

PLANT-ASSOCIATED MICROBE GENOME INITIATIVE

White Paper

Developed by the Plant Microbiology Research Community

**In association with the
APS Public Policy Board**

October 1, 2002

Drafting Committee:

Jan E. Leach

Plant Pathology, Kansas State University,
4024 Throckmorton,
Manhattan, KS 66506,
jeleach@ksu.edu

Scott Gold

Plant Pathology, University of Georgia,
Athens, GA 30602,
sgold@arches.uga.edu

Sue Tolin

Plant Molecular Biology Lab,
Room 102, 435 Old Glade Road,
Blacksburg, VA,
stolin@vt.edu

Kellye Eversole

Eversole Associates,
3208 Park View Rd,
Chevy Chase, MD 20815,
kellye@eversoleassociates.com

This document and listings of the meeting participants can be found at:

<http://www.apsnet.org/members/ppb/top.asp>

Executive Summary

Plant-associated microorganisms are critical to agricultural and food security and are key components in maintaining the balance of our ecosystems. Some of these diverse microbes, which include viruses, bacteria, oomycetes, fungi and nematodes, cause plant diseases, while others prevent diseases or enhance plant growth. Despite their importance, we know little about them on a genomic level. To intervene in disease and understand the basis of biological control or symbiotic relationships, a concerted and coordinated genomic analysis of these microbes is essential. Genome analysis, in this context, refers to the structural and functional analysis of the microbe DNA including the genes, the proteins encoded by those genes, as well as noncoding sequences involved in genome dynamics and function. The ultimate emphasis is on understanding genomic functions involved in plant-associations.

Members of the American Phytopathological Society (APS) developed a prioritized list of plant-associated microbes for genome analysis. With this list as a foundation for discussions, a *Workshop on Genomic Analysis of Plant-Associated Microorganisms* was held in Washington D.C. on April 9-11, 2002. The workshop was organized by the Public Policy Board of APS, and was funded by the Department of Energy (DOE), the National Science Foundation (NSF), USDA-Agricultural Research Service (USDA-ARS) and USDA-National Research Initiatives (USDA-NRI). The workshop included academic, industrial, and governmental experts from the genomics and microbial research communities and observers from the federal funding agencies. After reviewing current and near-term technologies, workshop participants proposed a comprehensive, international initiative to obtain the genomic information needed to understand these important microbes and their interactions with host plants and the environment.

Specifically, the recommendations call for a 5-year, \$500 million international public effort for genome analysis of plant-associated microbes. The goals are to (1) obtain genome sequence information for several representative groups of microbes; (2) identify and determine function for the genes/proteins and other genomic elements involved in plant-microbe interactions; (3) develop and implement standardized bioinformatic tools and a database system that is applicable across all microbes; and (4) educate and train scientists with skills and knowledge of biological and computational sciences who will apply the information to the protection of our food sources and environment.

Overview

Plant-associated microbes play critical roles in agricultural and food safety and security and in the maintenance of ecosystem balance. Despite their importance, as of September, 2002, less than six percent of the microbes whose genomic sequence has been completed and made publicly available were plant-associated microbes (<http://www.tigr.org/~vinita/PPwebpage.html>, http://www.jgi.doe.gov/JGI_microbial/html/index.html). To remedy this dearth of critical information, the microbe-plant interaction community, lead by the Public Policy Board of the American Phytopathological Society (APS), initiated a planning effort to develop a strategy to undertake the structural and functional genomic analysis of viruses, bacteria, oomycetes, fungi, and nematodes that interact with plants. This white paper, which is the product of a *Workshop on Genome Analysis of Plant-Associated Microbes* held in Washington D.C. (April 9-11, 2002), describes the rationale and recommendations for an international public initiative for genomic analysis of these plant-associated microbes.

Background

Microbes can be found on or within every higher organism, including plants. Plant-associated microbes play critical roles in plant health. Some microbes cause diseases whereas others prevent diseases or enhance plant growth. Microbes also interact with each other on or within plants with the result of more or less severe disease. Within each microbe group, the diversity of species that interacts with plants is tremendous. For example, more than 10,000 pathogenic species of fungi and viruses are estimated to cause diseases on plants, many of which are economically important. Pathogenic microbes continually erode our supply of food and fiber, causing losses to producers of over \$200 billion annually. Recent terrorist activities have heightened concern that, in the wrong hands, these agents could be used to threaten our food, environment or economic security (Schaad *et al.* 1999, <http://www.apsnet.org/online/feature/BioSecurity/Top.html>; Countering Agricultural Bioterrorism, 2002,

National Research Council, <http://www.nap.edu/catalog/10505.html>). The historic devastation wrought by the inadvertent introduction of exotic pathogens testifies to such potential. For example, the introduction of the chestnut blight pathogen, *Cryphonectria parasitica*, from Europe in 1904 effectively eliminated from North America the American chestnut, an important and beautiful timber tree. Similar potential damage is currently occurring with disease epidemics of forest trees including sudden oak death caused by *Phytophthora ramorum* and pitch canker of pine caused by *Fusarium circinatum*. Introductions of other pathogenic microbes have caused tremendous suffering to humans, including the famines that resulted from the *Phytophthora* late blight epidemic of potato in 1845-1846 in Ireland, and the *Cochliobolus* brown spot epidemic of rice in Bengal in 1943.

Despite their importance to our agricultural base and food security, we know little of the molecular genetics of these microbes. Our understanding of how these microbes cause disease or benefit plants is rudimentary, at best. With the recent advent of genomics, which includes the structural and functional analysis of the microbe genes and the proteins encoded by those genes, we see the opportunity to understand the nature of beneficial and pathogenic microbe-plant-associations at levels never before possible.

The Promise of Genome Analysis of Plant-Associated Microbes

Due to powerful automated sequence technologies and advances in bioinformatics, the task of sequencing entire genomes of organisms is now routine. The sequenced genomes of a relatively few model organisms are already enabling a wide variety of new discoveries, including new genes and metabolic pathways and insights into the mechanisms of microbial pathogenesis. The continuous development of ever improving bioinformatics tools is enhancing the discovery power of these sequences. With these exciting advances, the means to generate sequence and functional information and the tools to use that information to understand the basic biology of microorganisms that cause or prevent diseases on plants are now available.

Genome analyses will enhance our knowledge base exponentially and thereby provide tools to abrogate the problems caused by plant pathogenic microbes through genetically-based approaches and allow development of improved beneficial microbes. Historically, the molecular basis for interactions between plants and microbes was studied using a gene-by-gene approach and host resistance was a major control approach. Now, coupled with the increasing availability of plant host genomic sequences, structural and functional genomic analysis of plant-associated microbes will increase the speed of identification of genes involved in host-pathogen interactions and will allow genome-wide approaches to understanding the role of a gene or pathway in interactions with plants. Some genes will potentially be useful as sources of pathogen-derived resistance as has already been demonstrated for many viral diseases. Other genes may be involved in recognition to trigger host defense responses to the pathogen.

Comparisons of genomes of related strains or species will provide an understanding of the evolution of microbial genomes, particularly as they evolve in associations with plants. Researchers exploiting comparative genomics will be able to discern the molecular basis of how some microbes have evolved to form intimate biotrophic associations with plant cells, whereas others inhabit intercellular spaces and still others colonize only the vascular systems of their hosts. Comparisons of sequences within and between species will also provide information to develop accurate diagnostic tools. Such precision in diagnostics may be essential to track the source or origin of a pathogen (e.g., as for the anthrax strains recently released in the US, or plum pox virus recently introduced). This information may be essential for forensic purposes as well as to understand the genetic potential of the pathogen to become established or to cause an epidemic. Additionally, because of the evolutionary conservation between microbes of different genera, and because of the exchange of genetic information between some groups of microbes, genome analysis of plant-associated microbes will contribute to our understanding of how animal pathogens cause disease. For example bacterial genes involved in plant disease induction exhibit a great deal of similarity to bacterial genes involved in disease induction in animals and humans (Keen *et al.*, 2000, <http://www.pnas.org/cgi/content/full/97/16/8752>). Thus, identifying the full complement of genes involved in pathogenesis will help decipher the disease strategies of these organisms and will guide design of new control measures for use in both plants and animals.

The knowledge of how microbes interact with plants is essential to the development of effective and environmentally-sound, chemically-based strategies for disease control. In the US, over \$600 million

are spent annually on agricultural fungicides alone. Similar to what is being observed with the heavy use of antibiotics targeted to human or animal pathogens, changes in pathogen biology more and more frequently render some of these chemicals ineffective. Furthermore, increased regulatory policies are restricting the use of existing agrochemicals for pathogen control. With genome information, multiple tactics for control of pathogens could be developed. For example, the identification of more precise targets in the pathogen may allow for design of more specific and effective chemicals that are environmentally benign.

In sum, functional, structural and comparative genome analyses will provide insights into how microbes infect and establish either disease or beneficial interactions with host plants, how they reproduce and spread in the plants and how they persist and are disseminated in the environment. This information will be critical for the development of more effective, environmentally benign control measures and for accurate diagnostic and tracking methods.

Which plant-associated microbes should be sequenced?

Even with the costs of genome sequencing decreasing, it is clearly not feasible to sequence genomes of all plant-associated species at this time. Through a series of task forces with broad input, the APS established a prioritized list of species for genome sequencing. This list, published in a White Paper entitled “*Microbial Genomic Sequencing: Perspectives of the American Phytopathological Society*” (<http://www.apsnet.org/members/ppb/ps/top.asp>), was developed using the following criteria:

- Economic importance and relevance to U.S. and world-wide agriculture and forestry;
- Unique biological or environmental features;
- Broad interest to a significant community of scientists or agriculturists;
- Genetic tractability (i.e., the ease with which genetic studies, such as crosses or genome modifications, can be performed); and
- Availability of tools and other biological resources (e.g., gene libraries, genetic maps, and the genetic tractability of the host so that the system could be addressed from both sides of the host-pathogen interaction, etc.).

The intention at its publication was that the APS list, compiled in 2000-2001, would be subject to change. A forum was held at the APS meetings in July, 2002 to discuss the list criteria and content, and define a means to periodically update the list. As a result of these discussions, and to encourage even wider input and a more accurate up-to-date list, a mechanism was established whereby the research community can revisit the list criteria and content, and make recommendations through the chairs of each pertinent APS subject matter committee (<http://www.apsnet.org/members/com/reports.asp>). The subject matter committees will annually evaluate the list. In addition, input is invited on the need for specialized lists, such as the list of organisms that are potential introduced threats to U.S. agriculture (<http://www.apsnet.org/online/feature/exotic/>).

Community discussions preceding this white paper

A workshop was held in Washington, D.C. April 9-11, 2002 to develop a strategy to expedite genome analysis of plant-associated microbes. The workshop was organized by the APS Public Policy Board, and was funded by the Department of Energy (DOE), the National Science Foundation (NSF), USDA-Agricultural Research Service (USDA-ARS) and USDA-National Research Initiative (USDA-NRI). Participants included academic, industrial, and governmental experts from the genomics and microbial research communities and observers from the federal funding agencies. The participants agreed on several key points:

- The lack of publicly available information on the genomes of plant-associated microbes is a major impediment to research, and, as a result, to crop biosecurity;
- The APS list of microbes for sequencing is a good starting point, but requires periodic review and updating;
- Current technological and biological resources would allow for sequence analysis of a significant number of microbes from this very diverse group, given appropriate funding resources;

- The availability of genome sequence would enable and enhance the development of tools for functional and comparative genomics; and
- There is an urgent need for development and employment of flexible and broadly applicable bioinformatic tools and the establishment and continual updating of databases and genetic resource centers.

An advisory committee was appointed to oversee the development of a *Plant-Associated Microbe Genome Initiative*. Discussions toward this initiative were broadened by two means. Firstly, a draft of this position paper was published in several relevant society newsletters and through email lists, and the research community was invited to send ideas or comments by email (microbegenomics@scisoc.org). These comments were posted at the APS PPB website (<http://www.apsnet.org/members/ppb/top.asp>). Secondly, a public open forum was held at the Annual APS meetings in Milwaukee, WI (July 27, 2002). At this forum, the Initiative recommendations were summarized and feedback on them solicited. All comments and suggestions from these various sources was used to develop this final position paper. This final position paper, referred to as the *Plant-Associated Microbe Genome Initiative*, will be distributed to appropriate federal research funding agencies to garner support and guide program development for promoting genome analysis of these important microbes.

Recommendations for Genome Analysis of Plant-Associated Microbes

To remedy the paucity of genomic information available for plant-associated microbes, the workshop participants developed the *Plant-Associated Microbe Genome Initiative* and proposed a five-year concentrated effort. The recommendations for the initiative reflect the diversity of the microbes that interact with plants and the tremendous differences in the current status of the genome analysis of those organisms. Due to the diverse group of organisms, the recommendations are meant to provide a broad outline of how an international comprehensive effort might be approached and are not meant to provide specific technical detail of how the analyses should be accomplished. To be effective, the effort must be *collaborative, well-coordinated, and international*. Industrial and public partnerships should be encouraged with the stipulation that all information must be made rapidly available to the *public*. Finally, the effort must involve strong *education and training* programs.

Four areas are described below that should be targeted in this five-year initiative. The participants recommend a \$500 million investment to be distributed equally among these areas over the five years, except that in the first few years the proportion of resources necessarily will be greater to sequence analysis than to functional analyses.

1. Genome sequence analysis

Genome sequence information is the key to gene discovery and the first step towards learning the function of each gene. There are vast differences in the amounts of sequence information currently available for different plant-associated microbes. For example, due to their smaller genome size, many different plant viruses and a few plant bacteria have already been sequenced; whereas, no genomes of plant-associated oomycetes, fungi or nematodes have been sequenced to completion. These differences in available information, which vary widely within and between microbe groups, impact the entry level for analysis of each group. They also impact the cost of sequencing the genomes of different microbe groups.

No consensus was reached on the depth of sequence needed (i.e., draft versus complete genome sequences). One component of workshop participants suggested that to obtain useful SNP and genotype information more than one strain of an organism should be sequenced. As an example, this component recommended that the genomes of three strains of a given bacterial species should be sequenced completely followed by sequencing to 8X coverage of an additional 10-20 strains. Another component suggested that if a complete genome sequence were available from one type strain, then lower coverage draft sequence of additional strains would provide information sufficient for most purposes. Although there were other permutations of this discussion, the emerging conclusions are that the depth of sequencing will depend on the end user needs (the questions to be addressed), subject to the complexity of the microbial genome of interest and available resources.

Recommendation for the Plant-Associated Microbe Genome Initiative:

- Focus initial resources largely on sequence analysis to remedy the paucity in sequence information available for plant-associated microbes;
- Target 2,000 Mb of sequence total over all microbes over five years; and
- Determine the depth of sequence or coverage for each microbe on a case-by-case basis, based on the specific needs or questions, to be evaluated by the research community.

2. **Functional analysis**

Sophisticated technologies, including array-based technologies, are available for gene expression and proteomic and metabolomic analyses and will enable the dissection of the functions of genes and proteins involved in interactions of microbes with plants and the environment. Additionally, high-throughput genome-wide gene deletion and tagging methods are being developed. Although limited, there is publicly-available genome sequence information currently for a few plant-associated microbes that is sufficient to launch functional genomic studies. These microbes can serve as model systems for functional genomic efforts to identify genes/proteins and biochemical pathways that are involved in interactions with plants.

Recommendation for the Plant-Associated Microbe Genome Initiative:

- Focus initial resources on those few microbes with sufficient structural genomic and genetic resources available; and
- Invest in improvement of technologies for functional analyses.

3. **Standardized bioinformatic tools and database system**

The underpinning for successful structural and functional genome analysis is the ability to integrate, manipulate, analyze, compare, and store all of data generated by these technologies. It was clear from the intensity of discussions surrounding the bioinformatic tools and database systems that this is an area of major deficiency and concern. It was also clear that this is a problem not only for analysis and storage of data for the genomes of plant-associated microbes, but for all databases. Participants discussed the merits of fixing the existing systems as opposed to developing new, state-of-the-art systems, or using one major database as opposed to many smaller, interconnected databases. Critical features highlighted were that these tools must be (1) flexible (e.g., usable for comparisons among different genomes, including microbes and plants), (2) scalable, (3) inter-operable, (4) easily implemented, (5) readily updateable, and (6) stably funded and maintained. Guidelines should be developed for standardization of public databases, and these should include common (1) means for electronic submission of data, (2) terminology, and (3) criteria for quality of data submitted.

Recommendation for the Plant-Associated Microbe Genome Initiative:

- Increase resources allocated to the standardization of databases and bioinformatic tools, as these are essential to the total science of genomics and the application of that science;
- Develop mechanisms for discussions, planning, and research at an international level; and
- Develop a framework to finance and manage the maintenance and continual updating of databases and information resources.

4. **Train the next generation of plant microbiologists**

To capitalize on the information provided from the analysis of the genomes of plant-associated microbes, a key goal of the *Plant-Associated Microbe Genome Initiative* must be to train new and update current plant microbiologists to ensure a research community skilled in microbial and plant biology and in computational sciences. Additionally, these scholars will need the resources to interact and collaborate in an international research arena. International collaborations are particularly important for analysis of pathogens exotic to the US.

Recommendation for the Plant-Associated Microbe Genome Initiative:

- Train plant microbiologists in the skills of genomics and computational sciences (e.g., this training should involve scientists at all levels of their careers to ensure an up-to-date, skilled pool of plant microbiologists); and
- Provide funding and opportunities for domestic and international research collaborations.