

Meeting report
Sequence-based genomics
Andrew JG Simpson

Address: Ludwig Institute for Cancer Research, Rua Prof. Antonio Prudente, 01509-010 São Paulo, Brazil. E-mail: asimpson@ludwig.org.br

Published: 27 August 2002

Genome Biology 2002, **3(9)**:reports4029.1–4029.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/9/reports/4029>

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

A report on the Genome-Based Pathogen Biology meeting, Hinxton, UK, 7-10 July 2002.

It is extraordinary the extent to which genomics serves as a unifying force to aid our understanding of biology. This is most clearly evident at those congresses and symposia where the genome sequencing and analysis of a wide range of organisms is presented. It would have been hard to imagine, even a few years ago, that a single, small, focused meeting would even contemplate the simultaneous discussions of viral, bacterial and parasitological pathogenicity, but that is exactly what was covered at the recent meeting convened by the Wellcome Trust at the Sanger Institute entitled "Genome-Based Pathogen Biology: The First 25 Years And Beyond". Somehow, whatever the organism and its life style, using the structure, variation and immediate function of its genome as the basis for describing its biology and observed phenotype grabs the attention of everyone who is also a genome-based biologist. It is probably because we are naturally most interested in our immediate area of research and we assimilate best new knowledge that bears directly upon it. Through genomics, everything bears directly on everything else and intellectual connections of relevance are possible even across wide spaces of the phylogenetic spectrum. We can continually see similarities and convergences between even quite distant organisms and the organism with which we are currently involved, and hence we are continually kept on our toes.

In addition to a fascinating retrospective on the development and evolution of sequencing techniques, presented by Bart Barrell (Sanger Institute, Hinxton, UK), I was struck by the way in which genome sequence information itself (along with the functional genomics tools of microarrays, RNA interference, gene transfection and so on) is continuing to serve as a front-line research tool for studying all the pathogens that were discussed. The discussion of the use of sequence data in the study of organisms with genomes of

such different sizes brought home to me just how powerful sequencing, re-sequencing and subsequent comparative genomics continue to be in unveiling biology.

Viruses have the smallest genomes, and a representative complete genome sequence for many of these organisms has been with us for many years. And yet sequencing continues unabated. For example, the current sequencing of herpesvirus genomes, as presented by Duncan McGeogh (MRC Virology Unit, Glasgow, UK), is not only providing the basis of a more detailed understanding of the phylogenetic relationships of herpesviruses and their evolution but is also providing novel insights into the function of individual genes. Some genes are remarkably stable in terms of inter-species variability, whereas others are highly variable, indicating increased selective pressure. These include genes encoding membrane glycoproteins, with obvious implications for their role as targets of immune responses. In this area, it was pointed out that DNA sequencing had been responsible for the very discovery of some novel viruses. A massive comparison of gene sequences is proving instrumental in understanding basic aspects of the life cycle of HIV (Simon Wain-Hobson, Institut Pasteur, Paris, France). Knowledge of the replication error rate and the rate of change of the genome sequence within infected individuals has made it possible to deduce that there is a huge loss of virus through immune attrition and that very few infected cells actually produce progeny. The depth to which the host-virus relationship can thus be probed, simply by DNA sequencing, is quite remarkable and reveals the technology as being truly a tool for experimental biologists.

Insights into biology and evolutionary origins were not limited to viral sequencing projects. Mark Achtman (Max Planck Institut für Infektionsbiologie, Berlin, Germany) provided elegant examples of such studies (performed with Daniel Falush and Sebastian Suerbaum) in the context of *Helicobacter pylori*. The sequencing of selected gene fragments from paired bacterial isolates recovered from the same patient approximately two years apart revealed a range

of alterations. Although in most cases the sequences were found to have remained unchanged, others showed a variety of alterations resulting from point mutations and recombination events. From the characteristics of these events it has been possible to deduce that the last common ancestor of *H. pylori* existed at least 2,500 years ago, and possibly as long ago as 11,000 years. Equally fascinating was the use of *H. pylori* sequences for tracing human migration. This is possible because the bacterium is transmitted vertically in families and because different bacterial populations can be identified corresponding to human population groupings. Indeed, it was proposed that, because there is very slow exchange of bacteria between population groups, such ethnic grouping is actually superior to that exhibited by human mitochondrial DNA!

Gordon Dougan (Imperial College, London, UK), discussing enteric bacteria, highlighted the immense power of interspecies genome-sequence comparisons. Comparison of the genome of *Salmonella enterica* serovar Typhi with that of *Escherichia coli* has revealed hundreds of insertions and deletions, within a highly conserved enteric backbone. The insertions/deletions range in size from single genes to large islands, consistent with a strong influence of horizontal gene transfer on bacterial evolution. Particularly interesting was the observation of many pseudogenes of apparently recent origin where the functional gene is associated with pathogenicity. These are clear examples of change of molecular capacity identified purely using DNA sequence information. The loss of enzymatic function underlies both Typhi's relatively restricted host specificity and its dispensation with the need to survive within the intestine, because it can invade other tissues and thus ensure its transmission.

Other powerful insights into bacterial phenotypes derived from whole-genome sequencing and comparative genomics were reported by Stewart Cole (Institut Pasteur, Paris, France) on the basis of a comparison of *Mycobacterium tuberculosis*, *M. leprae* and *M. ulcerans*. Again, one of the most striking features to emerge from this comparison is the massive reductive evolution that has occurred in *M. leprae* when compared with *M. tuberculosis*, as a result of the generation of pseudogenes and gene loss. In addition, genome sequence analysis is indicating possible new drugs that could be used to treat tuberculosis. For example, it might be possible to utilize existing anti-fungal reagents to target the many cytochrome-P450-requiring enzymes that rely on substrates not utilized by man.

Fungal pathogens, protozoal parasites and helminthic parasites were also discussed at the meeting. No complete genome sequences are yet available for helminthic parasites, but genome sequencing of the protozoans *Plasmodium falciparum* and *Leishmania major* is well advanced and for many others extensive cDNA sequencing has been undertaken. The sequence-derived data from *P. falciparum*,

discussed by Dan Carucci (US Naval Medical Research Center, Bethesda, USA) are a source of information concerning virulence, drug resistance, host specificity, parasite evolution, vaccine development and novel therapeutic targets. In addition, these data also provide the basic information required to utilize the functional genomics arsenal of microarrays, serial analysis of gene expression (SAGE) and proteomics. The situation with *Leishmania*, outlined by Stephen Beverley (Washington University School of Medicine, St Louis, USA) is entirely analogous and the promise of the functional genomics approaches that are being adapted to work with protozoal and indeed helminthic parasites is impressive.

Thus, whatever the organism, whatever its genome size and without regard to whether or not a representative genome sequence has already been generated, the importance of DNA sequencing remains high. Whether the current focus is the initial sequence of a genome, or the sequencing of variants and closely related species, in order to permit the power of comparative genomics to come into play the generation of novel DNA sequences is a front-line research tool. While the tremendous power of functional genomics approaches is indisputable, and the importance of their integration with highly focused hypothesis-based research is undoubted, I would venture that sequence-based research will continue unabated for the foreseeable future, and that we are still very much at the beginning of the sequencing era despite the enormous progress that has already been made. Any discussion of the lessening of importance of dedicated sequencing centers and consortia and movement away from sequencing to other experimental approaches is thus, to my mind, probably highly premature.