Genetica **108:** vii–ix, 2000.

Guest editorial

Sequence to Function: The 7th conference on Small Genomes

Jizhong Zhou & Anthony V. Palumbo
Environmental Sciences Division, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831, USA

Key words: conference, genome, genomics, sequencing, microarray, bioinformatics, gene discovery

Background and history

Microbial genomic analysis is in a phase of 'exponential growth'. The gene sequences of many microorganisms have been determined and are available. Dozens of other microbial sequencing projects are in progress. However, determining the structure of entire genome sequences is only the first step in characterizing an organism. Elucidating the functions and interactions of the sequenced genes is a great challenge.

Knowledge of entire genetic sequences opens a whole new range of possibilities for research that is more efficient. Thus, many laboratories are addressing important questions in functional genomic research by integrating genetic, biochemical, and bioinformatic approaches. Consequently, areas in functional genomics and associated genomic technology are developing very rapidly. Annual conferences provide opportunities for investigators to exchange ideas, discuss their recent discoveries, and work toward solving common problems. Rapid exchange of knowledge and the establishment of critical collaborations are vital in remaining at the cutting edge of this field.

To meet the need for rapid exchange of information, several series of conferences have been organized including an annual conference on small genomes (primarily microorganisms). This series on small genomes originated in 1992 has proven to be an important forum for information exchange in microbial genomics. The meetings have met with great success and have engendered an integrated approach to the *Escherichia coli* genome, the chief focus of the first three meetings. The 4th meeting expanded the focus by including research on the sequencing of other small genomes. The 5th meeting celebrated the completion of the *E. coli* sequence and those of other small genomes, and it further extended the scope of functional genomics. The 6th meeting addressed genetics and physiological studies related to the *E. coli* sequence, as well as the sequencing of other microbial genomes. All of these meetings have attracted leading scientists and institutions involved in genome sequencing and microbial functional genomics.

The 7th conference on Small Genomes

The 7th Conference on Small Genomes was held on November 13–17, 1999, in Arlington, Virginia, U.S. This conference focused on:

(1) defining gene functions and regulatory networks using integrative multidisciplinary approaches and
(2) exploring genome sequence information to understand various biological processes.

Seven areas of microbial genomics were highlighted at this conference including:

(1) microbial diversity and microbial genome sequencing;
(2) bioinformatics;
(3) genomic technology;
(4) genetic and biochemical technologies for defining gene functions;
(5) cellular processes, regulatory networks, and evolution;
(6) systematic functional analysis of genome sequences; and
(7) applied functional genomics.

The following people served as members of the advisory committee for the 7th Conference: Dr. Jizhong Zhou (Chair) from Oak Ridge National Laboratory; Dr. Jeffrey H. Miller from the University of California, Los Angeles (Co-Chair, and the chairperson of the 4th, 6th, and 8th meetings at Lake Arrowhead in 1996, 1998, and 2000, respectively); Dr. George Weinstock

from the University of Texas at Houston; Dr. Monica Riley from the Marine Biological Laboratory; Dr. Elisabeth Raleigh from New England BioLab; Dr. Jennie Hunter-Cevera from Lawrence Berkeley National Laboratory; and Dr. Richard Mural and Dr. Anthony V. Palumbo (Co-Chair) from Oak Ridge National Laboratory.

The conference was designed by the organizers to be broadly representative of the microbial genomics research community. More than 100 professionals from various countries presented their research in oral presentations and posters. The audience consisted of more than 250 researchers, postdoctoral associates, students, and others. With the support from U.S. government agencies and private industry, this conference contributed to the education of the next generation of genomic experts. Fellowship awards were used to support the attendance of 43 graduate students, postdoctoral associates, and young faculty members at this conference.

Conference papers in this mini-special issue

The papers presented in this issue touch on several of the focus areas of the meeting. We hope that these papers convey some of the diversity of information and approaches being used in microbial genomics. The papers presented in this issue cover areas ranging from development of specific tools for sequencing, bioinformatics, and comparative genomics to applied functional genomics of pathogens.

The development of tools for genomics was a common theme of the meeting. Hoffman et al. (this issue) presented a transposon method for direct sequencing of microbial genomes using transpoase-transposon complexes. They point out that their method could be used in providing sequence data from bacteria for which whole genome sequences are not available. Their paper illustrates that despite the availability of whole genome sequences from shotgun cloning, other approaches can still be useful in providing specific data. Other papers presented at this meeting, but not included in this issue, looked at other specific techniques for sequencing closely related microorganisms, such as subtractive hybridization.

The utilization of sequence data in gene expression studies has been greatly facilitated by high throughput techniques, such as microarrays. Kuklin et al. (this issue) presented tools for more efficiently gathering data generated in microarray studies. There are many issues that must be addressed when taking a microarray image and reducing it to quantitative data that can be analyzed and manipulated. In the early stages of microarray development, many of these issues had to be addressed by tedious human operations. The development of tools such as those presented by Kuklin et al., should facilitate more rapid analysis than was previously possible using less automated image analysis techniques.

Makarova et al. (this issue) presented an analysis of the *Deinococcus radiodurans* genome that focused on the identification of protein families. This paper is representative of the bioinformatics approach to extracting useful information from published microbial genomes. It also represents a significant research interest in *Deinococcus radiodurans* that was evident at the meeting and in the literature. For example, there were several other papers presented at the meeting that were related to *Deinococcus radiodurans*. This organism represents fertile ground for gene discovery because of its unique ability to withstand environmental stress and DNA damage. It may also have importance in bioremediation activities. The authors highlighted the discovery of protein families that are overrepresented in this organism, and some of these are clearly related to the unique abilities of the organism. Their study identified proteins that are potential targets for mutagenesis and gene expression analysis. This was one of several papers presented at the meeting that used bioinformatics and comparative approaches to examine genome sequences.

Although comparative approaches are the basis for many current gene identification systems, developments in gene recognition are still occurring (e.g., Natale et al., this issue). The authors designed a systematic approach using a database of ‘Clusters of Orthologous Groups of proteins or COGs’ to assist in gene recognition. These clusters are used to help identify species that are missing members of specific COGs, and these species become the target for additional analysis to identify the apparently novel gene that codes for the missing COG. Using this approach, they have identified new sequences that were unrecognized by other approaches. One reason for the previous lack of identification was the small size of some of these candidate proteins. This work points out the problem posed in trying to identify short gene products in DNA sequences.

Friis et al. (this issue) have combined comparative genomics approaches and bioinformatics to yield

new insights into the origin of pathogenicity. They focused on the visualization of pathogenicity regions or 'islands' as a tool in deciphering gene relationships. To assist in this analysis, they created 'Genome Atlases' for many sequenced organisms and have also made them available on the Internet. They illustrate this visualization approach using three published sequences with known virulence factors. When applied to two plasmids (one from a pathogenic *E. coli* strain and the other from *Yersinia pestis*) with virulence genes or genes for toxins, the approach illustrated the difference in A + T content as well as structural properties in the virulence genes.

Another example of a comparative genomics approach presented was a study of gene order in eubacteria and eukaryotic organelles by Nikolaickik and Donachie (this issue). The authors showed that the gene order present in specific gene clusters, such as ribosomal proteins, and in those proteins involved in cell division is highly conserved. As the authors pointed out, this is a ripe area for future research, because the reasons for this conservation are unknown.

A paper presented by Fernández et al. (this issue) focused on indicators for drinking water quality in Argentina. The study focused on *Aeromonas hydrophila*. The authors used both conventional microbiological techniques to isolate strains and molecular techniques to look at the diversity of genes and gene products. They demonstrated that despite chlorination this pathogen can be detected in municipal water systems.

Acknowledgements

We gratefully acknowledge the support of the U.S. Department of Energy, the U.S. National Science Foundation, the U.S. Office of Navy Research, Oak Ridge National Laboratory, Merck Research Laboratory, and GSI Lumonics. We would especially like to thank the DOE Office of Environmental and Biological Research for the seed money to initiate the planning of this conference. The organizers would also like to thank the editors for providing the opportunity for publishing this mini-special issue in *Genetica.* Oak Ridge National Laboratory is managed by University of Tennessee-Battelle LLC for the Department of Energy under contract DE-AC05-00OR22725.

References

1. Kuklin, A., S. Shams & S. Shah, 2000. High throughput screening of gene expression signatures. Genetica 108: 41–46.
2. Friis, C., L.J. Jensen & D.W. Ussery, 2000. Visualization of pathogenicity regions in bacteria. Genetica 108: 47–51.
3. Fernández, M.C., B.N. Giampaolo, S.B. Ibañez, M.V. Guagliardo, M.M. Esnaola, L. Conca, P. Valdivia, S.M. Stagnaro, C. Chiale & H. Frade, 2000. Aeromonas hydrophila and its relation with drinking water indicators of microbiological quality in Argentine. Genetica 108: 35–40.
4. Natale, D.A., M.Y. Galperin, R.L. Tatusov & E.V. Koonin, 2000. Using the COG database to improve gene recognition in complete genomes. Genetica 108: 9–17.
5. Hoffman, L.M., J.J. Jendrisak, R.J. Meis, I.Y. Goryshin & W.S. Reznikoff, 2000. Transposome insertional mutagenesis and direct sequencing of microbial genomes. Genetica 108: 19–24.
6. Makarova, K.S., L. Aravind, M.J. Daly & E.V. Koonin, 2000. Specific expansion of protein families in the radioresistant bacterium *Deinococcus radiodurans*. Genetica 108: 25–34.
7. Nikolaichik, Y.A. & W.D. Donachie, 2000. Conservation of gene order amongst cell wall and cell division genes in Eubacteria, and ribosomal genes in Eubacteria and Eukaryotic organelles. Genetica 108: 1–7.