INTERNATL MICROBIOL (1999) 2:1–2

EDITORIAL

Enrique Herrero

What is microbiology gaining from genomics?

Correspondence to:
Department of Basic Medical Sciences.
School of Medicine. University of Lleida.
Rovira Roure, 44. 25198 Lleida. Spain.
Tel.: +34-973702409. Fax: +34-973702426.
E-mail: enric.herrero@cmb.udl.es

In recent years, a large number of microbiologists have expressed their worries about the decreasing role of microbiology as a biological discipline in the context of the academic world. An example of this tendency can be seen in American universities as the diverse microbiology departments merge into larger units. Carl Woese claimed some years ago that the "denaturation of microbiology was due to the invasion of molecular or cell biology-based approaches into the heart of this discipline" (*Microbiol. Rev.* 58:1–9, 1994). He emphasized the need for a return towards a discipline standing at the basis of all biological sciences. In other words, he longed for a discipline that would exalt the diversity of living strategies found in the microbial world and where phylogeny-based thinking would impregnate all studies. While in the 1960s and 1970s microbiology (or perhaps the study of *Escherichia coli*) presented the groundwork for the subsequent expansion of molecular biology, at present some funding agencies (and many of our undergraduate students!) contemplate microbiology as little more than an insignificant subject which cannot be compared to other disciplines more directly related to human molecular biology.

Jacques Monod once proposed that "when we understand *E. coli* we will understand elephants". Two decades ago it could have been reasonably foreseen that once in depth knowledge of the molecular aspects of *E. coli* and other "standard" bacteria had been attained, the interests of most molecular biologists would obviously turn towards higher organisms (even though most *E. coli* molecular biologists would claim this as a fallacy). Although this is what has generally happened, we should acknowledge that true comensalism or symbiotic associations have been established between microbiology and other biological disciplines. For example, it is obvious that molecular biology has deeply revolutionized clinical microbiology diagnostics; or that research on the molecular basis of microbial pathogenesis benefits from studies on the biology of "higher" cells. Those dramatic changes will probaby lead to consider cellular microbiology as a new discipline. Perhaps it is high time to realize that microbiology is no longer at the center of the biological universe!

But now genomics has emerged. By the end of 1998, seventeen prokaryotic genomes had been completely sequenced (13 from Bacteria and 4 from Archaea), along with the genome of two Eukarya, the yeast *Saccharomyces cerevisiae* and *Caenorhabditis elegans*, an animal. During 1999 as many as fifteen sequences of bacterial genomes and three of an archaeon are expected to be released (see http://www.tigr.org/tdb/mdb/mdb.html), besides an additional number of microbial sequences that may be determined by private institutions without publicly releasing their information. Although a bias exists towards preferentially sequencing genomes from bacterial pathogens or extremophiles, free-living mesophiles (*Bacillus subtilis*) and autotrophs (*Synechocystis* sp.) are also on the list of the seventeeen already performed. Thus, microbial biodiversity is reasonably represented in the range of sequenced genomes. It is a current opinion in specialized forums that microbial genomics will have an enormous impact in such areas as industrial microbiology or in the fight against infectious diseases. However, it is also emphasized that more potent tools will be required to deal with the information being generated. Not only will it be necessary to manage the enormous volume of data, but also to uncover new sequences or structural traits to which they must adscribe potential functions. Conventional studies on protein chemistry, molecular genetics and microbial physiology, among others, will continue to be essential.

But how do these great expectations on microbial genomics accommodate to the present reality? The case of the *S. cerevisiae* genome may be taken as an example. It is obvious that the publication of the complete sequence of the yeast genome (*Nature* 387 (suppl.):5–105, 1997) has changed many aspects of the techniques of the yeast researchers community. Cloning and characterization of genes based on complementation of a particular phenotype is already a "prehistoric" practice in the history of yeast molecular biology. Genes and disruptant mutants can be isolated in a few days, and reverse genetics makes it much easier to study particularly complex biological processes. As it was initially designed, the technology of the high density DNA arrays on chips, itself a direct consequence of the yeast DNA sequencing project, has revolutionized the studies on yeast gene expression in specific growth conditions and/or genetic backgrounds. More importantly, the technology can be extrapolated to other areas such as diagnostic microbiology. Easy

automatization of the technique may facilitate its use in clinical laboratories. Unfortunately, however, the function of more than one third of the close to 6000 genes of *S. cerevisiae* remains to be determined.

Systematical determination of the function of genes discovered through a sequencing project is carried out mainly through one of these two strategies: either the phenotypic study of mutants for a particular gene or the determination of sequence homologies with orthologous genes from other organisms for which the function is known. The presence of short motives in the protein sequence with adscribed biological activity may also be useful for function determination. But several difficulties arise in this prospect. Genetic redundancy (a common trait in the yeast genome) may hide the phenotype alterations of individual mutants inside a gene family. Overexpression of a gene may not result in an obvious phenotype that could be related to a function. Most importantly, many biological activities result from the functional interaction of a number of gene products. This biological complexity difficults the unequivocal establishment of a relationship between mutation and phenotypical alteration. With respect to sequence comparisons, gene divergence may result in non-homologous sequences sharing the same function (more potent sequence comparison programs may help to solve this difficulty). The contrary may also happen, and significant homologies affecting internal regions of larger protein molecules do not always guarantee sharing similar functions. Sharing diverse functions in a single polypeptide molecule difficults comparative analyses.

In the four years since the first prokaryotic genome was fully sequenced, bacterial genomics has been undoubtly useful in establishing a genetic basis for the biological properties of particular microbial species (see *Curr. Opin. Microbiol.* 1:562–566, 1998, for a more general discussion). For example, that parasites usually have smaller genomes (and therefore a more limited collection of genes) than opportunistic bacteria can be traced to the fact that they colonize a more limited range of ecological niches. The initial analysis of the *Mycobacterium tuberculosis* genome sequence (*Nature* 393:537–544, 1998), revealed the existence of an unusually large number of genes involved in lipid biosynthesis and degradation. This remarks the importance of lipid metabolism in the life cycle of this human parasite, and therefore, suggests possible therapeutic targets against it. The sequence analysis also showed the existence of two previously unknown protein families which contain a large number of members. Their structure suggests an antigenic variation which could explain how the parasite evolved into a coexistence with the host's immune system during chronic infections. Comparison of the genomes of *Mycobacterium tuberculosis* and *Mycobacterium leprae* show that although both species share a common ancestor, *M. leprae* contains many silenced genes and has lower gene density. This genetic shrinkage could explain the extreme host dependence of the leprae bacillus. Differences in isoniazide sensitivity between both bacilli can also be explained at the molecular level.

Comparative genomics has also confirmed the independent evolutionary root of archaea with respect to bacteria and eukaryotes. These studies have demonstrated that the proteins involved in archaea's primary metabolism are closely related to those in bacteria, while the archaea proteins involved in DNA replication and gene transcription and translation display a closer sequence relationship with eukaryotes. Nevertheless, plasticity of the genomes by horizontal gene transfer may difficult the establishment of evolutionary trees based on single gene homologies. In some cases, comparative genomics does not reveal any differences in pathogenicity properties nor in life strategies. For example, the *Mycoplasma pneumoniae* genome contains 236 kilobase pairs and about 300 open reading frames more than *Mycoplasma genitalium*. However, substractive comparative analysis does not help to explain the differences in tissue specificity between both species during host invasion (*Curr. Opin. Microbiol.* 1:572–579, 1998). Since substractive analysis may be crucial to uncover putative pathogenic genes and drug targets against virulent species, as well as possible immunoreactive antigens, it will be essential to know the results of these studies when applied to the genome of a virulent *Mycobacterium tuberculosis* strain against a non-virulent one, or to sequences of laboratory and clinical *Candida albicans* strains.

In summary, microbial genomics is providing us with many information pieces not always obviously interrelated. Sometimes these pieces help to explain the genetic basis of some biological properties and of the differences among species. Phylogenetic relationships could be clarified and better understood with this new information. In other cases, the information can be directly used for biotechnological processes, which might explain the efforts devoted to sequencing thermophile species. But unless new bioinformatic tools come to completely replace more conventional approaches (and I must admit I am skeptical about it), microbial genomics should be envisaged as an enormous source of data on which more sophisticated, less systematic studies to understand microbial processes should be applied. Microbial genomics is facilitating our task, and it will probably guide us in the best way to carry out different studies, but we will have to rely on other microbiological skills to put together the pieces of the puzzle.

Enrique Herrero

Professor of Microbiology at the School of Medicine of the University of Lleida, he works on yeast molecular biology, having participated in the European Union-funded projects on sequencing and functional analysis of the yeast genome. He was awarded the second Biennal Prize of the Spanish Society for Microbiology, in 1987.