

March 2007

REPORT  
IN BRIEF

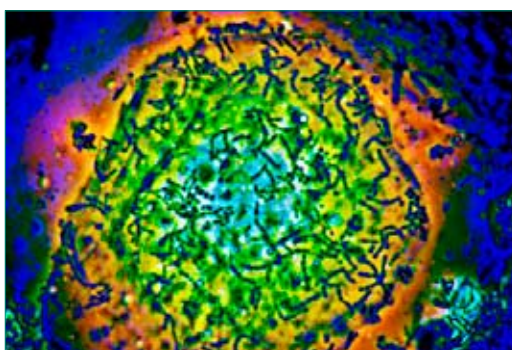
## The New Science of Metagenomics

### Revealing the Secrets of Our Microbial Planet

Although we can't see them, microbes are essential for every part of human life—indeed all life on Earth. The emerging field of metagenomics provides a new way of viewing the microbial world that will not only transform modern microbiology, but also may revolutionize understanding of the entire living world.

Every part of the biosphere is impacted by the seemingly endless ability of microorganisms to transform the world around them. It is microorganisms, or microbes, that convert the key elements of life—carbon, nitrogen, oxygen, and sulfur—into forms accessible to other living things. They also make necessary nutrients, minerals, and vitamins available to plants and animals. The billions of microbes living in the human gut help humans digest food, break down toxins, and fight off disease-causing pathogens. Microbes also clean up pollutants in the environment, such as oil and chemical spills. All of these activities are carried out not by individual microbes but by complex microbial communities—intricate, balanced, and integrated entities that have a remarkable ability to adapt swiftly to environmental change.

Historically, microbiology has focused on single species in pure laboratory culture, and thus understanding of microbial communities has lagged behind understanding of their individual members. Metagenomics is a new tool to study microbes in the complex communities where they live and to begin to understand how these communities work. Traditional microbiological approaches have already shown how useful microbes can be; the new approach of metagenomics will greatly extend scientists' ability to discover and benefit from microbial capabilities.



**Bacteria in human saliva.** Trillions of bacteria make up the normal microbial community found in and on the human body. The new science of metagenomics can help us understand the role of microbial communities in human health and the environment. *(Image courtesy of Michael Abbey)*

### What is Metagenomics?

The emerging field of metagenomics presents the greatest opportunity—perhaps since the invention of the microscope—to revolutionize understanding of the living world. Through metagenomics, scientists can apply the power of genomic analysis (the analysis of all of the DNA in an organism) to entire communities of microbes, bypassing the need to isolate and culture individual community members. In Greek, *meta* means “transcendent.” Metagenomics transcends the individual organism to the “meta level” of the community. Moreover, as both a research paradigm, comprising many inter-related approaches and methods, and an emerging research field, metagenomics transcends classical approaches to microbiology.

Metagenomics involves studying the genetic makeup of many microbes in an environment simultaneously, and makes accessible the many types of microbes that cannot be grown in the

**THE NATIONAL ACADEMIES**  
*Advisers to the Nation on Science, Engineering, and Medicine*

laboratory and therefore cannot be studied using the central tool of classical microbiology. Metagenomics also enables the study of entire microbial communities, offering a window to intact microbial systems in which all of the parts can be examined individually or as a working whole.

In 2005, the National Science Foundation, five Institutes of the National Institutes of Health, and the Department of Energy asked the National Research Council to assemble a committee to address the current state of metagenomics. Their charge was to identify obstacles current researchers are facing, explore the potential of metagenomics, and determine how best to stage its development.

### Applications of Metagenomics

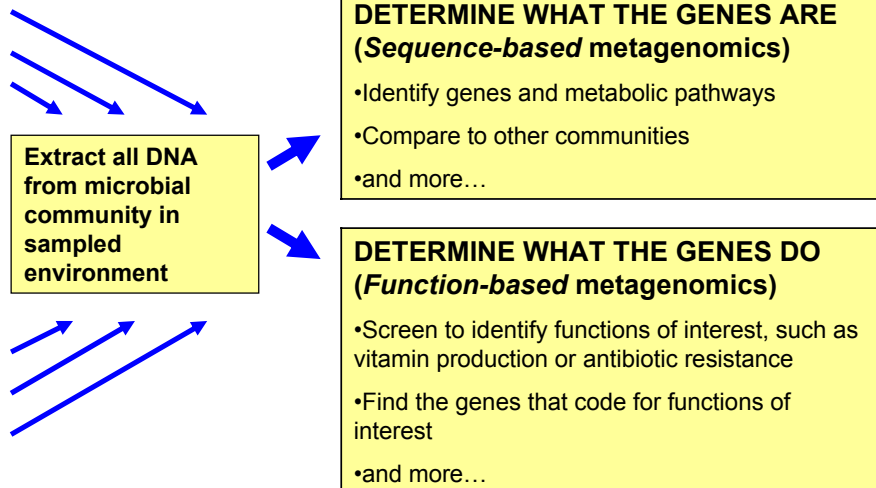
Cracking the secrets of some of the earth's countless microbial communities will reveal solutions to myriad challenges in human health, agriculture, and environmental stewardship. The metagenomics process (see figure below) has already been used to identify novel antibiotics and has revealed proteins involved in antibiotic resistance, vitamin production, and pollutant degradation. Metagenomics has been applied to many environments, including the oceans,

#### What is a microbe?

*Microbes* are living things too small to see with the naked eye—generally smaller than about 0.2 mm (most are around .001 mm). Microbiologists have specific names for the various types of microbes, which include bacteria, archaea, viruses, and small eukaryotes and fungi. This report focuses on metagenomics projects involving bacteria, archaea, and viruses. While most metagenomics projects exclude microbes that have large genomes, such as those called eukaryotes, they will likely be included in more projects as technology advances. Most of the diversity of life on Earth is microbial. Metagenomics offers a new way to study the microbial world—most of which has never before been accessible to scientists.

soils, thermal vents, hot springs, and the human mouth and gastrointestinal tract. Areas in which metagenomics offers particular value include: medicine, alternative energy, environmental remediation, biotechnology, agriculture, biodefense, and forensics.

In addition to the practical applications of knowledge gained through metagenomics, it is likely that fundamental biological concepts will also be reshaped. Increased understanding of microbes will lead to new concepts of genomes, species, evolution, and ecosystem robustness that will have effects beyond the



**The Metagenomics Process.** First, DNA is extracted from all the bacteria and archaea living in a particular environment. Typically, laboratory bacteria are then induced to take up and replicate the fragments of the extracted DNA, creating a “library” containing the genomes of all the bacteria and archaea found in the sampled environment. (New technologies facilitate studying a community’s DNA directly, bypassing the creation of a library.) The DNA of the community can then be studied in several ways. In *sequence-based* metagenomics, researchers analyze the DNA to identify genes and metabolic pathways by comparing the DNA with genes found in other communities or samples. In *function-based* metagenomics, researchers screen the DNA library for specific functions, such as vitamin production or antibiotic resistance. When a function of interest is detected, then the DNA coding for that function is sequenced and compared with DNA from other organisms or communities. While the metagenomics process has much in common with classical genomics, it involves studying the genome of an entire community rather than that of an individual species.



The microbial communities in soil and on plants play a central role in the health and productivity of crops. Metagenomics enables the study of these complex communi-

ties without culturing individual species. Agriculture is one of the many areas in which metagenomics has practical applications.

specific field of microbiology. Metagenomics can help scientists address questions such as “How do microbes evolve?” “What is the role of microbes in maintaining the health of their hosts?” and “How diverse is life?”

## Challenges of Metagenomics

Much has been learned from early metagenomics studies and it is starting to become clear which steps in the process commonly present difficulties and obstacles. The metagenomics research community—which will include scientists working on a very broad range of habitats and funded by many different agencies—should work together to disseminate advances, agree on common standards, and develop guidelines on best practices in metagenomics that would be of use to all of the funding agencies interested in supporting metagenomics research. The report specifically addresses the following issues:

**Interdisciplinary Collaboration:** Interdisciplinary collaboration will enhance the value of metagenomics projects and aid in the use of new knowledge for practical applications. The involvement of experts on the particular environment being sampled will amplify the value of metagenomics data. Expertise that would contribute to collaborations include: atmospheric, ocean, soil and water studies; geology; medicine; veterinary science; agricultural science; environmental science; and bioengineering. Virtually all biologists—whether they work on evolution, development, ecology or cancer and whether they study yeast, plants, corals, or mammals—will find that greater understanding of microbial communities has something to contribute to their research.

**Governmental Stakeholders:** Because the application areas are so broad, the governmental stakeholders in metagenomics are numerous. The report recommends that an interagency working group take responsibility for ensuring that the development of the field of metagenomics occurs in the context of ongoing com-

munication and coordination among the interested government agencies.

**Methodological Challenges:** Various sampling methods and DNA extraction techniques present a challenge for the standardization of metagenomics methods. It is essential to consider sampling issues and limitations at the beginning and throughout any study of a complex community. Data obtained from metagenomic analysis of any community will only be as good as the procedures used for the extraction of DNA from an environmental sample. Sampling schemes and DNA extraction methods must inform the interpretation of results. Additionally, developing ways to use different laboratory hosts for detecting the functions of genes in a community’s genome will be a challenge for function-based metagenomics projects.

**Data Analysis/Bioinformatics:** Challenges such as determining how to use pooled sequence data to determine the complete genome of individual community members, comparing the diversity of various environments, and assessing changes in diversity, need to be addressed. Improvements in bioinformatics tools, culturing techniques, and physical separation methods, along with the generation of complete genome sequences for model microorganisms, will all make it easier to interpret metagenomic sequence data and in some cases assemble whole genomes.

**Data Archiving:** The enormous amounts of data generated by metagenomics studies should be made publicly available in international archives as rapidly as possible. The report recommends the establishment of specialized, peer-reviewed, and continuously maintained databases for storing and sharing metagenomics data. These databases should store not only sequences generated by metagenomics projects, but also information about the sampling and DNA extraction methods used to obtain the data and the computational and algorithmic methods used to analyze the data. They should include specialized tools that enable deposited data to be manipulated in different ways by different researchers, and thus to add value to the data.

## Establishing a “Global Metagenomics Initiative”

The report recommends the establishment of a Global Metagenomics Initiative that includes a small number of large-scale, comprehensive projects that use metagenomics to understand model microbial communities, a larger number of middle-sized proj-



ects, and many small projects. Large-scale projects would explore a few microbial communities in great depth, exploring a habitat with attention to variation, commonalities, and detailed characterization. Medium-sized projects would provide centers of excellence in metagenomics that can be more diverse than the large-scale projects, but would include multidisciplinary approaches to the study of a community. The small-scale projects would be single-investigator initiated and would examine a slice of a community, a particular function in multiple communities, or a specific technical advance.

#### **Large-Scale Metagenomics Projects**

Large-scale metagenomics studies would establish methods, approaches, and conceptual insights that could be applied to ever more complex and dynamic systems. The report recommends the establishment of a small number of large-scale projects representing a breadth of habitat types, including:

- Microbial communities in a natural environment such as seawater or soil.
- Host-associated microbial communities such as the microbial inhabitants of the human gut or oral cavity.
- Microbial communities in a “managed-environment” such as a sewage treatment plant or an aquaculture facility.

#### **Conclusions**

Traditional microbiology has revealed how critical microbes are to life on earth—from helping humans digest food to cleaning up hazardous waste; they perform many of the functions that

are essential to the habitability of the planet. The new approach of metagenomics will greatly extend scientists’ ability to discover the incredible capabilities of microbial communities. The landscape of metagenomics is as expansive as microbiology itself. Defining the metagenomic characteristics of microbial communities is a critical first step in understanding their contributions to the health of the planet, their roles in the well-being of humans, and the environmental consequences of human activities.

As metagenomics develops, researchers and funding agencies will need to address a variety of technical and structural challenges. The field will benefit from a framework of interdisciplinary coordination, new bioinformatics and data management tools, effective systems of data sharing, and strong methodological standards. The report recommends the establishment of a “Global Metagenomics Initiative” to drive advances in the field in the same way that the Human Genome Project advanced the mapping of the human genetic code. This “Global Metagenomics Initiative” would be an effective way to advance the field and enable the application of information about microbial communities to a variety of areas, including medicine, energy, biotechnology, agriculture, and many others.

The more that is known about microbes, the greater value metagenomics data will have. Thus, basic microbiology research should not be neglected, but instead be strengthened and deepened. An active dialogue between metagenomics researchers and other microbiologists and their representatives in funding agencies will help guide the fields in complementary directions.

---

#### **Committee on Metagenomics: Challenges and Functional Applications**

**Jo Handelsman (Co-chair)**, University of Wisconsin-Madison; **James M. Tiedje (Co-chair)**, Michigan State University; **Lisa Alvarez-Cohen**, University of California, Berkeley; **Michael Ashburner**, University of Cambridge; **Isaac K. O. Cann**, University of Illinois, Urbana-Champaign; **Edward F. DeLong**, Massachusetts Institute of Technology; **W. Ford Doolittle**, Dalhousie University; **Claire M. Fraser-Liggett**, The Institute for Genomic Research; **Adam Godzik**, The Burnham Institute; **Jeffrey I. Gordon**, Washington University School of Medicine; **Margaret Riley**, University of Massachusetts, Amherst; **Molly B. Schmid**, Keck Graduate Institute; **Ann H. Reid (Study Director)**, National Research Council.

**This report brief was prepared by the National Research Council based on the committee’s report.** For more information, contact the Board on Life Sciences at [bls@nas.edu](mailto:bls@nas.edu) or visit <http://nationalacademies.org/bls>. *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet* is available from the National Academies Press, 500 Fifth Street, NW, Washington, D.C. 20001; (800) 624-6242; [www.nap.edu](http://www.nap.edu). Support for this publication was provided by the Presidents’ Circle Communications Initiative of the National Academies.