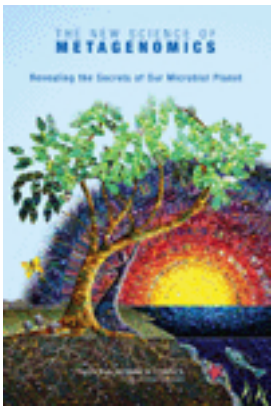
Free Executive Summary



The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet

Committee on Metagenomics: Challenges and Functional Applications, National Research Council

ISBN: 978-0-309-10676-4, 170 pages, 6 x 9, paperback (2007)

Although we can't usually see them, microbes are essential for every part of human life -- indeed all life on Earth. The emerging field of metagenomics offers a new way of exploring the microbial world that will transform modern microbiology and lead to practical applications in medicine, agriculture, alternative energy, environmental remediation, and many others areas. Metagenomics allows researchers to look at the genomes of all of the microbes in an environment at once, providing a "meta" view of the whole microbial community and the complex interactions within it. It's a quantum leap beyond traditional research techniques that rely on studying -- one at a time -- the few microbes that can be grown in the laboratory. At the request of the National Science Foundation, five Institutes of the National Institutes of Health, and the Department of Energy, the National Research Council organized a committee to address the current state of metagenomics and identify obstacles current researchers are facing in order to determine how to best support the field and encourage its success. The report recommends the establishment of a "Global Metagenomics Initiative" comprising a small number of large-scale metagenomics projects as well as many medium- and small-scale projects to advance the technology and develop the standard practices needed to advance the field. The report also addresses database needs, methodological challenges, and the importance of interdisciplinary collaboration in supporting this new field.

Summary

THE DAWNING OF A NEW MICROBIAL AGE

Microbes run the world. It's that simple. Although we cannot usually see them, microbes are essential for every part of human life—indeed all life on Earth. Every process in the biosphere is touched by the seemingly endless capacity of microbes to transform the world around them. It is microbes that convert the key elements of life—carbon, nitrogen, oxygen, and sulfur—into forms accessible to all other living things. For example, although plants tend to get credit for photosynthesis, it is in fact microbes that contribute most of the photosynthetic capacity to the planet. All plants and animals have closely associated microbial communities that make necessary nutrients, metals, and vitamins available to their hosts. The billions of benign microbes that live in the human gut help us to digest food, break down toxins, and fight off disease-causing microbes. We also depend on microbes to clean up pollutants in the environment, such as oil and chemical spills. All these activities are carried out by complex microbial communities—intricate, balanced, and integrated entities that adapt swiftly and flexibly to environmental change. Some of the communities, like those in soil, may contain thousands of interdependent kinds of microbes. Microbial communities not only are key players in maintaining environmental stability and the health of individual plants and animals, they can also live in extreme environments, at temperatures, pressures, and pH levels in which no other organisms can survive. Microbes have developed countless strategies for survival, their genomes contain the directions for countless biochemical transformations, and their communities have

adapted through countless individual generations and billions of years of environmental change. In addition to their essential activities throughout the biosphere, microbes have been the source of numerous technologies that have improved the human condition. They are used commercially to produce most of the antibiotics and many other drugs in clinical use, to remediate pollutants in soil and water, to enhance crop productivity, to produce biofuels, to ferment many human foods, and to provide unique signatures that form the basis of microbial detection in disease diagnosis and forensic analysis.

Historically, the study of microbes has predominantly focused on single species in pure laboratory culture, and so understanding of microbial communities lags behind understanding of their individual members. Only recently have the tools become available to study microbes in the complex communities where they actually live and thus to begin to understand what they are capable of and how they 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.

The opportunity that stands before microbiologists today is akin to a reinvention of the microscope in the expanse of research questions it opens to investigation. Metagenomics provides a new way of examining the microbial world that not only will transform modern microbiology but has the potential to revolutionize understanding of the entire living world. In metagenomics, the power of genomic analysis is applied to entire communities of microbes, bypassing the need to isolate and culture individual bacterial community members. The new approach and its attendant technologies will bring to light the myriad capabilities of microbial communities that drive the planet's energy and nutrient cycles, maintain the health of its inhabitants, and shape the evolution of life. Metagenomics will generate knowledge of microbial interactions so that they can be harnessed to improve human health, food security, and energy production.

Metagenomics combines the power of genomics, bioinformatics, and systems biology. Operationally, it is novel in that it involves study of the genomes of many organisms simultaneously. It provides new access to the microbial world; the vast majority of microbes cannot be grown in the laboratory and therefore cannot be studied with the classical methods of microbiology. Although community ecology is not new to microbiology, the ability to bring to bear the power of genomics in the study of communities initiates an unparalled opportunity.

WHAT IS METAGENOMICS?

Like genomics, metagenomics is both a set of *research techniques*, comprising many related approaches and methods, and a *research field*. In Greek, *meta* means "transcendent." In its approach and methods, metagenomics overcomes the twin problems of the unculturability and genomic diversity of most microbes, the biggest roadblocks to advancement in clinical and environmental microbiology. *Meta* in the first sense means that this new science seeks to understand biology at the aggregate level, transcending the individual organism to focus on the genes in the community and how genes might influence each other's activities in serving collective functions. In the second sense, *meta* also recognizes the need to develop computational methods that maximize understanding of the genetic composition and activities of communities so complex that they can only be sampled, never completely characterized.

Metagenomics, still a very new science, has already produced a wealth of knowledge about the uncultured microbial world because of its radically new ways of doing microbiology. All metagenomics studies take the same first step: DNA is extracted directly from all the microbes living in a particular environment. The mixed sample of DNA can then be analyzed directly, or cloned into a form maintainable in laboratory bacteria, creating a library that contains the genomes of all the microbes found in that environment (see Box S-1). The library can then be studied in several ways, based primarily either on analyzing the nucleotide sequence of the cloned DNA or on determining what the cloned genes can do when they are expressed as proteins. It is important to recognize that the library is not organized into neat volumes, each containing the genome of one community member. Instead,

BOX S-1
Clones and Libraries

The word *clone* can have several different meanings in biology. In the context of this report, the word is used to describe a process whereby fragments of DNA isolated from a microbial community are inserted—or *cloned*—into circular pieces of DNA called plasmids. Laboratory bacteria can be manipulated to take up all the plasmids; when the bacteria subsequently divide, they replicate the plasmid along with their genomic DNA. When a large collection of plasmids containing all the DNA fragments from a given community is cloned into a bacterial culture, the resultant collection of bacteria is called a *library*—a living repository of all of the DNA from a microbial community.

it consists of millions of clones, each holding a random fragment of DNA. A metagenomics library is like thousands of jigsaw puzzles jumbled into a single box—putting the puzzles together again is one of this new science's great challenges. The metagenomics approach is now possible because of the availability of inexpensive, high-throughput DNA sequencing and the advanced computing capabilities needed to make sense of the millions of random sequences contained in the libraries.

Sequence-based metagenomics captures a massive amount of information on the microbial community under study. A study of the metagenome of the microbial inhabitants of the Sargasso Sea, for example, generated sequences of about a million genes and revealed whole classes of genes that were more diverse than could ever have been anticipated on the basis of studies of cultured organisms. At the other end of the spectrum, studies of a simple microbial community that lives in the extremely acidic water draining from metal mines demonstrated the potential of metagenomics to dissect detailed interactions among microbial-community members.

Metagenomics, however, is more than just large-scale sequencing. In *function-based metagenomics*, millions of random DNA fragments in a library are translated into proteins by bacteria that grow in the laboratory. Clones producing "foreign" proteins are then screened for various capabilities, such as vitamin production or antibiotic resistance. This enables researchers to access the tremendous genetic diversity in a microbial community without knowing anything about the underlying gene sequence, the structure of the desired protein, or the microbe of origin. New antibiotics and resistance mechanisms have already been discovered using function-based metagenomics.

STAGING THE FUTURE OF METAGENOMICS

The landscape of metagenomics is as expansive as microbiology itself. Microbial communities live virtually everywhere, and we are largely ignorant of their inhabitants and ecology; so there are literally millions of potential metagenomics projects. Each project would generate massive amounts of DNA sequence and functional data. To understand the potential of this new field and to determine how best to stage its development and encourage its success, several US government agencies—the National Science Foundation, five institutes of the National Institutes of Health, and the Department of Energy—asked the National Research Council to undertake an 18-month study of the emerging field of metagenomics. The Committee on Metagenomics: Challenges and Functional Applications was charged with describing the current state of the field and identifying obstacles that current researchers are facing. The committee was also asked to recommend the most promising directions for future metagenomics research and pos-

sible mechanisms for addressing infrastructure needs and improving communication and collaboration among groups studying different microbial communities. The committee met four times in 2006, including two short workshops: one on the implications of the massive amount of data generated by metagenomics and one on the questions of how and whether the nonbacterial members of environmental communities could be included in metagenomics studies (see Statement of Task, Appendix A).

Until recently, the complex microbial communities inhabiting nearly every environment and organism on Earth have essentially been invisible. With metagenomics, the astonishing genetic and metabolic diversity of the microbial world will be increasingly revealed. The practical applications of knowledge of these previously unseen realms of nature will be only part of the result. It is likely that as new biological strategies are brought to light, fundamental biological concepts will be affected. Basic ideas that organize biologists' understanding of the living world may need refinement in the face of greater understanding of how microbial communities function. New concepts of genomes, species, evolution, and ecosystem robustness will have effects beyond the specific field of microbiology. The questions that must be asked are "deep" ones, but answers will in all cases inform and guide the work of putting increased knowledge of microbial communities to practical use.

MAJOR ACADEMIC, GOVERNMENTAL, AND COMMERCIAL STAKEHOLDERS

There are many potentially beneficial collaborations among various academic disciplines in metagenomics projects, including atmospheric, ocean, soil, and water studies; geology; medicine; veterinary science; agricultural science; environmental; and bioengineering. It is, however, perhaps the field of biology that will be most affected by increasing knowledge of microbes. Virtually all biologists—whether they work on evolution, development, ecology, or cancer and whether they study yeasts, plants, corals, insects, birds, or mammals—will find that greater understanding of microbial communities has something to contribute to their research.

Because the applications are so broad, the government stakeholders in metagenomics are numerous. Metagenomic study of microbial communities has the potential to contribute to the missions of many government agencies. Fortunately, there is already a mechanism for 12 US government agencies with interests in microbiology to share information about their activities. The Microbe Project is an interagency working group formed in August 2000. The mission of the Microbe Project 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 committee hopes that this existing mechanism will prove useful in ensuring that the development of the field of metagenomics occurs in the context of continuing communication and coordination among the interested government agencies. Besides the United States, metagenomics projects are also under way in the European Community, Canada, China, Brazil, Singapore, South Korea, and Japan, and including these and other international groups in planning for the field of metagenomics would be worthwhile.

DIFFICULTIES FACING CURRENT RESEARCHERS

The sequence-based metagenomics approach has already been applied to many environments, including the ocean, many soils, coral reefs, whale carcasses, thermal vents, and hot springs. The microbial communities associated with different organisms—including humans, termites, aphids, and worms—have been studied. Function-based metagenomics has been used to identify novel antibiotics and proteins involved in antibiotic resistance, vitamin production, and pollutant degradation. Much has been learned from the early efforts, and it is starting to become clear which steps in the process commonly present difficulties and obstacles.

The starting material for a metagenomics study is a mixture of DNA from a community of cells that may include bacterial, archaeal, eukaryotic, and viral species at different levels of diversity and abundance. In some projects, sample collection may be confounded because too little DNA is present or because compounds are present that interfere with DNA extraction. Contaminating DNA from a microbial community's host or from eukaryotic members of a community needs to be excluded from current metagenomic analyses because the amount of DNA they contain overwhelms both sequencing capacity and computational analysis. The quality and completeness of data obtained from metagenomic analysis of any community will be only as good as the procedures used for the extraction of DNA from an environmental sample.

Determining how best to sample a microbial community for metagenomics is also fraught with challenges. Change in habitats over time is one of the most interesting aspects of communities, and their responses to changing conditions are central to understanding community structure, function, and robustness. Similarly, understanding the role of host-associated microbial communities in host development and health requires not only sampling from the same host over time, but also understanding host-to-host variation. But habitat and host variability exacerbate the sampling conundrum. Over time, as biological and computational methods become more efficient, we will be able to draw more robust conclusions from more complex communities in more variable habitats. No matter the power of

the methods now or in the future, it is essential to consider sampling issues and limitations at the beginning and throughout any metagenomic study of a complex community, and the sampling scheme must inform the interpretation of results.

Extracting maximal information from metagenomic libraries will continue to be challenging, primarily because of the massive size and complexity of the datasets. Determining the complete genome of any individual community member from pooled sequence data is extremely difficult and currently achievable only for very simple communities. The problem is exacerbated by the uneven abundance of members of microbial communities, which leads to sampling the most abundant organisms over and over and often missing the rare ones entirely. New technologies that allow much greater depth of sequencing or that remove redundant DNA would make it possible to detect important members that may be rare. Finally, improvements in bioinformatics tools, culturing techniques, and physical separation methods—with the generation of complete genome sequences for model microbes—will all make it easier to interpret the metagenome sequence data and in some cases to assemble whole genomes from metagenomic sequence data.

Function-driven metagenomics has already unearthed many proteins that would not have been recognized by their sequences alone. The potential for discovery is staggering but would greatly benefit from the development of new techniques and host organisms to allow genes from a wide variety of microbes to be expressed in the laboratory.

RECOMMENDATIONS

The opportunity afforded by metagenomics to study microbial communities in their natural state represents an endless frontier. Given the intense competition for science funding, some priority-setting is necessary to ensure that the most possible value is gained from early metagenomics investments. The diversity of habitats on Earth, the complexity of microbial communities, and the myriad functions governed by microbes suggest that highly productive metagenomics research will be possible in decentralized, *small-project settings*. However, no individual researcher is likely to have the capability and resources to achieve a comprehensive characterization of a complex microbial community. Therefore, there is also a substantial need for *medium-sized, collaborative projects* that involve multiple investigators. Both mechanisms of funding are tested and proven effective in advancing new fields of science. The mixture of single- and multi-investigator projects maximizes the diversity of scientific approaches, assures that many avenues of research are pursued simultaneously, presents an opportunity to study many habitats, and engages a broad community, thereby utilizing

the creativity of many investigators. All these benefits are essential for the advancement of the field.

Metagenomics, however, differs from much of the science that precedes it in its complexity, multidisciplinarity, and in the magnitude of its unknowns. Its very nature departs from each of the fields—microbiology, ecology, and genomics—that fuse to form this new science. Consequently, metagenomics presents a number of conceptual and technical obstacles that limit the productivity of all metagenomics researchers. The committee believes that the needs of the metagenomics field are not entirely met by current funding mechanisms. Encouraged by the example of the human and other model organism genome projects, the committee believes that the best way to spur these advances is through a multi-scale approach. The committee 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 projects, and many small projects.

The committee believes that the field of metagenomics would be greatly advanced by the establishment of a few large, internationally coordinated projects with the goal of characterizing in great detail a small number of carefully chosen microbial communities. These large-scale model metagenomics projects would enable collaboration and coordination that are difficult to achieve in smaller projects. Large-scale projects could unite scientists of multiple disciplines around the study of a particular sample, habitat, function, or analytical challenge—an approach that is more likely to illuminate themes and advance technical approaches than would a disparate group of small projects by researchers with different goals and nonuniform methods. These large-scale projects would also serve as incubators for the development of novel technologies, analytical techniques, and community databases and would equip smaller-scale projects with the knowledge to design efficient sampling schemes, make informed choices about habitats to study, and identify fruitful strategies for identifying specific functions. Moreover, large projects would furnish the basis for developing a new conceptual framework for microbial ecology, as well as a new community of young scientists, that will guide the design of predictive models about community behavior.

Because the study of microbial communities has the potential to contribute to the missions of so many government agencies, it is likely that each will support a portfolio of small-scale metagenomics projects relevant to its particular mission. However, the metagenomics research community, which will include scientists working on a broad array of habitats and funded by many agencies, should be encouraged to work together to disseminate advances, agree on common standards, and develop guidelines on best practices in metagenomics that would be of use to all the funding agen-

cies interested in supporting metagenomics research. This should include attention to bringing sample collection into alignment with international agreements and local values.

Information from metagenomics studies will be exploited fully only if appropriate data management and analysis methods are in place. Furthermore, metadata—information on the sampling method, sample treatment and data about the sampled habitat—are essential for the analysis of metagenomics sequence data. If metagenomics data are to be used to their fullest advantage, a metadata infrastructure is an urgent need. No metadata standard will be appropriate to all habitat types, but there should be close collaboration and coordination among the communities of scientists developing metadata standards.

In the genomic-sequencing community, many of the major species being studied have special community genomics databases, for example, *FlyBase* for the fruitfly *Drosophila*,[1] and *TAIR* for the model plant *Arabidopsis*.[2] This model—community databases organized to accommodate metagenomics data from particular environments or organisms—appears to be a promising approach to providing convenient access to the data of metagenomics projects.

One major challenge faced by metagenomics databases in contrast with "conventional" genomics databases will be the demand for community input into the annotation process. Annotation is the process of assigning functional, positional, and species-of-origin information to the genes in a database. In conventional genomics, primary responsibility for annotating data falls on the authors, and annotations are not often updated. In metagenomics projects, annotations will change as additional data (or metadata) are collected by other groups and an annotation database must be able to accept and integrate individual and large-scale (computational) annotations of metagenomic data continually. The need for dynamic and flexible annotation may make it essential that community metagenomics databases be provided sufficient resources to support ongoing, professional curation.

The analysis of genomics data is absolutely dependent on computer software. In general, grants for metagenomics projects will require an even higher percentage of funds for bioinformatic and statistical support than have genomics projects or than may be typical for other kinds of biological research. It is common for software developed for a particular project gradually to find widespread use in the community. Providing a mechanism whereby analytical tools that have proved their value to the community can be brought up to robust, engineered, documented form would be very

[1] *http://www.flybase.org/.*
[2] *http://www.arabidopsis.org/.*

worthwhile. This is a pipeline that is poorly supported by traditional grant-funding mechanisms.

The rise of genomics has been characterized by both technological and scientific innovations and by novel practices in data dissemination. In the early 1980s the scientific community in Europe and the United States established community archives for nucleic acid sequence data. These data immediately became accessible in a form suitable for computer analysis and were freely available, without impediment to all researchers, whether in academe or in industry. It is no exaggeration to state that without these publicly accessible databanks, the success of the Human Genome Project and similar genome projects would not have been possible. It is vital that the metagenomics community continue to adhere to the practice of publicly depositing, in a timely manner, all relevant data.

It should also be remembered that the more is known about microbes, the greater value metagenomics data will have. Thus, it is extremely important that basic microbiology research not be neglected, but instead be strengthened and deepened. Active communication between metagenomics researchers and members of other subdisciplines of microbiology and their representatives in funding agencies will help to guide the various fields in complementary directions.

TRAINING AND PUBLIC OUTREACH

Metagenomics presents some specific challenges for training experts and some global opportunities for educating the public about microbiology. The interdisciplinary nature of the science of metagenomics necessitates deployment of new training programs to encourage scientists to broaden their skills beyond those learned in their own disciplines. Graduate programs, intensive courses, fellowship programs, and sabbatical support are all mechanisms that can be used to develop investigators with the necessary configuration of skills and knowledge. Metagenomics also offers an opportunity to integrate public communication into graduate training. Each metagenomics project should design ways of teaching graduate students the principles of effective public outreach and then provide opportunities for them to use their new skills.

The dazzling power and opportunity of metagenomics as well as the "Big Science" nature of the large-sized projects in the Global Metagenomics Initiative will attract public interest in microbiology. The sense of delving into a truly unknown world, the potential for deriving human benefit from microbes, and the sheer power of microbes to influence just about every earthly function provide an irresistible draw for the public. Therefore, both large and small projects can be used as catalysts for teaching microbiology. Each large project should have a budget for developing materials that

explain its scientific basis and implications in accessible and interesting ways. All metagenomics scientists should be encouraged to teach about their science in their local communities. In turn, these outreach efforts would provide a training ground for a new generation of scientists who are skilled in communicating science to the public.

THE NEW SCIENCE OF METAGENOMICS

Revealing the Secrets of Our Microbial Planet

Committee on Metagenomics: Challenges and Functional Applications

Board on Life Sciences
Division on Earth and Life Studies

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract MCB—0544539 between the National Academy of Sciences and the National Science Foundation (NSF), Contract N01-OD-4-2139 between the National Academy of Sciences and the Department of Health and Human Services, National Institutes of Health (NIH), and Contract DE-AT01-05ER64072 between the National Academy of Sciences and the Department of Energy (DOE). The content of this publication does not necessarily reflect the views or policies of NIH, NSF, or DOE, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

International Standard Book Number-13: 978-0-309-10676-4
International Standard Book Number-10: 0-309-10676-1

Cover: Design by Francesca Moghari; artwork by Nicolle Rager Fuller (*www.sayo-art.com*).

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, NW, Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON METAGENOMICS: CHALLENGES AND FUNCTIONAL APPLICATIONS

JO HANDELSMAN *(Cochair)*, University of Wisconsin, Madison
JAMES TIEDJE *(Cochair)*, Michigan State University, East Lansing
LISA ALVAREZ-COHEN, University of California, Berkeley
MICHAEL ASHBURNER, University of Cambridge, United Kingdom
ISAAC K. O. CANN, University of Illinois, Urbana-Champaign
EDWARD F. DELONG, Massachusetts Institute of Technology, Cambridge
W. FORD DOOLITTLE, Dalhousie University, Halifax, Nova Scotia, Canada
CLAIRE M. FRASER-LIGGETT, University of Maryland School of Medicine, Baltimore
ADAM GODZIK, Burnham Institute for Medical Research, La Jolla, CA
JEFFREY I. GORDON, Washington University School of Medicine, St. Louis, MO
MARGARET RILEY, University of Massachusetts, Amherst
MOLLY B. SCHMID, Keck Graduate Institute, Claremont, CA

Staff

ANN H. REID, Study Director
FRANCES E. SHARPLES, Director, Board on Life Sciences
ANNE F. JURKOWSKI, Senior Program Assistant
MERC FOX, Program Assistant
NORMAN GROSSBLATT, Senior Editor

Acknowledgments

This report has been reviewed in draft form by persons chosen for their diverse perspectives and technical expertise in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of the independent review is to provide candid and critical comments that will assist the institution in making the published report as sound as possible and to ensure that the report meets institutional standards of objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following for their review of the report:

Gary Anderson, Lawrence Berkeley National Laboratory, Berkeley, CA
Jeffrey Dangl, University of North Carolina, Chapel Hill
Julian E. Davies, University of British Columbia, Vancouver, BC, Canada
Jed Fuhrman, University of Southern California, Los Angeles
Dennis Mangan, University of Southern California School of Dentistry, Los Angeles
Victor Markowitz, Lawrence Berkeley National Laboratory, Berkeley, CA
Randall Murch, Virginia Polytechnic Institute and State University, Blacksburg
Norman R. Pace (NAS), University of Colorado, Boulder
David Relman, Stanford University, Stanford, CA
Edward Rubin, Lawrence Berkeley National Laboratory, Berkeley, CA
George Weinstock, Baylor College of Medicine, Houston, TX

Although the reviewers listed above have provided constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of the report was overseen by **John Wooley,** University of California, San Diego. Appointed by the National Research Council, he was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of the report rests entirely with the author committee and the institution.

The committee benefited from briefings provided by several speakers. At its second meeting, on May 2, 2006, the committee was briefed by: **Michael Gray** (by telephone), Professor and Department Head, Canada Research Chair in Genomics and Genome Evolution, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada; **Mitchell Sogin,** Senior Scientist and Director of the Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, The Woods Hole Biological Laboratory, Woods Hole, MA; and **Robert Edwards,** San Diego State University and Burnham Institute, San Diego, CA. At its third meeting, on July 27, 2006, the committee was briefed by: **David J. Lipman,** Director, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Rockville, MD; **Rolf Apweiler,** Head of Sequence Database Group, European Bioinformatics Institute, Cambridge, UK; **Victor Markowitz,** Head of Lawrence Berkeley National Lab's Biological Data Management and Technology Center, Berkeley, CA; **Paul Gilna,** Executive Director, CAMERA, San Diego, CA; and **Amaranth Gupta,** Associate Research Scientist, Director Advanced Query Processing Lab, San Diego Supercomputer Center, University of California, San Diego.

The committee extends heartfelt thanks to Ann Reid who served as Study Director for this report. The product reflects both Ann's attention to our charge and her ability to provoke us into addressing it thoroughly. Her outstanding editing contributed greatly to the clarity and logic of the report. We also thank Anne Jurkowski for her dedication to this report and its authors. Throughout the process, the committee relied on Anne's administrative prowess and her willingness to do whatever was necessary to get the report done or the committee on track. Anne's aesthetic intuition and visual acuity shaped the report as well as its derivative materials.

We thank Dr. Patrick Schloss for his assistance in building the metagenomics bibliography and Dr. Luke Moe, Snow Brook Peterson, and Dr. Ainslie Little for helpful discussion and Christina Matta for assuring historical accuracy.

Contents