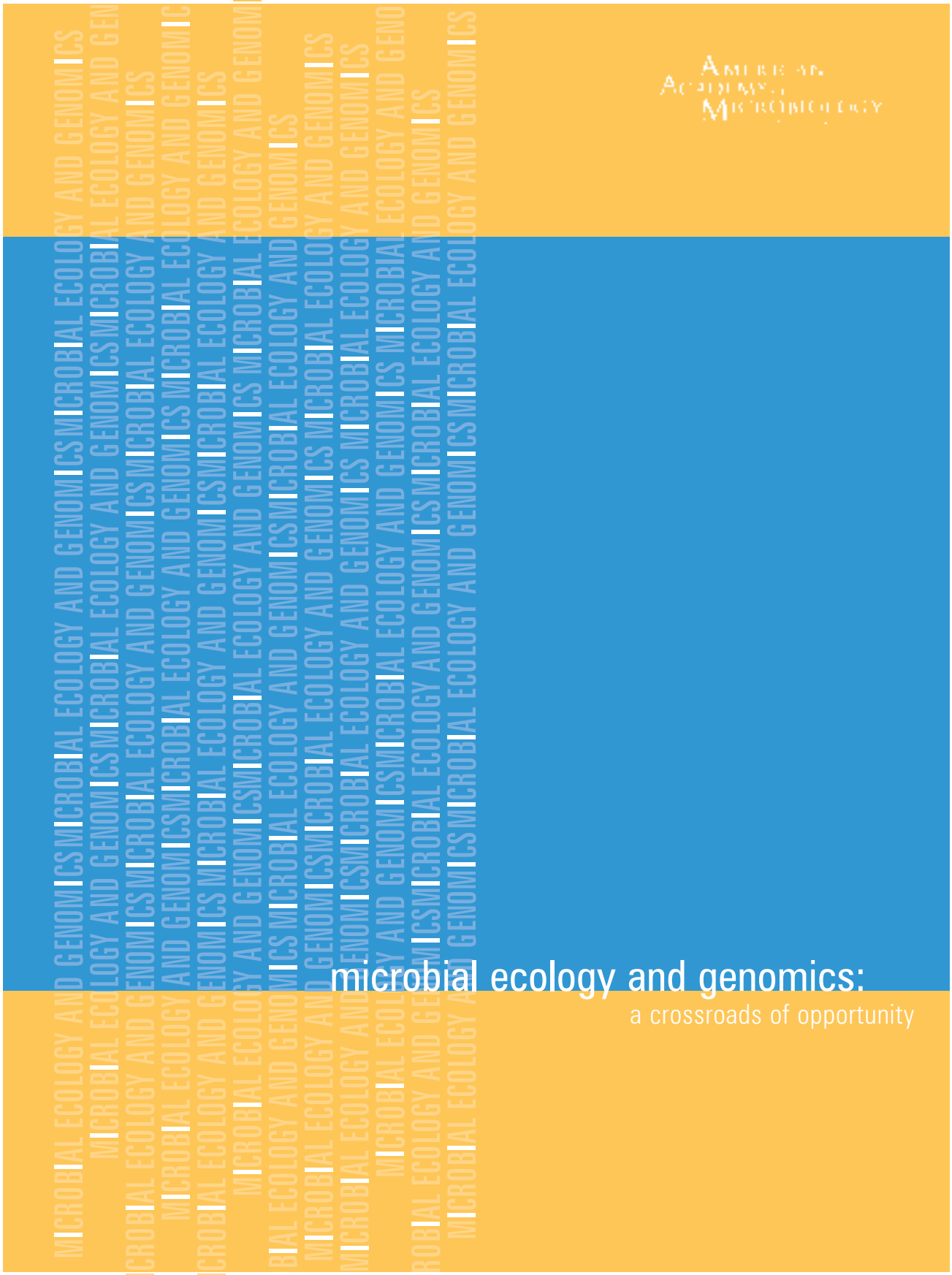


American Academy of Microbiology

microbial ecology and genomics:

a crossroads of opportunity

This report is based on a colloquium sponsored by the American Academy of Microbiology held February 23-25, 2001, in Singer Island, Florida.

The American Academy of Microbiology wishes to express its gratitude for the generous support of this project from the following organizations:

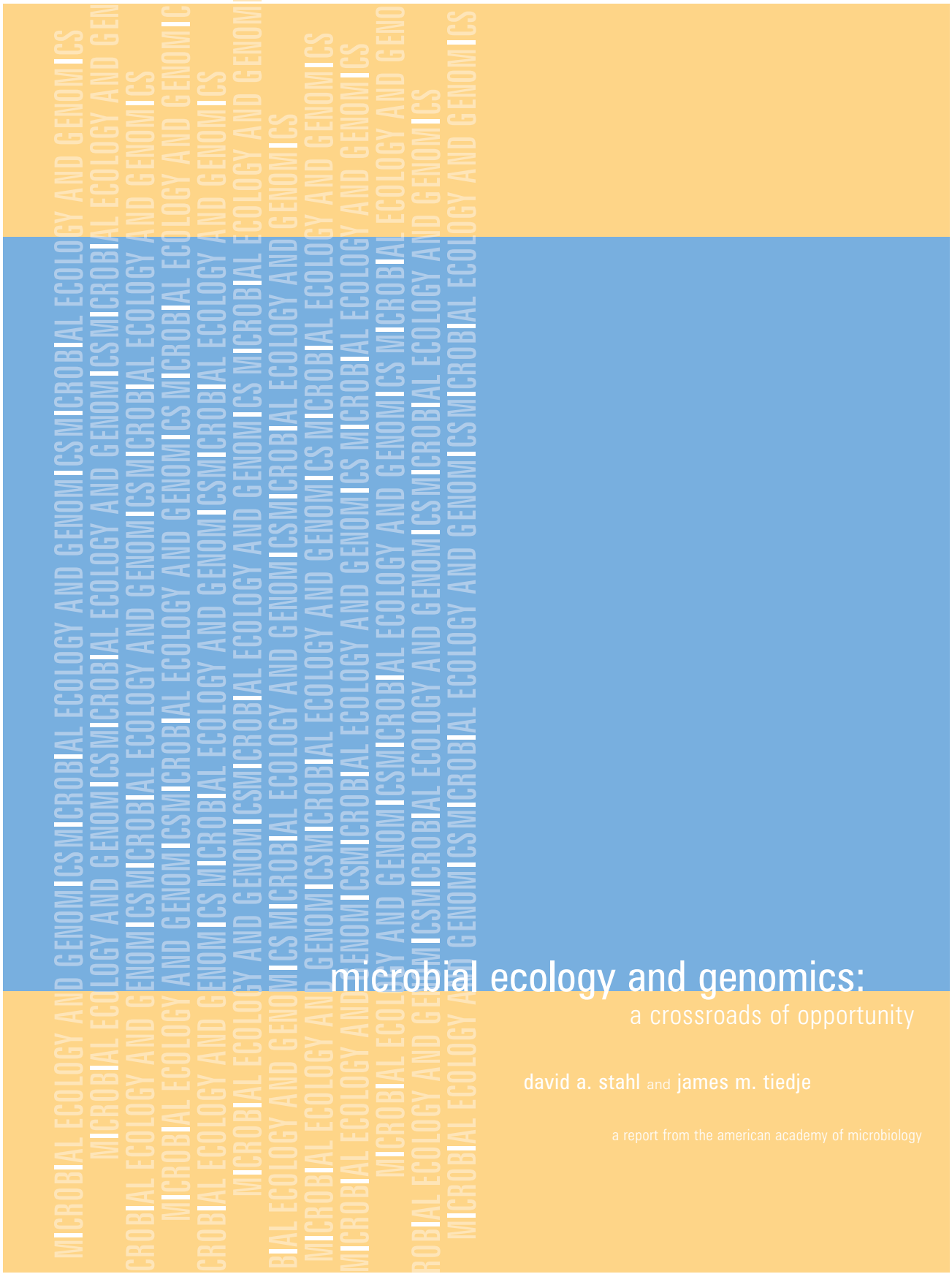
American Society for Microbiology
National Institute of Dental and Craniofacial Research,
National Institutes of Health, 1 R13 DE14267-01
National Science Foundation, DEB-0090711
United States Department of Energy
United States Environmental Protection Agency

American Academy of Microbiology
1752 N STREET, NW
WASHINGTON DC 20036

This report and others from the American Academy of Microbiology are available on-line:
www.asmusa.org

microbial ecology and genomics:
a crossroads of opportunity

A REPORT FROM THE AMERICAN ACADEMY OF MICROBIOLOGY



microbial ecology and genomics:

a crossroads of opportunity

david a. stahl and james m. tiedje

a report from the american academy of microbiology

Colloquium Steering Committee

David A. Stahl, Ph.D.,
University of Washington (Co-Chair)
James M. Tiedje, Ph.D.,
Michigan State University (Co-Chair)
Penny S. Amy, Ph.D.,
University of Nevada, Las Vegas
David L. Balkwill, Ph.D.,
Florida State University
George M. Garrity, Ph.D.,
Michigan State University
Gary J. Olsen, Ph.D.,
University of Illinois
Erko S. Stackebrandt, Ph.D.,
DMS, Braunschweig, Germany
James T. Staley, Ph.D.,
University of Washington
Carl Woese, Ph.D.,
University of Illinois
Carol A. Colgan, Director,
American Academy of Microbiology

Board of Governors, American Academy of Microbiology

Eugene W. Nester, Ph.D. *(Chair)*,
University of Washington
Joseph M. Campos, Ph.D.,
Children's National Medical Center, Washington, DC
R. John Collier, Ph.D.,
Harvard Medical School
Marie B. Coyle, Ph.D.,
Harborview Medical Center, University of Washington
James E. Dahlberg, Ph.D.,
University of Wisconsin, Madison
Julian E. Davies, Ph.D.,
Cubist Pharamaceuticals, Inc., Vancouver, BC, Canada
Arnold L. Demain, Ph.D.,
Drew University
Lucia B. Rothman-Denes, Ph.D.,
University of Chicago
Abraham L. Sonenshein, Ph.D.,
Tufts University Medical Center
David A. Stahl, Ph.D.,
University of Washington
Judy D. Wall, Ph.D.,
University of Missouri

Colloquium Participants

Rudi Amann, Ph.D.,
Max-Planck Institute for Marine Microbiology, Bremen, Germany
Penny S. Amy, Ph.D.,
University of Nevada, Las Vegas
David L. Balkwill, Ph.D.,
Florida State University
Tom Bruns, Ph.D.,
University of California, Berkeley
Colleen M. Cavanaugh, Ph.D.,
Harvard University
Sallie W. Chisholm, Ph.D.,
Massachusetts Institute of Technology
Edward F. Delong, Ph.D.,
Monterey Bay Aquarium Research Institute, CA
Charles Delwiche, Ph.D.,
University of Maryland
Richard Devereux, Ph.D.,
U.S. Environmental Protection Agency, Gulf Breeze, FL
Sorel Fitz-Gibbon, Ph.D.,
University of California, Los Angeles
George M. Garrity, Ph.D.,
Michigan State University
Michael A. Huston, Ph.D.,
Oak Ridge National Laboratory, TN
Matthew D. Kane, Ph.D.,
National Science Foundation
Stephen Lory, Ph.D.,
Harvard University
Dennis Mangan, Ph.D.,
NIDCR, National Institutes of Health
Shahid Naeem, Ph.D.,
University of Washington
Karen E. Nelson, Ph.D.,
The Institute for Genomic Research, Rockville, MD
Gary J. Olsen, Ph.D.,
University of Illinois
Frederick A. Rainey, Ph.D.,
Louisiana State University
David A. Relman, M.D.,
Stanford University School of Medicine
Margaret A. Riley, Ph.D.,
Yale University
Monica M. Riley, Ph.D.,
Marine Biological Laboratories, Woods Hole, MA
Abigail A. Salyers, Ph.D.,
University of Illinois
Karl H. Schleifer, Ph.D.
Technical University of Munich, Germany
Mitchell L. Sogin, Ph.D.,
Marine Biological Laboratories, Woods Hole, MA
Erko S. Stackebrandt, Ph.D.,
DMS, Braunschweig, Germany
David A. Stahl, Ph.D.,
University of Washington
James T. Staley, Ph.D.,
University of Washington
James M. Tiedje, Ph.D.,
Michigan State University
Jennifer J. Wernegreen, Ph.D.,
Marine Biological Laboratories, Woods Hole, MA

Executive Summary

Microbes have dominated life on Earth for most of its 4.5 billion-year history. They are the foundation of the biosphere, controlling the biogeochemical cycles and affecting geology, hydrology, and local and global climates. All life is completely dependent upon them. Humans cannot survive without the rich diversity of microbes, but most microbial species can survive without humans.

Extraordinary advances in molecular technology have fostered an explosion of information in microbial biology. It is now known that microbial species in culture poorly represent their natural diversity—which dwarfs conventions established for the visible world. This was revealed over the last decade using newer molecular tools to explore environmental diversity and has sparked an explosive growth in microbial ecology and technologies that may profit from the bounty of natural biochemical diversity. Several colloquia and meetings have helped formulate policy recommendations to enable sustained research programs in these areas. One such colloquium organized by the American Academy of Microbiology ("The Microbial World: Foundation of the Biosphere," 1997) made two key recommendations: (1) develop a more complete inventory of living organisms and the interagency cooperation needed to accomplish this goal, and (2) develop strategies to harvest this remarkable biological diversity for the benefit of science, technology, and society. Complete genome sequence information was identified as an essential part of strategy development, and the recommendation was made to sequence the genome of at least one species of each of the major divisions of microbial life.

At the time the 1997 AAM report was written, the recommendation to sequence one genome from each major division was an ambitious objective. Now, with the remarkable advances in sequencing technology and bioinformatics, genome sequencing is nearly routine. Comparative genomics is a robust and rapidly growing research arena. In addition, other technologies are coming to the fore. For example, DNA arrays now provide unprecedented insights into the physiology of an organism. The scientific and technological developments of the last several years have been so rapid as to bring us to a new crossroads of opportunity—analysis of the tremendous complexity of natural microbial systems in more complete terms. Thus, there is a pressing need to reexamine and clearly articulate new opportunities in microbial biology.

In February 2001, a selected group of scientists attended a three-day colloquium sponsored by the American Academy of Microbiology entitled, "Microbial Ecology and Genomics: A Crossroads of Opportunity." The colloquium participants discussed where the field should be going next in light of the rapid and heretofore unanticipated development of genomics sequencing and proteomics technologies. Just as knowledge of the human genome promises to revolutionize medical science, the application of genomic technologies to microbial evolution and environmental biology promises to revolutionize microbial biology. The rewards of carrying out such work are not just increased knowledge, but also economic ones.

A more complete understanding of microbial diversity and the environmental processes they control will require much more than a biotic inventory. It will require a deeper understanding of the basic units of organization and their interactions. Communities, not total biomass, control net process rates driving the biogeochemical cycles sustaining the biosphere. Thus, descriptions of the temporal and spatial dimensions of microbial community structure and the complex gene expression patterns that underlie trophic interactions are fundamental to a more complete understanding of our biosphere. In turn, this understanding will be incomplete without knowledge of the fundamental mechanisms contributing to genetic variation and speciation. Genome sequencing has revealed totally unexpected genetic plasticity within and among named microbial species, and horizontal DNA exchange is now recognized to be a major force in the shaping of their genomes and fostering biochemical innovation.

Rapid advances in genome science must be complemented by investment in systematics, developing a taxonomy better adapted to genomic information. Traditional taxonomic concepts (*i.e.*, species, genus, family) do not serve microbial systematics, in which problems of horizontal gene transfer and mechanisms of speciation and evolution are varied and complex. Currently, the

strain is the unit of taxonomy and research, but this needs to be refined, not only for more effective ecological and evolutionary research, but also for better storage and retrieval of cultures, genomic data, identification of organisms in samples, patenting, and bioprospecting.

Although the general patterns of macroorganismal diversity are relatively well known, spatial patterns of microorganismal diversity are completely unknown. Intensive microbial genetic/biodiversity surveys, covering the full range of environmental conditions and geological/ evolutionary histories, will be required to determine the patterns that exist. This is a prerequisite to developing hypotheses to explain these patterns and linking patterns to processes at local and global levels.

These studies will generate very large amounts of data, as is now most clearly shown by the explosive growth of genome sequence information. However, reducing sequence data to a more useful form through annotation lags far behind the accumulation of sequences. As yet there is no provision for systematic inclusion of environmental information in sequence repositories. Thus, an important need is the ability to get data out in usable form. Although a universally accessible database would be a step toward sharing data and communicating information to the scientific community, more specific steps need to be taken.

Ten years from now, we hope that we will have integrated genome sequence information with the environment. The environment is the context in which genomes evolved, function, and continue to evolve. It is the only context in which they can be fully understood. The future includes a 10-year plan, in which techniques, outreach and training, and targeted areas for specific research programs will provide a road map for a structured, rapid integration of genomics with microbial systematics, evolution, and ecology.

Genomics, including analysis, is tool driven. Thus, mechanisms need to be established to encourage and reward development of new technology needed to efficiently and broadly deploy these new techniques.

Recommendations

- Advance ecogenomics (the intersection of ecology and genome science) to explore the many levels of biological organization that sustain the biosphere.
- Enhance the study of genome plasticity to reveal the ecological significance of genome content and organization, polymorphisms, and genetic redundancy. Such studies should include comparative genomics of both closely and distantly related taxa. We recommend obtaining genome sequences for approximately 10 species within a well-defined species cluster and for 10 strains within one or more individual species to address the question of how separate are gene pools of putative species.
- Evaluate a radical restructuring of the prokaryotic paradigm, especially below the genus level, using genomic information.
- Support the expansion of culture collections that includes a central, well-ordered facility for the maintenance of type strains and their associated data.
- Develop integrated universal databases that include genomic, phenotypic, habitat, and geographic information.
- Establish a minimal set of data descriptors that can be embedded into a database and used in the generation of content for distribution (*e.g.*, through XML-enabled browsers) so that the information is more readily available and more easily maintained as additional environmental information comes available and taxonomic rearrangements occur.
- Develop new technologies, including (i) methods for measuring the activity of microorganisms in the environment (at the level of populations and single cells); (ii) approaches to cultivating currently uncultivable species; and (iii) methods for rapid determination of key physiological traits and activities. Establish mechanisms to encourage the necessary instrument development.
- Develop Genome Resource Centers to advance microbial genomic science, including a user-oriented approach for sharing microarray technologies, data analysis, and proteomic analysis. Smaller scale resources, such as primers for open reading frame (ORF) sequences, probes for genes, etc., could be made available through the American Type Culture collection or a similar organization.
- Conduct more extensive surveys of microbial diversity using 16S rRNA genes and other molecular methods to reveal how community structure is patterned with environment.
- Develop more training opportunities for students, including such disciplines as ecology, evolution, genomics, bioinformatics, and computational sciences. Cross-enrollment programs, summer internships, and virtual courses could be mechanisms to hone skills in emerging and rapidly changing fields.

Introduction—Genomics Sets the Stage

Microorganisms generally live as part of complex communities. Although community is the environment in which microbes (and their genomes) evolved, continue to evolve, and function, a comprehensive understanding of community structure and activity has been beyond the reach of traditional microbiology. The human genome project has set a benchmark for biological system complexity, yet most microbial communities are far more complex. The environmental playing field is immense. Not only is microbial species diversity vast, and mostly undefined, but also natural systems are highly dynamic, their population structure and activities shifting with changing environmental circumstance. In the past, comprehensive studies of environmental microbial systems simply have been beyond the reach of conventional tools. Today we see that rapidly evolving technology is changing the scope of research questions amenable to study. The powerful tools of modern high-throughput genomic and post-genomic technology (*e.g.*, functional genomics, expression and activity analyses, proteomics) are having a profound effect on our science. Just as knowledge of the human genome promises to revolutionize medical science, the application of genomic technologies to microbial evolution and environmental biology promises to revolutionize microbial biology.

Even though the human genome is nearly completely sequenced, the end of the genome sequencing era is not in sight. Thousands of plant and animal species remain to be sequenced. Many insect species remain unknown. But such tasks pale in comparison to the task of performing a quantitative estimate of global microbial gene content. There may be 30,000 human genes, but estimates suggest there are in excess of 10^{10} globally distributed bacterial genes! Enumerating and characterizing this genetic landscape is an immense challenge. Even were we to elucidate the genomes of the 5,000 described species of microbes, this alone is a daunting task, as there are about 2,000–4,000 genes per organism. Add to this that we are able to cultivate only about one percent of all known bacterial species, rendering studies of the genomes of the noncultivatable species more difficult. Thus, we have a very long way to go before we have exhausted the value of sequencing Earth's biota.

The rewards of carrying out such work are not just increased knowledge, but also economic ones. These microorganisms and their communities provide key nutrients for macrobiotic organisms and sustain essential nutrient cycling. Also, many of these microorganisms are mutualists that carry out essential biochemical functions for their hosts. For example, we will have a better understanding of human and animal illnesses arising from presumed ecological disturbances; the function of microorganisms in agriculture; and how microbes develop antibiotic resistance. They also have great use in commercial products, such as producing food and beverages, enzymes, biopolymers, or drugs and vaccines. And, of course, sequencing the microbial world will result in discovery of hitherto unknown microorganisms and hitherto unknown functions of microorganisms. Both can lead to new commercial products. For the money invested, we get much more information about the evolution of life on our planet by sequencing bacterial genomes than we do the genomes of higher organisms, especially about the origins of metabolic innovation and the biochemical underpinnings of our biosphere. Microorganisms have been powerful geological and evolutionary agents over the course of our planet's history. They created the oxygen in the atmosphere and deposited a significant fraction of the fossil fuels that we

exploit today. Their biochemical pathways provide the basic design for metabolism in all life forms.

Why Now?

Extraordinary advances in molecular technology have fostered an explosion in genomics and bioinformatics. These events have led to a renaissance in microbial biology, especially prokaryotic microbial ecology, opening a virtually limitless frontier of scientific opportunities that await the integration of modern-day genomics into contemporary microbial systematics, evolution, and ecology.

This report outlines a framework for the integration of genomic technologies with technologies that can elucidate microbial physiology and evolutionary biology. We will highlight the areas in which integration is most critical and will further the revolutionary advances genomics has provided and will provide in both the fundamental and applied aspects of microbial ecology. Not only must we increase our knowledge using available technologies, but we also need to develop new technologies that will expand information on microbial (prokaryal, fungal, and protistal) evolution and ecology to encompass entire communities, ecosystems, biomes, and the biosphere. We also need significant public outreach to communicate our findings and their value to scientists in other fields and to the general public.

Integration of genomics with evolutionary biology will bring forward two of the most fundamental questions in biology: What is life and where did we come from? We will learn much more about "what is life" from the genomic information from the inventory and identification of all of the genes and gene products of a free-living cell—their functions and their interactions. With complete genomic sequences, we can undertake to learn what every gene product does and how such products interact in a concerted fashion to maintain and reproduce a living cell. This task is now well under way for model microorganisms, but completing the picture will require information from new, focused biological experimentation, both for the model systems and for a sizable roster of important microorganisms. Genomics will help us to formulate the questions that still need to be answered experimentally. It also will allow new avenues of exploration, such as defining what minimal sets of genes are adequate to sustain life under various defined biological conditions.

We now are in a position to address questions that we were previously unable to imagine. The answer to the question "Where did we come from?" will be informed by detailed evolutionary reconstruction of the tree of life, facilitated by rapidly accumulating genomic data. We can deduce the paths of evolution from ancient ancestors by analysis of contemporary genes and gene products. Comparative sequence analysis of suitable proteins will allow us to trace life back to its origins, to the theoretical last common ancestral cell and perhaps even before, to a time of generation of protein families. Some protein families of today are related to their ancestral antecedents. Others are separated from the vertical path by horizontal transfer, losses of genes, changes in cellular function of homologous gene products in different organisms, and the like. By identifying and excluding these exceptional cases, truly homologous genetic trees generated from the corresponding protein sequences will be useful in generating valid phylogenies. The resulting tapestry of a tree of life with some network properties will be complex, but it will reflect the patterns and mechanisms of evolution that have given rise to all forms of life today. That will provide an important part of the picture of where we came from.

Did You Know?
There are more microbes than any other group of organisms. Microbes have been on Earth for billions of years. Consider that the Earth is about 4.5 billion years old, and you realize that the Earth has been populated by microbes for most of its existence, while vertebrates have only been on Earth for a relatively short time. Microbes are essential for maintaining our planet's quality of life: they produce greenhouse gases; they can be used to create "green," eco-friendly chemistry; and they have a major role in countering the events leading to global climate change. Greenhouse gases are curbed by microbial processes. For example, as oceans warm they can release methane from hydrate melting that would have a severe impact on life. This methane release could possibly be mitigated by microbial action. If we understand the basis of how the world's ecosystem works—the role of gene shuffling, the diversity of genomic information and its biological meaning, the effect of environmental disturbances on genetically based processes, the function of green chemistry—we can have a better idea of how human activities will affect life.

Why Is This a Changing View of Microbial Biology?

Standard taxonomic methods do not usually tell us the necessary story about microbial evolution and relationships. Initial selections of microbes for genomic sequencing have not been done in an organized fashion; rather it reflected phylogenetic disparity in a search for genetic novelty. At this point, only an important but limited set of microbial genomes have been sequenced. One amazing lesson from this exercise is the plasticity of genomes from organisms adjudged closely related by standard taxonomic yardsticks. Yet, within these closely related species with different genomic patterns there are recognizable genetic signatures that testify to the close evolutionary history of these organisms. Clearly, we are naïve in our predictions of evolutionary closeness and what makes these organisms related; we need to sample more broadly. In fact, we need to sample the entire spectrum of microorganisms, from pathogens to commensals, to parasites, to free-living organisms that make up microbial consortia.

An example that surprised us was the genomic information on two strains of *Escherichia coli*—K-12 and the pathogenic H0157. There are substantial genomic differences between these two organisms. About one-third of the genes are different, scattered throughout the genomes in regions of unique sequence. About 10% of the H0157 genome arose recently by acquisition. In contrast, strain K12 lost genetic information. We have to ask ourselves, "What is the role of the donated sequences? What organisms are the donors? What are the mechanisms and context of transfer?"

We need more than genomic evidence to address these questions. Even for *E. coli*, on which there are 100 papers published per month, 17% of the genes have functions that are unknown. Understanding of the consequences of gene exchange demands a better understanding of their functional significance. Part of our problem is lack of technologies and techniques to study function. Microbiologists have primarily relied on very gross, culture-based measurements of phenotypic differences among microorganisms, mostly determined by investigator insight. We need to be able to study function on an operon or regulon level to demystify the link between gene and function. New techniques, for example, may allow us to identify protein function (*e.g.*, amino acid biosynthesis) for ORFs of unknown function.

Also, it is evident that these questions require population studies of genomes. At this time we have only fragmentary information needed to conceptualize this information. Notably, not all species so far characterized show such tremendous genomic plasticity as observed for *E. coli*. The various strains of *Chlamydia* do not show this variety, nor do the strains of *Buchnera*, an endosymbiont that lives within the guts of aphids. How many strains must we sequence to observe and interpret these kinds of patterns? How does organism lifestyle influence genomic plasticity?

Recommendation: We need to sequence more related strains, such as different *E. coli* isolates, to begin to see the genomic patterns that are not obvious from using standard phylogenetic techniques.

Initial findings from the comparative genomics of two strains of *Prochlorococcus* (a cyanobacterium that numerically dominates the oceans) are illustrative of the problems involved in determining relationships.

The two strains (which differ by roughly two percent at the rRNA locus) have significantly different physiologies (phenotypes). One, the high-light-adapted strain, grows optimally at light intensities at which the other, the low-light-adapted strain, is photoinhibited. The low-light strain can grow at extremely low light intensities, which will not sustain the growth of the high-light strain. Representatives of these two strains have been shown to preferentially dominate different layers of the ocean: the high-light-adapted at the surface and the low-light-adapted in deeper waters.

Analysis of the complete genomes of these strains reveals significant differences, some of which are very consistent with the gross physiological properties observed. First, the genomes are significantly different in size. The low-light strain has 2,300 genes (900 of which are not found in its cousin) and the high-light strain has 1,700 genes (300 of which are unique). Neither strain can use nitrate as a nitrogen source, which is consistent with the complete absence of a nitrate reductase gene in both genomes. The low-light strain, however, can grow on nitrite and has the nitrite reductase gene. This distribution of genes is consistent with the vertical distribution of the organisms in the ocean, as the low-light strain can often be found dominating the nitrite-maximum layer in the deeper waters, and the high-light strain dominates the surface waters, where regenerated ammonia is the dominant source of nitrogen. In keeping with this genotype/phenotype congruence is the observation that some of the genes that are unique to the high-light strain are involved in ultraviolet damage repair and high-light-induced proteins that have been found in closely related organisms.

Horizontal Gene Transfer

The analysis of complete microbial genomes has made it clear that horizontal gene transfer—transfer of genes from one microbial species to another—has been an important factor determining the content of microbial genomes, but the extent of its impact on evolution remains to be determined. While it is clear that the dominant mode of inheritance is vertical (*i.e.*, from ancestor to descendant), most microbial genomes show a number of genes that could only have been acquired by horizontal gene transfer. It remains unclear, however, whether all genes are equally likely to undergo transfer and whether all microbial lineages participate equally in gene transfer. Several mechanisms for gene transfer have been identified, but the relative importance of these processes in nature, and the ways in which the mediating process influences the characteristics of gene transfer, remain poorly understood. Because of the evident importance of gene transfer in microbial evolution, it is imperative to determine the factors that influence the frequency and character of gene transfer, to characterize their role under natural conditions, and to understand the ways in which horizontal gene transfer influences microbial ecology and evolution.

How do we develop a detailed understanding of gene transfer in the environment? Could high density DNA microarrays allow us to monitor gene content and expression in natural settings? By linking expression with environmental and community processes we should learn more about the 47% of genes for which the function is currently unknown.

The Role of Genomics in Microbial Evolution and Ecology

Although we can now rapidly sequence the genomes of microbes, it is still too early to know the many impacts of this genomic knowledge. However, there is a sea change occurring in microbial ecology and evolution as these fields begin to respond to the genomics revolution. The following sections elaborate on some of the most immediate of opportunities. In particular, we see genome science serving a major role in drawing together the disciplines of evolution and ecology.

Scientists who wish to understand the origins of the major, deep-branching phyla of Bacteria, Archaea, and Eukarya should have the opportunity to analyze the genomes of representatives of all the major lineages. The hope of such an approach is that enough information still remains in the contemporary microbial lineages of these three domains so that it will be possible to reconstruct the evolutionary steps that led to their origin and diversification from the last common ancestor. If successful, it may be possible to infer the phenotypic features of the Last Common Ancestor and its evolved descendants. Indeed, it may be possible to reconstruct the genomes of representatives of presumed ancient lineages to study their phenotypes.

Role of Genomics in Understanding Evolution

Microbial evolution provides at least two major challenges. First, it has occurred over more than 3.8 billion years. Second, horizontal gene transfer complicates the understanding of lineages and phylogenies. The resolution of early evolutionary events and an understanding of the role that horizontal gene transfer has played in shaping the evolutionary process will greatly benefit from genomic information.

Phylogenetic trees, at this time, are primarily based on rRNA similarities among organisms. Genomic data will allow the informed choice of alternative molecules for more fine-tuned phylogenetic reconstructions. Microbial diversity and history are dominated by the evolution of novelties in biochemical pathways, thus leading to clearer links among genomics, proteomics, and evolution. While traditional evolutionary frameworks may serve as starting points, models are likely to range from those that are population genetics-based to ones that are gene-geneology-based. Using these models as frameworks for investigation, genomics will provide the information to develop an effective understanding of, and tools for, studying microbial evolutionary biology.

The smaller genome sizes of microbes, especially of prokaryotes, provide exciting opportunities to link environmental function, physiology, and evolutionary biology much more closely than is possible in the Metaphyta and Metazoa. The synthesis of genomics and evolutionary biology will advance our knowledge of life and how it arose. It will aid us in deducing minimal genomes, which are of interest both scientifically and for industrial purposes, because minimum-gene organisms may provide more effective tools for biotechnology and gene product manufacturing. It will assist in quantifying the extent of horizontal gene transfer and determining the character of ecological and physiological factors that affect it. It will help us determine the history of events that produced modern organisms and will yield an understanding of symbiotic relationships and how they have contributed to metazoan and metaphytan evolution.

The sequence of the human genome has reinforced what scientists have long suspected: humans are not as distant from their microbial ancestors as most had believed. Increased knowledge of microbial genomics and its role in determining evolutionary biology may benefit human health and welfare by providing an understanding of DNA repair and genetic mechanisms of cancer, the emergence of variation (with implications for emerging diseases), and of the etiology of antibiotic resistance.

The original notion that *E. coli* would serve as a general model for life is now recognized as insufficient. However, the utility of model organisms can't be denied and genomics should provide a new framework for identifying the next generation of "model" organisms. For example, the "tree of life" derived from rRNA sequences has suggested a new category for model organisms, the Archaea, because of their closeness to the eukaryotic lineage. DNA repair in the Archaea could well provide insights into the defects in repair that may have led to mutations resulting in human cancers. A new look at the Archaea as a model for eukaryotes will potentially be rewarding.

Another critically important area of evolutionary studies is the study of the evolution of microbial pathogens. Currently, scientists are only able to react to newly discovered pathogens as they appear. Is it possible that we could predict what effect human-initiated changes, such as pollution and global warming, are likely to have on future pathogen load? A better understanding of how newly emerging bacterial pathogens develop could help us move in that direction.

Microbial resistance to the antimicrobial agents that once controlled them is a chilling recent development. The public has witnessed failures by scientists to find cures, which they find upsetting and troublesome; the loss of current cures would be a new and frightening phenomenon that could conceivably undermine public confidence in the medical establishment. The evolution of antimicrobial-resistant microbes, from *Staphylococcus aureus* to *Plasmodium falciparum*, needs to be understood for two reasons. First, it may be possible to slow or stop the increases in resistance by timely interventions were we to know what factors drive this particular aspect of microbial evolution. Second, it is still not clear how antimicrobial resistance evolves. In many cases, housekeeping genes appear to be recruited for this new function. Understanding how this occurs could provide new and important insights into how microbes adapt to the many new challenges and opportunities provided by human manipulation of their environment.

Genomic Information Will Refine Models of Evolution

The evolutionary synthesis provides a common framework for considering evolutionary processes. Although the synthesis was developed primarily in consideration of macro-organismal evolutionary issues, we believe that the same processes apply in microbial evolution. Further, we expect that the primary distinction between macro- and microevolutionary processes will relate to the following issues:

- *The rate of microbial evolution.* The short generation times of many microbes suggest that microbes may evolve, on average, more rapidly than do macroorganisms. There is some evidence that suggests more rapid rates in some circumstances. Microbes experience their environment more directly and respond to environmental perturbations more rapidly than do macroorganisms. Rapid responses to selection result in rapid changes in gene frequencies and, again, more rapid evolutionary rates.

- *Microbes have much larger population sizes than do macroorganisms.* Evolutionary forces that are sensitive to effective population size, for example, random genetic drift, may play a less dominant role in microorganismal evolutionary dynamics than is the case for macroorganisms.

- *Microbes possess a sophisticated set of genetic exchange mechanisms.* Many of these mechanisms are specific to the microbial world, with no macrobiology counterparts. This enhanced ability to transfer information may result in increased evolutionary rates.

Genomic information provides a window into the patterns of microbial evolution and will allow us to refine and extend current evolutionary models. Only with extensive comparative and population-level genomic information will we be able to assess the relative importance of various evolutionary forces in microbial evolution and determine whether common evolutionary models apply across biological diversity.

Factors Driving Speciation and Extinction

The environment of early Earth provided a variety of energy sources for microbial life, including organic compounds as well as reduced inorganic compounds such as hydrogen gas ammonium and ferrous iron. Based on our current understanding of microbial diversity, it seems likely that the availability of these potential energy sources provided niches for the speciation process. Likewise, today speciation processes must be driven also in part by the availability of new niches in which novel energy sources become available. Examples of such energy sources include new organic compounds originating from plant and animal speciation events—resulting in unique niches for new species, including loosely and tightly associated symbionts. Thus, the origins of bacterial symbionts, as revealed by genomic sequences, may provide an important perspective for understanding the speciation process.

Interestingly, we do not know what effect major disturbances, such as the early asteroid bombardment period, have had on the evolution of bacteria. Clearly, bolide impacts have had profound effects on higher organisms. Even the rather recent K-T impact, which was small by cosmic standards, resulted in the extinction of 90% of animal and plant species. Presumably, the microbial symbionts of these species were also driven to extinction.

Microbial extinction events must be occurring continuously as organisms adapt to the changes in the communities in which they

live. Presumably, the rate of extinction will depend on the rapidity with which niches in the community change. The most readily documented extinction events may occur with the symbionts of plant and animal species that are threatened with extinction, so that these tightly coupled systems may be the most accessible candidates in which to study microbial extinction events.

The impetus to study microbial extinction is the need to understand its more general ecological and evolutionary significance. However, the process of microbial extinction now lacks clear paradigms. What is the definition of extinction among microbes? Is it comparable to extinction among macro-organismal forms or is the loss of a gene or metabolic pathway the more relevant extinction event? Genes and pathways may be preserved in microbial systems via horizontal gene transfer, even though the original host becomes extinct. This scenario raises many additional questions about selective advantage conferred by transfer, retention of original function, and the possibility for continued expansion of genomic complexity in the absence of expansion in the numbers of microbial species. We suggest that the study of extinction among microbes is uniquely dependent upon advances in genome science.

Ecology and Ecogenomics

Microbes represent possibly the largest component of biodiversity, but lack of adequate tools for quantifying diversity, in either model or natural communities, has hindered progress in microbial ecology, especially microbial community ecology. Genomics and modern molecular methods have dramatically altered this situation and will enable community ecology to address questions that are important scientifically and result in a better understanding of environmental issues. The integration of genomics with ecology will provide dramatic advances in all areas of microbial ecology.

One of the greatest challenges of microbial ecology is the spatial and temporal heterogeneity of communities and their environments. The equipment, supplies necessary for, and time costs of current study methods have limited studies to single sets of conditions, generally conducted at one point in time. From the level of the individual through the level of the community, genomics provides a means for quantifying diversity and activity over a range of conditions. New studies should include chemostats run at a range of conditions and thorough examinations, over reasonable spatial and temporal scales, of localized, specific communities.

Several areas represent important first steps in the continuing revolution of microbial ecology. These include understanding which organisms are present in a given community, what they are doing, and how they interact. This can be studied *in vitro* by using model communities or *in situ* using natural communities. However, it is essential that model communities be constructed in consideration of the natural world. Thus, biotic surveys and inventories are critical beginning points in ecological studies. From such lists, microbial ecologists will be able to search for patterns in microbial biodiversity or the distribution and abundance of species, however they may be defined. Basic information, such as the distribution and abundance of microbial biodiversity with respect to latitude, altitude, among biomes, and over many other ecological gradients, is important. It is important, not only for bioprospecting for organisms that can produce products or drugs we can utilize, but for comparison with patterns of distribution and abundance of plant and animal species.

Integration of genomics and ecology will help us predict the effects of environmental perturbation on microbial communities. Such perturbations include the major features of global change, including modifications of biogeochemical processes (*e.g.*, enhanced levels of CO_2 and doubled rates of N deposition), land use change, acidification, desertification, ozone depletion, and climate change.

This process will also add to understanding the relationship between microbial biodiversity and biogeochemical processes. The majority of biogeochemical or ecosystem processes, such as decomposition, nutrient cycling, and carbon cycling, are governed by microbial processes. These processes determine soil and water quality, atmospheric composition, and serve as the basis for most ecosystem services. The role of the extraordinary diversity of microbial communities identified in these processes is virtually unknown. Is the majority of this diversity redundant, or does such diversity play an important role in the stability or reliability of biogeochemical processes? Community ecology has made substantial progress in studying these questions, but the emphasis has been exclusively on plants and animals—not microbes.

It is essential to understand community structure and community dynamics in microbial communities. The fundamental structure (pattern of interactions) among species in communities provides important information about the complexity of microbial communities. Microbial genomics can be used to identify structure and complexity in natural and managed (*e.g.*, wastewater treatment reactors) microbial communities by identifying key trophic groups. Are microbial communities dominated by positive or facilitative interactions, or are antagonistic interactions (*e.g.*, interspecific competition for limiting resources) the dominant features of microbial community structures? Such data may provide information about sensitivity of microbial communities to perturbation or insights into the role of community complexity in community dynamics.

The benefits of focusing our understanding of microbial genomics on ecological factors are economic, as well as social. It could provide for new tools for bioremediation and biorestoration. It may aid in managing ecosystems and biogeochemical ecosystem services and in predicting global change through microbially based mechanisms. Such research may also result in an improved understanding of disease dynamics (old and emerging) and invasions.

By using sequencing information to ask questions about microbial processes, we may be able to propose answers to broad ecological, population biology, and comparative microbiology questions. But first we need to learn to ask the questions before we can propose answers. Among the questions that must be asked is how to measure disturbance or stability of ecosystems. Can we integrate organismal/functional arrays for physical/chemical environmental process measurements *in situ*? We need to understand how the environment (or host) responds to microbial community composition and activity. This will aid us in gaining predictive capabilities from these response patterns.

Case Study – Example of Environment in Which Genomics Has Been Useful and Could Be Far More Useful

At least two deep terrestrial subsurface environments in which microbial communities are probably independent of photosynthesis and driven by hydrogen-oxidizing chemoautotrophic bacteria have been discovered. One of these is deep ground water, accessible through deep (3.6 km below land surface) South African gold mines. The water is under high pressure, at approximately 60°C, and contains H_2 as a possible energy source for autotrophic metabolism. Modern methods for extraction of DNA, followed by PCR amplification and cloning and sequencing of 16S rRNA genes, have provided a preliminary assessment of this microbial community. This information was used to devise strategies to culture microbes at this site. Some of the clone groups were closely related to organisms with physiological characteristics likely to influence the geochemistry of the ground water (*e.g.*, sulfate reducers). However, five clone groups were very-deep-branching organisms that could represent new phyla within the Eubacteria. Although the latter results are very interesting, they provide no clue as to how such organisms could be cultured.

Genomic approaches (community genomics) could be very useful in elucidating the functioning of this community. Information on the types of metabolic capabilities that are present would most likely provide clues on how to culture microbes from this unusual environment (e.g., by identifying degradative pathways, etc.). Such information might also be analyzed to find possible dependencies between different species (*e.g.*, it may show that one species requires a metabolic product of another as an energy/carbon source or growth factor), which could also explain why some organisms cannot be cultured. Knowing the metabolic capabilities that are present within the microbial community would help elucidate the energy sources and mechanisms by which nutrients are cycled among microbes that are not dependent on photosynthesis for primary production.

Genomics is also the emerging tool to increase our knowledge of previously known, prominent but so-far-uncultured lineages such as OP11, SAR86, or the marine Archaea groups I and II. By retrieving large chromosomal DNA fragments directly from the environment, we can learn much more about the genetic potential of these "new" groups. However, we note that this will not substitute for the need to cultivate the respective organisms.

The integration of traditional studies of physiology and genetics with modern genomics presents us with unprecedented opportunities for making major advances in the study of microbial evolution and understanding how microbes provide the foundation of our biosphere. Despite the central role that microbes play in biotic processes, we know little about their actual diversity or how they regulate key biogeochemical cycles. Process-oriented studies, in which the microbial world is treated as a "black box," alert us to interactions with macrobiotic forms and the formation of feedback loops between biogeochemical gradients and structured microbial populations. Yet, there are no comprehensive descriptions of underlying biochemical and genetic mechanisms that govern these processes.

By taking advantage of high-throughput genomics technology, we can now completely or almost completely determine the microbial species composition of an ecosystem (at least a simple one) and employ gene expression profiling techniques to link process-oriented observations with specific metabolic activities. An entire microbial community might be treated as a complex, mixed genome. Using this strategy, investigators could link biogeochemical measurements to metabolic processes, community-wide gene expression patterns, and microbial population structures in natural settings. Perhaps the most significant result of such holistic studies would be a definition of the factors that sustain the remarkable diversity of microorganisms.

Recommendation: Through ecogenomics (the intersection of genome science and ecosystem interactions), we can explore many levels of biological organization.

Endosymbionts

A similar experimental approach can be applied to studies that target how microbes affect the diversity and distribution of macrobiotic forms, including the interactions between microbes and more complex multicellular organisms. Bacterial endosymbionts are fundamental to the survival of most multicellular species and, therefore, are among the most ecologically significant microbes on Earth. Without bacterial symbionts, most animals would not survive. These genome-genome interactions are intrinsically less complicated than those in diverse microbial communities, but their impact upon ecosystem structure and productivity is equally important. Like most of the microbial world, many obligate symbionts are not cultivable and, thus, not tractable by traditional genetic and microbiological approaches. We know little about their specific functional roles or patterns of evolution. The same technology and experimental approach for studies of complex microbial environments will be applied to studies of bacterial symbionts of eukaryotic unicellular and multicellular species. Similar to microbes in complex communities, symbionts perform key biochemical reactions that mediate dynamic processes. Some obligate bacterial mutualists mediate a key process in most terrestrial ecosystems: the consumption of plants by insects. These mutualists carry out essential biochemical reactions for their insect hosts (*e.g.*, the biosynthesis of essential amino acids or the recycling of uric acid nitrogen) and supply nutrients that are lacking in an unbalanced plant diet. With the powerful tools of genomic science, bacterial enzymes that catalyze such key biochemical reactions of functional importance can be identified.

Goal: A balance of resources will be required for large-scale DNA sequencing, postgenomic expression studies, underlying bioinformatics, and elucidating metabolic functions. To maximize use of this information, as well as to develop a more complete understanding of microbial evolution, we also need to develop comprehensive ecological databases that can supply us with information about the ecological context of molecular, physiological, and genetic data.

How Can Genomics and Environmental Biology Inform Us about Emerging Infectious Diseases?

Over the past two decades, population biology and evolutionary biology have contributed fundamental insights into the nature of infectious diseases. Molecular phylogeny studies based upon comparisons of rRNAs have identified the phylogenetic distribution of most important human pathogens and provided the database for design of diagnostic probes. These same studies provide rationales for selecting optimal model systems for studying diseases caused by microbial parasites. For example, *Pneumocystis*, which infects nearly 90% of all immunocompromised patients (*e.g.*, AIDS patients), was widely described as a protozoan, probably related to the malarial parasites. Molecular studies demonstrated that this pathogen actually is an uncultured member of the fungi. This finding has important implications for selection of model systems for exploring the biology of human pathogens and suggests that new drug treatment should include new-generation antifungal drugs. Molecular phylogenetic information provides a powerful means to explore the ecological context of emerging pathogens.

Studies have been carried out on pathogenic mycobacteria, which may have arisen independently three times. In the search for reasons for drug resistance and the genes responsible for pathogenesis, several mycobacterial genomes have been fully sequenced. Binary comparisons with pathogenic strains have searched for differences that might underlie virulence, but were compared to the genome only from a nonpathogenic mycobacterial strain. This approach is insufficient. The use of several nonpathogenic genomes in simultaneous comparisons of all available pathogenic and nonpathogenic sequences will simplify the search required to identify potential therapeutic targets. Hence, inference of the genetic basis of key phenotypic traits will require multiple genome sequences.

Systematics – Linkages with Ecology and Evolution

How will genomics impact taxonomic classification? The American Society for Microbiology (ASM) Systematics Division early responded to the revolution in genomics by renaming itself the Division of Evolutionary and Genomic Microbiology. The renaming of this division is a clear signal of a changing role of systematics in microbial biology. The genomics revolution has placed systematics at the interface between microbial evolution and microbial ecology. This will infuse systematics with newfound excitement and reemphasize its central position in microbial biology. In turn, the science of systematics should play a key role in bringing essential order to genome science.

For the life scientist, taxonomy traditionally provided a common language for classifying and describing organisms, and this was based on the presence or absence of specified phenotypic characteristics. Phylogeny is a classification system that attempts to reflect the evolutionary history of the considered organisms. As new ideas and revolutionary technology were introduced, novel phenotypic characteristics were recognized and that has frequently led to shifts in both taxonomic schemes and phylogenetic reconstructions of evolutionary history. In this sense, taxonomy and phylogeny driven by technology are dynamic descriptions. Nowhere in biology is this more evident than in the microbial world, in which the richness of directly observable characteristics is modest in comparison to that of multicellular organisms.

Today, traditional taxonomic concepts (*i.e.*, species, genus, family) do not serve microbial systematics, in which problems of horizontal gene transfer and mechanisms of speciation and evolution are varied and complex. A new framework for taxonomy, one better adapted to genomic information and microbial taxa, needs to be derived. Currently, the strain is the unit of taxonomy and research, but this must be refined, not only for more effective ecological and evolutionary research, but also for better storage and retrieval of cultures, genomic data, and identification of organisms in samples, patenting, and bioprospecting. For both scientific and practical reasons, we need to develop a new framework for taxonomy, one that fully embraces the new insights of genome science.

The Relevance of Classical Taxonomy of the Postgenomic Era

Classical taxonomy, through the adoption of an agreed-upon code of nomenclature, provides a means of summarizing the features of any named organism or group of organisms with a set of simple descriptors (*e.g.*, Latinate binomials). These names provide indices into the scientific and patent literature as well as the sequence databases.

With the introduction of molecular evolution techniques and the more recent sequence determinations of entire genomes, traditional taxonomic descriptions have become even more tenuous. Yet, taxonomy is far more than an outdated means of classification. It provides a common language for describing microbial forms in the context of a rich literature about their physiology, metabolism, and life history. Molecular phylogenetics has forced us to reevaluate how organisms are related without requiring us to discard traditional taxonomic views. Not only would the elimination of traditional taxonomic descriptions lead to the potential loss of information about species richness, but the substitution of new taxonomic descriptions based upon molecular phylogenies would not arrive at a single satisfactory description of how organisms are related to each other. If we have learned no other lesson from molecular phylogeny, the lesson learned should be that different genes and different inference techniques do not necessarily describe congruent evolutionary relationships. Yet, we do not consider this to be an insurmountable problem.

An important feature of sequence-based classification schemes is that they provide a universally applicable and cost-effective method, eliminating much of the ambiguity arising from earlier systems. In addition to providing information about evolution and phylogenetic relationship, sequences will ultimately be mapped to specific phenotypic and ecological characteristics of an organism. To facilitate this mapping, it is essential that genomic research is complemented with research designed to more fully elaborate nongenomic information. We must begin with the question of how current phenotypic and ecological information maps to genomic information and then ask, has our strategy for genome selection furthered this mapping? We suggest that these considerations must be paramount in continued genomic studies, directing the selection of organisms and defining scope of coverage, both related and unrelated taxa. In order to achieve this ultimate goal, and also to enable decision making, it is essential to make the data more useful and accessible, to both systematists and end users of any resulting classification(s) (*e.g.*, clinical microbiologists, regulatory agencies, ecologists). A more generalized structure for describing and linking phenotypic, ecological, genomic data is essential. Using modern, centralized databases, it is possible to accommodate multiple phylogenies and taxonomies for collections of microorganisms. We embrace a technological solution that allows us to view rapidly changing phylogenies in the context of traditional taxonomies. This ensures the retention of descriptive information in the literature for organisms that are assigned new taxonomic status.

Recommendation: A minimal set of data descriptors that can be embedded into a database and be used in the generation of content for distribution through XML-enabled browsers so that the information is more readily available and more easily maintained as taxonomic rearrangements occur should be defined. These databases must also incorporate relevant ecological and environmental information.

Recommendation: Genome studies need to sample both closely and distantly related taxa.

Recommendation: New methods need to be explored to cultivate uncultivated microbes, and new methods are needed to provide rapid, systematic analyses of existing and newly isolated microorganisms.

Recommendation: Culture collections should be expanded, and a central, well-ordered facility should be put into place for the maintenance of type strains.

The Role of Genomics in Higher Order Classification

The framework of prokaryotic systematics is now based on 16S rRNA molecular phylogeny. The comparative analysis of 16S rRNA was, and still is, very useful in revealing the higher taxonomic structure within the microbial world. These data have been used to generate the outlines of the domains Archaea and Bacteria and enable scientists to determine the evolutionary relationships to eukaryotic species (Domain Eukarya). Within each prokaryotic domain, numerous deeply branching lines of descent reveal the enormous phylogenetic diversity of prokaryotic organisms. A natural domain-to-strain hierarchy appears to exist. Further subdivisions lead to recognition of phylogenetically coherent groupings, many of which are described as classes, orders, and families. For most of these taxa, phenotypic properties are included in the description of the respective taxon. Certain phyla, such as the *Spirochaeta* or the *Cyanobacteria*, have a common phenotype: a defined cellular ultrastructure in one case and photoautotrophy in the other case. Groups of an intermediate taxonomic level, such as the ammonia-oxidizing chemolithoautotrophic beta-subclass *Proteobacteria*, may also be highly coherent; all members share a distinct physiology that is key to their ecological function. There are, however, physiologically very diverse rRNA-defined groups, such as the *Nitrobacter*-branch of the alpha-subclass of *Proteobacteria*. For groups like the latter, of which there are many, genomics may be key in defining whether they indeed share a distinct common gene pool and in explaining the sometimes huge phenotypic differences.

The general outline, emerging from the 16S rRNA gene sequence analyses, is generally supported by the analyses of sequences of other genes. We expect that comparative analyses of whole genome sequences will refine this framework by better resolving the branching points to determine the evolution of lineage-specific gene sets and to facilitate the molecular identification of members of taxa between the ranks of domains and families.

Genomics will Inform the Species Concept

In contrast to defining higher taxonomic ranks, bacterial systematics has taken a pragmatic, polyphasic approach to describing taxa at the ranks of genus and species. These two ranks are the most important ones in microbiology. Without their proper descriptive identification, classification will fail. As now defined, a species constitutes an assemblage of strains that share high genomic similarities as well as a range of common physiological and cultural properties. Different species within a genus are currently defined as sharing approximately 70% DNA-DNA similarities among their members, as well as differences in physiological properties, and, if applicable, in their ability to cause disease. We anticipate that genome science will provide the foundation for a rational basis for having a species (genes, gene clusters, and structural features that, at the molecular level, will demonstrate the distinctness of the hitherto pragmatically defined taxon). One possible resolution will be based on an assessment of gene flow between and among species. Ultimately, genome science will assist in resolving one of the most contentious questions in microbial biology - does the natural entity of "species" exist in the microbial world?

Macro-organismal diversity is discrete, rather than continuous. Available studies also suggest that microbial diversity may also be discrete as reflected by the existence of distinct gene pools, which, by definition, makes up a species. These gene pools evolve in response to a particular set of evolutionary and ecological pressures. We propose that, just as in macro-organisms, a workable microbial species definition must be informed, if not dictated, by the presence of discrete gene pools. Such a genetically based species definition predicts that the levels of gene flow between members of a species (or within a gene pool) will be greater than the level of gene flow between species.

We Can Apply Genomic Information to Directly Address Level of Gene Flow Between Members of a Species

Recommendation: We can obtain genomic-level information from a population sample (perhaps 10 isolated genomes) for a sample of approximately 10 species within a "well-defined" species cluster. These genomic data will answer the question of how separate are the gene pools of our putative species. It will resolve the levels of "within" versus "between" genetic mixing or communication. It may also inform us as to what is the minimum genome content that distinguishes each of these gene pools, or species, and, further, what unites them as a species cluster.

Ideally, this same approach could be used for clusters of species that we would not predict to conform to such a species definition. For example, biochemical and molecular data suggest that members of the Enterobacteriaceae may conform closely to our proposed species definition, while Streptomyces species may not. Does that mean a single species concept will not apply? Comparative genomics that incorporate population-level sampling will allow us to resolve these apparent distinctions and refine our microbial species definition.

The Synthesis of Genomics and Systematics: Key to Understanding the Process of Diversification (Speciation)

Systematics has long been regarded by many scientists as a dry subject that has little to do with the really exciting areas of research. This perception is about to change with the marriage of systematics and genome science. The combination will provide the fullest view that has yet been possible of the essential features of microorganisms. This synthesis will provide essential insight into the evolutionary dynamic: what are the factors that drive bacterial diversification (speciation)? How far does an organism have to vary to become a new species? Comparative genomics, focused at the proper evolutionary time scales, will provide a fundamentally new set of data with which to begin to address this complex issue.

This new set of insights will make possible a new era of predictive biology. For example, over the past few decades, scientists have witnessed a bewildering succession of emerging infectious diseases, most of which have arisen as a result of changes in human habits and habitats that have created new niches for bacterial pathogens to colonize. These human alterations range from surgery to air conditioning. As the characteristics of bacterial pathogens come into better focus, it may well become possible to predict what environmental and ecological factors that drive the emergence of new species. It will provide a foundation to predict the habitat range and activities of characterized organisms—for example, whether a particular bacterium is likely to be able to cause disease in humans, animals, or plants.

An essential aspect of this higher level understanding will derive from maintaining a connection to ecology and promoting an appreciation for the importance of studying the natural history of microorganisms. In this regard, it is possible that *E. coli*, as a commensal bacterium, has co-evolved with its mammalian hosts. If so, there is no evidence for this from the current literature, although it is known that specific serotype groups exist. We also know that horizontal gene transfer has played a role in determining the genetic composition of strains. Therefore, if *E. coli* was selected as a model system of study, the selection of the strains for sequencing is extremely important. From which mammalian hosts should they be isolated? Only primate hosts or should hosts from other mammalian orders (*e.g.*, Rodentia and Artiodactyla) be studied? Is there a difference in *E. coli* strains among herbivore, carnivore, and omnivore hosts? What genetic markers do the *E. coli* strains contain that would tell us something about their habitats? These difficult questions are central to constructing appropriate road maps of genomic analyses.

The genus *Simonsiella* may also be interesting as a test organism for studying speciation and the bacterial species concept. This filamentous bacterium is a beta-*Proteobacteria* that, like *E. coli*, lives as a commensal in mammals. However, *Simonsiella* species live in the oral cavities of mammals. Evidence from studies of 16S rDNA sequencing indicates that the strains from humans (Order Primates), sheep (Order Artiodactyla), dogs (Order Carnivora) and

cats (Order Carnivora) each form separate clades. Furthermore, the dog and cat strains are monophyletic, supporting the idea that there is no horizontal gene transfer within the carnivore group. Studies need to be performed to examine strains from well-known, closely related species pairs of host mammals (*e.g.*, primate species such as humans and their closest relative, the chimpanzee), to determine if they have co-evolved with their hosts. If they have, then genomic sequencing of four strains from each of these hosts should suffice in circumscribing, at the genetic level, the bacterial counterpart for well-known mammalian species.

The two examples of commensal bacteria cited above serve as illustrative examples for studying bacterial species concepts. However, it will be important to use members of different phyla and domains for these studies, as the processes and patterns of evolution may vary among prokaryotes. The best candidates would appear to be those species that are already well known. It is important to keep in mind that speciation in free-living bacteria is likely to be more complex than that in commensal bacteria because there are fewer constraints on their evolution. For example, commensal bacteria live in an almost constant environment in the host, whereas free-living bacteria are exposed to diurnal and seasonal changes and much more complex biological environments.

Similarly, a greater understanding of the genomic features of bacteria that make them important for agriculture should improve our ability to predict the utility of various microbes for agriculture. There is no question that the soil microbiota are critical for the success, not only of commercially important plants, but for restoring habitat for native plants. A fusion of information about genes and function will increase our ability to prescribe treatments that will have a widespread effect on various aspects of agriculture.

Industry has long appreciated the microbial world as a source of new products and processes. Such technical advances as PCR and the ligase chain reaction (LCR), along with methods of production of vitamins and amino acids, and discovery of new enzymes for the food and detergent industries are only a few examples of the benefits to be gained from mining microbial diversity. At present, such discoveries have been haphazard, but a new synthesis of systematics and genome science could help to refine the search for microbes that will be useful industrially.

Recommendation: Evaluate the radical idea of redoing systematics, using genomic information to redefine families, genera, and species. We can begin by reexamining our present concepts of systematics using genomics.

Symbionts Versus Free-Living Bacteria: A Model for Addressing Speciation

There is a broad range in the degree of association in microbial symbiosis. In some associations, such as that of the mitochondrion and the chloroplast with the eukaryotic cell, the bacteria have evolved to such a degree that they have become an essential component of the eukaryotic cell. Even in the case of *Buchnera*, the bacterial endosymbiont of aphids, an obligatory association has evolved during 150 to 200 million years. In such tight associations, the process of reductive evolution has pared the bacterial genome enormously from an ancestral genome of about 4 Mbp to one of about 650 Kbp. Thus, many of the genes that are nonessential to the host-symbiont association have been eliminated. Genome sequencing of such genomes will help lead to a better understanding of how this process occurred.

Connections to Macro-organismal Speciation

At what spatial scales does speciation occur in different types of microorganisms and in different types of environments? Are spatial (regional and global) patterns of speciation rates in microorganisms correlated with spatial patterns of speciation rates in terrestrial plants? How can rates of evolution in micro- and macroorganisms be evaluated? The fundamental contrast between microorganismal and macroorganismal generation times (and potentially speciation rates) may be critical for understanding the structure of taxonomic hierarchies in microbes and the temporal and phylogenetic levels at which species should be defined.

How does microorganismal diversity compare to macroorganismal diversity in terms of taxonomic hierarchies? Mechanisms of gene transfer (intra- and interspecific) among microbes may alter the

dynamics of speciation when compared with that of macroorganisms. Potentially high rates of short-distance gene transfer, as well as high rates of long-distance movement of individuals in certain environments (by wind or water transport), may reduce the potential for genetic isolation and allopatric speciation. However, the small size of microbes, as well as the difficulty of long-distance physical movement especially in soils and sedimentary and geological substrates, may facilitate genetic isolation and allow allopatric speciation at very small spatial scales. Consequently, speciation rates may vary greatly among microbes from different environments, or with different life history adaptations. Likewise, lineage-specific differences in gene transfer mechanisms and rates may produce variation in speciation rates among different phylogenetic groups.

One particularly interesting contrast is between relatively free-living microbes (*e.g.*, saprophytes) and those that are symbiotic or commensal with higher organisms, such as with vertebrates. The potential genetic isolation of microbial populations associated with different individuals, populations, and species of higher organisms may offer a continuum of genetic isolation and, thus, opportunities for speciation. In addition, the relatively constant and resource-rich environment experienced by symbionts is likely to result in very different dynamics of natural selection and speciation than those that occur in free-living microbes, many of which will experience less temporal stability in environmental conditions, as well as lower resource availability.

These problems leave many questions unanswered. How do the potentially higher rates of gene flow among microbes interact with their potentially greater opportunities for physical isolation, affecting speciation rates among microbes of different types and in different environments? How do variations in speciation rates interact with potential variations in extinction rates of species or lineages to affect the level of microbial diversity in different environments? Do higher rates of genetic transfer tend to produce more generalized or more specialized species? Do parasites and endosymbionts evolve faster (and have higher speciation rates) than microbes in less stable and/or resource-rich environments? What are the life history, physiological, energetic, and biochemical tradeoffs that constrain microbial evolution? How general can a generalist microbe be?

The same conditions that potentially affect microbial speciation rates are likely to also influence the dynamics of ecological processes, such as competition, facilitation, predation, etc. There are two factors that are thought to be important in regulating the competitive interactions and species diversity of plants (along the terrestrial to aquatic continuum): the rate of resource diffusion in the substrate and the rate of resource supply. Variation between these two factors creates a range of conditions from completely mixed chemostats (analogous to classical Lotka-Volterra competition models in ecology) to complete physical isolation of individual organisms, and thus the impossibility of competitive exclusion through resource competition (Huston and DeAngelis, 1994). In microbial communities, these same two environmental properties are likely to influence both resource availability (and potential competition) and gene flow between populations.

Is variation in these environmental properties associated with variation in microbial population dynamics, competitive interactions, speciation rates, and/or diversity patterns in microbes? Can experimental manipulation of diffusion rates and resource availability be used to create model systems for the study of microbial competition and coexistence and for the study of microbial speciation rates? Are these processes analogous to the competitive interactions and evolutionary dynamics of macroorganisms?

How are microbial communities structured in terms of energy flow (trophic structure), biogeochemical cycles, or autogenic and allogeneic environmental gradients? What is the relative importance of positive interactions (facilitation, complementarity, mutualism) versus negative interactions (competition, predation, allelopathy, toxicity) in different microbial environments? To what extent do microbes interact directly (genetic exchange or chemical signaling) versus indirectly through their effects on their common environment (such as allelopathy or toxicity)?

Goal: Intensive microbial genetic/biodiversity surveys, covering a range of environmental conditions and geological/evolutionary histories will be required to determine the extant patterns of microbial biology on Earth.

Technology Uses and Needs

New technologies have led to important recent advances in microbial ecology and evolution. Fluorescence *in situ* hybridization (FISH) allows for the identification, quantification, and localization of populations in complex environmental samples. Directed sorting will allow the sampling of selected fractions of the "metagenome" in an environment. This is an area in which technological tools, such as flow cytometry sorting, FISH, high density DNA hybridization arrays, and yet to be developed tools link with genomics. Use of these techniques has enabled, for the first time, a census of microbial diversity that is unrestricted by our current inability to cultivate the majority of microorganisms. Hybridization techniques yield reliable quantification of the more abundant species (>1%) and allow us to target those species for further studies. An in-depth genomic analysis might give important hints on what those organisms are doing in the environment. Within a few years, we should be able to study changes in gene expression of still uncultured bacteria in response to environmental change by using high density DNA hybridization arrays.

Although the role of technology is secondary to the research questions, advances in knowledge in microbial ecology are particularly dependent upon new technologies. However, new technology can have greater resolution than our ability to understand the implications from its use.

New technologies are needed that can rapidly compare species and that do not require whole genome sequencing. For example, *E. coli* and *Shigella* sp. are species that are co-linear for much of their sequences, then differ dramatically in some regions. Newer technologies will help us delineate such differences and determine their value in systematics and ecology without necessitating full-genome sequencing. We need rapid screening for taxonomic groups to help reduce the number of unknown microorganisms.

Recommendation: Use new technology to resolve discrepancies of classification and to inform evolution and ecology of microbial ecotypes. New technology is especially needed to probe for physiological traits and ecological activities of microbes in their native communities.

Our techniques both expand our horizons at the same time they limit us. Genomic data gathering has outpaced development of genomic theory, whereas microbial community ecology has not been outpaced by data. Collection of data on natural communities is still far behind. For genomics and community microbial ecology to converge, we need theory to provide guidance. Enzyme activities have been discovered for which we have yet to find the responsible genes. Information on these genes is missing due to our lack of appropriate techniques and sensitive enough assays.

Goal: A 10-year plan for using new technologies to study microbial identification and genomics should be integrated with theory development.

Need for Integrated Database Structure

One of the important needs is the ability to get data out and in usable form. A recent major biodiversity initiative in Germany began in response to the communication problems between scientists who study macroorganisms and those who study microorganisms. They devised an integrated database that allows researchers who study all ends of the organismic spectrum to communicate with one another.

Although genomic data are easy to distribute and make available and a number of databases exist for this, it is more difficult to search from an ecological perspective. For example, one can't search a current sequence database for all marine organisms within 20° of the equator, or all organisms that survive at a particular temperature. For example, one-meter GIS information would be helpful for habitat correlations with many organisms (*e.g.*, mycorrhizal fungi). But to have these kinds of data available, researchers would have to agree on required fields and format for the data. To encourage populating such a database, funding agencies and journals could require deposition of both habitat data and sequences before publication. Such a database could allow one to correlate single principal investigator data to large group data through links and updates of information. It could include shared data, publications, and electronic data that could be queried. Such a database would need long-term funding and

provide international access. It could aid educational outreach of microbial science to high schools and colleges.

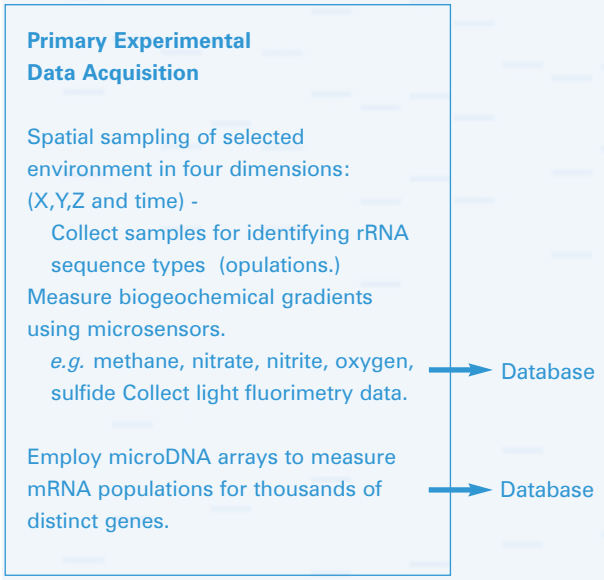


Figure 1. Data Analysis

Figure 1 illustrates the types of data that would be inserted into a database.

Although a universally accessible database would be a step toward sharing data and communicating information within the scientific community, there are more specific steps that need to be taken. One of the needs is the development of a metalanguage (*e.g.*, XML-DTD) to make this information more easily accessible and searchable in primary literature. ChemML/MathML are examples of such specialized languages.

Goal: Within 10 years, develop a large, international, integrated database that not only will include microbial data, including genomics, but also can include habitat information (collection site, time; geological and ecological data).

Instrumentation Development and Facilities

As mentioned earlier, microbial science, especially microbial ecology, relies heavily on new technologies to inform us about diverse, small organisms living in unique niches. Hence, collaborations with device engineers, chemists, physicists, and computational scientists are needed to propel the science forward. Without this type of encouragement, an opportunity is lost to tie technology development to scientific concepts. There are a handful of systems being tested. A few companies are making arrays to detect organisms. One of these is the Rapid Array Characterization of Microbial Ecosystems (RACEME). But significantly more work is needed. Nano- and micro- scale technologies need to be developed, as well as spectral analysis equipment with the ability to detect fine level gradients and minute concentrations of molecules. Sensitivity of spectroscopic equipment needs improvement so that instrumental or laboratory noise does not significantly compete with its functioning.

Genomics is technique driven. Thus, mechanisms need to be established to encourage and reward the development of new techniques and the necessary technology needed to efficiently and effectively deploy new techniques.

While contemporary methods, such as PCR and cloning, have provided tremendous advances, means to minimize difficulties associated with their bias are required. Quantitative PCR, quantitative hybridization, and *in situ* quantification represent new technologies that may be used to better study microbial biodiversity, but the costs, in equipment and time, need to be reduced. Microarrays provide much promise, but their specificity and sensitivity need to be improved to resolve closely related populations and to identify rare species.

All methods will, of course, have limitations, but these limitations only constrain inference. Techniques and new technologies serve to generate hypotheses by which future research is structured. Simplification, miniaturization, cost reduction, reliability, and portability for fieldwork are all necessary directions that need to be taken in technology development. Effective partnerships with industry need to be established to promote development of new technologies.

Recommendation: Develop Genome Resource Centers to advance technology and to provide a service-oriented approach for sharing microarray technologies and expression or genomic analyses, or proteomics research equipment. Smaller scale resources, such as primers for open reading frame (ORF) sequences, probes for genes, etc., could be made available through the American Type Culture Collection or a similar organization.

Recommendation: Encourage instrumentation development through collaborations with device engineers, chemists, physicists, and computational scientists, since uncovering the diversity and activities of the microbial world is dependant on such advances.

What Is the Future? The Research Road Map

We suggest a 10-year plan in which microbial diversity is more fully explored at the community, function and genome levels and will provide a road map for a structured, rapid integration of genomics to inform microbial systematics, evolution, and ecology (Figure 2).

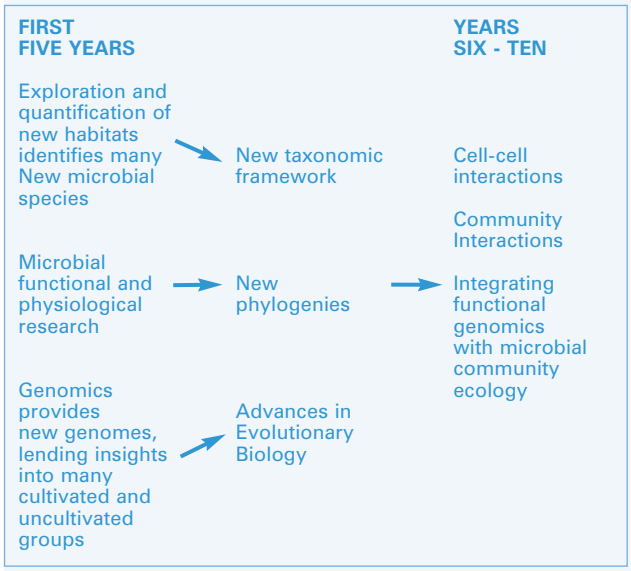


Figure 2. An Approximate 10-Year Plan in which microbial genomics informs microbial systematics, evolution, and ecology.

Education and Public Awareness

To inform the future public about the power and impact of genomics, we suggest K-12 partnerships, in which research laboratories partner with elementary, middle, or high schools. Media partnerships are also essential. The latter may be accomplished by holding informational or training sessions for the media, giving them hands-on experience with laboratory equipment. Courses, such as the Marine Biology Laboratory's in which scientists work with science writers, help inform the media and hence better communicate the science to the public.

The American Society for Microbiology (ASM), too, has particular responsibilities in communicating the benefits of increasing genomic research on microorganisms. They can engage in outreach activities with the general public, stressing the role that microbes play in maintaining the health of ecosystems, in public health, and in commerce. They should also foster a dialogue on ecogenomics with members of other professional societies through sponsorship of joint meetings.

Interdisciplinary Training

Interdisciplinary training can be subdivided into several broad categories based on the target audience. First, there is a continuing need to provide educational and retraining opportunities for senior and junior faculty members who want to adopt genomic approaches in their research. These needs could be met by the support for participants in summer training programs (*e.g.*, Woods Hole Summer Courses) or two- to three-month mini-sabbaticals.

Second, we encourage expansion of training grants and fellowships to provide funding for undergraduate, graduate, and postgraduate students that ensure future scientists will have the multidisciplinary knowledge and skills to realize the benefits of the genomic age. Besides core knowledge in microbiology, knowledge in ecology, evolution, genomics, bioinformatics, and computational sciences is needed in various combinations in different graduates. Such programs should also encourage participation by smaller schools. One way to meet this goal is through the use of multi-institutional grants with cross-enrollment.

set of defined tagging rules. Mark-up schemes have already been developed by the mathematics (MathML) and chemistry (ChemML) communities to expedite dissemination of research findings, both in print and electronic form.

Recommendation: Develop multidisciplinary training opportunities for microbiology students and faculty, including disciplines such as ecology, evolution, genomics, bioinformatics, and computational sciences. These could include training grants, summer short courses, cross enrollment programs, and virtual courses.

Third, there is a need to bridge the chasm between the communities of eukaryotic and prokaryotic systematists and ecologists. These communities have historically been divided for a variety of reasons. However, for larger initiatives (*e.g.*, Tree of Life) to have meaning, we need to encourage greater communication. Funding should be provided to sponsor symposia and roundtables in which these groups can interact at national and international meetings. We would also encourage greater participation of microbiologists on the NSF (National Science Foundation) systematics panels and more involvement of macrobiologists in programs such as NSF's Microbial Observatories. One suggestion is the use of "floating" panelists who bring microbiological expertise and perspectives to grant review panel discussions. We believe that this would augment the use of outside reviewers, but provide opportunities to clarify issues that are raised during the discussion of proposals. This would also provide an "educational opportunity" for the panelists who would not usually be exposed to concepts in microbiological ecology.

With genomics permeating all of biology, we anticipate changes in the curricula of most microbiology programs. Within a short time, introductory and advanced courses in genomics and bioinformatics will be required to ensure that students are adequately prepared for their careers. The manner in which these courses are accommodated will need careful consideration to avoid discontinuation of other course offerings that are part of the core curriculum in a microbiology program (*e.g.*, microbial physiology, ecology). There is also a need for quality teaching materials in these areas. While some texts exist in the field of bioinformatics, there are no comparable offerings in genomics at this time.

Genomic information should be communicated in a variety of formats, based on the intended audience. Clearly, journal articles remain a principal vehicle for communicating research findings. While sequence information is readily accessible, alignments, masks, updated annotations, and links to other types of data would greatly enhance the value of the data by making them more accessible directly from the literature. Much of the technology (mark-up languages such as SGML and XML) to accomplish this goal is already available, but it would require the development of a

Timetable and Research Questions

The participants suggest a 10-year plan with a progression from simpler challenges to the more complex in later years, organized according to questions addressed. We also identify the technology and intellectual challenges to be met to effectively address each question.

Time Scale	Level of Complexity	Questions Answered by Genomics	Technology/ Intellectual Limitations
3 yrs.	Species	Additional known genomes Evaluate the species concept Identification of core genes most useful for systematics	Advance annotation and improve bioinformatic tools Conceptualize the species and higher taxonomic groupings using genomic data
5 yrs.	Metapopulations Communities I (very simple) Communities II (simple)	Speciation, mechanisms of evolution, minimal genome	Comparative genomics Subtractive hybridizations
10 yrs.	Communities III (complex) Microbial mats Intestinal tract Oral cavity Marine plankton Simple terrestrial habitats	Genomic linkage to biogeochemical activities Minimal collective genome for sustainable system	Techniques for efficient screening of genes and assembly of genomes in communities Regulatory elements and network control Elucidate signaling Noninvasive methods for ecological activities Monitoring activities of single cells

Suggested Program Development

General Program Goals

Achieve sufficient and balanced resources for large-scale DNA sequencing and postgenomic environmental studies (including genome resource centers, specimen repositories and international integrated databases). In order to maximize the use of genome information, as well as to develop a more complete understanding of microbial evolution, it is essential to develop comprehensive ecological databases that can supply us with information about the ecological context of molecular, physiological, and genetic data.

- Develop a new theoretical framework for microbial taxonomy, underlying systematics with more extensive genomic, proteomic, and metabolomic data.
- Complete genome sequence analyses of free-living and commensal species (10-15 strains of each) selected on the basis of their environmental or medical relevance. This information is essential to evaluating the concept of species and to furthering understanding of the process of speciation. We recommend that *E. coli* be included as one of the reference species.
- Complete extensive microbial genetic/biodiversity surveys, covering a range of environmental conditions and sampling at the appropriate temporal and spatial scales needed to resolve patterns of distribution and establish connections with system-level processes.
- In 10 years have up and running a full database linking microbial and macroorganismal genomics, phenotypes, and environmental data.
- Develop technology and analysis capability to study microbial communities and symbioses holistically, measuring system-wide expression patterns (mRNA and protein) and activity measurements at the level of populations and single cells.

Suggested Reading

Béjà, O, EN Spudich, JL Spudich, M Leclerc, and EF DeLong. 2001. Proteorhodopsin phototrophy in the ocean. Nature 411:786-789.

Carpenter, SR. 1999. Microcosm experiments have limited relevance for community and ecosystem ecology: comment. Ecology 80:1085-1088.

Drenner, RW, and A Mazumder. 1999. Microcosm experiments have limited relevance for community and ecosystem ecology: comment. Ecology 80:1081-1085.

Gürtler, V, and BC Mayall. 2001. Genomic approaches to typing, taxonomy and evolution of bacterial isolates. Int. J. Syst. Evol. Microbiol. 51, 3-16.

Hess, WR, G Rocap, CS Ting, F Larimer, S Stilwagen, J Lamerdin and SW Chisholm. 2001. The photosynthetic apparatus of Prochlorococcus: Insights through comparative genomics. Photosynthesis Research 70: 53-71.

Huston, MA. 1993. Biological diversity, soils, and economics. Science 262:1676-1680.

Huston, MA. 1999. Local processes and regional patterns: appropriate scales for understanding variation in the diversity of plants and animals. OIKOS 86:393-401.

Huston, MA. 1999. Microcosm experiments have limited relevance for community and ecosystem ecology: synthesis of comments. Ecology 80:1088-1089.

Huston, MA, and DL DeAngelis. 1994. Competition and coexistence: the effects of resource transport and supply rates. Am. Nat. 144:954-977.

Koonin, EV, 1 Kira, S Makarova and L Aravind. 2001. Horizontal Gene Transfer in Prokaryotes: Quantification and Classification. Annu. Rev. Microbiol. 55: 709–42.

Liu, W-T, and DA Stahl. 2001. Molecular approaches for the measurement of density, diversity, and phylogeny. In: Manual of Environmental Microbiology, Second Edition. ASM Press, Washington, DC, pp. 114-134.

Marsh, TL, P Saxman, J Cole and J Tiedje. 2000. Terminal restriction fragment length polymorphism analysis program, a web-based research tool for microbial community analysis. Applied and Environmental Microbiology 66:3616-3620.

McFall-Ngai, MJ. 2002. Unseen forces: The influence of bacteria on animal development. Developmental Biology 242: 1-14.

Orphan, V J, CH House, K-U Hinrichs, KD McKeegan and EF DeLong. 2001. Methane-Consuming Archaea Revealed by Directly Coupled Isotopic and Phylogenetic Analysis. Science 293: 484-487.

Perna, NT, G Plunkett, V Burland, B Mau, et al. 2001. Genome sequence of enterohemorrhagic Escherichia coli O157:H7. Nature 409:529-533.

Relman, DA. The human body as microbial observatory. 2002. Nature Genetics 30: 131-133.

Relman, DA, and E Strauss. 2000. Microbial Genomes: Blueprints for Life. The American Academy of Microbiology, Washington, D.C. On-line at: http://www.asmusa.org/acasrc/aca1/html.

Rondon, MR, PR August, AD Bettermann, SF Brady, TH Grossman, MR Liles, KA Loiacono, BA Lynch, IA MacNeil, C Minor, CL.Tiong, M Gilman, MS Osburne, J Clardy, J Handelsman and RM Goodman. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Applied and Environmental Microbiology 66: 2541-2547.

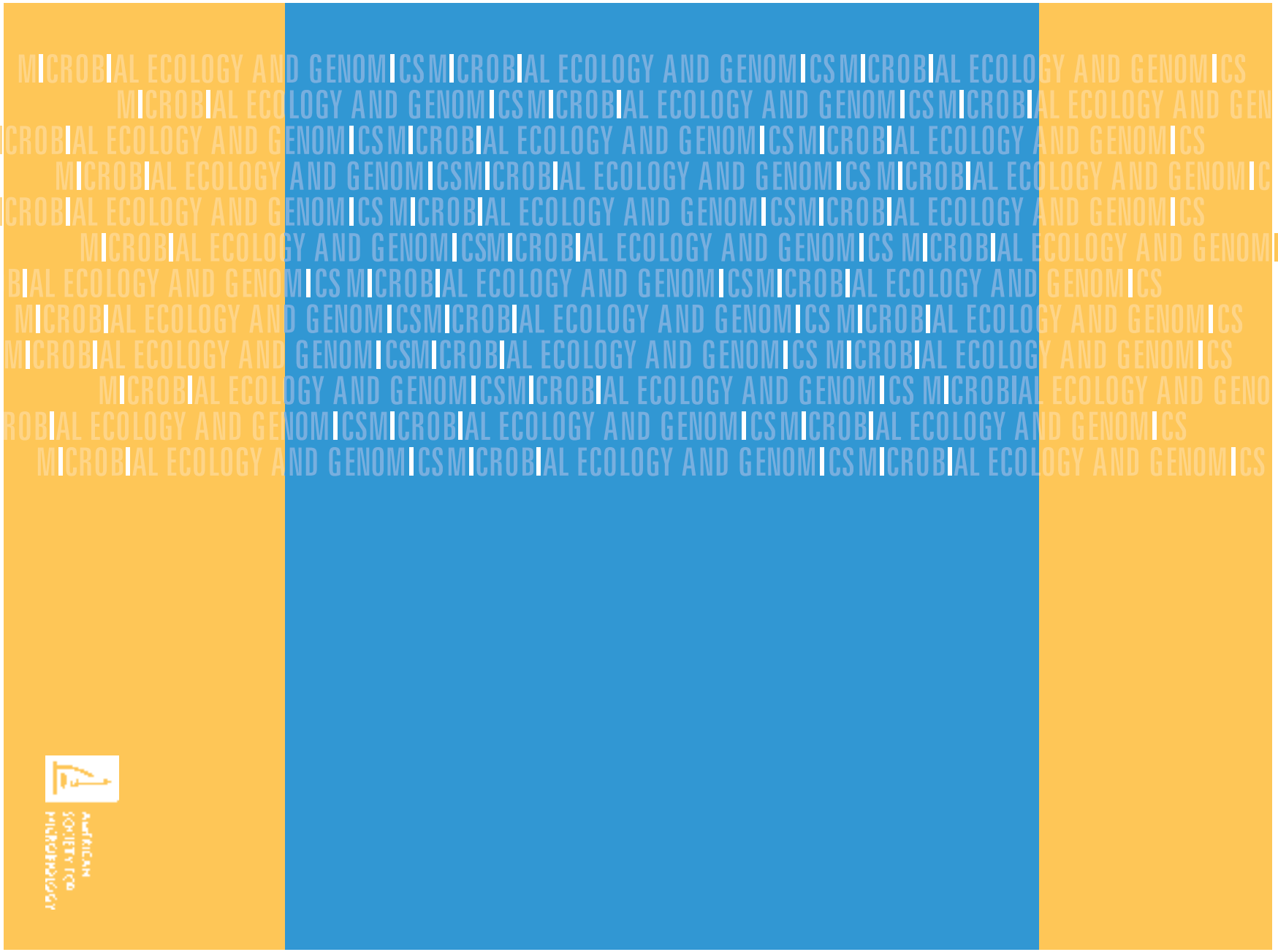
Stackebrandt, E, W Frederiksen, GM Garrity, P Grimont, P Kämpfer, MCJ Maiden, X Nesme, R Rosselló-Mora, J Swings, HG Trüper, L Vauterin, AC Ward and WB Whitman. 2002. Report of the Ad Hoc Committee for the Re-Evaluation of the Species Definition in Bacteriology Int. J. Syst. Evol. Microbiol. 52: 1043-1047.

Staley, JT, RW Castenholz, RR Colwell, JG Holt, et al. 1997. The Microbial World: Foundation of the Biosphere. The American Academy of Microbiology, Washington, DC. On-line at: http://www.asmusa.org/acasrc/aca1/html.

Tiedje, JM. 2000. 20 years since Dunedin: The past and future of microbial ecology. In, CR Bell, M Brylinsky and P Johnson-Green, eds. Microbial Biosystems: New Frontiers. Proceedings of the 8th International Symposium on Microbial Ecology. Atlantic Canada Society for Microbial Ecology, Halifax, Canada.

Woese, CR. 1998. A manifesto for microbial genomics. Current Biology 8: R781-R783.

this report designed by:
Pensaré Design Group
WWW.PENSAREDESIGN.COM


MICROBIAL ECOLOGY AND GENOMICS
AMERICAN SOCIETY FOR MICROBIOLOGY