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.

[1]

[1]