APPENDIX

Materials sent to participants in advance of the workshop.

The Microbe Project: A BIO Advisory Committee Workshop

Conveners:

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

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

Objectives of the Workshop

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 TIGR published the first complete genome sequence for a free-living organism, *Haemophilus influenzae* (2), and since that first report another twenty-seven microbial genome sequences have been published, with more than 100 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*.

All of the organisms that have been studied 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 under-

standing of the biology and the evolutionary relationships among microbial species. Microbiologists are now beginning to exploit this vast amount of new information in both basic and applied areas of research. The payoffs from these efforts will be significant and will promote advances in drug and vaccine design and in industrial and environmental processes.

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 expand cooperation and coordination among the governmental agencies that are funding this work; address priorities for future sequencing projects; determine the role of small vs. large sequencing groups in the overall enterprise; define and adopt standards in gene annotation, develop consistent and fair policies governing data release by sequencing groups and its use by the scientific community, establish a long-term investment in databases and software for data mining and manipulation, meet challenges and opportunities in the area of functional genomics, and implement new programs to train the next generation of genome investigators.

Following on from the Interagency Report on the Federal Investment in Microbial Genomics, a new interagency working group will be 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 some agencies will collaborate and develop joint programs.

Statement of Purpose and a List of Topics

The purpose of the proposed BIO AC workshop will be to provide advice that will help crystallize NSF's role in The Microbe Project. The workshop will summarize the accomplishments, challenges, and opportunities in the microbial genomics field that are relevant to the NSF; provide some direct advice such as providing criteria for selection of microbes that should be sequenced with NSF support; and identify issues that should be addressed in greater detail including informatics, standards for annotation, and infrastructure needs in future workshops or other venues. The list of topics that could be discussed at this workshop include:

Where are we today in microbial genome sequencing — summary of funded microbial genome sequencing projects.

Where should we go from here – discussion of areas of research not yet funded that could increase the understanding of the microbial world including its biochemical and metabolic diversity; the evolution of microbial species; and the application of this information in basic biology, human and animal health, agriculture, the environment, and biotechnology.

How should priorities for future sequencing projects be determined? What criteria should be used to evaluate the choice of microorganisms for genome sequencing projects? How should genomics of larger organisms be addressed in the future?

What is the role of small vs. medium vs. large sequencing centers in meeting NSF objectives in microbial genomics? What is the role of academic/government/industrial partnerships? Does a virtual genomics facility made up of multiple participants make sense? How should future NSF-funded sequencing projects be coordinated to involve as many interested scientists as possible? How can sequencing infrastructure needs best be addressed?

What is required to further develop the genomic information infrastructure (sequences, databases, software) to enable the largest number of scientists to carry out genome-enabled science? What is the role of large and generalized vs. small and specialty databases? How is funding for these activities coordinated among multiple funding agencies?

What is the current state of the art in techniques related to functional genomics (DNA microarrays, proteomics, computational biology, and structural biology)? What are the challenges and opportunities in the coming years in the area of functional genomics? What is the role of large vs. small genomics facilities in elucidating the relationship between sequence and function on a large-scale? What are the most urgent tools and resources needed to drive genomic-level research? What is the role for NSF in establishing functional genomics centers?

What are the most urgent workforce issues that need to be addressed in order to maximally exploit breakthroughs that will come from genomics?

Participants

The invitees represent a cross-section from the fields of microbiology, environmental microbiology, evolutionary microbiology, microbial genomics, and bioinformatics. Many of the participants have had in-depth experience in the application of genomics to research questions.

BIO AC Members, Co-Conveners

Claire Fraser - The Institute for Genomic Research
John Wooley – University of California, San Diego

Invited Participants

Farooq Azam - Scripps Institute of Oceanography
Colleen Cavanaugh - Harvard University
Penny Chisholm – Massachusetts Institute of Technology
Ed DeLong – Monterey Bay Aquarium Research Institute
W. Ford Doolittle - Dalhousie University
Horst Feldbeck – Scripps Institute of Oceanography
Ken Halanych – Woods Hole Oceanographic Institute
Peter Karp - SRI International
Chad Nusbaum – Massachusetts Institute of Technology
Jim Oliver – University of North Carolina at Charlotte
Gary Olsen – University of Illinois
Mitch Sogin – MBL
Dieter Soll - Yale University
Jim Tiedje - Michigan State University
Maggie Werner-Washburne - University of New Mexico
Owen White – The Institute for Genomic Research

NSF Observers

Rita Colwell – NSF Director
Mary Clutter – Assistant Director for the Biological Sciences
Maryanna Henkart – Division Director, Division of Molecular and Cellular Biosciences
Matthew Kane – Program Director, Systematic Biology

Location of Meeting and Dates

The workshop will be held at the J. Erik Jonsson Center of the National Academy of Sciences in Woods Hole, MA from August 10 (all day) until early afternoon on August 11, 2000. We plan to start the first day with a series of short presentations by selected participants to give a feel for the current state of microbial genomics, technological capabilities, and projected needs. We hope that this will lead into the afternoon's discussions on how a NSF microbial genomics program should be structured and supported. Following lunch on the first day we will break into groups to discuss some of the specific questions outlined above and begin to compile a list of priorities and needs. Day 2 will start with short presentations by the break-out group leaders which should prompt further discussions. If necessary, the groups will reconvene for finalization of their documents. A final session of the second day will be for a wrap-up session. Our hope is to be able to produce a white paper that will summarize how this group envisions the present and future needs of the scientific communities in the broad field of microbial genomics and how these needs can best be met by NSF alone and in partnership with other agencies. The recommendations made in this document, we hope, will also serve as a platform in support of additional workshops and symposia related to specific issues.

References

Staley, J. J., Castenholz, R. W., Colwell, R. R., Holt, J. G., Kane, M. D., Pace, N. R. et al. The Microbial World. The foundation of the biosphere. American Society for Microbiology, 1997.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae*. Science 269: 496-512 (1995).
http://www.tigr.org/tdb/mdb/mdb.html

Schedule

August 10, 2000

8:30 – 9:00	Informal breakfast at the Jonsson Center
9:00 – 12:30	Presentations by participants
9:00	Welcome – Rita Colwell, Director, NSF
9:10	Introduction and objectives of workshop Mary Clutter / Claire Fraser / John Wooley
	Microbial genome sequencing projects: state of the field and where do we go from here?
9:20	Ed DeLong (MBARI)
9:40	Ford Doolittle (University of Dalhousie)
10:00	Farooq Azam (SIO)
10:20	Mitch Sogin (MBL)
10:40	Break
	Genome databases: state of the field and where do we go from here?
11:00	Owen White (TIGR)
11:20	Peter Karp (SRI)
	Functional genomics: state of the field and where do we go from here?
11:40	Maggie Werner-Washburne (University of New Mexico)
12:00	TBD
	Summary of report of Marine Microbial Genomics Workshop held in April 2000
12:20	Penny Chisholm (MIT)
12:40 – 1:40	Lunch
1:40 – 2:00	Discussion of workshop objectives / Charge to participants
2:00 – 5:00	Break-out group discussions (with break at 3:30 p.m.)
5:00 – 6:00	Reception at the Jonsson Center
6:00 – 8:00	Dinner at the Jonsson Center

August 11, 2000

8:30 – 9:00	Informal breakfast at the Jonsson Center
9:00 – 10:30	Presentations by break-out groups (30 minutes each)
10:30 – 12:00	Break-out groups reconvene for final write-ups
12:00 – 1:00	Lunch
1:00 – 2:00	Final meeting of all participants and review of recommendations

The intention of the break-out groups is to address a limited number of specific topics in the areas of (1) genome sequencing, (2) bioinformatics and infrastructure, and (3) functional genomics. During the first afternoon, each break-out group will have the charge to discuss their area, address the questions, and produce a synopsis of recommendations for presentation to the entire group the next morning.

Following the presentations on day two, the break-out groups will reassemble and, using the revised list of recommendations, prepare a document that will be included into a white paper for presentation to BIO.