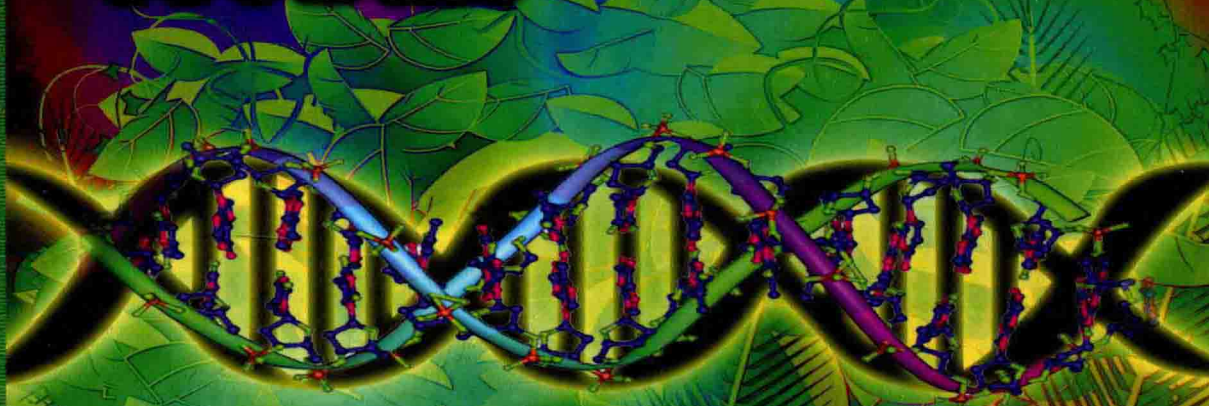
Genetics - Research and Issues

METAGENOMICS AND ITS APPLICATIONS IN AGRICULTURE, BIOMEDICINE AND ENVIRONMENTAL STUDIES



Robert W. Li
Editor

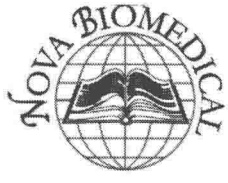
Novinka

Nova Biomedical

GENETICS – RESEARCH AND ISSUES

METAGENOMICS AND ITS APPLICATIONS IN AGRICULTURE, BIOMEDICINE AND ENVIRONMENTAL STUDIES

ROBERT W. LI
EDITOR



New York

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA
Metagenomics and its applications in agriculture, biomedicine, and
environmental studies / [edited by] Robert W. Li.
p. ; cm.
Includes bibliographical references and index.
ISBN: 978-1-62808-644-7 (softcover)
1. Microbial genomics. 2. Microbial ecology. I. Li, Robert W.
[DNLM: 1. Metagenomics--methods. 2. Environmental Microbiology. QU 58.5
M587 2010]
QH434.M48 2010
660.6'5--dc22
2010025793

Published by Nova Science Publishers, Inc. † New York

GENETICS – RESEARCH AND ISSUES

METAGENOMICS AND ITS APPLICATIONS IN AGRICULTURE, BIOMEDICINE AND ENVIRONMENTAL STUDIES

GENETICS – RESEARCH AND ISSUES

Additional books in this series can be found on Nova's website under the Series tab.

Additional e-books in this series can be found on Nova's website under the e-book tab.

FOREWORD

John C. Wooley

Evidence ultimately revised our view of some fundamental aspect of life. Each revolution in our understanding of life provided not just insight but, even in the midst of significant progress, an initial shock. Then, upon reflection we have been reminded how little we really know about biology and the processes of life.

A molecular approach was often the framework on which discoveries, such as genes-in-pieces, overturned our textbooks, let alone the words of sages. Remarkable details of cellular organization and processes subsequently emerged. In the past two decades, combining the vast improvements in DNA sequencing with equivalent computational advances led to the first complete genome sequence and its paradigm-shattering analysis. Today, we enjoy a never-ending parade of newly completed genomes numbering in the thousands. The genomics scale analysis of messenger RNA (transcriptomics) and of protein species (proteomics), and their subsequent computational analysis is leading to an understanding of biology at a systems level.

Some revolutions in biology occur literally at our feet, or under our noses. Metagenomics, a nascent interdisciplinary science within biology, presents just such a revolution. Metagenomics is the study of microbial communities sampled directly from their natural environment, without prior culturing. DNA sequencing (or alternatively, an assay of functional properties) is applied to the culture-independent analyses of complex and diverse ("meta") populations of microbes. Accordingly the research scale is one of potentially complex populations or communities, not of individual species (although the study of a single species in an environmental setting, such as the ocean or acid mine drainage, is generally included in the range of what is defined as metagenomics research). In contrast to traditional whole genome sequencing efforts, the first level of metagenomics research aims to determine directly the entire assemblage of genes in a sample (i.e., the metagenome) from some (biological/host specific or physical) environment, and subsequently, to analyze their biochemical activities and complex interactions. Nonetheless, genomics, whenever an individual species can be cultured, will help cement metagenomics and population analyses, and thus, is a singular instantiation of metagenomics. Correspondingly, all of the experimental approaches already used in genomics, from functional biology to transcriptomics and proteomics, are or will be applied to metagenomics.

Toward the end of the 20th century, research on microbial species distribution and populations provided the first glimpse into this most recent revolution in biology. Among the early discoveries was an unexpectedly vast diversity of microbes, in contrast to our narrow understanding of microbial variety based on studies in culture or laboratory-based microbiology. The new environmental microbiology sampled our unbelievably diverse earth and provided an essential perspective: this planet is a microbial world. Indeed, the notion that we inhabit a microbial world is a common theme of 21st century microbiology. However, the field of metagenomics is still being defined, as two decades of pioneering research are enhanced by ever-advancing (and cheaper) sequencing methodologies, correlated computational hardware, and innovative software tools for data reduction and analysis. This sustained, convergent technology growth in sequencing and computing enabled the collection and analysis of large ocean samples from the Sargasso Sea and a worldwide voyage. Such sustained explorations in microbial ecology and population-level microbiology, with ultra-scale sampling, provided both publicity and a timely nudge to microbiology and biological research.

The essence and extent of this revolution, including the techniques used and the potential range of impacts and scientific challenges, were first summarized and authenticated in a National Research Council report: The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet (National Academy Press: http://www.nap.edu/catalog.php?record_id=11902). Its release enhanced opportunities for the research community and opened the door for federal initiatives. In reading early research papers, the reader should note that several terms were in use, such as community genomics, ecological genomics and environmental population genomics, but metagenomics has become the established term for this new science.

An early, detailed definition is found in the report from the American Academy of Microbiology: THE GLOBAL GENOME QUESTION: Microbes as the Key to Understanding Evolution and Ecology, Merry R. Buckley: http://academy.asm.org/. Metagenomics "entails large-scale sequencing of pooled, community genomic material, with either random or targeted approaches, assembly of sequences into unique genomes or genome clusters, determination of variation in community gene and genome content or expression over space and time, and inference of global community activities, function, differentiation, and evolution from community genome data." The NRC study notes that the term "meta" (from the Greek) means transcendent; and further, that metagenomics "seeks to understand biology at the aggregate level, transcending the individual organism to focus on the genes in the community and how genes might influence each other's activities in serving collective functions."

Consider the broad range of traditional sciences, i.e., the major disciplines such as physics, chemistry and microbiology, which were the primary fields for the scientists who established molecular biology. The biological sciences frequently benefited from such external infusions, and not just from the slight differences in their perspectives. Metagenomics is an exceptionally interdisciplinary endeavor and brings together fields such as biogeochemistry, oceanography, agricultural science, climatology, and ecological, evolutionary and environmental biology, as well as a wide range of research in microbiology and genomics. Characterized by the participation of multiple fields of study, the flavor of the experimental studies in metagenomics represents a springboard for innovation and reflects the sense of revolution that permeates the field.

The rapid growth of metagenomics entails an expanded focus on microbes in a wide range of basic and applied research fields, ranging from microbiology to agriculture, biomedicine to bioenergy, and from veterinary to clinical medicine. Microbes not only dominate the physical environment, multicellular organisms are colonized by vast numbers of microbes, leading to the concept of supraorganisms (whose genetic capacity is the aggregate of contributions from the host and at least some of their microbes). Entirely new research questions are presented, such as the co-evolution of hosts and their microbes. Metagenomics observations have also led to an expanded range of clinical disciplines incorporating microbiology research, whereas medical microbiology had been restricted to "subspecialties" in clinical and biomedicine. Specialties like gastroenterology now need to interact with the basic biosciences in order to characterize the microbes in humans, or the human microbiome.

The wide range of potential applications for metagenomics originates from the capacity of microbes to live in almost any environment. They thrive in biofilms, fouling ships and human teeth, adapt to extreme conditions like acid mine drainage and the mammalian gut, and successfully colonize both the shallow and deep oceans, from near freezing temperatures to boiling. Besides being dispersed throughout the vast diversity of soils and directly participating in rhizobial interactions, microbes succeed within hosts as distinct as the bovine rumen and the termite hindgut. While these are obviously novel environments for the host microbiota (or novel microbiomes), the potential capacities of microbes allow them to grow in many different environments. It will be important for future researchers to be aware of all metagenomics studies, not just those from a given habitat. By providing ready access to all relevant data on all microbes in all environments, the community knowledge resources and their funding sources can ensure the entire research field has such access. Cooperation and collaboration will enable powerful comparative studies and the most rapid progress possible.

Like those previous revolutions in biology, metagenomics, beginning with the early microbial phylogenetic studies on environmental communities, has often reminded us how little we know, even about the biology of microbes and their impact on the environment. At the same time, the convergent computing and sequencing technologies that opened up this new research approach for basic biology already demonstrated the vast diversity of proteins and provided insight into microbial ecology and evolution in the physical world. Similarly, the principles and theory of evolution and ecology are being applied to microbes living in hosts, including characterizing population dynamics and colonization models. A particular focus has been on the microbes living in the human gut and looking at the nature of human host – microbe interaction in terms of human disease. Besides the insight into diseases such as Crohn's and cystic fibrosis provided by metagenomics, a new path has been opened for biomedicine: exploration of the beneficial roles of microbes living in humans, collectively termed the human microbiome. The new approaches in research have introduced a new perspective for 21st century medical microbiology, extending the subject beyond pathogens and diseases. Expanding the potential for study of microbes from about six thousand in culture collections to the millions of microbial species in the physical environment will facilitate a more comprehensive exploration of the capacity of microbial chemistry, and in turn, the potential for new drugs among the vast range of compounds produced by microbial communities. The metagenomics sequence universe (and their RNA and protein products) opens new avenues for discovery-based science (viewing metagenomics as a genetic resource), from synthetic biology to drug discovery (and from bioremediation to bioenergy research). Even more abundant than bacteria, viruses add another perspective. Virus

metagenomics provides information on the viral activities per se, the impact of viruses in various ecosystems (such as on microbial population dynamics), and insight into the mechanisms for the vast genetic diversity of microbes and their various means for energy metabolism.

The reach of metagenomics extends well beyond basic biology and biomedicine. Microbes play essential roles in the environment, from photosynthesis to converting carbon dioxide to release oxygen and yield carbon compounds more readily metabolized by many organisms. Such chemical transformations occur as a result of the collective contributions of a microbial community. Metagenomics approaches will also provide insights into how gene product exchange can serve to enable community level production. Understanding exchange and community level production may yield baseline data and identify the genes with which to engineer microbes for a specific applied goal. An application along these lines is being pursued in bioenergy research, to see how our knowledge of the genomes and overall biology of plants and microbes may produce alternative biofuels sources as alternatives to petrochemicals. Current work focuses on characterizing molecular details, adapting all of the contemporary methods of genomics. Analysis of the proteins in metagenomic samples (metaproteomics) will provide novel impetus, yielding insight into community function as well as the direct products of microbial activity. In agriculture, resident soil microbial communities have been shown to protect plant health, carrying out many biochemical transformations necessary for plant growth. Similarly, metagenomics provides new insights into agricultural animals such as cattle. For example, rumen microbiota respond significantly to dietary changes, so analysis by metagenomics opens up opportunities for manipulating rumen fermentation.

The importance, excitement about, and potential impact of metagenomics are shown clearly by individual discoveries and their publication. However, the emergence of a new science often means that current literature is widely dispersed, on one hand; and on the other, that only individual papers and specialized reviews are available. A volume collecting inclusive and comprehensive technical reviews across the vast range of metagenomics research is very much needed.

In your hands you now hold the timely and requisite response that reflects the existing opportunities in environmental metagenomics and microbiomes. This superbly selected collection constitutes an engaging, informed dialogue on metagenomics and its promising and diverse applications. It brings together the relevant and requisite content of microbiology, with research contributions introduced from other life science disciplines, to encourage future research directions and the establishment of interdisciplinary collaborations, to recruit young scientists, and in general, to accelerate growth of the metagenomics research community. For reasons that will be more fully articulated below, I am enthusiastic, even keen or upbeat, about METAGENOMICS and its applications in agriculture, biomedicine, and environmental studies.

My grey beard, so to speak, cannot manage to think or write the full title frequently, so FYA (for your amusement - as well as mine), I think of this work as "METAGENOMICS – the book," with "METAGENOMICS – the movie" playing in nature, and for simplicity, will refer below to the volume simply as METAGENOMICS. Thank you, contributors, for describing the new science of metagenomics.

Following an applied theme for this collection is both practical and of high value in that the broad range of applications of metagenomics is itself extraordinary. The volume most

importantly provides a broad knowledgebase that will enable and encourage a growing dialogue among the disparate domains of environmental, ecological, molecular, and medical microbiology. METAGENOMICS includes considerations of requisite technology and computational tools, while focusing on approaches that answer important questions in molecular biology; environmental, population and evolutionary biology, ecology; biomedicine and clinical medicine. Providing a balanced representation among research domains, METAGENOMICS will enhance interactions across the wide distribution of the research community.

While next generation (DNA) sequencing fuels the metagenomics revolution, the key to progress in science, especially for an emerging science, is invariably the next generation of researchers. This highly interdisciplinary, broadly conceived and comprehensive volume, METAGENOMICS, will undoubtedly serve to excite young scientists to participate in the new science. At the same time, it will inform established researchers from disciplines so disparate they would not typically manage to interact. Such interactions will enable novel research directions and collaborations that answer the challenges made apparent through the new perspective known as metagenomics. Completely unexpected findings in metagenomics will translate into new insights across all fields of interest, and the breadth and depth of coverage presented herein is much appreciated and applauded.

I will leave the details of the chapters in this volume for the editor and his introduction, but point briefly to the attributes of the collection that directly reflect the nature of the challenges and opportunities in metagenomics. My sense of the immediate opportunities across the range of the new science is constructed and drafted from my own viewpoint, based notably on the International Congress for Metagenomics that I have been organizing since 2006 (see http://metagenomics.calit2.net/ and links from that page). Naturally, when I turned to this volume, I was quite pleased and energized to see that these opportunities are covered; and aspects that I consider particularly important are highlighted. I recommend the discussions on the state of the art of the research, the algorithmic and computational challenges, a consideration of functional biology as well as of sequencing, and the range of the new science as well as the breadth of its potential application. METAGENOMICS provides an informed guide for the journey of discovery in metagenomics, one that will keep scientists well entertained with its challenges for the foreseeable future. I highly recommend the volume to all members of our community and all those thinking about joining the adventure.

As the NRC report states: "Microbes run the world. It is that simple." A whole suite of scientific questions beyond our imagination today, questions about the microbial world (and its impact on the human world) in agriculture, in the environment, and in health and disease, will be answered as this new science matures. I anticipate that METAGENOMICS, as an insightful and inclusive study on metagenomics and its applications, will have played an important role.

CONTENTS

Foreword	*John C. Wooley*	**vii**
Chapter 1	Metagenomics: Methods, Applications, and Challenges *Martina M. Ederer*	**1**
Chapter 2	Towards a More Complete Metagenomics Toolkit *Michael E. Sparks and Robert W. Li*	**33**
Chapter 3	Analyzing Metagenomic Data: Inferring Microbial Community Function with MG-RAST *Dionysios A. Antonopoulos, Elizabeth M. Glass and Folker Meyer*	**47**
Chapter 4	Automatic Annotation of Microbial Genomes and Metagenomic Sequences *Victor Solovyev and Asaf Salamov*	**61**
Chapter 5	Next-Generation and Future DNA Sequencing Technologies and Metagenomics *Peiming Zhang, Jian Gu, Jin He, Weimin Gao, Weiwen Zhang, Stuart Lindsay and Deirdre R. Meldrum*	**79**
Chapter 6	Comparison of De novo Short Read Assemblers on Simulated Metagenomic Data *Ying Huang, Weizhong Li, David Perkins and Robert W. Li*	**107**
Chapter 7	Submetagenomics: Selective Sampling of Metagenomic Subsets of Interest *Yi Vee Chew and Andrew J. Holmes*	**121**
Chapter 8	Dynamics of the Rumen Microbiota *Robert W. Li, Michael E. Sparks and Erin E. Connor*	**135**

Chapter 9 Detection of Functional Shifts in the Rumen Microbiota in Response to Propionate Intake in Cattle **165**
Michael E. Sparks, Ying Huang, Ransom L. Baldwin, Weizhong Li, Erin E. Connor, Congjun Li, Tad S. Sonstegard, Steven G. Schroeder, Brian J. Bequette and Robert W. Li

Chapter 10 Carbohydrate Actives Enzymes Derived from Metagenomes: From Microbial Ecology to Enzymology **185**
Renaud Berlemont and Moreno Galleni

Chapter 11 An Evolutionary Ecology Perspective on Comparative Metagenomics **211**
Dieter Ebert and Weihong Qi

Chapter 12 Interactions of the Gut Microbiota and its Host in the Context of Gastrointestinal and Metabolic Diseases **237**
Rodrigo Bibiloni, Lutz Krause and Chieh J. Chou

Chapter 13 Metagenomic Analysis of Human Gut Microbiota **255**
Volker Mai, Mihai Pop and Mark Schell

Chapter 14 Emerging Relevance of Metagenomics in Medicine **277**
Vasuki Venkat, Ying Huang, Patricia Finn and David Perkins

Chapter 15 Metagenomic Gene Discovery **287**
Rolene Bauer, Moola S. Mutondo, Robert J. Huddy, I. Marla Tuffin and Don A. Cowan

Chapter 16 Emerging Techniques for Metagenome Screening **321**
Takefumi Shimoyama and Kazuya Watanabe

Chapter 17 Viral Metagenomics: From Fish Slime to the World **337**
Dana Willner, Christelle Desnues and Forest Rohwer

Chapter 18 Exploration of a Microbial Community for Novel Genetic Resources Using Activity-Based Screening of a Metagenomic Library **367**
Hikaru Suenaga and Kentaro Miyazaki

Chapter 19 Microbial Communities in Crude Petroleum Oils and Metagenomics **381**
Kunio Yamane, Hideaki Maki, Shigeru Matsuyama and Hiroo Uchiyama

Chapter 20 Environmental Metaproteomics: Identifying the Protein Machinery that Impact the Activities and Interactions of Microbes in Isolation or Communities **407**
Richard J. Giannone, Martin Keller and Robert Hettich

Contributors **437**

Index **441**

In: Metagenomics and its Applications…
Editor: Robert W. Li
ISBN: 978-1-62808-644-7

Chapter 1

METAGENOMICS: METHODS, APPLICATIONS, AND CHALLENGES

Martina M. Ederer

ABSTRACT

The new science of metagenomics is transforming the traditional, reductionist approach of microbiology. Rather than focusing on one bacterial strain in pure culture under specific laboratory conditions, metagenomics takes a community approach studying the ecosystem as a whole. Ideally the data should lead to a detailed analysis of the microbial community of a particular environment; however, so far the results have been mostly descriptive. Generally, three different habitat types can be distinguished, i.e., naturally occurring sites (ocean, soil, etc.), managed sites (sewage treatment plants, bioremediation sites, etc.), and host specific sites (animal gut, mouth, etc.). From these environments samples are taken and community nucleic acids are isolated. 16S rRNA gene libraries and/or shotgun libraries are constructed and sequenced by Sanger sequencing. Alternatively, shotgun sequencing is conducted by next-generation sequencing techniques (454 pyrosequencing). Since sequencing costs have been decreasing drastically recently, vast amounts of sequencing data are being generated. Metagenomics has already led to a number of significant discoveries, regarding both genome reconstructions and metabolic and functional potentials of specific environments. Here we take a look at the history of metagenomics, survey some of the most common metagenomic applications and approaches, and examine some of the problems that are being encountered by metagenomics researchers. Finally, we explore what steps can and must be taken to move metagenomics in the direction of hypothesis-driven science.

INTRODUCTION

In the beginning there was the pure culture … . Much has happened since this paradigm was the center of microbiology. Metagenomics, the study and analysis of community DNA sequence data, has since revolutionized the field. The idea of metagenomics was prefigured by Joshua Lederberg referring to the "World Wide Web" of microbial genomes (Morse S.J., 2008); however, the term itself was not coined until 1998 (Handelsman et al., 1998) during a search for novel antibiotics from soil-associated microbial communities.

For most of its history the field of microbiology was primarily centered on the pure culture. It was long believed that microbial organisms could only be classified if they could be cultured (Society of American Bacteriologists, 1923) and as more and more bacteria were isolated and grown in pure culture, some microbiologists even became optimistic about completing the inventory of microorganism by proclaiming that "great strides are being made characterizing the microbial world" (Waksman and Starkey, 1931). Today, we know that we have scratched only the surface of what constitutes the Earth's microbial biodiversity, and a renewed interest in the identification of microorganisms and characterization of microbial consortia has become evident.

The biology of eukaryotes as well as microbiology thus far has focused on reducing the field to the study of its parts by characterizing individual cells, genes and molecules, rather than considering larger assemblages of organisms or complete biological systems. During the last century many important discoveries were made using this reductionist approach; however, for the more complex challenges of the 21st century such as emerging human diseases, requirements for novel, renewable energy sources, environmental pollution problems, and global climate change, a more systems-oriented approach seems to be appropriate (Little et al., 2008).

In the 1980s scientists realized that microbes were much more ubiquitous, diverse, and numerous than previously thought. Staley and Konopka (1985) referred to the "great plate count anomaly" describing the reoccurring phenomenon that enumeration of bacteria by microscopic analysis yielded much higher numbers than was observed when these communities were plated. Since then it has become clear that to date most of the bacterial species on earth, probably > 99.9%, have not been cultured or otherwise described.

Estimates of bacterial diversity are >10^7 different species (Curtis et al., 2002; Eisen, 2007) with the largest diversity expected in the oceans (2 x 10^6 distinct species) and in the soil (4 x 10^6 species per ton of soil) (Curtis et al., 2002). The total number of bacterial cells on earth is estimated to be in the order of 4 - 6 x 10^{30} (Whitman et al., 1998). Microbes store about 50% of the total carbon present in all living organisms, and harbor the largest pool of cellular N and P, 10-fold higher than all plants together. Bacteria have been around for at least the last three billion years, but only about 5,000 species of bacteria have been classified at the species level (Staley, 2006). Considering that about 350,000 different beetle species, which evolved in much less evolutionary time, have been described, it becomes obvious that the vast majority of bacteria and archaea, resistant to current isolation and culturing techniques, has yet to be discovered, described and eventually cultured (using novel, innovative techniques) and classified.

With this realization of how little is actually known, microbiologists needed a more comprehensive systems-oriented approach for the characterization and classification of

microbes, describing them in their environment, by their potential interactions with community members, and by their metabolic potential. In addition, alternative bacterial classification methods, not based on pure culture metabolic differences, were required.

WHERE DID IT ALL START?

The Origin Of Metagenomics

The origins of molecular evolutionary study lie in the 1960s when Zuckerkandl and Pauling (1965) used amino acid sequence differences in hemoglobin from different organisms to look at evolution. These researchers postulated that amino acid sequence differences can be used as a direct measure of evolutionary distance and coined the term "molecular clock." However, since hemoglobin is not universally present in all organisms, this approach is not universally applicable. In the 1970s Carl Woese looked at another molecule, the 16S rRNA (Woese and Fox, 1977). The 16S rRNA (or 18S rRNA) molecule is a much better candidate since it is a structural component of the protein synthesis machinery and thus universally present in all organisms. It shows a high amount of conservation in all living organisms due to its sequence constraints as a structural component of the ribosome. Woese looked at T1 ribonuclease digestion patterns of the 16S rRNA molecule isolated from different organisms. His research resulted in the discovery of a new, distinct form of life, the *Archaea* (initially termed archaebacteria). The tree of life was branched dichotomously into prokaryote and eukaryote. The prokaryote consisted of the kingdom Monera with the subkingdoms Eubacteria and Archaebacteria. The eukaryotes were subdivided into the kingdoms Animalia, Plantae, Fungi and Protista. Together they represented the five kingdoms of life. The discovery of the Archaea and the realization that they were a distinct separate group had a huge impact on the structure of the tree of life and led to a "new world order" – specifically the three domains of life: the Eukarya, the Bacteria and the Archaea.

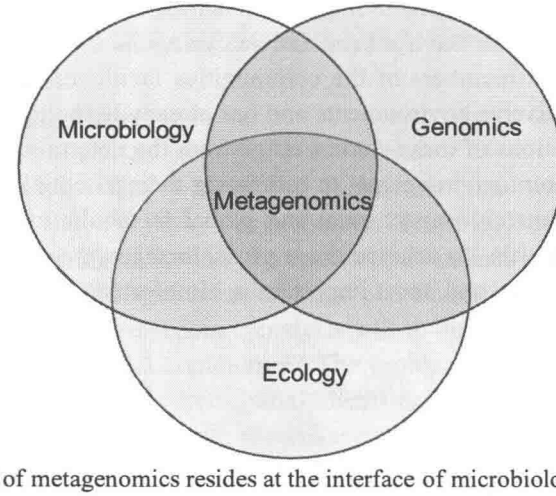


Figure 1. The new field of metagenomics resides at the interface of microbiology, genomics and ecology.

Pace and colleagues, interested in bacterial diversity and classification but no longer satisfied using "simple morphology" of bacteria and "sometimes ambiguous" physiological traits to characterize bacterial species, were looking for alternative, non-traditional methods to characterize bacterial species (Pace et al., 1986). They started using 16S rRNA sequence differences as a means for classifying bacteria. By that time sequencing technology had made great leaps forward (Sanger et al., 1977), so 16S rRNA gene sequence analysis became the most important tool to study microbial diversity. Applications extended from identification of bacteria in clinical samples, surveying bacterial contaminations in different environments, to detection of novel bacteria in the environment. Ultimately, these kinds of studies led to the recognition of the new field of metagenomics.

Metagenomics

While genomic analyses of the 1990s targeted the genomes of individual plants, animals and microorganisms, the current focus is now shifting toward the genomes of whole communities of organisms. However, this kind of analysis still faces many obstacles. The concept of metagenomics was officially introduced in 1998 (Handelsman et al.) and since has developed into a new field of study. Metagenomics, the study "beyond genomics" or the study of the "genome of all combined genomes," represents the interface between microbiology, ecology and genomics (Figure 1). Metagenomics aims to characterize the microbial diversity, genetic complexity and metabolic potential of a community in a particular environment by analyzing the community nucleic acid composition. Community diversity can and has been assessed using 16S rRNA DNA PCR amplicons generated from environmental DNA samples. These studies are aimed at answering the question, "Who is there?" Metagenomics has taken this approach one step further. It is aimed at the answering the question, "What are they doing?" Community DNA obtained from an environmental sample is cloned as a small or large insert clone library and subjected to nucleotide sequence analysis, or it is directly sequenced with a shotgun approach using novel next-generation sequencing (Figure 2). This systems-biology approach of analyzing/sequencing community DNA directly from a particular environment and analyzing the metagenome without the need to cultivate the individual members of the communities facilitates surveys of genomic and metabolic potentials in diverse environments and has already led to many new and important discoveries. The applications of these studies range from the determination of biodiversity of microorganisms in different environments to estimating their potential metabolic activity and how these activities ultimately impact local and global metabolic cycling, i.e., asking how microbes affect life on Earth. To achieve these goals, metagenomics requires the interaction of genomics, systems biology and, most importantly, bioinformatics.

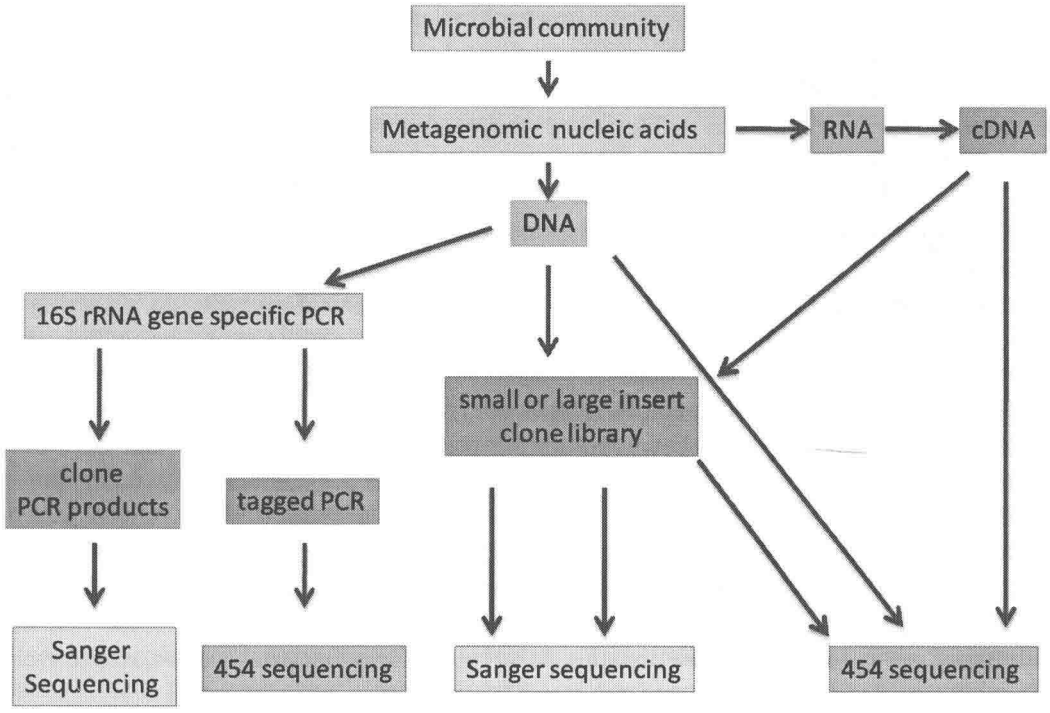


Figure 2. Outline of the scope of metagenomic applications.
Microbial diversity estimations result from the analyses of 16SrRNA gene comparisons (blue). Alternatively, total community nucleic acids can be subject to analysis. Shot-gun DNA sequencing can be done directly by new-generation sequencing (like 454 pyrosequencing), or first cloned as small insert or large insert clone libraries that are sequenced by Sanger sequencing or 454 sequencing technology. Analysis of the data is either gene-centric or genome-centric. Community RNA is first reverse transcribed to cDNA, which can be sequenced directly or cloned first. These data lead to functional analysis and suggest which enzymatic activities are present in a community.

How Can we Make Metagenomic Studies More Meaningful?

Bacterial Species Definition

In order to determine "Who is there?" it is essential that we effectively recognize, identify and correctly classify who we encounter. However, ample disagreement and controversy exist regarding the definition of a species, particularly in the domains of Bacteria and Archaea. If we cannot find agreement on what constitutes bacterial or archaeal species, how can we then expect that metagenomic analysis will to be able to adequately describe the members of a particular microbial community? The problem begins not at the species level but at the base of the phylogenetic tree. Initial classification efforts placed the monera (all prokaryotes) at the base of the tree, with the eukaryotes considered more complex and thus further up the trunk of the tree, giving rise to the rest of the Earth's biodiversity. This dichotomous classification system, eukaryotes vs. prokaryotes, uses a positive description for eukaryotic organisms, but defines the prokaryote by what it is not (*i.e.,* it does not have a