

Meeting report

Small genomes and big science

Jeremy Edwards

Address: Molecular Genetics and Microbiology, Cancer Research and Treatment Center, University of New Mexico Health Sciences Center, and Chemical and Nuclear Engineering, University of New Mexico, Albuquerque, NM 87131, USA. Email: jsedwards@salud.unm.edu

Published: 13 March 2006

Genome Biology 2006, **7**:308 (doi:10.1186/gb-2006-7-3-308)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2006/7/3/308>

© 2006 BioMed Central Ltd

A report of the 13th Annual International Conference on Microbial Genomes, Madison, USA, 11-15 September 2005.

The presentations from the 2005 Annual Conference on Microbial Genomes focused on diverse areas of microbial genomics - from the evolution of enterobacteria to structural genomics and systems biology. An overriding theme of the meeting was the importance of new technologies and tools for functional genomics and how they are being used to understand microbial physiology. This meeting took a big step forward in showing how to take advantage of the increasing availability of microbial genomes to fill in the gap between functional genomics and physiology. This report discusses a few of the many highlights of the meeting in the fields of metagenomics, structural genomics, new genomics technologies and systems biology.

Metagenomics and community biology

A number of presentations focused on the metagenomics of species groups and the analysis of microbial communities, rather than on individual species or strains. Jeremy Glasner (University of Wisconsin, Madison, USA) discussed the parallel evolution of pathogenicity in enterobacteria. From his results, a new view of genome evolution in the enterobacteria emerges - one in which the genomes of species are incredibly dynamic and genes are exchanged between strains and species. Glasner's work analyzes sequences that have been completed in more traditional genome-sequencing projects, where individual strains are analyzed in isolation. A complementary approach is that of Jizhong Zhou (Oak Ridge National Laboratory, Oak Ridge, USA), who described his work on the metagenomic analysis of microbial communities in uranium-contaminated groundwaters. Zhou shotgun-sequenced the DNA isolated from a mixed community of microbes and analyzed the sequence in an attempt to understand this complex community. He and his colleagues

sequenced 60 Mb from the uranium-contaminated soil samples and identified a composite sequence of approximately 6 Mb in 879 contigs. They were unable to determine exactly how many species made up the community, but they did identify *Azoarcus* species (at least four) as well as other bacterial species via analysis of 16S rDNA. This type of metagenomic analysis promises to provide a critical insight into the biology of microbes in their natural environment, but the amount of the sequences that need to be analyzed begs for new technology, some of which is described later.

Staying with the metagenomics theme, Garth Ehrlich (Allegheny General Hospital, Pittsburgh, USA) described the 'distributed genome' hypothesis. His group sequenced ten *Haemophilus influenzae* strains and identified a 'supergenome' of approximately 3,300 genes, which is about twice the gene complement of any single strain. They hypothesized that there are contingency genes spread across the population that provide improved population survival. This hypothesis could have a tremendous impact on how we should study microbes and could reshape our understanding of a species and how we define it.

Structural genomics and new genomics technologies

No meeting on genomics would be complete without a discussion of structural genomics. The keynote address from Sung Hou Kim (University of California, Berkeley, USA) provided a global view of the protein universe and the evolution of protein fold classes. Kim and colleagues have analyzed all the protein structural motifs that have been experimentally determined out of the potential 10^{12} proteins and approximately 10,000 structural motifs on Earth. The analysis revealed a protein-structure universe map which was clearly defined by the four major fold classes (alpha, beta, alpha + beta, alpha/beta), and the map could be interpreted with respect to the molecular functions of a protein, such as metal binding. The map provides a simplifying organization that

can be applied to analyzing and understanding structural data, and may provide methods for linking structure to function.

A number of other talks highlighted the importance of structural genomics in understanding microbial physiology. George Phillips (University of Wisconsin, Madison, USA) described the use of structural genomics to understand the function of unknown proteins, an approach that has been termed 'reverse structural biology'. Reverse structural genomics is likely to be another important tool in the daunting task of elucidating the function of all the proteins encoded by a genome. Scott Lesley (Scripps Research Institute, La Jolla, USA) described the progress of *Thermotoga maritima* structural genomics. He and colleagues have cloned the entire *T. maritima* proteome and have studied this clone set for optimal protein expression systems and crystallization conditions with a view to X-ray crystallography. This project could serve as a model for future 'crystal-lome' studies and could provide a key insight into microbial physiology because of the thermostability of the *T. maritima* proteins.

The genomics technologies that have already been developed have allowed the microbiologist to "think outside the box" and collect data that were unimaginable just a decade ago. But despite the technological progress with tools such as DNA and protein microarrays and high-throughput sequencing, new genomics tools are needed to facilitate additional types of studies in microbial genomics. Maithreyan Srinivasan (454 Life Sciences Corp., Branford, USA) described a sequencing-by-synthesis technology using picoliter-scale reactions. Recently reported in *Nature*, this is a critical tool for microbial genomics as it opens up the possibility of any lab being able to generate a genome sequence for their favorite organism at a very reasonable cost. Tom Albert (NimbleGen, Madison, USA) described a method of genome resequencing using dense arrays of oligonucleotides. Scott Jackson (Food and Drug Administration, Laurel, USA) described optical mapping of *Escherichia coli* O157; for this, whole-genome maps are constructed from genomic DNA molecules directly extracted from the bacteria by creating ordered restriction maps using individual DNA molecules mounted on surfaces. And I described the applications to microbial genomics of polony technology, a method for the parallel analysis of large numbers of individual DNA molecules in a high-throughput manner. Describing an application of these technologies, Bernhard Palsson (University of California, San Diego, USA) discussed the utilization of NimbleGen arrays and mass spectrometry for the resequencing of evolved *E. coli* strains. His group was able to identify mutations that provided a selective advantage and to interpret these results utilizing metabolic modeling. It is clear that new technologies are being developed that will continue to push the limits of microbiology.

Systems biology

The emergence of genomics has led to the emergence of systems biology, which encompasses research areas such as synthetic biology, metabolic engineering and computer modeling of biological processes. The ultimate goal of synthetic biology is to generate designer organisms. Synthetic biology is regarded as part of systems biology, as in order to design an organism one must have a detailed understanding of all the 'parts' and how these parts operate together in a complex system. One of the major obstacles to a completely artificial organism is the construction of the genome that will code for all the parts. Clyde Hutchinson (University of North Carolina, Chapel Hill, USA) described the attempt to eliminate this bottleneck by defining the 'minimal genome', which is the genome that is generated by removing all unnecessary genes until only those genes essential for supporting life remain. He described bioinformatics and transposon mutagenesis approaches to identifying the minimal genome as a prerequisite to constructing an artificial genome. Hutchinson and colleagues have identified between 310 and 388 essential genes for the minimal genome, depending on the method for calculating this number.

The Hutchinson method for constructing a minimal genome is providing tremendous insight into the function of microbes but, as pointed out by Drew Endy (Massachusetts Institute of Technology, Cambridge, USA), without constructing the genome *de novo* and understanding and controlling such features as gene orientation, one will not have a complete systems-level understanding of the organism. It is not technically possible yet for synthetic biology to work at the whole-genome scale for a free-living organism, but Jingdong Tian (Duke University, Durham, USA) presented a method that may allow the synthesis of megabases of DNA. He has developed a genome-synthesis method using DNA microchips, and he and his colleagues have used the method to synthesize a 14 kb 21-gene operon with an error rate of 1 in 1,400 bp. They suspect that they will be able to reduce this error rate to 1 in 30,000 bp in the short term and have an ultimate goal of an error rate less than 1 in 10^6 .

Metabolic engineering is another aspect of systems biology, and Costas Maranas (Pennsylvania State University, University Park, USA) has been developing optimization tools that can be used in the design of microbial metabolism. He and colleagues have developed computational methods that can be used to identify key points in metabolic pathways for genetic engineering. Many of the tools and techniques that Maranas has developed are incorporated into a commercial software package available from Genomatica (San Diego, USA). Christophe Schilling of Genomatica described these software tools and how they can be utilized to guide the engineering of metabolic pathways. He also described how they have been used to aid the annotation of microbial genomes such as that of *Geobacter*. Such tools will clearly accelerate the design and construction of bacterial strains for industrial

applications. Lisa Laffend (DuPont, Wilmington, USA) described work at DuPont on the construction of a strain of *E. coli* for the industrial production of 1,3-propanediol, a tremendous example of a successful industrial strain. The 1,3-propanediol project is an innovative bio-based method that uses corn, rather than petroleum-based processes, to make monomers for the production of clothing, carpets, automobile interiors, for example. Although this strain was constructed without the benefit of the tools developed by Maranas and Schilling, they would clearly have been very useful.

Also in the general area of metabolic engineering, Jay Keasling (University of California, Berkeley, USA) described the many steps in the development of a strain of *E. coli* for the production of isoprenoids for use as antibacterial, anti-fungal and anticancer drugs. They have made great progress in producing compounds of medical value that cannot be obtained by any other method.

Overall, the conference highlighted the main directions of microbial research in the post-genomic era. In order to move forward and make the maximum use of the available data, both traditional biologists and those focused on high-throughput approaches will need to interact and collaborate with engineers and computer scientists. Bringing together tools developed by a diverse group of researchers is likely to push the field ahead at an even greater pace.