Ninth Genome Sequencing and Analysis Conference

By Darrell Doyle, The Institute for Genomic Research

A similar report on this conference by Darrell Doyle appeared in *Microbial and Comparative Genomics* 2(4), 251 57 (1997). Conference details and abstracts can be found in *Microbial and Comparative Genomics* 2 (3), 1997.

The Ninth Genome Sequencing and Analysis Conference attracted over 1000 participants to Hilton Head Island, South Carolina, on September 13-16, 1997. In addition to beautiful weather, innovative science kept most of the participants at the conference for its entirety. The conference cochairs were Drs. Andre Goffeau (Universite Catholique de Louvain), Leroy E. Hood (University of Washington), and J. Craig Venter [The Institute for Genomic Research (TIGR)] who initiated these conferences in 1989.

Plenary sessions were devoted to Microbial Genomes, Microarraying Technology, Functional Analysis, Eukaryotic Genomes, Genome Biology, Advanced Technology and New Approaches, and Analysis of Human Genomes. There were three well-attended breakout sessions on Informatics, Technology, and Genomics as well as poster sessions from 1 to 4 p.m. on Sunday, Monday, and Tuesday.

Microbial Genomes

Progress in the sequencing and analysis of microbial genomes was presented at the first plenary session on Saturday evening. At least 10 microbial genomes have already been completely sequenced, and more than 25 others are in progress, with many close to completion. These organisms represent a broad spectrum of bacterial and Archaeal life.

Claire M. Fraser (TIGR) discussed progress in a comparative genome analysis of two spirochetes, *Treponema pallidum* and *Borrelia burgdorferi*. These organisms are human pathogens that cause syphilis and Lyme disease, respectively. Over 1 million new cases of syphilis per year are caused by *T. pallidum*, an organism that is difficult to grow in culture. Interestingly, before the advent of antibiotics, it was recognized that *T. pallidum* was very sensitive to high temperatures and that patients with syphilis could be infected with *Plasmodium falciparum*, the malaria-causing organism. The fever induced by malaria was sufficient to kill *T. pallidum*, thus curing the syphilis, and the drug chloroquine was effective in treating the malaria.

The genome sizes of *T. pallidum* and *B. burgdorferi* are about 1.15 and 1.3 Mb, respectively, but their structural organizations are markedly different. *B. burgdorferi*, which has now been completely sequenced (http://www.tigr.org), has a large linear genome but also contains up to 20 plasmids. The right end of the linear chromosome is relatively sparse in open reading frames (ORFs) specifying genes coding for proteins or RNA. Both *T. pallidum* and *B. burgdorferi*, similarly to *Mycoplasma genitalium*, are limited in their biosynthetic capabilities. *B. burgdorferi* in culture requires large concentrations of N-acetylglucosamine, which forms the chitin polymer. Fructose 6-phosphate and N-acetylglucosamine are interconvertible in the pathway of glycolysis. The proton-generating ATP synthase is present and needed to pump protons across the membrane.

Both *T. pallidum* and *B. burgdorferi* have spiral shapes due to the presence of internal flagella inside the cell wall in the periplasmic space. A significant percentage of each organism's genome is concerned with genes for flagellar structure and function, suggesting that motility is important to the survival of these organisms.

The complete genome sequence of *Bacillus subtilis*, a gram-positive bacterium, has been determined by a consortium of Japanese and European laboratories. As summarized by Antoine Danchin (Institut Pasteur), the 4,214,807-bp genome, of which 87% is coding, can specify more than 4000 genes. The task of obtaining this complete genomic sequence was made difficult because the *B. subtilis* DNA is very toxic in *Escherichia coli.* About half the ORFs specify single-copy genes. Around 60% of the genes could be assigned a sure or probable role or biological function. Another 25% had possible [probability lower than e10(-10) or putative e10(-6)] assignments, leaving about 15% with no role or function present in the searched databases. About 100 operons are conserved between *B. subtilis* and *E. coli*. Transcription of the *B. subtilis* genome is mainly in the same direction as replication. A significant proportion of the biosynthetic capacity of

B. subtilis is for utilization of carbohydrates, for interaction of the bacterium with plants, and for metabolizing branched chain acids. The *B. subtilis* genome also contains ten plasmids which, as in *B. burgdorferi*, may be involved in lateral gene transfer. (Consult http://genolist.pasteur.fr/SubtiList/)

Mycobacterium tuberculosis causes about 8 million new cases and 3 million deaths per year. Two strains of this organism are being sequenced: the virulent reference strain H37Rv (at Institut Pasteur and Sanger Centre) and a recent isolate CSU#93 (at TIGR). Robert D. Fleischmann (TIGR) and Stewart T. Cole (Institut Pasteur) reported progress in analyzing this major cause of death worldwide. Comparisons of completed sequences of the two strains' 4.4-Mb genomes reveal a high degree of similarity. There is a difference in IS sequences (16 in H37Rv vs 4 in CSU#93) and in direct repeats around the IS region. There is a large number of protein kinases and phosphatases in *M. tuberculosis* as well as multiple drug-resistance genes, virulence genes, ABC transporters, phospholipases, and 1300 to 1400 genes with no known function. Hence, there are many potential targets for new drugs and for the development of new vaccines.

Frederick R. Blattner (University of Wisconsin) made good use of the sequence of the *E. coli* K-12 laboratory strain, which his group completed recently, in comparisons with a highly pathogenic strain of this organism that usually resides in the intestines of cattle. *E. coli* 0157:H7 is responsible for many of the recent episodes of fast-food poisoning associated with hemolytic uremic syndrome. Only one to ten organisms per hamburger can produce symptoms due to the intestinal mucosa's invasion by two closely related shigella-like toxins that affect the kidney's glomeruli. The 0157:H7 strain has about a megabase of DNA that is not present in K12. Homology searches reveal a plethora of matches to virulence genes of a variety of species. Within a 350-kb segment surrounding a pathogenicity island, LEE, are 12 insertions ranging from 1 to 50 kb and 3 deletions. E. coli strain 0157:H7 (Strain EDL933) appears to be a complex mosaic of genetic elements superimposed upon a conserved backbone and not K12 plus a few additional plasmids.

Rickettsia prowazekii, an obligate intracellular parasite, causes typhus fever. This organism is an alpha proteobacterium and a member of the rickettsiacea, from which mitochondria may have evolved. Siv G.E. Andersson (Uppsala University) reported on analysis of the genomic sequence of *R. prowazekii*. About 700 kb of this organism's 1.1-Mb genome has been sequenced and annotated by shotgun sequencing of genomic libraries. Comparison of rickettsial sequences with the recently completed sequence of the mitochondrial *Reclinomonas americana* genome, which has the most eubacterial characteristics of any mitochondrial DNA, suggests that an ancestor of the mitochondrion was in fact derived from within the rickettsiacea.

Microarraying Technology

An entire session at the meeting was devoted to the new microarraying technology and its use in assessing gene expression, polymorphisms, disease, gene discovery, and mapping. Three main techniques are used to deposit small volumes of analytical probe onto a surface to produce ordered microarrays of high density, allowing massive parallel analysis to be done in a reasonably short time. Material is deposited by photolithography, mechanical spotting, or by piezoelectric "ink-jetting." Representatives from Affymetrix, Inc. (Robert J. Lipshutz) and Molecular Dynamics (Richard Johnston) reported on the different technologies being developed commercially. At Affymetrix, light-directed chemical synthesis is used to generate miniature high-density arrays of oligonucleotide probes. Instrumentation and software have been developed for all steps in the procedures, including hybridization to the arrays, fluorescent detection, and data acquisition and analysis. Microarraying technology has been used successfully to compare gene expression in cells growing in enriched Vs minimal medium, in resequencing already sequenced genes and expressed sequence tags (ESTs), and for identifying polymorphisms and disease-causing genes. A search of 15,000 sequence tagged sites, for example, revealed 1650 polymorphisms.

At Molecular Dynamics, DNA fragments are deposited directly onto low-fluorescent glass slides. This fragment-based fluorescent microarray technology using 9500 fragment spots per slide has been used, for example, to study gene expression in *E. coli* growing at 37, 42, and 48 degrees C. An analysis of 200 ORFs showed heat-shock gene expression increasing at 42 degrees, including groEL, and genes being shut down at the higher temperature.

Mark Schena (Stanford University) reported on biological microarraying of more than 10,000 analytical elements to examine expression of human genes. The detection limit was about one transcript out of a million. Two-color

fluorescent detection of gene expression in myelogenous compared to lymphoblastoid leukemia cells revealed an expression difference of only about 270 genes of the 10,000 analyzed in the two cell types.

Alan P. Blanchard (University of Washington, Seattle) showed that oligonucleotides synthesized using standard phosphoramidite chemistry will bond covalently to a glass surface after delivery by drop-on-demand ink-jet heads. Arrays of 1000 19-mer oligonucleotides could be generated on a standard glass microscope slide in 2 hours. Ultimate array densities are expected to be more than 5000 per square centimeter. Single-base mismatches were readily detected using these microarrays.

The take-home message of this session was that there is an exponential increase in the number of identified human genes, which in the very near future will approach 100,000 or the entire complement of expressed human genes. Microarraying technology will be needed to monitor or assay for the expression of these genes in normal development and aging, and many diseases or environmental challenges will manifest characteristic patterns of gene expression. The development of this technology is well under way.

Functional Analysis

At the session concerned with functional analysis, Goffeau reviewed the effort by 643 scientists worldwide to complete the genomic sequence of *Saccharomyces cerevisiae*. He emphasized that one-third or about 2000 of the identified yeast ORFs specify polypeptides with no known or a poorly defined function. EUROFAN, a European consortium of 134 laboratories, has begun to accelerate the functional analysis of these novel proteins, using disruption of each orphan gene as one component of the program. Peter Philippsen (Biozentrum, University of Basel, Switzerland) is coordinating this effort.

One question of interest to yeast biologists is why the yeast genome is so large. One way to examine essential yeast genes is to compare yeast with other free-living eukaryotic organisms that are syntenic but have smaller genomes. One such organism is *Ashbya gossypii*, a filamentous fungus with only seven chromosomes. A knockout introduced into this organism's genome resulted in slow growth and elongated cells, with effects on the spindle pole body. Reintroduction of a normal gene resulted in a normal phenotype. The mutated gene seemed to have two effects on the cell's phenotype, primarily on nucleus distribution during the cell cycle with a secondary-effect oncytokinesis, illustrating that interpretation of single-gene knockouts also can be complicated.

Rajeev Aurora (Johns Hopkins University) discussed the need for protein structural analysis as a tool for identifying function. The percentage of sequence similarity of homologous proteins in different organisms decreases as a function of evolutionary time. When this sequence similarity reaches a limit of about 30%, three-dimensional structure is required to assess homology. For example, a protein consisting of 100 amino acids could be formed by any combination of between 20 and 100 amino acids. Predictive methods for protein structure using primary amino acid sequence, such as threading or profile methods, have been developed and improved in recent years. These methods work best if the unknown structure is part of a family identified by multiple sequence alignment and work even better when the fold of another family member is already known.

At Johns Hopkins, a method based on a single sequence analysis is being developed to predict structure. With this method, no multiple sequence alignments or prior structural information is needed. The method was able to predict correctly the structure and homologs of triose phosphate isomerase with a beta-barrel fold. Order of the secondary structure, loop size of an unknown protein, and comparison of selected amino acids in the unknown protein with other proteins already in databases allows prediction of fold. Assigning a function to an identified or predicted fold can be difficult. For example, thymydylate synthase of *Methanococcus* has the fold identified as involved in folate recognition in homologous proteins, but Archaea do not have folate. Clearly, much work still needs to be done to identify biological functions for the large number of ORFs specifying proteins with unknown functions that are being identified in completely sequenced genomes.

Eukaryotic Genomes

Other eukaryotic genomes discussed at the meeting included the simple flowering plant *Arabidopsis thaliana* [Chris Somerville (Carnegie Institution of Washington at Stanford University) and Mike Bevan (John Innes Center, Norwich,

U.K.)]. Richard K. Wilson (Washington University, St. Louis) spoke on behalf of the C. elegans Genome Consortium, and Bruce Kimmel represented the Drosophila Genome Center, Berkeley, California. At least 15 Mb or about 20% of the relatively small *Arabidopsis* genome already has been completed, and sequencing efforts of six different international groups are well ahead of schedule. A density of about one ORF per 4 to 4.8 kb of DNA is expected for around 20,000 to 25,000 genes, with between 5 and 6 introns per gene. About 40% of the ORFs are represented in ESTs. Several approaches, including BAC end sequencing, are being used to ensure rapid completion of this plant's entire genome sequence, which will provide a background of information that will be useful in analyzing important crop plants.

About 70 Mb of the 100-Mb *Caenorhabditis elegans* genome has been sequenced. The DNA of this model organism is organized on 5 autosomes and the X chromosome with about 10,000 genes specifying polypeptides and about 400 RNA genes predicted. About half the ORFs specify polypeptides with database matches. Comparative genome analysis with a related nematode, *C. briggsae*, shows extensive conservation of exxon sequences and also possible regulatory regions outside the structural gene regions.

The Drosophila Genome Center has initiated a collaboration with Motorola to increase efficiency and accuracy by applying established manufacturing principles in the high-throughput complete genome sequencing of this model organism. The time required to finish the sequence of a *Drosophila* P1 clone (up to 1 year) has been one of the bottlenecks, but increasing the sample size in the shotgun random phase of cloning has reduced the cycle time to 3 months. Improvements in other program aspects should speed up the entire process significantly while maintaining a low error rate.

Genome Biology

A plenary session on genome biology included related and complementary lectures on membrane proteins and transport functions by Karen Ketchum (TIGR), Goffeau, and Milton Saier (University of California at San Diego). Based on a comparative analysis of already sequenced eubacteria and Archaea, transport functions clearly have been conserved throughout the evolution of prokaryotes. These functions are required both to bring things into cells and remove things from cells to the external milieu, including (1) amino acids, peptides, and amines; (2) carbohydrates, organic alcohols and acids, and ions; (3) cations, and (4) anions. They are needed to extrude or secrete such materials as toxins that are poisonous to the host of a pathogen and material that may be poisonous to the microorganism producing it. These systems are vital in energy production and in maintaining pH and other physiologically important conditions.

All microorganisms examined have sequences encoding transporters for potassium and the uptake of inorganic phosphate. They also have systems for extruding divalent cations. Autotrophic organisms have a higher proportion of transport systems involved in the regulation of cation and anion homeostasis. In contrast, in heterotrophs the uptake of organic substrates appears more important in regulating solute transport. Analysis of the hypothetical class of proteins for possible membrane localization by hydrophobicity or hydrophilicity ratios predicts that about 30% of proteins with unknown functions in all organisms are localized in cellular membranes, suggesting that many other transport functions await identification.

Of 5885 ORFs specifying polypeptides of 100 or more amino acids in *S. cerevisiae*, 2300 can be assigned as putative membrane proteins. There are many redundant genes in yeast; about 40% of the genome is duplicated. There are 130 yeast protein kinases, of which only one corresponds to a protein kinase C. There are no tyrosine protein kinases. A large number of proteins seem to be involved in multidrug resistance.

There are as many as 16 members of the ABC transporter superfamily. Four are implicated in multidrug transport, PDR1 (a transcription factor), PDR3, and ABC transporters PDR5, SNQ2, and YDR1. PDR5 is involved in extrusion of many different materials from the yeast cell, including detergents and anticancer drugs. Many compounds toxic to yeast are pumped out of the cell by more than one transporter. In some cases, two of three transport genes can be deleted and the remaining gene is still able to pump toxic material from the cell. There are genes for 180 permeases in yeast, 28 of which are involved in multidrug resistance but are not ABC transporters. Also in yeast are many homologs to human disease genes, including membrane proteins implicated in cystic fibrosis and Zellweger syndrome. Insight into how these proteins function can be gained by expressing the human disease gene in yeast, but overexpression often results in

a "traffic jam" in the endoplasmic reticulum while en route to the plasma membrane. These problems, however, should not be too difficult to solve, and this system should become an important resource for determining biological roles for disease-causing proteins.

An analysis of six sequenced prokaryotic organisms (*M. genitalium*, *M. pneumoniae*, *H. influenzae*, a species of the blue-green bacterium *Synechocystis*, *E. coli*, the archaeon *Methanococcus jannaschii*, and the eukaryote *S. cerevisiae*) for known and putative transport proteins allowed the identification of over 120 families of transport functions. These systems were classified by Saier and his colleagues according to mode of transport and mechanism of energy coupling (W), phylogenetic family (X), phylogenetic cluster or subfamily (Y), and substrate specificity (Z). This TC classification system W,X,Y,Z is comparable to the EC system of classifying enzymes in which the primary basis is the most fundamental function performed by the transport protein or enzyme. The ATP Binding Cassette (ABC) superfamily of transporters, with a 6 or 12 membrane-spanning domain topology, together with the Major Facilitator Superfamily (MFS) with a similar topology in the membrane, account for nearly 50% of all transporters in these organisms. Most transporters are present in *E. coli*, but 4 of the 62 identified are present in all prokaryotes. The number of solute transporters is roughly proportional to genome size, and a significant percentage of the solute transporters are also drug transporters.

The genomics program at Amgen is concerned with identifying novel gene products that could be targets for therapeutic intervention. Sid Suggs discussed the Amgen program to identify secreted proteins by using signal sequence homology and signal-trap methodology together with an EST-based strategy. The genes for these proteins are then expressed in transgenic mice to assess biological function. Expression, designed to mimic intravenous injection of the secretory protein, is driven by the APOE promotor and its liver-specific enhancer. Using this approach, several growth factors have been identified, including a secreted factor in the tumor necrosis factor superfamily and a novel secreted protein involved in bone resorption. This protein in transgenic mice inhibits the terminal differentiation of osteoclasts resulting in increased bone density. When injected into normal mice, the recombinant protein has the same effect. The approach demonstrates the identification of new therapeutic agents, in this case a possible agent for treatment of osteoporosis.

Nancy Hopkins [Massachusetts Institute of Technology (MIT)] is using the zebrafish as the model and insertional mutagenesis as the tool to identify and clone genes essential for early development. Zebrafish are a valuable model because early embryos are transparent and developmental abnormalities can be seen. Chemical mutagenesis already had identified over 2400 genes which, when mutated one at a time, affected development. Some 70% of these mutants were nonspecific, 20% were involved in apoptosis, and 50% had multiple effects. Mutations in 375 genes had specific effects on phenotype, but none affected cerebellum or blood. About 300,000 insertional mutations would be needed to profile completely the zebrafish's early development. Mouse retroviral vector with a VSV envelope protein marker could infect the zebrafish germline. In a pilot study, 1 of 70 insertions caused a mutant phenotype, and the insertional mutations resemble those caused by chemical mutagenesis. Multiple defects in liver, gut, eyes, and other sites were observed often. Based on the frequency of mutant phenotypes produced in the pilot screen, about 20% to 30% of embryonic lethal mutations in zebrafish could be isolated in the next few years.

Leroy Hood (University of Washington, Seattle) reported on progress in the continuing analysis of the three complex multigene families that encode for the alpha-delta, beta, and gamma polypeptides of the T-cell receptor (TCR). TCR beta loci in human and mouse consist of two interspersed gene families: trypsinogen genes and segments of TCR. The latter are 45 functional variable segments, 7 joining (J) segments. and two diversity (D) segments. The human and mouse TCR beta loci are the same overall size, but the human TCR variable segments have undergone a massive duplication compared to the mouse. Sequence similarity among TCR beta locus segments is a function of the evolutionary forces of gene duplication, gene deletion and mutation, and gene conversion acting over the course of 80 million years. Mutation and gene conversion acting in opposite ways make it difficult to reconstruct the molecular evolution or archaeology of this region. The 1.1-Mb alpha-delta contains 48 variable segments, of which 45 are expressed, and 5 olfactory genes. Comparison of this region with the same region in mouse is also yielding useful information on the organization of genomic DNA with respect to gene duplication and interspersion of repetitive elements. Finally, a dot-matrix comparison of the major histocompatibility genes of mouse and human reveals a high degree of similarity. The region is GC rich and gene dense.

Advanced Technology and New Approaches

Advanced Technology and New Approaches was the topic of one session at the meeting. These approaches included increasing the throughput of DNA sequencing and mapping using parallel array capillary electrophoresis (J. Bashkin, Molecular Dynamics), development of high-sensitivity dyes (BigDye) to fluorescently label primers and improve the quality of data from DNA sequencing (Michael W. Hunkapiller, PE Applied Biosystems), optical mapping as a single-molecule approach for generating ordered restriction maps from single molecules of DNA (David C. Schwartz, New York University), and arraying DNA for genome analysis (David R. Cox, Stanford University). Continued improvement of present methodology, development of such new technologies for genomic analysis as those presented at this session and in many of the posters, and improvements in bioinformatics lead to a feeling of confidence that complete DNA sequence analysis of very complex organisms is an attainable goal in a reasonable length of time.

Analysis of Human Genomes

The final day of the meeting was devoted to progress reports on human genome sequencing and analysis. Mark Adams (TIGR) reported on the collaborative effort of TIGR and the California Institute of Technology to sequence the short arm of chromosome 16. Bacterial artificial chromosomes (BACs) are being used as clones for sequencing. New and revised software programs are being developed for more efficient and faster assembly of sequences, closure, and annotation (see http://www.tigr.org). About 4 Mb of the chromosome 16 short arm has been sequenced. IS elements are present in four of the BACs currently in closure. One of these has 20 copies of a 60-bp tandem repeat. There are also duplicated regions greater than 20 kb long. The 16 p region is relatively gene poor, with a gene density of less than one gene per 100 kb of DNA.

LaDeana Hiller discussed progress at the Genome Sequencing Center at Washington University School of Medicine, St. Louis, Missouri. This group has concentrated their efforts primarily on chromosome 7, also using a BAC and PAC cloning approach. About 30 Mb of chromosome 7 sequence is available, with the largest mapped contig consisting of 2.2 Mb (http://genome.wustl.edu).

Richard A. Gibbs reported that the Baylor College of Medicine Human Genome Sequencing Center has deposited more than 4 Mb of complete sequence in the public databases, has 7 Mb in the pipeline, and expects to have 15 Mb by April 1998. The program's ultimate goal is to generate 100 Mb of sequence per year. This group is using the less expensive BODIPY fluorescent dye-labeled primers to generate most of this data. They are also comparing syntenic regions of human and mouse chromosomes.

Christopher Martin and Dan Drell gave a progress report on the Department of Energy's program to combine the genome-analysis efforts at Los Alamos National Laboratory, Lawrence Livermore National Laboratory, and Lawrence Berkeley National Laboratory into one unified Joint Genome Institute located in Walnut Creek, California.

Yoshiyuki Sakaki reported on the efforts of the sequencing team at the Human Genome Center, Institute of Medical Science, University of Tokyo, and the JST Sequencing Laboratory at Kitasato University, Tokyo. These groups are using a strategy based on P1 and cosmid-PAC cloning to sequence human chromosome 21. This chromosome contains a 1.6-Mb region, from D21S167 to ERG, that is critical for Down syndrome. The complete sequence of the region revealed seven genes, two of which are specifying unknown functions. Strategies being used produce 3 to 4 Mb of sequence per year, and efforts are being applied to extend these analyses to other parts of chromosome 21.

Bruce Roe (University of Oklahoma) gave a progress report on sequencing the proximal half of the q arm of chromosome 22. Almost 4 Mb of sequence has been generated using a P1, PAC, and BAC cloning strategy. Regions of this chromosome already sequenced include the Cat Eye region, the DiGeorge critical region, four meningioma deletion regions, the immunoglobulin light chain region, the GNAZ-BCR region, and the telomeric end of the q arm. The entire 1.5 Mb of DNA covering 22q11, which is involved in DiGeorge syndrome resulting in cleft palate and mental retardation, has been sequenced. About two dozen developmentally regulated genes in the region include two overlapping genes, several pseudogenes, repeated sequences, and alternatively spliced genes. An analysis of mouse syntenic chromosome 16 shows conservation of some sequences, but in human there is a clathrin heavy chain gene not present in the mouse.

Eric Lander (MIT) discussed the evolution of human variation studies, beginning with restriction fragment length

polymorphism (RFLP) as the first-generation method, then simple sequence length polymorphism, to the single nucleotide polymorphism (SNP) level of analysis. In 1.85 Mb of DNA analyzed by a reversed dot-blot method using Affymetrix chips, 1659 candidate SNPs were identified. Resequencing using PE Applied Biosystems methodology revealed three missed SNPs and about 11% false positives. In fact, a recheck of SNPs identified by sequencing also showed an 8% false-positive rate. Despite errors in SNP analysis by different methods, the study does indicate a large amount of polymorphism in the human population. A survey of different ethnic populations, including Caucasian, African, Asian, and Amerindians showed twice as many SNPs in the full sample vs the Caucasian and more variation in the African group than in the whole sample.

At the final session of the meeting, outlines of progress in human genome sequencing were presented by Glen Evans (chromosome 11 at the Genome Science and Technology Center, University of Texas Southwestern Medical Center), Jane Rogers (chromosomes 22, 6, 1, 20, and X at the Sanger Centre, http://www.sanger.ac.uk), Shawn Iadonato (chromosome 7 at the Human Genome Center, University of Washington, Seattle), Richard Myers (Stanford Human Genome Center), and Andre Rosenthal (chromosomes 7, 21, and X at the Department of Genome Analysis, Institute of Molecular Biotechnology, Jena, Germany, http://genome.imb-jena.de). Considerable progress is being made by all the groups in sequencing, accuracy (a single base-pair error rate of 0.01% or less), and developing hardware and software to keep pace with the goal of completing the human sequence early in the next century.

This conference was supported by a grant from the Department of Energy (DOE DE FG02 ER61892A.003). Support and contributions were also provided by Amersham Pharmacia Biotech; Amgen, Inc.; Mary Ann Liebert, Inc.; PE Applied Biosystems; TIGR and TIGR Science Education Foundation; Beckman; Biotechniques; Compugen, Ltd.; Digital Equipment Corporation; Genetics Computer Group, Inc.; Hewlett Packard; Merck Genome Research Institute; and Silicon Graphics, Inc.

Future Meeting

The Tenth International Genome Sequencing and Analysis Conference will held September 17-20, 1998, in Miami Beach, Florida. Cochairs will be J. Craig Venter (TIGR), Eric Lander (Whitehead Institute-MIT), Chris Somerville (Carnegie Institution, Stanford), and Mathias Uhlen (Royal Institute of Technology, Stockholm, Sweden).