Genomics
Editorial overview

Howard Ochman

Addresses
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85521, USA; e-mail: hochman@u.arizona.edu

Current Opinion in Microbiology 1999, **2**:517–518

1369-5274/99/$ – see front matter

Many microbiologists, evolutionary biologists, and geneticists maintain a distorted view of bacterial genomics because it appears that many of the conclusions from such studies could have been resolved with more limited sequence information, or by applying an alternative (i.e. less expensive) approach. This view is further compounded by the fact that once one has determined the complete genomic sequence of an organism, there seems to be nothing left to do except to move on to the next strain or species. And, finally, there are the concerns about the amount of information generated from such studies: how will it be possible to manage, annotate and disseminate — let alone interpret — these data at anywhere near the rate that they are being produced?

The topics considered in evolutionary and comparative of genomics presently fall into two broad categories: local problems, involving the analysis of individual genomes, or those of closely related organisms; and global problems, focusing on comparisons over a vast phylogenetic scale. As little as one year ago, it would have been difficult to investigate even the most simple of the local problems because the sequenced organisms were too distantly related, and their genomes too divergent, to allow for meaningful comparisons. Such local problems have traditionally been addressed by examining limited portions of the genome, or by techniques that did not involve nucleotide sequencing. However, the complete sequences of genetically distinct strains within a species, such as *Helicobacter pylori*, *Escherichia coli* and *Salmonella enterica*, have, or will soon, become available, and this will allow for a direct appraisal of the rates and patterns of chromosome evolution in bacteria.

Because of phylogenetic distribution of microorganisms for which complete sequence information has already appeared, there have been numerous studies concerning some of the more global issues in the genome evolution of bacteria. In this regard, two papers, which have not been discussed in any of the articles in this issue, deserve special mention.

The first appeared well over a year ago [1], and is still the best example of how the evolutionary history of bacterial chromosomes can be reconstructed from complete genome sequences. In this paper, the authors set out to examine the fraction of genes shared among organisms, as well as gene order and their relative positions, in relation to phylogenetic distance. This seems to be a rather simple task; however, the identification of orthologous genes (i.e. genes whose independent evolution result from speciation rather than duplication [2]) from such divergent organisms is not at all straightforward. After proceeding through numerous cautionary measures, Huynen and Bork settled on a set of 34 orthologs that were shared among all organisms compared, and established the relative rates of genome evolution at several levels of organization. Over a broad evolutionary timescale, higher order features of the genome, such as gene order and the number of shared orthologs, decay faster than the degree of protein identity. This reinforces the need to examine closely related organisms in order to analyze such less conserved features of the genome.

It would be logical to assume that genes that are shared among all organisms represent those essential to cell survival, and this reasoning led Arigoni *et al.* [3] to study the role of conserved open reading frames (ORFs) of unascertained function in *Escherichia coli*, *Bacillus subtilis* and yeast. First, these authors identified 26 genes of unknown function common to *E. coli* and *Mycoplasma genitalium* (which possesses only about 10% of the coding capacity of *E. coli*), and, by constructing in-frame deletions of each region, they recovered six that were essential to *E. coli*. This gene knockout strategy was extended to the homologs of these six ORFs in *B. subtilis* and *Saccharomyces cerevisiae*, and the results were indeed surprising. Five of the six *E. coli* homologs were essential to *B. subtilis* and only one of the six was essential to yeast; but, inexplicably, the single gene that was non-essential to *B. subtilis* was the only one of these six conserved genes found to be essential to yeast. Apparently, evolutionary conservation does not serve as a reliable guide to the indispensability of a sequence since these organisms harbor different sets of genes that specify redundant functions.

These papers, along with the contributions to this issue, illustrate the diversity of ways that the information assembled through large-scale sequencing projects has been applied. Despite relatively little variation among these microorganisms in the overall amount of genetic information — the genome sizes of the organisms covered in this issue span but a single order of magnitude — the novelty in genome structure, content and organization is remarkable. In sum, these studies have shown that by looking within and beyond the sequence of nucleotides that constitute a genome, it is possible to establish the factors

molding and diversifying genomes, and to discover new principles about the biology of microorganisms.

References

1. Huynen MA, Bork P: **Measuring genome evolution.** *Proc Natl Acad Sci USA* 1998, **95**:5849-5856.

2. Fitch WM: **Distinguishing homologous and analogous proteins.** *Syst Zool* 1970, **19**:99-113.

3. Arigoni F, Talabot F, Peitsch M, Edgerton MD, MeldrumE, Allet E, Fish R, Jamotte T, Curchod M-l, Loferer H: **A genome-based approach for the identification of essential bacterial genes.** *Nat Biotechnol* 1998, **16**:851-857.