

The Microbe Project
Interagency Working Group
Coordination Plan
March 2008

EXECUTIVE OFFICE OF THE PRESIDENT
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL
WASHINGTON, DC 20502

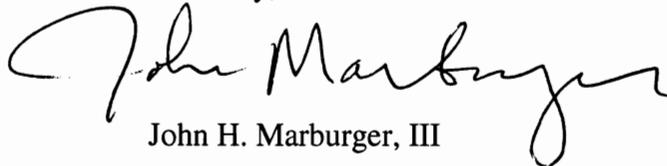
Dear Colleagues,

The study of microbes and complex microbial communities is an important challenge even at the outset of the 21st century. While microbes were discovered centuries ago, only in recent history have sophisticated molecular genetic studies of microbes and diverse microbial communities become possible. As a 2007 National Research Council report noted, "Microbes run the world. It's that simple." Their importance in human, plant and animal health, and also to other global processes, underscores the need to better understand how microbes function at all levels, from the molecular to the ecosystem.

The 2008 *Coordination Plan of the Microbe Project Interagency Working Group* (MPIWG) describes how federal agencies will work together to promote and exploit the latest discoveries in genome-enabled microbial science to achieve common goals and address national needs. The plan highlights two major areas for future coordination: 1) infrastructure and education to provide the foundation for genome-enabled microbial science across all agencies; and 2) genome-enabled microbial research of interest to multiple agencies.

The MPIWG will continue to serve as a forum to foster coordination and collaboration among federal agencies while working closely with the broader scientific community in the U.S. and internationally. This includes working toward implementing the goals identified in this report to maximize the opportunities offered by genome-enabled microbial science to benefit science and society.

Sincerely,

A handwritten signature in black ink, reading "John H. Marburger, III". The signature is fluid and cursive, with the first name "John" being particularly prominent and stylized.

John H. Marburger, III

Table of Contents

Executive Summary	4 - 6
Introduction	7
I. Purpose and Accomplishments of the Microbe Project IWG (2001-2006)	8 - 11
II. Interagency Coordination Plan (2008-2018)	12 - 21
III. MPIWG Next Steps and Implementation Plan	22
Table 1. Roles of MPIWG Agencies in Infrastructure and Education	23
Table 2. Roles of MPIWG Agencies in Coordinated Research Areas	24
Glossary	25
References	26 - 27
MPIWG Membership Roster	28
Acknowledgements	29
Appendix: Training Programs Relevant to Genome-Enabled Microbial Science	30 - 32

Executive Summary

The microbial world is vast and diverse, and the metabolic versatility of microbes is unparalleled. Microorganisms have colonized virtually every environment on Earth ranging from deep sea thermal vents, polar sea ice, desert rocks, guts of termites, roots of plants, to the human body. However, less than 1% of the microbes on Earth have been cultured and studied in the laboratory (R.I. Amann, W. Ludwig, and K.H. Schliefer, *Microbiological Reviews* 59 (1), 143 (1995)). The unknown microbial world represents a potentially enormous resource for new processes and products in biotechnology, human health, energy, agriculture and environmental management (C. Harwood and M. Buckley, *The Uncharted Microbial World: microbes and their activities in the environment*, American Academy of Microbiology, 2008).

Each microorganism contains a genome with the complete set of genetic material often referred to as its “genetic blueprint.” The advent of genomics has created a new era of scientific opportunity, fueling discoveries and innovations for the benefit of society as well as contributing to our fundamental understanding of the natural world. Examples of products from genome-enabled microbial science include new vaccines, detection, diagnostic and treatment technologies, as well as bioremediation and bioenergy strategies.

The Microbe Project Interagency Working Group (MPIWG) was formed in 2000 under the aegis of the Subcommittee on Biotechnology Research, the Committee on Science, of the National Science and Technology Council. According to its Terms of Reference, the mission of the MPIWG is to:

“Maximize the opportunities offered by genome-enabled microbial science to benefit science and society through coordinated interagency efforts to promote research, infrastructure development, education and outreach.”

The MPIWG fosters partnerships, coordination, and information exchange among federal agencies in the area of genome-enabled microbial science. Accomplishments to date include the development of: A) recommendations for microbial sequencing to strengthen the Nation’s biodefense efforts; B) guiding principles for release of microbial genome sequence data prior to publication, and recommendations for database accessibility and interoperability; C) collaborative activities among federal agencies on microbial genome sequencing; and D) a Microbe Project Web Site for improving public awareness of the Federal government’s support of research in the microbial sciences.

This **2008 MPIWG Coordination Plan** describes how the federal agencies will work together to promote and exploit the latest discoveries in genome-enabled microbial science to achieve common goals and address national needs. The MPIWG has identified two major areas for future coordination: 1) **Infrastructure and Education to provide the foundation for genome-enabled microbial science across all agencies**; and 2) **Research in genome-enabled microbiology of interest to multiple agencies**.

Within the category of Infrastructure and Education, the MPIWG identified four areas which are considered to be critical to providing the foundation for genome-enabled microbial science. These include: 1) Microbial Sequencing; 2) Databases and Bioinformatics Infrastructure; 3) Experimental Tools, Techniques and Biological Resources; and 4) and Education and Training. Within each of those areas, the MPIWG also established a set of near-term goals, which will be met by agency and interagency activities.

1. **Microbial Genome Sequencing**: Continue to support the **genome sequencing and functional analysis** of a diverse set of microbial genomes and metagenomes by promoting the development and application of novel sequencing and computational techniques to rapidly sequence and analyze entire microbial genomes.

2. Databases and Bioinformatics Infrastructure: Coordinate the development of **informatics tools and applications** for the management, collection, storage, retrieval, and analysis of large data sets generated by genomics, metagenomics, proteomics and metabolomics studies; promote the **integration** of “-omics” data with clinical, epidemiological and environmental data; support **standards of interoperability**, data validation, quality assurance; and promote **accessibility** of published data relevant to the microbial sciences.
3. Experimental Tools, Techniques and Biological Resources: Enable (and manage within the constraints of biosecurity concerns) **access to biological resources** such as specialized cells and cell lines, strains, and large-insert libraries such as Bacterial Artificial Chromosome (BAC) libraries; and promote and foster the development of **novel culture techniques** for microbes, such as extremophiles and other fastidious microbes, that come from a range of environments.
4. Education and Training: Promote **education and training** of students, scientists and the public about genome-enabled microbial science and promote opportunities that help prepare the next generation of researchers to work in interdisciplinary teams.

The six areas identified under the category of Research include: 1) functional genomics; 2) proteomics; 3) comparative genomics; 4) microbe-host and microbe-microbe interactions; 5) microbial forensics; and 6) environmental genomics. The MPIWG has identified the following goals for each of those areas:

1. Functional Genomics: Promote the **identification and determination of function** of the complete set of expressed genes and regulatory elements in diverse microbes including those involved in disease, environmental remediation, carbon processing and biotechnology; promote the development and use of **low cost, high-throughput methods for assigning function**, particularly of hypothetical genes.
2. Proteomics: Promote the development and use of high throughput methods for producing proteins and native protein complexes with retention of structural and compositional integrity; support the development of novel experimental and computational techniques with high sensitivity and specificity to **detect, quantify, and amplify protein complexes**; and encourage research to elucidate **metabolic networks and pathways**, 3D imaging of metabolic process in a cell, and development of proteomic resources and reagents.
3. Comparative Genomics: Encourage the scientific community to establish a framework for the number of genomic sequences needed to represent a specified level of genomic organization, and **criteria for selection of strains for total genome sequencing** within strain groupings; and promote the development and evaluation of technologies for rapid assessment of genomic diversity to guide the choice of strains for whole genome sequencing.
4. Microbe-Host and Microbe-Microbe Interactions: Promote the identification of **common genes and signal molecules involved in communication** by diverse bacteria, microbial genome sequences involved in pathogenesis, and host factors involved in response to microbial pathogens, symbionts and/or commensals for model hosts and promote the development of tools and techniques needed for **large scale metabolomics studies of microbes and hosts**.
5. Microbial Forensics: Promote development of **techniques and criteria for microbial strain identification and attribution** to distinguish natural from introduced pathogens and support the continued development of models of outbreaks that allow comparison of microbial forensics and molecular epidemiology paradigms.
6. Environmental Genomics: Encourage the coordination of **metagenomic data from a range of environments, including the human body**, and support the development of informatics tools to integrate metagenomic data with other environmental data.

The MPIWG will continue to serve as a forum to foster coordination and collaboration among the agencies, and will work toward implementing the goals that have been identified in this report. In particular, the Microbe

Project will: 1) Conduct cross-agency briefings on high priority areas in infrastructure, education and research areas to coordinate and leverage investments; 2) Sponsor or otherwise encourage workshops to identify critical gaps, opportunities and next steps to advance infrastructure, education and research; 3) Continue to coordinate Federal agencies to promote education and training of students, scientists and the public on genome-enabled microbial science; and 4) Maintain, update and encourage the scientific community to make use of the Microbe Project web site.

Introduction

The microbial world is vast and diverse. The Earth is home to over 10^{30} microorganisms, which represent the largest component of the planet's biomass (W. B. Whitman, D. C. Coleman, and W. J. Wiebe, Prokaryotes: The Unseen Majority. *Proceedings of the National Academy of Sciences*, 95: 6578-6583. 1998)). Microbes include bacteria, archaea, mollicutes, fungi, microalgae, viruses and protozoa---organisms with a wide range of morphologies and lifestyles. Microorganisms have colonized virtually every environment on Earth ranging from deep sea thermal vents, polar sea ice, desert rocks, the guts of termites, the roots of plants, to the human body. While microbes are often feared for the diseases they can cause, the vast majority of characterized microbes mediate the essential biogeochemical cycles of key elements (carbon, nitrogen, iron, phosphorus, sulfur and others) that make our planet habitable. Ancient lineages of microorganisms may hold the key to understanding the earliest history of life on Earth.

The metabolic diversity of microbes is unparalleled. Microbes are able to convert toxins to benign products, to invade organisms and cause disease, to photosynthesize under extremely low light, to thrive in the absence of oxygen, to metabolize inorganic materials to organic compounds, and to survive under extreme conditions including sub-zero temperatures, dessication as well as high levels of radiation. Many aspects of microorganisms -- their metabolic versatility, their ability to colonize seemingly inhospitable environments, and their ability to switch between benign and virulent forms--- still remain a mystery. It is estimated that less than 1% of the microbes on Earth have been cultured and studied in the laboratory (C. Harwood and M. Buckley, **The Uncharted Microbial World: microbes and their activities in the environment**, American Academy of Microbiology, 2008). The unknown microbial world represents a potentially enormous resource for new processes and products in biotechnology, human health, energy, agriculture and environmental management.

The genome of an organism is the complete set of genetic material and is often referred to as a “genetic blueprint” of an organism. The advent of genomics has moved us into a new era of scientific opportunity, fueling discovery and innovation for the benefit of society as well as contributing to our fundamental understanding of the natural world. Examples of products from genome-enabled microbial science include new drugs and vaccines, detection and treatment technologies, and bioremediation strategies, and many other products.

At this time (February 2008), more than 2500 microbial genomes had been partially or completely sequenced. To take full advantage of these data, teams of biologists, information scientists, chemists, engineers and biotechnologists will need to work together to mine the information contained in the genetic code. Knowledge of the genome sequence of diverse microorganisms is enabling advances in microbial science and is, in fact, changing the way that science is conducted. Still, the function of approximately 40% of the genes in sequenced microbial genomes remains unknown. These unknown genes may hold the key to scientific advances leading to new products and processes that would benefit society.

Recently a new approach to the study of natural microbial communities has been developed. This approach is known as “metagenomics.” In metagenomics the genomic content of an entire microbial community is sequenced and studied. Metagenomics takes advantage of high performance computing to compare the community-derived sequences with the extensive databases of sequence information from the genomes of individual, identified microbes. This dramatic new approach is already beginning to revolutionize microbial research. A recent report (*The New Science of Metagenomics*, The National Academies Press, 2007) describes metagenomics approaches in detail as well as the way in which they will contribute to many areas of science and society.

I. Purpose and Accomplishments of the Microbe Project IWG (2001-2006)

The Microbe Project Interagency Working Group (MPIWG) was formed in 2000 under the aegis of the National Science and Technology Council's Committee on Science, Subcommittee on Biotechnology. The mission of the MPIWG is to:

Maximize the opportunities offered by genome-enabled microbial science to benefit science and society through coordinated interagency efforts to promote research, infrastructure development, education and outreach.

Each agency in the MPIWG supports fundamental microbial science and/or applied research leading to the implementation of new technologies and approaches in support of the agency's mission. The MPIWG fosters partnerships, coordination, and information exchange among federal agencies in the area of genome-enabled microbial science. The purpose of the 2008 Coordination Plan is to identify the areas in which the federal agencies need to work in the coming years to eliminate technical and scientific barriers and to promote and exploit the latest discoveries in this important area of science.

In its eight-year history, the MPIWG has accomplished the coordination of selection and sequencing of a number of microbial genomes and has also worked to advance technologies that have increased the speed and reduce the cost of sequencing. In addition, it has launched a website through which the group now shares information about research findings and opportunities with the scientific community and the general public. The Microbe Project has made important contributions to the fields of biodefense, information technology, and molecular biology through the work supported by member agencies as well as through interagency efforts.

A. Biodefense

The events of September 11, 2001 and the anthrax attacks the next month in the U.S. served to highlight the importance of developing a broad knowledge of microbes and their roles in human health, agriculture and the environment. Response to an outbreak of a pathogenic agent requires the

ability to detect, diagnose, and treat infected patients. The MPIWG initially addressed biodefense-related issues focusing on the status and coordination of microbial genomics research, then strongly recommended that genomic sequence information be obtained for additional strains of pathogenic organisms.

Subsequently, the MPIWG prepared an extensive inventory of current and planned microbial sequencing projects focusing on human, plant and animal pathogens and their relatives, to aid in the coordination of biodefense-related research planning within the federal government. The information compiled by the MPIWG also contributed to the coordination of interagency efforts in other mission areas of national importance, such as energy, agriculture, environmental management, food safety, and human health.

B. Data and Infrastructure Issues

In order to expand the microbial sequencing efforts, it was necessary to first strengthen the Nation's bioinformatics capability, with the specific goal of developing large-scale, flexible and searchable databases that house and characterize sequence information. Several important issues were identified by the MPIWG as needing additional attention. These included the proliferation of stand-alone databases, the need to increase sequencing of microbes that have not been studied before, and the need to maintain and increase access to biological research tools and facilities. The MPIWG has made substantial progress in addressing each of these issues, as summarized below.

Data-related issues

Database accessibility and interoperability are critical for the efficient use of microbial sequence data. Interoperability across multiple databases facilitates comparisons among microbial genomes by allowing identification of orthologs in related and unrelated species. In addition, interoperability allows the association of genes with ongoing work in proteomics, protein structural characterization, enzymology, biochemistry, and metabolic

engineering to be informed by genetic information from a variety of sources.

To improve the state of microbial databases, the MPIWG collaborated on three workshops:

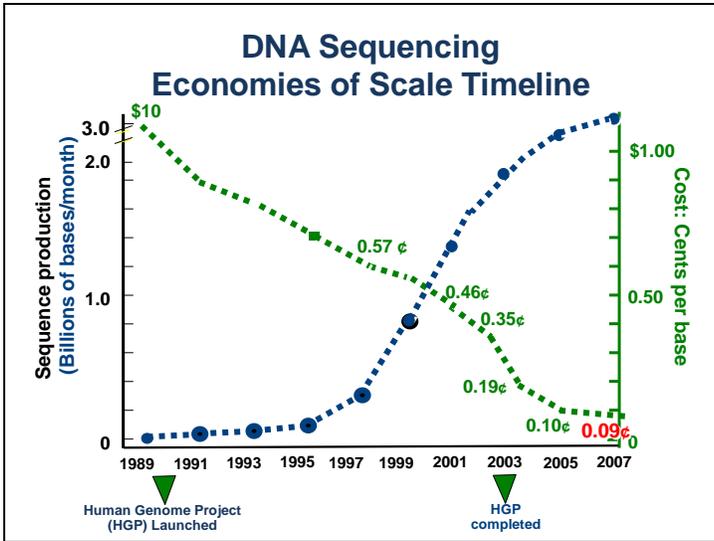
- Microbial database compatibility issues were explored in the first workshop led by the Department of Energy (DOE) that was held on November 28, 2001, in conjunction with the Computational Biology Workshop sponsored by The Institute for Genomic Research.
- A second workshop, held in October 2003 at the National Science Foundation (NSF), focused on technical issues associated with database management, interoperability, and sustainability (http://dpis.engr.iupui.edu/NSFWorkshop/Workshop_report.doc).
- A third workshop, funded by several agencies in the MPIWG, was convened by the American Academy of Microbiology in July 2004. The workshop focused on the problems faced in annotating the many completed microbial genome sequences because of the high frequency of genes of unknown function (<http://www.asm.org/Academy/index.asp?bid=32664>).

Interagency collaboration on genome sequencing

The Microbe Project serves as a highly effective forum for sharing information about genome sequencing projects supported by Federal agencies. Communication of ongoing sequencing projects across the Federal agencies has minimized overlap and allowed for leveraging of funds. Coordination through the MPIWG resulted in an ongoing partnership between USDA and NSF in microbial genome sequencing, through the issuance of an annual joint solicitation on microbial genome sequencing beginning in FY2001. This NSF-USDA partnership has resulted in the sequencing of more than 100 new microbial genomes of importance to agriculture, the environment, and basic science. For example, a research project funded by the joint USDA/NSF microbial genome sequencing program recently resulted in sequencing of *Phytophthora ramorum*, the causative agent of sudden oak death, and its relative *Phytophthora sojae*, an important soybean pathogen. This work

was performed at the DOE Joint Genome Institute. In a similar collaboration, National Institutes of Health (NIH) partnered with NSF to provide support and informatic resources to investigators sequencing the genome of the model ciliate *Tetrahymena thermophila*, the first protozoan to be sequenced. NIH, NSF and the Office of Naval Research (ONR) collaborated to sequence the genome of several isolates of *Bacillus anthracis*, a bioterror agent, as well as non-pathogenic relatives of this important microbe. Genome sequence data are proving to be of great value to the advancement of microbial science. The MPIWG, therefore, believes that “pre-publication” (e.g. “draft” or unfinished) microbial genome sequence data should be made freely and publicly available *via* deposition in publicly searchable databases as rapidly as possible, without restriction. Users of these data have an obligation to respect normal standards of scientific etiquette and “fair use” within the broad scientific community in using the pre-publication data, and to recognize and protect the contribution of sequence data producers.

To help promote these concepts, the MPIWG articulated guiding principles for the release and use of pre-publication microbial genome sequence data; these principles were approved by the NSTC Committee on Science in Spring 2004. The guiding principles have been distributed to the agencies for implementation as appropriate for each and are already in use. For example, DOE now requires all microbial sequencing projects at the Joint Genome Institute to follow these guidelines (<http://www.jgi.doe.gov/sequencing/collaborators/databarelease.html>). Similarly, NIH/NIAID includes the guidelines in all contracts for the new Microbial Sequencing Centers (<http://www.niaid.nih.gov/dmid/genomes/mscs/databarelease.htm>).



Since its opening in 1997, the throughput of the DOE Joint Genome Institute sequencing center has increased by ten-fold while the cost has decreased by almost six-fold. This mirrors trends at other sequencing centers and reflects the impacts of both new technologies as well as the production line ("factory") organization of the steps involved. Graph supplied courtesy of the Department of Energy's Joint Genome Institute.

Microbe Project Interagency Web Site

In 2001, the MPIWG issued a report entitled: "The Microbe Project" which outlined coordinated Federal efforts in microbial genomics focused on three goals: to build needed infrastructure, to promote research, and to develop human resources and an informed public. One of the recommendations of this report was to create an Interagency Microbe Project Web Site, to provide information about individual and interagency programs, program announcements, and requests for applications could be listed. Shortly after the report was published, the anthrax attacks in the U.S. highlighted the need for public awareness of the Federal government's support of research in the

microbial sciences, and the role of the MPIWG in facilitating interagency coordination and collaboration. The MPIWG web site (<http://www.microbeproject.gov>) thus provides information about the Microbe Project and its member agencies, and serves as a portal to the Federal government's interests and investment in genome-enabled microbial science.

HOME RESEARCH EDUCATION SEQUENCING

The MICROBE PROJECT

A Portal to U.S. Federal Efforts in Microbial Research

Member Agencies

- Department of Agriculture (USDA)
- Department of Defense (DoD)
- Department of Energy (DOE)
- Department of Homeland Security (DHS)
- Department of Interior (DOI)
- Environmental Protection Agency (EPA)
- Food and Drug Administration (FDA)
- National Aeronautics and Space Administration (NASA)
- National Institutes of Health (NIH)
- National Institute of Standards and Technology (NIST)
- National Oceanic and Atmospheric Administration (NOAA)
- National Science Foundation (NSF)

Microbial Sciences Tools and Resources

- The Institute for Genomic Research, Microbial Database
- American Society for Microbiology
- DOE Joint Genome Institute (JGI)

Reports

- The Microbe Project Report, May 2001
- Interagency Report on the Federal Investment in Microbial Genomics, June 2000

The Microbe Project

"to maximize the opportunities offered by genome-enabled microbial science to benefit science and society, through coordinated interagency efforts to promote research, infrastructure development, education and outreach"

The microbial world is vast and diverse. The Earth hosts over 10^{30} microorganisms, representing the largest component of the planet's biomass. Microbes include bacteria, archaea, mollicutes, fungi, microalgae, viruses and protozoa---organisms with a wide range of morphologies and lifestyles. Microorganisms have colonized virtually every environment on Earth ranging from deep sea thermal vents, polar sea ice, desert rocks, guts of termites, roots of plants, to the human body. While microbes are often feared for the diseases they may cause, other microorganisms mediate the essential biogeochemical cycles of key elements that make our planet habitable. Ancient lineages of microorganisms may hold the key to understanding the earliest history of life on Earth.

More >>

In The Spotlight

-  **Marine Microorganism Suspected to Play Role in Global Carbon and Nitrogen Cycles**
Scientists successfully grow "dwarf belonging to the sea" in laboratory
Updated September 23, 2005
-  **Genome Streamlining in a Cosmopolitan Oceanic Bacterium**

II. Interagency Coordination Plan (2008-2018)

Although microbiology is evolving rapidly, there remain a number of scientific and technical challenges that must be overcome as we move into the post-genomics world of proteomics, metabolomics and eventually personalized medicine. This 2008 Coordination Plan seeks to identify those challenges, as well as identify opportunities for interagency collaboration towards finding solutions and promoting the continued growth and development of this exciting field.

The MPIWG developed this coordination plan to guide the activities and priorities for the next decade. The MPIWG identified two broad areas for planning: 1) infrastructure and education; and 2) research. *Infrastructure and education* includes programs that provide the foundation for genome-enabled microbial science across all agencies. Four areas of Infrastructure and Education were identified that were important for the activities of all the agencies: Microbial Genome Sequencing, Databases and Bioinformatics Infrastructure, Experimental Tools, Techniques and Biological Resources, and Education and Training. *Research* areas include genome-enabled microbial science of interest to multiple agencies, and which would benefit from enhanced interagency coordination. Six Research areas were identified, which include Functional Genomics; Proteomics; Comparative Genomics; Microbe-Host and Microbe-Microbe Interactions; Environmental Genomics (incorporating metagenomics); and Microbial Forensics.

A. *Infrastructure and Education in Genome-Enabled Microbial Science*

Microbial Genome Sequencing

Critical to advancing the field of microbial science and facilitating the most efficient use of Federal dollars is the role that the MPIWG plays in sharing of information about planned and ongoing microbial genome sequencing projects. As noted in the Accomplishments section, in 2002 the MPIWG produced a complete inventory of all planned and ongoing microbial genome sequencing projects to support the federal government's biodefense strategic planning. The MPIWG agencies plan to

continue to partner through joint programs to facilitate the sequencing of microbial genomes. In addition, all of the agencies in the MPIWG will continue to routinely share plans for upcoming sequencing projects, to avoid unnecessary duplication of effort and to identify new opportunities for collaboration. The MPIWG is the primary venue for coordinating agency planning in this area, and will continue to play this role in the future.

Databases and Bioinformatics Infrastructure

Information infrastructure as well as bioinformatics tools and applications are critical to all areas of genome-enabled science. Interagency coordination will leverage investments in databases and bioinformatics. Bioinformatics and computational infrastructure capabilities for microbial sciences currently support research scientists; in the future, genomic data will need to be accessible to an even broader group of users including clinicians, public health professionals, forensic scientists, veterinarians, farmers, environmental managers, and users in the biotech and private sectors.

Currently, genome-enabled microbial science has critical infrastructure needs. These include hardware, database architectures, as well as and software tools to query, maintain and update terabytes of data. In addition, infrastructure will be needed to support the storage, analysis and integration of genomics, proteomics, metabolomics, system biology, and imaging data. Computational infrastructure is needed to support the determination of protein conformation, drug design and discovery, and *in silico* experiments that are time efficient and cost effective. While a number of Federal agencies contribute to infrastructure, additional coordination and collaboration is needed to provide a stronger national infrastructure for microbial science.

Experimental tools, techniques and biological resources

Development of tools and techniques, as well as maintenance of collections of biological research resources, is critical to the success of genome-enabled science. Improvement of genome sequencing and experimentation continues to be

critically important for the rapid advancement of knowledge. For example, the number of microbial genome sequencing projects has increased considerably as a result of dramatic reductions in sequencing costs. In addition, assaying gene expression across an entire genome was a relatively early technological development that followed the completion of the first few microbial genome sequences. These analyses have become standard techniques for genome-enabled research and have already contributed tremendous amounts of new information. Increasingly sophisticated technologies offer the opportunity to rapidly analyze complex patterns of gene expression and their organization into regulatory networks.

Genome-enabled microbial research also depends upon the development and availability of a variety of biological research resources, such as culture collections of living cells, plasmids, viruses, clone libraries, etc. Culture collections provide a valuable source of biological material for experimentation by microbiologists. However, new methods are needed for culturing environmental isolates of some of the currently unculturable organisms to enable more detailed and distributed studies of their biology. Microbes that require highly specific conditions to grow and/or live at extreme temperatures (“extremophiles”) are often difficult to maintain under standard culture regimens. Microbial communities that are representative of a range of environments need to be preserved as a resource for scientists to study in the future, while accommodating biosecurity issues. It is essential

that coordination of microbial culture collections occurs internationally and that the cultures be available and accessible to scientists world-wide.

Education and Training

There are many options open to individuals seeking research training opportunities in most areas of biological science. Research grants to academic institutions include, as a rule, funds that support education and training for graduate students and postdoctoral associates, and most formal training, including microbial genomics, occurs within this sphere. Since many of the relevant Federal agencies that support biological research issue grants to universities and colleges, an individual seeking training has, in principle, a variety of sources to support this training. Microbial genomics, which is widely supported across the agencies participating in the MPIWG, exemplifies the training opportunities inherent in the support of research grants. There are also targeted training programs and courses at the individual and institutional levels, to which potential trainees may apply, although there are few training programs exclusively targeted to microbial genomics *per se*. Directly relevant, conceptual and technical training is often a part of general training programs in genomics science. These programs would benefit from international collaboration, particularly with developing countries. The Appendix includes a description of training programs that can provide the basis for professional development in microbial genomics.

Goals for Infrastructure and Education

1. Continue to support the **genome sequencing** of a diverse set of microbial genomes which enhance the knowledge base for examining the functional analysis of genes; promote the development and application of novel sequencing and computational techniques to sequence and analyze rapidly entire microbial genomes.
2. Coordinate the development of tools and applications for the management, collection, storage, retrieval, and analysis of large data sets generated by genomics, metagenomics, proteomics and metabolomics studies. Promote the integration of “-omics” data with clinical, epidemiological and environmental data. Support standards of interoperability, data validation, quality assurance. Promote **accessibility** of published data relevant to the microbial sciences (where the data types are not only genome sequence data).
3. Encourage **access to biological resources** such as specialized cells and cell lines, strains, BAC libraries, etc. The MPIWG is aware that access to biological resources must be viewed and managed in the context of biosecurity concerns. Promote and foster the development of **novel culture techniques** for microbes from a range of environments including extremophiles and other fastidious microbes.
4. Promote **education and training** of students, scientists and the public on genome-enabled microbial science, and promote opportunities that help prepare the next generation of researchers to work in interdisciplinary teams.

B. Research in Genome-Enabled Microbial Sciences

Assigning Functions to Genes

Once a microbial genome is sequenced, the genes and the function of the gene products need to be identified as a first step in understanding the biology of the microbe. Thus the first priority area addresses opportunities to use microbial genomics, microbial and human DNA sequences, and high throughput or large-scale genomic technologies to determine the functions of genes in whole microbial genomes and proteomes. Decoding microbial genome information can facilitate new insights into pathogenicity, host-microbe interactions (including susceptibility to infection and response to treatment such as therapeutics or vaccines) and microbial diversity. Such scientific advances hold promise for the development of new drug targets, diagnostics, and vaccines, tools for

phylogenetic analysis, new industrial catalysts, methods for processing food and ensuring food safety, strategies for environmental management and remediation and potentially, even the production of bioenergy and the reduction of greenhouse gases.

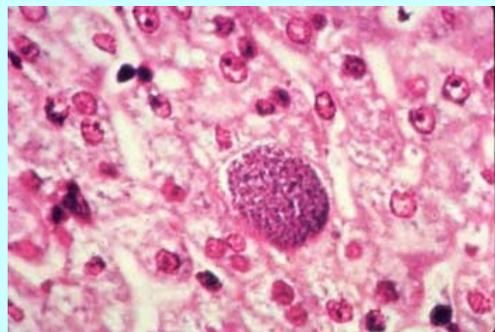
The application of emerging, high throughput, genomic technologies that will identify the complete DNA sequence of a microbe, determine the function and structure of each microbial protein, and identify metabolic pathways will facilitate achieving the mission and goals of the MPIWG member agencies. Therefore, the development of tools and technologies will be important to agencies that share an interest in genome-enabled microbial science, whether their mission addresses biodefense, environmental management, climate change, agriculture or increasing our fundamental knowledge of microbial processes.

Genomics Provides Promising New Targets for Antimalarial Drugs

Approximately 40% of the world's population is at risk of malaria. The disease is responsible for 300 million acute illnesses and at least one million deaths annually. In Africa, approximately 90% of the deaths are young children; an African child dies every 30 seconds from malaria. Although potent drugs and aggressive mosquito control once appeared to have malaria under control, widening drug resistance and ineffectiveness of insecticides have been among leading reasons why malaria today is considered one of the major public health challenges in the poorest countries of the world.

This life threatening human disease is caused by parasites, including *Plasmodium falciparum* and three other closely related species, and transmitted from person to person by the mosquito, *Anopheles gambiae*. In the mid-1990s, an international consortium of scientists and funding agencies was organized to determine the genomic sequence of *Plasmodium falciparum*. In 2002, the completed sequence was published along with the genome of the mosquito, *Anopheles gambiae*, the result of another international consortium.

Publicly available genome sequence data have provided scientists all over the world with the genetic blueprint for this deadly disease and identified promising new targets for antimalarial drugs. Very rapidly, scientists scanned the parasitic genome and identified biochemical pathways in the parasite that were potential locations where drugs might inhibit the life cycle of this deadly parasite. A drug called fosmidomycin and other closely related compounds were found to be effective in inhibiting one of these newly discovered parasitic biochemical pathways needed for making steroids. Evaluation of these drugs for effectiveness and safety in treatment of malaria is underway.



Legend: *Plasmodium falciparum*: Pre-erythrocytic liver stage. (Credit: WHO/MAP/TDR)

Microbial Proteomics

Microbial genomes encode thousands of proteins, yet we know the functions of a relatively small number of these. Microbial proteomics uses microbial genomics to elucidate and characterize the composition, structures, inventory, dynamics, regulation and functions of both individual proteins and protein complexes in microbial cells. Typically, many (if not most) of these proteins that are critical to cellular function work as part of protein complexes or "molecular machines," dynamic in time and subcellular localization. Understanding proteins involved in metabolism, energy utilization, growth and division, environmental adaptation, pathogenesis, and biosynthesis are all high priorities to be pursued by individual agency programs in accord with their missions.

Astonishingly, while more and more microbes are being sequenced, (2494 bacterial and archaeal projects were listed in GenomesOnLine (<http://www.genomesonline.org>) as of February 2008) the number of potential genes of unknown function is *not* decreasing. This observation suggests that current approaches to defining and describing ("annotating") microbial proteins are far from satisfactory and/or that the repertoire of unknown (and hitherto unsuspected) proteins in the microbial world is even greater than expected. To exploit the increased microbial genomic sequence data, we need to elucidate the functions of the products of open reading frames identified by high throughput sequencing projects.

One third or more of the predicted proteins in a microbial cell remain obscure as to their function, and because a microbial cell contains many dynamic multi-protein complexes or "molecular machines", the challenge only increases when multi-protein complexes are considered. With its 51 distinct polypeptides and three rRNAs, the bacterial ribosome may be considered the prototype of a molecular machine. While the role of the ribosome in the synthesis of polypeptides has been well studied and characterized, questions remain about ribosome processing, regulation, and functioning. Thus, characterization of hundreds of *other* microbial machines (a detailed inventory does not yet exist) will be challenging. A coordinated, interagency effort in microbial proteomics would improve our understanding of the repertoire of

bacterial molecular machines, including their composition, inventory, assembly, dynamics, regulation, function, and responses to environmental challenges. Target microbes might include those involved in pathogenesis, environmental remediation, carbon processing in the terrestrial and ocean environments, biotechnology and mission areas of federal agencies.

Microbial "Extremophiles" Survive pH <1 at Iron Mountain

Iron Mountain Mine in Northern California was once



Legend: A biofilm comprising a small number (~5) of discrete species of bacteria collaborating to survive in the harsh (pH <0.5) environment of Richmond Mine, Iron Mountain, CA. (Credit: Rudy Carver, Iron Mountain Operations)

the largest source of toxic metals in the United States, a location where a century of mining for gold and other metals unearthed rocks that, when exposed to the air, released acid and heavy metals into the water underground. Miners who accidentally left their shovels at the site overnight were likely to return the next day to find half a shovel eaten away by the corrosive water. At

points inside the mine, microbes are living, hundreds of feet underground where sunlight is absent and nutrients are scarce. These microbes live in water that is 120° F with high levels of arsenic, zinc and copper. In places, the pH values are less than 1.0 (more caustic than battery acid) yet microbes live there. From DNA sequencing data, the microbes were determined to be a mixture of bacteria and Archaea, an ancient domain of life that is separate from animals, plants, and bacteria. A team of scientists is now extracting and analyzing DNA and the resulting proteins from the entire set of organisms living in this extraordinary environment and characterizing the microbial community to understand its abilities to survive such extreme conditions.

The clear imperative is that improved approaches and technologies are needed for all phases of protein and protein complex identification, characterization, and measurement. While genomics can determine the potential “parts list” for a microbe, it is the proteins that form the structures that enable cell functions.

Comparative Genomics

Comparative genomics is the analysis and comparison of genomes among and within species. In the diverse world of microbes, comparative genomics offers the possibilities of gaining a better understanding of how species have evolved, determining the level of genetic diversity within species or sub-species groupings, examining commensal and pathogen relationships, and helping to determine the function of genic and non-genic regions of the genome. Comparative genomics perhaps offers the best opportunity to expose sites of horizontal gene transfer and to provide details on the process of recombination in the evolution of

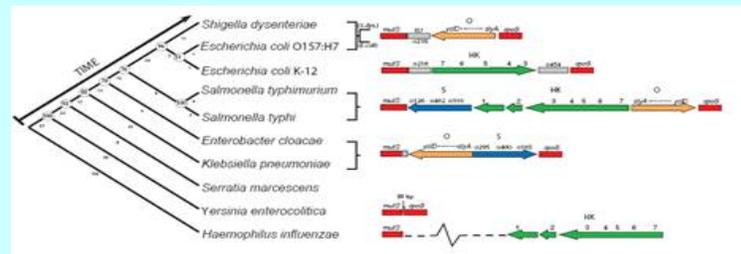
microbial genomes.

Different questions in microbial genomics can be addressed by comparing genomes at different phylogenetic (evolutionary) distances. At short distances, very similar genomes could reveal key sequence differences that may account for the phenotypic differences observed among organisms at the sub-species levels, which represent genetic sequence changes under positive selection. At moderate phylogenetic distances, both functional and nonfunctional sequences are found within the conserved DNA, the functional sequences having changed less than the nonfunctional or neutral DNA. Broad insights about types of genes and gene functions might be achieved by genomic comparisons at very long phylogenetic distances, where the order of genes and the sequences regulating their expression are generally not conserved. Comparative genomics is thus a powerful approach that will continue to be more informative as microbial sequence data and multiple comparisons accumulate.

What Makes A Microbe Pathogenic?

As the wealth of information from bacterial genomics unfolds, comparison of genomes exposes the truly mosaic nature of the microbial genome. Striking examples are the comparative analyses of the genome of the food-borne pathogen *Escherichia coli* O157:H7 with the sequence of a non-pathogenic laboratory strain of *E. coli* K-12. *E. coli* is a versatile enteric and usually a harmless member of the intestinal flora of higher vertebrates. *E. coli* O157:H7 is indeed harmless in cattle and other ruminants, but in humans is a worldwide health concern as a contaminant of beef and cause of hemorrhagic colitis. Comparative genomics revealed that the O157:H7 genome is 1 Mb larger than that of K-12, largely from 177 segments of DNA greater than 50 base pairs present in O157:H7 but not in K-12. Further analysis is revealing that all O-islands may not be O157-specific, but rather unique to the strains that were sequenced. Recognition of this amazing diversity in a bacterium once thought to be ‘clonally structured’ emphasizes a central role for recombination in shaping the bacterial chromosome. It causes us to rethink the approaches we may need to evaluate the true genomic diversity of microbial species. Of course, there is concern that information about what makes a particular organism pathogenic could be used for nefarious purposes by those with bioterrorist intent.

The National Science Advisory Board for Biosecurity (NSABB) has been established to provide advice to federal departments and agencies on ways to minimize the possibility that knowledge and technologies emanating from vitally important biological research such as this will be misused to threaten public health or national security (<http://www.biosecurityboard.gov/>).



Legend: Comparative genomic analysis of the mutS-rpoS gene region in enteric bacteria suggests that the region has expanded and contracted through the evolution of species, highlighting how horizontal gene transfers have forged a mosaic structure for the bacterial chromosome. (Credit: Kotewicz et. al., Trends Microbiol. 11:2-6 (2003), Elsevier)

RNA Interference



Legend: Infection of this white-leaved plant with a plant virus modified to produce RNA from a plant pigment gene triggered RNA interference, which inhibited pigment production. (Credit: Larry Grill, Large Scale Biology Corp)

through an RNA-mediated surveillance system that is directed against invading viruses, limiting viral accumulation and movement. This surveillance system, known as RNA-interference or RNAi, detects and breaks down invading RNA, preventing viruses from taking control of normal cellular processes. Some viruses have developed ways of evading the surveillance system. Many of the components of the molecular machinery involved in the process of RNA interference still need to be identified and characterized. However, as scientists learn more about RNAi, they are discovering that its role in the physiology of a wide range of hosts, including plants, worms, flies and humans, goes well beyond defense against viral invasion.

RNAi plays a critical role in normal host development and regulation of host gene expression. Research on RNAi and the ways that viruses evade detection is rapidly producing new insights and potential applications. For example, scientists are exploring the use of RNAi as a tool for genome-wide studies of plants and animals, for large-scale genetic screens, for drug discovery and for disease therapy.

How do eukaryotic organisms (plants and animals) defend themselves against viral infection? The importance of animal antibodies and the immune system in viral defense is well known. The role of ribonucleic acid (RNA) in eukaryotic organisms had traditionally been thought to be that of a passive intermediary in the process of protein production from genetic templates. By studying how plants and animals defend themselves from invading viruses, scientists have discovered that RNA plays a much more versatile and active role in living cells than was previously thought.

One means by which plant and animal hosts defend themselves is

Microbe-Host Interactions and Microbe-Microbe Interactions

This priority area addresses opportunities to use microbial genomics, proteomics and metabolomics to elucidate mechanisms of microbe-host and microbe-microbe interactions. Examples of mechanisms to be explored include those involved in 1) microbial pathogenesis and host disease-resistance, 2) microbe-host symbiosis, parasitism or commensalism, 3) communication mechanisms used by microbes involving release of chemical signals that lead to changes in a recipient organism's (microbial or non-microbial) gene expression (e.g., quorum sensing), and 4) microbe-microbe and microbe-host horizontal gene transfer. The microbes may be bacteria, archaea, mollicutes, fungi, viruses and protozoa; hosts may include animals, plants and other microbes.

Common mechanisms for secretion pathways, bacterial regulatory proteins, toxins and other virulence factors need to be compared across diverse microorganisms (e.g. pathogens of humans, animals and plants) as well as significant differences in host range and the gene sequences responsible for those differences. We must also develop a better understanding of how microbe-microbe and microbe-host signaling result in changes in gene expression. Challenges to increased understanding in these areas include the mechanistic complexity of these processes and the need for new tools and technologies to deal with the complexity, particularly in a high-throughput manner. The large number of microorganisms that can not currently be cultured increases the challenges associated with understanding of microbe-microbe and microbe-host interactions.

Microbial Forensics and Molecular Epidemiology

Microbial forensics is a new scientific discipline that seeks to analyze and interpret microbial genetic data to determine their potential links to a criminal or terrorist event. Microbial forensics is similar in many respects to the field of molecular epidemiology. Both strive to identify and recognize patterns in outbreaks of agronomic, environmental, and medical concern, communicate those patterns to scientists and the general public, determine the microbe(s) involved, contain the

outbreak and trace the microbe to its source. The assumption, at onset, that a biocrime has been committed (and pursuit of the perpetrator), however, distinguishes the emerging field of microbial forensics from a molecular epidemiology investigation.

Whereas epidemiology seeks a causal link between an outbreak of disease and a microbial source, a forensic investigation analyzes and interprets physical evidence for purposes of attribution, that is, the assignment to a source of origin, usually with the intent of criminal prosecution. As such, the criteria for identification of microbes for attribution purposes, likely subject to litigation, are much stricter and require a high degree of certainty and specificity. The quality of data needed to establish attribution, in the eyes of the court, reflect the

resources employed and the kinds of methodologies utilized. Forensic tools, both genomic and non-DNA based, are needed to establish attribution of a suspect strain. The tools employed will likely differ for the pathogen encountered and the range of diversity within its pathogenic group. In developing a suite of molecular technologies for discriminating strains of microbial pathogens, the issues of resolving power, validation, quality control, and statistical applications must therefore be treated with each application. It is recognized that these issues are also considered by interagency groups such as the FBI's Scientific Working Group on Microbial Genetics and Forensics; the role of the MPIWG is to identify research needs and to collaborate to address those needs.

Antibiotic-Resistant Strains: Implications for Public Health, Forensics and Attribution

Current interest in the growing challenge posed by microbial resistance to available antibiotics has focused on the public health implications. This has fostered increased research targeted on the development of the next generation of antimicrobials to meet the growing challenges of pathogen resistance to current antibiotics. However, this research may not be adequate to address the needs associated with analysis for purposes of forensics, in which the genetic basis for pathogen resistance to antibiotics must be identified. Specifically, forensics must have a robust set of tools, providing methods that can characterize, identify, and potentially attribute the organism to a specific perpetrator. Tools essential for dealing with antibiotic resistant strains will need to distinguish naturally occurring genetic drift from intentional surreptitious modification. An aggressive research program on the identification of genetic modification is essential in development of analytical tools to combat potential use of genetic manipulation for next generation biological threat agents.

Environmental Genomics

Environmental Genomics is made possible by Metagenomics.

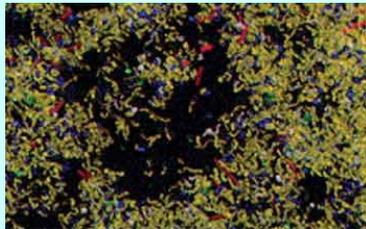
Metagenomics is the analysis of genome sequence data extracted directly from environmental samples. The growing availability of genome sequences of known microbes has enabled the development of environmental genomics. The application of metagenomics to the study of microorganisms in their environments provides unique opportunities to address the long-standing questions about microbes: “What organisms are out there, and what are they doing?” In the context of concerns about emerging infectious diseases and biodefense, it is important to know what microbes are found in “normal” environments, and how human intervention, either inadvertently or by design, affects the composition and function of microbial communities. A wide range of environments is of interest, from the surfaces of plant roots and animal digestive tracts to hot springs and polar ice, from agricultural lands and forests, to waste dumps and coastal waters and open oceans. Several agencies are beginning to support research to address these problems in the context of their

missions; however, a vast range of environments still need to be explored, and many environments are relevant to the mission of more than one agency. Thus, coordination and collaboration in the nascent field of environmental genomics approached by metagenomics is very important.

Metagenomics is beginning to reveal astounding new information relevant to the missions of a number of agencies. However, sequence information alone does not reveal how microbes cooperate or compete with each other, how new metabolic capabilities may be acquired *via* horizontal gene transfer processes, or how community composition and function change with time. New methods are being developed and applied to link information about genetics and genomics with metabolic activities of microbes in environmental samples.

For more information see *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet*, National Academies Press, 2007.

Chemical Communication among Bacteria and Biofilms



Legend: The picture shows a biofilm in which different specific fluorescent dyes are used to label different members of the microbial community. (Credit: Gene Tyson, UC Berkeley)

Bacteria have been found to produce and release, and to detect and respond to signaling molecules that accumulate in their environment as the cell population increases. Unlike humans, bacteria have been found to produce separate chemical signals (“odors”) to communicate with members of their own species and to communicate with other species. Genomic analysis has revealed that genes involved in such bacterial signaling responses are found in over thirty species of bacteria. Thus, communication by these types of signaling systems may be a common mechanism that all bacteria employ for interacting in natural environments.

Bacteria respond to chemical signals by turning on or off specific genes. The products of these genes may set off chains of events in which the bacteria synthesize and build complex structures in which they take up residence. These structures are called “biofilms.” A biofilm may be made and inhabited by one or more types of bacteria. When there is more than one type, each contributes to the structure and commerce of the complex biofilm community.

Biofilms are found on nearly every kind of moist surface imaginable, from the insides of water pipes, to stones in mountain streams, the bottoms of ocean-going ships, and human teeth. Although building and living in biofilms is advantageous for bacteria, biofilms are often undesirable from the human point of view. Biofilms are responsible for the plaque that encourages tooth decay and contribute to the friction of ships moving through the water, adding to the cost of fuel. However, despite their importance, we still know very little about the mechanisms by which bacteria produce biofilms and the complex processes that go on in biofilm communities. A better understanding of the development and function of biofilms will provide us with tools to modify them for human benefit and to reduce their economic costs.

Goals for Research

1. Functional Genomics. Promote the **identification and determination of function** of the complete set of expressed genes and regulatory elements in diverse microbes including those involved in disease, environmental remediation, carbon processing and biotechnology. Promote the development and use of **low cost, high throughput methods for assigning function**, particularly of hypothetical genes.
2. Proteomics. Support the development of novel experimental and computational techniques to **detect, quantify, and amplify protein complexes** with high sensitivity/specificity, including post translational modifications; promote the development and use of high throughput methods for producing proteins and native protein complexes with retention of structural and compositional integrity. Encourage research on **metabolic networks and pathways**, 3D imaging of metabolic processes in a cell, and development of proteomic resources and reagents.
3. Comparative Genomics. Encourage the scientific community to establish a framework for the number of genomic sequences needed to represent a specified level of genomic organization, and **criteria for selection of strains for total genome sequencing** within strain groupings. Promote the evaluation of technologies for rapid assessment of genomic diversity to guide the choice of strains for whole genome sequencing.
4. Microbe-Host and Microbe-Microbe Interactions. Promote the identification of common genes and signal molecules used for communication by diverse bacteria, microbial genome sequences involved in pathogenesis, and host factors involved in response to microbial pathogens, symbionts and/or commensals for model hosts. Promote the development of tools and techniques needed for **large-scale metabolomics studies of microbes and hosts**.
5. Microbial Forensics. Promote development of **techniques and criteria for microbial strain attribution** to distinguish natural from introduced pathogens. Support the continued development of models of outbreaks that allow comparison of microbial forensics and molecular epidemiology paradigms.
6. Environmental Genomics. Encourage the collection, coordination, and analyses of **metagenomics data from a range of environments** and support the development of informatics tools to integrate genomic data with other environmental data.

III. MPIWG Next Steps and Implementation Plan

The microbial world is incredibly vast and directly and indirectly impacts global health (plant, animal, human), the economy and national security. These areas are relevant to the missions of numerous agencies and departments of the Federal government. The development of new tools and technologies have led to rapid advances in microbial genome sequencing, creating increased opportunities and incentives for Federal agencies and departments to leverage resources and expertise. To this end, the Microbe Project will continue to serve as a forum to foster coordination and collaboration among the agencies in implementing the goals that have been identified in this report.

Several times in this report, particularly for research needs, we call for the development of new tools and techniques. This can be (and is being) steadily accomplished by the research programs at the MPIWG member agencies but should be accelerated. Clear steps that the MPIWG can take to encourage this are:

1. Conduct cross-agency briefings on current agency investments in infrastructure/education and research to coordinate and leverage investments.
2. Sponsor or otherwise encourage workshops to identify gaps, opportunities and next steps important for advancing infrastructure/education and research.
3. Continue to coordinate among Federal agencies to promote education and training of students, scientists and the public on genome-enabled microbial science.
4. Maintain, update and encourage a wide distribution of the Microbe Project web site.

Understanding the microbial world is of vital importance to the Nation to ensure the health and well-being of all its citizens. The MPIWG will continue to coordinate microbial genomic activities among federal agencies and look for opportunities to coordinate with international agencies. Mechanisms may include joint programs, joint workshops, joint planning, memoranda of understanding, as well as a close coordination of relevant programs. The MPIWG recognizes the importance of linkage and integration of microbial genomics, transcriptomics, proteomics, metabolomics, and other “omics” to promote an integrated systems biology approach to the study of microbes.

Finally, the MPIWG remains strongly committed to public accessibility of genomic and genome-related data, while maintaining an awareness of biosecurity issues. The deposit of genomic and proteomic data into publicly accessible, databases for use by the broad scientific community will be strongly encouraged. Public access to data is critical to understanding the diversity and evolution of the microbial world; the microbial species involved infectious diseases and environmental processes; and the biotechnological potential of microorganisms. In addition, the MPIWG will continue to serve as a resource for the U.S. Government on national biosecurity issues.

The roles of the participating agencies in coordinated research and infrastructure areas are summarized in Tables 1 and 2.

Table 1. Roles of MPIWG Agencies in Infrastructure and Education

(c=contributor of genome-enabled microbial science and engineering infrastructure and education; u=user of microbial science and engineering infrastructure and education)						
Agency Mission relevance*	1. Microbial Genome Sequencing	2. Metagenomics	3. Databases and Bioinformatics Infrastructure	4. Experimental Tools and Techniques	5. Culture Collections and Biological Resources	6. Education and training
DOC – NIST						
DOC – NOAA						
DOD	u/c	u	u/c	u/c	u/c	u/c
DOE	u/c	u/c	u/c	u/c	u/c	u
DHHS – CDC	u/c	u	u/c	u/c	u/c	u/c
DHHS – FDA	u/c	u	u/c	u/c	u	u/c
DHHS – NIH	u/c	u/c	u/c	u/c	u/c	u/c
DHS	u/c	u	u/c	u/c	u/c	u/c
DOJ – FBI	u/c	u	u/c	u/c	u/c	u/c
EPA	u	u	u/c	u/c	u/c	u/c
IC**	u/c	u	u/c	u	u	u
NASA	u	u/c	u/c	u/c	c/u	u
NSF	c	c	c	c	c	u/c
USDA – ARS	c/u	c/u	c/u	c/u	c/u	c/u
USDA – CSREES	c/u	c/u	c/u	c/u	c/u	c/u

Contributors have resources, facilities and technical programs that generate capabilities or knowledge that can be harnessed and brought to bear to advance microbial science and engineering, even if such concerns are generally peripheral to their mission.

Users will derive benefit from research, technology and engineering that originates in or with support from their own or other agencies to apply to their agency’s mission in genome-enabled microbial science and engineering, even if they do not actively perform such research.

*Each Agency’s mission related to genome-enabled microbial science and engineering is summarized on The Microbe Project web site. <http://www.microbeproject.gov>.

**IC stands for the agencies represented in the Intelligence Community.

Table 2. Roles of MPIWG Agencies in Coordinated Research Areas

(c=contributor of microbial science and engineering knowledge; u=user of microbial science and engineering knowledge)								
Agency Mission relevance*	1. Functional Genomics	2. Proteomics	3. Comparative Genomics	4. Microbe-Host and Microbe-Microbe Interactions	5. Microbial Forensics	6. Environmental Genomics	7. The Human Microbiome	8. Metagenomics
DOC – NIST								
DOC – NOAA								
DOD	u/c	u/c	u/c	u/c	u	u/c	u	
DOE	u/c	u/c	u/c	u/c	u/c	u/c	u/c	u/c
DHHS – CDC						u	u	u
DHHS – FDA	u/c	u/c	u/c	u/c	u/c	u	u	u
DHHS – NIH	u/c	u/c	u/c	u/c	u/c	u/c	u/c	u/c
DHS	u/c	u/c	u/c	u/c	u/c	u/c	u	
DOJ – FBI							u	
EPA	u/c	u/c	u/c	u/c	u/c	u/c	u	u
IC							u	u
NASA						u/c	u	u
NSF	u/c	u/c	u/c	u/c	u/c	u/c	u/c	u/c
USDA – ARS	u/c	u/c	u/c	u/c	u/c			u
USDA – CSREES	u/c	u/c	u/c	u/c	u/c	u/c	u/c	u/c

Contributors have resources, facilities and technical programs that generate capabilities or knowledge that can be harnessed and brought to bear to advance microbial science and engineering, even if such concerns are generally peripheral to their mission.

Users will derive benefit from research, technology and engineering that originates in their own or other agencies to apply to their agency’s mission in genome-enabled microbial science and engineering, even if they do not actively perform such research.

*Each Agency’s mission related to genome-enabled microbial science and engineering is summarized on The Microbe Project web site: <http://www.microbeproject.gov>.

Glossary

Annotation:	The process of associating a likely protein product or function with a probable gene in a DNA sequence.
Archaea:	A major branch of single celled organisms. Like bacteria, cells of archaea lack an organized nucleus. Archaea were once thought to be bacteria, but have been found to be different in a number of fundamental ways.
BAC:	Bacterial artificial chromosome—a construct for propagating segments of DNA used in genome sequencing to prepare convenient-sized segments of DNA for sequencing.
Commensal:	A microbe living in a host, without harming the host.
Enteric:	A type of microbe that inhabits areas of the digestive tract of an animal.
Extremophile:	Microbes that tend to live in what seem to humans, extreme environments, such as heat, cold, acid, high salt
Genomics:	The study of the set of DNA-coded instructions (the genome) of organisms. Studies based on the use of information from the genome sequences of an organism
Horizontal gene transfer:	The process in which an organism transfers genetic material to another cell that is not its offspring
Metabolomics:	Study of the combined metabolic processes of an organism, generally based on deductions of function from the genome sequence
Metagenomics:	Study of the functions and capabilities of a microbial community by analysis of the DNA sequences extracted from the entire community obtained, for example, from an environmental sample.
Mollicute:	A type of microbe, generally small, with a small genome and generally pathogenic to plants or animals
Ortholog:	Genes from two different organisms that seem sufficiently similar in DNA sequence to suggest that they have the same ancestral origins.
Parasitism:	A relationship between different species of organisms in which one organism, the parasite, benefits and the other organism, the host, is harmed.
Phylogenetic:	A study of the family relationships among organisms
Proteomics:	Analyzing the entire complement of all proteins from a particular cell, tissue or organism, with or without experimental manipulations of the sample
Protozoa:	One-celled eukaryotes (cells with defined nuclei bounded by a membrane); typically larger and more complex than bacteria.
Quorum sensing:	The regulation of gene expression in response to fluctuations in cell-population density.
Symbiote:	A microbe (or other type of organism) that lives in a mutually beneficial relationship (symbiosis) with a host

References

Assigning Function

S. Borrmann *et al.* Fosmidomycin-Clindamycin for the treatment of *Plasmodium falciparum* malaria. The Journal of Infectious Diseases 190:1534-1540. (2004).

M. J. Gardner *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419: 498-511. (2002).

R.A. Holt *et al.* The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 297: 129-49 (2002).

Jomaa *et al.* Inhibitors of Nonmevalonate Pathway of Isoprenoid Biosynthesis as Anti-malarial drugs. Science 285: 1573-1576. (1999)

Proteomics

R.J., VerBerkmoes, N.C., Thelen, M. P., Tyson, G.W., Baker, B.J. Blake, R.C. II, Shah, M., Hettich, R.L. and Banfield, J.F. Community proteomics of a natural microbial biofilm, Science, 308:1915-1920. (2005)

Comparative Genomics:

S. D. Bentley and J. Parkhill. Comparative genomic structure of prokaryotes. Annu. Rev. Genet. 38: 771-791 (2004).

M. McClelland *et al.* Comparison of the *Escherichia coli* K-12 genome with sampled genomes of a *Klebsiella pneumoniae* and three *Salmonella enterica* serovars, Typhimurium, Typhi and Paratyphi. Nucleic Acids Res. 28: 4974-4986 (2000).

N. T. Perna *et al.* Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409: 529-533 (2001).

T. Hayashi *et al.* Complete genome sequence of entero-hemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain, K-12. DNA Res. 8: 11-22 (2001).

Microbe-Microbe and Microbe Host Interactions:

Eckburg, P. *et al.* Diversity of the human intestinal microbial flora. Science 308:1635-38.

Mathesius, U. *et al.* Extensive and specific responses of a eukaryote to bacterial quorum-sensing signals. Proc. Nat. Acad. Sci. 100: 1444-1449 (2003).

Teplitski, M. *et al.* *Chlamydomonas reinhardtii* secretes compounds that mimic bacterial signals and interfere with quorum sensing regulation in bacteria. Plant Physiology 134: 137-146 (2004).

Microbial Forensics

B. Budowle *et al.* Toward a system of microbial forensics: from sample collection to interpretation of evidence. Appl Environ. Microbiol. 71: 2209-2213 (2005).

T. A. Cebula *et al.* Molecular applications for identifying microbial pathogens in the post-9/11 era. Expert Rev. Mol. Diagn. 5: 431-445 (2005).

Environmental Genomics

Allen, E.E. and Banfield, J.F. Community genomics in microbial ecology and evolution. *Nature Reviews Microbiology*, 3(6): 489-498. (2005)

J Handelsman. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews* 68: 669-685. (2004)

General

Breitbart, M. and Rohwer, F. Here a virus, there a virus, everywhere the same virus. *Trends in Microbiology* 13:278-284. (2005)

Claverie, J.-M. Fewer Genes, More Noncoding RNA. *Science* 309: 1529-1530 (2005).

Kumagai, M. H. et al. Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA *Proc. Nat. Acad. Sci.* 92: 1679-1683. (1995).

Lindbo, J. and Dougherty, W.G. Plant Pathology and RNAi: A Brief History. *Annu. Rev. Phytopathology* 43: 191-204 (2005).

National Research Council. *The New Science of Genomics: Revealing the Secrets of our Microbial Planet.* The National Academies Press. (2007)

Pace, N.R. A molecular view of microbial diversity and the biosphere. *Science* 276:734-740. (1997)

Tyson, G.W. and Banfield, J.F. Cultivating the uncultivated: a community genomics perspective. *Trends in Microbiology* 13:411-415. (2005)

Woese, C. A new biology for a new century. *Microbiology and Molecular Biology Reviews* 68:173-186. (2004)

Zamore, P.D. and Haley, B. Ribo-genome: The big world of small RNAs. *Science* 309: 1519-1524 (2005)

Web Site

The Microbe Project Web Site is a portal to U.S. Federal Agency efforts in genome-enabled microbial sciences. <http://www.microbeproject.gov>

MPIWG Membership Roster

Daniel Drell (Chair)
Biological and Environmental Research
Department of Energy

Ann Lichens-Park (Executive Secretary)
Cooperative State Research, Education, and
Extension Service
U.S. Department of Agriculture

James Anderson
National Institute of General Medical Sciences
National Institutes of Health

Kay Marano Briggs
U.S. Geological Survey
Department of Interior

Michael Broder
U.S. Environmental Protection Agency

Thomas Cebula
Food and Drug Administration
Department of Health and Human Services

Linda Chrisey
Life Sciences Research Division
Office of Naval Research

Valentina Di Francesco
National Institute of Allergies and Infectious
Diseases
National Institutes of Health

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

Maria Giovanni
National Institute of Allergies and Infectious
Diseases
National Institutes of Health

J. Eugene LeClerc
Food and Drug Administration
Department of Health and Human Services

James Lindsay
Agricultural Research Service
U.S. Department of Agriculture

Pamela McInnes
National Institute of Dental and Craniofacial
Research
National Institutes of Health

Shawn McLaughlin
National Oceanic and Atmospheric Administration

Dale Nordenberg
Center for Disease Control and Prevention
Department of Health and Human Services

Frederick Pearce
Office of the Deputy Under Secretary of Defense
for Science & Technology
Department of Defense

Peter Pesenti
Science & Technology
Department of Homeland Security

Scott Steele
Office of Science & Technology Policy
Executive Office of the President

Gerard Stelma
U.S. Environmental Protection Agency

Joanne Tornow
Division of Molecular and Cellular Biosciences
National Science Foundation

Acknowledgements

The MPIWG recognizes and acknowledges the hard work and major contributions of Dr. Anna Palmisano, Chair in 2006, during the process of generating and producing this Coordination Plan.

The MPIWG also acknowledges the contributions and assistance of:

- Thomas A. Cebula, Food and Drug Administration, Department of Health and Human Services, MPIWG Chair in 2007
- Joanne Tornow, National Science Foundation, current member and Executive Secretary from 2000 to 2005
- Matthew Kane, National Science Foundation, Executive Secretary in 2007
- Bruce Budowle, Federal Bureau of Investigation, Department of Justice
- Maryanna Henkart, National Science Foundation
- Diane Jones, Office of Science and Technology Policy, Executive Office of the President
- Pushpa Kathir, United States Department of Agriculture, Cooperative State Research, Education and Extension Service
- Bart Kuhn, Office of the Deputy Under Secretary of Defense for Science and Technology, Department of Defense
- Dennis Mangan, National Institute of Dental and Craniofacial Research, National Institutes of Health
- Jane Silverthorne, Office of Science and Technology Policy, Executive Office of the President
- Ronald Weiner, National Science Foundation
- Kurt Zuelke, Office of Science and Technology Policy, Executive Office of the President

Appendix: Training Programs Relevant to Genome-Enabled Microbial Science

For regularly updated information, refer to The Microbe Project Web Site (www.microbeproject.gov).

Department of Defense (DoD)

1. The US Army supports a variety of training programs that can include those relevant to microbial genomics:
 - a. Summer Faculty Research and Engineering Program (<http://www.battelle.org/army/sfrep/default.htm>)
 - b. Science and Engineering Apprentice Program: (<http://wrair-www.army.mil/>)
 - c. Student Training in Advanced Research Skills: (<http://wrair-www.army.mil/>)
2. The Uniformed Services University for the Health Sciences offers graduate programs in genetics, microbiology, cell biology, immunology, molecular and cell biology that also can include training in aspects of microbial genomics (<http://www.usuhs.mil/>).
3. Office of Naval Research supported training is listed at the Microbe Project web site.

Department of Energy (DOE)

DOE has a number of training programs, the majority of which are administered by the Oak Ridge Associated Universities in Oak Ridge, Tennessee, including:

1. Minority Institutions Research Fellowship Program

For information on these programs, go to <http://www.ornl.gov>.

Department of Homeland Security (DHS)

Science and Technology Scholars Program: Undergraduate scholarships and graduate fellowships. (<http://www.ornl.gov/dhsed>).

Environmental Protection Agency (EPA)

EPA offers three fellowships which are intended to encourage promising students to obtain advanced degrees and pursue careers in environmentally related fields, including microbiology. For information about the Science to Achieve Results (STAR) Graduate Fellowships, Greater Research Opportunities (GRO) Graduate Fellowships, and the GRO Undergraduate Fellowships, go to <http://es.epa.gov/fellow/>.

National Institutes of Health (NIH), Department of Health and Human Services (HHS)

1. The National Institute of Allergy and Infectious Diseases (NIAID), through its Microbiology and Infectious Diseases Training Grants, supports training that includes genomic scale research on microbes of importance to disease; similarly, individual fellowships are awarded for training in related areas, including AIDS research. Further information can be found on the website, <http://www.niaid.nih.gov/ncn/training/default.htm> and by contacting Milton Hernandez, Ph.D. (mhernandez@niaid.nih.gov), Office of Special Populations and Research Training, NIAID.
2. The National Institute of Dental and Craniofacial Research (NIDCR) also provides training grants on genomic scale microbial research training. Further information can be found on the website, <http://www.nidcr.nih.gov/Funding/Training/default.htm>, and by contacting Kevin Hardwick, DDS, Ph.D. at Kevin.Hardwick@nih.gov. Training in bioinformatics and genomic database management can also be found at the NIDCR Oralgen project at the Los Alamos National Laboratories (<http://www.oralgen.lanl.gov>).
3. Several institutes, and the Library of Medicine (NLM), provide training in bioinformatics that is relevant to microbial genomics:
 - a. The National Institute of General Medical Sciences (NIGMS), through its training programs in bioinformatics and computational biology, provides training of particular relevance to microbial genomics. NIGMS also supports training programs in biotechnology, systems and integrative biology, and genetics. Individual training grants of these programs provide relevant training in genomic approaches to microbial research. Information on NIGMS programs can be found at the website,

- <http://www.nigms.nih.gov/funding/trngmech.html> and by contacting John Norvell, Ph.D. at norvellj@gml.nigms.nih.gov.
- b. The National Human Genome Research Institute (NHGRI) provides training grants and fellowships that emphasize bioinformatics relevant to all aspects of the Human Genome Project, including microbial genomics. Information on training at the NHGRI can be found on the website, http://www.nhgri.nih.gov/Grant_info/Funding/Training/ and by contacting Bettie Graham, Ph.D. at grahamb@odder.nhgri.nih.gov.
 - c. NLM offers both institutional training grants and fellowships in informatics training, relevant to microbial genomics. Training information at the NLM can be found at the website, <http://www.nlm.nih.gov/pubs/factsheets/extrapro.html> and by contacting Carol Bean, Ph.D., M.L.S. at beanc@mail.nlm.nih.gov.
4. The Fogarty International Center (FIC) supports training for scientists from low- and middle-income countries that is relevant to the microbe project through two of its international training programs:
 - a. Informatics Training for Global Health supports training in bioinformatics related to microbial and parasitic genomes. Further information can be found on the website <http://www.fic.nih.gov/programs/itgh.html> or by contacting Flora Katz, Ph.D. at katzf@mail.nih.gov.
 - b. The Global Infectious Disease Research Training Program supports research- related training on infectious diseases that are predominately endemic in or impact upon people living in developing countries, including training in bioinformatics, genomics, and genome-enabled science for diseases of microbial and parasitic origin. Further information can be found on the website, <http://www.fic.nih.gov/programs/infectiousdisease.html>, or by contacting Barbara Sina, Ph.D. at sinab@mail.nih.gov ."

National Science Foundation (NSF)

1. Postdoctoral Research Fellowships in Biology. This program seeks to encourage research and training at the postdoctoral level in selected areas of biology supported by the Biological Sciences Directorate, to encourage independence early in an individual's research career and to permit the pursuit of the individual's research and training goals in the most appropriate research locations regardless of the availability of funding for the Fellows at that site. In FY2009, the program supported fellowships to broaden participation in the biological sciences, and also supported fellowships for training in biological informatics. For information, go to: http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=12720&org=DBI .
2. Integrative Graduate Education and Research Traineeship (IGERT). This program was developed to meet the challenges of educating Ph.D. scientists and engineers with the multidisciplinary backgrounds and the technical, professional, and personal skills needed for the career demands of the future. Integrative Graduate Education and Research Traineeship (IGERT): For information, go to http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=12759&from=fund.
3. Research Experiences for Undergraduates (REU). The REU program, a Foundation-wide program for support of active research participation by undergraduate students, seeks to expand student participation in all kinds of research -- whether disciplinary, interdisciplinary, or educational in focus -- encompassing efforts by individual investigators, groups, centers, national facilities and others. For information, go to http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5517&from=fund
4. Undergraduate Research and Mentoring in Biology (URM). The intent of this activity is to provide support for talented students to gain research experiences in biological sciences and to foster an enriched and culturally diverse research and educational environment. For information, go to http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf06591

Agricultural Research Service, United States Department of Agriculture (USDA)(Intramural Funding)

ARS does not conduct any formal training programs, however, the Agency mentors students for masters &/or doctoral degrees, and provides additional training through its postdoctoral research awards program.

<http://www.afm.ars.usda.gov/divisions/hrd/hrdhomepage/vacancy/pd962.html> Internally, ARS provides its scientific and technical support staff “detail” opportunities for training at other research locations. ARS conducts extensive collaborative research with external (national and international) research entities. Ph.D level scientists are provided “visiting scientist” status which provides and often includes a training component.

Cooperative State Research, Education and Extension Service, United States Department of Agriculture (USDA)(Extramural funding)

1. Agriculture and Food Research Initiative (AFRI) Competitive Grants Program (<http://www.csrees.usda.gov/fo/fundview.cfm?fonum=1112>). This program is charged with funding research and Integrated (Research, education and Extension) activities on key problems of national and regional importance in biological, environmental, physical, and social sciences relevant to agriculture, food, and the environment.
2. Capacity Building Grants Program: 1890 Institution Teaching and Research Capacity Building Grants Program (<http://www.csrees.usda.gov/fo/fundview.cfm?fonum=1060>). The focus of the program is to strengthen teaching and research programs in the food and agricultural sciences by building the institutional capacities of the 1890 Land-Grant Institutions, Tuskegee University and West Virginia State University.
3. Food and Agricultural Sciences National Needs Graduate and Postgraduate Fellowship Grants Program (<http://www.csrees.usda.gov/fo/fundview.cfm?fonum=1280>). This program awards grants to train students for master's &/or doctoral degrees & provide additional postdoctoral training for Fellows who have completed their doctoral degrees at colleges and universities that have demonstrable teaching and research competencies in the food & agricultural sciences.
4. Higher Education Multicultural Scholars Program (<http://www.csrees.usda.gov/fo/fundview.cfm?fonum=1110>). This undergraduate scholarship grant program is designed to support the multicultural students in food and agricultural sciences.
5. Tribal Colleges Research Grants Program (<http://www.csrees.usda.gov/fo/fundview.cfm?fonum=1133>). This program supports Tribal Colleges to conduct research in the area of food and agricultural sciences.