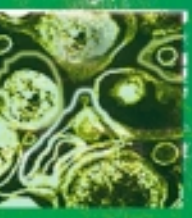
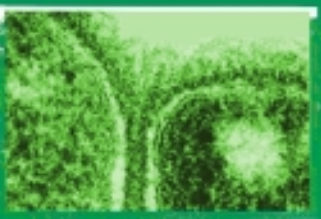
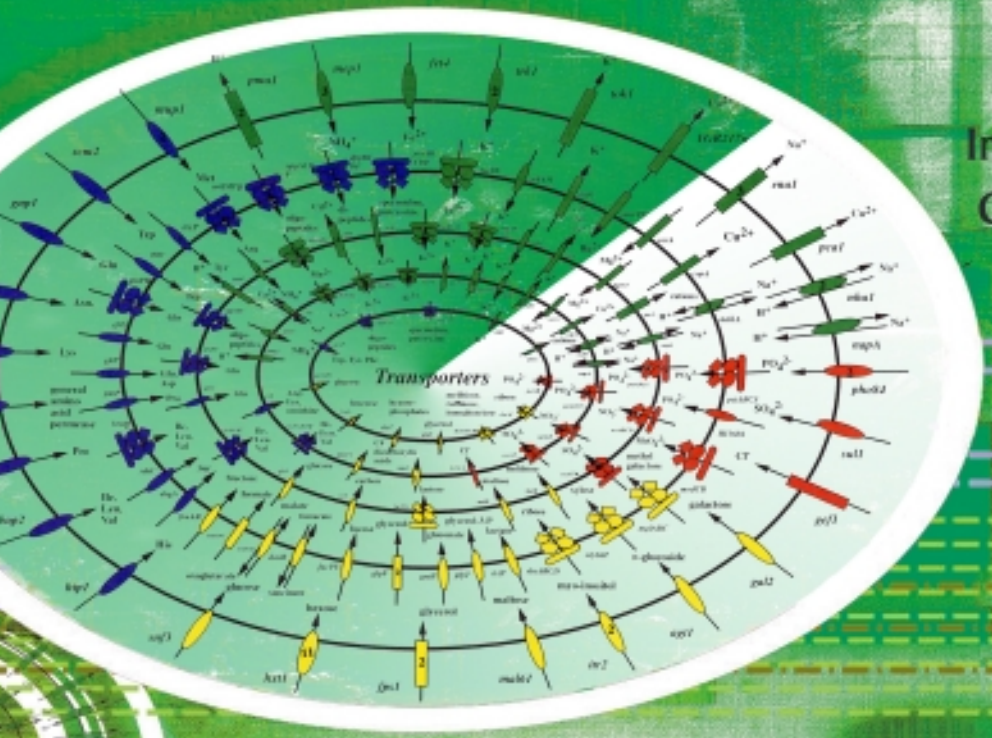
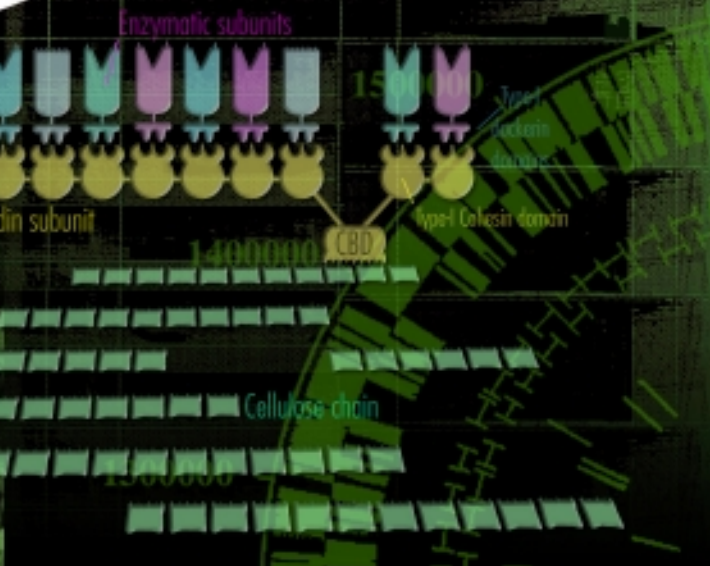
THE MICROBE PROJECT





A Report from the Interagency Working Group on Microbial Genomics





National Science and Technology Council

Committee on Science

Subcommittee on Biotechnology

January 2001

About the National Science and Technology Council

The National Science and Technology Council (NSTC) was established by Executive Order on November 23, 1993. This cabinet-level council is the principal means for the President to coordinate science, space, and technology policies across the Federal Government. NSTC acts as a virtual agency for science and technology (S&T). The President chairs the NSTC. Membership consists of the Vice President, Assistant to the President for Science and Technology, Cabinet Secretaries and Agency Heads with significant S&T responsibilities, and other White House officials. Through the NSTC, Federal departments and agencies work cooperatively to ensure that Federal science and technology investments support national goals. NSTC Committees prepare R&D strategies that are coordinated across the Federal government to form a comprehensive investment package.

Call 202-456-6100 to obtain additional information regarding the NSTC, or see our web site at: http://ostp.gov/NSTC/html/NSTC_Home.html.

About the Office of Science and Technology Policy

The Office of Science and Technology Policy (OSTP) was established by the National Science and Technology Policy, Organization and Priorities Act of 1976. OSTP's responsibilities include advising the President in policy formulation and budget development on all questions in which S&T are important elements; articulating the President's S&T policies and programs; and fostering strong partnerships among Federal, state and local governments, and the scientific communities in industry and academe. The Director of OSTP also serves as Assistant to the President for Science and Technology and manages the NSTC for the President.

Call 202-456-7116 to obtain additional information regarding the OSTP, or see our web site at: http://www.ostp.gov

Note: This document does not represent the final determination in an overall Administration budget decision-making process. The programs presented in this report will have to compete for resources against many other high-priority Federal programs. If these programs compete successfully, they will be reflected in future Administration budgets.

Photo reproduction permissions: Membrane transport functions identified by analysis of coding regions in five complete microbial genomes (*Mycoplasma genitalium, Methanococcus jannaschii, Sunechocystis PCC6803, Haemophilus influenzae,* and *Saccharomyces cerevisiae*), reprinted by permission from Nature 387:459-462 copyright 1997 Macmillan Magazines Ltd. Electron micrograph of *Ruminococcus albus* and a schematic representation of the extracellular cellulosome associated with *R. albus* in animal rumens that aids in the digestion of cellulose, provided courtesy of Mark Morrison at Ohio State University and Ed Bayer at the Weizmann Institute, respectively. Electron micrograph of a salt-sensitive soil bacterium, S34, provided courtesy of the American Society for Microbiology.

The Microbe Project Report

January 2001

National Science and Technology Council
Committee on Science
Subcommittee on Biotechnology
Interagency Working Group on Microbial Genomics

EXECUTIVE OFFICE OF THE PRESIDENT
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL
WASHINGTON, D.C. 20502

January 2001

Dear Colleague:

The attached report provides a rationale and interagency plan for the Microbe Project, a coordinated Federal effort in microbial genomics. While it is clear that microbial genomics research offers unprecedented opportunities, a 1999 inventory of Federal programs showed that 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. The National Science and Technology Council, Committee on Science, Subcommittee on Biotechnology's Microbe Project Interagency Working Group was charged with developing a coordinated plan to address research, infrastructure and human resource gaps, which is described within. Implementation of this plan will greatly advance discoveries based on microbial genomics research, leading to exciting new opportunities in the basic sciences, biotechnology, agriculture, human health, energy, and the environment. The private sector and international community are moving forward aggressively in this area; a vigorous public sector program will help support these efforts and ensure that microbial genomic data and resources are open and accessible to all scientists. This is essential to promote future scientific breakthroughs and new practical applications.

We thank the Interagency Working Group and the many individuals who contributed to the development of this report. The Microbe Project will contribute to the scientific enterprise that provides a high quality of life for us and for future generations.

Sincerely,

Rita R. Colwell
Co-Chair
Committee on Science
Director
National Science Foundation

Ruth L. Kirschstein
Co-Chair
Committee on Science
Principal Deputy Director
National Institutes of Health

Arthur Bienenstock
Co-Chair
Committee on Science
Associate Director for Science
Office of Science and Technology Policy

Microbe Project Interagency Working Group
Subcommittee on Biotechnology
Committee on Science
National Science and Technology Council

Anne Vidaver (Chair)
Chief Scientist
National Research Initiative
Competitive Grants Program
U.S. Department of Agriculture

Joanne Tornow (Executive Secretary)
Program Director in Molecular and
Cellular Biosciences
National Science Foundation

Janet Dorigan
Senior Scientist
Directorate of Science and Technology
Central Intelligence Agency

Robert Foster
Director, BioSystems
Office of the Deputy Under Secretary of
Defense for Science & Technology
Department of Defense

Marvin Frazier
Director, Life Sciences Division
Office of Biological and
Environmental Research
Department of Energy

Dennis B. Fenn
Associate Director for Biology
U.S. Geological Survey
Department of the Interior

David Reese
Environmental Scientist
Environmental Protection Agency

Thomas Cebula
Director, Division of Molecular Biology
Food and Drug Administration

Guy Fogleman
Acting Director, Biomedical and
Human Support Research Division
National Aeronautics and
Space Administration

Maria Giovanni
Assistant Director for Microbial Genomics
National Institute of Allergy and Infectious Diseases
National Institutes of Health

Gregory B. Vasquez
Scientific Advisor
Chemical Science and
Technology Laboratory
National Institute of Standards
and Technology

Linda Kupfer
Program Director for Marine Biotechnology
National Oceanic and
Atmospheric Administration
Department of Commerce

Maryanna Henkart
Director, Division of Molecular and
Cellular Biosciences
National Science Foundation

Marc Garufi, David Radzanowski
Mike Holland, and Noah Engelberg
Office of Management and Budget

Rachel Levinson
Assistant Director for Life Sciences
Office of Science and Technology Policy

Acknowledgments

The MPIWG was assisted greatly by the hard work and rapid responses of the following individuals:

MPIWG alternates
Eric Kaufman (CIA), Bart Kuhn (DoD), Daniel Drell and Anna Palmisano (DOE), Robert Menzer (EPA), Joseph E. LeClerc (FDA), David Liskowsky (NASA), Carole Heilman and James Anderson (NIH), Prasad Reddy (NIST), Shawn McLaughlin (NOAA), Matthew Kane and Phil Taylor (NSF), Leland Ellis and Peter Johnson (USDA).

Other contributors
Eric Eisenstadt (DARPA), Bernard A. Schwetz and Elizabeth Jacobson (FDA), Michael Meyer, Tod Companion and Bette Siegel (NASA), Sally Rockey (USDA), Elise Feingold (NHGRI, NIH), Eugene Koonin (NLM, NIH), Dennis Mangan (NIDCR, NIH), Louise Ramm (NCRR, NIH) and Janet Shoemaker (American Society for Microbiology).

Table of Contents

Executive Summary

The Need for a Coordinated Federal Effort in Microbial Genomics

Microorganisms have been present for over 3.8 billion years; we have known about their existence for over 300 years. Yet, despite the fact that microbes comprise most of the earth's biomass, maintain its environments, and hold the key both to understanding the history and health of life on Earth and to exploiting the full potential of biotechnology for myriad applications, we still know almost nothing about most of them. Now, with the advent of genomics, we are entering a new era of scientific discovery. Recognizing the broad importance of microbial genomics research, in 1999 an interagency task group conducted an informal inventory of Federally-supported research in microbial genomics. While it is clear that genomics offers unprecedented opportunities, this inventory showed that 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. A coordinated interagency (and international) effort is needed to seize the opportunities offered by genome-enabled microbial science. In recognition of this need, the Microbe Project Interagency Working Group was convened in August 2000, and charged with developing a coordinated interagency action plan for microbial genomics activities.

Goals of the Coordinated Effort

The Microbe Project has three broad goals: to build needed infrastructure, to promote research, and to develop human resources and an informed public.

- The three major components of infrastructure needed to support microbial genomics research are 1) genome sequences, 2) tools, technologies and biological resources, and 3) databases and bioinformatics.
- Genome-enabled microbial research holds enormous promise for understanding life at its most basic level, and for enabling breakthrough applications in health, agriculture, biotechnology, the environment, and national defense.
- The education and training of students, scientists, and the public in genome-enabled microbial biology, and assuring a diversity of participants in this area, is essential.

Recommendations

- Microbial genome sequencing should be expanded to include scientifically important but as yet understudied microbes.
- Individual agencies should continue or, as necessary, increase support for research on technique and tool development.
- The Federal government should initiate a deliberate planning effort to address the issue of providing sustained support for and access to microbial genomic resources.
- Develop standardized bioinformatics tools for the analysis of microbial genomes.
- Database issues (including standardized annotation, inter-operability, and long term support) must be resolved through an interagency effort with planning activities to begin immediately.
- Each agency, as its mission directs, should encourage and support genome-enabled microbial research objectives, as described in this report.
- 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.
- Continue coordination across agencies of all Microbe Project activities, in part through the development of an interagency Microbe Project web site.

Introduction

A vast and diverse microbial world occupies every nook and cranny of the globe, from the deepest depths of the ocean to the highest mountain peaks, living in the water, soil, and air that surround us, on and in the food that we eat, on and within our own bodies. Microbes (including viruses, bacteria, fungi, protozoa and microalgae) comprise most of the earth's biomass, maintain its environments, and hold the key to understanding the history of life on Earth. Microorganisms have been present for over 3.8 billion years; we have known about their existence for over 300 years. Yet, incredibly, with some notable exceptions, we still know almost nothing about most of them. Now, with the advent of genomics (the study of an organism's entire DNA complement and its function), we are entering a new era of scientific discovery that holds great promise for understanding the complexities of the microbial world.

The DNA sequence of an organism's genome is often referred to as its genetic blueprint. Analysis of microbial genome data available thus far has already yielded surprising discoveries. In each microbial genome that has been sequenced, 40 to 50% of the putative genes encode proteins of unknown function, and 20 to 30% encode unknown proteins apparently unique to that species. Genomic analysis also suggests that less than 1% of the microbes on Earth have been cultured and studied in the laboratory. Because of the unique properties of microbes already known, and the almost incomprehensible number of microbes on Earth yet to be studied, these organisms represent an untapped and extremely valuable resource for the basic sciences, biotechnology, agriculture, human health, energy, and the environment.

While a genomic sequence can yield a great deal of information, genome sequencing is only the first step toward achieving an understanding of a microbe's biological capabilities. Learning how the genomic information predicts the functions of an organism's genes and therefore predicts the organism's biology is a challenge that can now be met. Genome-enabled studies will lead to breakthroughs such as improved vaccines and better disease-diagnostic tools, identification of new drug and chemical targets in pathogens, discovery of new industrial catalysts, more accurate identification of microorganisms *in situ* in ecosystems from polar ice to soils to oceans, phylogenetic analyses of microorganisms, a general understanding of the Earth's microbial diversity, and perhaps clues to the origins of life on earth.

Recognizing the broad importance of research based on microbial genomics, in 1999 an interagency task group conducted an informal inventory of Federally-supported research in microbial genomics. The resulting "Interagency Report on the Federal Investment in Microbial Genomics" was published in spring 2000, and described in detail the ongoing infrastructure, research, and training activities of the Federal government related to microbial genomics. It is clear that this area of research supports the missions of many agencies, each of which is investing in microbial genomics-related projects, as their resources allow. In the short time since the last report, there has been increased activity by the Federal agencies, both in investments made and in the number of agencies interested in and contributing to development of microbial genomics infrastructure, research, and training.

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 (as will be described in the Gaps and Opportunities section). A coordinated inter-agency (and international) effort, is needed to seize the opportunities offered by genome-enabled microbial science. In recognition of this need, the current Interagency Working Group, named "The Microbe Project", was convened in August 2000, and charged with developing a coordinated interagency action plan for microbial genomics activities. The following section highlights the importance of infrastructure, research

and human resources for progress in The Microbe Project. Current activities in each of these areas are highlighted, by agency, in Appendix 3.

The Challenges of Infrastructure

The Federal government has a critical role, not only in support for the National research effort, but also in building the enabling infrastructure. Microbial genomics relies upon three major components of infrastructure: genomic sequencing, new tools, technologies and biological resources, and databases and informatics tools. With support from several Federal agencies, this needed infrastructure is being assembled, although the task is far from completed.

Genomic Sequencing

Obtaining the complete genome sequence of an organism is the foundation upon which all other genomics-related research is built. By examining a genomic DNA 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. The tools and technologies of sequencing have advanced tremendously in recent years, driving costs down and production rates up. The National Institutes of Health (NIH) and the Department of Energy (DOE) have been the two major Federal investors in microbial genome sequencing thus far. These agencies develop lists of priority organisms whose genome sequence is needed to advance their respective agency missions. To date, the other agencies have made only limited investments in microbial genome sequencing, as discussed in the Gaps and Opportunities section.

Tools, Technologies and Biological Resources

The development of new tools and technologies to improve experimentation is critically important for the rapid advancement of knowledge. For example, the use of microarrays ("gene chips") to assay gene expression of the entire genome was an early technological development that followed the completion of the first few microbial genome sequences. Microarray analyses are rapidly becoming standard techniques for genome-enabled research, and have already contributed tremendous amounts of new information. Genome-enabled research also depends upon the development of a variety of biological resources, such as specialized cells and cell lines, strains, clone libraries, etc. Equally important to the development of these tools, technologies, and resources is ensuring that they are accessible to all the communities of academic scientists whose research progress depends upon them. Individual Federal agencies have recognized these needs, and supported these efforts in a variety of ways. For example, NIH and the National Science Foundation (NSF) have supported programs to provide state-of-the art instrumentation (such as DNA sequencers, gene-chip equipment, and mass spectrometers) to researchers for genome-enabled research. Also, NIAID/NIH is establishing a Pathogen Functional Genomics Resource Center to distribute genomics resources and technology to the research community.

Databases and Informatics Tools

Genomic sequence data are being generated at exponentially increasing rates. One of the most pressing infrastructure needs of genomics research is the development of robust databases and informatics tools to store and analyze these data. Accurate annotation (i.e., identifying each gene in the genome with a name and putative function), analysis of global genomic data (such as the data generated by microarray experiments), and the synthesis of these data to decipher complex metabolic pathways and evolutionary relationships, all

rely on the ability to access and interpret information stored in sequence databases. The awesome potential of these computational analyses will never be realized as long as the incompatibilities between individual databases due to the absence of common standards, which currently prevent any significant cross-talk, continue to exist. Individual agencies have recognized this need, and have begun to address it, but much more will be needed in the near future (see Gaps and Opportunities section).

The Promise of Microbial Genomics Research

Even as the infrastructure is being built (but is by no means finished), the Federal agencies have begun to invest in genome-enabled microbial research. Areas of research with the potential to benefit from a continuing investment include human and animal health, agriculture, aquaculture, the environment, biotechnology and fundamental research.

Human Health

Protecting and improving human health drives the microbial genomics-related research at the NIH, and is also important to the missions of the Department of Defense (DoD), the Department of Agriculture (USDA), the National Aeronautics and Space Administration (NASA), the Environmental Protection Agency (EPA), the National Oceanic and Atmospheric Administration (NOAA), and the Food and Drug Administration (FDA). NIH has made significant investments in large-scale genomic sequencing, which is complemented by further investments in functional and structural genomics projects that incorporate and build on the genomic sequence data. The most obvious anticipated health benefits from these efforts include identifying microbial genes that represent new targets for antibiotics and vaccines, new diagnostic tests, and disease-specific markers. Microbial genomics research will also protect health in other ways, for example it will improve our ability to detect toxin-producing microbes (such as harmful algae or bacteria) that cause food poisoning.

Agriculture and Aquaculture

The study of microbes and their interactions with terrestrial and aquatic plants and animals, both harmful and beneficial, is very important for the missions of USDA, FDA, and NOAA. NSF also supports fundamental research in this area. Microbial genomics will have a major impact on the ability of the U.S. to continue to produce nutritious and safe food, while preserving the environment and wild stocks, and sustaining the economic stability of the agricultural enterprise. To date, very few agricultural or aquaculture microbes have been, or are in the process of being, sequenced. Consequently, agriculture lags behind other fields, such as human health and energy production, with respect to microbial genomics.

Energy and the Environment

The DOE, the EPA, NOAA, USDA and the U.S. Geological Survey (USGS) at the Department of the Interior (DOI) have complementary responsibilities in protecting the environment. Among other environmental research efforts, each of these agencies supports microbial genomics-related research that impacts the environment. Examples of environmental applications include identifying and harnessing the metabolic processes of microbes and microbial consortia to clean up organic compounds and heavy metal environmental pollutants, and developing sensor technology to assess the levels and effects of microbial and viral pathogens on the health of coastal ecosystems. DOE, in particular, is also supporting research to explore the mechanisms by which diverse microbes capture, transform, store and utilize energy for potential human use,

and sequester CO_2 as a very important part of the global carbon cycle.

Biotechnology

Microbes and their manufacturing capabilities offer a wealth of potential new products and processes for biotechnology. A number of agencies are interested in conducting or supporting research to explore and adapt them for numerous purposes. Examples of such products range from new polymers, to heat and cold-resistant catalysts, to antibiotics. DoD, DOE, FDA, NIH, NIST, NOAA, NSF and USDA have interests and investments in this area.

Understanding Our World

Genomics truly is the key to understanding the inhabitants of the microbial world. From a small sample of sea water or soil, DNA can be extracted and analyzed to tell us what sorts of organisms live there, and much about their potential for interacting with the environment and with other living things. Genomics can also be used to delve into the biological mysteries within the microbial cell, to understand what genes are involved in different metabolic and regulatory pathways and how those pathways connect to support a living cell. Without the tools of genomics, such insights into the microbial world and the individual cell would be unimaginable.

The Importance of Human Resources

To take advantage of the opportunities offered by the application of genomics to the study of microbes, a well-trained workforce and an educated public will be required. Future students must not only be thoroughly grounded in the concepts of biological sciences, but must be well trained in quantitative thinking and facile with computational tools. Current investigators should also have opportunities to update their skills at the interface of biology and quantitative and computational sciences to stay at the leading edge of research. The public must be educated to understand both the research efforts involved in microbial genomics, and the outcomes and impacts of such research.

Demographics show that in the near future, some groups that currently are in the minority will represent the majority of the U.S. workforce. To capitalize on the promise of genome-enabled microbial science, it will be necessary to mobilize all the best minds in the Nation and to tap all of the diverse components of our population.

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 on the following pages.

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]

[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).

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
- 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.

Broader Issues

Several broader issues still must be addressed in the near future, to maximize the impact of the investment in microbial genomics by both the public and private sectors. These issues include:

Access to Biological Resources. Mechanisms are needed to enable access of small research units to current and future Federal resources. The MPIWG believes that to promote scientific progress nationwide, it is essential to capitalize on human resources and provide state-of-the-art technology training for students and professionals at all levels of their careers.

Data release and intellectual property. Some Federal agencies require rapid release of microbial genome sequence data that are generated using public funds, because early release of unfinished sequence has proven useful in accelerating the pace of experimental discovery. On the other hand, the MPIWG recognizes that rapid release policies have to be balanced with other concerns, namely scientific fairness (allowing time for those who sequenced the genomes to do a first analysis of the information contained therein) and intellectual property, and recommends further discussion of these issues.

Implications of genomics with respect to pathogens and genetically modified microbes. Genomic sequence data is essential for enhancing our understanding of microbial life and for developing beneficial technologies such as rapid diagnostics, new therapeutics and vaccines. The MPIWG recommends that U.S. agencies support research on the scientific, environmental and ethical issues associated with the use of genetically modified microbes and engage in frank and open discussion about the ethical, legal and social implications of making public the complete DNA sequences of pathogens. With respect to the latter, the MPIWG notes that the White House Office of Science and Technology Policy has initiated such discussions, in the context of the security implications of fundamental biological and biomedical research.

Industry. Despite a significant private sector investment in microbial genomics, there are at least two compelling reasons for a strong public sector investment. First, industry's interests in microbial genomics

are focused, understandably, on commercial value, including targeting of genes related to pathogenesis, possibilities for acquired pathogen resistance, industrial and food-grade enzymes, and probiotics (to encourage beneficial microbes) for animals and humans, all for wide-scale distribution and use. Public access to genome sequences and functional genomics data held by industry is expected to be limited. Thus, for some microbes, the MPIWG considers it necessary, in the public interest, to support research that will add to the data in the public domain. Second, industry itself is supportive of a more enhanced role for the public sector in microbial genomics. Many small biotechnology companies do not have the resources to do the critical basic research needed in microbial genomics. Without a strong public research base, many of these companies will not be able to receive the financing necessary either to get started or to survive.

International collaborations. International foundations, as well as private and publicly supported institutions are active in the field of microbial genomics. These are found in Belgium, Brazil, Canada, China, the United Kingdom, France, Germany, Japan, Norway and Sweden. U.S. scientists, supported by Federal agencies, have international collaborations with a number of these institutions and organizations to do microbial sequencing and functional genomics, primarily for microbes associated with human, animal and plant diseases. These efforts include microbes of public health and bioterrorism concern that are not being addressed by the private sector. It must be recognized that different governments have differing views on which microbes should be addressed, by whom and what resources to allocate. Nonetheless, international collaborations have already shown themselves to be very fruitful for other genomics efforts (e.g., the Human Genome Project and international plant genome projects including Arabidopsis and rice), and should be encouraged in the microbial genomics arena as well. International conferences and workshops regularly serve as fora for enhancing interactions among the public, non-governmental organizations, and private sector.

Summary Recommendations

The MPIWG makes the following recommendations with respect to infrastructure, research and human resources:

Infrastructure:

- Microbial genome sequencing should be expanded to include under-studied microbes as described above, either in complementary efforts by individual agencies, in multi-agency joint efforts, or international collaborations.
- Individual agencies should continue or, as necessary, increase support for research on technique and tool development. All such efforts should be coordinated through the MPIWG.
- The Federal government should initiate a deliberate planning effort to address the issue of providing sustained support for genomic resources and ensuring access to these resources for the academic, government, and not-for profit research communities.
- Resolving the database issues (including standardization of annotation and inter-operability) should be a top priority, and this should be undertaken as an interagency effort with planning activities to begin immediately.
- Standardized bioinformatics tools for the analysis of microbial genomes should be developed.

Research:

- Each agency, as its mission directs, should encourage and support genome-enabled microbial research objectives such as those described above.

Human Resources:

- 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.

Follow-on Activities by the MPIWG

As first steps in acting on these recommendations, the MPIWG plans to:

- Create an Interagency Microbe Project Web Site, where all individual and interagency programs, program announcements, and requests for applications will be listed. The MPIWG recommends that, where appropriate, future individual agency program announcements include a statement indicating that the activity is part of a coordinated Federal effort in microbial genomics, and provide a link to the Interagency Microbe Project web site.
- Hold workshop(s) to address database incompatibilities, annotation standardization, long-term support, and other database issues.
- Hold workshop(s) to address the competing priorities of rapid release of sequence data vs. scientific fairness and intellectual property concerns.
- Hold workshop(s) to address issues related to Biological and Genomic Resource Centers, such as establishing standards for deposition, quality assurance, access, and distribution, and developing mechanisms for short- and long-term support for such resources.

Recommended Investment

It is estimated that an annual investment of $230 million across a dozen agencies is needed for the foreseeable future to make significant progress on the infrastructure, research and human resources objectives outlined above. This estimate is based on the following needs:

- $60 million to support an expanded and broader microbial genome sequencing effort, including the development of new technologies to drive costs down.

- $20 million to provide increased support for research on technique and tool development.

- $40 million to address the issues of providing sustained support for Biological Resource Center(s) for genomic resources, ensuring access to these resources for the academic, government, and not-for-profit research communities, and resolving database issues.

- $100 million to enhance support for genome-enabled microbial research.

- $10 million for human resource development.

NATIONAL SCIENCE FOUNDATION
4201 WILSON BOULEVARD
ARLINGTON, VIRGINIA 22230

OFFICE OF THE
ASSISTANT DIRECTOR
BIOLOGICAL SCIENCES

July 5, 2000

From: Mary Clutter, Chair, Subcommittee on Biotechnology MJC

Subject: Establishment of an Interagency Working Group on Microbial Genomics

Recent advances in genomics technology are ushering in a new era of scientific research and discovery for many areas of biology. Microbial genomics research is at the forefront of this developing science, and has become increasingly important in both the public and private sectors because of its potential impact on a wide variety of fields. In 1999, an interagency task group conducted an informal inventory of Federally-supported research in microbial genomics, and found that 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. There is a need for a coordinated interagency effort to maximize the opportunities offered by genome-enabled microbial science.

To address this need, I am establishing an Interagency Working Group called "The Microbe Project" under the direction of the National Science and Technology Council, Committee on Science, Subcommittee on Biotechnology. The IWG will have two primary functions:
1) identify primary gaps and opportunities in microbial genomics across the government; and
2) develop a coordinated interagency action plan.

An IWG report should be prepared by mid-December for communication to the Subcommittee on Biotechnology.

The first meeting of the IWG is scheduled for Tuesday, August 1, 2000, from 1:30 to 3:30 pm, in room 472 of the Old Executive Office Building. Principals are requested to attend, but additional staff are welcome.

DISTRIBUTION:

Members of the Microbe Project Interagency Working Group

Anne Vidaver, USDA (chair)
Joanne Tornow, NSF (executive secretary)
Janet Dorigan, CIA
Robert Foster, DOD
Marvin Frazier, DOE
Dennis Fenn, DOI
David Reese, EPA
Tom Cebula, FDA
Guy Fogleman, NASA
Maria Giovanni, NIH
Gregory B. Vasquez, NIST
Linda Kupfer, NOAA
Maryanna Henkart, NSF
Marc Garufi, OMB
David Radzanowski, OMB
Mike Holland, OMB
Noah Engelberg, OMB
Rachel Levinson, OSTP

Telephone (703) 292-8400 FAX (703) 292-9154

Distribution of the Federal Investment in Microbial Genome Sequencing

Sequencing is the platform upon which all other genome-enabled science is built; the investment in sequencing is only one measure of an agency's involvement in microbial genomics. The following table lists best estimates of FY99 and FY00 funding for large-scale sequencing projects only. To date, approximately 50% of the completed microbial genome sequences have been funded in whole or in part by U.S. government agencies.

Agency	FY99 funding	'99 Agency Total	FY00 funding	'00 Agency Total	Organisms*
DOE (OBER)	$12,400,000			$13,700,000	
DOE (OBES)	$100,000			$100,000	
DOE Total		**$12,500,000**		**$13,800,000**	45
NIAID	$12,871,000		$18,222,000		
NIDCR	$3,989,000		$2,105,000		
NICHD	$1,243,000		$1,233,000		
NIH Total		**$18,103,000**		**$21,560,000**	27
USDA	$440,000		$6,365,000		
(CREES) USDA (ARS)	$0		$0		
USDA Total		**$440,000**		**$6,365,000**	11
NSF Total		**$1,736,000**		**$3,710,000**	4
TOTAL		**$32,779,000**		**$45,435,000**	87

*Approximate number of organisms. Includes both completed and in progress through FY2000.

Highlights of Agency Activities in Microbial Genomics

Department of Defense (DoD) (http://www.dtic.mil/ddre/):

The DoD investment in microbial genomics is driven by biomedical and non-biomedical interests and by military operational requirements. In the biomedical area DoD is interested in developing technologies that provide health support and services to military personnel and that counter the threat of endemic infectious diseases and biological warfare (BW) agents. A major focus of the DoD investment in microbial genomics is, therefore, directed at developing genomic-based information about infectious agents, that can be exploited for the rational design of therapies, vaccines, detection, and medical diagnostic strategies. DoD, along with the NIH and international partners Wellcome Trust and Burroughs Wellcome Fund, is supporting the sequencing of the entire genome of the malaria parasite *P. falciparum.* DoD has also recently begun a new effort to determining the sequences of nine, novel plasmids found in marine sediment microbes. Other microbes and pathogens are sequenced as needed to assist in the development of vaccines, drugs, and/or diagnostic tests. In an interagency effort, DoD, DOE, and NIH are supporting the effort to sequence the genome of *Bacillus anthracis,* the causative agent of anthrax and a potential biowarfare threat agent.

In the non-biomedical area, DoD is interested in biotechnological approaches for developing new materials and managing the impact of DoD operations on the environment. Information emerging from functional genomics research should enable new technologies to develop novel biosynthetic schemes for producing materials of interest to DoD, and should provide a better understanding of processes governing the fate and effects of contaminants in marine and terrestrial sediments. Genomic information about novel, naturally occurring plasmids could enable the development of new biotechnology-based tools for manipulating microbes. To benefit fully from the information provided by genomic sequence analysis requires tools that enable prediction of the structure, function, regulation, and physiological impact of gene products. To this end, DoD pursues programs that fully integrate genomic sequence and functional genomics research.

Department of Energy (DOE) (http://www.doe.gov/):

The DOE Microbial Genome Program (http://www.ornl.gov/microbialgenomes/index.html), established in 1994, was the first U.S. government effort supporting the sequencing of microbial genomes and continues to provide microbial DNA sequence information to further the understanding and application of microbiology relating to DOE's mission areas of energy production, chemical and materials production, environmental carbon sequestration, and environmental cleanup. To date, the Microbial Genome Program has supported the complete genomic sequencing of 17 microbes (9 published) with 29 additional microbes (including one fungus) in various stages of progress. The closely integrated Natural and Accelerated Bioremediation Research Program in the Office of Biological and Environmental Research (OBER) provides much of the rationale for the microbes that DOE selects for whole genome sequencing, and separately supports several microbial research projects on bioremediation of radiation, heavy metals and chelating agents. The elucidation of microbial genome sequences remains a natural outgrowth of past and current BER Programs, including DNA sequencing from the Human Genome Program, structural biology studies utilizing BER-supported facilities and synchrotrons located at DOE laboratories, microbial physiological and biochemical studies supported by the Basic Energy Sciences program, and molecular microbiological research supported by BER environmental programs. The MGP benefits directly from capabilities at DOE national laboratories, DOE and NIH Human Genome Centers, the NCBI at the NIH, and the capabilities of universities and non-profit organizations. Over the last 7 years, sequencing of microorganisms that live in extreme

environments (including the deep subsurface, geothermal environments, hypersaline environments, high-radiation environments, and toxic waste sites) has provided a considerable information base for scientific research related not only to DOE missions but also to other Federal agency missions, and U.S. industry.

To date, the focus of the MGP has been on high-throughput microbial whole genome sequencing. The MGP is now shifting its emphasis to the elucidation of the biological information content of those sequences in order to address DOE mission challenges. The new thrusts comprise: whole-genome functional analyses, bioinformatics applied to microbial genome sequences, characterization of microbial genomic plasticity, novel microbial sequencing approaches, and the characterization of the diversity of microbial consortia and/or hard-to-culture microbes that mediate processes of relevance to the DOE. Candidate microorganisms of interest to the DOE MGP can include archaea, bacteria, or communities made up of bacteria and/or archaea that mediate or catalyze metabolic events of energy or environmental importance. Microbes for which complete or near-complete genomic sequencing information in the public domain exists can be viewed at http://www.ornl.gov/microbialgenomes/organisms.html. Additional microbes that are presently being sequenced or have been sequenced to "high draft" (about 8x) coverage at the DOE Joint Genome Institute can be viewed at http://spider.jgi-psf.org/JGI_microbial/html/. In general, priority is given to studies on those microbes that generate potential energy compounds (e.g. fuels, chemicals such as hydrogen or methane), can bioremediate metals and radionuclides, can degrade significant biopolymers such as celluloses and lignins or are involved in environmental carbon sequestration, e.g. CO_2 fixation. Finally, microbes that participate in consortia with already-sequenced species are of interest. Strict pathogens or parasites are usually not considered.

The DOE also supports microbial research that addresses its energy mission through the Energy Biosciences program, the Energy Efficiency program and the Fossil Energy program. The Energy Biosciences program supports mechanistic research on fundamental biological processes related to capture, transformation, storage and utilization of energy. The Energy Efficiency program supports a variety of projects focused on cellulase biochemistry, advanced development of cellulase enzyme systems to more cost effectively convert cellulose to sugars, and the characterization and modeling of microbial and fungal cellulase action on biomass. The Fossil Energy program supports several activities exploring the bioprocessing of high sulfur crude oil, potential biodesulfurization of diesel fuels, and the use of microbial cultures for the removal of contaminants from petroleum feedstocks.

Department of Interior (DOI), U.S. Geological Survey (USGS) (http:// biology.usgs.gov):

The USGS has initiated more than a dozen individual research efforts in the last few years that develop and apply microbial genetic information to natural resources research. These include identifying microbes that could be used to control invasive species such as the brown tree snake, examining the effects of commercial additives in the treatment of municipal sludge to determine the effect on microbial populations required for successful anaerobic digestion of municipal wastewater, and documenting the distribution of known and potential microbial pathogens that may be affecting the sustainability of the health of the Salton Sea ecosystem.

Environmental Protection Agency (EPA) (http://www.epa.gov):

EPA has a broadly mandated mission to protect human health and the health of the nation's ecosystems. Under this mandate, the EPA's Office of Research and Development (ORD) supports a broad array of intramural and extramural basic and applied research programs aimed at assessing and reducing the risks to

humans and ecosystem biota from exposure to pollutants. A segment of the ORD's intramural research program focuses on understanding the biology of harmful and beneficial microorganisms found in the environment. Included in their portfolio are studies on: fungi that are associated with sick-building syndrome, asthma and acute pulmonary hemorrhage/hemosiderosis, and infant mortality; parasitic protozoan water and food contaminants responsible for a variety of human diseases that have been increasing in incidence; organisms associated with harmful algal blooms which cause toxicity in both humans and wildlife; complex microbial communities found in biofilms associated with water distribution systems; bioremediation studies on anaerobic bacteria that colonize the roots of aquatic and wetland plants and contribute to the beneficial effects of wetlands in removing toxic contaminants from water; thermophilic methanotrophic bacteria which have the ability to degrade chlorinated solvents and other organics; bacterial species which sequester lead and other heavy metals.

The ORD also supports extramurally funded microbial research through its National Center for Environmental Research. Project descriptions can be found on the Center's web site (http://es.epa.gov/ncerqa/grants/). In addition ORD's research activities, the EPA's Office of Prevention, Pesticides and Toxic Substances has a strong interest in microbial genomics because of its need to differentiate genetically modified organisms at the species level and to evaluate patterns of lateral gene transfer and rearrangements in support of regulatory decision making.

Food and Drug Administration (FDA) (http://www.fda.gov):

FDA conducts research and surveillance on microbial pathogens and provides educational outreach on microbial hazards as fundamental parts of its mission to protect the public health. Both fundamental and applied research are focused on pathogens that endanger human health and the safety of food animals. Since a current science base underpins the regulatory roles of the Agency, the outgrowths from information on microbial genomes will continue to streamline the interaction of regulated industries with FDA and better serve the public health. FDA has provided extramural support for genome-scale comparisons among pathogenic strains of *E. coli,* aimed at identifying sequences that contribute to unique characterisistics in these pathogens. Within FDA, the development of DNA chip technologies has been coordinated among the Centers of the Agency in order to foster the development and use of standardized methods that will support the regulatory mission of FDA. Application of these technologies to a vast array of microbial sequences will make possible:

- Use of novel sequence elements for the rapid classification of pathogens in the clinical, community, or environmental setting;
- Understanding of pathogen emergence as determined by gene acquisitions that modify the traits of familiar—and currently controlled—pathogens;
- Identification of new targets for antimicrobial agents or treatments, especially as a means to overcome antimicrobial resistance;
- Using genomic information for surveillance of drug resistant organisms and development of treatment strategies to combat the emergence of antibiotic resistance.

National Aeronautics and Space Administration (NASA) (http://www.nasa.gov):

NASA seeks to understand the nature of life in the universe, and to assure astronaut health and productivity for increasing periods of time and at greater distances beyond Earth. NASA's interest in Microbial Genomics lies in the functional genomics of organisms in extreme environments, including the space environment. While NASA does not develop fundamental genomic technologies, NASA investigators regularly

use techniques including computational biology, bioinformatics, in situ genomic analyses, medical genomics, and genomics as a basis for engineering. NASA has a continuing interest in using the tools of genomics to enable correlation of environmental changes with changes in gene expression, gene products, metabolic effects, and structural changes over multiple generations. Through its programs in Astrobiology and the Office of Biological and Physical Research, we expect to fund continued use of this technology.

NASA's Astrobiology program seeks to understand the origin, evolution, distribution and destiny of life in the universe. NASA's investment in microbial genomics within this program is centered on discovering phylogenetic relations between organisms to determine our last common ancestor, to investigate life's earliest metabolic capabilities, and to infer Earth's earliest environments. Functional genomic studies of microbes in extreme environments on Earth, especially those conducted in situ, provide models for understanding the limits of life and the nature of habitable environments. Within the area of Biological and Physical Research, NASA uses microgravity and other characteristics of the space environment to enhance our understanding of fundamental biological processes, and to develop technological foundations for a human presence in space. Microbial genomics is essential to our understanding the response of terrestrial microbial life to the space environment, and to support human exploration beyond our planet. The Fundamental Biology Program and its Biomedical Research and Countermeasures Program, within the Life Sciences Division, have strong interests in supporting genomics research, focused on integrated and functional genomics as tools to understand complex biological pathways and systems, and how their interactions might support or interfere with human spaceflight

NASA has a health-related mission interest as well. Microbes represent a health hazard to human exploration crews, either in their natural state or through possible mutations brought about by novel selection pressures, including the closed environment of the spacecraft, microgravity effects, and radiation induced changes. For the future, habitable artificial ecologies designed to operate beyond earth over decade-long time periods will almost certainly employ microbes as part of their life support strategies, including those that are bioengineered for specific functions.

National Institute of Standards and Technology (NIST) (http://www.nist.gov/):

The National Institute of Standards and Technology was established by Congress "to assist industry in the development of technology ... needed to improve product quality, to modernize manufacturing processes, to ensure product reliability ... and to facilitate rapid commercialization ... of products based on new scientific discoveries." In regard to microbial research, NIST is involved in a number of projects that directly impact microbial research and an even greater number that broadly affect genomic research in general. This research can be divided into five broad technology areas: bioinformatics, structural genomics, protein/metabolic engineering, standardization, and DNA diagnostics.

In the area of bioinformatics, NIST, with Rutgers University and UCSD, is part of the Research Collaboratory for Structural Bioinformatics (RCSB, http://www.rcsb.org), and is the managing organization of the Protein Data Bank (http://www.pdb.org), the international repository for all 3-dimensional biological macromolecular structure data. NIST's role in the PDB is to assure uniformity of the data for accurate querying, and archiving the database. Additionally, NIST maintains the "Thermodynamics of Enzyme Catalyzed Reactions" database, NIST Standard Reference Database 74 (http://wwwbmcd.nist.gov:8080/enzyme/enzyme.html), a database that contains the thermodynamic properties of numerous enzymatic reactions that are currently primarily along the aromatic amino acid synthesis pathway.

Structural genomics, or proteomics, is the study of the function of newly discovered gene products or proteins through the structure of those proteins. NIST has partnered with the University of Maryland Biotechnology Institute and The Institute for Genomic Research in the first NIH awarded structural genomics program project (http://s2f.carb.nist.gov/; http://www.structuralgenomics.org/) beginning in 1998. As of May 2000, 31 proteins of unknown function were expressed and purified from open-reading frames of the *Haemophilus influenzae* genome. Of these, 21 proteins have been crystallized (9 of which had diffraction-quality crystals), and 9 structures have been solved (7 by X-ray crystallographic methods and 2 by NMR methods), in this ongoing research to date.

There is a great effort to incorporate biological production of chemical products in order to make better drugs more economically, to avoid costly toxic chemical processes, and to make new environmentally-friendly, biodegradable materials. To accomplish this, proteins have to be re-engineered to work in industrial environments or to process non-native substrates, which is termed protein engineering. NIST's efforts in the protein engineering area include helping to re-engineer a generic protease from *Bacillus subtilis*, subtilisin, to work more efficiently in laundry detergent, thus replacing environmentally unsafe chemical detergents. Additionally, organisms may have to be re-engineered so that their metabolic pathways will overproduce either native or non-native products to be used commercially, such as drugs, dyes or polymer precursors, which is called metabolic engineering. NIST is currently doing research to evaluate the kinds and amounts of information required to effectively pursue metabolic engineering, focusing on the aromatic amino acid synthesis pathway, responsible for the production of a number of industrially important chemicals, such as aspartame, indigo dye and nylon precursors. This work compliments the "Thermodynamics of Enzyme Catalyzed Reactions" database research.

Currently, there is interest in obtaining a "standard" cell, a fully characterized organism, for testing environmental, process and growth conditions in a reproducible manner. The Advanced Technology Program is attempting to evaluate and eventually fund solutions to this problem in both eukaryotic and prokaryotic systems. This standard is expected to be very important for industrial calibration, process design and product uniformity. NIST has already established a number of DNA diagnostic standards in the areas of forensics, some specifically for the evaluation of PCR methods, which makes them applicable to the area of microbial genomic studies as well.

As data, information and research with respect to microbial research becomes increasingly important to industry and the National economy, NIST will be focusing more of its efforts and funding in this area.

National Institutes of Health (NIH) (http://www.nih.gov):

NIH supports and conducts biomedical research that will uncover new knowledge leading to better health. NIH conducts research in its own laboratories and supports the research of non-Federal scientists in universities, medical schools, hospitals and research institutions throughout the country and abroad, helps in training research investigators, and supports fostering communication of medical information. NIH is comprised of 25 separate Institutes and Centers, and has a budget of more than $17.8 billion in 2000.

National Institutes of Allergy and Infectious Disease (NIAID, at http://www.niaid.nih.gov) supports research on microbial pathogens that are responsible for diseases of public health importance both domestically and globally, spanning basic biomedical research, such as studies of microbial physiology and antigenic structure, to applied research, including the development of diagnostic tests and the conduct of clinical trials to evaluate experimental drugs and vaccines. NIAID supports projects on microbial genomics, which are

expected to enhance understanding of the pathogen's biology and its ability to cause disease, leading to new strategies to prevent and treat infections. NIAID has supported the sequencing of 11 microbial pathogens, and is supporting the sequencing of more than 30 other microbial genomes. NIAID also funds work in the bioinformatics and functional genomics of human pathogens, is establishing a Pathogen Functional Genomics Resource Center to distribute genomic resources and technology to the research community, and currently supports an Orthopoxvirus Genomics and Bioinformatics Resource Center and, through an Interagency Agreement with DOE, a database for sexually transmitted pathogens (http://www.stdgen.lanl.gov).

The National Institute of Dental and Craniofacial Research (NIDCR, at http://www.nidcr.nih.gov) supports basic, clinical, translational, epidemiological and developmental research on infectious diseases of the oral cavity. NIDCR supports sequence analysis of entire microbial genomes, which promises to yield a comprehensive picture of the structure and function of microorganisms. Genome analysis may be able to elucidate previously unrecognized pathogenic mechanisms that can be blocked by drug therapies, and immunogenic components ideal for vaccine development. NIDCR currently supports the complete sequencing of five oral pathogenic bacteria and a yeast (for the last, NIDCR is partnering with The Wellcome Trust).

The mission of the National Human Genome Research Institute (NHGRI, at http://www.nhgri.nih.gov) is to head the Human Genome Project (HGP) for the NIH (an international research effort to characterize the genomes of human and selected model organisms through complete mapping and sequencing of their DNA), to develop technologies for genomic analysis, to examine the ethical, legal, and social implications of human genetics research, and to train scientists who will be able to utilize the tools and resources developed through the HGP. The NHGRI has supported complete genome sequencing of two model microbes, the prokaryote Escherichia coli and the eukaryote Saccharomyces cerevisiae. With the completion of the sequence of these and other whole genomes of model organisms, NHGRI has begun to develop programs in the analysis of genomic sequences. NHGRI's specific interest in microbial genomics is in the analysis of the genome of the baker's yeast, S. cerevisiae, including support of large-scale functional analyses and the Saccharomyces Genome Database (SGD). NHGRI has a particular interest in supporting large-scale functional studies in S. cerevisiae as a model of these types of projects in other eukaryotic organisms, both from a technical point of view and because of the challenge that analyzing this type and scale of data poses. NHGRI is also putting a major emphasis on research to reduce the cost of DNA sequencing. Other relevant activities include efforts to develop or improve technologies for functional analyses, including analysis of RNA and protein expression, protein interactions, genetic mapping and sequence variation, and mutagenesis. Emphasis is on technologies that can be used on a large scale, are efficient and are capable of generating complete data for the genome as a whole.

The mission of the National Institute of General Medical Sciences (NIGMS, at http://www.nigms.nih.gov) emphasizes the importance of understanding fundamental life processes in the most advantageous systems available. Because of the wealth of genetic and molecular information available on microbes, particularly that resulting from current genomics advances, NIGMS supports extensive research on microbes. One such study resulted in the first genome-scale description of protein-protein interactions in yeast. Currently, NIGMS supports an effort to determine the function of all open reading frames in the *E. coli* genome, and is supporting a number of microbial projects that have developed genomics approaches. NIGMS recognizes that to reach the stated research goals, research training in computation and bioinformatics will be needed, and NIGMS is committed to support training programs in the area of systems and integrative biology, bioinformatics and computational biology and fellowships in quantitative biology.

The National Center for Research Resources (NCRR, at http://www.ncrr.nih.gov) has a responsibility

at NIH to develop critical research technologies and to provide cost-effective, multidisciplinary resources to biomedical investigators across the spectrum of research activities supported by the NIH. NCRR provides a broad array of technologies, tools, and materials to carry out research in microbial genomics. For example, NCRR is supporting two mass spectrometry centers that are developing new techniques for directly identifying proteins from large complexes based on knowledge of their molecular masses deduced from complete genome sequences. The Yeast Resource Technology Center has been established to exploit the yeast genome sequence, and integrates a set of state-of-the-art analytical technologies, including mass spectrometry, two-hybrid analysis, and microscopy. The Shared Instrument Grant Program (SIG) provides key instruments needed to analyze microbial genomes including high-throughput protein and DNA sequencers, sequence detector systems, and DNA chip technologies.

The National Center for Biotechnology Information (NCBI at http://www.ncbi.nlm.nih.gov), a component of the National Library of Medicine (NLM), is a national resource for molecular biology information and creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. NCBI's interests lie in the computational analysis of microbial genomes. In addition to curating and maintaining GenBank, NCBI supports an in-house effort in computational biology and bioinformatics focused on the analysis of microbial genomes (http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html). The core of the NCBI's effort in this direction is the database of Clusters of Orthologous Genes (COGs).

National Oceanic and Atmospheric Administration (NOAA) (http://www.noaa.gov):

Marine microbial genomics plays a critical role in enabling NOAA to satisfy its mission of Environmental Stewardship. Specifically, it plays a role in NOAA's goals of sustaining healthy coasts, building sustainable fisheries, and recovering protected species. Microbial genomics provides NOAA with new opportunities to effectively address the challenge of improving the health and productivity of oceans, predicting ecosystem changes, and providing fundamental information for use in improving management of fisheries and coastal habitats. To accomplish these objectives, NOAA supports peer-reviewed intra- and extramural research, education, and outreach in the area of genome-enabled science to study and monitor marine microbiota. The goals of these efforts are to assess the levels and effects of viruses, bacteria, protozoa, microalgae, and parasites on the health of coastal ecosystems and natural resources, and to discover and characterize novel marine natural products, drugs and processes and employ them in a myriad of important applications.

NOAA currently supports peer-reviewed intra- and extramural research on bacteria and parasites that impact coastal ecosystems, aquaculture, and wild fish and shellfish harvest. Microbial pathogens are a significant source of morbidity and mortality in fish and shellfish raised in hatcheries, aquaculture, and in programs aimed at the restoration of endangered species of Pacific salmon. For example, NOAA has a program that is characterizing virulence determinants of salmonid bacterial pathogens at the genetic level with a goal towards developing more effective therapeutics or vaccines, as well as improved diagnostic or molecular differentiation tools. The development of molecular techniques to effectively monitor the safety of seafood is a priority. NOAA has also initiated research to identify and characterize microbial pathogens of corals to determine the underlying molecular and cellular cause(s) of declines in coral health.

NOAA is currently supporting several small projects to develop specific molecular probes for organisms that cause harmful algal blooms, such as brown tide and toxic dinoflagellates, and to develop probes for

and control of toxic events since many of the toxins produced can be concentracted in the food chain and harm animals and humans. NOAA is also investigating the potential for *in situ* bioremediation of marine sites contaminated by toxic chemicals. The research shows that *in situ* bioremediation is promising and can be less disruptive to an ecosystem.

NOAA currently has a small effort to identify novel marine organisms in extreme environments (deep-sea vents and sea ice) with the ultimate goal of identifying new products and processes from these very primitive archaea and bacteria. To date, NOAA has concentrated on recovering, identifying and characterizing organisms that live in extreme temperatures. NOAA scientists are working with other agencies and other scientists to understand the ecology and requirements of the various organisms. Several industrial partnerships are now under development, as there may be considerable industrial potential for enzymes found in the high temperature organisms. Low temperature organisms, and the external polymers that coat them – have potential application in industrial processes involving below-freezing processes or products, for refrigeration, improved shipment for fish, fish eggs etc.

The emergence of antibiotic resistance of many bacterial pathogens highlights the importance of developing different antibiotics. Bacteria from the marine environment have the potential to produce these novel substances because they live in unique systems, vastly different from their terrestrial counterparts. NOAA has supported a small effort to discover and characterize novel antibiotic producing marine microorganisms, as well as characterizing of symbiont microbes involved in the production of many marine products.

National Science Foundation (NSF) (http:www.nsf.gov):

NSF supports microbial research in a broad range of areas, including environmental and evolutionary biology, metabolic biology and engineering, genetics, and oceanography, all may involve collaborations with mathematics, computer sciences, and chemistry. NSF's interest in microbial genomics parallels its support of microbiological research, including microbes of interest in basic research, microbes that occupy critical or compelling environmental or evolutionary niches, microbes developed for metabolic engineering, and microbes that interact with or are models for the higher eukaryotic systems supported by NSF, such as plants.

NSF supports microbial genomics through a number of established programs in multiple directorates, including the Biology, Engineering, and Geosciences Directorates, and the Office of Polar Programs. Several Foundation-wide initiatives (Microbial Observatories, The Plant Genome Program, Biocomplexity in the Environment and Information Technology Research) also have components that address issues of importance to the study of microbial genomics. Finally, unsolicited proposals involving microbial genome sequencing are received and reviewed through the regular NSF programs. By this mechanism, NSF funded the complete sequencing of a salt-loving archaea (now finished) and the filamentous fungus *Neurospora crassa*, and a project to develop genomic clones in preparation for sequencing a dinoflagellate.

Through its Computational and Database Activities in the Biological Sciences programs, NSF invests in the ever-increasing computational and database needs that provide infrastructure not only for microbial genomics but for many areas of life science. Examples include projects dealing with genomic mapping data, the Ribosome Database Project (RDP), a microbial biodegradation database. The NSF also supports the Protein Data Bank (PDB), funded jointly with NIH, DOE and NIST. All seek to accommodate the rapid growth in storable information, allow complex querying, and facilitate access to data as well as linkage to

and integration with other databases.

NSF also funds research facilities and centers in the area of microbiology, including the Science and Technology Center for Microbial Ecology at Michigan State University, and the Advanced Microbe Isolation Laboratory at Oregon State University. Researchers there will develop automated approaches for culturing and identifying novel microorganisms from natural ecosystems.

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 the IGERT program. A new microbial biology postdoctoral program was initiated in FY2000 and is expected to continue for at least 3 years. The goal of this program is to develop a cohort of scientists trained in non-model microbial systems and microbial systematics. NSF also provides postdoctoral fellowships in Biological Informatics, to address the need for scientists trained in computational biology and bioinformatics. In addition, programs in the Education and Human Resources Directorate (EHR) have supported laboratory research, curriculum development, and education in the area of microbiology and genomics.

Department of Agriculture (USDA) (http:// www.usda.gov/):

The USDA supports research, education and extension in the biological, environmental, physical, and social sciences to address regional and national problems and opportunities relevant to agriculture, food, forestry, and the environment. Microbial genomics is a high priority investment area, because it is essential for carrying out the mission of the USDA. Research in this area is critical for advances in food safety, food security, biotechnology, value-added products, human nutrition and functional foods, plant and animal protection and furthering fundamental research in the agricultural sciences. To maintain this nation's competitiveness, the USDA has identified four major objectives: (1) Assure that the complete nucleic acid sequences of high priority beneficial and detrimental agricultural microorganisms are available in public databases; (2) Assure that the agricultural research community has adequate resources and facilities available for the functional analysis of agricultural microbes (e.g., expression array technologies; proteomics; relational databases and other bioinformatics tools) so that practical benefits are not delayed; (3) Support training and extension for microbial genomics and its evolving technologies; and, (4) Foster U.S. interests through national and international public and private partnerships in microbial genomics, and through such partnerships, facilitate capacity development in the U.S. and abroad that ensures public access and appropriate use of intellectual property.

In FY2000, the extramural research arm (CSREES) of the USDA launched a new microbial genomics initiative, through the National Research Initiative (NRI) and The Initiative for Future Agriculture and Food Systems (IFAFS). Through this initiative, CSREES supported genome sequencing for six animal pathogens and two beneficial rumen microbes, expressed sequence tag (EST) projects for three plant pathogens, and genomics and bioinformatics projects related to agricultural microbes. The intramural Agricultural Research Service (ARS) also supports a number of microbial studies, which are integral components of the USDA national programs in animal health, food animal production, food safety, plant and microbial genomics, and plant diseases. ARS activity on functional genomics is primarily in conjunction with NSF Plant Genome Research Program projects, which involve EST sequencing, identification of 'unigene sets' for each species, and DNA microarrays. The technology base thereby developed in agency laboratories and their collaborators will facilitate planned studies on the genomes of important agricultural microbes and pathogens.

USDA established the ARS Bioinformatics Working Group (ABWG) to help facilitate access for its scientists to bioinformatics databases and the tools required to effectively utilize genome information, and coordinate this effort across species (animals, plants, insects, microbes) and agency programs. A key element of the ABWG strategy is training (initially via a series of quarterly workshops), which will bring together ABWG experts and USDA-ARS students, scientists, and staff.

Glossary

Annotation: The assembling of information of several distinctive types, starting with DNA sequence data and extending to varying degrees of complexity. For example, DNA sequence information may be segmented into distinct intervals that may be identified as encoding specific types of "product," such as proteins, transfer RNAs, and phage sequences. At a higher level of annotation, a protein that is encoded by a particular gene may be annotated in terms of its physical attributes, such as molecular weight, membrane spanning regions, structural domains, or three-dimensional structure. Moreover, annotation at the level of comparative biology may include information linking a particular protein from a specific microorganism to similar proteins from other organisms or to members of similar protein families.

Biogeochemical Cycles: The circulation of chemical components through the biosphere from or to the earth, atmosphere, or bodies of water.

Bioremediation: The process by which living organisms act to degrade or transform hazardous organic contaminants.

cDNA: DNA that is synthesized from a messenger RNA (mRNA) template. mRNA is copied from the chromosomal DNA, and contains only the protein-encoding information of a gene.

Clone: This term can refer to genetically identical cells produced by mitotic divisions from one original cell, genetically identical organisms all descended from the same single parent by asexual processes, or DNA molecules derived from one original length of DNA sequences and produced by a bacterium or virus using genetic engineering techniques.

Comparative Genomics: The practice of comparing the gene or protein sequences of different organisms with the goal of elucidating functional and evolutionary significance.

DNA (deoxyribonucleic acid): The molecule that encodes genetic information. DNA is a double stranded polymer, the subunits of which are called nucleotides. Nucleotides have three parts: a sugar, a phosphate and a base. Only four nucleotides are used to build a DNA molecule, which differ by the base they contain: adenine (A), guanine (G), cytosine (C), or thymine (T). The two strands are held together by weak bonds between the bases of the nucleotides. In nature, base pairs form only between A and T and between G and C; thus the base sequence of each single strand can be deduced from that of its partner.

DNA Library: An unordered collection of clones (i.e., cloned DNA from a particular organism), whose relationship to each other can be established by physical mapping.

EST: Expressed Sequence Tag: A unique, short DNA sequence derived from a cDNA library. ESTs are useful for localizing and orienting the mapping and sequence data reported from many different laboratories and serve as identifying landmarks on the developing physical map of a genome.

Expression Pattern: Gene expression is the process by which a gene's coded information is converted into the structures or molecules present and operating in the cell. Expression pattern refers to a set of genes expressed under a set of conditions (e.g., genes expressed in microbes grown in the presence of oxygen may differ from those expressed in microbes grown in the absence of oxygen).

Functional Genomics: Studies of the relationship between the structure and organization of the genome and the function of the genome as it directs growth, development, physiological activities, and other life processes of the organism.

GenBank: A public database where DNA sequences are deposited. It is operated and supported by the National Library of Medicine, part of the National Institutes of Health, and is part of an international consortium of gene sequence databases.

Gene: The fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides located in a particular position on a DNA molecule located in a particular chromosome that encodes a specific functional product (i.e., a protein or RNA molecule).

Genetic Map: A map of the relative positions of genetic loci on a chromosome, determined on the basis of how often the loci are inherited together.

Genetics: The study of the inheritance of specific traits.

Genome: All the genetic material in the DNA of a particular organism; its size is generally given as its total number of base pairs.

Genome Project: Research and technology development effort aimed at mapping and sequencing some or all of the genome of human beings and other organisms.

Genomics: Activities associated with genome mapping and sequencing, as well as the use of information derived from genome sequence data to further elucidate what genes do, how they are controlled, and how they work together.

High Throughput Biology: An experimental approach that generates massive amounts of raw data at the production scale using highly automated technologies such as genome sequencing technology or microarray technology, and processes the data using computational and other information management tools.

Informatics: The application of computer and statistical techniques to the management of information. In genome projects, informatics includes the development of methods to search databases quickly, to analyze DNA sequence information, and to predict protein sequence and structure from DNA sequence data.

Microarray Technology: Microarray technology is one of several developing approaches to comparatively analyze genome-wide patterns of gene expression. Terms to describe this technology include, but are not limited to: biochip, DNA chip, DNA microarray, gene chip, and gene array. DNA microarrays are fabricated by high-speed robotics, such that thousands of different DNA sequences get attached to a solid support in an orderly pattern, like checkers on a board. These pieces of DNA act like probes. When used to study transcription, an investigator collects cells of interest, isolates the mRNA from the cells, labels it with a fluorescent dye, and passes it over a chip. The mRNA grabs onto ("hybridizes to") the gene it came from. The hybridization site is detected by the fluorescent tag, and reveals the identity and expression level of genes expressed specifically in the test cells. Microarray technologies are also being developed to look at levels of protein expression, protein modifications, and protein interactions.

Microbe: For purposes of this report, microbes (or microorganisms) are organisms that are too small to be seen by the naked eye, and include viruses, bacteria, fungi, protozoa, and microalgae.

Phylogenetics: The study of the evolutionary history of a group of organisms to identify how they are related to each other via common ancestry. Often this is depicted as an evolutionary tree.

Physical Map: A map of the physical locations of identifiable landmarks on DNA (e.g., restriction enzyme cutting sites, genes); distance is measured in base pairs. The highest resolution map would be the complete nucleotide sequence of the chromosomes.

Proteomics: The identification and quantification of the tens of thousands of proteins in a given organism to define patterns of protein expression. This information can then be used to characterize functional cellular processes, such as those involved in development, the cell cycle and cell death, in response to pharmaceutical intervention or extracellular stimuli and toxic agents, and in disease.

Sequencing: Determination of the order of nucleotides (sequence of bases) in a DNA or RNA molecule, or the order of amino acids in a protein.

Structural Genomics: Studies to determine the 3-dimensional structures of all proteins encoded in a genome. May also refer to studies of the structure and organization of the genome itself.

Symbionts: Organisms that form close, often mutually beneficial, associations with other organisms.

Technology Transfer: The process of converting scientific findings from research laboratories into useful products by the commercial sector.

References

Recent Literature References:

1. K.E.Nelson, I.T. Paulsen, J.F. Heidelberg and C.M Fraser. 2000. Status of genome projects for nonpathogenic bacteria and archea. *Nature Biotechnology* 18:1049-1054.

2. Claire M. Fraser, Jonathan A. Eisen, Steven L. Salzberg. 2000. Microbial Genome Sequencing. *Nature* 406:799-803.

URLs

Reports

Interagency Report on the Federal Investment in Microbial Genomics http://www.ostp.gov/html/microbial/start.htm

American Academy of Microbiology Report *Microbial Genomes: Blueprints for Life* http://www.asmusa.org/acasrc/pdfs/genome.pdf

A Workshop on Marine Microbial Genomics: Advice to the NSF http://www.ocean.udel.edu/genomics/genomicsindex.html

ASM Report *Recommendations Related to Microbial Genome Sequence Analysis and Annotation* http://www.asmusa.org/pasrc/microbialgenome.htm

Agency Microbial Sequencing Web Pages

DOE Microbial Genome Program	http://www.ornl.gov/microbialgenomes/index.html
NIAID Pathogen Genome Program	http://www.niaid.nih.gov/dmid/genomes/default.htm
NIDCR Microbial Genome Projects	http://www.nidcr.nih.gov/research/extramural/Pol_Microbial_Genome_Seq_Projects.htm

Lists of Sequenced Microbes (with links to individual microbe pages)

TIGR Microbial Database	http://www.tigr.org/tdb/mdb/mdbcomplete.html
NCBI Entrez Genomes	http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html
Sanger Center Microbial Genomes	http://www.sanger.ac.uk/Projects/Microbes/
DOE Microbial Genome Program	http://www.ornl.gov/microbialgenomes/organisms.html

General Reference

Microbe World	http://www.microbeworld.org/
The Genomics Lexicon	http://209.52.56.28/lexicon/index.html
The “Bad Bug Book”	http://vm.cfsan.fda.gov/~mow/intro.html

EXECUTIVE OFFICE OF THE PRESIDENT

Office of Science and Technology Policy

Washington, D.C. 20502