

The Microbe Project:

A BIO Advisory Committee Workshop

Report of a Workshop on Microbial Genomics to
Develop Recommendations for the
BIO Directorate of the National Science Foundation



Workshop Dates: August 10-11, 2000

Conveners:

Claire M. Fraser

The Institute for Genomic Research
cmfraser@tigr.org

John Wooley

University of California, San Diego
jwooley@ucsd.edu

About the Workshop

The Microbe Project: A BIO Advisory Committee Workshop

August 10-11, 2000

Report of a Workshop on Microbial Genomics to develop recommendations for the BIO Directorate of the National Science Foundation.

Conveners:

Claire M. Fraser
The Institute for Genomic Research
cmfraser@tigr.org

John Wooley
University of California, San Diego
jwooley@ucsd.edu

“The purpose of the BIO Advisory Workshop was to provide advice that will help crystallize NSF’s role in the Microbe Project.”

Microbes were the first organisms on earth and predated animals and plants by more than 3 billion years. They are the foundation of the biosphere — both from an evolutionary and an environmental perspective (1). It has been estimated that microbial species make up about 60% of the Earth’s biomass. The genetic, metabolic, and physiological diversity of microbial species is far greater than that found in plants and animals. And yet the diversity of the microbial world is largely unknown, with less than one-half of 1% of the estimated 2-3 billion microbial species identified. Of those species that have been described, their biological diversity is extraordinary, having adapted to grow under extremes of temperature, pH, salt concentration, and oxygen levels.

Perhaps no other area of research has been so energized by the application of genomic technology than the microbial field. It was just five years ago that the first complete genome sequence for a free-living organism was reported (2), and since that first report more than thirty-five microbial genome sequences have been published, with more than one hundred other projects underway (3). This progress has represented, on average, one completed genome sequence every two months and all indications point to this pace continuing to accelerate. Included in the first completed microbial projects are many important human pathogens; the simplest known free-living organism; “model” organisms *E. coli* and *B. subtilis*; thermophilic bacterial species that may represent some of the deepest branching members of the bacterial lineage; five representatives of the archaeal domain; and the first eukaryote, *Saccharomyces cerevisiae*. While this level of progress may seem impressive, it should be stressed that this is only the tip of the iceberg in terms of microbial diversity.

The application of genomics to the study of microorganisms provides a unique opportunity to learn about the unity and diversity of life on this planet. All of the organisms that have been studied to date by whole genome analysis are species that can be grown either in the laboratory or in animal cells. It is important to remember that the vast majority of microbial species cannot be cultivated at all, and these organisms, which live in microbial communities, play essential roles in the overall ecology of the planet. Nevertheless, the study of “laboratory-adapted” microbes has had a profound impact on our understanding of the biology and the evolutionary relationships among microbial species. For example, these efforts have uncovered entirely new metabolic pathways, have accelerated the study of gene regulation in microbial species, have revealed that approximately one-half of all predicted coding sequences are of unknown biological function, and have suggested that lateral gene transfer among organisms has played a significant role in the evolution of microbial species and microbial diversity. None of these insights would have been possible without genome sequencing and analysis. This vast amount of new information has provided an entirely new starting point for investigations in both basic and applied areas of research. The payoffs from these efforts will be significant and will promote advances

Agencies involved in the Microbial Genome Project at time of workshop: NIH, USDA, DOE, NASA, and NSF.

in microbiology and related disciplines, systems biology, drug and vaccine design, and industrial and environmental processes.

Following on from the Interagency Report on the Federal Investment in Microbial Genomics, a new interagency working group is being convened to develop a coordinated, interagency effort, now called "The Microbe Project." NIH, USDA, DOE, and NASA in addition to NSF are involved thus far, and other agencies may join. Each agency's mission will dictate its primary role in the Microbe Project. The NSF role clearly is basic science related to microbial diversity, including microbes in the environment. It seems likely that in future years joint programs may be developed through collaborations among agencies.

Given the rapid progress in the field of microbial genomics, it is perhaps not surprising that a number of challenges have emerged, and certain areas that deserve attention have been overlooked. These include, for example, the need to address priorities for future sequencing projects, to determine the role of small vs. large sequencing groups in the overall enterprise, to define and adopt standards in gene annotation, to develop consistent and fair policies governing data release by sequencing groups and its use by the scientific community, to establish mechanisms for long-term investment in databases and software for data mining and manipulation, to meet the challenges and opportunities in the area of functional genomics, and to train the next generation of genome investigators. To address all these challenges will require expanded cooperation and coordination among the governmental agencies that fund this work.

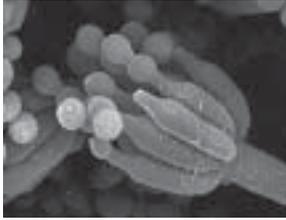
The purpose of the BIO AC workshop was to provide advice that will help crystallize NSF's role in The Microbe Project. The workshop summarized the accomplishments, challenges, and opportunities in the microbial genomics field that are relevant to the NSF; provided some direct advice such as criteria for selection of microbes that should be sequenced with NSF support and areas for future development in the bioinformatics arena; and identified issues that should be addressed in greater detail in future workshops or other venues including informatics, standards for annotation, and infrastructure needs.

Recommendations

The specific recommendations to the BIO Directorate of the NSF resulting from this workshop can be summarized as follows:

1. Support genomics-based research to elucidate the content and organization of genes in the biosphere. These programs will answer such questions as:
 - How are genes distributed among organisms, and why?
 - How do genes define the interactions of organisms with the environment?

The Microbe Project: A BIO Advisory Committee Workshop



Penicillium. Photo courtesy of Richard Edelman.

- How is biological diversity between and among organisms achieved and maintained?
- How do organisms interact with each other in the environment?

It is anticipated that large-scale DNA sequencing and analysis will play a critical role in these projects. A wide range of research activities were recommended as a means of addressing questions that included:

- complete genome analysis of cultivated prokaryotic species to achieve phylogenetic breadth
- rDNA and genome inventories of multiple environments to understand species distributions and the effect of environment on organism activities such as nutrient and energy cycling

Both viral and lower eukaryotic protists should be included in these activities to gain a comprehensive view of environmental diversity and species relationships.

2. Support development of accurate, automated genome analysis tools.

Support creation of community-specific databases that integrate genomic and experimental information, undergo continuous curation by experts in the field, and interoperate with other databases.

Define and further develop standards for genome annotation that facilitate comparative genomics studies and database interoperability.

- ### 3. Support further development of readily-affordable technologies for genome-enabled science and the requisite tools for handling ever-increasing amounts of functional genomics data. Specific goals include the development of technologies to interrogate every open reading frame in an organism/population, to systematically determine the function of unknown genes/proteins, to measure gene activities in the environment at the level of a single cell, and to study the functional interactions of all organisms in the biome.
- ### 4. Support training programs that foster collaboration and cross-training of scientists from a variety of disciplines including but not limited to microbiology, ecology, genomics, computer science, and bioinformatics. Training programs should be created not only for undergraduates, graduate students, and post-doctoral fellows, but also for established investigators and may consist of training grants, symposia, and intensive summer courses, as examples.
- ### 5. Support educational outreach programs that communicate the excitement and practical importance and benefits of microbial genomics to other scientific communities and to the general public. All too often good science is not funded because the relevance of the research to the non-specialist is not understood. This is a time and an area with potentially great pay-off for society, and it is imperative that this message be effectively communicated.