# Metagenomics and Microbial Communities

**Jo Handelsman,** *University of Wisconsin–Madison, Madison, Wisconsin, USA*

Metagenomics is the genomic analysis of microbial communities. This new science provides access to organisms that are recalcitrant to culturing, as are the vast majority of microorganisms on Earth. Metagenomics thereby offers a first peek at the wide variety of life that has never been studied, thus providing new insight into the structure and function of ecosystems as diverse as the oceans and the human gastrointestinal tract.

Advanced article

Article Contents



doi: 10.1002/9780470015902.a0020367

## Introduction

Microorganisms comprise the overwhelming majority of life forms on Earth. The diversity of microorganisms reflects a staggering array of functions, providing services that are necessary for all other life to exist. Microorganisms drive most of the chemical cycles on Earth, process waste and promote growth and reproduction of plants and animals. In addition, microorganisms serve human beings by producing drugs, fermenting food and maintaining our health. **See also**: Bacterial Ecology

Most microorganisms live in communities, or assemblages of more than one species. In these communities they carry out their own functions and contribute to and depend on the activities of other microorganisms. Only recently, however, have microbiologists begun to appreciate the true magnitude of diversity within microbial communities. This understanding was late in coming because much of microbiological research is based on culturing organisms in the laboratory, and it was discovered only in the last few decades that most microorganisms do not grow under standard culturing conditions. Consequently, microbiologists were stumped about how to understand these teeming communities that could not be 'seen' by traditional microbiological methods.

Metagenomics offers a way to access unculturable microorganisms because it is a culture-*independent* way to study them. By pooling and studying the genomes of all of the organisms in a community, all of the functions encoded in the community's deoxyribonucleic acid (DNA), or metagenome, can be studied. Metagenomics has revolutionized microbiology because it offers a window on an enormous and previously unknown world of microorganisms.

## Microbial Services

Microorganisms are the most important organisms on Earth, but they receive little credit for their contributions because they cannot be seen with the naked eye. Most people take for granted the air we breathe, the food we eat and our bodies' most basic functions, and very few people stop to thank the world of microorganisms for their contributions to our welfare. Despite our ingratitude, we are entirely dependent on microorganisms. There are numerous examples of essential services provided by microorganisms (see **Figure 1**) (Madigan and Martinko, 2005); this article includes only a few.

### Genesis of oxygen

A dramatic step in the history of life on Earth was the transition from anaerobic to aerobic life made possible by photosynthetic bacteria. When life on Earth began, the environment was anaerobic, or oxygen-free. About 1 billion years ago, some of the microorganisms in the Earth's oceans evolved the ability to use light energy from the sun to drive chemical reactions that resulted in the release of oxygen. Oxygen accumulated in the atmosphere until it was sufficiently abundant to support the emergence of organisms that require oxygen. Once the oxygen was at a relatively high concentration, oxygen molecules ($O_2$) began to collide and they produced ozone ($O_3$). Eventually a thick layer of ozone accumulated in the stratosphere, $\sim 25$ km from the Earth's surface. Ozone is special because, unlike oxygen molecules, it absorbs ultraviolet light. The ozone layer, therefore, is critical to life on land because it filters out many of the sun's ultraviolet rays, protecting the surface of Earth from radiation that damages cells and their DNA and that had previously made life on land impossible. Following the formation of an oxygenated atmosphere and the ozone layer in the stratosphere, the diversity of life exploded because many different habitats on land became livable environments. Oxygen and the ozone layer, which are required for human life on Earth, were first generated entirely by microorganisms and today are maintained by photosynthetic bacteria as well as plants. **See also**: Bacterial Evolution

### Conversion of nitrogen

Another service provided by microorganisms is supplying nitrogen in a form that is usable by plants and animals. One of the great paradoxes of life is that although Earth is surrounded by an atmosphere that is 79% nitrogen, most organisms cannot use it and many are nitrogen-starved. The nitrogen in the atmosphere is $N_2$, a molecule whose atoms are bonded very tightly because they are linked by three

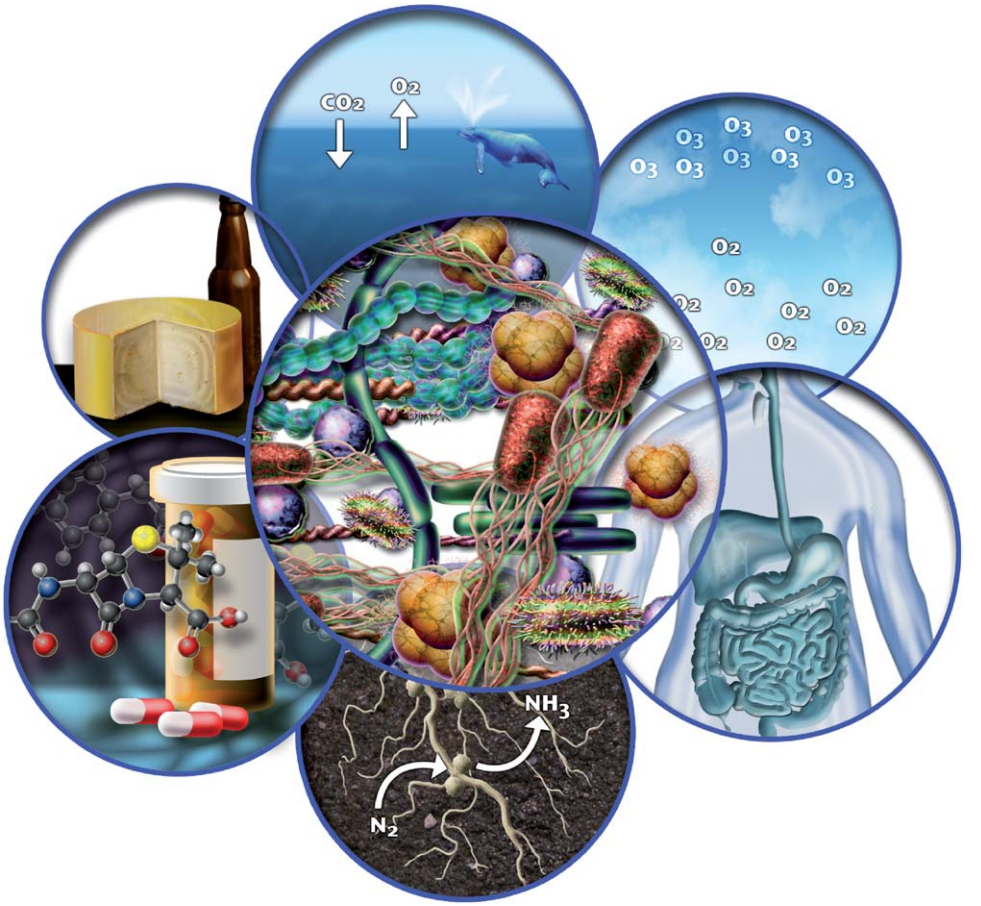



**Figure 1** Examples of services provided by microorganisms. Clockwise from top: microorganisms in the sea remove carbon dioxide from the atmosphere and produce oxygen; the ozone layer in the stratosphere, which protects Earth from harmful ultraviolet rays, was created by the reaction of oxygen molecules produced on the early Earth by microbes; the human body contains more bacterial cells than human cells and most of these live in the gastrointestinal tract; certain bacteria convert dinitrogen gas to ammonia in special structures on the roots of leguminous plants; soil bacteria are the source of most of the antibiotics and many other drugs used in medicine today; microbes ferment many foods; centre, microorganisms come in many shapes and sizes. Graphic: Nicolle Rager Fuller.

covalent bonds. Because this triple bond is strong, it is hard to break and only a few organisms have the biochemical ability to do so. In fact, certain bacteria are the only organisms that can split this triple bond and 'fix' nitrogen. They split the bond and then conduct chemical reactions that incorporate the nitrogen into cellular constituents. These can then be passed through the food chain, providing nitrogen to other microorganisms, plants and animals.

## Maintenance of human health

The microbial communities in and on the human body are essential to our health. Bacteria in the gastrointestinal tract protect us from invading pathogens; when the balance of gut microbial communities is compromised, many diseases, including inflammatory bowel disease, colon cancer and perhaps even obesity and diabetes, are thought to follow. These diseases are not incited by a single organism, but are mediated by community processes. In some respects, the gut community functions as a unit in which each microorganism is dependent on others for its livelihood and the effective functioning of the system is dependent upon interactions and cooperation among its members.

The bacteria in the mouth determine dental health. The mouth contains a complex community that develops in a prescribed succession of organisms that depend upon each other to survive in the mouth. These include bacteria that cause dental caries and periodontal disease as well as those that keep the detrimental bacteria in check. In addition to their local effects on oral health, certain bacteria from the mouth can cause serious heart disease if they get loose in the bloodstream. Understanding human-associated communities – known as the human microbiome – that contain thousands of organisms and millions of genes is one of the major frontiers of research in human health.

The communities associated with the human body are subject to two evolutionary processes. First, they have co-evolved with us over millennia of sustained symbiosis, so that our genes are tuned to support and rely on certain organisms and their genes are tuned to the home we

provide. Second, the communities inside each of us evolve over our lifetimes. Evolution of microbial communities with the human species as well as with each individual produces an intertwined web of dependency and communication. Consequently, these microorganisms must be studied in their community context, not just in pure culture in the laboratory.

### Synthesis of antibiotics

Microorganisms are the source of most of the antibiotics that we use to manage infectious diseases. In the 1920s, Alexander Fleming noted that a fungus produced a chemical that inhibited bacterial growth; the chemical was penicillin. That finding led to a concerted effort to discover and commercialize antibiotics for medical and agricultural uses. Many bacteria, as well as certain fungi, produce antibiotics with a wide range of chemical and functional properties, providing the array of drugs we have available today. It is not known why bacteria and fungi produce antibiotics, but the widespread capability suggests that they are important for microbial life. One favoured hypothesis is that the microorganisms use antibiotics to inhibit growth of their competitors for nutrients; another hypothesis is that they use antibiotics as signal molecules to alert other species to important events within communities. Whatever their roles in microbial communities, antibiotics have transformed human existence by providing outstanding treatment for infectious disease. World War II, the first war during which antibiotics were available, was the first war in history in which more people died of wounds inflicted by weapons than infectious disease.

### Other services

The list of services provided by microorganisms is long. It includes many other examples of nutrient cycling and medical innovations than are presented here. In addition, microorganisms play a role in food production through fermentation, provide industrial enzymes and polymers, clean up toxic waste and protect humans and other animals and plants from disease. Although microorganisms are often associated with disease, pathogens represent only a tiny fraction of the total microorganisms on Earth and the beneficial microorganisms often keep the pathogens in check. As microbial communities are studied with metagenomics, other microbial services are likely to be revealed.
**See also**: Bacteriology; Microorganisms

## Microbial Communities

Most microorganisms live in communities – assemblages of multiple species. Many of these communities are quite complex, containing thousands of interacting members. Within communities, microorganisms compete for nutrients and space and also cooperate to accomplish functions that one species could not by itself. Some of the communities that have been studied in detail include: communities in soil that degrade gasoline, which leaks from buried tanks all over the world; communities in soil that protect plants from disease and provide them with necessary nutrients; communities in the oceans that conduct photosynthesis and regulate the carbon balance of the planet and the community that lives in the human gastrointestinal tract that provides nutrients, assists with digestion and protects the gut from invasion by pathogens.

Because of their ubiquity and centrality to life, microbial communities demand our attention. It will not be possible to understand fully the many services they perform without knowing which organisms are present and how each contributes to community function. One of the daunting challenges in studying microbial communities is gathering information from members that cannot be cultured under standard laboratory conditions. In some environments such as soil and ocean water, as few as 0.01–0.1% of the community members can be cultured. Culturing has been the main practice in bacteriology for over a century, but it has also resulted in gaps in knowledge about the myriad microorganisms that cannot be cultured.

In the 1980s, new methods became available to assess the membership of microbial communities without culturing (Pace *et al*., 1985). One approach provides the nucleotide sequence of the community members' small subunit ribosomal ribonucleic acid (rRNA) genes, which are carried by and provide a signature for every organism on Earth. The analysis of small subunit rRNA genes (16S rRNA for prokaryotes and 18S rRNA for eukaryotes) served to generate awareness of the amazing diversity of organisms in many different microbial communities, but provided little insight into what these organisms were doing. Therefore, new tools were needed to discover and analyse the functions of these disparate microorganisms. Metagenomics presented precisely such a tool.

## Metagenomics – A New Tool and a Way of Thinking

Metagenomics is the genomic analysis of microbial communities (National Research Council, 2007). It involves extracting DNA directly from an environmental sample – e.g. seawater, soil, the human gut – and then studying the DNA sample. Metagenomic DNA is complex since it is a pool of genomes from many different organisms, making its analysis challenging. It can be captured and maintained in a surrogate host, such as *Escherichia coli*, and the genes embedded in the metagenomic DNA can be studied either by determining their nucleotide sequences or characterizing the properties they confer on a surrogate host.

Metagenomics was conceived in the 1990s with various goals in mind (Stein *et al*., 1996; Handelsman *et al*., 1998). It was first and foremost a method to learn about the contributions to the biosphere made by the community members that could not be cultured. It was also designed for

practical gains, such as discovery of new genes and gene products that would lead to new medicinal chemistry, agricultural innovations and industrial processes. These two goals led to two parallel methodological approaches. In the first approach, known as 'sequence-driven metagenomics', DNA from the environment of interest is sequenced and subjected to computational analysis (**Figure 2**). The metagenomic sequences are compared to sequences deposited in publicly available databases such as GENBANK. The genes are then collected into groups of similar predicted function, and the distribution of various functions and types of proteins that conduct those functions can be assessed.

In the second approach, 'function-driven metagenomics', the DNA extracted from the environment is also captured and stored in a surrogate host, but instead of sequencing it, scientists screen the captured fragments of DNA, or 'clones', for a certain function (**Figure 2**). The

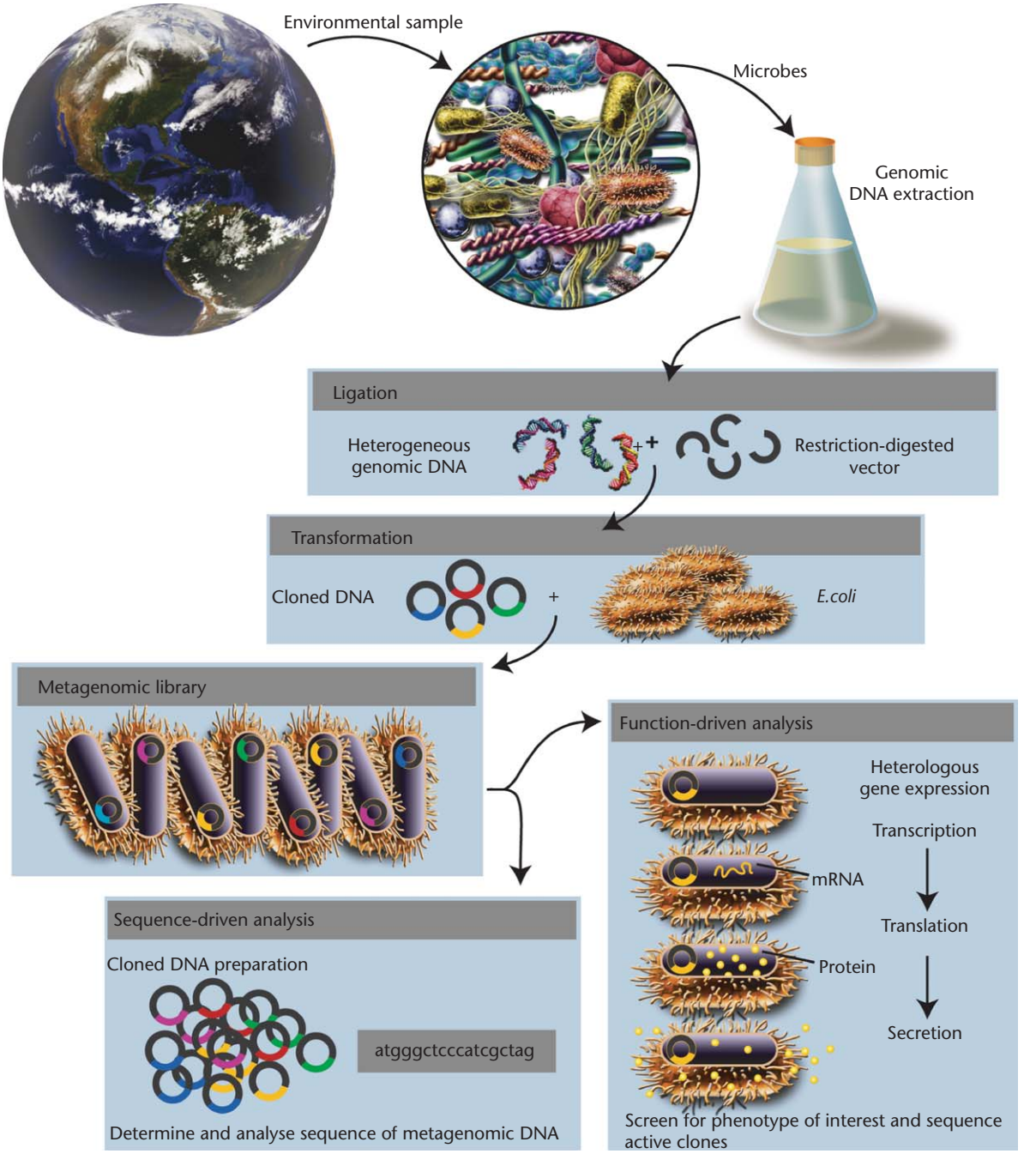



**Figure 2** Construction and analysis of metagenomic libraries. Graphic: Nicolle Rager Fuller.

function must be absent in the surrogate host so that acquisition of the function can be attributed to the metagenomic DNA.

Both approaches have advantages and limitations, and both have been highly effective in informing us about diversity of function of the microbial world. The sequence-driven approach is limited by existing knowledge: if a metagenomic gene does not look like a gene of known function deposited in the databases, then little can be learned about the gene or its product from sequence alone. The function-driven analysis can identify genes that are not related to anything previously studied because genes are identified by their expressed function, rather than sequence, but the drawback is that most genes from organisms in wild communities cannot be expressed easily by a given surrogate host. Therefore, the two approaches are complementary and should be pursued in parallel.

In addition to providing tools to study the unculturable majority of microorganisms, metagenomics presents a new way of thinking. Genomics brought with it a new approach to human genetics that relies on knowledge of all of the genes of the organism and an understanding of some of the interactions among the genes and their products. Genes can now be approached as nodes in a network that modulates the activities and success of the host. Likewise, metagenomics presents a system-level view of microbial communities. Instead of studying single organisms or single functions, metagenomics examines the entire complement of genes in a community, enabling construction of a scaffold of genes and functions on which to build principles about community structure and function.

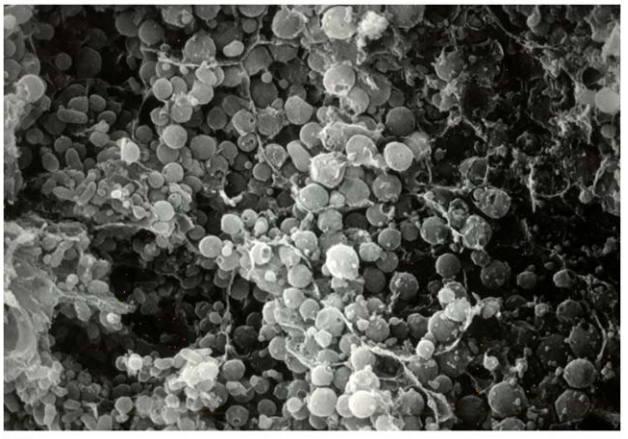
## Metagenomics and symbioses

Many microorganisms that have tight associations or symbiotic relationships with their hosts are difficult to culture away from the host; thus such symbiotic organisms were prime candidates for metagenomics. In a symbiosis between an insect and a bacterium, the aphid and *Buchnera*, metagenomic analysis pointed to the reason that *Buchnera* has been recalcitrant to culturing (Moran and Degnan, 2006). Although this is a simple microbial community – it includes only a single species – it is included as a metagenomic study because it is the first example of genomics on an uncultured microorganism. The analysis of the *Buchnera* genome produced the stunning realization that the bacterium had lost almost 2000 genes since it entered the symbiotic relationship 200–250 million years ago. As a result, it contains a tiny genome consisting of only 564 genes (approximately one-tenth the size of the *E. coli* genome), and does not conduct many of the life functions required for independent growth. Instead, these functions are provided by the insect host. The genomic study of *Buchnera* helped biologists refine their understanding of symbiosis. **See also**: Endosymbionts

Another symbiosis with remarkable biological features is that of the deep-sea tube worm, *Riftia pachyptila*, and a bacterium (Boetius, 2005). These creatures live in harsh environments near thermal vents 2600 m below the ocean surface. In this case, the bacterium lives in a feeding sac inside the tube worm and comprises about one-half of its host's body weight. Although the bacterium has not been cultured, metagenomic analysis has provided insight into this extraordinary symbiosis. The tube worm provides the bacterium with carbon dioxide, hydrogen sulfide and oxygen, which it accumulates from the seawater. The bacterium, in return, converts the carbon dioxide to amino acids and sugars needed by the tube worm, using the hydrogen sulfide for energy (**Figure 3a** and **b**).

The communities of organisms that live symbiotically with human beings, known collectively as the 'human microbiome', are extraordinary in their complexity and influence on health. Metagenomic studies have revealed patterns that were invisible in studies relying on culture-based methods of study. For example, metagenomics and other culture-independent methods have demonstrated that each person carries a unique microbial community in his or her gastrointestinal tract; in fact these communities have been called a 'second fingerprint' because they provide

(a)



(b)

**Figure 3** (a) Tube worms containing bacterial symbionts. Courtesy: National Science Foundation (Courtesy: National Science Foundation). (b) Bacterial symbionts inside the tube worm (Courtesy of Colleen M Cavanaugh). **See also**: Hydrothermal Vent Communities

a personal signature for each of us. Unlike fingerprints, however, microbial communities are dynamic and change over our lifetimes because their composition is influenced by numerous factors. People who live together develop gut microbial communities that share certain features, and identical twins have remarkably similar communities. This indicates that both environment and genetics dictate community membership (Bäckhed *et al*., 2005). The human microbiome is attracting substantial attention from scientists and the National Institutes of Health because of some extraordinary research that suggests that the composition of the human gut microbial community is associated with obesity, heart disease, colon cancer and asthma. The next decade will be an exciting one as research teases apart the roles of our microbial partners in health and disease.

## Metagenomics and geochemical cycles

Metagenomics has provided insight into many habitats and functions on Earth, and only a few can be highlighted here. Metagenomic analyses of seawater revealed some interesting aspects of ocean-dwelling microorganisms. Through these studies, more than one million genes were sequenced and deposited in the public databases. Among these are genes closely related to well-known genes, genes that fall into known gene families but are not closely related to known homologues, and genes that have no known homologues.

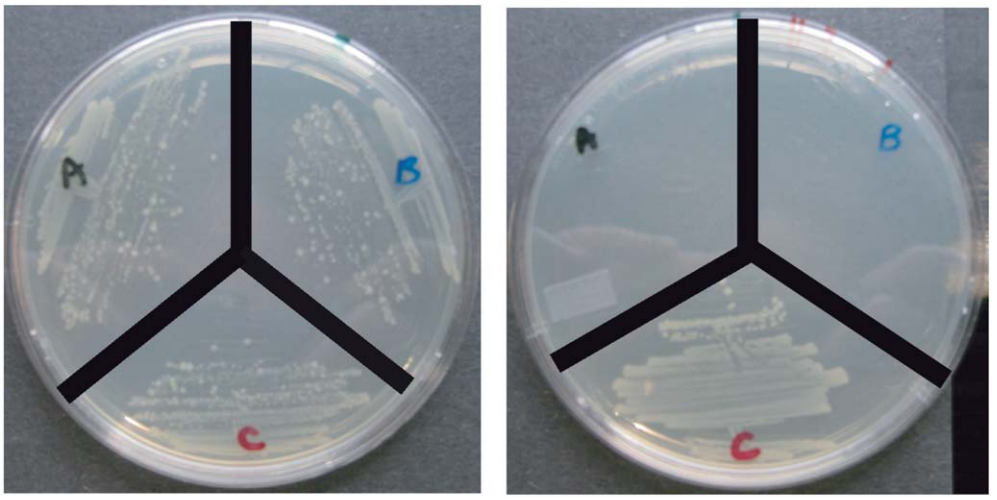
One of the challenges with metagenomic information is to link functions with the organisms in which they normally reside in the community. When a 16S rRNA gene, which provides a taxonomic signature, and a functional gene are found on the same DNA fragment, the prokaryotic organism from which the DNA was isolated can be inferred by the 16S signature (Stein *et al*., 1996). Based on such analysis, some surprises emerged from the seawater metagenomic studies. Groups of bacteria that were not previously known to transduce light energy appear to contain genes for such a function. Rhodopsins, proteins long known to be involved in light harvesting, had been found in cultured Archaea, one of the three major groups of microorganisms. Metagenomics revealed a vast diversity of rhodopsin genes and many were clearly isolated from Bacteria, another group of microorganisms. The grand diversity of these proteins and their presence in both Archaea and Bacteria led to a reassessment of the distribution and activity of light-harvesting complexes in seawater.

Metagenomics produced many surprises in the study of an acid mine drainage sample (Tyson *et al*., 2004). Abandoned ore mines often produce drainage that is highly acidic and rich with toxic metal ions. The pH of the material can be 0 or sometimes even lower! This toxic waste is considered one of the worst by-products of mining. Scientists used methods from classical geochemistry and microbiology combined with metagenomics to study the Iron Mountain Mine in California, leading to a comprehensive understanding of the chemical reactions that produce the toxic waste and of the organisms that catalyse these reactions. The microorganisms in the mine form biofilms, or floating mats, in which they are densely packed together. Metagenomic analysis of the biofilm, including a massive sequencing effort, led to the computer-based reconstruction of the genomes of some of the community members. This was possible because the community is simple – it contains five to seven species – and because highly redundant sequencing was conducted. From the reconstructed genomes and the rest of the metagenomic information arose a model for the cycling of carbon, nitrogen and metals in the acid mine environment. A particularly interesting result was the association of genes for nitrogen fixation, the microbial process of converting nitrogen gas to a biologically assessable form, with a low-abundance organism, *Leptospirillum* group III, which was not previously thought to be a nitrogen fixer. Members of this group had never been cultured, but the knowledge that they fix nitrogen provided a means to culture them. The biofilm was cultured with dinitrogen gas ($N_2$) as the sole nitrogen source and *Leptospirillum* group III grew under these conditions. This is an example of how metagenomics can reveal the unexpected and lead to methods to coax previously uncultured organisms to grow in laboratory culture.

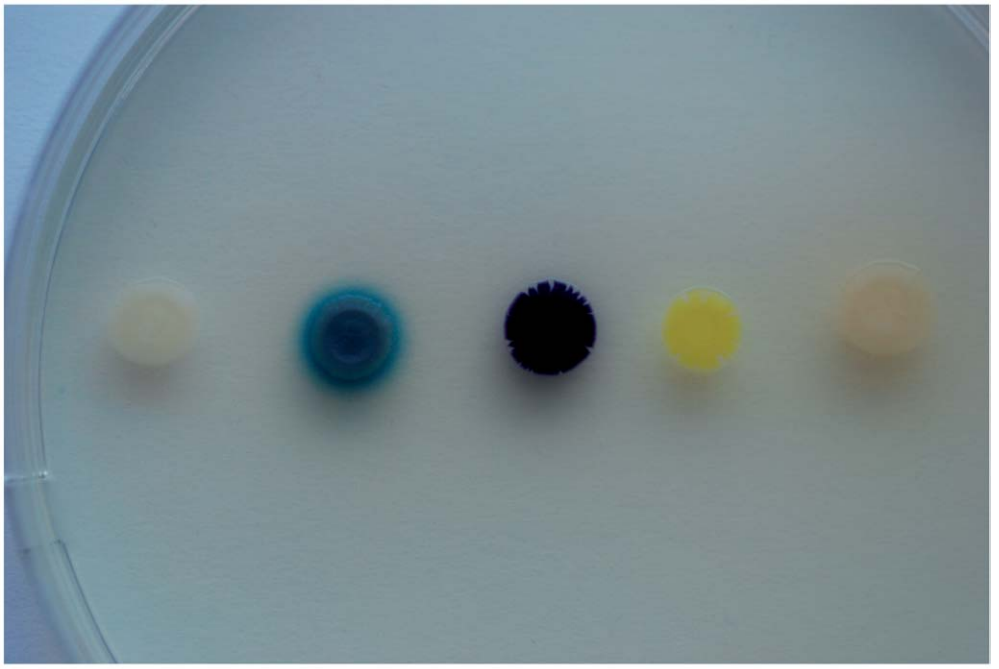
## Metagenomics and antibiotic resistance

Function-driven metagenomics offers a different path to understanding microbial communities. In this version of metagenomics, the clones containing DNA from the community are screened initially for a function or activity they confer on a surrogate host, such as *E. coli*. (**Figure 4a** and **b**). The key advantage of the function-driven approach is that the genes discovered do not have to be part of a family of genes of known function to be informative. In fact, a critical outcome of this approach is the assignment of function to genes of unknown function. This is important to the advancement of metagenomics, and genomics generally, because many genes in the databases have no homologues of known function. These genes can account for as many as 60% of the genes in metagenomic libraries. If the genes responsible for the activity have been captured on a large fragment of DNA, then their sequence and the flanking sequence can provide clues to the identity of the organism from which the DNA originated.

Function-driven metagenomics has led to the discovery of new enzymes and small molecules, such as antibiotics. It has been very productive in the discovery of antibiotic resistance genes from soil. Applying metagenomics to surveying antibiotic resistance genes in soil is of particular interest because antibiotic resistance determinants are believed to be abundant in the environment where antibiotics are produced. Many soil microorganisms are difficult to culture, and although antibiotic resistance is a growing health concern, little is known about its origins in the environment and even less is known about the role of the unculturable microorganisms as reservoirs of resistance genes. Antibiotic resistance is an ideal candidate for metagenomic analysis because clones carrying the resistance




(a)



(b)

**Figure 4** DNA cloned from soil confers many traits on *E. coli*. (a) Left plate contains no antibiotics, right plate contains β-lactam antibiotic. Clone 'C' carries a gene from the soil metagenome that makes *E. coli* resistant to the β-lactam antibiotic. (b) Clone on left contains no heterologous DNA. The other four clones carry DNA fragments from soil metagenomic that confer on *E. coli* the ability to produce pigments. Photos: Heather Allen and Lynn Williamson.

genes can be readily detected in large libraries. If libraries are constructed in a surrogate host bacterium that is sensitive to the antibiotic of interest, then when the libraries are cultured on media containing the appropriate antibiotic, the only clones that should grow are those containing genes for resistance.

Metagenomics reveals a surprising diversity of resistance genes (Riesenfeld *et al.*, 2004). Resistance to aminoglycosides (a group that includes antibiotics such as streptomycin) is conferred by genes that encode enzymes that modify the antibiotics by acetylation. These enzymes are a well-known class of aminoglycoside resistance determinants, but those from the soil metagenome form a new subgroup of their own. Similarly, the soil metagenome contains genes for resistance to β-lactams (the group of antibiotics that includes penicillin) that fall into known classes of resistance genes, but diverge deeply from the genes from cultured organisms. Moreover, certain resistance determinants do not fall into known classes of resistance genes and therefore may represent new mechanisms of resistance. Whether

these genes find their way to clinical settings is unknown, but knowledge of their existence makes it possible to begin tracking their movement.

## Opportunities in metagenomics

Metagenomics is a new field of biology and is full of promise and uncertainty. It provides a window on a world that was unseen before, so it is hard to predict what will be found. A certainty is that it will expand our view of the diversity of genes and functions in the microbial world. It also promises to provide a more complete understanding of the global cycles that keep the biosphere in balance, offer clues to the basis for many diseases, lead to development of new antimicrobial therapies and present solutions to environmental and biotechnical challenges. The expansive genomic landscape presents an opportunity that can be harvested with the new frontier of metagenomic analysis.

## References


Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA and Gordon JI (2005) Host-bacterial mutualism in the human intestine. *Science* **307**: 1915–1920.

Boetius A (2005) Microfauna–macrofauna interaction in the seafloor: lessons from the tubeworm. *PLoS Biology* **3**: e102, doi:10.1371/journal.pbio.0030102.

Handelsman J, Rondon MR, Brady SF, Clardy J and Goodman RM (1998) Molecular biology provides access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry Biology* **5**(10): R245–R249.

Madigan M and Martinko J (2005) *Brock Biology of Microorganisms*, Chaps. 1, 8 and 10. San Francisco, CA: Benjamin Cummings.

Moran NA and Degnan PH (2006) Functional genomics of *Buchnera* and the ecology of aphid hosts. *Molecular Ecology* **15**(5): 1251–1261.

National Research Council (2007) *New Science of Metagenomics Revealing the Secrets of Our Microbial Planet*. Washington, DC: National Academies Press.

Pace NR, Stahl DA, Lane DJ and Olsen GJ (1985) Analyzing natural microbial populations by rRNA sequences. *ASM News* **51**(1): 4–12.

Riesenfeld CS, Goodman RM and Handelsman J (2004) Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environmental Microbiology* **6**: 981–989.

Stein JL, Marsh TL, Wu KY, Shizuya H and DeLong EF (1996) Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology* **178**(3): 591–599.

Tyson GW, Chapman J, Hugenholtz P *et al*. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.


## Further Reading


Delong EF (2007) Sea change for metagenomics? *Nature Reviews Microbiology* **5**: 326.

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* **68**: 669–685.

Rohwer F (2007) Phage metagenomics. *Annual Review of Genetics* **41**: doi:10.1146/annurev.genet.40.110405.090628.

Turnbaugh PJ, Ley RE, Mahowald MA *et al*. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1031.

Woyke T, Teeling H and Ivanova NN *et al*. (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**: 950–955.