Paths of Discovery
Pontifical Academy of Sciences, Acta 18, Vatican City 2006
www.pas.va/content/dam/accademia/pdf/acta18/acta18-vicunia.pdf

MICROBIAL BIODIVERSITY: A NEW VOYAGE OF DISCOVERY

RAFAEL VICUÑA

INTRODUCTION

When I entered basic school exactly half a century ago, I was taught that all living organisms could be ascribed to either the animal or the plant kingdoms. Some years later, it had become evident that this was a rather simple way to assess the extraordinary richness of the biosphere. Hence, the former two kingdoms were extended to five: animals, plants, fungi, protists and monera (bacteria). Moreover, the members of the first four kingdoms could be grouped under the general category of the eukaryotes, with cells possessing membrane systems and organelles that are absent in bacteria. The latter are in turn grouped as prokaryotes. Of note is that this much broader classification does not include viruses. Whether these macromolecular structures should be considered living entities remains a matter of controversy until today.

Of these five kingdoms, plants and animals persist in concentrating most people's attention, including that of a large proportion of scientists. This attitude can be easily understood. When we contemplate nature, the immense variety of mammals, birds, fish, insects, trees, plants and flowers cannot be anything more than overwhelming. We spontaneously tend to associate biodiversity with the life forms that are accessible to our eyes. Biology textbooks contribute to this way of thinking, with illustrations of magnificent whales, lions, antelopes, zebras, tropical fish, ferns, old elms, butterflies, etc. The number of described species seems enormous to us: about one million arthropods, 270,000 vascular plants, 75,000 fungi, 70,000 mollusks, 45,000 chordates and 40,000 species of algae. In contrast, only 5,000 bacterial species have been identified.

However, the recent advent of powerful molecular techniques for the exploration of the microbial world is leading us to the realization that our

criterion to assess biodiversity has been enormously restricted. First of all, we have learned that only between 0.1 and 10 per cent of bacteria normally present in the environment can actually be cultivated under laboratory conditions.[1] Therefore, any projection of total biodiversity on Earth that considered only those bacteria able to grow in culture had to be erroneous. Through both direct and indirect determinations, we are becoming aware that the dominance of bacteria in the biosphere is tremendous. In terms of biomass, the amount of carbon deposited in bacterial cells is at least equal, if not larger, to the total carbon present in animals and plants.[2] According to recent estimates, two-thirds of the bacterial biomass manage to survive within marine sediments, while the majority of the remainder reside in soil and the terrestrial subsurface. Most likely, large quantities of bacteria are also found in aquatic habitats, although they amount to only 1 per cent of the total.[3] To most people, it may come as a surprise to learn that an average human individual harbors as much as 1 kg of bacteria.[4] The number of bacterial cells contributing to this biomass is about 10^{14}, a figure roughly similar to that of the total number of cells composing the human body.

Paths of Discovery

Bacteria were first observed in 1676 by Antonie van Leeuwenhoek (1632-1723), a Dutch amateur microscope builder with little formal education. The primitive microscope designed by van Leeuwenhoek consisted of a lens mounted in a brass plate, adjacent to the tip of an adjustable focusing screw. The name for this novel instrument, although with a different shape, was proposed in 1625 by the physician-naturalist John Faber, who was a member of the *Accademia dei Lincei*: 'The optical tube ... it has pleased me to call, after the model of the telescope, a microscope, because it permits a view of minute things'.[5]

[1] Amann, R.I., Ludwig, W. and Schleifer, K-H., 'Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation', *Microbiol. Rev*. 59, 143-169 (1995).

[2] Whitman, W.B., Coleman, D.C. and Wiebe, W.J., 'Prokaryotes: The unseen microbial majority', *Proc. Natl. Acad. Sci. USA* 95, 6578-6583 (1998).

[3] *Idem*.

[4] Abbot, A., 'Gut Reaction', *Nature* 427, 284-286 (2004).

[5] Cited by Daniel J. Borstin in *The Discoverers* (Random House, NY, 1983).

The Royal Society of London, which by that time had been established in England for the communication of scientific work, published a series of letters describing van Leewenhoek's work, up until his death in 1723. Microscopic fungi had been discovered earlier by Robert Hooke (1635-1703), a British scholar of broad scientific interest and founding member of the Royal Society. Hooke, who also coined the word 'cell' for the first time after observing the texture of cork, published in 1665 a book entitled *Micrographia, or some physiological descriptions of minute bodies made by magnifying glasses with observations and inquiries thereupon.*[6] Apparently, this book had strong influence in van Leeuwenhoek's subsequent discoveries and therefore it would be fair to say that Hooke and van Leeuwenhoek share the credit for founding microbiology.[7]

Due to the lack of innovation in the design of new microscopes, early progress in microbiology was slow. It was only after the middle of the nineteenth century that a change in circumstances contributed to move forward this young science. The German botanist Ferdinand Cohn (1828-1898) made the first attempts to classify bacteria and introduced the use of cotton plugs to prevent contamination of sterile culture media. Also by that time, the French chemist Louis Pasteur (1822-1895) and the English physicist John Tyndall (1820-1893) provided definitive proof that micro-organisms do not arise as a result of spontaneous generation from lifeless matter. This was not the only controversial issue that Pasteur helped to clarify. He also demonstrated that decomposition of food and other organic materials were not merely chemical processes, but required the participation of micro-organisms. Moreover, fermentations revealed the remarkable fact that life can proliferate in the absence of oxygen. Tyndall and Cohn further discovered the formation of spores resistant to heat, a finding that led to the improvement of sterilization procedures.

However, proof that micro-organisms cause disease provided the greatest drive for the development of microbiology. Discoveries in sanitization procedures, made earlier by the British surgeon Joseph Lister, found their scientific support in 1876, through the conceptualization of the 'germ theory of disease' by Robert Koch (1843-1910). The so-called Koch postulates were formulated to demonstrate that a specific type of micro-organism causes a particular disease. Another key contribution of this German coun-

[6] J. Martín and J. Allestry, Printers to the Royal Society, London.

[7] Gest, H., 'The discovery of micro-organisms revisited', ASM News 70, 269-274 (2004).

try doctor was the development of inventive methods for obtaining the growth of bacteria in pure cultures. Previously, Pasteur had used rather simple liquid media to grow fermentative micro-organisms. However, Koch by adding new ingredients modified these media to establish nutrient broths, in both their liquid and solid versions, suitable for growing disease-producing bacteria.

While the role played by micro-organisms as agents of infectious diseases was of central interest in the late decades of the nineteenth century, studies on the participation of bacteria in the cycle of matter on earth began to receive growing attention. Relevant scientists in this issue were the Russian Sergius Winogradsky (1856-1953) and the Dutchman Martinus Willem Beijerinck (1851-1931), who discovered that the metabolic versatility of bacteria is much broader than that of plants and/or animals. Bacteria were shown to grow in completely inorganic cultures and to play an irreplaceable role in the carbon cycle. These scientists went on to show that bacteria are responsible for the fixation of atmospheric nitrogen into organic compounds. Another major contribution of Winogradsky and Beijerinck was the concept of enrichment culture, a miniature application of the principle of natural selection. It consists of a culture of defined composition that is inoculated with a sample containing a complex microbial population. After a few generations, those bacteria having the ability to grow faster in this specific medium will predominate. Enrichment cultures have been utilized since to isolate a large variety of bacteria with specific nutrient requirements.

The experimental methods developed by the afore-mentioned pioneers of microbiology allowed us to gain insight in a new world previously inaccessible to the naked eye. During the twentieth century we learned that bacteria show an amazing ability to adapt to all sorts of environmental conditions. They can grow at temperatures as high as 106ºC (*Pyrolobus fumarii*) or as low as 4ºC (*Polaromonas vacuolata*). Some require a very acidic pH to proliferate (*Picrophilus oshimae*), whereas others prefer extreme alkaline settings (*Natronobacterium gregory*). Bacteria have been shown to withstand pressures of up to 700 atmospheres (MT41) and to require salt concentrations above 25% (*Halobacterium salinarum*). Just a few months ago, *Nature* magazine reported the presence of a community of bacteria living in a lake beneath an Icelandic glacier.[8] This lake, placed inside the crater of

[8] *Nature* news published on line (13 July 2004).

the Grimsvötn volcano, is 100 meters deep and it is covered by 300 meters of a thick ice-sheet. This abundance of extremophiles makes us wonder about the possibilities of extraterrestrial microbial life. Paradoxically, there seems to be a place on the surface of the Earth where there is no life at all, namely, the Atacama Desert in Chile.[9]

Bacteria also became the organisms of choice to elucidate different metabolic pathways as well as to unveil the mechanisms of gene expression. These studies showed that in terms of metabolic versatility, bacteria are unsurpassed. As previously mentioned, only bacteria have the capacity to reduce the rather inert molecule of nitrogen, a task performed with a large expense of chemical energy. Most of the methane or 'natural gas' encountered in the outer few kilometers of the Earth's crust or in the atmosphere is the product of bacteria that use carbon dioxide as electron acceptor in their metabolism. On the other hand, bacteria conduct one-fifth of the photosynthesis occurring in the planet.[10] On a yearly basis, this represents twice as much as the fossil fuel energy used by human consumption the world over.[11] Some bacteria can obtain their energy by oxidizing inorganic molecules such as H_2, H_2S, Fe^{+2}, NH^{+4} and uranium^{+4}, among several others. Bacteria are able to consume the hydrocarbons present in oil, using them as carbon source for survival. They also have the capacity to degrade highly toxic xenobiotic compounds, such as pesticides, polychlorinated biphenyls, munitions, dyes and chlorinated solvents. Thus, as stated by the late Stephen J. Gould, 'bacteria represent the great success story of life's pathway. They occupy a wider domain of environments and span a broader range of biochemistries than any other group. They are adaptable, indestructible and astoundingly diverse'.[12]

A Hidden Microbial World

In spite of the great usefulness of pure cultures, the fact that most bacteria defy cultivation in the laboratory markedly restricted our knowledge of the microbial world. A key breakthrough which delivered a profound

[9] C. McKay, as quoted by J. Whitfield in *Nature* 430, 288-290 (2004).

[10] Dr. Dave J. Scanlan, personal communication.

[11] Lehiniger, A.L., Nelson, D.L. and Cox, M.M., *Principles of Biochemistry*, 2nd edition (Worth Publishers, New York, 1993).

[12] Gould, S.J., 'The evolution of life on earth', *Sci. Am.* 271(4), 85-91 (1994).

impact on our perception of microbial diversity took place about 30 years ago with the introduction of molecular sequences as a criterion to relate organisms. Because the number of sequence differences in a molecule is proportional to the number of stable mutational changes fixed in the encoding DNA, evolutionary distances can be measured by differences in the nucleotide or aminoacid composition in homologous nucleic acids or proteins, respectively. Based on this concept, the American microbiologist Carl Woese proposed the use of ribosomal RNA (rRNA) sequences as evolutionary chronometers. Ribosomes are cytoplasmatic particles made up of two subunits containing RNA and protein, which are part of the protein synthesizing machinery of the cell. Ribosomal RNAs constitute an adequate macromolecule for measuring phylogenetic distances because they are functionally constant, universally distributed and moderately well conserved. By aligning small-subunit RNA sequences from various organisms and counting the corresponding nucleotide differences among them, Woese constructed a phylogenetic tree that could be used to relate all organisms in the biosphere. Hence, it became known as the tree of life. Woese's tree showed three primary lines of evolutionary descent that span across the formerly known five kingdoms of living organisms. These lines are called urkingdoms or domains, and comprise the Eucarya (eukaryotes), Bacteria (previously called eubacteria) and Archea (initially called archeobacteria).

In the following years, complete genome sequences confirmed the existence of Archea as a separate evolutionary line that branched from the Eukarya. In addition, widespread sequencing revealed several short rRNA sequences that are unique to certain groups of bacteria. Linked to a fluorescent dye, these so-called signature sequences have been widely used as probes for the detection of bacteria (FISH: fluorescent in situ hybridization). Woese's tree has been now widely accepted, although it must be taken with some caution. First, due to multiple events of lateral gene transfer or even to intermixing of genomes in the course of evolution, the rRNA tree is not always congruent with phylogenetic trees based on metabolic genes. In addition, not all rRNAs are detected with the same sensitivity (the universal primers used in PCR experiments do not amplify all rRNA genes with the same efficiency). On the other hand, due to rRNA genes varying in number among prokaryotes, estimations of species diversity and abundance in a particular sample may not be highly reliable.

A simple glimpse at the tree of life makes it evident that the large organisms in which scientists have concentrated their attention constitute an extremely limited fraction of the diversity of life. For example, the related-

ness of the human lineage to any kind of insect is closer than that of any two bacteria belonging to different phylogenetic groups. The tree also lends support to the hypothesis that mitochondria and chloroplasts derive from bacterial symbionts, most likely proteobacteria and cyanobacteria, respectively. On the other hand, the tree shows that the eukaryotic nuclear line extends as deep in the history of life as that of bacteria, and that it later gave rise to the archeal lineage.

By 1987, Woese and colleagues had completed a tree with cultivated micro-organisms delineating 11 major phyla or lineages,[13] which thereafter became 12 as a result of the separation of the gram-positive into two branches (*Firmicutes* and *Actinobacteria*). Among these original phyla are the gram-negative *Proteobacteria*, which include *Escherichia* and *Pseudomonas*, as well as the photosynthetic *Cyanobacteria*. Since then, about 14 additional phyla of bacteria growing in cultures have been defined. Several of these (*Thermodesulfurobacteria*, *Coprothermobacteria*, etc.) are thermophilic.

By the same time Carl Woese was sorting out his tree of life, his colleague Norman Pace proposed the use of modern molecular techniques to gain knowledge of the microbial diversity in natural environments. His rationale was based on the fact that with the sequence-based taxonomic criterion, a sole gene sequence suffices to ascribe an organism to a known phylum or to define a new one. Ribosomal RNA genes from uncultivated micro-organisms can be obtained directly from environmental samples by amplification with the polymerase chain reaction (PCR). Due to the relatively high conservation of rRNA genes, primers can be designed in such a way that they will anneal to sequences that are shared by representatives of all three domains. The fragments thus obtained are resolved by cloning and then sequenced. This approach has proven highly successful, leading to the identification of about 26 new phyla that contain no known cultivated representatives. These are found in a variety of habitats and some are highly abundant, especially in the previously unexplored Earth's crust. Most of the latter rely on energy provided by redox reactions of inorganic compounds, as opposed to those that depend either on the harvesting of sunlight or on the metabolism of organic compounds as sources of energy. Most often, the new lineages are distantly related to previously characterized ones and the shape of the tree reveals that bacterial diversity arose as a result of a radiation of lineages rather than from a sequential divergence from a main line.

[13] Woese, C.R., 'Bacterial evolution', *Microbiol. Rev*. 51, 221-271 (1987).

To-date, more than 60,000 small subunit rRNA sequences from a wide variety of prokaryotes have been reported.[14]

Further Innovations of the Molecular Methods of Analysis

The new molecular techniques proposed by Pace have become highly refined. They have even been focused to decipher functional information of the unmasked microbes, a characteristic that is rarely provided by phylogeny based on rRNA. In a recent publication,[15] Craig Venter and collaborators analyzed bacterial samples from the Sargasso Sea, a nutrient-limited, open ocean environment, by applying a whole genome shotgun sequencing method. In this approach, total DNA from a complex bacterial mixture is isolated and sheared into tiny fragments that are cloned and sequenced from both ends. Based on the sequences obtained, the fragments are pieced back together into their proper genomic arrangement with computer guidance. In this study, Venter and colleagues obtained more than 1,000 billion base pairs of non redundant sequences, equivalent to approximately 775 complete microbial genomes. Based on sequence relatedness, they estimated that the DNA analyzed was derived from 1,800 genomic species, including 148 previously unknown bacterial phylotypes. Other phylogenetic trees constructed with markers such as the elongation factor Tu or the heat shock protein 70 gave more conservative estimates, ranging between 341 and 569, of species richness.

This difference in the number of species obtained by using diverse criteria brings us to a complex issue: the definition of species in the prokaryotic world. For most animals, plants and fungi, a species corresponds to an interbreeding population. However, although bacteria interchange genetic material by various means, they do not have formal sex. Therefore, identification of bacterial species is a matter of consensus among microbiologists. A common rule of thumb is that two bacteria classify in the same species if their identity in rRNA is higher than 97%. Common phenotypic

[14] Cole, J.R., *et al.*, 'The ribosomal database project (RDP-II): previewing a new autoligner that allows regular updates and the new prokaryotic taxonomy', *Nucleic Acids Res*. 31, 442-443 (2003).

[15] Venter, J.C., *et al.*, 'Environmental genome shotgun sequencing of the Sargasso Sea', *Science* 304, 66-74 (2004).

characters and overall genomic coherence serve as additional criteria to confirm the rRNA sequence based diagnostic.

The study conducted by Venter and collaborators also led to the identification of 1.2 million previously unidentified predicted genes, of which about 70,000 are novel. This represents about an order of magnitude over the number of sequences presently archived in public databases such as SwissProt and REM-TrEMBL (about 140,000). The predicted genes could be classified by function, i.e. central intermediary metabolism, DNA metabolism, transcription, signal transduction, transport, etc. Due to the large heterogeneity of the microbial population, reconstruction of near complete genomes was not possible. However, the identity of some genes provided key information about the physiological characteristics of the community. For example, it identified the dominating cyanobacteria species performing photosynthesis (*Prochlorococcus*). Also, the identification of about 800 new genes encoding bacteriorhodopsin confirmed that coupling of light energy harvesting and carbon cycling through a non-chlorophyll based pathway is an important biochemical process in the ocean.

A similar shotgun approach can be applied for objectives that go beyond bacterial identification. For example, Gene Tyson *et al.*[16] have reassembled genomes from microbes occurring in an acidic ecosystem. Analysis of each reconstructed genome led to the identification of the corresponding bacteria, but also revealed pathways for carbon and nitrogen fixation, as well as for energy generation. It also provided insights into survival strategies in this extreme environment. In this particular case, this strategy was successful due to the small number of species population and the low frequency of genomic rearrangements, gene insertions and deletions. Sequencing of only 76 million base pairs sufficed to reconstruct two near complete genomes and three other partial genomes of the bacteria thriving in this habitat.

The culture-independent genomic analysis of microbial communities has been termed metagenomics.[17] It involves extracting DNA directly from some habitat, cloning it into suitable vectors and transforming it into a culturable host cell. If the aim is to study genome organization and genes encoding metabolic pathways, libraries are constructed in vectors that hold

[16] Tyson, G.W., *et al.*, 'Community structure and metabolism through reconstruction of microbial genomes from the environment', *Nature* 428, 37-43 (2004).

[17] Schloss, P.D., and Handelsman, J., 'Biotechnological prospects from metagenomics', *Curr. Opinion Biotechnol*. 14, 303-310 (2003).

large fragments of DNA that are subjected to sequencing. Alternatively, functional expression of individual genes can be attained by constructing libraries in expression vectors that allow insertion of the foreign gene next to a promoter. Thus, metagenomics aims to establish a link between phylogeny and function. Functional analysis has the potential to expand the range of known functions and elucidate functions of genes with no known homologs. Metagenomic analysis has been particularly useful to study soil micro-organisms. Several genes and gene products have been discovered in DNA extracted directly from soil, and of particular interest are those involved in pathways for biosynthesis of or resistance to several novel antibiotics.[18] In spite of the promising results obtained in this function-driven analysis, the construction of metagenomic libraries in expression vectors faces some drawbacks. Firstly, some *Bacteria* and *Archea* are resistant to lysis and therefore their DNA is not represented in the libraries. Secondly, some genes or their products are toxic to the host cell, typically *Escherichia coli*, or the foreign gene fails to be expressed due to its phylogenetic distance to the host cell machinery.

Concluding Remarks

New molecular techniques are manifesting the short-sightedness in our estimations of bacterial diversity and versatility. In spite of their small size and limited amount of genetic material, bacteria have evolved incredible sophisticated strategies to survive and proliferate in the most varying of environmental niches.

Analysis of DNA from uncultured micro-organisms has doubled the number of bacterial phyla and although major lineages will probably increase still further, the rate of expansion of the tree of life will probably be less explosive than in recent years. However, in spite of the importance of reaching an accurate knowledge of biodiversity, it is likely that new efforts will concentrate on the identification of novel genes. The ultimate goal should be to identify all genes in the biosphere. Because all bacteria share a certain percentage of genes, as more complete genomes and shotgun sequences become available, knowledge of the global genetic comple-

[18] Handelsman, J., *Soils – The metagenomic approach, Microbial biodiversity and bioprospecting*, A.T. Bull (ed.), ASM Press, Washington, pp. 109-119.

ment will be approached asymptotically. Major advances are expected to come in the functional predictions based on genome sequences. Although at first it may seem unobtainable to assign a specific role to each of such a vast array of novel genes, this task may be less daunting as this discovery of genes is likely to be far smaller than the variety of microbial species.[19]

We may wonder how realistic is it that we will ever reach an accurate assessment of Earth's microbial biodiversity. The massive lateral transfer of genes and the bacteria's rapid adaptation to an every changing environment represents great challenges for the future. However, it has become clear, on this new voyage to discovery, that we have encountered and will encounter still, more wonders than Antonie van Leewenhoek could ever have dreamed of, when he first took his primitive microscope and ventured upon this new world.

[19] Cases, I. and de Lorenzo, V., 'The grammar of (micro)biological diversity', *Environ. Microbiol*. 4, 623-627 (2003).