BIOLOGY CHEMISTRY PHYSICS ENGINEERING MATHEMATICS

Human Genome news

A Forum for Human Genome Information Exchange

Sponsored by the U.S. Department of Energy

ISSN: 1050-6101 Vol. 8, No. 1, July–September 1996

Justice Faces the Genome: Trials and Tribulations

Educating the Courts to Meet the Body of Changing Scientific Evidence

For most visitors to Cape Cod, parades, picnics, and fireworks are the usual topics of conversation during the Fourth of July week. Among the throngs of vacationers this summer was a group of judges, science advisors, and others bent on more serious discussion during a week-long meeting sponsored by the Ethical, Legal, and Social Issues (ELSI) component of the DOE Human Genome Program. The topic: DNA testing and the massive impact it soon will have in courtrooms across the nation. The implications of DNA testing have the potential to overwhelm the legal system and shake its centuries-old foundations.

Consider this scenario:

Jeffrey Chase was adopted at 4 months in 1958, when adoption privacy was paramount and no information on biological parents was released. Jeffrey began exhibiting personality problems when he was 2 years old. Through the preschool, elementary, and high school years he displayed increasingly severe behavioral disorders. At 17, he assaulted a 4-year-old girl, choking and traumatizing her. He was tried for attempted murder, convicted, and sentenced to 30 years in prison. After serving 21 troubled years, he was released. Within 6 weeks he shot and killed his adoptive parents.

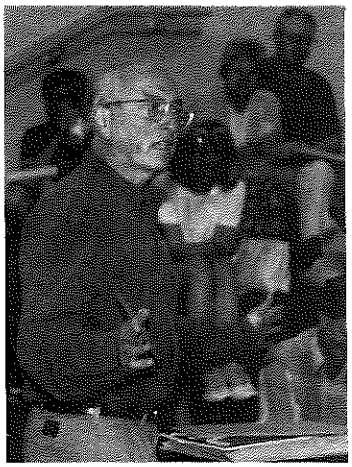
Could Jeffrey's genetic endowment have influenced his behavior? If researchers pinpoint genetic influences on aberrational behavior more precisely, will this rationale be used to absolve people of personal responsibility for their actions—the cornerstone of 600 years of Anglo-American law? Or will proof of an unalterable predisposition to violence cause a jury to choose the death penalty for a convicted murderer on the grounds that a person's genetic makeup precludes any possibility of rehabilitation in prison? What other ways might scientific advances be used to justify imposing the state's power?

Other Issues Looming

"Science is moving so fast the courts must scramble to cope," says Franklin Zweig, president of the nonprofit Einstein Institute for Science, Health, and the Courts (EINSHAC), a think tank housed on the life sciences corridor in Bethesda, Maryland. While advances accelerate in the Human Genome Project and the broader

(see Courts, p. 2)

See page 2, "Briefing Judges for Flood of Novel Cases"



Participants at the '96 Working Conversation on Genetics listen to Franklin Zweig (EINSHAC) introduce the program at the July meeting on Cape Cod, Massachusetts.

Third Branch of Life Confirmed

Researchers Present Archaea Genome Sequence

In a major scientific breakthrough, a team of DOE-funded researchers reported in the August 23 issue of *Science* (**273**, 1058–73) that they had sequenced the first complete genome of a microorganism that confirms the existence of the third major branch of life on earth [see *HGN* **7**(6), 12–13]. For the first time, researchers can trek boldly across uncharted terrain to make large-scale comparisons among the three domains of life at the genomic level. Researchers from The Institute for Genomic Research (TIGR); University of Illinois, Urbana (UIU); and Johns Hopkins University presented the sequence for *Methanococcus jannaschii*, a member of the Archaea domain of life. The other two major life groups are prokaryotes (bacteria) and the more complex

(see Archaea, p. 7)

In This Issue

Courts *(from p. 1)*

genomics communities, questions continue to mount about privacy, the fair and valid use of DNA information, and intellectual property. EINSHAC developed the "working conversation" meetings to allow judges to begin considering these issues before being confronted with them in court.

"A growing tide of evidence based on genetics threatens to engulf our legal system," Zweig said. "More than 30,000 judges across the country soon could be struggling with new questions: What genetic information is valid as evidence in criminal and civil cases? Does an individual have the right not to disclose results of a genetic test to family members, insurers, and employers? Whose privacy rights are more important in adoption cases? How might a parental predisposition to disease affect custody decisions? When should genetic tests be ordered in health-care decisions?"

Courts at the Front Lines

Legislatures and administrative agencies are attempting to address some aspects of these thorny issues (see sidebar, p. 3), but the courts do not have the luxury of time for extended debate. Court access is unlimited, and people will turn to litigation to resolve the growing list of dilemmas arising from the increased availability of genetic knowledge (see sidebar, p. 4). Early encounters with evidence for identifying DNA in criminal cases already have provoked controversy and confusion and resulted in over 150 varying DNA-admissibility decisions.

Because of the 1993 U.S. Supreme Court decision, *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, judges have greater responsibility to weed out unreliable "junk" science to ensure that the expert testimony is relevant and valid. The vast majority of judges, however, have no knowledge of genetics and are skeptical about the often-biased testimony of expert witnesses hired by opposing parties in legal cases.

A Changing Paradigm

Traditionally, courts looked to past precedents to guide their decisions, but the personal and familial nature of genomic information poses many unique and complex questions for which past decisions offer little collective wisdom. Judges will need assistance to effectively manage the highly complex genetics-related cases for which they will make rulings that chart new legal territory, Zweig says.

Judge Pauline Newman (U.S. Court of Appeals for the Federal Circuit in Washington, D.C.) says the problem is one of growing urgency. "We are reaching a critical threshold. If we understand the fundamentals of this new and powerful science, we can better weigh what the experts tell us in a partisan setting. It's our responsibility to become educated enough to decide the issues correctly—and wisely." Newman is the chairperson of EINSHAC's judicial advisory-and-review committee for the Genetics Adjudication Resource Project (GARP).

New Tools

EINSHAC established GARP in an attempt to develop solutions before courts become bogged down in genetic issues. GARP is supported in part by the ELSI component of the DOE Human Genome Program. GARP's dual goals over the next few years are to alert 1000 judges to the potential surge of cases involving genetic questions and provide resources to help them evaluate and adjudicate these issues. Resources include working conversations with judges, lawyers, and science advisors and a Web site for rapid information exchange and links to the courts. Future efforts will involve identifying changes the legal system could make to help courts apply genetics-related evidence more effectively.

The hoped-for bottom line: Judges who can keep junk science out of the courtroom. "We want an informed group of people in the courts who can recognize a bill of goods," says Daniel Drell, director of the DOE ELSI program. "The planning and educational efforts at the Cape Cod meeting represent an important step toward achieving that goal." *[Denise Casey, HGMIS]* ◊

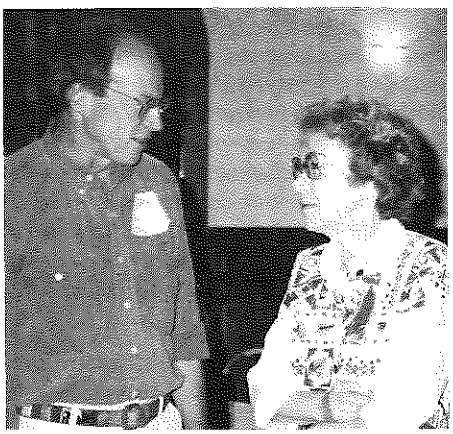
Briefing Judges for Flood of Novel Cases

'96 Working Conversation on Genetics

Judges, science advisors, and policymakers began to grapple with the tangle of emerging genetic issues and their impact on the courts in "working conversations" held during the past 2 years. These unique seminars were designed to introduce judges to genetic concepts while developing materials and recruiting faculty for as many as nine large regional judicial and science conferences beginning next year.

The 1996 working conversation, attended by about 35 federal and state judges and a dozen science advisors, was held July 1–7 on Cape Cod, Massachusetts, in the town of Orleans. Also taking part in the dialogue were legislators and other policy leaders who reviewed the efforts of Congress and administrative agencies to anticipate and handle ethical, legal, and social (ELSI) questions confronting them (see sidebar, p. 3). The meetings, supported by the DOE Human Genome Program's ELSI component, were organized by the Einstein Institute for Science, Health, and the Courts (EINSHAC) as part of a major judicial education effort (see lead article, p. 1). Highlights of this summer's meeting follow.

Aristides Patrinos, director of the DOE Human Genome Program, convened the meeting with a discussion of the Human Genome Project's growing impact, citing such spin-offs as the Microbial Genome Project (see



Judges William Rhea (District Court of Texas) and Pauline Newman (U.S. Court of Appeals for the Federal Circuit in Washington, D.C.).

Archaea article, p. 1). He emphasized the importance of identifying and attempting to resolve ELSI concerns arising from data produced by genomic research.

Conversation '96 Sidebar

Genetic Privacy and Property: Perspectives from Capitol Hill

Genetic information is highly personal and unique. The potential for its misuse threatens to penetrate many aspects of life, including employment, insurance, law enforcement, finance, and education. "Social policy needs to keep up with Human Genome Project discoveries," said U.S. Rep. Louise Slaughter (D-NY), participant in the 1996 Cape Cod working conversation on genetics and cosponsor of H.R. 2748, a bill that would outlaw health-care discrimination based on genetic information.

"Everyone is vulnerable. We are drafting laws to make sure people will not be left behind," she said. Slaughter believes an individual's right to privacy and control will be recognized as a new civil right.

Privacy Law

Dan Freeman (Committee on the Judiciary staff, U.S. House of Representatives) summarized recent congressional activity to protect genetic information. Several bills calling for uniform national standards have been introduced in Congress; H. R. 3103, the Health Care Insurance Portability and Accountability Act of 1996, was signed into law in August. It prohibits the use of genetic information in certain health-insurance eligibility decisions and requires the Department of Health and Human Services to establish standards in preparation for enforcing the health information privacy provisions.

"We have the opportunity to protect human rights before technology overwhelms us," Freeman said, "and we may be a bit ahead of the curve. The challenge is to minimize the harms."

Employment Protection

Last year the Equal Employment Opportunity Commission (EEOC) issued guidelines regarding genetic discrimination under the Americans with Disabilities Act [*HGN* 7(2), 4]. Speaking at the meeting, Peggy Mastroianni, EEOC Associate Legal Counsel, noted that the commission's interpretation still leaves gaps that the courts will be asked to fill.

Patenting Issues: Biotech Controls

What role should the federal government have in biotechnology issues? Tom Mooney (Committee on the Judiciary staff, U.S. House of Representatives) discussed controversies over patenting genes and products of genetic recombinant technology such as animals harboring foreign genes. "The science is intermingled with politics, business, and the nation's economy," he said, noting over $9.3 billion in annual sales by 1300 U.S. biotechnology companies. Mooney also acknowledged public concern over unique ethical, religious, and safety issues related to biotechnology.◊

In addition to intense discussions focusing on legal and scientific issues, popular features at the meeting included a laboratory demonstration on DNA fingerprinting (see sidebar, p. 5) and a hands-on computer workshop showcasing a prototype version of the new online journal *CASOLM* (*Courts and Science On-Line Magazine, http://www.ornl.gov/courts/*). Created by Hazel Witte (EINSHAC) with technical assistance from *HGN* staff, *CASOLM* features simple explanations of science, hypothetical cases, and links to more detailed information. A core of 20 pilot courts will review *CASOLM* and provide feedback and materials for a year, after which the magazine will be distributed to 21,000 federal and state courts and will be available to all parties via the Internet.

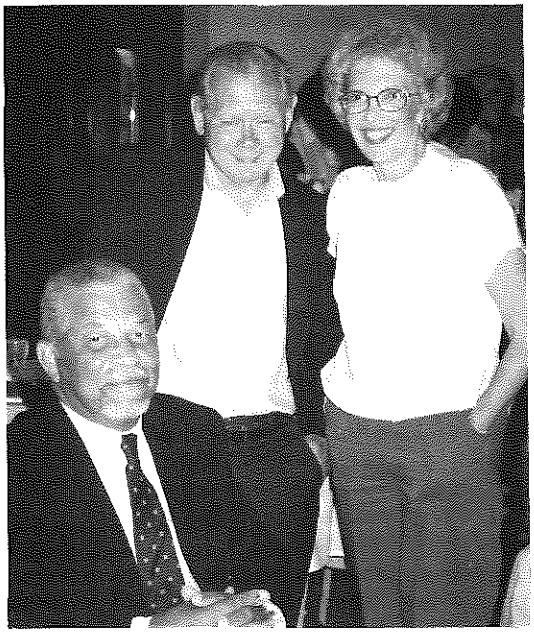
Linking Law and Science

At the meeting, legal and genetic points were considered in the context of imaginatively written case scenarios, most based on composites of actual cases. "There is little common language between the scientific and legal communities," notes EINSHAC's President Franklin Zweig, "but case histories can cut through and demystify the subject matter for both. They provide a common focus for discussion and help participants empathize with problems that put ordinary people in extraordinary cases."

Reviewing the facts for the first case history, Judge Gladys Kessler (U.S. District Court for the District of Columbia) observed that while the science is new, the fundamental legal concepts are not. She encouraged participants to try to meld the science with the traditional legal framework and analysis methods as much as possible. Building bridges between law and science will not be easy, she warned, but it is essential if people are to have confidence in our legal system.

"The law is years behind where the science is now," Kessler said, "and the span will lengthen as science jumps farther ahead. We must struggle to find a way to deal with the disparity, or legal solutions will become irrelevant. The law is a staid and lumbering mechanism for social control, while science moves at lightning speed. We need to find ways to harmonize the slow pace of law with the rapid pace of science."

Thus charged, participants turned their attention to analyzing the spectrum of genetic, legal, and ethical issues posed



Chief Judge Eugene Hamilton (Superior Court for the District of Columbia), at left. At right are Judge Gladys Kessler (U.S. District Court for the District of Columbia) and her husband Arthur Mackwell.

by the hypothetical case histories. Science advisors explained relevant concepts of molecular biology and genetic testing, and small groups formed to hash out the issues.

Case Scenario: Fired After Alzheimer's Disease Diagnosis

Catherine F., a model employee who was 2 years from retirement with full benefits, was fired abruptly. The reason: her employer had been informed by the company's health insurance carrier of an increase in premiums due to Catherine's recent diagnosis of Alzheimer's disease by her family physician. Her court-appointed lawyer filed suit for discrimination under the Americans with Disabilities Act. Catherine's daughter filed a petition for guardianship. Functional, physical, psychiatric, and genetic tests were performed, all with ambiguous results.

A key issue: Is genetic-testing evidence admissible in this case?

Reviewing the facts for the group, Judge Barbara Rothstein (U.S. District Court for the Western District of Washington) asked participants to consider the case in the context of a pretrial hearing to determine admissibility of scientific evidence. In federal and some state courts, this type of hearing is called a Daubert hearing. After careful consideration of the reliability and usefulness of the evidence offered, judges decide whether juries should be allowed to hear it.

Under Daubert, new or established scientific evidence must be based on demonstrably valid methodologies and principles. Other factors, such as the extent of peer review and the older Frye rule's standard, may also be used.

Based on a 1923 District of Columbia lie-detector case, Frye gives paramount importance to general acceptance in the scientific community. In a Frye hearing (unlike Daubert), judges perform little threshold evaluation of the reliability of scientific opinion evidence. Frye is used in many state courts.

Breakout groups agreed that genetic testing for Alzheimer's Disease would not pass either Daubert or Frye criteria because the scientific and medical communities currently consider this test unreliable for either diagnosing the disease or predicting a susceptibility to it.

Ongoing Battles: Admissibility in Criminal Cases

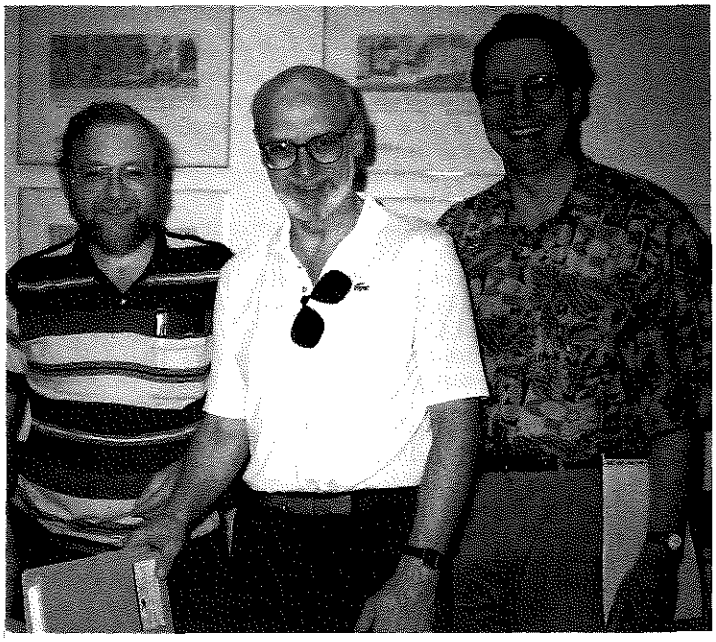
Judges at the Cape Cod meeting were particularly interested in one of the most powerful and controversial uses of DNA technology in the legal arena: its applicability to criminal jurisprudence, specifically for comparing the DNA found at a crime scene with that of a person suspected of the crime. David Bing and Janice Williamson (Center for Blood Research, Boston) demonstrated how forensic scientists generate DNA profiles.

Conversation '96 Sidebar

A Look at Future Cases: Courting Disaster?

Seminar participants agreed that, once lawyers understand the use that can be made of genetic information, the impact on the courts will be enormous. A cornucopia of cases anticipated by the group includes the following:

- Adoptions: Rights of adoptive parents to know the child's genetic makeup; the natural parents' (and on occasion grandparents') privacy rights; rights of the child.
- Criminal cases: DNA identification and defenses involved in genetic predisposition to criminal behavior; a defendant's right to have an independent examination of evidence; review of changing DNA-identification technology as the science evolves away from the use of DNA markers and toward direct genomic sequencing.
- Parent and child cases: Balancing the parent's right to know the child's likelihood of developing a disease against the child's right not to know; decisions to abort based on prenatal testing (whose choice and who represents the child?); effects of disease predisposition on custody decisions.
- Civil rights: Relief from perceived discrimination because of inherited genomic characteristics, based on the Americans with Disabilities Act, Civil Rights Act of 1964, and various other federal and state laws.
- Personal injury: Genetic testing, prenatal and adult gene therapy, and safety issues involving new biotechnology products.
- Health care: Assessment of genetic diseases and predispositions in court-ordered health-care cases; malpractice cases from failure to provide carrier or prenatal testing or failure to offer genetic testing where known history indicates disease potential.
- Patents: Proposals to integrate ethical and religious considerations with patent laws and decisions.◊



From left, Daniel Drell (DOE Human Genome Program), Daniel Freeman (U.S. congressional staff), and Eric Fischer (National Academy of Sciences) at the Cape Cod working conversation on genetics and the courts.

In the debate over the admissibility of DNA forensic evidence, experts on both sides agree that the techniques and scientific principles underlying genetic testing for identification, called DNA profiling, are valid (see sidebar below). For years, bitter disagreement over the admissibility of DNA evidence centered on calculating the statistical probability that two people could share the same set of markers and produce the same DNA profile.

What Are the Odds? New NRC Recommendations. In an attempt to resolve this question, this spring the National Research Council (NRC) published a second set of recommendations for performing calculations, handling DNA samples, and other aspects of using DNA as a forensic tool. [*The Evaluation of DNA Forensic Evidence*. NRC, 1996. Orders: National Academy Press (800/624-6242 or *http://www.nap.edu/bookstore/*)]

Eric Fischer (National Academy of Sciences) explained the updated recommendations of this report (called NRC2), particularly those focusing on interpretation of matching DNA profiles. Regarding the latter, NRC2 recommends use of the "product rule" (based on population-genetics principles) to calculate the probability that a match between two DNA profiles is due to chance.

In applying the product rule, scientists determine how frequently a specific marker occurs in a particular population by using databases of DNA profiles from black, Caucasian, or Hispanic populations. Individual probabilities are multiplied to obtain the overall probability that the composite profile will occur in the population. Because the reference databases are now quite

Conversation '96 Sidebar

Creating and Comparing DNA Profiles

Only one-tenth of a single percent of DNA (about 3 million bases) differs from one person to the next. Scientists can use these variable regions to generate a DNA profile of an individual, using samples from blood, bone, hair, and other body tissues and products.

In criminal cases, this generally involves obtaining samples from crime-scene evidence and a suspect, extracting the DNA, and analyzing it for the presence of a set of specific regions of DNA (markers).

Scientists find the markers in a DNA sample by designing small pieces of DNA (probes) that will each seek out and bind to a complementary DNA sequence in the sample. A series of probes bound to a DNA sample creates a distinctive pattern for an individual. Forensic scientists compare these DNA profiles to determine whether the suspect's sample matches the evidence sample. A marker by itself usually is not unique to an individual; if, however, two DNA samples are alike at four or five regions, odds are great that the samples are from the same person.

If the sample profiles don't match, the person did *not* contribute the DNA at the crime scene.

If the patterns match, it means that the suspect *may have* contributed the evidence sample. While there is a chance that someone else has the same DNA profile for a particular probe set, the odds are exceedingly slim. The question is, How small do they have to be when conviction of the guilty or acquittal of the innocent lies in the balance? Many judges consider this a matter for a jury to take into consideration along with other evidence in the case. Experts point out that using DNA forensic technology is far superior to eyewitness accounts, where the odds for correct identification are about 50:50. According to Eric Fischer (National Academy of Sciences), DNA-profiling reliability now falls between testing for blood groups and analyzing dermal (skin) fingerprints.

The more probes used, the greater the odds for a unique pattern and against a coincidental match, but each additional probe adds greatly to the time and expense of testing. NRC2 (see "What Are the Odds?" in article above) recommends using four to six probes. Within a year, testing with several more probes will become routine, observed John Hicks (Alabama State Department of Forensic Services). He predicted that, within the next 3 to 5 years, DNA chip technology (in which thousands of short DNA sequences are embedded in a tiny chip) will enable much more rapid, inexpensive analysis using many more probes, raising the odds against coincidental matches.

Forensic scientists look forward to the day when DNA sequencing technologies have progressed to the point where direct characterization of very large DNA segments, and possibly even whole genomes, will become feasible and practical. Then base-by-base comparison of unique genomes finally will enable scientists to declare a perfect match.◊

David Bing (Center for Blood Research, Boston) demonstrates DNA profiling to the '96 Working Conversation attendees during a laboratory workshop. Observing, from left, are Judges Ricardo Urbina and Gladys Kessler (U.S. District Court for the District of Columbia) and Barbara Rothstein (U.S. District Court for the Western District of Washington).

large, they provide confidence that the frequencies obtained are representative of those in the actual populations.

New Focus on Quality Control. Early reaction to NRC2 recommendations has been positive [*Science* **272**, 803–4 (May 10, 1996)]. Major critics of the first NRC report (1992) on DNA forensic technology now say the focus for concern has moved from statistical calculations to laboratory quality and accreditation. Although the NRC2 report calls for laboratory proficiency testing, it also recommends splitting DNA samples to enable the defense to verify results independently. The report also notes a need for research into jury understanding of DNA evidence.

Burden on the Court. NRC2 still places the onus for evaluating matching DNA profiles squarely on the courts. When can a DNA profile be considered "unique enough?" When the probability of a chance match is 1 out of 1000, 10,000, or even more? Some would like scientists to make this call, but the scientists say it isn't their job.

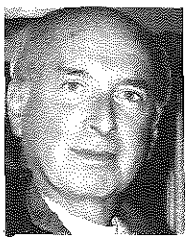
Underlying the friction rests a common misperception of scientific knowledge as a body of immutable truths. Scientists never make this claim, knowing that even the most carefully crafted and tested hypothesis can be disproven by a single example. As Joe McInerney (Biological Sciences Curriculum Study) noted, "Nature has a tendency to embarrass us when we make absolute statements." A robust system of rigorous hypothesis testing and peer review does, however, enable newfound knowledge to gain acceptance in the scientific community.

Justice Victoria Lederberg (Supreme Court of Rhode Island) questioned the concern over numbers, asking whether it is wise to seek absolute consensus from science. After all, courts have dealt with very complex material before, and judges usually have not defined "beyond a reasonable doubt" in terms of a numerical standard. She observed that although DNA technology is new and complex, it still can be absorbed during a trial and can be helpful to the jury in making decisions. DNA tests usually are not the only evidence available, she added. Statistical probabilities affect the weight to be accorded the evidence, not its admissibility, and that weight is rightly decided by the jury.

If judges demand a consensus to facilitate their role as evidence gatekeepers, they may want to consider creating a forum similar to one used by NIH to generate a consensus on a controversial scientific or medical topic, suggested John Ferguson (NIH Medical Applications Research). His advice to judges: Review all the data and have it interpreted by independent experts who have not performed the research.

Scaling Up the Conversation

As the end of the holiday week drew near, Zweig reaffirmed the value of the Cape Cod working conversations in building the foundation needed to deliver these ideas within a conversational format to 1000 judges. He emphasized the importance of this daunting task. "We depend on science for technological power and on the courts for the power of enforcement. For civil society to endure, the courts must prevail," he said. "Helping the courts cope with the challenges and master the perturbations created by genetics is a worthy and achievable goal." *[Denise Casey, HGMIS]* ◊



Franklin Zweig

Hazel Witte

For more information about the working conversations, contact Franklin Zweig or Hazel Witte at EINSHAC; Three Bethesda Metro Center, Suite 750; Bethesda, MD 20814 (301/961-1949, Fax: /913-5739, *einshac@aol.com*).

Conversation '96 Sidebar

Searching for Truth in the Court System

In 1995, over 15 million criminal and civil cases were filed in U.S. federal and state courts.

The federal court system is small, with about 550 trial judges who preside in U.S. district courts and around 180 appellate judges. Federal courts, whose judges are appointed for life, hear about 2% of all cases filed. The remaining cases fall to about 29,000 judges presiding over separate and autonomous state court systems. Many of these judges are elected to office for a specified term. The overwhelming majority of cases involving law enforcement, commerce, and human relations are decided in state trial courts.

Observers are concerned that neither court system is prepared to handle cases involving complex genetic information. If courts can't screen and present the science so that juries can understand and process it in a fair way, verdicts may appear to lack foundation and create a crisis of public confidence in the courts. "The battle of experts often impedes the search for truth," notes Judge Ronald Reinstein (Superior Court of Arizona). "Judges have a responsibility to be proactive and mandate clarity and simplicity."

Arizona courts have adopted far-reaching jury reforms that permit jurors to ask questions, take notes, and discuss cases as they develop. Most courts do not, said Franklin Zweig (EINSHAC). "Too frequently," he added, "courts blind, gag, and otherwise limit jurors. Jury reform must keep pace with judicial screening of scientific evidence." In 1997, Arizona is expected to host the first large regional working conversation on genetics for judges and lawyers.◊

Archaea *(from p. 1)*

eukaryotes, which include plants, animals, and humans.

"This brings to closure the question of whether Archaea are separate and distinct life forms," said Craig Venter (TIGR). "In decoding the genetic structure of Archaea, we were astounded to find that two-thirds of the genes do not look like anything we've ever seen in biology before." According to the *Science* paper, only about 38% of *M. jannaschii* genes match a gene with a known cellular function already entered in sequence databases. Comparisons also were made with genes found in the genomes of the other two complete microbial genomes sequenced by TIGR, those of *Mycoplasma genitalium* and *Haemophilus influenzae*.

More extensive gene comparisons will help scientists better understand the evolution of all three branches of life. Early analysis points to a closer shared evolutionary heritage for the Archaea and eukaryotes. This is particularly evident in the genes controlling information processing: transcription, translation, and DNA replication. Features shared with bacteria include the lack of a nucleus and strongly similar metabolic genes. "We can look at the Archaea as the living fossils of our prokaryotic [bacterial] ancestors," observed Carl Woese (UIU).

Random Sequencing Strategy

Sequencing of the 1.7-Mb *M. jannaschii* genome, which consists of three distinct genetic elements, was completed by a team led by Carol Bult (TIGR) in just over a year. Researchers applied a whole-genome random approach, using both a small-insert (2.5-kb average) plasmid library and a larger-insert (16-kb average) lambda library. All clones were sequenced from both ends, with the lambda library used to verify contigs built from assembled plasmid sequences. Data, including double the number of genes and proteins previously known for Archaea, are available on the Web (see box).

Completely Sequenced Genomes*

- *Haemophilus influenzae* (bacterium), 1.9 Mb
- *Mycoplasma genitalium* (bacterium), 0.580 Mb
- *Methanococcus jannaschii* (Archaea), 1.7 Mb
- *Saccharomyces cerevisiae* (eukaryote), 12 Mb

*Does not include viruses.

M. jannaschii and Other Archaea on Web

- DOE press release (*http://www.ornl.gov/hgmis/archive/methanoc.html*)
- Sequence data: Genome Sequence Data Base (accession numbers L77117, L77118, L77119)
- Annotated sequence and clone data: TIGR site (*http://www.tigr.org/*)
- Phylogeny, metabolism, alignments: PUMA database from Argonne National Laboratory (*http://www.mcs.anl.gov/home/compbio/PUMA/Production/puma.html*). Integrates genomic sequence and biochemical data from Archaea and other organisms within a functional context.
- "Earth & Sky" Web sites (*http://www.earthsky.com/1995/es951026.html* and *http://www.earthsky.com/1995/es951027.html*)

Fulfilling a DOE Mandate

DOE and its predecessor agencies have a long history of support for genetic research growing out of their legislative mandate to understand the health effects of nuclear energy and radiation and the byproducts of other forms of energy production. DOE funded the Archaea research as part of its Microbial Genome Program (MGP), a complementary project to the Human Genome Program. MGP is headed by Jay Grimes of the DOE Office of Health and Environmental Research (*darrell.grimes@oer.doe.gov*).

Begun in 1994, MGP's goal is to sequence microorganisms of interest to DOE's energy and environmental cleanup programs. *M. jannaschii* is the second genome completed in MGP. The first, *M. genitalium*, is a bacterium thought to be the simplest known self-replicating and free-living life form.

(see Archaea, p. 8)

Archaic Overachiever Thrives in Hostile Environments

First discovered almost 20 years ago by Carl Woese and Ralph S. Wolfe (both of University of Illinois, Urbana), the Archaea domain (whose name means "ancient" in Greek) is believed to have separated from true bacteria over 3 billion years ago. Archaea once were thought to live only at extreme environmental conditions of temperature and pressure but now are believed to be far more common and to make up a significant part—perhaps half—of the world's biomass. They are suspected of playing important but still unknown roles in the earth's ecology, including its carbon and nitrogen cycles.

The single-celled, 1738-gene *M. jannaschii* was isolated from a sample collected in over 8000 feet of water at the base of a deep-sea thermal vent on the floor of the Pacific Ocean. It is named for Holger Jannasch of the Marine Biological Laboratory in Woods Hole, Massachusetts, who collected the sample. Thriving at pressures that would crush a conventional submarine, this heat-loving, methane-producing microbe lives without sunlight, oxygen, or organic carbon.

Instead, it uses carbon dioxide, nitrogen, and hydrogen expelled from the thermal vent for its life functions. Analysis of the microbe's genome will provide researchers with valuable information for understanding how organisms can make life's building blocks from inorganic sources and under such extreme conditions.

Practical Payoffs

With its unusual characteristics, *M. jannaschii* has the potential to supply fuel and other ingredients for products from plastics to pharmaceuticals. Commercial interests now have the opportunity to develop such heat-resistant products as detergent additives or stable enzymes for the textile, paper, and chemical industries.

Methane (CH_4) causes both ozone production and depletion but with a net production of ozone. This mean that more knowledge about bacterial/archaeal methane production could lead to better understanding of global-warming processes.

Some of the following areas may benefit from *M. jannaschii* applications.

- Transportation: Develop "biological" vehicles.
- Energy: Generate large supplies of safe, renewable power.
- Weather: Understand and control methane's contribution to global warming.
- Environmental cleanup: Use biological methods to clean up hazardous waste sites.
- Household use: Manufacture biodegradable detergents and cleaners.◊

BAC End-Sequencing Projects Initiated

New Strategy Bypasses Contig Mapping

Assembling ordered, overlapping sets (contigs) of high-quality, sequence-ready clones has long been considered an essential step toward human genome sequencing. Not only do the clones provide uniform materials for sequencing, but, because they have been mapped to precise genomic locations, the DNA sequence obtained from them can be located on the chromosomes with minimal uncertainty. Very useful low-resolution maps have been produced by several methods. [See, for example, summary maps at the MIT–Whitehead Institute (*http://www-genome.wi.mit.edu/*) and T.J. Hudson et al., *Science* **270,** 1945–54.] However, only 3 of 24 chromosomes (16, 19, and 22) are substantially covered by contigs of sequence-ready clones.

The availability of newer, more stable clone resources containing large human DNA inserts (up to 150,000 bases) has stimulated an alternative strategy to contig building for complete genome sequencing. In this new approach, described in *Nature* (**381**, 364–66), about 500 bases of sequence are obtained from both ends of BAC or PAC clone inserts. BAC-PAC end sequencing is performed on clones from a deep (about 20-fold), arrayed library, and all clone names, end sequences, and other useful data are entered into a public database. When a given clone is sequenced, any researcher can search for and identify additional clones with overlapping sequences. These "hits" can then guide the choice of the next overlapping BAC or PAC clone to be sequenced. In this way, contig building (determining overlapping pieces and ordering all the clones) is bypassed; it happens as a consequence of sequencing, not as a prerequisite.

Proponents of this strategy assert that it will be a simpler, faster, and cheaper way to obtain the total genome sequence—the ultimate goal of the Human Genome Project. Also, because clone resources and their end-sequence data can be made available worldwide, this approach will support geographically dispersed participation in genome sequencing as well as easier access to clones for other genome-related purposes.

The general BAC-PAC end-sequencing strategy was discussed at several recent meetings, including a December 1995 DOE-sponsored BAC workshop and the February 1996 Bermuda conference on high-throughput sequencing sponsored largely by the Welcome Trust [see *HGN* **7**(6), 19]. This past summer, DOE reviewed end-sequencing applications and made 2-year awards to the following teams to undertake pilot projects for testing the feasibility of this strategy, its technologies, and its economics:

- Mark Adams [The Institute for Genomic Research (TIGR)], Leroy Hood (University of Washington, Seattle), and Melvin Simon (California Institute of Technology); and
- Glen Evans with Harold Garner [University of Texas Southwest (UTSW) Medical Center], Pieter de Jong (Roswell Park Cancer Center), and Julie Korenberg (Cedars Sinai Medical Center).

[Marvin Stodolsky, DOE (301/903-4475, marvin.stodolsky@oer.doe.gov*)]* ◊

Abstracts of the pilot projects and related research can be accessed through the BAC Workshop WWW page currently being updated (*http://www.ornl.gov/meetings/bacpac/95bac.html*) and participating sites:

- California Institute of Technology (*http://www.tree.caltech.edu/*)
- Cedars Sinai Medical Center (*http://www.csmc.edu/genetics/korenberg/korenberg.html*)
- TIGR (*http://www.tigr.org/*)
- UTSW Medical Center (*http://mcdermott.swmed.edu/gestec/mission/*)
- University of Washington, Seattle (*http://weber.u.washington.edu/~mbt/*)

Archaea *(from p. 7)*

With a genome of just 580 kb, it provides researchers with a model for the minimum number of genes and protein products necessary for independent existence.

MGP Research Providing Clues to Understanding Genetics

Private companies, universities, and DOE laboratories are now sequencing an additional ten microbial genomes for MGP, including that of the most radiation-resistant life form ever found. This organism, *Deinococcus radiodurans*, was first discovered in spoiled beef thought to have been sterilized by radiation. The microbe is potentially useful for cleanup of radioactive wastes as well as for adding to the understanding of sensitive enzymes responsible for monitoring and repairing damage to DNA caused by radiation and other environmental agents.◊

DOE, NCHGR Issue Human Subject Guidelines

On August 19, Aristides Patrinos, Director of the DOE Human Genome Program, and Francis Collins, Director of the NIH National Center for Human Genome Research, issued a document providing investigators with guidance in the use of human subjects for large-scale sequencing projects. The guidance recommends the following principles:

- Derive the initial version of the complete human DNA sequence from multiple donors,
- Ensure that donors can make informed, unpressured decisions about DNA contributions,
- Protect donor privacy and confidentiality,
- Obtain institutional review-board approval before work is initiated, and
- Rapidly introduce new libraries constructed in accordance with this guidance.

The guidance emphasizes numerous ways to preserve anonymity of donors and suggests that they should be selected from diverse pools of individuals, including females as well as males. Recruiting from laboratory staff is discouraged.A copy of the guidance is available from HGMIS (see address, p. 12) or via WWW (*http://www.ornl.gov/hgmis/archive/nchgrdoe.html*).◊

Immune System Genes Reveal Surprises

Scientists analyzing the sequence of the longest (685-kb) continuous segment of human DNA published to date have uncovered powerful information about the human immune system that may help doctors prevent autoimmune diseases such as arthritis and multiple sclerosis.

Analysis of the human beta T-cell receptor (TCR) locus comprising a complex family of genes was reported in the June 21, 1996, issue of *Science* (**272**, 1755–62). The locus contains a cluster of genes that play a vital role in recognizing foreign viruses, bacteria, and cancer cells and in triggering the body's defense mechanisms to destroy these invaders.

Authors Lee Rowen and Leroy Hood (both at University of Washington, Seattle) and Ben F. Koop (University of Victoria, Canada) reported identifying and classifying all related genes at the locus. This information will enable development of tests specific to each individual gene, allowing easier identification of genes involved in autoimmune diseases. "This may ultimately lead to a new form of medicine focusing on preventing, not curing, diseases," said Hood. The research was funded by a grant from DOE with support from NIH.

Unexpected findings include identification of genes encoding trypsinogen, an important enzyme for digesting protein-rich food. The discovery raises the possibility that trypsinogens and immune receptors may work in concert. Researchers also confirmed that a piece of the immune receptor gene has been copied and moved from chromosome 7 to chromosome 9, providing evidence of evolutionary transfer of genes. The *Science* cover depicts chromosomes 7 and 9 with the TCR and trypsinogen genes identified.

The authors attribute their dual achievements—sequencing and analyzing the entire TCR locus—to advances in automated DNA sequencing and development of computational tools for sequence assembly and analysis.

Importance of Computational Tools

According to Rowen, "Computational tools are critical—they are used to store and manage large amounts of data and conduct sequence comparisons. They are also essential for detecting and organizing large-scale patterns within long stretches of sequence."

Sequences can be analyzed in a number of ways, she said. For example, two sequences covering the same chromosomal region in different people can be compared for variations, some of which may correlate with disease susceptibilities. A sequence also can be compared against other sequences in large public databases, for example, gene-coding sequences (cDNAs or ESTs) against genomic sequences, sequences from one species against those in another species, or amino acid sequences against protein motif databases.

The laboratory at the University of Washington is one of six centers chosen recently by NIH to participate in a pilot study of large-scale human sequencing in the Human Genome Project [*HGN* **7**(6), 20]. It is also taking part in the DOE pilot study of BAC end sequencing (see p. 8).◊

The August 2, 1996, issue of *Science* (**273**) features a special section on the booming field of bioinformatics—the use of computers in biology. The Web site (*http://www.sciencemag.org/*) describes the section and links to relevant WWW sites.

☛ Software Finds Genes Across Species

New software called PROCRUSTES, described in the August 20 issue of the *Proceedings of the National Academy of Sciences*, can identify with remarkable accuracy human versions of genes found in other life forms. The product of a collaboration between an American and two Russian researchers, PROCRUSTES is considered far more useful than existing techniques if a related pattern is known.

"With this method, anything alive can serve as a template to find human genes. Mouse, chicken, frog—the species doesn't matter," said coauthor Pavel Pevzner (University of Southern California).

Pevzner and his Russian collaborators, Mikhail S. Gelfand (Russian Academy of Science) and Andrey Mironov (Russian National Center for Biotechnology), devised a spliced-alignment algorithm and software tool that overcomes formidable obstacles. Human genes, which average about 2000 bp, are broken up into smaller segments called exons. The exons can be separated by millions of bases of noncoding DNA that sometimes mimic the exons.

As Pevzner explains, searching for exons is like trying to follow a magazine article that appears on pages 1, 16, 21, 74, and 87, with almost identical advertisements and other articles appearing between. PROCRUSTES helps by constructing a list of all the "pages" that are part of the "story," then automatically combining them into the set that makes the best fit.

The technique works best when a "target protein" from the nonhuman sample is available to guide the search. With such guidance, the method's accuracy often approaches 100%, the authors report. The new tool should prove particularly useful for researchers trying to pinpoint elusive human versions of cancer-causing genes already sequenced in mice and other species.

Articles on PROCRUSTES have appeared in *Business Week*, *Investor's Business Daily*, and *BioWorld Today*. The research was supported by grants from DOE, the Russian Fund for Fundamental Research, the Russian Human Genome Program, and the National Science Foundation's Young Investigator Program. [Contact: Pevzner (213/740-2407, *ppevzner@hto.usc.edu*), PROCRUSTES: *http://www-hto.usc.edu/software/procrustes/*] ◊

New Web Site on Genetics and Public Issues

The National Center for Genome Resources (NCGR) of Santa Fe, New Mexico, publishes a wide array of resources in its Genetics and Public Issues site on the Internet (*http://www.ncgr.org/*). The site includes information on such inherited illnesses as breast cancer and Alzheimer's disease; full-text documents including congressional bills relating to genetics and privacy; and articles about genetics and medical care, confidentiality and privacy, discrimination, support groups, treatments, and gene therapy. NCGR's Genetics and Public Issues program was created in response to the need for information about medical and social implications of genetic research.◊

Commercial Strategies Aim to Spin DNA Threads into Gold

Earlier this year, HGN *7(5) carried an article describing the collaboration between the pharmaceutical giant Merck & Co. and the Genome Sequence Center at Washington University (St. Louis) to develop a publicly available data set of partial cDNA sequences (called expressed sequence tags, or ESTs).*

Because cDNAs represent genome coding areas (i.e., genes), databases containing these sequences offer researchers a way to speed-read through the genome in the hunt for disease genes, bypassing billions of base pairs of noncoding genomic DNA.

In the following guest article, intellectual-property lawyer Rebecca Eisenberg (University of Michigan Law School, rse@umich.edu*), a DOE-funded researcher, analyzes the different cDNA database-usage strategies undertaken by Merck and two major genome-sequencing companies.*

Intellectual-property issues have been unusually conspicuous in the recent history of genomic advances, even by the standards of the patent-weary genetics and molecular biology communities. Controversy has been particularly acute over intellectual-property rights in the results of large-scale cDNA sequencing.

Beginning in 1991 with NIH's filing of patent applications on the first batch of ESTs from Craig Venter's laboratory, each new development has been met with lively speculation about its strategic significance from an intellectual-property perspective. Are cDNA fragments of unknown function patentable, or must they undergo further research or characterization before they satisfy patent-law standards? Will patents on such fragments promote commercial investment in product development or interfere with scientific communication and collaboration and retard the overall research effort?

In the absence of patent rights, how might the owners of private cDNA sequence databases earn a return on their investment while still permitting other investigators to obtain information access on reasonable terms? What are the rights of those who contribute such resources as the cDNA libraries that are used to create the databases and of those who formulate appropriate queries to identify interesting sequences from the morass of information? Will the disclosure of ESTs in the public domain preclude patenting of subsequently characterized full-length genes and gene products? And why would a commercial firm invest its own resources in generating an EST database for the public domain?

Two factors have contributed to the fascination with intellectual-property issues in this setting. First is a perception that some pioneers in large-scale cDNA sequencing have sought to claim intellectual-property rights that reach far beyond their own actual achievements to cover the future discoveries of others. For example, the controversial NIH patent applications claimed not only the ESTs for specified sequences but also the corresponding full-length cDNAs and smaller portions that might not even include the disclosed ESTs. More recently, private owners of cDNA sequence databases have conditioned data access on advance agreements offering either a license or right of first refusal to any resulting intellectual property. These actions raise questions about the fairness and efficiency of the system to protect intellectual property. Such concerns are particularly compelling to research scientists, who have more than commercial interests at stake.

Second is the surprising alignment of interests in the data. NIH, a public institution, initially took an aggressive position in favor of patenting discoveries that some industry representatives thought were unpatentable and should remain unpatented. Merck & Co. ultimately took on the quasigovernmental function of sponsoring a university-based effort to place comparable information in the public domain. These topsy-turvy positions raise intriguing questions about the proper roles of government and industry in genomics research and about who stands to benefit—and to lose—from the private appropriation of genomic information.

Promoting R&D Through Exclusive Rights

Research scientists who work in public institutions often are troubled by the concept of intellectual property because their norms tell them that science will advance more rapidly if researchers enjoy free access to knowledge. By contrast, the law of intellectual property rests on an assumption that, without exclusive rights, no one will be willing to invest in research and development (R&D).

Patenting provides a strategy for protecting inventions without secrecy. A patent grants the right to exclude others from making, using, and selling the invention for a limited term, 20 years from application filing date in most of the world. To get a patent, an inventor must disclose the invention fully so as to enable others to make and use it. Within the realm of industrial research, the patent system promotes more disclosure than would occur if secrecy were the only means of excluding competitors. This is less clear in the case of public-sector research, which typically is published with or without patent protection.

The argument for patenting public-sector inventions is a variation on the standard justification for patents in commercial settings. The argument is that postinvention development costs typically far exceed preinvention research outlays, and firms are unwilling to make this substantial investment without protection from competition. Patents thus facilitate transfer of technology to the private sector by providing exclusive rights to preserve the profit incentives of innovating firms.

Nonpatent Strategies for Commercial Exploitation

No patents have been issued so far on cDNA fragments of unknown function, although a number of private firms have pending patent applications that claim thousands of such fragments. Meanwhile, three firms—Human Genome Sciences (HGS), Incyte Pharmaceuticals, and Merck—are pursuing different nonpatent strategies for exploiting the value of these sequences as unpatented information. These strategies are exclusive licensing, nonexclusive licensing, and dedication to the

public domain, and it is still too early to tell how each will pay off. We can see, however, how different firms are placing their bets, and we also have some idea of the sizes of those bets.

HGS and Incyte are exploiting their databases commercially by controlling access to them, in effect using contracts and trade secrecy to protect their intellectual property. The viability of these strategies may be limited by Merck's sponsorship of a competing cDNA sequencing effort at Washington University dedicated to the public domain. The commercial value of the private databases is likely to decline as public-domain information increases. Although public-domain databases are growing rapidly, the private ones remain larger at this point and claim to offer superior products. These products include longer sequences of contiguous cDNA fragments; more complete sequence annotations, including information about expression in different types of tissues; high-powered bioinformatics capabilities; and user-friendly software.

A significant limitation on the value of public-domain databases is the pending patent applications of private database owners. If these applications ripen into issued patents, they could preempt the use of any covered sequences, even if those sequences were disclosed publicly before the patent was issued, as long as the patent applicants are able to establish their priority.

U.S. patent applications are confidential until a patent is issued, so determining which sequences are the subject of patent applications is impossible. Those who use sequences from public databases today risk facing a future injunction if those sequences turn out to be patented by HGS or Incyte on the basis of previously filed patent applications. The same uncertainty applies to sequences obtained from private databases; for example, sequences that are obtained from the Incyte database may turn out to be covered by a previously filed HGS patent. Because the Merck initiative got off to a late start, its sequences are more likely to be covered by other firms' prior patent applications.

Exclusive Licensing. For $125 million over a 3-year period plus royalties on product sales, HGS has licensed exclusive rights to access its database to SmithKline Beecham (SB). SB also gained the right of first refusal to develop and market protein therapeutic and diagnostic products from information in the database. HGS has entered into separate collaborative agreements with other research partners for gene-therapy and other DNA-based product development.

During the period of SB's exclusive license, investigators in academic and nonprofit institutions may obtain access to some of the same sequence information through a separate database maintained by The Institute for Genomic Research (TIGR) under the terms of a Database Access Agreement. The TIGR database includes sequences that are similar to previously published sequences and accessible to nonprofit investigators with minimal restrictions on use. It also includes proprietary sequences that are accessible only to those who sign more restrictive agreements giving HGS rights to prepublication review and an option to negotiate a license to any resulting inventions. Some academic investigators also have obtained access to sequences in the separate HGS proprietary database by signing a Materials Transfer Agreement granting HGS "a sole and exclusive worldwide right and license" to develop any resulting products on terms to be negotiated in the future.

An obvious advantage of this exclusive licensing strategy for HGS is that it has generated a lot of revenue; SB placed what appeared to be a very large bet 3 years ago. An obvious concern is that restricting database access to such a degree may limit the value that can be extracted during the term of the license. Perhaps this concern motivated SB and HGS to enter into collaborative agreements announced this past summer to share the database with four additional pharmaceutical firms [Takeda Chemical Industries, Merck KGaA (not related to Merck & Co.), Schering Plough, and Synthelabo SA]. With the signing of these agreements, SB appears to have made its money back even before bringing any new products to market—the agreements call for payments totaling $140 million plus royalties on product sales.

Nonexclusive Licensing. Incyte has offered nonexclusive licenses to as many firms as will take them, at a much cheaper price than SB paid for its exclusive deal with HGS. So far ten pharmaceutical firms have signed on as subscribers, including Pfizer, Pharmacia & Upjohn, Novo Nordisk, Hoechst, Abbott Laboratories, Johnson & Johnson, BASF AG, Hoffmann-La Roche, Zeneca, and Schering AG Berlin. Financial terms for most of these agreements have not been disclosed, but press accounts report that they

➠

Industry Moves DNA Patenting Forward

Huge Backlog Challenges U.S. Patent Office

Quick public access to sequence data remains a hallmark of the Human Genome Project for many genome researchers in the United States and worldwide [see *HGN* 7(6), 19 and 20]. At the same time, private companies are filing applications to patent DNA sequences at unprecedented rates.

These new patent applications are challenging the capacity of the U.S. Patent and Trademark Office (PTO) to review them. Companies specializing in DNA sequencing have applied for patents on hundreds of thousands of sequences, including genes and gene fragments. PTO examines all sequence applications for fulfillment of four major patenting criteria: novelty, nonobviousness, usefulness, and enablement (i.e., detailed enough to enable one skilled in the field to use it for the stated purpose).

Earlier this year PTO held public hearings to gather ideas on streamlining the time-consuming and expensive examination process (*http://www.uspto.gov/web/uspto/info/seq-hear.txt*). Some possible changes include requiring more background research by applicants, setting new limits on applications, and prescreening sequences for usefulness before examining them further. [For more information see *Science* **272**, 643 (May 3, 1996).] ◊

total more than $160 million, excluding contingent payments such as milestones and product royalties.

Although each Incyte subscriber has placed a smaller bet than SB did, in the aggregate they may well provide more funds for the development of Incyte's genomic databases. From a broader social standpoint, of course, the more interesting question is not the size of the bets but the ultimate payoffs. Which approach will yield more discoveries or more commercial products?

Public Domain. The Merck strategy of putting sequence information into the public domain is the newest approach and, at first glance, the most puzzling. How does this strategy advance Merck's own interests? By placing data in the public domain, Merck can generate the sequence information more cheaply—indeed, almost unbelievably cheaply. Merck is placing a very small bet, somewhere under $10 million, but by positioning itself as a public benefactor, the company is able to take advantage of existing infrastructure at Washington University, put in place with public funds, for its sequencing efforts.

Apart from generating sequence information more cheaply, Merck expects to promote research and derive more benefit by distributing the data widely. As Merck sees it, sequence information will not yield products for commercial development until further fundamental research is done to understand functions and biological pathways associated with the partially sequenced genes. Merck's interest is in developing specific drugs at a later stage in the R&D process. Nothing obligates researchers to bring any potential products to Merck for commercial development, but Merck is confident that it can capture an adequate share of resulting products to justify the company's modest investment in generating the database.

Some observers have suggested the more cynical possibility that Merck may seek to undermine the value of its commercial competitors' investments in existing sequence databases. HGS and Incyte will be dependent on patents to protect their proprietary positions in the long run, and Merck may be betting that the two companies will not obtain much in the way of patent rights.

Preliminary indications suggest that the public data is generating considerable interest, with EST-database accessions showing a dramatic increase. A big part of the increase has come in daily anonymous FTP downloads of the entire database, a form of query likely to be popular with commercial users who do not want to leave an electronic record of what they are looking for.

The most obvious benefit of disseminating information in the public domain is that free availability encourages widespread use of information, minimizes transaction costs, and makes R&D cheaper and faster. Of particular relevance to research science, a vigorous public domain can supply a meeting place for people, information, and ideas that might not find each other in the course of more organized, licensed encounters.

Finally, information in the public domain is accessible to users who otherwise would be priced out of the market. In emphasizing intellectual-property rights in the past, we may have underestimated the value of a rich public domain to private as well as public sectors. We may need now to reconsider the limits of private appropriation of new information as a means of promoting commercial development.◊

A similar article by Eisenberg was published in Elsevier's Trends in Biotechnology *[Vol. 14, 302–7 (August 1996)].*

Human Genome news

This newsletter is intended to facilitate communication, help prevent duplication of research effort, and inform persons interested in genome research. Suggestions are invited.

Human Genome Management Information System
Oak Ridge National Laboratory
1060 Commerce Park, MS 6480
Oak Ridge,TN 37830
423/576-6669, Fax: /574-9888,
http://www.ornl.gov/hgmis/

Managing Editor
Betty K. Mansfield
bkq@ornl.gov

Editors/Writers/ Designers
Anne E. Adamson
Denise K. Casey
Sheryl A. Martin
Judy M. Wyrick

Production Assistants
Murray Browne
Marissa D. Mills
Laura N. Yust

U. S. Department of Energy Office of Health and Environmental Research
Ari Patrinos, Associate Director

Health Effects and Life Sciences Research Division, OHER
Marvin E. Frazier, Acting Director

http://www.er.doe.gov/production/her/hug_top.html

Contact: Daniel W. Drell
301/903-6488, Fax: -8521
Daniel.Drell@oer.doe.gov or *genome@er.doe.gov*

DNA Quiz on WWW

As part of a basic course in molecular biology at Göteborg University in Sweden, a self-paced quiz is available on the Web (*http://www.medkem.gu.se/ln/molbio/gene/*). Questions are presented one at a time; only by giving the right answer can the user proceed to the next question. Most questions are related to DNA and protein sequence information, including building blocks of nucleic acids and proteins, nucleic acid structure, nucleotide and amino acid sequences, restriction enzymes, transcription, and the genetic code and translation.◊

¶ *Genomes, Molecular Biology, and Drug Discovery*

Genomes, Molecular Biology, and Drug Discovery is based on the Seventh SmithKline Beecham International Symposium held in Cambridge, England, in March 1995. The symposium was devoted to consideration of what the various genome projects can deliver and how that information can be used to generate new therapeutics for superior and better-targeted health care. Edited by Michael J. Browne and Peter Thurlby of SmithKline Beecham, the book includes 11 technical papers on speeding up the discovery of human genes, particularly disease genes, and the identification of new targets for drug development. Such subjects as ESTs, protein structure prediction, monoclonal antibodies, and retroviral vectors are explored, and an appendix of poster abstracts is included. 196 pp., 1996. [Academic Press; 1260 Sixth Ave.; San Diego, CA 92101 (619/699-6742)].◊

¶ Genome Journal

Genome Science & Technology is a quarterly peer-reviewed journal, edited by Craig Venter (The Institute for Genomic Research). The journal is devoted to comprehensive coverage of genome and cDNA sequencing; mapping; informatics; biological interpretations; human genetic disease; ethical, legal, and social issues; new technologies; and computational approaches. [Mary Ann Liebert, Inc. (inside New York State: 914/834-3100; outside New York State: 800/654-3237), *liebert@pipeline.com*] ◊

Mutation Database Initiative Under Way

Over the last 5 years, the Human Genome Project has had an enormous impact on the scientific community, which is witnessing an explosive increase in the description of genes and disease-causing mutations. As a result of this overwhelming expansion of data, many problems have arisen in describing and cataloguing sequence alterations and making them accessible to researchers. Journals cannot publish all the mutation reports, and central databases such as Online *Mendelian Inheritance in Man* (OMIM) have found it necessary to limit the data they attempt to capture. Mutation information currently is fragmented and incomplete, and, for most genes, no database exists at all.

A few locus-specific databases (such as those for Factor IX, cystic fibrosis, PAH, and *BRCA1*) have developed from the need of individuals or consortiums to keep track of data for their research. However, the usefulness of these smaller databases is limited by their lack of uniformity in design, content, and nomenclature—making access, communication, and analysis difficult.

This scarcity of up-to-date gene-mutation listings hampers researchers and clinicians in determining whether a particular mutation has been described or not. Also, investigators lack comprehensive information on the loss of function of different mutations in specific genes, and clinicians are unable to draw on experiences of others who have patients with similar mutations.

Mutation Lists Vital

Complete and current mutation lists are vital for studying genotype-phenotype relationships, treating patients with similar phenotypes, and analyzing amino-acid residues important in the function of gene products. Mutation lists also are needed to verify the existence of modifier genes and assist in their identification, monitor mutagenic environmental influences in somatic oncogenes and tumor-suppressor genes, design diagnostic protocols, assist in manuscript review, and examine the spectrum and type of gene mutations.

Although rare, disease-associated mutations are powerful probes of genetic variations in populations. Knowledge of these mutations allows researchers to provide information and services to those affected and to characterize mutation effects in populations.

Because researchers believe mutations in the human genome are likely to number in the millions, a systematic approach to collecting and maintaining mutation data is needed. Some progress toward that goal has been achieved: Some researchers have made their databases available via the Internet (e.g., phenylketonuria, cystic fibrosis, mutations in factor IX, and P*53*), and central databases such as OMIM have compiled partial listings of mutations identified in specific genes. The Human Gene Mutation Database of David Cooper and Michael Krawczak (Institute of Medical Genetics, Cardiff, Wales) contains mutations and an index of where they are published (*http://www.cf.ac.uk/uwcm/mg/hgmd0.html*).

Database Association

To make comprehensive mutation lists available for research, investigators are now forming an association of curator-driven, locus-specific databases whose standardized content can be downloaded onto central databases.

New Genome Center in Australia

The newly established Australian Genome Research Facility was awarded a $10-million grant in late 1995 from the Australian Major National Research Facilities Program. The center will concentrate on sequencing (John Mattick), mapping and linkage (Simon Foote), and mutation detection (Richard Cotton, author of the article on this page). Funding is for hardware only, and operating costs are expected to be generated from supporting organizations and other subscribers. Although several Australian groups have played major roles in genome projects worldwide, this program is the first effort to capitalize on opportunities offered by the discovery of genome sequences. More information can be found on its Web site (*http://www.cmcb.uq.edu.au/agrf/*).◊

Mutation Initiative Contacts

To help create a mutation database or participate in a working group, contact any of the three collaborators:

- Charles Scriver 514/934-4417, Fax: -4329, *mc77@musica.mcgill.ca*
- Victor McKusick 410/955-6641, Fax: -4999, *mckusick@gdb.org*
- Richard Cotton +61-3/9288-2980, Fax: -2989, *cotton@ariel.its.unimelb.edu.au*

Mutation Database Web Site

- *http://ariel.ucs.unimelb.edu.au/~cotton/mut_database.htm*

Mutation Newsgroup

Communication can be established by subscribing to the mutation newsgroup. Send a message to *biosci-server@net.bio.net* with *subscribe mutation* in the message body.

Related Web Site

- Mutation Research Genomics Initiative: *http://www.ornl.gov/molgen/*

Many believe these databases can serve the community better because they usually are set up and maintained by researchers, contain numerous unpublished mutations, have a greater likelihood of being complete, and avoid the limitations of central databases. Their bulletin-board components contain useful information on primers, methods, and scanning laboratories. A directory of locus-specific databases will be added to the Web site. Another directory comprises listings in OMIM.

Project History

At an October 1994 meeting in Montreal, called to consider nomenclature, further discussion led to an agreement to work toward placing up-to-date mutation lists on the Internet. Attendees felt that the curator-driven system should be simple and contain both published and unpublished mutations. This approach was supported by Victor McKusick (OMIM), Charles Scriver (McGill University, Montreal), Haig Kazazian (University of Pennsylvania), Lap-Chee Tsui (Hospital for Sick Children, Toronto), Aravinda Chakravarti (Case Western Reserve University), and Douglas Wallace (Emory University) as well as the American Society of Human Genetics (ASHG), Human

(see Database, p. 14)

ELSI Working Group Under Review

An 11-member committee was appointed in June to review the function and structure of the NIH-DOE Joint Working Group on the Ethical, Legal, and Social Implications (ELSI) of Human Genome Research. As planning begins for the next 5 years, NIH and DOE Human Genome Program staff consider this an appropriate time to determine how best to provide for objective advice on ELSI issues. The committee expects to finish its review and issue a report by January 1997.

The ELSI working group was one of several formed in 1989 to guide the two agencies on the Human Genome Project's research agenda. All the other working groups have finished their tasks, even though the goals have not been completed, and guidance in those areas is now being provided through ad hoc committees.

The ELSI working group initially helped to define the parameters of the research grants program, sponsored regional town meetings to educate the public about the Human Genome Project and ELSI issues arising from research, oversaw the Task Force on Genetic Information and Insurance, and is playing a similar role with the Task Force on Genetic Testing. The working group has also commented on such issues as informed consent, privacy, discrimination, and genetic testing for cystic fibrosis and breast cancer.◊

Database *(from p. 13)*

Genome Organisation (HUGO), and March of Dimes (MOD).

A formal meeting in Minneapolis in October 1995, supported by MOD and organized by HUGO, included systems experts and representatives from prominent mutation databases, genetic and cancer societies and journals, publishers, and central databases. Attendees endorsed the idea of an alliance of locus-specific database curators who would keep their own databases current and assist in implementing a unified and uniform approach. Working groups were established on Software and Content (Chair, Scriver), Central Database [Chairs, McKusick and Jim Ostell (NCBI)], Nomenclature [Chair, Stylianos Antonarakis (University of Geneva)], and Mutation Database Association (Chair, Richard Cotton).

At a satellite meeting of HGM '96, Michael Ashburner (University of Cambridge) outlined a proposal for a new central mutation database that would supplement individual mutation databases. The central database would contain core information rather than locus-specific data and allow faster analysis of stored data. In preparation for a half-day meeting to be held October 29 in conjunction with the ASHG meeting in San Francisco, HGM '96 attendees agreed to place the working groups' discussion documents on the Web site and to initiate a mutation electronic mailing list. The mutation-database project is now a HUGO initiative and collaboration among Scriver, McKusick, and Cotton (Chair). *[Richard G.H. Cotton, Mutation Research Centre, St. Vincent's Hospital, Victoria, Australia]* ◊

GSDB Conversion Complete

Genome Sequence Data Base recently announced that the GSDB conversion is complete. GSDB 1.0 includes all public DNA sequence and feature data, which are accessible via their individual GSDB accession numbers or their international collaboration accession numbers. GSDB 1.0 also supports a variety of such new data types as discontiguous sequences, alignments, sequence confidence values, and analysis results.

Bulk submissions can be made with the new GSDB Input-Output (GIO) format (specifications: *http://www.ncgr.org/gsdb/GIO-doc.html*). Individual sequence submissions or feature updates can be made through the GSDB Annotator (see below). Basic sequence and feature retrieval and advanced query methods are available from the Web site. GSDB staff are available to help with submissions or Web queries.

The new GSDB Annotator (beta version) is a graphical client server application for Macintosh or Sun that allows users direct account-controlled read-and-write access to the database. [Information: *http://www.ncgr.org/* or 505/982-7840] ◊

¶ *Mutation Research* Journal Adds New Section

In 1997 *Mutation Research* will devote a new section to the union between genomics and mutation research. Managing Editor Anthony Carrano (Lawrence Livermore National Laboratory) says *Mutation Research Genomics* will publish papers on the nature and consequences of genome variations in humans and appropriate model systems. Its focus will be on experimental approaches, instrumentation, and informatics technologies useful for measuring and characterizing genetic variation. The new section will be produced in traditional hard copy and online, initially in one volume of four issues. The September 1996 issue of *Forum*, also a part of *Mutation Research*, contains an editorial by John Wassom (Oak Ridge National Laboratory) giving background information on development of the field of mutation genomics. [Mutation Research Genomics Initiative: *http://www.ornl.gov/molgen/hmepg.html*] ◊

The Gene Letter Now Online

Volume 1, Issue 1 of *The Gene Letter* (*http://www.geneletter.org/*) went online in July. Developed and published by the Shriver Center with a 2-year grant from DOE, the bimonthly electronic newsletter is designed to inform consumers and professionals about advances in genetics and to encourage discourse about emerging policy dilemmas. Regular columns are Science, Medicine, Ethics, Law, International Developments, Student Corner, and Resources. Editors are Philip Reilly and Dorothy Wertz (Shriver Center) and Robin Blatt (Massachusetts Department of Public Health). *The Gene Letter* also operates an uncensored chatroom (*http://www.geneletter.org/genetalk.html*).◊

¶ Video on Genetic Testing

Promise and Perils of Biotechnology: Genetic Testing is the third videotape in the *Winding Your Way Through DNA* educational series. An outgrowth of the 1992 symposium of the same name, the series was developed in response to teachers' interest in videotapes and curriculum materials based on the symposium's topics. The 25-min. classroom video and teacher's guide educate students about inherited disorders, their prevalence in society, and the benefits and drawbacks of genetic testing. Through the narration of a genetic counselor, the documentary follows three people: one young woman with a family history of Huntington's disease who decides to be tested and a mother and daughter who change their lifestyles to deal with familial hypercholesterolemia. For high school, college, and public education classes in genetics, biotechnology, and bioethics. [Contacts for scientists and individuals: Cold Spring Harbor Laboratory Press (800/843-4388 or 516/349-1930, *cshpress@cshl.org*); teachers and educational institutions: Pyramid Media (800/421-2304 or 310/828-7577, *rwright@pyramedia.com*)] ◊

Carrano a HUGO Vice-President

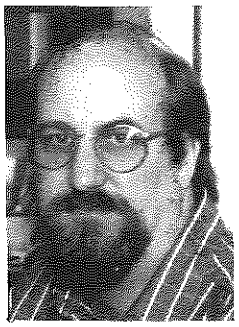
Anthony Carrano

Anthony Carrano, Director of the DOE Human Genome Center at Lawrence Livermore National Laboratory (LLNL), became Vice President for the Americas of the Human Genome Organisation (HUGO) in March 1996. Grant Sutherland (Australia) is President of HUGO, which represents nearly 1000 members from 50 countries.

The three HUGO regional vice-presidents, who are also members of the 18-person International Council, serve as liaisons between their regions and the council. The Americas office representing North, South, and Central America is located in Bethesda, Maryland, and headed by Susan Wallace. The other two main offices are HUGO Europe (London), whose vice-president is Gert Jan Van Ommen, and HUGO Pacific (Osaka, Japan), for which Yoshiyuki Sakaki is the new vice-president.

Carrano is also Associate Director for Biology, Biotechnology, and Healthcare Research at LLNL. His lifelong interest is mutation research, specifically in the area of DNA repair genes; chromosome 19, which was mapped by the LLNL genome center (*http://www-bio.llnl.gov/bbrp/genome/genome.html*), is particularly rich in such genes. Among the accomplishments of Carrano's research team, which is involved in numerous international collaborations, are discovery of the myotonic dystrophy gene and development of DNA research substrates for distribution to other laboratories. The group is now concentrating on sequencing chromosome 19.◊

Palazzolo Named Genome Center Director



Michael Palazzolo

Michael Palazzolo has been appointed Director of the Human Genome Center at Lawrence Berkeley National Laboratory (LBNL). In making the announcement, LBNL Director Charles Shank said Palazzolo "has developed and adapted new technologies to directly sequence DNA on a massive scale. His innovations have been key to placing the Berkeley Genome Center at the cutting edge of genomic sciences worldwide." Palazzolo succeeds Acting Director Mohandas Narla, who assumes leadership of a new department in the Life Sciences Division that will combine studies of membrane proteins with innovative microscopies.

Palazzolo received his B.A. in chemistry and M.D.-Ph.D. in Medicine and Physiology from Columbia University. From 1985 to 1990 he did postdoctoral research at California Institute of Technology and in 1990–91 was a research assistant professor with the Department of Genetics at Washington University School of Medicine, St. Louis.

In a cooperative effort with Lawrence Livermore National Laboratory and Los Alamos National Laboratory to sequence nearly one-third of the total 3 billion base pairs of human genes, the LBNL center has completed more than 5 million base pairs at a current rate of 500,000 a month. The center is also engaged in a major collaboration with the University of California, Berkeley, and Carnegie Institution to map, sequence, and characterize the genome of the fruit fly *Drosophila melanogaster*.

Palazzolo said he looks forward to leading the center (*http://www-hgc.lbl.gov/GenomeHome.html*) through its next phases, including final sequencing of the human genome.◊

Dovichi Wins ACS Award

Norman J. Dovichi (University of Alberta, Canada) recently received the American Chemical Society Award in Chemical Instrumentation, sponsored by the Dow Chemical Company Foundation. Dovichi was recognized for his research in ultrasensitive instrumentation for analytical chemistry, combining laser spectroscopy with high-performance chemical separation techniques. His recent research into DNA sequencing by capillary electrophoresis was cited as especially noteworthy. Dovichi's research is supported by the DOE Human Genome Program. [Contact: Roland Hirsch (301/903-3213)] ◊

Hollaender Fellows Named

DOE has announced the award of five 1996 Alexander Hollaender Distinguished Postdoctoral Fellowships for up to 2 years of research at DOE laboratories having substantial programs supportive of the Office of Health and Environmental Research's mission. The mission is to understand health and environmental effects associated with energy technologies and to develop and sustain research programs in life, biomedical, and environmental sciences.

Fellowship winners were chosen from a field of applicants who received their doctoral degrees after April 30, 1994. Listed below are each fellow's name, university of doctoral degree, host laboratory and research mentor, and research topic.

- **Cymbeline Culiat** (University of Tennessee, Knoxville): Oak Ridge National Laboratory; Lisa Stubbs. Cloning of a mouse gene causing severe deafness and balance defects.
- **Bruce Hungate** (University of California, Berkeley): Smithsonian Environmental Research Center, Edgewater, Maryland; Bert Drake. Effects of elevated CO_2 on soil hydrology: Links to N cycling and long-term ecosystem responses to rising CO_2.
- **Michael Mann** (Yale University): University of Massachusetts, Amherst; Raymond Bradley. Investigation of patterns of organized large-scale climatic variability during the last millennium.
- **Steven Ripp** (Oklahoma State University): Center for Environmental Biotechnology, University of Tennessee, Knoxville; Gary Sayler. Potential for transduction and pseudolysogeny in a soil ecosystem.
- **Tau-Mu Yi** (Massachusetts Institute of Technology): Laboratory of Structural Biology and Molecular Medicine, Los Angeles; James Bowie. Structure-function analysis of alpha-factor receptor.

Application deadline for the next round of Hollaender Fellowships is January 15, 1997. For information on this and other research funding, see box, p. 19.◊

Genome FAQs

Relatively Speaking

The many visitors (some 7000 a month) to the Human Genome Project Information Web site ask us interesting questions about genetics and the Human Genome Project. Some we answer directly, and others we refer to leading researchers in relevant fields. From time to time, *HGN* will print answers to selected frequently asked questions (FAQs). These and other FAQs will also be posted to the Web site (*http://www.ornl.gov/hgmis/*).

The following reply, supplied by researcher Lisa Stubbs [Oak Ridge National Laboratory (ORNL)], was prompted by a recent query sent by a New Zealand ethicist. Stubbs conducts research on mouse-human comparative genome mapping for the DOE Human Genome Program.

FAQ: How closely related are mice and humans? What percentage of genes are the same?

Mice and humans (indeed, most or all mammals including dogs, cats, rabbits, monkeys, and apes) have roughly the same number of nucleotides in their genomes—about 3 billion base pairs. This comparable DNA content implies that all mammals contain more or less the same number of genes, and indeed our work at ORNL and the work of many others have provided evidence to confirm that notion.

I know of only a few cases in which no mouse counterpart can be found to correspond to a particular human gene, but otherwise we see essentially a one-to-one correspondence between genes in the two species. The exceptions generally appear to be of a particular type—genes that arise when an existing sequence is duplicated and changed enough to perform a new function. These make up a small percentage of the total genes, in my opinion. We won't know for certain until both genomes are sequenced, but I believe the number won't be significantly larger than 1 to 5%.

The differences between mice and humans are not in the *number* of genes we each carry but in the *structure* of genes and the *activities* of their protein products. Gene for gene, we are very similar to mice. What really matters is that around 100,000 very subtle changes add together to make quite different organisms. Further, genes and proteins interact in complex ways that multiply the functions of each. In addition, a gene can produce more than one protein product through alternative splicing or post-translational modification. A gene can produce more or less protein in different cells at various times in response to developmental or environmental cues, and many proteins can express disparate functions in various biological contexts. Thus, subtle distinctions are multiplied by the more than 100,000 estimated genes.

Lisa Stubbs, a researcher at Oak Ridge National Laboratory, displays one of the many furry residents of ORNL's "Mouse House," the world's largest collection of mutant mouse strains.

The often-quoted statement that we share over 90% of our genes with apes actually should be put another way. That is, we share virtually all our genes with apes. However, on average, a single related set of ape and human genes would differ in DNA sequence by about 10%. For mouse, it is more like 20 to 30%, with a lot of variation from gene to gene in those differences (e.g., some mouse and human gene products are almost identical). Some of those 10 to 30% nucleotide changes would be "neutral" and would not result in production of a significantly altered protein. Others, but probably only a relatively small percentage, would introduce changes that could substantially alter what the protein does.

Put these alterations in the context of known human inherited diseases: If a certain nucleotide is changed in a particular gene, for example, a person can develop sickle cell disease, cystic fibrosis, or breast cancer. A single nucleotide difference can alter protein function in such a way that it causes a terrible tissue malfunction. However, many other single-nucleotide changes in the same gene would do nothing harmful at all. Evolutionary changes are the same way—some are neutral, some subtle, and some dramatic. Add them all together, and they can make quite an impact and account for the huge differences we see among organisms.◊

☛ Tool for Analyzing Multigene Families

FINEX, developed by Stephan Beck and colleagues at the Imperial Cancer Research Fund in London, is a novel tool for identifying and analyzing multigene families, even in the absence of significant sequence similarity. FINEX compares strings of exons delimited by intron-exon boundary phases against a database of fingerprints (*J. Mol. Biol.* **249**, 342–59, 1995). FINEX access: e-mail (*finex@biu.icnet.uk* with *help* in the message body) and WWW (*http://www.biu.icnet.uk/projects/finex/index.html*).◊

¶ Genetic Principles

Genetics and You, by medical geneticist John F. Jackson (University of Mississippi Medical Center), explains in layman's terms the genetic principles that underlie genetic disorders. From these principles, the author discusses genetic counseling, family pedigrees, prenatal diagnosis, the role of environment in birth defects, early detection and prevention, and reproductive options. 104 pp., 1996. Hardcover and softcover. [Humana Press, Totowa, New Jersey (201/256-1699)] ◊

Fasman and Letovsky to Assume New GDB Roles

Effective September 1, Ken Fasman became Director of Research and Development for the Genome Database. He will be based at the Whitehead Institute for Biomedical Research–MIT Center for Genome Research in Cambridge, Massachusetts. While at the center, Fasman will focus on long-range design issues, particularly the integration of mapping and sequencing data. This move presents an excellent opportunity for GDB to benefit from an active collaboration with a major center for human genome mapping and sequencing.

Stanley Letovsky, formerly Deputy Director of GDB Informatics, takes over as Director. Letovsky has done pioneering research on software to integrate databases into the World Wide Web as well as on methods for combining different sources of genome mapping data. As Informatics Director, he will focus initially on extending GDB with gene-function information and on supporting enhanced queries that combine positional and functional constraints.◊

GDB Links Human Genes to *Drosophila* Homologs

Over 600 gene entries in GDB are now linked to homologous *Drosophila* genes in the FlyBase database, which contains genetic and other information from all current and past scientific studies on the fruit fly. All *Drosophila* species are represented in FlyBase, so the Homology links from any given human gene may contain several FlyBase links. Gene symbols of the class Nnnn* (e.g., Dgua\Hsp22) are from species other than *Drosophila melanogaster*.

GDB thanks those at FlyBase who made it possible to provide these links to the community, especially Michael Ashburner, Don Gilbert, and Wayne Rindone.◊

Map Manager

Map Manager (*http://mcbio.med.buffalo.edu/mapmgr.html*) is a program for the Macintosh personal computer that helps analyze the result of genetic-mapping experiments using backcrosses, intercrosses, or recombinant inbred strains. The specialized program allows easy storage, retrieval, and display of information from such mapping experiments and also has tools for statistical analysis of experimental results. Two new versions of Map Manager are under development:

- Map Manager QT (*http://mcbio.med.buffalo.edu/mmQT.html*), with functions for analysis of quantitative traits, is now at Version QTb8.
- Map Manager XP (*http://mcbio.med.buffalo.edu/mmXP.html*), for Windows and Macintosh OS, allows analysis of intercrosses with dominant markers and crosses with mixed segregation patterns. Map Manager XP is expected to be available for user testing by January 1997. When it is ready for general release, all current Map Manager users will receive it by mail.

Map Manager was created and is maintained by Kenneth Manly, Robert Cudmore, Jr., and Greg Kohler at Roswell Park Cancer Institute. It is supported by a grant from the Rockefeller Foundation and a subcontract from the Jackson Laboratory Mouse Genome Informatics Project funded by NIH.◊

GDB Access Via WWW

The GDB Web server is available directly at the following URLs:

- **United States** *http://gdbwww.gdb.org/*
- **Australia** *http://morgan.angis.su.oz.au/gdb/gdbtop.html*
- **France** *http://gdb.infobiogen.fr/*
- **Germany** *http://gdbwww.dkfz-heidelberg.de/*
- **Israel** *http://gdb.weizmann.ac.il/*
- **Japan** *http://www2.gdb.gdbnet.ad.jp/gdb/gdbtop.html*
- **Netherlands** *http://www-gdb.caos.kun.nl/gdb/gdbtop.html*
- **Sweden** *http://gdb.embnet.se:gdb/*
- **United Kingdom** *http://www.hgmp.mrc.ac.uk/gdb/gdbtop.html*

GDB User Support Offices

UNITED STATES
Baltimore, Maryland
help@gdb.org

AUSTRALIA
Sydney
bucholtz@angis.su.oz.au

FRANCE
Villejuif
gdb@infobiogen.fr

GERMANY
Heidelberg
gdb@dkfz-heidelberg.de

ISRAEL
Rehovot
lsprilus@weizmann.weizmann.ac.il

JAPAN
Tokyo
mika@gdb.gdbnet.ad.jp

NETHERLANDS
Nijmegen
post@caos.caos.kun.nl

SWEDEN
Uppsala
help@gdb.embnet.se

UNITED KINGDOM
Cambridge
admin-gdb@hgmp.mrc.ac.uk

Searching for Genes in GDB

One way to search for genes in GDB is to access the U.S. GDB home page and choose "Search by Gene Name or Symbol." Searches can be done using an entire or partial gene symbol or name.

Results are presented in tables from which additional searches can be performed. Initial results provide cytogenetic location and other names (aliases) by which the gene is known. Additional information for each gene includes nearby genes and markers, contacts for reagents (PCR primers, clones, ASOs) to assay for the gene, citations, polymorphisms, and mutations.

Links are also provided to external databases for information on homologous genes in other species, DNA sequences, human disease phenotypes, and enzyme function, if applicable. [Questions and suggestions: *findgene@gdb.org*] ◊

GDB Has the Numbers

Want to get a quick count of the genes in GDB? Or clones, amplimers, citations? Are you curious about how many genes have been localized to a particular chromosome? Help is on the way.

The Genome Database recently added a Reports and Statistics section to its Web site. There you'll find weekly counts of a number of classes in the database. Currently only a few categories are listed, but GDB plans to expand these reports to include other classes such as maps and polymorphisms. The Reports and Statistics section is available via the U.S. GDB home page or directly (*http://gdbwww.gdb.org/gdb/report.html*). GDB welcomes your suggestions (*help@gdb.org*).◊

This newsletter is prepared at the request of the DOE Office of Health and Environmental Research by the Biomedical and Environmental Information Analysis Section of the Health Sciences Research Division at Oak Ridge National Laboratory, which is managed by Lockheed Martin Energy Research Corp. for the U.S. Department of Energy, under Contract DE-AC05-96OR22464.◊

Calendar of Genome and Biotechnology Meetings*

November 1996

4–5. Gene Localization; San Diego [CHI, 617/630-1300, Fax: -1325, *chi@healthtech.com, http://www.healthtech.com/conferences/*]

4–6. Intl. IEEE Symp. on Intelligence and Systems; Washington, DC [N.G. Bourbakis, 607/777-2165, Fax: -4464, *bourbaki@bingsuns.cc.binghamton.edu*]

7–8. Natl. Conf. on Preparing Schools for the Genetic Revolution; Lincoln, NE [G. Wright, 402/472-8881, Fax: -8412, *gwright@unl.edu, http://nncf.unl.edu/conf/call.html*]

8. Fifth Annu. Health Law Symp.: Communities of Color and Genetic Testing; Newark, NJ [K. Boozang, 201/642-8871, Fax: -8194, *boozanka@lanmail.shu.edu*]

8–10. Intl. Workshop on Chromosome 5; Manchester, U.K. [M. Dixon, +44-161/275-5620, Fax: -3915, *mdpmjd@mh1.mcc.ac.uk*]

14. TIGR/NRC/DOE Distinguished Speaker Series: James Crow (Univ. of Washington); Washington, DC [D. Hawkins, 301/838-3501, Fax: -0209, *dhawkins@tigr.org, http://www.tigr.org/conference/speakers/ds9697.html*]

14–17. ASM Yeast Genetics and Human Disease Conf.; Baltimore [L. Nalker, 202/942-9254, Fax: -9340, *lnalker@asmusa.org, http://www.asmusa.org/mtgsrc/yeast1.htm*]

14–17. 4th Meeting of European Working Group on Human Gene Transfer and Therapy; Leiden, Netherlands [H. van Gennep, +31-71/514-8203, Fax: /512-8095, *r.c.hoeben@biochemistry.medfac.leidenuniv.nl*]

17–21. 24th Aharon Katzir-Katchalsky Conf.: Bioinformatics-Structure; Jerusalem [R. Goldstein, +972-8/934-2148, Fax: /947-4425, *http://bioinformatics.weizmann.ac.il/conf/pdb25sw10/*]

21–23. Nucleic Acids: Integrating Molecular Diagnosis and Therapy; San Diego [AACC, 202/857-0717, Fax: /833-4576]

December 1996

9–10. Biological Approaches and Novel Applications for Molecular Nanotechnology; San Diego [IBC, 508/481-6400, Fax: -7911, *inq@ibcusa.com, http://www.io.org/~ibc/*]

11–13. Functional Genomics; IBC, San Diego [see contact: Dec. 9-10]

January 1997

6–7. Gene Quantification; CHI, San Diego [see contact: November 4-5]

6–9. Pacific Symp. on Biocomputing; Hawaii [F. De La Vega, 525/747-7000, ext. 5355, Fax: -7100, *fvega@gene.cinvestav.mx, http://www.cgl.ucsf.edu/psb/*]

7–10. BioEast '97; Washington, DC [H. Matysko, 914/834-3100, Fax: -3689]

8–9. Genomic Targets for Drug Development; CHI, San Diego [see contact: November 4-5]

8–10. Physical Mapping of Plant Chromosomes; Aberystwyth, Wales, U.K. [N. Jones, +44/1970622-230, Fax: -307, *rnj@aber.ac.uk, http://scaffold.biologie.uni-kl.de/Beanref/physmapchr.htm*]

12–16. Plant and Animal Genome V Conf.; San Diego [D. Scherago, 212/643-1750, Fax: -1758, *pag5@scherago.com, http://pgenome.arsusda.gov:8000/pag5draft.html*]

16. TIGR/NRC/DOE Distinguished Speaker Series: Arthur Caplan (Univ. of Penn.); Washington, DC [see contact: Nov. 14]

18–22. LABAutomation '97: Laboratory Robotics and Automation Conference; San Diego [P. Stojadinovic, 619/646-8263, Fax: /452-6653, *petar@sequana.com, http://labautomation.org/*]

20–22. 1st Annu. Intl. Conf. on Computational Molecular Biology; Santa Fe, NM [S. Istrail, 505/845-7612, Fax: -7442, *scistra@cs.sandia.gov, http://www.cs.sandia.gov/recomb97/*]

25–28. Small Genomes: Sequencing, Functional Characterization, and Comparative Genomics; Hilton Head, SC [C. Sadler, 301/838-3509, Fax: -0229, *seqconf@tigr.org, http://www.tigr.org/conference/conference.html*]

26–30. Ninth Intl. Symp. on High Performance Capillary Electrophoresis and Related Microscale Techniques; Anaheim, CA [S. Schlessinger, 312/527-2011, *http://www.beckman.com/biorsrch/sympo/hpce97.htm*]

February 1997

1–5. Miami Nature Biotechnol. Winter Symp.— Advances in Gene Technology; Fort Lauderdale, FL [Meeting Coordinator, 800/642-4363, Fax: 305/324-5665, *mbws@mednet.med.miami.edu*]

16–21. Structure, Function, Expression, and Regulation of Genes and Proteins; Santa Fe, NM [Keystone Symp., 800/253-0685 or 970/262-1230, Fax: -1525, *keystone@symposia.com, http://www.colorado.net/symposia/*]

17–21. 19th Annu. Conf. on Organisation and Expression of the Genome; Lorne, Victoria, Australia [R.A. Sturm, +617/3365-1831, Fax: -4388, *r.sturm@mailbox.uq.edu.au*]

24–26. Human Genome Project: Commercial Implications; CHI, San Francisco [see contact: Nov. 4–5]

27. TIGR/NRC/DOE Distinguished Speaker Series: James Watson (CSHL); Rockville, MD [see contact: Nov. 14]

27–28. Genetic Screening and Diagnosis of Human Diseases; CHI, San Francisco [see contact: Nov. 4–5]

28–Mar. 2. 4th Joint Clinical Genetics Meetings: MOD and ACMG; Ft. Lauderdale, FL [M. Greenfield, 301/571-1887, Fax: -1895, *mgross@genetics.faseb.org, http://www.faseb.org/genetics/acmg/ann-meet.htm*]

March 1997

4–5. Chromosome 16 Workshop; Toronto [N. Doggett, 505/665-4007, Fax: -3024, *doggett@gnome.lanl.gov*]

6–8. HGM '97; Toronto [HUGO, 301/654-1477, Fax: /652-3368, *http://hugo.gdb.org/hgm97.htm*]

13. TIGR/NRC/DOE Distinguished Speaker Series: David Botstein (Stanford Univ.); Washington, DC [see contact: Nov. 14]

16–19. Fourth Intl. Conf. on Automation in Mapping and DNA Sequencing; Heidelberg, Germany [I. Fatscher, +49-6224/929-025, Fax: -026, *fatscher@embl-heidelberg.de, http://www.embl-heidelberg.de/CourseInfo/AMS97/AMS97.html*]

17–18. Symp. on Genomic Medicine; Hilton Head, SC [see contact: Jan. 25–28]

21–23. Intl. Workshop on Chromosome 10; Leeds, U.K. [N. Spurr, +44-171/269-3846, Fax: -3802, *spurr@icrf.icnet.uk*]

31–April 3. 11th Intl. Conf. on Math. and Computer Modeling & Scientific Computing; Washington, DC (abs. deadline: Oct. 31) [X.J. Avula, 573/341-4585, Fax: /364-3351, *avula@umr.edu*]

April 1997

4–5. 15th Annu. SERGG Meeting; Atlanta [M. Lane, 404/727-5844, Fax: -5783, *mrl@ru.ped.emory.edu, http://www.cc.emory.edu/PEDIATRICS/sergg/meeting/meeting.htm*]

13–19. Molecular & Cellular Biology of Gene Therapy; Keystone Symp., Snowbird, UT [see contact: Feb. 16–21]

14–19. 9th Intl. Cong. on Genes, Gene Families, and Isozymes; San Antonio, TX [Barr Enterprises, J. Cunningham, 301/898-3772, Fax: -5596]

16–20. 38th Annu. Drosophila Res. Conf.; Chicago [M. Ryan, 301/571-1825, Fax: /530-7079, *mryan@genetics.faseb.org*]

17–18. 5th Intl. Nature Genetics Conf. — Functional Genomics: From Genes to Drugs; Washington DC [Cambridge Symp., 617/630-1399, Fax: -1395, *symposia@xensei.com, http://www.cambridge.org/symposia/*]

May 1997

14–18. Genome Mapping and Sequencing; Cold Spring Harbor, NY [CSHL, 516/367-8346, Fax: -8845, *meetings@cshl.org, http://www.cshl.org/*]

October 1997

28–Nov. 1. ASHG; Baltimore [M. Ryan, 301/571-1825, Fax: /530-7079]◊

Training Calendar*

November 1996

15–16. Practical Biotechnol. for Teaching Lab.; Norwalk, CA [J. Boyle, 310/860-2451, ext. 2682, Fax: /467-5005, *jsbhitek@aol.com*]

January 1997

15–16. Applications of Biotechnology for Society; Norwalk, CA [see contact: Nov. 15–16]

May 1997

4–7. Genetic Analysis Methods for Medical Researchers (focus: human genetic disease mapping); Durham, NC (app. deadline: Feb. 1) [V. Roberts, 919/684-6274, Fax: -6514, *genclass@genemap.mc.duke.edu, http://www.mc.duke.edu/depts/genetics/courses/index.html*] ◊

Extended calendars and a list of organizations offering training are available at *http://www.ornl.gov/hgmis/* or from HGMIS (see p. 12 for contact information).

*Dates and meeting status may change; courses may also be offered at other times and places; check with contact person.
**Attendance is either limited or restricted.

SBIR 1996 Human Genome Awards Announced

In July the DOE Office of Health and Environmental Research announced awards in human genome topics from the Small Business Innovation Research (SBIR) program. The highly competitive SBIR awards are designed to stimulate commercialization of federally funded research and development (R&D) for the benefit of both private and public sectors. SBIR emphasizes cutting-edge, high-risk research with potential for high payoff in hundreds of areas, including human genome research.

The SBIR program was initiated in 1982 to provide opportunities for science- and technology-based businesses with 500 employees or less to compete among themselves for federal R&D awards. In 1992 Congress reauthorized the SBIR Program until October 1, 2000. Eleven agencies, those with extramural R&D budgets of over $100 million, are required to maintain an SBIR program using a set-aside of a percentage of their budgets. The legislation provides for a gradual set-aside increase from 1.25% in 1992 to a maximum of 2.5% in FY 1997 and thereafter.

1996 Awards in Genome and DNA-Related Technologies

Phase I

- BIOS Laboratories, Inc. (New Haven, Connecticut): Directed Multiplex DNA Sequencing by Hybridization
- Promega Corporation (Madison, Wisconsin): An Engineered RNA/DNA Polymerase to Increase Speed and Economy of Deoxyribonucleic Acid Sequencing

Phase II

- ApoCom, Inc. (Oak Ridge, Tennessee): GRAIL-GenQuest: A Comprehensive Computational System for DNA Sequence Analysis
- CyberConnect Corporation (Storrs, Connecticut): A Graphical Ad Hoc Query Interface Capable of Accessing Heterogeneous Public Genome Databases

SBIR Phases

- Phase I: Awards for up to 6 months and $75,000 for a firm to explore the scientific and technical merit and feasibility of a research idea.
- Phase II: Awards for up to 2 years and $750,000 to expand on Phase I results and pursue further development. Only Phase I awardees are eligible for Phase II, which is the principal R&D effort.
- Phase III: Private or non-SBIR federal funding for commercialization of Phase II results.

SBIR Conferences

National SBIR conferences are held periodically to help small business firms identify R&D and marketing opportunities. Such subjects as procurement, auditing, finance, accounting, proposal preparation, and licensing are explored. Upcoming conferences are listed in the box at right.

Solicitation Announcements

SBIR Solicitation Announcements are available electronically (*http://sbir.er.doe.gov/sbir.htm*) or in hard copy (301/903-5707). The DOE SBIR program contact is Kay Etzler (see box at right). DOE SBIR information is included regularly in the funding box in each *HGN issue*.◊

NIGMS Catalog on Web

The NIGMS Human Genetic Mutant Cell Repository catalog is now available on the Web (*http://arginine.umdnj.edu/coriell/nigms.htm/*).◊

U.S. Genome Research Funding

Investigators wishing to apply for funding are urged to discuss projects with agency staff before submitting proposals.

DOE Office of Health and Environmental Research (OHER) Human Genome Program

- Contact for funding information or general inquiries: *genome@er.doe.gov* or 301/903-6488
- Relevant documents: *http://www.er.doe.gov/production/oher/hug_top.html*

Alexander Hollaender Distinguished Postdoctoral Fellowships (DOE)

Research opportunities are available in energy-related life, biomedical, and environmental sciences, including human genome, global change, and supporting disciplines.

- Next deadline: January 1997
- Contact: Barbara Dorsey, Oak Ridge Institute for Science and Education (423/576-9975, Fax: /241-5219)

NIH National Center for Human Genome Research (NCHGR)

Program announcements are listed in *NIH Guide for Grants and Contracts* (*http://www.nih.gov/*).

- NCHGR Program Contact: 301/496-7531, Fax: /480-2770, *http://www.nchgr.nih.gov/*
- ELSI: 301/402-4997

Small Business Innovation Research (SBIR) Grants

DOE and NIH invite small business firms (less than 500 employees) to submit grant applications addressing the human genome topic of SBIR programs. The two agencies also support the Small Business Technology Transfer (STTR) program to foster transfers between research institutions and small businesses. Contacts:

- Kay Etzler; c/o SBIR Program Manager, ER-16; DOE; Washington, DC 20585 (301/903-5867, Fax: -5488, *kay.etzler@oer.doe.gov*)
- Bettie Graham (see contact, NCHGR). NIH SBIR due April 15, August 15, and December 15. STTR, December 1

National SBIR/STTR conferences: Anaheim, CA (Nov. 13–15, 1996); Orlando, FL (April 2–4, 1997). Conference information: 203/205-6450.◊

HGN E-Mail Notification

Readers who would like to be notified when the latest issue of *HGN* is placed on the Web should send their request and e-mail address to *yustln@ornl.gov*. A table of contents and URL will be e-mailed to requestors when each issue is posted. This electronic version usually is available at least 2 weeks before readers receive the printed copy.◊

Human Genome Management Information System Subscription/Document Request* (Vol. 8, No. 1)

Name _____
(First) (MI) (Last)

Affiliation _____

Department/Division _____

Street/P.O. Box/Building _____

City/State/Zip Code _____

Country _____ **Area of Interest** _____

Phone _____ **Fax** _____

E-Mail Address (important to list if you have one) _____

1. ___ *Human Genome News* ___New Subscriber ___Change of Name/Affiliation/Address (circle all that apply) ___Drop Subscription
2. ___ Print copy of Santa Fe '96 Meeting Abstracts ___DOE *Primer on Molecular Genetics* (see *http://www.ornl.gov/hgmis/publicat/publications.html)*
3. ___ Reprint of "A New Five-Year Plan for the U.S. Human Genome Project" *(Science*, October 1, 1993) by Francis Collins and David Galas

*Please type, print carefully, or enclose a business card to ensure efficient shipping. To change name/address/affiliation or drop your subscription to *Human Genome News*, enclose your current *HGN* address label. Send to HGMIS address shown below and on p. 12.

SELECTED ACRONYMS

AACC Am. Assoc. for Clinical Chem.
ACMG Am. Coll. of Medical Genet.
AMIA Am. Med. Informatics Assoc.
ASM Am. Soc.for Microbiology
BAC bacterial artificial chromosome
bp base pair
CHI Cambridge Healthtech Inst.
cM centimorgan
CSHL Cold Spring Harbor Lab.
DHHS Dept. Of Health and Human Services
DOE Dept. of Energy
ELSI ethical, legal, and social issues
EST expressed sequence tag
FISH fluorescence in situ hybridization
HGM Human Genome Meeting
HGMIS Human Genome Management Information System
HUGO Hum. Genome Org.
IBC Intl. Bus. Communications
IEEE Inst.of Electrical and Electronics Engineers
MIT Mass. Inst.of Technology
MOD March of Dimes
NCHGR Natl. Ctr. for Human Genome Research
NIH Natl. Institutes of Health
NRC National Research Council
PAC P1 artificial chromosome
SCI Society of Chemical Industry
SERGG SouthEastern Regional Genetics Group
STS sequence tagged site
TIGR The Inst. for Genomic Res.
WWW World Wide Web
YAC yeast artificial chromosome

☆U.S. GOVERNMENT PRINTING OFFICE: 1996-549-259/40002

OAK RIDGE NATIONAL LABORATORY • MANAGED BY LOCKHEED MARTIN ENERGY RESEARCH CORP. • FOR THE U.S. DEPARTMENT OF ENERGY

Betty K. Mansfield
HGMIS
Oak Ridge National Laboratory
1060 Commerce Park, MS 6480
Oak Ridge, TN 37830, U.S.A.

Postmaster: Do Not Forward.
Address Correction Requested.
Return Postage Guaranteed.

BULK RATE
U.S. POSTAGE
PAID
OAK RIDGE, TN
PERMIT NO. 3

Printed on Recycled paper