strain designation. The phylogenetic analysis was performed as described in the text. Numbers above the nodes are bootstrap scores. Only bootstrap scores >60 are presented. The horizontal line below each phylogeny indicated a genetic distance of 0.2.

were performed with MrBayes (ver. 3.04b) (Huelsenbeck and Ronquist 2001) using the Jones-Taylor-Thornton model as a prior assumption. Each inference consisted of four Markov chains starting from random trees and running for 500,000 generations. Following the discard of the initial 500 burn-in trees, 1 tree was sampled every 100 generations, and a 50% majority-rule consensus tree was generated. Posterior probabilities were calculated to determine the phylogenetic confidence for each node. Phylogenetic trees were viewed in Tree-View (Page 1996).
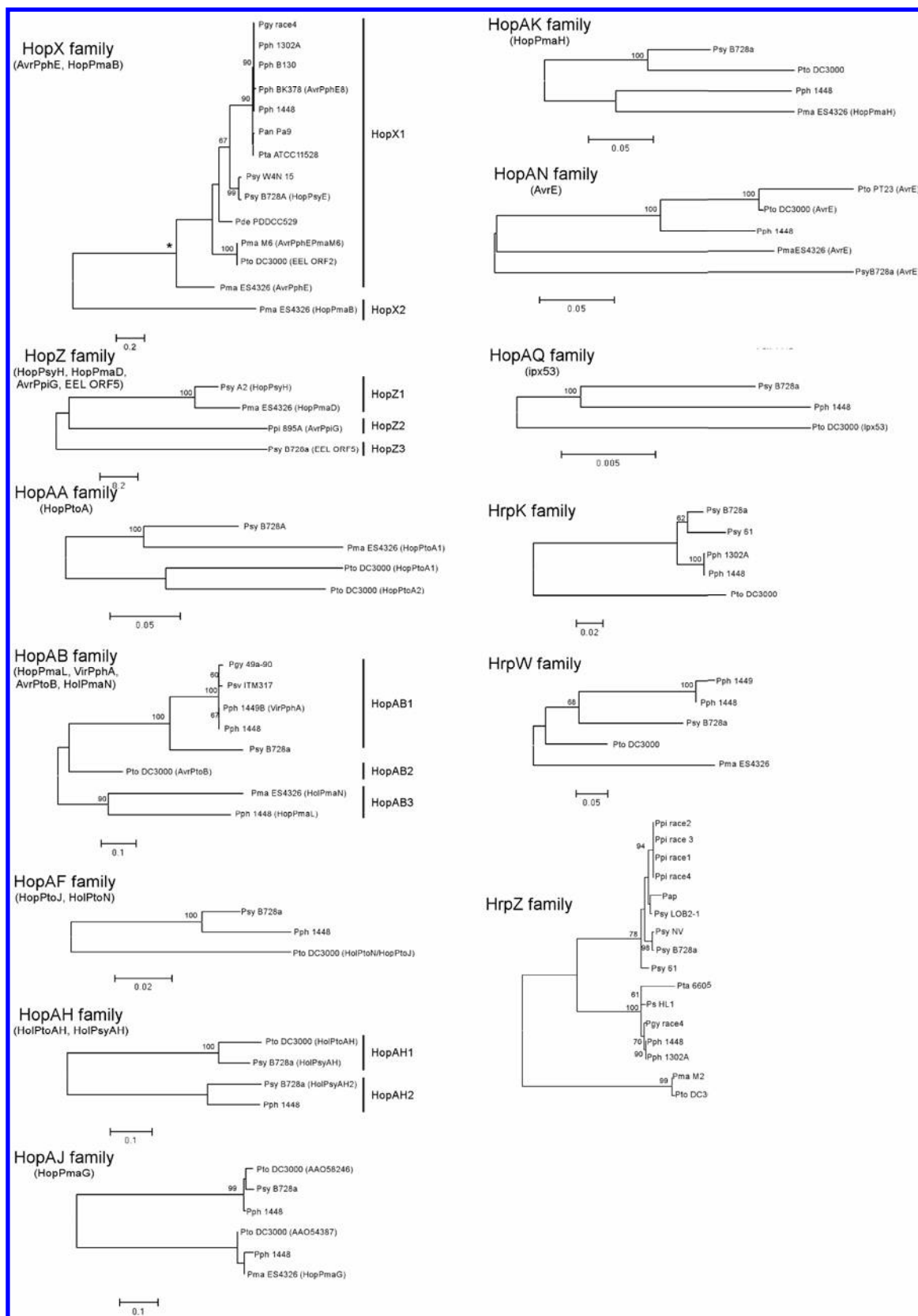

Fig. 1. *Continued.*

Effector homology families were defined based on sets of effectors that could be aligned across their entire length (global alignment). Homology families were divided into subgroups based on their phylogenetic clustering. The level of amino acid diversity within and between subgroups was used as the basis for determining whether a clade deserved a unique subgroup number within a homology family. Amino acid diversity was calculated with MEGA2 using the gamma model with the gamma parameter set at 2.25, and standard error calculation via bootstrapping.

Family and subgroup assignments of previously characterized Hops.

In nearly every case, the three methods (NJ, ML, and Bayesian) produced identical tree topologies. The only exceptions were AvrD1, HopX (AvrPphE and HopPmaB), HopA (Hop PsyA and HopPsyB1), and HrpW1, but none of these inconsistencies resulted in a significant change to the subfamily groupings. ML trees for all Hop families with three or more members are shown in Figure 1, with the exception of HopAM, for which all three members are identical.

The phylogenetic analyses were used to determine whether homology families should be subdivided into Hop subgroups based on amino acid sequence divergence. A homology family was subdivided if it was found to have within-group amino acid diversity <0.75 and between-group amino acid diversity >0.75. The only exception to this rule was in the HopAB family. These cutoff levels were chosen based on the quality of the alignments and robustness of the phylogenetic reconstruction at different levels of diversity. For those families that were divided up into subgroups, the average within-subgroup diversity was 0.23, whereas the average between-subgroup diversity was 1.26. For comparison, the average within-subgroup diversity for those families that were not subdivided was 0.13, whereas the between-subgroup diversity for these families was only 0.32. Examples of homology families that were split into subgroups include the HopZ family, which was subdivided into the HopZ1 subgroup (effectors formerly called HopPmaD and HopPsyH); the HopZ2 subgroup (AvrPpiG); and the HopZ3 subgroup (HopPsyV or EEL ORF5). Another example is the HopX family, consisting of HopX1 (AvrPphE) and HopX2 (HopPmaB) (Fig. 1). The HopAB family, which consists of Hops formerly known as VirPphA, AvrPtoB, HopPmaL, and HolPmaN, also was split into subgroups. This family proved particularly challenging because both the Pph 1448A HopPmaL allele and the Pma ES4326 HolPmaL allele appear to be truncated homologs of the Pma ES4326 HopPmaL allele, with the former similar to the C-terminal half and the latter is similar to the N-terminal half of Pma ES4326 HopPmaL. Consequently, the Pph 1448A HopPmaL and Pma ES4326 HolPmaL alleles do not overlap and, therefore, cannot be analyzed in the same phylogenetic analysis.

Many TTSS effectors share N- or C-terminal homology but are unrelated over other parts of the protein. These sequences were termed chimeric effectors and placed in separate homology families. In general, chimerics had regions with very high alignment quality but could not be aligned across their entire length. A rather arbitrary decision has been made to consider Hops that are unalignable over a contiguous stretch of at least 40% of their coding sequence as chimeras. Examples of chimeric effectors include HopK, the N-terminus of which shares similarity with AvrRps4, and HopD1, which shares similarity in the N-terminal region with HopAR1 (formerly HopPtoD2).

General conclusions are difficult to draw from the phylogenetic analyses due to the great diversity of effector families. There are families such as HopX (formerly AvrPphE and HopPmaB) where effectors appear to be evolving largely in concordance with the *P. syringae* core genome (Deng et al. 2003; Rohmer et al. 2004; Sarkar and Guttman 2004). There also are families such as HopAB in which large-scale mutational events, such as deletions of large regions of the coding sequence, appear to occur with some regularity. There are incredibly diverse families such as AvrB, which has been subdivided into three very distinct and strongly supported subgroups, and other families such as HopD (formerly HopPtoD and AvrPphD), which is extremely well conserved. If there is one fairly good generality to be made from these analyses, it is that effector evolution is a highly complex process that frequently is driven by the horizontal movement of selfish elements. These processes have left a pattern of fragmented coding sequences and chimeric molecules in their wake. They also present a tantalizing picture of the mechanisms by which new effectors evolve, and by which existing effectors can capture and co-opt other loci.

Suggested name reassignment of previously characterized Hops and Avrs.

Hops. Using the results of the phylogenetic analyses, Hops were assembled into families and subgroups and assigned names consistent with the new nomenclature guidelines. New names and subgroup designations for Hop families with three or more members are indicated in Figure 1. A comprehensive list of suggested new names for all previously published effectors, together with all known previous aliases, accession numbers, sequences, and evidence categories reflecting the criteria for Hop naming can be viewed on the Pseudomonas–Plant Interaction website. To aid in more rapid identification of new names, a table listing Hops alphabetically by their former names also is included. Tables posted on the website will be updated as more Hops are identified.

In the nomenclature system described here, most previously described Hops retain their original letter designations, with the pathovar abbreviations moved to the subscript. For example, the protein formerly published as HopPtoL is now HopL1$_{Pto DC3000}$ and HopPmaI$_{Pto}$ is HopI1$_{Pto DC3000}$. Unrelated Hops published under the same letter designation (e.g., HopPsyA, HopPtoA1, and HopPmaA) have been reviewed by their original discoverers and priority for the alphabetic character determined according to the date of original publication and number of publications related to the protein in question. As in the case of the HopAB family, we suggest that Hops found to be members of the same phylogenetic family have their names changed to reflect this relatedness.

Avrs. Integration of effectors bearing Avr names has presented a greater challenge. In addition to the obvious inconsistency with the Hop nomenclature system, Avr names imply a specific avirulence phenotype, which often is not shared by subsequently identified homologs. However, many of the Avr names are deeply imbedded in the literature and changing them would lead to a confusing break in well-established lines of investigation. In the absence of a clear consensus, it has been left to the original lab of discovery to decide whether a given Avr family should be incorporated into the proposed Hop nomenclature.

For AvrA, AvrB, AvrC, AvrD, AvrE, AvrPto, AvrRpm1, AvrRpt2, and AvrRps4, individual family members will be referred to by their original names. The family will be named after the founding Avr and subsequently identified homologs are to be named according to the family name. For example, AvrRpm1$_{Pma M2}$ will continue to be called AvrRpm1$_{Pma M2}$. It is considered part of the AvrRpm1 family and subsequent homologs will be named AvrRpm1$_{pv strain}$. Because AvrC is a member of the AvrB family, future homologs of AvrB and C will be named for AvrB.

The remaining Avrs, including previously published homologs of AvrPphB, AvrPphE, AvrPphF, VirPphA, AvrPpiB, AvrPpiC, AvrPpiG, and AvrPtoB, also will be referred to by their original names. However, the proposed family name will include both the name of the founding Avr and a new Hop designation. Subsequently identified homologs would be named according to the Hop name with reference to the founding Avr early in the text. For example, AvrPphE$_{Pto\ DC3000}$ has been assigned to the AvrPphE (HopX) family with subsequent homologs named HopX$_{pv\ strain}$.

A final group of TTSS substrates that will retain their earlier names because of an established literature and close genetic (and most likely functional) association with the Hrp system are HrpA (the Hrp pilus protein), HrpK (an apparent translocator), HrpZ (a harpin), and HrpW (another harpin). However, recently discovered harpin-like proteins encoded outside of the Hrp pathogenicity island have been given Hop designations.

Phylogenetic classification of new Hops.

To determine whether a protein sequence meets criterion A for Hop name assignment, or to select a name for Hops confirmed by criteria B, C, or D, the following steps are recommended.

Step 1: conduct BLASTP analyses to determine whether a given protein is similar to any previously characterized Hops. If it is not significantly similar to any other Hop, go to step 3. If it is significantly similar, go to step 2. In this case, significant similarity is roughly defined as a BLAST expect value of <10^{-5} and with alignment extending >60% of the length of the protein.

Step 2: at this stage, a protein can be named and published using the alphabetic characters of the family to which it is similar (guidelines have been suggested in "Name structure and selection"). However, additional phylogenetic analyses are strongly recommended in order to generate a more accurate alignment and assign the protein to a subgroup that reflects its relationship to other family members. Suggested procedures for the phylogenetic analyses are listed below, with detailed, step-by-step instructions provided online.

i) Obtain a file listing the protein sequences of all members of the Hop family in question from the Pseudomonas–Plant Interaction website. Add the new Hop sequence to the file, and perform a multiple sequence alignment using Clustal or Multalign. Clustal alignments are easily generated using the ClustalW tool available through the EMBL website. New Hops for which a contiguous stretch of ≥40% of the coding region is unique or derived from an unrelated Hop should be considered chimeras and assigned a novel alphabetic character (see step 3). Proteins that appear to be truncated or disrupted members of an established Hop family do not require novel alphabetic characters.

ii) Download the free MEGA2 package available online. Using utilities available through the pull down menu in MEGA2, convert the Clustal .aln file to MEGA format and perform neighbor-joining, minimum evolution, or maximum parsimony analysis with bootstrapping to determine how the new Hop clusters within the existing family and subgroups. Alternative phylogenetic methods such as maximum likelihood or Bayesian analysis also can be performed, although the software supporting these methods is not as transparent as MEGA2.

iii) If a new Hop phylogenetically clusters with an existing family and subgroup, it should be named using the same alphabetic character and subgroup number as its homologs.

iv) Alternatively, if the new Hop diverges from existing subgroups within a family, use MEGA2 to calculate the genetic distance of the new Hop from the established subgroups. As a general rule, when genetic distance is calculated using the gamma model, with the gamma parameter set at 2.25, the distance between Hops within a subgroup will be less than 0.75. Those in separate subgroups will have a distance greater than 0.75.

v) If it is determined that the Hop warrants creation of a new subgroup, it should be named with the alphabetic character of the homologous Hop family followed by the next available numeric designation. Proceed to step 4.

Step 3: in the absence of homology to a pre-existing Hop family, confirm Hop identity using criteria B, C, or D. Go to the list of Hop names online and select the next available alphabetic character and subgroup number of "1".

Step 4: submit the sequence and name of the new Hop to GenBank for immediate release.

Priority for name assignment.

Priority for naming will be determined according to the date on which the sequences and names of new Hops are made public in the GenBank database. Release via GenBank ensures that anyone subsequently identifying the same Hop will learn of the prior name assignment during routine BLAST searches. Researchers who do not wish their sequences to be public for an extended time before publication are free to delay both official name assignment and sequence release until the time of manuscript submission.

Database management.

The new nomenclature has been designed to ensure that newly discovered effectors are systematically assigned unique names. To minimize confusion over the names for previously characterized effectors, a database of all known effectors including former names and aliases is available on the Pseudomonas–Plant Interaction website. Names also have been updated in the *Pseudomonas syringae* pv. *tomato* DC3000 genome annotation at GenBank. Given that GenBank remains the ultimate repository of information, and that only the original depositors can change records, it is requested that other research groups similarly update their previous GenBank depositions.

LITERATURE CITED

Alfano, J. R., and Collmer, A. 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., and Collmer, A. R. 2004. Type III secretion system effector proteins: Double agents in bacterial disease and plant defense. Annu. Rev. Phytopathol. 42:385-414.

Badel, J. L., Nomura, K., Bandyopadhyay, S., Shimizu, R., Collmer, A., and He, S. Y. 2003. *Pseudomonas syringae* pv. *tomato* DC3000 HopPtoM (CEL ORF3) is important for lesion formation but not growth in tomato and is secreted and translocated by the Hrp type III secretion system in a chaperone-dependent manner. Mol. Microbiol. 49(5):1239-1251.

Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T., Gwinn, M. L., Dodson, R. J., Deboy, R. T., Durkin, A. S., Kolonay, J. F., Madupu, R., Daugherty, S., Brinkac, L., Beanan, M. J., Haft, D. H., Nelson, W. C., Davidsen, T., Zafar, N., Zhou, L., Liu, J., Yuan, Q., Khouri, H., Fedorova, N., Tran, B., Russell, D., Berry, K., Utterback, T., Van Aken, S. E., Feldblyum, T. V., D'Ascenzo, M., Deng, W. L., Ramos, A. R., Alfano, J. R., Cartinhour, S., Chatterjee, A. K., Delaney, T. P., Lazarowitz, S. G., Martin, G. B., Schneider, D. J., Tang, X., Bender, C. L., White, O., Fraser, C. M., and Collmer, A. 2003. The complete genome sequence of the Arabidopsis and tomato pathogen

Pseudomonas syringae pv. *tomato* DC3000. Proc. Natl. Acad. Sci. U.S.A. 100(18):10181-10186.

Casper-Lindley, C., Dahlbeck, D., Clark, E. T., and Staskawicz, B. J. 2002. Direct biochemical evidence for type III secretion-dependent translocation of the AvrBs2 effector protein into plant cells. Proc. Natl. Acad. Sci. U.S.A. 99(12):8336-8341.

Chang, J. H., Goel, A. K., Grant, S. R., and Dangl, J. L. 2004. Wake of the flood: Ascribing functions to the wave of type III effector proteins of phytopathogenic bacteria. Curr. Opin. Microbiol. 7(1):11-18.

Collmer, A., Lindeberg, M., Petnicki-Ocwieja, T., Schneider, D. J., and Alfano, J. R. 2002. Genomic mining type III secretion system effectors in *Pseudomonas syringae* yields new picks for all TTSS prospectors. Trends Microbiol. 10(10):462-469.

Cornelis, G. R., and Van Gijsegem, F. 2000. Assembly and function of type III secretory systems. Annu. Rev. Microbiol. 54:735-774.

Deng, W. L., Rehm, A. H., Charkowski, A. O., Rojas, C. M., and Collmer, A. 2003. *Pseudomonas syringae* exchangeable effector loci: Sequence diversity in representative pathovars and virulence function in *P. syringae* pv. *syringae* B728a. J. Bacteriol. 185(8):2592-2602.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package). Distributed by author. Department of Genetics, University of Washington, Seattle.

Fouts, D. E., Abramovitch, R. B., Alfano, J. R., Baldo, A. M., Buell, C. R., Cartinhour, S., Chatterjee, A. K., D'Ascenzo, M., Gwinn, M. L., Lazarowitz, S. G., Lin, N. C., Martin, G. B., Rehm, A. H., Schneider, D. J., van Dijk, K., Tang, X., and Collmer, A. 2002. Genomewide identification of *Pseudomonas syringae* pv. *tomato* DC3000 promoters controlled by the HrpL alternative sigma factor. Proc. Natl. Acad. Sci. U.S.A. 99(4):2275-2280.

Fouts, D. E., True, H. L., and Celander, D. W. 1997. Functional recognition of fragmented operator sites by R17/MS2 coat protein, a translational repressor. Nucleic Acids Res. 25:4464-4473.

Greenberg, J. T., and Vinatzer, B. A. 2003. Identifying type III effectors of plant pathogens and analyzing their interaction with plant cells. Curr. Opin. Microbiol. 6(1):20-28.

Guttman, D. S., Vinatzer, B. A., Sarkar, S. F., Ranall, M. V., Kettler, G., and Greenberg, J. T. 2002. A functional screen for the type III (Hrp) secretome of the plant pathogen *Pseudomonas syringae*. Science 295(5560):1722-1726.

Huelsenbeck, J. P., and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Innes, R. W., Bent, A. F., Kunkel, B. N., Bisgrove, S. R., and Staskawicz, B. J. 1993. Molecular analysis of avirulence gene *avrRpt2* and identification of a putative regulatory sequence common to all known *Pseudomonas syringae* avirulence genes. J. Bacteriol. 175:4859-4869.

Jin, Q., Thilmony, R., Zwiesler-Vollick, J., and He, S. Y. 2003. Type III protein secretion in *Pseudomonas syringae*. Microbes Infect. 5(4):301-310.

Kumar, S., Tamura, K., Jakobsen, I. B., and Nei, M. 2001. MEGA2: Molecular evolutionary genetics analysis software. Bioinformatics 17(12):1244-1245.

Mudgett, M. B., Chesnokova, O., Dahlbeck, D., Clark, E. T., Rossier, O., Bonas, U., and Staskawicz, B. J. 2000. Molecular signals required for type III secretion and translocation of the *Xanthomonas campestris* AvrBs2 protein to pepper plants. Proc. Natl. Acad. Sci. U.S.A. 97(24):13324-13329.

Page, R. D. 1996. TreeView: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12:357-358.

Petnicki-Ocwieja, T., Schneider, D. J., Tam, V. C., Chancey, S. T., Shan, L., Jamir, Y., Schechter, L. M., Janes, M. D., Buell, C. R., Tang, X.,

Collmer, A., and Alfano, J. R. 2002. Genomewide identification of proteins secreted by the Hrp type III protein secretion system of *Pseudomonas syringae* pv. *tomato* DC3000. Proc. Natl. Acad. Sci. U.S.A. 99(11):7652-7657.

Rohmer, L., Guttman, D. S., and 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.

Sarkar, S. F., and Guttman, D. S. 2004. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. Appl. Environ. Microbiol. 70:1999-2012.

Schechter, L. M., Roberts, K. A., Jamir, Y., Alfano, J. R., and Collmer, A. 2004. *Pseudomonas syringae* type III secretion system targeting signals and novel effectors studied with a Cya translocation reporter. J. Bacteriol. 186(2):543-555.

Shen, H., and Keen, N. T. 1993. Characterization of the promoter of avirulence gene D from *Pseudomonas syringae* pv. *tomato*. J. Bacteriol. 175:5916-5924.

van Dijk, K., Tam, V. C., Records, A. R., Petnicki-Ocwieja, T., and Alfano, J. R. 2002. The ShcA protein is a molecular chaperone that assists in the secretion of the HopPsyA effector from the type III (Hrp) protein secretion system of *Pseudomonas syringae*. Mol. Microbiol. 44(6):1469-1481.

Vivian, A., and Mansfield, J. 1993. A proposal for a uniform genetic nomenclature for avirulence genes in phytopathogenic pseudomonads. Mol. Plant-Microbe Interact. 6:9-10.

Wattiau, P., and Cornelis, G. R. 1993. SycE, a chaperone-like protein of *Yersinia enterocolitica* involved in the secretion of YopE. Mol. Microbiol. 8:123-131.

Wehling, M. D., Guo, M., Fu, Z. Q., and Alfano, J. R. 2004. The *Pseudomonas syringae* HopPtoV protein is secreted in culture and translocated into plant cells via the type III protein secretion system in a manner dependent on the ShcV type III chaperone. J. Bacteriol. 186:3621-3630.

Xiao, Y., and Hutcheson, S. 1994. A single promoter sequence recognized by a newly identified alternate sigma factor directs expression of pathogenicity and host range determinants in *Pseudomonas syringae*. J. Bacteriol. 176:3089-3091.

Yuan, J., Amend, A., Borkowski, J., DeMarco, R., Bailey, W., Liu, Y., Xie, G., and Blevins, R. 1999. MULTICLUSTAL: A systematic method for surveying Clustal W alignment parameters. Bioinformatics 15(10):862-863.

Yucel, I., Midland, S. L., Sims, J. J., and Keen, N. T. 1994. Class I and class II *avrD* alleles direct the production of different products in gram-negative bacteria. Mol. Plant-Microbe Interact. 7:148-150.

Zwiesler-Vollick, J., Plovanich-Jones, A. E., Nomura, K., Bandyopadhyay, S., Joardar, V., Kunkel, B. N., and He, S. Y. 2002. Identification of novel *hrp*-regulated genes through functional genomic analysis of the *Pseudomonas syringae* pv. *tomato* DC3000 genome. Mol. Microbiol. 45(5):1207-1218.

AUTHOR-RECOMMENDED INTERNET RESOURCES

EMBL-European Bioinformatics Institute clustalw webpage:
  www.ebi.ac.uk/clustalw
Journal of Bacteriology genetic nomenclature guidelines:
  jb.asm.org/misc/itoa.pdf
Molecular Evolutionary Genetics Analysis (MEGA) website:
  www.megasoftware.net
Pseudomonas–Plant Interaction website: pseudomonas-syringae.org