

Exploiting genomes: bases to megabases in 50 years

Petra Oyston & Dave Kelly

The Symposium *Exploiting genomes: bases to megabases in 50 years* is being organized by H. Jenkinson, P. Oyston, I. Sutcliffe, J. Parkhill and D. Kelly and will take place on Monday 8 and Tuesday 9 September 2003 at the 153rd Meeting of the SGM, at UMIST, Manchester. See www.sgm.ac.uk/meetings for the full programme, and to book online.

The whole of this meeting is a major SGM contribution to the DNA50 programme.

It is 50 years since Crick and Watson published in *Nature* the structure of DNA, and the SGM has planned a symposium to be held at the September 2003 meeting at UMIST, Manchester, to celebrate the impact that this discovery has made on microbiology. What little progress could have been made in microbiology without the ability to examine and manipulate microbial DNA, skills we now take for granted, seems obvious to us now in 2003. Using the tools of molecular biology, we can clone and mutate genes, express proteins to high yield in large-scale fermenters, determine the evolutionary relatedness of any given strain, and even identify microbes that are impossible to grow in the laboratory. Indeed, it is obvious that microbiology has been transformed by molecular biology in ways that must have been unimaginable 50 years ago.

Once the structure of DNA was known, it was only a relatively short time before the organization of genes and (in bacteria) operons was understood. However, due to the early cumbersome chemical sequencing techniques it was some time before a complete genome was sequenced. The first DNA genome to be sequenced was that of bacterial virus ϕ X174 by Fred Sanger's group at Cambridge in 1977, and for a long time sequencing of only the small genomes of viruses could even be contemplated. The advent of high throughput sequencing technology has since revolutionized microbial genomics. Genome sequence data have allowed the entire genetic repertoire of an organism to be examined, rather than just individual genes in isolation. An introductory perspective on 'Genomes and beyond' will be provided in the opening lecture of the symposium, to be given by G. Weinstock. The world was rather taken by surprise when *Haemophilus influenzae* was the first bacterium to be completely sequenced by Fleischman *et al.* in 1995, as most people had expected *Escherichia coli*, the laboratory workhorse, to claim that honour. But although the sequences of several small-genome, specialized pathogens were published in rapid succession after 1995, larger microbial genomes soon followed. There are currently 41 completed bacterial genome sequences (although this number will be out of date by the time this article is published!), and over 150 sequencing projects at different stages of completion. At the Symposium, some of the applications of this mass of genome data will be reviewed by Julian Parkhill (Sanger Institute). Siv Andersson (Uppsala University) and Emilio Garcia (Lawrence Livermore National Laboratory) will talk in more detail about the recent sequencing of *Bartonella* and *Yersinia* (two interesting and important pathogenic bacteria), respectively. Nowadays, the ability to sequence whole genomes is almost routine, so that now there are examples of several strains of a species being sequenced for comparison. However, the publication of a genome

sequence is still a matter of particular excitement, not only for researchers working on that organism, but often for a wider audience of microbiologists who can see relationships with their own work. Rather than the genome sequence being an end in itself, it is the start of a huge amount of work to put the information into context with regard to the biological significance of the organism. For example, sequence data is being exploited to reveal how species have evolved, the basis of strain diversity and how pathogens have both acquired virulence genes and lost non-essential genes in their adaptation to their hosts.

A key discipline which has been essential to the success of genomics is that of bioinformatics, by which the likely identities of genes and structures of gene products can be inferred and their evolution traced (T. Gaasterland, Rockefeller University). Predictions can also be made about the level of expression of genes from sequence information (D. Ussery, Denmark). An interesting project called the Minimal Microbial Genome which will be outlined by C. Hutchison (TIGR, USA) is taking a novel approach to analysing the significance of each gene in an organism. They are asking the question, 'How few genes does a bacterium actually require for viability?' To answer this, they intend to strip down an existing genome to a minimal set of essential genes, creating, in effect, a new organism.

In addition, genomics has allowed the rapid development of other areas, such as transcriptomics and proteomics, whereby the expression of all of the genes and proteins, respectively, in an organism can be monitored in response to varying stimuli. Infection results in massive changes in protein expression in both the micro-organism and the host. Viruses, by their very nature, hijack the machinery of the cell for their replication, so it is unsurprising that viral infection results in dramatic changes in cellular gene expression. This topic will be covered in two talks (P. Ghazel, Edinburgh University and P. Kellam, UCL), and

ATGGGAAGGA



herpesvirus genomics will be reviewed by A. Davison (University of Glasgow). However, until recently the effect of bacterial infection on cellular gene expression has been less well understood. The availability of fully annotated bacterial genomes represents a powerful resource to help understand the complex biology of host–pathogen interactions for a range of medically important bacterial pathogens. The use of microarrays to monitor whole-genome expression in bacteria (for example, talks by Phil Butcher, St George’s Medical School and Brendan Wren, London School of Hygiene and Tropical Medicine) is a powerful approach, and is resulting in identification of differentially expressed genes important in pathogenesis, as well as providing useful targets for the rational design of new drugs and vaccine candidates for important bacterial pathogens. The study of transcripts by array technology is complemented by proteomics. David O’Connor (Southampton University) will describe the most recent developments in the field of protein identification and quantitation, using *Salmonella enterica* var. Typhimurium as his model. The technology of proteomics is moving ahead rapidly, with new software and automated systems of protease digestion and mass spectrometry for protein identification. Linking proteome maps to genomic databases will create an invaluable resource for future studies of many microbial species.

Whole-genome screening techniques, such as identifying promoters necessary for *in vivo* gene expression

(Paul Rainey, Oxford) or following the fate of individually tagged mutants in mixed pools during infection are proving a powerful way of coupling genetic techniques to genomics. Approaches like these will allow us to apply some hypotheses to the vast amount of data that has been generated in recent years. In fact, if there is one criticism of the ‘genomic era’ it is that data generation may have over-shadowed efforts to understand the biological significance of the information. Even storing data, such as proteome maps and microarray files, in an accessible yet controlled way is a challenge that is just starting to be addressed.

Many pharmaceutical companies have invested heavily in genome sequencing and this reflects the contribution that the data will have in the search for novel antimicrobial targets and vaccine antigens. Historically, identification of antigens suitable for sub-unit vaccines or finding targets which when inactivated would result in attenuation, were reliant on luck as much as anything. With the availability of genome and proteome data, a more logical and informed approach can be followed. *In silico* antigen prediction allows the best candidates from the whole protein complement of the pathogen to be identified (R. Rappuoli, IRIS Research Centre). Likewise, targets which would be suitable for inhibiting with a new generation of antimicrobials can be identified from genome data (P. Rathod, University of Washington).

The organizers hope that this symposium will not only be a celebration of the remarkable progress that has been made since the discovery of the structure of DNA, but will give an overview of cutting-edge research in this area, as well as providing a useful perspective with which to view the current and future development of microbial genomics. We therefore strongly encourage SGM members to participate in this historic symposium at UMIST.

●*Dr Petra Oyston is in the Department of Microbiology, Dstl Biomedical Sciences, Porton Down, Salisbury SP4 0JQ, UK.*

●*Professor Dave Kelly is Chair of Microbial Physiology in the Department of Molecular Biology and Biotechnology at the University of Sheffield, Western Bank, Sheffield S10 2TN, UK.*

LEFT: James Watson (b. 1928) on the left and Francis Crick (b. 1916) on the right with their model of part of a DNA molecule in 1953.

PHOTO A. BARRINGTON BROWN / SCIENCE PHOTO LIBRARY

BELOW: The scrolling display at the Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. COURTESY RICHARD SUMMERS, WELLCOME TRUST SANGER INSTITUTE

Representations of the double helix

A collection of structural models of DNA is on display at the Whipple Museum of the History of Science, Cambridge, throughout 2003. It is part of an exhibition to celebrate the scientific and cultural impact of the DNA double helix. A star attraction is the full-size reconstruction of the first Watson and Crick model. See www.hps.cam.ac.uk/whipple

