

Meeting report

## The growing utility of microbial genome sequences

Nigel J Saunders\* and Alex C Jeffries

Address: Molecular Infectious Diseases Group, Institute of Molecular Medicine, University of Oxford, Headington, Oxford, OX3 9DS, UK.

\*E-mail: saunders@molbiol.ox.ac.uk

Published: 9 June 2000

Genome **Biology** 2000, 1(1):reports410.1-410.3

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2000/1/1/1/reports/410>

© Genome**Biology**.com (Print ISSN 1465-6906; Online ISSN 1465-6914)

---

A report from the Genome Biology session of the 4<sup>th</sup> annual conference on microbial genomes, Virginia, February 12-15, 2000.

---

Of the 21 bacterial genomes that have so far been sequenced in their entirety, 16 are from pathogenic species. One of the main attractions to the complete sequencing of microbial genomes has been the potential to understand more readily the molecular biology of these important pathogens, and thereby obtain crucial insights that would lead to the development of vaccines and antimicrobials. The move from the relatively simple annotation of genes in a genome to insights into their function, however, is not a trivial task. Diverse approaches to exploit the utility and influence of microbial genome sequences on whole organism biology were highlighted in this session; each speaker tackled the issue of how best to take advantage of the available genome sequence information in a different way.

The *Neisseria meningitidis* serogroup B strain MC58 sequence has been used to identify novel vaccine candidates that are now being actively pursued (Rino Rappouli, Chiron SpA). Using the recently completed sequence, genes that encode potentially surface-exposed proteins that had not previously been pursued as vaccine candidates were identified. These were then cloned, expressed, purified and tested for immunogenicity; the properties of the antibodies raised were then determined. Screening of 350 proteins, whose genes were expressed in *Escherichia coli* led to the discovery of 85 new surface-exposed proteins, of which 25 induced bactericidal antibodies. A recurring theme of the meeting in general was the way in which information derived from a single strain could be used to study the wider population: as reflected here by sequencing the strongest vaccine candidates and comparing the amino acid sequence across a set of strains selected to represent the sequence diversity of the

group B meningococcus population. This revealed that some of the surface proteins displayed substantial inter-strain variability, which is a common feature of surface exposed proteins that are under immunological selection, and is frequently the reason why vaccines that are raised using a single strain are less effective in protecting against unrelated strains. Some of the candidates were highly conserved across the population and these are being pursued further. The work described presented a problem-oriented strategy using a complete genome sequence to identify and test novel surface proteins and has identified vaccine candidates, unlikely to have been identified in the absence of the genome sequence information.

Frank Rosenzweig (University of Florida) presented work in yeast that used a microarray developed from the complete genome to study *in vitro* evolution as driven by nutritional stresses in chemostat cultures. Evolution in this model system starts to occur after 25 to 50 generations, with the rate of change being strain-dependent. The clear message from this study is, that although changes in gene function due to mutation may be slow, the potential to adapt to nutritional stress (in this case glucose limitation) is great, leading to a rapid and consistent change in the organization of metabolism. Alteration (indicated by two-fold increases or decreases in mRNA concentration) in the expression of approximately 10% (608) of the genes could be detected, of which 209 had unknown functions; 534 of these changes were consistent between different strains and experiments.

Continued developments in the use of molecular beacons in the typing and antimicrobial susceptibility testing of *Mycobacterium tuberculosis* were reported by David Alland (Montefiore Medical Center, New York). Molecular beacons are oligonucleotide probes that form stable stem loop structures, have sequence complementary to the sequence of interest in the loop, and have a fluorophore and a quencher at each end. When the probe is not bound to the target

sequence, the fluorescence from the fluorophore is quenched by the quencher, but when bound to the target sequence, the fluorophore and quencher are separated and a signal can be detected. The major strengths of this approach are that it can detect single base-pair polymorphisms and can be performed rapidly on PCR products. In contrast to the approaches looking at the genome as a whole, this research focuses upon the genes that are known to confer antibiotic resistance in *M. tuberculosis* and upon the 16S ribosomal sequences. This group's approach requires that they look at a large number of organisms for small differences in the genes of interest.

An important finding in this ongoing project is that the pattern of point mutations present in strains monoresistant to the antibiotic isoniazid differs significantly from that seen in multidrug resistant strains. Of the monoresistant strains, 44% did not have the mutations typical of multidrug resistant strains. They interpret this observation to suggest that it challenges the current model in which multidrug resistance is thought to occur through the accumulation of mutations during prolonged exposure to antimicrobials.

New developments using this approach, which may be widely applicable, are the use of multiple beacons in a single assay, and the detection of low numbers of mismatches between the target sequence and the probe loop. For example, resistance to rifampicin is usually due to one of several alternative point mutations in the *rpoB* gene. By using a combination of molecular beacons, each with different fluorophores and specific for different mutations, Alland and colleagues have been able to detect changes in the five most frequent locations in a single reaction. The development of probes that can bind to regions including mismatches is particularly ingenious. These 'sloppy' molecular beacons have a longer probe region in the loop (40 to 45 nucleotides rather than 15 to 21), and similar complementary arm lengths forming the stem. The fluorescence generated in an assay depends upon the concentration of the target, and upon the number of differences between the probe and the target sequence. The relative fluorescence of multiple probes in a single assay is not, however, dependent upon the concentration of the target. Using a series of probes directed against a variable region of the sequence encoding the 16S RNA they have successfully used the molecular beacons to distinguish different species of mycobacteria. The development of these methods is of practical importance in investigating the slow-growing mycobacteria in a clinical situation, and also demonstrates that this approach may have more general application in the investigation of sequence polymorphisms in other contexts.

Michael Laub (Stanford University) presented work that demonstrates the utility of genome sequences, even prior to their completion, to investigate cell-cycle regulation using *Caulobacter crescentus* and a microarray based approach.

*C. crescentus* has a life cycle that has clearly definable stages and these can be experimentally synchronized. This group has demonstrated temporal co-ordination of many genes over the life cycle of the cell. Although some of their observations are not entirely novel, such as the expression of the genes encoding flagellar components in the same order as they are assembled, the detail in which these processes can be observed in the cell as a whole is remarkable. The group went on to search the genome for consensus binding sequences for the gene regulator CtrA and identified 29 sites within the genome. Expression profiling, using a temperature sensitive *ctrA* mutant, showed a large number of transcripts that were affected by CtrA. Analysis of these transcripts showed that genes in 13 operons had a putative CtrA site upstream, had increased mRNA expression in response to CtrA, and were transcripts that were normally cell-cycle regulated. One pair of divergent operons included a response regulator in one direction and a sigma factor and a histidine kinase in the other, highlighting the integration of the expression control mechanisms. So far from this study, 402 new cell-cycle regulated genes have been identified and understanding of the specific role of CtrA has been advanced substantially.

Pierre Legrain (Hybrigenics, Paris) described an extremely powerful analysis of protein-protein interactions using the yeast two-hybrid assay to look at bacterial systems, using *Helicobacter pylori* as an example. A complex library was prepared using random genomic DNA or cDNA fragments (4 million fragments for *H. pylori*) and screening was performed using tens of protein 'baits' in parallel. The use of many overlapping fragments, rather than the full length protein is thought to reduce the number of false positives and negatives that occur due to mis-folding of the proteins under the conditions in which they are expressed, and is used to identify the 'selected interaction domains' (SIDs). Some 300 proteins were analyzed as baits and 1,200 selected interactions were used to infer connections between 800 different proteins; these were then used to assign functions to many genes. This approach is being used to identify potential new antimicrobial targets by identifying the genes that are essential and those that interact with them; this latter category of target is not readily identified using usual methods. A huge amount of information is generated from these experiments and the nature and specificity of the interactions that are identified will need to be confirmed. Nonetheless, this approach has the potential to make a substantial contribution to defining the metabolic pathways and networks that exist within the cell, and to identifying the roles of the substantial proportion of proteins for which function is currently unknown.

It is a feature of the maturation of the field that genome sequences are being used in so many contexts and in combination with so many different experimental approaches. Global screening, expression-based studies, comparison

with other strains to identify vaccine candidates, studies based on chemostat and co-ordinated cultures investigating changes within a single strain over single or multiple generations using microarrays, the development of new methods such as molecular beacons that facilitate comparisons between strains, and finally the use of a global protein-interaction study, demonstrate the range and diversity of uses for genome sequences. The results of these genome-based research projects are making dramatic contributions to their fields - all this only five years since the first genome sequence was completed. The studies presented addressed basic biology as well as applied use of the sequences to analyze the broader population, and the identification of both vaccine and drug targets. Whilst some of the uses of genome sequences necessarily involve the use of substantial resources, it is clear that genome sequences will contribute more and more to an increasing proportion of the research community in very diverse ways.