Symposium Session VI : Molecular Plant-Microbe Interactions II

SVI-1

Pseudomonas syringae: Functional genomics and plant pathogenicity

Alan Collmer(1), C. Robin Buell(2), Samuel Cartinhour(3), Candace W. Collmer(4), Mark D'Ascenzo(5), Vinita Joardar(2), Magdalen Lindeberg(1), and David J. Schneider(3)

(1)Cornell University, Ithaca, NY 14853, USA; (2)The Institute for Genomic Research, Rockville, MD 20850, USA; (3)U. S. Department of Agriculture - Agricultural Research Service, Ithaca, NY 14853, USA; (4)Wells College, Aurora, NY 13026 USA; (5)Boyce Thompson Institute for Plant Research, Ithaca, NY 14853, USA.

Introduction

Pseudomonas syringae is a bacterial pathogen noted for its diverse interactions with plants. Virulence effector proteins injected into plant cells by the type III secretion system appear to be key to *P. syringae*-plant interactions, and TTSS pathway *hrp* mutants are unable to elicit the defense-associated hypersensitive response (HR) in resistant plants or be pathogenic in host plants (1). Strains of *P. syringae* are divided into more than 50 pathovars based largely on host specificity, and phylogenetic analyses reveal that most pathovars cluster into three groups (2). Strains representing each of the three groups have now been sequenced. These strains are *P. s. tomato* DC3000 (bacterial speck of tomato and Arabidopsis), *P. s. syringae* B728A (brown spot of bean), and *P. s. phaseolicola* 1448A (halo blight of bean). DC3000 and 1448A were sequenced by The Institute for Genomic Research (TIGR) and B728a by the U.S. Department of Energy Joint Genome Initiative. DC3000 was sequenced first. An 8x draft of the genome was obtained by TIGR in January 2001 and immediately released for functional genomics research by the community, which enhanced subsequent annotation of virulence genes when the genome sequence was completed (3).

Use of the draft sequence of the *P. s. tomato* DC3000 genome to discover genes encoding TTSS effectors and other candidate virulence factors

TTSS effectors are known as Avr (avirulence) or Hop (Hrp outer proteins) proteins depending on the phenotype supporting their discovery (4). The avirulence phenotype is observed when an effector gene, heterologously expressed in a normally virulent strain, triggers the HR in a test plant carrying a cognate resistance *R* gene. The Hop phenotype is observed when an immunoblot analysis (often facilitated by an epitope tag on the test protein) reveals TTSS-dependent secretion in culture or when a translocation reporter reveals TTSS-dependent injection of a hybrid protein into plant cells during infection.

These methods provide ways to determine if a candidate Avr/Hop can travel the TTSS pathway, but they are too tedious for comprehensive genomewide identification of effectors. Furthermore, identification of TTSS effector genes by loss-of-virulence mutant screens has not been effective, apparently because of redundancy among the effectors. However, secretion/translocation assays can be used effectively with a subset of ORFs that have been identified by computational approaches as being Hop candidates worthy of such experimental attention. We employed an approach that involved several iterations of experimental and computational methods and exploited multiple characteristics of *hop* genes and their products to find such candidates, as depicted in Fig.1.

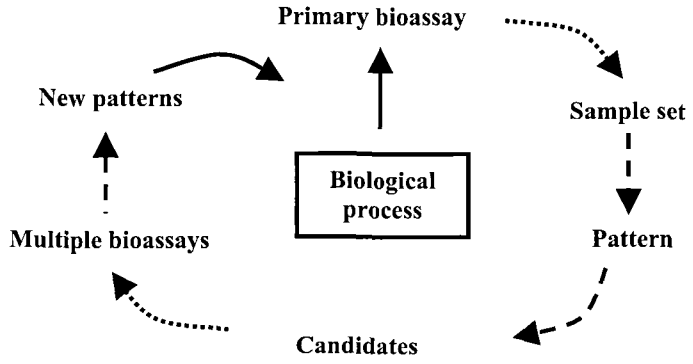

Fig. 1. The use of sequence patterns to identify in a bacterial genome complete sets of factors with similar function in a highly multifactorial process like pathogenesis. The approach is centered on the identification and characterization of factors underlying a target process, such as TTSS-dependent interactions with plants, and it involves cycles of experimental (dotted arrow) and computational (dashed arrow) methods. A primary bioassay, such as gain-of-function screens for *avr* genes, yields a sample set in which a pattern (e.g., Hrp box promoters) can be discerned. The pattern is then used to search the genome sequence for all candidates. The candidates can then be subjected to multiple assays (e.g., TTSS-dependent translocation), which can then be used to identify new patterns (e.g., TTSS targeting patterns). The new patterns can yield insights into biological function and additional candidates through further searches of the genome.

TTSS effector candidates were identified in the draft sequence of the DC3000 genome on the basis of the presence of "Hrp box" promoter sequences, which are activated by the HrpL alternative sigma factor and were previously found in association with *avr* genes characterized from various *P. syringae* strains. Secretion/translocation assays confirmed that several of the candidate novel effector genes in DC3000 encoded novel Hops (5). Our analysis of this newly enlarged set of *P. syringae* Hop proteins revealed TTSS substrate-associated patterns in the first 50 amino acids, including an aliphatic amino acid in position 3 or 4, a lack of acidic amino acids in the first 12 positions, and a high proportion of polar amino acids in the first 50 positions (6). Testing of additional ORFs associated with Hrp box promoters further confirmed the value of these patterns in predicting TTSS substrates (7). Through this combination of experimental and computational approaches with the draft genome of DC3000, genes for 38 confirmed TTSS substrates were identified (3).

The DC3000 genome contains 298 genes that we could identify as being implicated in virulence based on our analysis of TTSS effector genes and on previously reported experiments with DC3000 and other strains of *P. syringae* and similar plant pathogens (3). This includes genes associated with alginate, coronatine, and indole acetic acid production as well as motility, adhesion, degradation of plant cell walls, iron acquisition, and epiphytic growth. Importantly, 811 genes of unknown function were identified that were absent from the sequenced genomes of *P. aeruginosa* PAO1 and *P. putida* KT2440, which are not plant pathogens (3). These genes represent a new set of candidates worthy of special attention in exploring the genetic basis of *P. syringae*-plant interactions.

P. syringae genome biology resources for functional genomics research

To foster comparative analyses of multiple *P. syringae* genomes, we have sequenced to completion the genome of *P. s. phaseolicola* 1448A. A 9x draft sequence was made immediately available by TIGR to the research community through http://pseudomonas-syringae.org in December 2003. The closed genome sequence was similarly released in August 2004 and is now being annotated. The sequencing of *P. s. tomato* DC3000 and *P. s. phaseolicola* 1448A was part of the Pseudomonas-Plant Interaction (PPI) project, which is supported by the NSF Plant Genome Research Program and aims to provide a number of resources to the research community. These are available through the TIGR Comprehensive Microbial Resource, the PPI website http://pseudomonas-syringae.org, and a USDA/ARS website http://monod.cornell.edu developed by the computational biologists who led the pattern-based identification of TTSS effector genes in DC3000. The PPI site focuses more on community resources for functional genomics, and it features general information on *P. syringae*, a community registry, a database for TTSS effectors, a guide to various genome analysis tools, a genome viewing and community-based annotation tool, and instructions for using Artemis software (available through the Sanger Institute website) and special input files that can be downloaded from http://pseudomonas-syringae.org to enable viewing of the DC3000 genome from a pathobiology perspective.

Ongoing annotation is another important function of the PPI website, and this is supported through PeerGAD, a web-based, database-driven application that allows community-wide, peer-reviewed ongoing annotation of genome sequences (8). Importantly, PeerGAD supports the use of Gene Ontology (GO) terms in annotation. The goal of the GO Consortium is to "produce a controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells is accumulating and changing" (http://www.geneontology.org). The three structured networks of GO terms address the molecular function, biological process, and cellular component of individual gene products. Their use in the annotation of pathogen genomes could enable researchers working with Oomycetes, fungi, nematodes, and various bacteria to rapidly compare (by function rather than sequence similarity) genes involved in a specific pathogenic process, such as adhesion or plant cell wall degradation. PPI project members are working with the GO Consortium and other pathogen genome teams to develop GO terms for pathogenic processes.

The genetics underlying the virulence of *P. s. tomato* DC3000 are quite complex, which suggests that a satisfying understanding will require extensive comparative genomic analyses, the ongoing effort of a community of researchers applying both computational and experimental methods, and internet resources to enhance the sharing and interpretation of this new wealth of information.

References

1. Alfano, J. R. & Collmer, A. (2004) *Annu. Rev. Phytopathol.* **42,** 385-414.
2. Sawada, H., Suzuki, F., Matsuda, I. & Saitou, N. (1999) *J. Mol. Evol.* **49,** 627-644.
3. 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., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100,** 10181-10186.
4. Collmer, A., Lindeberg, M., Petnicki-Ocwieja, T., Schneider, D. J. & Alfano, J. R. (2002) *Trends in Microbiol.* **10,** 462-469.
5. 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., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 2275-2280.
6. Petnicki-Ocwieja, T., Schneider, D. J., Tam, V. C., Chancey, S. T., Shan, L., Jamir, Y., Schechter, L. M., Buell, C. R., Tang, X., Collmer, A., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 7652-7657.
7. Schechter, L. M., Roberts, K. A., Jamir, Y., Alfano, J. R. & Collmer, A. (2004) *J. Bacteriol.* **186,** 543-555.
8. D'Ascenzo, M. D., Collmer, A. & Martin, G. B. (2004) *Nucleic Acids Res.* **32,** 3124-3135.

Acknowledgments

This work was supported by NSF Plant Genome Research Program Cooperative Agreement DBI-0077622.