

Meeting report

The accelerating convergence of genomics and microbiology

Gary K Schoolnik

Address: Departments of Medicine and of Microbiology and Immunology, Stanford Medical School, Beckman Center, Stanford, CA 94305, USA. E-mail: Schoolnik@cmgm.stanford.edu

Published: 4 April 2001

Genome Biology 2001, **2**(4):reports4009.1–4009.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/2/4/reports/4009>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

A report from Microbial Genomes, a joint conference of the American Society for Microbiology and the Institute for Genomic Research, Monterey, California, USA, 28-31 January, 2001.

The rapid use of available genome sequences by biologists was brilliantly evident in many of this meeting's plenary lectures. The most pervasive theme was the relationship between genome sequence and adaptation of a microbe to a specialized niche. It is particularly instructive in this regard to contrast the largest and smallest bacterial genomes discussed.

The complex and resourceful life style of Myxobacteria as free-living soil organisms (Dale Kaiser, Stanford University, USA) includes organized social behavior during food-seeking excursions and, in periods of scarcity, the formation of fruiting bodies containing spores that can be carried by the wind to richer habitats. Fruiting body development has attracted the attention of developmental biologists because it provides an opportunity to study spatial differentiation in a genetically tractable and relatively simple organism. Recent information about the surprisingly small size of the human genome notwithstanding, it is reasonable to assume that complexity and genome size should be related. And so it seems to be: Barry Goldman (Monsanto, St Louis, USA) reported that *Myxococcus xanthus* has a genome of 9.5 Mb, making it the largest bacterial genome so far sequenced, and containing an estimated 7,500 open reading frames (ORFs).

By contrast, the genome of the obligate aphid endosymbiont *Buchnera aphidicola*, discussed by Siv Andersson (Uppsala University, Sweden), comprises only 583 ORFs, of which 579 are related to *Escherichia coli* genes. Thus, *B. aphidicola* probably evolved from a genetically well-endowed, free-living ancestor and, while occupying its niche within the cytoplasm of a eukaryotic cell, drastically contracted its genome size. Insight into how this occurred comes from

comparison of the sequences of two closely related *Buchnera* strains that co-evolved with their corresponding aphid hosts over the past 50 million years. On average, one gene has been eliminated each 5-10 million years. As the size of a genome is the sum of the rates at which genes are lost and gained, the on-going loss of *Buchnera* DNA was not compensated for by the acquisition of horizontally transferred DNA or by the duplication of indigenous genes. *Buchnera* probably illustrates a late stage in a multistep evolutionary process from free-living microbe, through the intermediate stages of facultative and then obligate intracellular parasitism, to its current status as an endosymbiont.

Some (but not all) microbiologists consider the cytoplasm to be an extreme environment and thus inhabited by extremophiles. None would doubt, however, that the capacity of, or indeed the requirement for, *Halobacterium* sp. NRC-1 to grow in 3-5 M NaCl would so qualify. The completion of its sequence, reported by Shiladitya DasSarma (University of Massachusetts, Amherst, USA), shows that *Halobacterium* contains 91 copies of insertion sequences and that its 2,682 ORFs are distributed between one large chromosome and two related minichromosomes. Transposition of its many IS repetitive elements would provide a mechanism for genetic plasticity through the fusion of smaller replicons. In turn, this may have led to the acquisition of essential genes by megaplasmids and hence to the possibility of competition between plasmid and chromosome. A second undisputed adaptation to an extreme environment is the capacity of hyperthermophilic archaea to grow at high temperature. One such organism, *Pyrococcus furiosus*, grows optimally at 100°C and is resistant to ionizing radiation owing, in part, to its capacity to reassemble chromosome fragments. Jocelyn DiRuggiero (University of Maryland Biotechnology Institute, Baltimore, USA) discussed the initiation of microarray hybridization experiments to identify genes that are differentially expressed during exposure of this organism to γ irradiation and to sub-lethal temperatures.

The oceans encompass surprisingly diverse environments, where different depths within the water column provide distinctive habitats within which adaptive evolution occurs. One such depth-dependent variable is light. Edward DeLong and colleagues (Monterey Bay Aquarium Research Institute, California, USA) used large-insert bacterial artificial chromosome (BAC) libraries to characterize uncultured marine bacterioplankton. Among the genes recovered was the coding sequence for a bacterial rhodopsin. Although this gene was found in the genome of an uncultivated γ proteobacterium, it is most similar in amino-acid sequence to archaeal rhodopsins, indicating its possible acquisition by horizontal gene transfer. Expression of this gene in *E. coli* and functional analysis of the corresponding protein showed a λ_{\max} corresponding to the wavelength of light available in the upper water column. By contrast, deeper-water rhodopsins were found to be blue-shifted, indicating that spectral tuning had occurred during the evolution of this protein family within bacteria living at different depths.

Evolutionary divergence by closely related strains living at different ocean depths is nowhere better demonstrated than in *Prochlorococcus*, a marine cyanobacterium whose photosynthetic activity contributes significantly to primary production in the oceans. Gabrielle Roca (Massachusetts Institute of Technology, Cambridge, USA) contrasted the sequence of two highly related strains of *Prochlorococcus* that exhibit strikingly different optimal light levels that adapt one for growth at the ocean surface and the other for growth in a deep water environment. Despite their closely related 16S rRNA sequences, the two genomes are substantially different with respect to size, GC content and codon usage. The strain living in the nitrogen-poor waters of the ocean surface lacks genes for nitrite and nitrate reductases, whereas the strain living in nitrogen-rich, deeper water possesses genes predicted to confer both activities. Thus, the sequencing of closely related strains living in different environments led to the identification of two genomic 'eco-types' that reflect both the constraints and advantages of their respective niches. How this diversity arose will be a matter of much future study, but surely one such mechanism is lateral gene transfer between natural populations of marine bacteria, as discussed by John Paul (University of South Florida, St Petersburg, USA).

Three fascinating genome sequence works-in-progress were presented. Steven Slater (Cereon Genomics LLC, Cambridge, USA) discussed *Agrobacterium tumefaciens*, a plant pathogen and the etiological agent of crown gall disease; Martin Odom (DuPont, Wilmington, USA) discussed the status and initial findings of the *Methylobacterium* project, a highly specialized aerobic eubacterium that can use only C-1 compounds (methanol/methane) for carbon and energy; and Malcolm Gardner (The Institute for Genome Research (TIGR), Rockville, USA) discussed *Theileria parva*, a tick-borne intracellular protozoan parasite of domesticated ruminants that induces malignant transformation of lymphocytes leading to fatal lymphosarcomas in afflicted animals.

The use of genome sequences to study microbial pathogenesis and epidemiology or to develop new vaccines and diagnostic assays was the focus of five talks. I presented my laboratory's microarray expression-profiling study of *Mycobacterium tuberculosis* growing within the macrophage phagosome and its transcriptional response to the nutrients and reactive oxygen and nitrogen intermediates that characterize this niche. Timothy Palzkill (Baylor College of Medicine, Houston, USA) described a phage-display method to express protein products corresponding to each of the 1,030 ORFs of the *Treponema pallidum* genome. This is a particularly worthwhile goal because this bacterium, the agent of syphilis, cannot be propagated *in vitro* and there is a compelling need for improved serological tests. The emerging drug resistance of *Staphylococcus aureus* is a major threat to all hospitalized patients and thus it was encouraging to learn that Uwe von Ahsen and colleagues (Intercell GmbH, Vienna, Austria) had used patients' sera and *in vivo* and *in vitro* methods to display all the proteins encoded by the staphylococcal genome, to identify genes coding for antigenic proteins expressed during infection and accessible to the immune system.

The movement of a genetically diverse infectious agent through a population can be tracked with the highest resolution by the combined use of conventional and molecular epidemiology. For the latter, Andreas Duesterhoeft (Qiagen GmbH, Düsseldorf, Germany) described a method for genotyping microorganisms by detecting single-nucleotide polymorphisms, and its use to study the epidemiology of tuberculosis. Genetic diversity can also occur during a chronic infection and can lead to mutants that are particularly well adapted to the host; these may be more virulent or more able to persist. Evgeni Sokurenko (University of Washington, Seattle, USA) addressed this important phenomenon by describing a method that can identify point mutations in total genomic DNA and is suitable for the analysis of genetic variability in bacteria isolated from infected tissues.

Genomics now and in the foreseeable future will both require and spawn novel analytical, data-mining and sequencing tools. Sung-Hou Kim (University of California, Berkeley, USA) forcefully admonished attendees that a crucial distinction exists between protein functions that are predicted from homology searches of databases, the proven role of a protein within a cellular pathway, and its molecular function as discerned by structure-function analysis. This is not merely the gap between genomics and proteomics - and the need to bridge it - as the latter term generally refers only to the 'proteome', that is, the complete protein expression capacity of the genome. Rather, as discussed by Kim, it will probably require innovative use of the rapidly expanding crystal-structure database to compile an encyclopedia of all protein folds and the development of tools that identify the set of folds that comprise the inferred protein products of newly sequenced genomes.