MEETING REVIEW

Microbial Genomics—Challenges and Opportunities: The 9th International Conference on Microbial Genomes

Jizhong Zhou[1]* and Jeffrey H. Miller[2]

Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831,[1] and Department of Microbiology & Molecular Genetics, University of California at Los Angeles, Los Angeles, California 90095[2]

BACKGROUND AND HISTORY

Genomic sequencing and analysis are in a period of "exponential growth." More than 60 eukaryotic and prokaryotic genomes have been completely sequenced, with analysis of more than 200 genome sequences currently under way (www.jgi.doe.gov; www.wit.integratedgenomics.com/GOLD/; www.tigr.org). The nearly complete human genome sequence is the cornerstone of genome-based biology and provides the richest intellectual resource in the history of biology. The availability of entire genome sequences marks a new age in biology, because it has the potential to open innovative and efficient research avenues (25).

Determination of entire genome sequences, however, is only the first step in understanding the inner workings of an organism. The next critical step is to elucidate the functions of these sequences and give biochemical, physiological, and ecological meaning to the information. Sequence analysis indicates that the biological functions of substantial portions of complete genomes are unknown. Defining the role of each gene in the complex cellular machine and network is a formidable task (5, 11, 25). In addition, genomes contain hundreds to thousands of genes, many of which encode multiple proteins that interact and function together as multicomponent systems or apparatuses for accomplishing specific cellular processes. The products of many genes are often coregulated in complex signal transduction networks, and understanding how the genome functions as a whole to give life to complete organisms presents an even greater challenge (5). In addition, gene functions, protein machinery, and regulatory networks cannot be identified solely by using traditional single-gene, single-protein approaches. A single laboratory or institution also would find it difficult to encompass the breadth and depth of biological and technical expertise that is required for a comprehensive functional characterization of genome sequences (18).

Knowledge of entire genetic sequences opens a whole new range of possibilities for more efficient research. Thus, many laboratories are addressing important questions in functional genomics research by integrating genomic, proteomic, genetic, biochemical, and bioinformatic approaches. Consequently, areas in functional genomics and associated genomic technology are developing very rapidly. Annual conferences provide valuable 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 to remaining on 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). The 9th International Conference on Microbial Genomes is the continuation of the conference series on small genomes that originated in 1992 and was last held in Lake Arrowhead, Calif., in 2000 (29). During the past 8 years, this meeting has achieved the status of a major high-quality microbial genomics conference and has proven to be an important forum for information exchange in microbial genomics. All of these meetings have attracted leading scientists and institutions involved in genome sequencing, microbial functional genomics, and genomic and proteomic technologies.

OVERVIEW OF THE 9TH INTERNATIONAL CONFERENCE ON MICROBIAL GENOMES

The 9th International Conference on Microbial Genomes was held from 28 October to 1 November 2001 in Gatlinburg, Tenn. This conference focused on (i) defining gene functions and regulatory networks using integrative multidisciplinary approaches and (ii) exploring genome sequence information to understand various biological processes. Seven areas of microbial genomics were highlighted: (i) microbial genome diversity, evolution, and microbial genome sequencing; (ii) bioinformatics and microarray-based genomic technologies; (iii) proteomics, cellular pathways, and regulatory networks; (iv) functional genomics of bioremediation and carbon management; (v) functional genomics of microbial pathogens and relative microorganisms; (vi) functional genomics of model microorganisms, extremophiles, and biofilms; and (vii) applied functional genomics.

The following people served as members of the advisory committee for the 9th International Conference on Microbial Genomes: Jizhong Zhou (chair), from Oak Ridge National Laboratory, Oak Ridge, Tenn.; Jeffrey H. Miller (cochair), from the University of California, Los Angeles; George Weinstock, from the University of Texas, Houston; Monica Riley,

* Corresponding author. Mailing address: Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831. Phone: (865) 576-7544. Fax: (865) 576-8646. E-mail: zhouj@ornl.gov.

from the Marine Biological Laboratory, Woods Hole, Mass.; Elisabeth Raleigh, from New England BioLabs, Beverly, Mass.; and Terry Gaasterland, from The Rockefeller University, New York, N.Y.

The conference was designed by the organizers to be broadly representative of the microbial genomics research community. About 100 professionals from various countries presented their research in oral presentations and posters. The audience consisted of more than 200 researchers, postdoctoral associates, students, and others from 14 countries. With generous support from the U.S. Department of Energy (DOE), the National Science Foundation (NSF), the U.S. Department of Agriculture, and New England BioLabs, this conference contributed to the education of the next generation of genomic experts. Fellowship awards were used to support the attendance of 30 graduate students, postdoctoral associates, and young faculty members at this conference.

HIGHLIGHTS OF THE CONFERENCE PRESENTATIONS

Microbial genome diversity and sequencing. Microorganisms inhabit almost every imaginable environment on earth. In contrast to plant and animal diversity, however, the extent of microbial diversity is largely unknown. Although microbial genome sequencing projects reveal enormous amounts of information about a particular microorganism, these projects only scratch the surface of microbial diversity in general (8). Clearly, insight into microbial diversity will require the sequencing of the genomes of many microbial species from various environments.

Several presentations focused on genome sequencing and comparative genomics. Comparisons of genome sequences of closely related pathogens and nonpathogens have the potential to provide rapid and effective methods for understanding pathogenesis (25). To understand the genetic basis of pathogenicity, Carmen Buchrieser (Institut Pasteur, Paris, France) sequenced and compared the genomes of *Listeria monocytogenes* ESD and *Listeria innocua* (10). The former is a major human food-borne pathogen responsible for outbreaks of listeriosis, whereas the latter is a nonpathogenic species present in food. Comparative genomic sequence analysis indicated 270 genes in *L. monocytogenes* that are absent in *L. innocua*. These genes could contribute to or be responsible for the pathogenicity of *L. monocytogenes*. Also, the genes unique to *L. monocytogenes* were scattered across 100 different regions of the 2.94-Mb chromosome, suggesting that the specific attributes of virulent species were acquired through multiple events of horizontal gene transfer (HGT) and gene deletion. Genes specific to *L. monocytogenes* serovar 4b, which is associated with outbreaks of invasive disease, were also identified by comparing genome sequences. A similar strategy was used by Julian Parkhill at The Sanger Center, Cambridge, United Kingdom, to compare the genome sequence differences among *Escherichia coli*, *Salmonella enterica* serovar Typhi, *S. enterica* serovar Typhimurium, and *Yersinia pestis*. It was shown that differences in pathogenicity and host range appear to result not only from the acquisition and loss of large genetic islands but also from small groups of genes and even single base-pair differences.

Frederick Blattner, of the University of Wisconsin (Madi-

son), reported the completion of the genome sequence (4.6 Mb) of *Y. pestis* KIM, the etiologic agent of bubonic and pneumonic plague that has caused widespread loss of human life during recurrent pandemics. A remarkable amount of genome rearrangement caused by multiple inversions of genome segments was found in the two very closely related strains *Y. pestis* KIM and *Y. pestis* CO92 (4). About 54% of KIM open reading frames (ORFs) were significantly similar to those in *E. coli* K-12, but a number of *E. coli* pathways and transport systems were not identified. New genes encoding candidate pathogenicity proteins, such as iron transport proteins, adhesins, and toxins, were identified in KIM-specific islands.

In a similar study, Vivek Kapur (University of Minnesota, St. Paul) presented results from the complete genome sequencing and functional genomic analysis of *Pasteurella multocida*, a major pathogen of livestock that can also infect humans through animal bites. Despite the long and rich history of experiments and clinical investigations of *P. multocida*, no effective vaccines are available. Through genome sequence comparison (14), Kapur and his colleague discovered a very large gene encoding a putative virulence factor, the filamentous hemagglutinin. Since inactivation of this gene results in more than a millionfold reduction in virulence, this gene could be a good candidate for vaccine development.

HGT. HGT is genetic exchange between different evolutionary lineages. Comparative analyses of genomic sequences from various microorganisms suggest that HGT is a major evolutionary force in microbial speciation (17). The importance of HGT in microbial evolution was highlighted in several presentations at this conference. Gerhard Gottschalk, of the University of Gottingen (Gottingen, Germany), reported the genome sequence of *Methanosarcina mazei*, a methane-producing archaeon. Of the 3,300 predicted ORFs, about 1,100 genes may have been acquired from bacteria. It is amazing that so many bacterial genes are possibly horizontally transferred. HGT among microbial species is often observed, but this is the first indicated that so many genes are possibly transferred (20).

Frank Robb, of the University of Maryland Center of Marine Biotechnology in Baltimore, Md., presented another extreme example of interdomain gene transfer. He focused on *Carboxydothermus hydrogenoformans*, an obligate, carbon monoxide (CO)-utilizing, hydrogen-forming, gram-positive bacterium. While the major portion of the genome (2.1 Mb) shares a common descent with low-GC gram-positive bacteria, a very high proportion (24%) of the ORFs appear to be of archaeal origin. The genes involved in the CO utilization pathway, such as carbon monoxide dehydrogenase and heterodisulfide reductase, are very similar to those from methanogenic archaea.

Caroline Harwood, of the University of Iowa, Iowa City, further demonstrated the phenomenon of HGT with *Rhodopseudomonas palustris*. *R. palustris* is a purple, nonsulfur, photosynthetic bacterium and is among the most metabolically versatile of known bacteria for using carbon, nitrogen, and energy sources. Genome sequence analysis revealed the presence of circadian rhythm genes, which were not thought to be part of the repertoire of bacteria or archaea except cyanobacteria. In addition, this organism had an unusual cluster of photosynthetic genes that were very similar to those of a rhizobium that infects soybean.

Since HGT is a rapid means by which novel cellular func-

tions may be conferred to recipient strains and allows rapid, effective, and competitive exploration of new ecological niches, it is believed that HGT is an important mechanism for bacterial diversification. However, the introduction of new genes to recipient cells appears to be restricted by environmental barriers. James Lake, of the University of California at Los Angeles, presented results from a recent study of the role of HGT in the evolution of new genes. He and his colleagues found that not all genes were transferred freely and equally among all microorganisms. For example, high-temperature organisms preferentially obtained new genes from other high-temperature organisms rather than from organisms that grow at low temperatures. It appeared that genes were preferentially exchanged among microorganisms that grew in similar environments.

Functional annotation. Reconstruction of metabolic pathways from genomic sequence data reveals that many biosynthetic pathways are conserved among different organisms. However, genes encoding hundreds of enzymes and other proteins in important metabolic pathways in various organisms still remain to be identified. Andrei Osterman, of Integrated Genomics in Chicago, Ill., presented an integrated approach for the identification and verification of the "missing genes" involved in the metabolism of vitamins and cofactors, such as NAD, coenzyme A, and FAD. He and his group identified some candidate hypothetical proteins by using chromosomal clustering and other comparative genomics tools available in the ERGO platform (wit.integratedgenomics.com/IGwit/) and experimentally verified the essential roles of nicotinamide/nicotinate mononucleotide adenylyltransferase in cell survival. They also used whole-genome transposon mutagenesis to identify several enzymes in the biosynthesis of coenzyme A in *E. coli* as potential antibacterial targets, including phosphopantethenate adenylyltransferase (PPAT) (9). They characterized recombinant PPAT from several microbial pathogens and confirmed its essentiality in both gram-negative and gram-positive model organisms. Notably, human PPAT has no overall sequence similarity to its bacterial counterpart, thus suggesting that bacterial PPAT should be a good target for the development of new antibiotics.

One of the challenges in sequence homology-based functional annotation is that the relationships of genes to enzymes are not only one to one but also many to one and one to many. Such complicated gene-enzyme relationships present difficulties for sequence-based functional annotation. Monica Riley, of the Marine Biological Laboratory, showed that some superfamilies of enzymes may carry out very different reactions and that sequence similarity does not always mean that enzymes catalyze the same type of reaction. She also showed that some individual catalytic sites of certain proteins are capable of catalyzing as many as three reactions by using different catalytic triads to affect different reactions with different chemistries at one shared site. These results suggested that functional annotation should incorporate information on protein biochemistry and protein structure to accurately assign gene functions to sequences.

Substantial portions of the ORFs in sequenced microbial genomes are not functionally annotated by sequence homology-based approaches. Defining their functions is a major undertaking, and any functional clues from sequence comparisons will help guide experimental design for studying their functions. Kelly Oliner, of Protein Pathways in Los Angeles, Calif., presented four new non-homology-based methods for functional annotation of hypothetical proteins. These methods linked functionally related proteins based on phylogenetic profiling, protein fusion, and operon reconstruction. Based on the linkages to known proteins, many previously unannotated ORFs in *E. coli* were annotated with fairly high confidence. The non-homology-based approach could be useful for functionally mapping hypothetical proteins.

One of the challenges presented by large-scale genome sequencing efforts is effectively displaying the information in a format that is accessible to laboratory scientists. Owen White, of the Institute for Genomic Research in Rockville, Md., presented the Comprehensive Microbial Resource (CMR) (http://www.tigr.org/tigr-scripts/CMR2/CMRHomePage.spl). The web presentation of CMR includes a comprehensive collection of bacterial genome sequences, curated information, and related informatics methodologies. The sequences, functional annotation, and searching results can be effectively displayed in various formats in a manner that is convenient for laboratory scientists.

Microarray-based genomic technology and applications. Microarrays (or microchips) are a recently developed genomic technology and are listed as one of 10 breakthrough technologies in 1998, along with genomics (16). Just as microprocessors have dramatically sped up computation, microarray-based genomic technologies have revolutionized the genetic analysis of biological systems. Microarray technology represents a powerful new tool that allows researchers to view the living cell under various physiological states from a comprehensive and dynamic molecular perspective. The widespread, routine use of such genomic technologies will shed light on a wide range of important research areas, such as how cells grow, differentiate, and evolve; the medical challenges of pathogenesis, antibiotic resistance, and cancer; agricultural issues of seed breeding and pesticide resistance; the biotechnological challenges of drug discovery; and the remediation of environmental contamination.

In oligonucleotide microarray technology, molecules either are synthesized as an array in situ or are presynthesized and then arrayed. Jorg Hoheisel (Deutsches Krebsforschungszentrum, Heidelberg, Germany) presented a new photo-controlled in situ synthesis type of oligonucleotide microarray. Oligonucleotide synthesis, array hybridization, scanning, and analysis are carried out with a single machine. The array design is very flexible and fits the needs of individual experiments. In contrast to the Affymetrix in situ synthesis method, a new chemistry method allows oligonucleotides to be synthesized in the 5′-to-3′ direction in an array format. Therefore, the microarrays can be directly used for single-base polymorphism analysis through primer extension with on-chip polymerase reactions (12).

The power of the cDNA approach for gene expression analysis was clearly demonstrated by Michael Adams (University of Georgia, Athens) in his effort to understand the biology of *Pyrococcus furiosus*, a hyperthermophile that grows optimally at about 100°C. This organism is able to reduce elemental sulfur and produce hydrogen. Sulfur reduction was previously considered trivial and nonspecific in *P. furiosus*. However, anal-

yses with partial microarrays containing genes involved in the primary metabolic pathways, energy conservation, and metal metabolism indicated that elemental sulfur or its metabolites play a major regulatory role at the transcriptional level. The reduction of elemental sulfur appears to be accomplished by a new type of enzyme system involving uncharacterized hypothetical proteins (22).

Microarray-based hybridization assays have generated considerable interest in the past few years, but skepticism with regard to this technology still remains. Many researchers consider microarray-based studies expensive and non-hypothesis-driven descriptive research, i.e., "fishing" experiments (3, 15). The effectiveness of microarrays for addressing biological questions was demonstrated by Alex Beliaev, of J. Zhou's group at Oak Ridge National Laboratory. His group examined the expression profile of the putative regulatory gene *etrA* mutant of *Shewanella oneidensis* MR-1, a gram-negative, metal-reducing bacterium. Although sequence comparisons suggested that the *S. oneidensis* MR-1 *etrA* gene, which exhibits 74% sequence identity to *E. coli* Fnr at the amino acid level, is involved in energy metabolism, no significant difference between the wild type and an *etrA* insertion mutant was observed in terms of growth and utilization of various electron acceptors (2). The absence of a discernible physiological effect for *etrA* mutants could be explained by the following possibilities: (i) the *etrA* gene is not functional in *S. oneidensis*; (ii) *etrA* is functional but is not involved in the regulation of anaerobic energy metabolism; (iii) similar to *fnr*, the *S. oneidensis etrA* gene is involved in the regulation of anaerobic energy metabolism but is not essential due to the presence of other dissimilar genes that encode proteins with similar functions; and (iv) the genes of *S. oneidensis* involved in energy metabolism have multiple copies, some of which are transcribed independently of EtrA. To test these hypotheses, microarrays containing genes involved in energy metabolism were constructed to compare the gene expression profiles of the mutant and wild-type cells. The results indicate that *S. oneidensis etrA* influences the transcription of genes with predicted functions in energy metabolism, transcriptional regulation, substrate transport, biosynthesis, and other cellular processes, but there are not sufficient data to distinguish hypotheses 3 and 4.

Detection, characterization, and quantification of microbial communities pose major difficulties for microbial ecologists because the majority of naturally occurring species are not culturable. Although microarray-based genomic technology has great potential for overcoming the limitations of traditional molecular methods for studying microbial communities, adapting this technology to environmental studies from pure cultures is a challenge (28). The utility of microarray-based genomic technology in environmental studies was demonstrated by Dorothea Thompson, of J. Zhou's group at Oak Ridge National Laboratory. The performance of the prototype microarrays containing functional genes important in biogeochemical cycles or the entire genomic DNA was systematically and rigorously evaluated in terms of specificity, sensitivity, and quantification within the context of environmental samples. The results indicated that glass-based microarray hybridization has potential as a specific, sensitive, and quantitative tool for analyzing microbial communities in natural environments (28).

Proteomics. Proteins are the main catalysts, structural elements, signaling messengers, and molecular machines of biological systems. Recent progress in whole-genome sequencing has provided information on the composition of many proteins from a variety of organisms, but how proteins confer on cells their capabilities, structure, and higher-order properties remains to be clarified. Because virtually all changes in cellular states are correlated with fluctuations in the mRNA levels, transcription patterns can be very useful in elucidating the function of unknown genes, regulatory pathways, and networks (5). Translational regulation and posttranslational modification are also important in determining protein abundance in a cell. To understand how a proteome (i.e., a complete set) is produced and the proteins interact with each other, high-throughput proteomic approaches such as mass spectrometry (MS), phage display, and protein arrays must be employed. Unlike high-throughput nucleic acid-based approaches, the development of high-throughput techniques for the analysis of proteins is a great challenge. Although proteomic methods have emerged in recent years, they are still in the early stages of development, and several presentations at this conference highlighted recent progress in this area.

MS is currently one of the most important tools in proteomics. With the availability of complete genomic sequences, proteins isolated from a given cell type no longer need to be identified by de novo sequencing but can instead be identified by correlating the molecular masses of short peptides with those predicted from sequence database. MS is the choice for peptide identification due to its exquisite sensitivity (femtomole level) and mass accuracy (6, 13). Richard Smith, of the Pacific Northwest National Laboratory in Richland, Wash., has developed a new sensitive technology for proteome-wide identification and quantitation of proteins using a two-phase approach (24). During the first stage, a database of accurate mass tags is generated for tryptic peptides obtained from global cellular protein digestions. In the second stage, the database is used for high-throughput expression studies, where stable isotope labeling is used to create a reference proteome to which all perturbations are compared. The approach was successfully tested on a radiation-resistant bacterium, *Deinococcus radiodurans*, for which almost half of the proteins were identified. This approach also allows the detection of modified proteins and provides quantitative measurements of protein changes under different environmental conditions.

Cells can be considered biological "factories" that carry out and integrate thousands of discrete, highly specialized processes using molecular "machines" consisting of different proteins and other molecules. Many proteins assemble into larger complexes to execute fundamental cellular functions and metabolic processes (e.g., DNA replication, transcription, translation, and protein degradation), mediate information flow within and among cells and within their environment (e.g., signal transduction, energy conversion, and cell movement), and build cellular structures. Systematic identification of protein interactions in such complex molecular machines is very important in understanding how a given proteome works. The two-hybrid system is one of the most widely used approaches for generating comprehensive protein interaction maps (26). The main advantage of this approach is that no protein manipulation is required, because proteins to be tested are ex-

pressed in vivo by the yeast cells. The two-hybrid system has been applied on a relatively small scale for generating yeast protein interaction maps (26).

Marc Vidal, of Harvard University (Boston, Mass.), presented his recent large-scale protein interaction mapping efforts with the nematode *Caenorhabditis elegans*. The first step of mapping protein interactions was to clone all protein-coding ORFs of an organism to allow exogenous expression of its protein for functional analysis. By developing an automated version of the Gateway cloning approach, about 97% of the ORFs (~19,000) in *C. elegans* were cloned. The two-hybrid system was used to map interactions of 27 proteins involved in vulval development in *C. elegans* (27). The resulting map revealed both known and new potential interactions and provided a functional annotation for about 100 uncharacterized gene products. The results indicated that mapping protein interaction on a genomic scale is feasible for *C. elegans*.

Phage display is another powerful technique for studying protein-ligand interactions (19). The method involves the fusion of peptides or proteins to a coat protein of a filamentous bacteriophage (23). Because the gene encoding the fusion protein is packaged within the same phage particle, there is a direct link between the phenotype, i.e., the ligand-binding characteristics of a displayed protein, and the DNA sequence of the gene for the displayed protein. This permits large libraries of peptides of random amino acid sequences to be rapidly screened for desired ligand-binding properties (23). The power of phage display for studying protein-protein interactions was demonstrated by Timothy Palzkill, of the Baylor College of Medicine, Houston, Tex. He and his group studied genome-wide protein-protein interactions of *Treponema pallidum*, the causative agent of syphilis. From a total of 1,030 predicted ORFs, 1,008 genes were cloned, and their protein products were expressed for phage display analysis. The phage display library was used to identify the dominant antigens for an anti-*Treponema* rabbit polyclonal antibody serum. After several rounds of binding enrichment, a common antigen was identified. These results suggest that phage display is a useful means of defining gene function based on ligand-binding interactions.

Protein arrays are another recently emerged, promising, high-throughput technology for monitoring protein expression and interactions. Different "bait" proteins, such as antibodies, are immobilized on the surface of a solid substrate. The surface is then probed with the samples of interest, and only the proteins that bind to the relevant bait proteins remain on the microarrays. The major difficulty in screening an entire proteome is generating the necessary clones as well as achieving the expression and purification of the bait proteins in a high-throughput fashion (30). Metin Bilgin, of Michael Snyder's group at Yale University (New Haven, Conn.), convincingly demonstrated the power of protein microarrays. About 93% of the yeast genes (~5,800 ORFs) were cloned and expressed. The protein products were purified and printed on glass slides to represent the entire yeast proteome. The protein microarrays were used for screening proteins for their ability to interact with proteins and phospholipids. Many new calmodulin- and phospholipid-interacting proteins, both previously known and uncharacterized, were identified (30). This study demonstrated that microarrays of entire eukaryotic proteomes could

be constructed and used for screening of various biological activities.

Cellular modeling. Predicting metabolic activities and cellular behavior is an ultimate goal of genome-based biology. However, modeling global cellular behavior is a major challenge due to the complexity of metabolic pathways and our lack of understanding of the dynamic behavior and regulatory mechanisms. The most thorough way to analyze metabolic networks is through a dynamic model of cellular metabolism. However, dynamic models require detailed kinetic information about enzymes and cofactors, as well as ranges of substrate concentrations. Even though biological knowledge is rapidly expanding, cellular metabolism is not understood in mathematical detail (1). To overcome the lack of detailed kinetic information on cellular and enzymatic processes, an alternative approach is to exclude possible metabolic or cellular behaviors by imposing known constraints.

Bernhard Palsson, of the University of California at San Diego, presented pioneering studies in developing and applying constraints-based modeling approaches to several microbial systems, such as *E. coli*, *Haemophilus influenzae*, and *Helicobacter pylori*. Under given constraints, optimal phenotypes can be computed and compared to experimental data. The constraints-based model for *E. coli* quantitatively predicted the growth and metabolic by-product secretion data in batch, fed-batch, and continuous cultures and accurately predicted the metabolic capabilities of 73 of 80 mutants examined (7). The constructed model for *H. pylori* was able to accurately predict the metabolic phenotypes of 10 of 17 deletion mutants (21). The results showed that this is a very promising approach for simulating and predicting metabolic capacities and behavior of microbial cells.

Microbial genomics-related programs. To address the challenges of functional genomics, the DOE Office of Science has launched various microbial genomics-related research programs, such as the Microbial Genome Program (www.ornl.gov /microbialgenomes/index.html), the Microbial Cell Project (www.microbialcellproject.org), the Natural and Accelerated Bioremediation Research Program (www.lbl.gov/NABIR), and the Experimental and Computational Structural Biology Program (www.science.doe.gov/ober/msd_struct_bio.html). Recently, an ambitious new program, Genomes To Life (GTL; www.DOEGenomesToLife.org), was initiated. Daniel Drell, of DOE, highlighted this new initiative at this conference. This program is designed to exploit whole-genome sequence information for the purpose of comprehensively identifying, characterizing, and predicting the diversity and dynamic behaviors of protein machinery and genetic regulatory networks of microorganisms and communities that are important for energy, the environment, and national security. This program was built on the continuing success of the international human genome-sequencing project (initiated by DOE) and on numerous microbial genome-sequencing projects. The Offices of Biological and Environmental Research and Advanced Scientific Computing Research within the DOE Office of Science are implementing the GTL program jointly.

The overall objective of this program is to provide a fundamental, comprehensive, and systematic understanding of how genomes are used to define and control life processes in both simple and complex organisms, as well as in communities. It

will support the DOE missions related to sustainable sources of energy, environmental management, carbon cycle and sequestration, national security, and human health protection. Since microbes are essential to DOE's mission, this program will initially focus on microbial genomes and communities. Towards this overall objective, the GTL program has the following four goals: (i) to systematically identify and characterize molecular machines and provide the molecular basis and knowledge for linking information on protein complexes, their dynamics, and their intracellular organization with cellular and organismic functions; (ii) to systematically identify and characterize genetic regulatory networks and develop theoretical and computational tools for simulating genetic regulatory dynamics and cell physiology; (iii) to dramatically extend the current understanding of genetic diversity and metabolic capabilities of microbial communities in the environment, especially those related to bioremediation, biogeochemical cycles, climate change, and energy production; and (iv) to develop the next generation of computational methods and capabilities for simulating complex cellular networks and behaviors.

The NSF has also launched microbial genomics research at all levels of biological organization and complexity (http://www.nsf.gov). The foundation seeks to develop a comprehensive agenda in the area of genome-enabled microbial science that builds on genomic sequence information and is unparalleled in breadth and scope. The ultimate goal is to achieve a complete mechanistic understanding of the molecular components, assemblages, machines, and systems operating in a single cell and to understand how microbes sense environmental signals, how they adjust their metabolic activity to survive assaults or exploit opportunities, how they share or acquire new genetic information and evolve into new species, how they communicate or interact with their relatives, neighbors, and other species to develop complex communities on which ecosystems depend, and how these communities interact and influence the global continuum of life on earth.

The NSF endeavors to coordinate its microbial research and infrastructure programs with other U.S. government and international agencies through the Interagency Microbe Project Working Group (for the full Microbe Project Report, go to http://www.ostp.gov/html/microbial/2000microbial/start.htm.). One of the project's goals is to develop a coordinated national effort to sequence microbial genomes of broad biological interest and importance. To address this goal, the NSF and the U.S. Department of Agriculture have established an interagency Microbial Genome Sequencing program (http://www.reeusda.gov/1700/funding/rfamgsp.htm). The purpose of this interagency program is to support high-throughput sequencing of microbial genomes that are of fundamental biological interest, as well as those that are important to our national security, to the productivity and sustainability of agriculture and forestry, and to the safety and quality of the nation's food supply. The data and resources generated by this program are expected to pioneer new and dramatic developments in the area of microbial biology.

ACKNOWLEDGMENTS

REFERENCES

1. **Bailey, J. E.** 1998. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. Biotechnol. Prog. **14:**8–20.
2. **Beliaev, A. S., D. K. Thompson, M. Fields, L. Wu, D. P. Lies, K. H. Nealson, and J. Zhou.** 2002. Microarray transcription. Profiling of a *Shewanella oneidensis etrA* mutant. J. Bacteriol. **184:**4612–4616.
3. **Brenner, S. E.** 1999. Errors in genome annotation. Trends Genet. **4:**132–133.
4. **Deng, W., V. Burland, G. Plunkett III, A. Boutin, G. F. Mayhew, P. Liss, N. T. Perna., D. J. Rose, B. Mau, S. Zhou, D. C. Schwartz, J. D. Fetherston, L. E. Lindler, R. R. Brubaker, G. V. Plano, S. C. Straley, K. A. McDonough, M. L. Nilles, J. S. Matson, F. R. Blattner, and R. D. Perry.** 2002. Genome sequence of *Yersinia pestis* KIM. J. Bacteriol. **184:**4601–4611.
5. **DeRisi, J. L., V. R. Iyer, and P. O. Brown.** 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science **278:**680–694.
6. **Dongre, A. R., J. K. Eng, and J. R. Yates.** 1997. Emerging tandem-mass-spectrometry techniques for the rapid identification of proteins. Trends Biotechnol. **15:**418–425.
7. **Edwards, J. S., R. U. Ibarra, and B. O. Palsson.** 2001. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat. Biotechnol. **19:**125–130.
8. **Fraser, C. M., and B. Dujon.** 2000. The genomics of microbial diversity. Editorial overview. Curr. Opin. Microbiol. **3:**443–444.
9. **Gerdes, S. Y., M. D. Scholle, M. D'Souza, A. Bernal, M. V. Baev, M. Farrell, O. V. Kurnasov, M. D. Daugherty, F. Mseeh, B. M. Polanuyer, J. Campbell, S. Anantha, K. Y. Shatalin, S. A. K. Chowdhury, M. Y. Fonstein, and A. Osterman.** 2002. From genetic footprinting to antimicrobial drug targets: examples in cofactor biosynthetic pathways. J. Bacteriol. **184:**4555–4572.
10. **Glaser, P., L. Frangeul, C. Buchrieser, C. Rusniok, A. Amend, F. Baquero, P. Berche, H. Bloecker, P. Brandt, T. Chakraborty, A. Charbit, F. Chetouani, E. Couve, A. de Daruvar, P. Dehoux, E. Domann, G. Dominguez-Bernal, E. Duchaud, L. Durant, O. Dussurget, K. D. Entian, H. Fsihi, F. Garcia-Del Portillo, P. Garrido, L. Gautier, W. Goebel, N. Gomez-Lopez, T. Hain, J. Hauf, D. Jackson, L. M. Jones, U. Kaerst, J. Kreft, M. Kuhn, F. Kunst, G. Kurapkat, E. Madueno, A. Maitournam, J. M. Vicente, E. Ng, H. Nedjari, G. Nordsiek, S. Novella, B. de Pablos, J. C. Perez-Diaz, R. Purcell, B. Remmel, M. Rose, T. T. Schlueter, N. Simoes, A. Tierrez, J. A. Vazquez-Boland, H. Voss, J. Wehland, and P. Cossart.** 2001. Comparative genomics of Listeria species. Science **294:**849–852.
11. **Hieter, P., and M. Boguski.** 1997. Functional genomics: it's all how you read it. Science **278:**601–602.
12. **Hirschhorn, J. N., P. Sklar, K. Lindblad-Toh, Y. M. Lim, M. Ruiz-Gutierrez, S. Bolk, B. Langhorst, S. Schaffner, E. Winchester, and, E. S. Lander.** 2000. SBE-TAGS: an array-based method for efficient single-nucleotide polymorphism genotyping. Proc. Natl. Acad. Sci. USA **97:**12164–12169.
13. **Mann, M., P. Hojrup, and P. Roepstorff.** 1993. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol. Mass Spectrom. **22:**338–345.
14. **May, B. J., Q. Zhang, L. L. Li, M. L. Paustian, T. S. Whittam, and V. Kapur.** 2001. Complete genomic sequence of Pasteurella multocida, Pm70. Proc. Natl. Acad. Sci. USA **98:**3460–3465.
15. **Mir, K. U.** 2000. The hypothesis is there is no hypothesis. The Microarray Meeting. Trends Genet. **16:**63–64.
16. **The news and editorial staffs.** 1998. Breakthrough of the year: the runners-up. Science **282:**2157–2161.
17. **Ochman, H., and N. A. Moran.** 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science **292:**1096–1099.
18. **Oliver, S.** 1996. A network approach to the systematic analysis of yeast gene function. Trends Genet. **12:**241–243.
19. **O'Neil, K. T., and R. H. Hoess.** 1995. Phage display: protein engineering and directed evolution. Curr. Opin. Struct. Biol. **5:**443–449.
20. **Pennisi, E.** 2001. Sequences reveal borrowed genes. Science **294:**1634–1635.
21. **Schilling, C. H., M. W. Covert, I. Famili, G. M. Church, J. S. Edwards, and B. O. Palsson.** 2002. Genome-scale metabolic models of less characterized organisms: a case study for *Helicobacter pylori*. J. Bacteriol. **184:**4582–4593.
22. **Schut, G. J., J. Zhou, and M. W. Adams.** 2001. DNA microarray analysis of

the hyperthermophilic archaeon *Pyrococcus furiosus*: evidence for a new type of sulfur-reducing enzyme complex. J. Bacteriol. **183:**7027–7036.

23. **Smith, G. P.** 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science **228:**1315–1317.

24. **Smith, R. D., G. A. Anderson, M. S. Lipton, C. Masselon, L. Pasa-Tolic, Y. Shen, and H. R. Udseth.** 2002. The use of accurate mass tags for high-throughput microbial proteomics. OMICS J. Integrative Biol. **6:**61–90.

25. **Strauss, E., and S. Falkow.** 1997. Microbial pathogenesis: genomics and beyond. Science **276:**707–712.

26. **Walhout, A. J., and M. Vidal.** 2001. Protein interaction maps for model organisms. Nat. Rev. Mol. Cell Biol. **2:**55–62.

27. **Walhout, A. J., R. Sordella, X. Lu, J. L. Hartley, G. F. Temple, M. A. Brasch, N. Thierry-Mieg, and M. Vidal.** 2000. Protein interaction mapping in C. elegans using proteins involved in vulval development. Science **287:**116–122.

28. **Wu, L., D. K. Thompson, G. Li, R. A. Hurt, J. M. Tiedje, and J. Zhou.** 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl. Environ. Microbiol. **67:**5780–5790.

29. **Zhou, J., and A. V. Palumbo.** 2000. Sequence to function: the 7th Conference on Small Genomes. Genetica **108:**vii-ix.

30. **Zhu, H., M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. Dean, M. Gerstein, and M. Snyder.** 2001. Global analysis of protein activities using proteome chips. Science **293:**2101–2105.