Gaps and Opportunities

Prior to 1999, a number of agency efforts in microbial genomics and genome-enabled microbial research have been developed independently and to different extents, as available resources allowed. Consequently, despite the intense interest of many Federal agencies in microbial genomics, there are important gaps in Federal support for research, infrastructure, and training in this important area.[1]

To seize the opportunities offered by genome-enabled microbial science, the Microbe Project has three broad goals: to build needed infrastructure, to promote research, and to develop human resources and an informed public for the future of this area. The gaps and opportunities identified with each goal, and potential coordinated agency activities to address these goals, are outlined below.

Goal 1: Infrastructure

Microbial genomics relies upon three major components of infrastructure: genome sequences, new tools and technologies, and databases and informatics tools.

A. Develop the genomic information infrastructure (genome sequences and primary databases) to enable further advances, focusing on microbes and microbial communities of scientific interest and practical importance.

Genome-enabled science depends upon the availability of genome sequence data. By examining a genome sequence and systematically comparing sequences among related and unrelated microbes one can learn fundamentally new information about the identity and function of a microbe's molecular anatomy.

To estimate the magnitude and distribution of support for research in microbial genomics, an informal survey was made of each agency's investment in microbial genome sequencing in fiscal years 1999 and 2000 (Appendix 2). The total Federal investment in large-scale sequencing of microbial genomes was approximately $33M in FY99, increasing to $45M in FY00, but the investments of the agencies have been unequal. Approximately 85% of the microorganisms whose genomes have been or are in the process of being sequenced are human pathogens and microbes of relevance to energy production and energy-related bioremediation, reflecting the larger and longer-term investments of NIH and DOE, respectively. Although the NSF and USDA investments increased in FY 2000, much remains to be done to fill the gaps in infrastructure represented by microbial genome sequences.

Gaps and Opportunities in Microbial Genome Sequencing:

Still missing from the Federal effort are significant and sustained investments to determine the genome sequences of:

- microbes of fundamental scientific interest such as those that may shed light on the history of life on earth
- microbes relevant to agriculture and aquaculture
- microbes that endanger human and animal health through food-borne routes
- microbes (including marine microbes and harmful algae) relevant to the environment and biogeochemical cycles

[1] Some of these gaps were identified in the first Interagency Report and in a recent report from the American Academy of Microbiology (see appendix for URLs).

- microbes inhabiting a wide range of ecological niches including symbionts, reef ecosystems, extreme environments or environments that may resemble that of early Earth or other planets
- microbes relevant to endangered and invasive species
- microbes under-represented in analyses and databases such as certain viruses, fungi, algae, difficult-to-culture microbes, and unique protozoa.
- microbes involved with bioremediation for improving the environment and bioindicator species for assessing environmental quality

Recommendations for Microbial Genome Sequencing:

- The Microbe Project Interagency Working Group (MPIWG) recommends that microbial genome sequencing be escalated and expanded to include microbes in the categories listed above, either in complementary efforts by individual agencies, in multi-agency joint efforts, or international collaborations. The MPIWG is aware that the private sector is involved in sequencing commercially important microbes. Issues associated with industry efforts are described in the Broader Issues section.

Planned Activities—Individual Agencies:

DOE's Microbial Genome Program will continue to support genomic sequencing of microbes relevant to DOE missions, to be carried out at the DOE Joint Genome Institute. The Joint Genome Institute is also making available a fraction of its genome sequencing capacity in support of projects of interest to other agencies. NIH has made a substantial investment in large scale sequencing projects already and will continue funding sequencing of human pathogens at the same level for the next few years. USDA plans to continue support for high-throughput sequencing of microorganisms that are important to agriculture, forestry, the environment, or the safety and quality of the nation's food supply. NSF is interested in supporting the sequencing of microbes of fundamental scientific interest, those that inhabit a wide range of ecological niches, those that will help to define the extent of microbial diversity, and microbes that may contribute to biotechnology. NSF accepts unsolicited proposals that include requests for microbial genome sequencing and support those that are deemed high priority based on merit review and available resources. FDA and NOAA are interested in the genome comparisons of pathogens that endanger food and seafood safety and their commensal (benign) counterparts. NOAA is interested in sequencing microbes that are important to coastal ecosystem health, that impact fishery resources by limiting harvest or by causing disease, and that can be exploited for bioremediation efforts or the production of novel compounds such as new antibiotics.

Planned Activities—Interagency Activities:

In FY2001, the NSF and USDA are planning a joint announcement to invite proposals for high throughput sequencing of microbial genomes of interest for fundamental biology and for agricultural applications.

For the future, a broader, multi-agency coordinated effort for soliciting microbial genome sequencing proposals is under consideration.

B. Develop the experimental tools and techniques and biological resources to expedite genome-enabled microbial research.

These infrastructure needs are common to all microbial genomics research across all the agencies, and in fact are needed to support all genomics research. Several agencies have tried to address these needs within

their available resources (for example the NIH Pathogen Functional Genomics Resource Center). However, there has been limited interagency coordination to date, even though this is an area that transcends mission boundaries. Under the Microbe Project, interagency coordination and collaboration will be facilitated, which will promote the development of these infrastructure needs dramatically.

Gaps and Opportunities in Tools, Techniques, and Biological Resources:

- Development of new experimental tools and techniques, including novel sequencing and characterization techniques (subtractive hybridization, etc.), functional genomics tools (gene chips, technologies, etc.), comparative genomics, proteomics tools, novel culture techniques, *in situ* analyses, and instrumentation.
- Development of biological resources needed to support genome-enabled research, such as specialized cells and cell lines, strains, BAC libraries, etc.
- Repositories for cells, strains, and genomics resources.

The issue of establishing and maintaining repositories is currently being debated nationally and internationally, by the international Organization for Economic Cooperation and Development Working Party on Biotechnology. Among the issues being discussed are whether there should be a "National Resource Center," or multiple centers to develop and distribute genomic resources, and whether the Federal Government should be responsible for linking existing resource centers. Also of concern are the establishment of standards for deposition, quality assurance, access, and distribution. Finally, the issue of how to provide short- and long-term support for such resources has yet to be resolved.

Recommendations for Tools, Techniques, and Biological Resources:

- Individual agencies should continue or, as necessary, increase support for technique and tool development. All such efforts should be coordinated through the MPIWG. It is expected that results of such research support will be published, including those required for patenting, in accord with the Bayh-Dole Act.
- The Federal government should initiate a deliberate planning effort to address the issue of providing sustained support for Biological Resource Center(s) for genomic resources and ensuring access to these resources for the academic, government, and not-for profit research communities.

C. Develop the databases and bioinformatics tools needed for optimal development of genome-enabled microbial science.

Gaps and Opportunities in Databases and Bioinformatics:

The creation of databases and bioinformatics tools to analyze the rapidly accumulating sequence data has raised a number of difficult issues. Because many of the microbial genome databases have been generated independently, there are incompatibilities and inconsistencies in the ways sequence data are stored, annotated, and released. Over the last year or so, the debate has increased in intensity. Among the issues being debated are whether there should be one "mega" database or a collection of linked "boutique" databases (each unique for a separate organism or limited set), whether there should be general standards for deposition, release, annotation, and accessibility of sequence data and if so, what the standards should be and how they could be enforced, and finally, how to manage short- and long-term support for databases and informatics facilities.

The American Society for Microbiology, in a November 9, 2000 report entitled *Recommendations Related to Microbial Genome Sequence Analysis and Annotation,* indicated that defining more consistent annotation definitions, and then developing standardized means to implement those definitions on dynamic datasets, was essential for the microbial community, and that a Federal interagency effort to this end was urgently needed. This committee also recommended that the value of moving toward a centralized or unified clearinghouse for thoroughly annotated sequence data be considered by the Federal government, and strongly reiterated the need to ensure that microbial genome datasets are made fully available to the public.
Recommendations for Databases and Informatics Tools:

- The MPIWG recommends that the responsible Federal agencies make resolving the database issues (including standardization of annotation and inter-operability) a top priority, and that this be undertaken as an interagency effort with planning activities to begin immediately. The results of these planning activities should guide the further development of agency programs to support databases and informatics tools development.
- The MPIWG recommends interagency coordination to maximize the investment and leveraging of resources to develop standardized bioinformatics tools for the analysis of microbial genomes.

Planned Infrastructure Activities—Individual Agencies

Several of the agencies are interested in supporting this goal, and a number of individual agency efforts are in development. For example, NIH is committed to building infrastructure that will expedite genome-enabled research and plans to continue this support as it is related to the mission of the different institutes, especially in the area of bioinformatics and access and distribution of tools and resources to the research community. The Microbial Genome Program at DOE is planning initiatives to develop novel strategies to avoid "starting from scratch" in sequencing microbes that are very closely related to others whose sequence already is known. DOE is also interested in developing new tools to study how groups of genes work together to produce specific products or determine particular behaviors, improving tools for annotation and analysis of sequence data, developing high-throughput methods for determining gene function and gene expression, and developing methods for examining protein-protein and protein-nucleic acid interaction. NSF programs consider technique development proposals in the context of the relevant biological research area. The Information Technology Research initiative at NSF has a new component in FY2001 to include biological IT applications. USDA is interested in supporting microarray/chip development for gene expression analysis for agricultural microbes, bioinformatics, and proteomics, and the establishment of centralized facilities for resource distribution. USDA is also interested in supporting the development/enhancement of bioinformatics tools with specific application to agricultural microbial genomic data. FDA is interested in the development and standardization of microarray analysis for both diagnostics and surveillance of a wide variety of microbes impacting human and animal health.

Planned Infrastructure Activities—Interagency

In the coming year the MPIWG will sponsor workshops to:

- evaluate the issues associated with biological resource centers for microbial genomics, and to guide planning for potential interagency activities.
- evaluate the issues surrounding standards and long term support for microbial genome sequence and higher order databases, and to guide planning for potential interagency activities to optimize the content and access to information infrastructure for genome-enabled microbial science.

Goal 2: Research

Enhance support for new, genome-enabled research using the tools, resources, and concepts of genomics.

Microbes are an essential and vast segment of the biological world about which we still know very little. Microbes build the natural environment in which we live, sustain its biological economy, and are essential for its decay. Microbes make us sick and keep us healthy, affect the health of animals and the safety of the foods we eat, and have enormous potential for providing new pharmaceutical and environmental products. At the most basic level, microbes can tell us how life began and how it is sustained today. With the advent of genomics, we are poised to make tremendous strides forward in our understanding of these diverse organisms.

While it is clear that genomics offers unprecedented opportunities, there are major areas of research as yet untouched that would increase our understanding of the broader microbial world, its diversity, and its potential applications.

Examples of Research Gaps and Opportunities include:

- Mining the information implicit in microbial genomes to deduce the biology of microbes including the structure and function of a cell.
- Exploiting the available genome sequence data from microbes to develop new strategies for diagnosis and treatment, such as defining new targets for drugs and vaccines for humans and animals.
- Using comparative genomics to look for variation in commensal and pathogenic strains.
- Determining how to take advantage of the diversity of microbes inhabiting the human body to promote health.
- Using microbial genome analysis to understand a microbial evolutionary tree, and then determine how this may relate to the evolutionary lineage of multicellular organisms and the emergence of beneficial and pathogenic species.
- Using microbial genome analysis to understand the frequency of, and constraints upon, lateral gene exchange (the acquisition of genes en bloc by one microbe from the genome of another).
- Analyzing microbial metabolic diversity and function, and applying findings in bioprocess engineering for environmentally friendly manufacturing and conversion of agricultural wastes.
- Studying beneficial as well as harmful microbes relevant to agricultural crops, aquaculture, fisheries, farm animals, food and food processing to develop the knowledge base for managing them.
- Elucidating the ecology of microbes in the wide range of habitats on Earth, including the contributions of microbes to biogeochemical cycles in the environment.
- Studying microbes in extreme environments to understand the potential for life elsewhere in the universe.
- Studying marine microbes (including harmful algae) to understand their effects on the health of the marine environment.

Recommendations for Research:

- The MPIWG recommends that each agency, as its mission directs, encourage and support genome-enabled microbial research objectives such as those listed above. In some cases enhanced resources must be sought to realize this goal.

Planned Activities—Individual Agencies

NIH, NSF and DOE are continuing current programs and developing new plans to exploit genomic information and tools to understand the biology of a microbial cell. The Microbial Genomics Program at DOE, for example, plans to initiate a new program called the Microbial Cell Project. This new initiative, beginning in FY2001, will support research that uses genomic approaches to integrate the extensive but fragmented molecular and cell biology data about cellular processes with the ultimate goal of understanding and modeling the complex functioning of a prokaryotic microbial cell. NIH is moving forward with initiatives in functional genomics, taking advantage of emerging DNA sequence information, and in addition, plans to place special emphasis on "complex systems" approaches to understanding the cell. These computationally intensive approaches will rely in great part upon genomics and to a significant extent will target microbial systems. The research areas directly related to microbial genomics will include: determination of the "wiring diagrams" and control logic of metabolic pathways; signal transduction pathways; macromolecule synthesis and degradation pathways; growth related mechanochemical processes; and quantitative modeling of system dynamics. NSF's Biocomplexity in the Environment Initiative, the new Quantitative Systems Biotechnology Program, and a planned Genome-Enabled Sciences emphasis will also have components directed toward integrating separate physiological systems and pathways into understanding of a complex whole. NOAA's Ocean Exploration Initiative will look at microbes in ocean ecosystems never before accessed or studied. The existence of a coordinated Federal Microbe Project will provide the opportunity to unite these agency efforts and promote greater coordination and collaboration.

Many of the research areas listed above directly address the missions of NIH, NSF, USDA, EPA, FDA, DOE, NASA and NOAA. These agencies are very interested in supporting genome-enabled microbial research, either through existing programs or through the development of new, possibly multi-agency, initiatives.

Goal 3: Human Resources

Promote education and training of students, scientists, and the public for genome-enabled microbial biology. Promote the diversity of participants in genome-enabled microbial biology.

Human Resources Gaps and Opportunities:

- Education at the interface of microbial biology, genetics, biotechnology, engineering, math, and computer science
- Training of new generations of genome-enabled microbial biologists, including systematists and physiologists.
- Full participation of the diverse U.S human resources in the advancement of genome-enabled microbial science.
- Education of the public to increase awareness of the power of genomics and importance of microbial biology in their lives.

Recommendations for Human Resources:

The MPIWG recommends that:

- Individual and interagency activities initiated as part of the Microbe Project should contain elements that encourage training and/or educational activities, and include efforts to enhance the diversity of participants in all aspects of each activity. Interagency coordination of the development and distribution of training

materials should be encouraged.

Planned Activities—Individual Agencies

NIH recognizes that to reach the stated research goals, research training in computation and bioinformatics will be needed, and is committed to support pre- and post-doctoral training programs in the area of systems and integrative biology, bioinformatics, and computational biology, and fellowships in quantitative biology. The USDA and NOAA are interested in supporting training/education/outreach activities for microbial genomics and its evolving technologies targeted to the research community, K-university level students, and the general public. The FDA is also particularly interested in public education, and provides outreach programs to educate the public on current microbial hazards and the ways that individuals can best safeguard themselves from exposure or infection. NSF has a strong commitment to integration of research and education, and has funded projects in the area of genomics curriculum development, and a number of interdisciplinary programs for graduate training in bioinformatics and genomics through its Integrative Graduate Education and Research Traineeship (IGERT) program. Postdoctoral fellowships in microbial biology and biological informatics have been designed to address the need for scientists trained in non-model microbial systems and microbial systematics using genomic tools, computational biology and bioinformatics. In addition, programs in the Education and Human Resources Directorate (EHR) have supported laboratory research, curriculum development, and undergraduate education in the area of microbiology and genomics.