

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.