Human Genome news

Sponsored by the National Institutes of Health and the U.S. Department of Energy

ISSN: 1050-6101

Vol. 5, No. 3, September 1993

Bioinformatics "Highway" Needed

DOE-Supported Meeting Stresses Software, Databases for Map and Sequence Data

The DOE Office of Health and Environmental Research (OHER) convened a meeting of informatics experts in Baltimore on April 26–27 to assess current OHER bioinformatics efforts and obtain advice on essential planning and coordination of future activities. The meeting was a continuation of a long-term planning strategy initiated in 1992 with an OHER review of the DOE portfolio of independent genomic informatics projects and core activities at genome centers [*HGN* **5**(1), 3–4 (May 1993)]. OHER supports research in genome informatics, structural biology, and other programs requiring integrated applications of biological data. Meeting topics included the biological research community's informatics needs, especially for databases, and actions to ensure that these needs will be met. Although this meeting grew out of an ongoing review of DOE-supported activities, the report of the meeting is being circulated to the general community (see sidebar). The National Center for Human Genome Research (NCHGR) and OHER expect to use the report and the community comments to develop plans for maintaining their coordinated support of research and development of genome informatics tools and systems.

OHER is releasing a report of this meeting as a part of its ongoing planning process. With input from the genome community, the agency hopes to improve and expand the report into a white paper. The report is available from HGMIS (see address, p. 12) and through the Johns Hopkins University Gopher at *gopher.gdb.org* under the Mathematics and Biology heading.

Community Databases

A general discussion focused on the role of community databases in facilitating OHER-supported research. More than mere archives, genomic databases provide analysis tools for project bench work. Reviewers concluded that existing community databases fall short of meeting community needs. The problems stem both from rapidly changing requirements and from conceptual and technical idiosyncracies in the design of current systems.

Investigators require access to databases containing protein sequences, crystallographic structures, nucleotide sequences, and mapping data (genes, maps, and probes). Databases managing this information include

- PIR (Protein Information Resource) and SWISS-PROT;
- PDB (Protein Data Bank);
- GenBank®, EMBL (European Molecular Biology Laboratory) Nucleotide Sequence Database, DDBJ (DNA Data Bank of Japan); and
- Species-specific mapping databases such as GDB (Genome Data Base) for humans, FlyBase for *Drosophila*, and MGD (Mouse Genome Database).

Database Interoperability

Wide-ranging and vigorous discussion was held on requirements for DNA sequence and mapping databases. Participants felt

In This Issue. . .

Genome News

Topics:

- Community Databases
- Database Requirements
- User Needs
- Data Submission, Curation
- Coordination
- Tech Transfer, Sharing
- Training

Coordination Among Databases, Internet Access Are Essential

strongly that interoperability is critical for conducting bioinformatics research and ensuring that information in biological databases is optimally useful to other research areas.

Coordination among genome databases and other informatics systems must be the highest priority. Software and database projects must interact with and consider other efforts, or the cost of systems development will be too high. In addition, projects that do not successfully interact with other databases will lead to inefficiencies associated with unlinked data. Users must be able to retrieve related data from multiple databases such as GDB, PIR, Medline, PDB, and GenBank without having to query each database separately and integrate the results themselves.

In addition to database interoperability within a specific research domain (e.g., the Human Genome Project), many workers also need integrated access to a variety of biological data areas. For example, studies of gene products and their functions may be outside genome project scope, but the value of genome data would be maximized by linking project results to databases dedicated to understanding gene products and their functions.

Community databases should be designed generically for interoperability, requiring both semantic and technical consistency among projects. For minimum semantic linkage, biological objects in all interoperating databases must be identified and cross-referenced via accession numbers.

To achieve minimum technical linkage, participating systems must at least present similar application programming interfaces through the Internet. Currently, this is achieved most cost-effectively when all interoperating databases are implemented as relational databases that support Internet standard query language (SQL) queries. Other architectures may eventually be optimal, but for now all databases need to support current community standards for data access (e.g., remote data access or SQL). Also, participating databases should be (1) self-documenting (offering an online data dictionary and other documentation), (2) stable (undergoing schema change rarely and only after ample warning), and (3) consistent (based on federation-wide semantics). However, goals of stability and consistency conflict with the need for maximum responsiveness to changing community needs.

Local incentives now often work against interoperability. Usually, community genome-relevant databases are not funded by the same agencies, advised by the same advisors, or otherwise coordinated. Community database funding is always limited, and interoperability is not often top priority; federal agencies and others will need to provide incentives to advance these priorities.

A truly federated information infrastructure cannot meet community needs without a minimum level of semantic consistency among data submitters and databases. In addition, an infrastructure will be needed to permit a client to make automatic queries to multiple databases at different locations. These databases should appear to an end user as one integrated database.

Database User Requirements

Users and producers of genome information generally fall into several institutional categories with different capabilities and needs. They include (1) genome centers; (2) independent laboratories at major research institutions; (3) individual investigators and small laboratories; and (4) other users, such as those in small clinical settings, homes, and high schools.

In discussing user requirements, the group agreed that Internet access will become an essential highway for routine, up-to-date data submission, retrieval, and analysis. Investigators and administrators should communicate with those in other scientific disciplines to increase awareness of Internet's importance in data communication and distribution. Genome centers and research institutions have different local requirements, so the need for local software tools and database designs will also vary; therefore, robust support will be required for onsite acquisition and handling and automated submission of mapping and sequencing data among sites.

Local user-support systems and academic or commercial software should be developed for interaction with community databases, and dispersed user communities will need easy-to-use data-access tools. Meeting the needs of various user classes will require widely accepted application programming interfaces to databases and associated software.

Users will need databases and software analysis tools capable of reviewing millions of base pairs of new sequence each

month. Community software tools should at minimum allow machine-readable input and output. Analysis tools with differing data formats create unnecessary difficulties for automated analysis.

Access to individual databases will require analysis algorithms, interfaces, and other tools incorporated into easy-to-use suites of software and servers that can run across the network. Users will also need integrated views of data from multiple databases and interfaces as well as database-linking software that is scaleable to the large amount of biological data to be incorporated.

Data Submission and Curation

Even if great success attends bench research in the Human Genome Project, the project will not be successful unless all researchers can retrieve answers to their genome-related questions. Systems for improved data submission and curation will enable investigators to submit data to robust community databases promptly and easily and receive useful and trustworthy data.

Genome researchers, funding agencies, and the entire biological community should ensure that an infrastructure for capturing data is in place. Discussants expressed skepticism that data capture by journal scanning will be adequate for total data submission. Many journals require database submission of sequence entries and protein structural coordinates before publication, but subsequent information about the sequence or protein is not always added. Discussants suggested that journals may also need to require submission of other types of information and to reference all relevant accession numbers if the article adds annotation information.

Database curation, and probably community-based curation, will become increasingly important. A possibility raised was a new professional job category, similar to museum curators, to the maintain databases. Routine capture of data from nongenome laboratories will also be needed, and users from these laboratories should be able to add routinely to the information.

Coordination of Informatics Efforts

Attendees noted that previous Human Genome Project 5-year goals focused more on the community than on the core-support local-user domain. To achieve mapping and sequencing goals, core informatics support will be needed at genome centers and for center-to-center dispersed collaborations and robust connections.

Participants stated that more information should flow between users and programmers at different centers. Computer demonstrations at the 1993 DOE genome contractor-grantee workshop in Santa Fe, New Mexico, were praised as a step in the right direction because they allowed hands-on informatics interaction between biologists and computer experts. Other suggestions included a sequence-analysis software fair in addition to the data fair at the Hilton Head genome sequencing meeting; joint NIH-DOE genome project–wide meetings similar to the DOE Santa Fe meetings; NIH-DOE informatics workshops with experts from genome centers and major databases; and more visiting and interchange among centers, perhaps through short-term exchange of personnel.

The need for coordinating the DOE genome program with NCHGR efforts was emphasized, as was cooperation with other NIH institutes, National Science Foundation, U.S. Department of Agriculture, and foreign efforts. Attendees felt that the rate of new-software implementation should be increased and that strategies should be considered for funding more software-development research, maintaining resource databases, supporting servers at various sites, and integrating diverse software into common servers or sets of tools.

Technology Transfer, Software Sharing

Attendees noted varying degrees of transfer and sharing, from exchanging experiences in developing software and databases to sharing source codes and schemata.

Genome center informatics core support activities such as databases and software tools cannot be transferred easily among centers because of deeply embedded differences in experimental methods. Nonetheless, off-site and some on-site projects are explicitly funded to provide research results and resources to the wider community. Participants felt that informatics tools should be made more widely available through software libraries and resource listings. They also noted the need to implement community-based research into easily accessible tools.

(see Informatics, p. 4)

Genome News

Information Exchange Encouraged Between Users, Programmers

Software Libraries, Resource Listings Needed

✍ Non-U.S. Subscribers:

If you have not requested to remain on the *HGN* mailing list, this is your last issue!

Genome News

Taq Polymerase Available at Reduced Rate

A new pricing agreement negotiated by NIH between federally supported genome programs and a commercial supplier of AmpliTaq® will allow genome researchers to purchase large quantities of the enzyme at a reduced rate. Manufactured by Roche Molecular Systems, Inc., and sold by Perkin-Elmer Corporation, *Taq* polymerase is a key enzyme used in the polymerase chain reaction (PCR). Each year investigators working in a single Human Genome Project laboratory perform up to 200,000 reactions that would cost about $100,000 under previous pricing agreements.

Perkin-Elmer and Roche have also established a matching volume-for-volume grant program for the 17 Genome Science and Technology Centers (called GESTECs) supported by the NIH National Center for Human Genome Research (NCHGR). DOE-designated genome centers are also eligible under the agreement. In addition, Perkin-Elmer will supply AmpliTaq to institutions in bulk form with a pricing schedule dependent upon the amount purchased each year.

"The new agreement will enhance our ability to carry out the genome project in a cost-effective way," said NCHGR Director Francis Collins. "We are fully behind any reasonable efforts to save public dollars." Collins also stated that the two companies are offering to collaborate with genome centers to help develop new applications of PCR and DNA analysis and are not requesting special rights to any technologies in exchange for the grants.◊

HGMIS Hosts DOE Honors Program Students

As part of the ninth DOE High School Science Students Honors Program, the Human Genome Management Information System (HGMIS) and the Transgenic Animal/Targeted Mutation Database (TBASE) group were hosts to 13 young people for 2 weeks this summer at Oak Ridge National Laboratory (ORNL). About 60 selected high school science students representing all the states and territories, the District of Columbia, and several foreign countries came together at ORNL for this program, which is designed to encourage promising students to pursue science as a career.

While at ORNL, the combined HGMIS and TBASE student groups researched the Human Genome Project and gene function literature and visited several laboratories to observe work and receive instruction. The students also conducted a telephone interview with Daniel Drell, who oversees the Ethical, Legal, and Social Issues Program of the DOE Human Genome Program. At the end of their stay, they produced a paper and oral presentation on aspects of the Human Genome Project and applications of transgenic animal research to the understanding of gene function.

Other DOE facilities taking part in the honors program included Fermi National Accelerator Laboratory; Lawrence Berkeley Laboratory; and Lawrence Livermore, Brookhaven, Argonne, and Sandia national laboratories. The project is managed by the University and Science Programs section of the DOE Office of Science and Technology. *[Anne Adamson, HGMIS]* ◊

DOE Program Encourages High School Students To Pursue Science Careers

Informatics *(from p. 3)*

The entire genome community and individual centers will at times need to compare results and approaches through the exchange of data and analyses performed with common standard analysis tools. A mechanism was proposed by which one or more servers could house a suite of analysis tools incorporating research results and software design of several projects.

Training Programs

Training programs, particularly institutional training grants that permit sites to develop courses and support students, are necessary to produce multidisciplinary people who support informatics. Attendees felt that individual fellowships should be maintained or increased. The possibility of short-term travel fellowships to allow for greater exchange of ideas and results was discussed. *[Jay Snoddy and Robert Robbins, DOE OHER, and Anne Adamson, HGMIS]* ◊

☛ ***Resource Available***

Update Mouse Nomenclature by E-Mail

Weekly or biweekly mouse locus nomenclature updates are available by e-mail from the Mouse Genome Informatics Project at Jackson Laboratory. These updates include lists of new, revised, withdrawn, and reserved locus symbols, names, and references. For more information or to be added to the mailing list, send an e-mail message to *davidnaman@jax.org. [Janan T. Eppig, Jackson Laboratory]* ◊

Galas Leaves OHER for New Biotech Company

After serving more than 3 years as Associate Director of the DOE Office of Health and Environmental Research (OHER), David J. Galas has joined Darwin Molecular Corporation, a biotechnology company located in Seattle, Washington. He also plans to resume some of his academic activities. Aristotle Patrinos, Director of the OHER Environmental Sciences Division, has been named Acting Associate Director to replace Galas.

Galas had been appointed to his OHER position under the Intergovernmental Personnel Act, which is used by some federal agencies to enlist individuals from academia and state and local governments on a temporary basis. Galas was on leave from the University of Southern California, where he was Director of Molecular Biology and a Professor of Biological Sciences. His best-known scientific accomplishment is the "footprinting" technique, a method widely used to define DNA sequence-specific sites for protein binding.

While at OHER, Galas supervised five divisions with an annual budget of about $350 million. He provided overall leadership for the DOE Human Genome Program and three other interdivisional programs and expanded genome research areas to include more biological components. Under his direction, a comparative study was implemented to map and partially sequence human and model organism genomes (especially the mouse) to obtain greater understanding of genome organization and function.

Galas also initiated and led the effort to coordinate and integrate all government-sponsored biotechnology research, which became known as the Presidential Initiative in Biotechnology Research. Recommendations from *Biotechnology for the 21st Century*, a special report generated through the initiative by the Committee on Life Sciences and Health, were included in the FY 1993 presidential budget. A strategic goal cited in the report is accelerated transfer of biotechnology research to commercial applications, a process seen as crucial in maintaining U.S. leadership in the rapidly growing biotechnology industry.

Galas will direct research and development efforts at Darwin Molecular Corporation, one of many new companies formed to pursue commercial opportunities arising from genome research. Darwin plans to establish a large-scale facility to lower costs and increase by 10- to 100-fold the sequencing throughput of disease-associated human genome regions . Gene targets for drug design and evolution and their appropriateness for further development will be determined through structural information generated by computer analysis of these sequences.

Biotech Companies Use New Approach to Drug Discovery

To create and screen huge numbers of diverse molecules for potential therapeutic use, Darwin and other biotechnology companies are using promising new techniques such as directed molecular evolution. This method allows the rapid construction in test tubes of large peptide, protein, and oligonucleotide libraries, which are then subjected to functional screening. Molecules that best perform a desired task (i.e., the "fittest" molecules) are reproduced by the millions and tested again. Screening is repeated generation after generation until new designer-made molecules with unique desired properties are produced, usually within a few weeks. This approach to drug discovery has proven very powerful in generating new chemical entities with the ability to bind to a number of therapeutically useful targets active in heart disease, inflammation, and AIDS.◊

✍ Non-U.S. Subscribers:

If you have not requested to remain on the *HGN* mailing list, this is your last issue!

David J. Galas

Galas Expanded Biological Components of the DOE Human Genome Program

Genome News

SBIR Set-Aside Funds Increased

OHER, NCHGR Grant New Awards for Human Genome Research

The DOE Office of Health and Environmental Research (OHER) and the NIH National Center for Human Genome Research (NCHGR) recently announced the latest awards in human genome topics of the Small Business Innovation Research (SBIR) program. Four Phase I and three Phase II awards for this cycle were identified by OHER, with NCHGR announcing five in Phase I and six in Phase II.

SBIR awards are designed to stimulate commercialization of new technology for the benefit of both private and public sectors. The highly competitive SBIR program emphasizes cutting-edge, high-risk research with potential for high payoff in hundreds of areas, including human genome research. SBIR human genome topics concentrate on innovative approaches and experimental technologies for carrying out the goals of the Human Genome Project—to map and sequence genes and genetic regions and develop databases for storing the resulting data. NIH and DOE SBIR awards support these goals, including improvements in technology, mapping resources, and DNA sequencing and in enhanced storage, processing, and analysis of genetic data.

Prospective genome project applicants are urged to discuss their plans with DOE and NIH staff before preparing formal proposals. For more information on SBIR genome programs, contact Kay Etzler or Bettie Graham at the addresses in the side bar.

SBIR Program History

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 research and development (R&D) awards. Since 1982, $2.765 billion has been distributed in 21,427 awards to firms from all 50 states, Puerto Rico, and the District of Columbia. In FY 1991, 3341 SBIR awards totaling about $483.1 million were made to small businesses.

In 1992 Congress reauthorized the SBIR Program until October 1, 2000. The legislation provides for a gradual increase in the SBIR set-aside funds of each participating federal agency's extramural research or R&D budget from 1.25% in 1992 to a maximum of 2.5% in FY 1997 and thereafter. For FY 1994, about $700 million is expected to be allocated government wide for SBIR, an amount that will grow to around $1 billion by FY 1997.

SBIR Program Phases

The general governmental SBIR program consists of three phases.

Phase I: Awards for up to 6 months and $100,000 (DOE and NIH, $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 (DOE and NIH, $500,000) 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. Subjects such as procurement, auditing, finance, accounting, proposal preparation, and licensing are explored. Upcoming conferences are in Washington, D.C. (October 13–15); Seattle (November 15–17); and Houston (April 26–28, 1994). For registration or further meeting information, call the conference Hotline (407/791-0720).

Pre-Solicitation Announcements

The SBIR Pre-Solicitation Announcement (*PSA*) is published by the Small Business Administration (SBA) each quarter, usually in March, June, September, and December. This announcement provides advance information on SBIR solicitations to be released by the 11 participating federal agencies. To be added to the *PSA* mailing list, contact the Office of Innovation, Research, and Technology; SBA; 409 Third Street, SW (8th Floor); Washington, DC 20416 (202/205-7777).◊

CONTACTS

■ **DOE:**
Kay Etzler
c/o DOE SBIR Program Manager
ER-16
Washington, DC 20585
301/903-5867
Fax: -5488

■ **NIH Genome Program:**
Bettie Graham
Bldg. 38A, Rm. 610
NIH
9000 Rockville Pike
Bethesda, MD 20892
301/496-7531
Fax: /480-2770

DOE and NIH SBIR information is included in the funding box on page 15 of each *HGN* issue.

Officers, Activities Given for HUGO

Officers of the international Human Genome Organization (HUGO), elected in the fall of 1992, assumed their duties in January of this year. C. Thomas Caskey (Baylor College of Medicine) became HUGO's third President; Ronald G. Worton (Hospital for Sick Children, Toronto), Vice President for the Americas; and Diane Hinton (HUGO Americas), Secretary. Other Vice Presidents serving are Andrei Mirzabekov (Englehardt Institute, Moscow) and Kenichi Matsubara (Osaka University). Robert Sparkes (University of California, Los Angeles) is Treasurer.

With over 500 members representing 32 countries, HUGO continues to be the only private organization devoted exclusively to facilitating coordination of worldwide genome research efforts.

Activities initiated recently by HUGO include:

- Research travel awards (see next column).
- Genome summit planned for January 1994 in Houston to enhance international communication through a gathering of scientific and administrative leaders.
- Ethics yearbook to create an international resource of scholarly articles and timely bulletins related to activities such as legislative initiatives and meetings [Alexander Capron (University of Southern California) and Bartha Knoppers (University of Montreal)]. Funding for pilot planning is being provided by DOE and the Canadian Genome Project.
- Human Genome 94 Meeting (in collaboration with *Science*) to be held October 3–5, 1994, in Washington, D.C.
- Intellectual property rights meeting for the United States and Canada planned for 1994, with an international conference to follow.

HUGO continues to work closely with international funding agencies on Single Chromosome Workshops, Chromosome Coordinating Meetings (CCMs), and the Human Genome Mapping (HGM) Workshop. CCM 93 and HGM 93 are being held in Japan on November 10–12 and 14–17, respectively.

HUGO Travel Awards

HUGO has announced new travel awards for short-term laboratory visits by junior investigators who wish to transfer technology or conduct collaborative research with genome laboratories. The awards will cover up to $1500 for travel expenses only. For further information, contact the HUGO Europe or HUGO Americas office.

HUGO Europe
One Park Square West
London NW1 4LJ, U.K.
+ 44/71-935-8085, Fax: -8341

HUGO Americas
7986-D Old Georgetown Road
Bethesda, MD 20814
301/654-1477, Fax: /652-3368 ◊

HUGO Council

Terms Ending 1995

Walter Bodmer	United Kingdom
C. Thomas Caskey	United States
David Cox	United States
Victor McKusick	United States
Gert-Jan van Ommen	Netherlands
Ronald Worton	Canada

Terms Ending 1994

Kay Davies	United Kingdom
Eric Lander	United States
Jean-Louis Mandel	France
Kenichi Matsubara	Japan
Ulf Pettersson	Sweden
Nobuyoshi Shimizu	Japan

Terms Ending 1993

Francis Collins	United States
John Evans	United Kingdom
Leroy Hood	United States
Andrei Mirzabekov	Russia
Grant Sutherland	Australia
James Wyngaarden	United States

HUGO Officers Elected for 1993

- **Thomas Caskey**
- **Ronald Worton**
- **Diane Hinton**

Activities

- **Genome Summit, January 1994**
- **Ethics Yearbook**
- **Human Genome 94 (with the journal *Science*)**
- **Intellectual Property Rights Meeting**

Genome News

Gopher Permits

- **Quicker Access to Newsletter Text for Foreign Subscribers**
- **Electronic Searching of Current and Back Issues**

HGN Available on JHU Gopher

Current and back issues of *Human Genome News* are now available in searchable format on the Johns Hopkins University Computational Biology Gopher system. This will allow much quicker access to newsletter text than with mail distribution. Detailed instructions on electronic searching follow. (To obtain a Gopher client, see item on p. 9).

Accessing and Using Gopher

If Internet and Gopher capabilities are present, point the gopher client at **gopher.gdb.org** (Port 70) and select

5. Genome Project/

2. Human Genome News/

The following menu will then appear:

1. About the Human Genome Newsletter.

2. Search All Issues of the Human Genome Newsletter?

3. Browse the Current Human Genome Newsletter (September - 1993)/

4. Browse Through All Issues of the Human Genome Newsletter/

Searching Help (For Selection 2: Search All Issues). To search all *HGN* issues, a word or series of words can be entered and articles containing at least one of the words will be listed. The search can be further refined by connecting the words with the Boolean terms **and**, **or**, and **not**. When search words are connected with **and**, records containing all the words will be listed. Use of **or** usually yields more results and will produce records containing one word or the other. Using **not** will show records containing the first but not succeeding words. Wildcards (*) may be used to search for partial words. For example, to search for **genome**, **genomes**, **genomic**, or **genomics**, use **genom***. For very restrictive searching, a phrase (e.g., **"cDNA sequencing"**) may be used.

Users without a Gopher client can obtain access through telnet. At the system prompt, type **telnet**, then **consultant.micro.umn.edu.** Use **gopher** as login. When the menu appears, select

8. Other Gopher and Information Servers/

8. North America/

3. USA/

20. Maryland/

1. Computational Biology (Welchlab—Johns Hopkins University)/.

The Gopher version of *HGN* was developed by Dan Jacobson at Johns Hopkins University. For more information, e-mail to *danj@mail.gdb.org.*◊

☛ *Resources Available:* FTP Sites

Chromosome 4. The Human Genome Mapping Center (HGMC) at Stanford University has established an anonymous FTP site for rapid dissemination of chromosome 4 mapping data to interested investigators. To access the site, use the command *ftp toolik.stanford.edu* or *ftp 36.159.0.5.* Use *pub/uploads* to deposit material, and send e-mail to *hgmcinfo@camis.stanford.edu* to notify HGMC of the deposit, so it can be processed rapidly. Data is deposited mainly as ASCII (text) or postscript files in the *pub/hgmc directory* and its associated subdirectories. For more information about deposited data or the FTP site or to be informed through an electronic mailing list when new data is deposited, send a detailed e-mail message including name, affiliation, and full contact details to *hgmcinfo@camis.stanford.edu.* [HGMC; Department of Genetics; Stanford University; 855 California Avenue; Palo Alto, CA 94304 (415/812-1915, Fax: -1916).] ◊

Genetic Linkage Analysis. An anonymous FTP site is being made available for use by people interested in genetic linkage analysis. All software programs from the laboratory of Jurg Ott (Columbia University) are being uploaded to the FTP site. To access the site (preferably not in peak afternoon hours), use the command *ftp york.ccc.columbia.edu* or *ftp 128.59.97.32.* When asked for a login name, enter *anonymous.* A password is not required. Reports of problems should be sent to Joe Terwilliger at *joe@york.ccc.columbia.edu.* [Material taken from *Linkage Newsletter* **7**(2) (June 1993). To subscribe to the newsletter or for more information on programs, contact Katherine Montague (212/960-2507, Fax: /568-2750, Internet: *jurg.ott@columbia.edu*, BITNET: *ott@nyspi.bitnet).*] ◊

What Is Gopher, Anyway?

Gopher was born as a distributed campus information system at the University of Minnesota (home of the "Golden Gophers"). The online service was designed so each university entity (e.g., administration office, athletic department, academic departments) could have control over its own data and server. Gopher's developers organized the university's individual databases by topic into a system resembling one large database; they then created a special application to guide users to the information they wanted.

With the Internet to connect servers all over the world, the Gopher concept was enlarged to an international system that began with 1 site in May 1991 and has grown to around 2000 sites. Gopher's main function is to "go fer" things, making its name even more appropriate.

[Information taken from *The Whole Internet: User's Guide and Catalog* by Ed Krol (O'Reilly & Associates, Inc., 103 Morris Street, Suite A, Sebastopol, CA 95472 (800/998-9938 or 707/829-0515, Internet: *nuts@ora.com* or *uunet!ora!nuts*).] ◊

GDB E-Mail Addresses Change

New GDB e-mail addresses are listed below.

Technical Support
New User Address: *help@gdb.org*
(formerly *help@library.welch.jhu.edu*)

GDB Staff
New Staff Address: *staffname@gdb.org*
(formerly *staffname@library.welch.jhu.edu*)

Welchlab Server, using telnet, rlogin, or rsh
New Hostname: *gdb.org*
(formerly *welchlab.welch.jhu.edu*)

FTP
New Hostname: *ftp.gdb.org*
(formerly *mendel.welch.jhu.edu*)

FTP Command
New Command: *ftp ftp.gdb.org*
(formerly *ftp mendel.welch.jhu.edu*)

GOPHER
New Hostname: *gopher.gdb.org*
(formerly *merlot.welch.jhu.edu*)

WAIS
New Hostname: *wais.gdb.org*
(formerly *welchlab.welch.jhu.edu*)

OMIM
New Hostname: *omim.gdb.org* (OMIM may still be accessed through the GDB address; this is another OMIM access option.)
(formerly *welchlab.welch.jhu.edu*) ◊

GOPHER NEWS

Obtaining a Gopher Client

Users on the Internet are encouraged to obtain a Gopher client for direct use from the desk top. A Gopher client program runs on the user's local PC, Macintosh, workstation, or mainframe; it uses custom features of the local machine, allows the display of images, and takes advantage of features such as mouse, scroll bars, local printers, and local hard disk. Gopher client programs for Macintosh, DOS, Windows, OS/2, VAX/VMS, NeXTstep, X Windows (including Motif), UNIX, and several other systems are available without charge by anonymous FTP from *boombox.micro.umn.edu* in the *pub/gopher* directory.◊

Gopher Now on GDB Menu

GDB modem or telnet users can now select Gopher from the GDB menu:

5. Additional Services (MLC, Gopher)
3. Internet Gopher Access.◊

GDB USER SUPPORT, REGISTRATION

To become a registered user of GDB and OMIM, contact one of the User Support offices listed at right (a user may register to access both Baltimore and a remote node). Questions, problems, or user-registration requests may be sent by telephone, fax, or e-mail. User-registration requests should include name, institutional affiliation, and title (if applicable), street address (no P.O. box numbers), telephone and fax numbers, and e-mail address.

GDB, OMIM Training Schedule

Comprehensive hands-on training courses on the use of GDB and OMIM will have at least one computer workstation for two participants. Registrants will receive at least 3 weeks notice if insufficient registration causes class cancellation.

The general course for scientific users provides a basic understanding of the databases and relationships among different types of data.

Courses are free, but attendees must pay their own travel and lodging expenses. Hotel information and directions will be mailed with registration materials.

As interest in GDB continues to grow, organizations around the world will offer training that requires access to GDB in Baltimore. Notifying GDB User Support about planned training activities will enable the staff to ensure database availability by scheduling maintenance and repairs at other times.

COURSE REGISTRATION INFORMATION

Contact U.S. GDB User Support Office (at top right).

COURSE SCHEDULE

General User Classes will be held in Baltimore on December 6–7 and in 1994 on February 14–15, April 18–19, and June 13–14.

SCHEDULED EXHIBITIONS

ASHG, New Orleans, Oct. 5–9

USER SUPPORT OFFICES

UNITED STATES
GDB User Support
Genome Data Base
Johns Hopkins University
2024 E. Monument Street
Baltimore, MD 21205-2100
410/955-7058, Fax: /614-0434
Internet: *help@gdb.org*

The Help Line is staffed from 9 a.m. to 5 p.m. EST for information on accounts and training courses, technical support, and data questions. Calls received after hours will be forwarded to the appropriate voice mail and returned as soon as possible. To obtain a user's local SprintNet (Telenet) number for locations within the United States: 800/736-1130.

UNITED KINGDOM
Christine Bates
Human Gene Mapping Program Resource Center
CRC, Watford Road
Harrow, Middx HA1 3UJ, U.K.
+ 44/81-869-3446, Fax: -3807
Internet: *cbates@uk.ac.crc*

GERMANY
Otto Ritter
Molecular Biophysics Dept.
German Cancer Research Center
Im Neuenheimer Feld 280
D-6900 Heidelberg, Germany
+ 49/6221-42-2372, Fax: -2333
Internet: *dbk261@cvx12.dkfz-heidelberg.de*

AUSTRALIA
Alex Reisner
ANGIS
Electrical Engineering Bldg. J03
University of Sydney
Sydney, N.S.W. 2006, Australia
+ 61/2-692-2948, Fax: -3847
Internet: *reisner@angis.su.oz.au*

NETHERLANDS
GDB User Support
CAOS/CAMM Center, Faculty of Science
University of Nijmegen
P.O. Box 9010
6500 GL NIJMEGEN, Netherlands
+ 31/80-653391, Fax: -652977
Internet: *post@caos.caos.kun.nl*

SWEDEN
GDB User Support
Biomedical Center, Box 570
S-751 23 Uppsala, Sweden
+ 46/18-174057, Fax: -524869
Internet: *help@gdb.embnet.se*

GDB E-Mail Addresses Change

New GDB e-mail addresses are listed below.

Technical Support

New User Address: *help@gdb.org*
(formerly *help@library.welch.jhu.edu*)

GDB Staff

New Staff Address: *staffname@gdb.org*
(formerly *staffname@library.welch.jhu.edu*)

Welchlab Server, using telnet, rlogin, or rsh

New Hostname: *gdb.org*
(formerly *welchlab.welch.jhu.edu*)

FTP

New Hostname: *ftp.gdb.org*
(formerly *mendel.welch.jhu.edu*)

FTP Command

New Command: *ftp ftp.gdb.org*
(formerly *ftp mendel.welch.jhu.edu*)

GOPHER

New Hostname: *gopher.gdb.org*
(formerly *merlot.welch.jhu.edu*)

WAIS

New Hostname: *wais.gdb.org*
(formerly *welchlab.welch.jhu.edu*)

OMIM

New Hostname: *omim.gdb.org* (OMIM may still be accessed through the GDB address; this is another OMIM access option.)
(formerly *welchlab.welch.jhu.edu*) ◊

GDB USER SUPPORT, REGISTRATION

To become a registered user of GDB and OMIM, contact one of the User Support offices listed at right (a user may register to access both Baltimore and a remote node). Questions, problems, or user-registration requests may be sent by telephone, fax, or e-mail. User-registration requests should include name, institutional affiliation, and title (if applicable), street address (no P.O. box numbers), telephone and fax numbers, and e-mail address.

GDB, OMIM Training Schedule

Comprehensive hands-on training courses on the use of GDB and OMIM will have at least one computer workstation for two participants. Registrants will receive at least 3 weeks notice if insufficient registration causes class cancellation.

The general course for scientific users provides a basic understanding of the databases and relationships among different types of data.

Courses are free, but attendees must pay their own travel and lodging expenses. Hotel information and directions will be mailed with registration materials.

As interest in GDB continues to grow, organizations around the world will offer training that requires access to GDB in Baltimore. Notifying GDB User Support about planned training activities will enable the staff to ensure database availability by scheduling maintenance and repairs at other times.

COURSE REGISTRATION INFORMATION

Contact U.S. GDB User Support Office (at top right).

COURSE SCHEDULE

General User Classes will be held in Baltimore on December 6–7 and in 1994 on February 14–15, April 18–19, and June 13–14.

SCHEDULED EXHIBITIONS

ASHG, New Orleans, Oct. 5–9.

USER SUPPORT OFFICES

UNITED STATES
GDB User Support
Genome Data Base
Johns Hopkins University
2024 E. Monument Street
Baltimore, MD 21205-2100
410/955-7058, Fax: /614-0434
Internet: *help@gdb.org*

The Help Line is staffed from 9 a.m. to 5 p.m. EST for information on accounts and training courses, technical support, and data questions. Calls received after hours will be forwarded to the appropriate voice mail and returned as soon as possible. To obtain a user's local SprintNet (Telenet) number for locations within the United States: 800/736-1130.

UNITED KINGDOM
Christine Bates
Human Gene Mapping Program Resource Center
CRC, Watford Road
Harrow, Middx HA1 3UJ, U.K.
+ 44/81-869-3446, Fax: -3807
Internet: *cbates@uk.ac.crc*

GERMANY
Otto Ritter
Molecular Biophysics Dept.
German Cancer Research Center
Im Neuenheimer Feld 280
D-6900 Heidelberg, Germany
+ 49/6221-42-2372, Fax: -2333
Internet: *dbk261@cvx12.dkfz-heidelberg.de*

AUSTRALIA
Alex Reisner
ANGIS
Electrical Engineering Bldg. J03
University of Sydney
Sydney, N.S.W. 2006, Australia
+ 61/2-692-2948, Fax: -3847
Internet: *reisner@angis.su.oz.au*

NETHERLANDS
GDB User Support
CAOS/CAMM Center, Faculty of Science
University of Nijmegen
P.O. Box 9010
6500 GL NIJMEGEN, Netherlands
+ 31/80-653391, Fax: -652977
Internet: *post@caos.caos.kun.nl*

SWEDEN
GDB User Support
Biomedical Center, Box 570
S-751 23 Uppsala, Sweden
+ 46/18-174057, Fax: -524869
Internet: *help@gdb.embnet.se*

GOPHER NEWS

Obtaining a Gopher Client

Users on the Internet are encouraged to obtain a Gopher client for direct use from the desk top. A Gopher client program runs on the user's local PC, Macintosh, workstation, or mainframe; it uses custom features of the local machine, allows the display of images, and takes advantage of features such as mouse, scroll bars, local printers, and local hard disk. Gopher client programs for Macintosh, DOS, Windows, OS/2, VAX/VMS, NeXTstep, X Windows (including Motif), UNIX, and several other systems are available without charge by anonymous FTP from *boombox.micro.umn.edu* in the *pub/gopher* directory.◊

Gopher Now on GDB Menu

GDB modem or telnet users can now select Gopher from the GDB menu:

5. Additional Services (MLC, Gopher)
3. Internet Gopher Access.◊

Meeting Reports

Cold Spring Harbor Genome Mapping and Sequencing Meeting

The Sixth Genome Mapping and Sequencing Meeting was held May 12–16 at Cold Spring Harbor, New York. Organized by Rick Myers (Stanford University), Bob Waterston (Washington University, St. Louis), and David Porteous [Medical Research Council (MRC), Edinburgh], the conference was attended by about 400 participants. Some 300 posters and platform presentations provided a rich and lively basis for discussion of recent advances and outstanding problems in the field.

Human Disease Genes

The first session, on human disease genes, opened with James Gusella (Massachusetts General Hospital) presenting the recent success of the Huntington's Collaborative Research Group in cloning and identifying the predisposing mutation in this important but hitherto elusive gene. The mutation is identified as a triplet expansion in a gene whose abnormal DNA sequence sheds little light on its function.

Model Organisms

Although most abstracts concerned human genome mapping, the session devoted to model organisms was one of the highlights. Gerald Rubin (University of California, Berkeley) reminded attendees of the role played by *Drosophila melanogaster* geneticists in laying the foundations for the Human Genome Project. The ability to correlate *Drosophila's* very high resolution cytogenetic map with a vigorously developed genetic mutational map illustrates the power of this model system.

Last year, attendees heard how chromosome 3 of *Saccharomyces cerevisiae* had been sequenced by the European Yeast Genome Collaborative Project; now, chromosomes 2 and 9 are close to completion.

Greg Elgar (Cambridge University, U.K.) introduced the project initiated by Sydney Brenner (Scripps Research Institute) to map and sequence the genome of the puffer fish or Fugu. This fish is distinguished by having the most compact vertebrate genome (~400 Mb), with a coding complexity very similar to that of mammalian DNA. Its sequence similarity to mammalian gene homologues is high, and its exon structure is highly conserved. With considerably smaller introns, this organism yields much more coding information from genomic sequencing than others do. As yet, little genetic data has been published for Fugu.

The *Caenhorabditis elegans* project, led jointly by Waterson and John Sulston [Laboratory of Molecular Biology (LMB), Cambridge, U.K.], continues to set the pace with 2 Mb of finished sequence, half of which is contiguous, at an average gene density of 1 per 4.5 kb. Sequencing of an additional 3 Mb of the estimated 100-Mb genome is expected this year. Finishing the sequence of this 16,000-gene organism, with its elegant genetics and completed cell-fate map to track each cell's movement during development, is a major near-term goal.

Comparative studies of mouse and human molecular composition, genetics, development, and pathophysiology is of powerful and mutual benefit to both communities. Steve Brown (St. Mary's Hospital Medical School, U.K.) of the European Collaborative Interspecific Backcross project illustrated this point by explaining how the 500-kb-resolution mouse genetic map can complement and significantly improve the human meiotic map constructed by typing microsatellite markers in the Centre d'Etude du Polymorphisme Humain (CEPH) pedigrees [Jean Weissenbach (Institut Pasteur); James Weber (Marshfield Medical Research Foundation)].

Sequencing Technologies

While shotgun sequencing—the generally favored large-scale sequencing method—is becoming more efficient, other technologies are also being explored. Possible improvements include the use of inexpensive walking primers assembled from hexamers or pentamers [Levi Ulanovsky (Weizmann Institute of Science, Rehovot, Israel)] or selected from small libraries of nonamers [Jerry Slightom (Upjohn Company)]. Use of genomic coupled amplification and sequencing eliminates the need for cloning and prior amplification by polymerase chain reaction and may be valuable for studying sequence variation [Gualberto Ruano (BIOS Laboratories)]. However, modifications of conventional protocols will probably not provide routine 1-kb reads in the near future. Andre Rosenthal (LMB) suggested that investigators need to move away from gel casting and consider such alternatives as the use of fluorescently

Use of Genomic Coupled Amplification and Sequencing Eliminates Need for Cloning

This newsletter is prepared at the request of the DOE Office of Health and Environmental Research and the NIH National Center for Human Genome Research by the Biomedical and Environmental Information Analysis Section of the Health Sciences Research Division at Oak Ridge National Laboratory, which is managed by Martin Marietta Energy Systems, Inc., for the U.S. Department of Energy, under Contract DE-AC05-84OR21400.0

labeled nucleotides to sequence by addition, removal, and extension of primed template DNA.

Informatics

The informatics session addressed challenges and recent successes in handling and interpreting the vast wealth of new DNA sequence information and the complex data being generated by meiotic mapping and contig-assembly projects. Ingenuity and enthusiasm within the informatics community has made possible the accommodation of chimeric clones and assembly of coherent maps [Steven Lincoln, Massachusetts Institute of Technology (MIT); Bruno Lacroix, CEPH]. Remarkably accurate gene identification is now achievable from raw sequence data [Gary Stormo, University of Colorado; David Searls, University of Pennsylvania; Phil Green, Washington University]. Existing databases may already contain representatives of most ancient conserved regions of protein sequence (Green).

Mapping

Sessions on cDNA and other mapping methods illustrated the need and potential for further technical innovations, including (1) microdissection and amplification of single cuts from metaphase chromosomes for use as painting probes [Jeff Trent (University of Michigan, Ann Arbor)]; (2) targeting of human dispersed repeats by homologous recombination for chromosome tagging and manipulation in somatic cells [Viv Watson (MRC, Edinburgh)]; (3) direct cDNA selection by genomic DNA carried in yeast artificial chromosomes (YACs) or cosmids [Mike Lovett (University of Texas Southwestern Medical Center, Dallas)]; and (4) construction of a truly normalized gene library by selective cloning of CpG islands using methyl-CpG binding protein [Sally Cross (University of Edinburgh)]. The presence of chimerism in CEPH mega-YAC libraries has been criticized, but users repeatedly defended the enormous value of the libraries for building maps and bridging gaps. Eric Lander (MIT) conveyed preliminary but very encouraging data suggesting a minimal level of chimerism in mega-YACs cloned in a new recombination-defective yeast strain. Nevertheless, cloning large (up to 350 kb) DNA fragments in bacteria is a promising alternative because of the simple insert purification offered by circular recombinant molecules [Hori Shizuya (California Institute of Technology)].

The closing session on large maps focused on efforts by human genome center investigators, who reported that large-insert clone physical maps of several chromosomes are either complete or nearing completion. Recurrent themes were the need for (1) multiply redundant cosmid and YAC libraries and (2) considerable duplication of effort—particularly because of significant levels of chimerism. Glen Evans (Salk Institute) described genomic sequencing strategies based on obtaining very dense cosmid coverage of chromosome 11 and exploitation of this coverage. His plans include using only T3 and T7 primers and end sequencing ~500 bp of every ordered cosmid for rapid construction of a frame or outline for the whole chromosome. He believes the resulting map of sequence tagged sites will provide a solid platform for biological investigations.

Participants expressed a growing sense that established methods and reagents are capable of yielding high-resolution clone maps of the human genome, chromosome by chromosome. However, some serious problems remain in making a global map, as illustrated by the computational difficulties in assembling coherent contig maps from the fingerprint analysis of the CEPH/Genethon mega-YAC library [Ilya Chumanov (CEPH)]. In collaboration with CEPH/Genethon, Lander plans to acquire and type microsatellite repeats on an industrial scale, but most of the meiotic events represented by the pedigrees are identifiable. Establishing relative order and distance with meiotic intervals requires a different approach. Integration with the physical map will depend heavily on overcoming the chimerism problem, either computationally or by constructing additional YAC libraries in recombination-deficient hosts.

David Cox (Stanford University) presented his latest mapping results using whole-genome radiation hybrids. The radiation map has the dual advantages of significantly increased resolution (now ~500 kb) and portability (~100 independent hybrids harboring a total of ~7000 independent and essentially random breaks) while still showing complete congruence with the meiotic map. His work and that of many other participants emphasized the importance of continued efforts in conceptual and experimental innovation and integration. *[David Porteous, MRC, Edinburgh; Bob Waterston, Washington University; and Rick Myers, Stanford University]* ◊

Gene Identification Now Achievable from Raw Sequence

labeled nucleotides to sequence by addition, removal, and extension of primed template DNA.

Informatics

The informatics session addressed challenges and recent successes in handling and interpreting the vast wealth of new DNA sequence information and the complex data being generated by meiotic mapping and contig-assembly projects. Ingenuity and enthusiasm within the informatics community has made possible the accommodation of chimeric clones and assembly of coherent maps [Steven Lincoln, Massachusetts Institute of Technology (MIT); Bruno Lacroix, CEPH]. Remarkably accurate gene identification is now achievable from raw sequence data [Gary Stormo, University of Colorado; David Searls, University of Pennsylvania; Phil Green, Washington University]. Existing databases may already contain representatives of most ancient conserved regions of protein sequence (Green).

Mapping

Sessions on cDNA and other mapping methods illustrated the need and potential for further technical innovations, including (1) microdissection and amplification of single cuts from metaphase chromosomes for use as painting probes [Jeff Trent (University of Michigan, Ann Arbor)]; (2) targeting of human dispersed repeats by homologous recombination for chromosome tagging and manipulation in somatic cells [Viv Watson (MRC, Edinburgh)]; (3) direct cDNA selection by genomic DNA carried in yeast artificial chromosomes (YACs) or cosmids [Mike Lovett (University of Texas Southwestern Medical Center, Dallas)]; and (4) construction of a truly normalized gene library by selective cloning of CpG islands using methyl-CpG binding protein [Sally Cross (University of Edinburgh)]. The presence of chimerism in CEPH mega-YAC libraries has been criticized, but users repeatedly defended the enormous value of the libraries for building maps and bridging gaps. Eric Lander (MIT) conveyed preliminary but very encouraging data suggesting a minimal level of chimerism in mega-YACs cloned in a new recombination-defective yeast strain. Nevertheless, cloning large (up to 350 kb) DNA fragments in bacteria is a promising alternative because of the simple insert purification offered by circular recombinant molecules [Hori Shizuya (California Institute of Technology)].

The closing session on large maps focused on efforts by human genome center investigators, who reported that large-insert clone physical maps of several chromosomes are either complete or nearing completion. Recurrent themes were the need for (1) multiply redundant cosmid and YAC libraries and (2) considerable duplication of effort—particularly because of significant levels of chimerism. Glen Evans (Salk Institute) described genomic sequencing strategies based on obtaining very dense cosmid coverage of chromosome 11 and exploitation of this coverage. His plans include using only T3 and T7 primers and end sequencing ~500 bp of every ordered cosmid for rapid construction of a frame or outline for the whole chromosome. He believes the resulting map of sequence tagged sites will provide a solid platform for biological investigations.

Participants expressed a growing sense that established methods and reagents are capable of yielding high-resolution clone maps of the human genome, chromosome by chromosome. However, some serious problems remain in making a global map, as illustrated by the computational difficulties in assembling coherent contig maps from the fingerprint analysis of the CEPH/Genethon mega-YAC library [Ilya Chumanov (CEPH)]. In collaboration with CEPH/Genethon, Lander plans to acquire and type microsatellite repeats on an industrial scale, but most of the meiotic events represented by the pedigrees are identifiable. Establishing relative order and distance with meiotic intervals requires a different approach. Integration with the physical map will depend heavily on overcoming the chimerism problem, either computationally or by constructing additional YAC libraries in recombination-deficient hosts.

David Cox (Stanford University) presented his latest mapping results using whole-genome radiation hybrids. The radiation map has the dual advantages of significantly increased resolution (now ~500 kb) and portability (~100 independent hybrids harboring a total of ~7000 independent and essentially random breaks) while still showing complete congruence with the meiotic map. His work and that of many other participants emphasized the importance of continued efforts in conceptual and experimental innovation and integration. *[David Porteous, MRC, Edinburgh; Bob Waterston, Washington University; and Rick Myers, Stanford University]* ◊

Gene Identification Now Achievable from Raw Sequence

Meeting Reports

Human Genome news

National Center for Human Genome Research

This newsletter is intended to facilitate communication among genome researchers and to inform persons interested in genome research. Suggestions are invited.

Human Genome Management Information System:

Managing Editor
Betty K. Mansfield

Editors/Writers
Anne E. Adamson
Denise K. Casey
Kathleen H. Mavournin

Production Manager/Editor
Judy M. Wyrick

Production Assistants
K. Alicia Davidson
Larry W. Davis
Sheryl A. Martin
Laura N. Yust

Correspondence Address:
Betty K. Mansfield
ORNL
P.O. Box 2008
Oak Ridge,TN 37831-6050
615/576-6669
Fax:/574-9888
%3BITNET:
Internet: *bkq@ornl.gov*

Sponsor Contacts:

Daniel W. Drell
DOE Program Office
Germantown, MD 20545
301/903-6488, Fax: -8521
Internet: *daniel.drell%er@mailgw.er.doe.gov*

Leslie Fink
NIH National Center for Human Genome Research
Bethesda, MD 20892
301/402-0911, Fax: -4570
Internet: *lsf@cu.nih.gov*

First International Conference on Intelligent Systems for Molecular Biology

The First International Conference on Intelligent Systems for Molecular Biology was held July 6–9 at the Lister Hill Center of the National Library of Medicine (NLM) in Bethesda, Maryland. Over 200 biologists and computer scientists from 13 countries gathered to consider applications of artificial intelligence (AI) and related technologies to challenges in molecular biology, primarily sequence analysis.

The meeting was organized by Lawrence Hunter (NLM), David Searls (University of Pennsylvania), and Jude Shavlik (University of Wisconsin, Madison). Support was provided by NLM, DOE, the Biomatrix Society, and the American Association for Artificial Intelligence (AAAI).

The conference was preceded by a series of tutorials, including introductions to AI for biologists and to molecular biology for computer scientists. Other tutorials covered genetic algorithms, neural networks, and linguistic methods for sequence analysis. The meeting was opened each day by invited speakers: Temple Smith (Boston University) on classification of protein structure "cores," Leroy Hood (University of Washington) on sequencing technology and recent results in immune-system–gene regions, and Harold Morowitz (George Mason University) on a new theory of evolutionary development of intermediary metabolism. Some 25 posters and 27 talks, accepted after rigorous review, are each represented by a paper in the published proceedings (see box, above right).

The meeting focused on three major areas:

- Predicting protein secondary structure and classifying or clustering tertiary substructures into families. The repertoire of machine-learning and probabilistic techniques being applied to these problems is expanding dramatically; papers were presented on constructive induction, probabilistic networks, case-based reasoning, hidden Markov models, megaclassification techniques, and neural networks.
- Nucleic acid sequence analysis at a wide variety of levels, from base reading in sequencing gels to map integration. Novel approaches were presented to problems of sequence assembly and restriction-site mapping, gene identification, interpretation of DNA crystallographic data, knowledge discovery in sequence databases, and RNA structure prediction.
- A variety of other AI techniques such as constraint reasoning and qualitative modeling, biochemical applications including simulation of metabolic pathways, the study of gene regulation, and automated analysis of biological literature.

Conference proceedings ($45 plus shipping, ISBN 0-929280-47-4) are available from AAAI Press of Menlo Park, California (415/328-3123).

Planning is under way for the second conference, to be held in Seattle at about the same time next year. To receive advance information on this meeting, send a message to *ismb@nlm.nih.gov*. *[David Searls (University of Pennsylvania)]* ◊

University of Iowa Program in Biomedical Ethics Fellowships

The University of Iowa Program in Biomedical Ethics invites applications for its visiting fellowships in molecular and clinical genetics. This project is part of the ethical, legal, and social implications (ELSI) core of the Cooperative Human Linkage Center, one of ten genome centers funded by the National Center for Human Genome Research. The fellowship program is intended for philosophers, historians, attorneys, journalists, nurses, and other professionals who are not biological scientists but have demonstrated a strong interest in the ELSI aspects of human genetics.

Working in laboratory and clinical settings, fellows will analyze implications of genetic research and observe the increasing connections among molecular genetics, clinical genetics, and ELSI-related issues. The fellowship program will begin in fall 1994 (for 1994–95) and continue in fall 1995 (for 1995–96). A maximum of three fellows each year will receive a monthly stipend of $3500 for a period of 2 to 4 months.

Application deadline: January 1, 1994. Inquiries and requests for applications: Jay Horton; Program in Biomedical Ethics; University of Iowa College of Medicine, 112 MEB; Iowa City, IA 52242 (319/335-6706).◊

☛ Resources Available

HUGO Mouse Resources Database

The Human Genome Organization (HUGO) Mouse Resources Database has been mailed in hard copy to all HUGO members. The database, which was compiled by Mouse Committee Chairman Steve Brown, is designed to (1) provide global information on mouse genome mapping resources and projects not generally available through established outlets and (2) foster many new points of contact between human and mouse genetics communities. The database will be updated and circulated every 6 months.

The database is compiled on Hypercard, provided with most Macintosh computers. Information from each laboratory is contained on an individual card carrying a number of fields that cover different resource areas. Hypercard allows users to carry out a variety of functions, including searching for specific strings; printing available fields or subsets of fields in the chosen format; and even modifying the database to suit individual needs.

To receive a copy of the database on disk, send a letter or fax with all contact information to HUGO Europe; One Park Square West; London NW1 4LJ (+ 44/71-935-8085, Fax: -8341). To make corrections, additions, or new entries, contact Brown at the Department of Biochemistry and Molecular Genetics; St. Mary's Hospital Medical School; London W2 1PG; U.K. (+ 44/71-723-1252, ext. 5484; Fax: -706-3272; Internet: *s.brown@sm.ic.ac.uk*).◊

Genetic Mapping Directories, Files

The Cooperative Human Linkage Center (CHLC), directed by Jeffrey Murray, was established by the National Center for Human Genome Research in the fall of 1992 to develop high-heterozygosity genetic maps. CHLC has projects located at the University of Iowa, Fox Chase Cancer Center, Marshfield Medical Research Foundation, and Harvard Medical School.

The following directories and files are available from CHLC via anonymous FTP to *ftp.chlc.org* and through a CHLC Gopher Server addressed *gopher.chlc.org*.

- ***chlc/newsletters:*** CHLC newsletters in plain text and postscript. [Volume 1(1) of *CHLC Report* was published in May. To subscribe to the hard copy version of the newsletter, contact CHLC Administration; #440 EMRB; University of Iowa; Iowa City, IA 52242.]
- ***chlc/genotypes/tables:*** Tabular descriptions of marker systems in chromosome-specific data sets.
- ***chlc/genotypes/typing:*** Chromosome-specific genotype sets in CRIMAP file format.
- ***chlc/maps/framework:*** Framework maps of all markers currently mapped by CHLC (including markers from other sources).
- ***chlc/maps/skeletal:*** Maps generated using stringent map-building algorithms.
- ***chlc/markers/chlc:*** CHLC-produced marker data.
- ***chlc/markers/marshfield:*** Marker data produced by Marshfield Medical Research Foundation.

Useful e-mail addresses include the following.

- ***help@chlc.org:*** Answers from CHLC staff to questions about CHLC services.
- ***info-server@chlc.org:*** Descriptive information about the CHLC project and data.
- ***linkage-server@chlc.org:*** Server that will genetically map submitted markers using CHLC data sets.

Note: All CHLC postings will be presented via an appropriate BIOSCI newsgroup (currently BIOSCI/GENETIC-LINKAGE). Users with USENET news can access the newsgroup with this address: *bionet.molbio.gene-linkage*. Requests for adding or canceling e-mail subscriptions to BIOSCI should be sent to *biosci@net.bio.net* (Americas and Pacific Rim) or *biosci@daresbury.ac.uk* (Europe, Africa, and Central Asia).◊

Biotechnology Pharmaceuticals

Sales of pharmaceuticals derived from biotechnology research and development are expected to rise from $2.38 billion in 1992 to $9.2 billion by 2000, according to *Biotechnology in the U.S. Pharmaceutical Industry*, a new report from the Institute for Biotechnology Information. Institute Director Mark Dibner said, "There are two notable findings from our analysis. First, the increase is flat at 10 to 26% growth per year, not rising exponentially. Second, sales of these compounds will rise from 4.6% of U.S. drug and vaccine sales to just over 13% by 2000."

The special 400-page report contains details on 23 marketed drugs (for 31 indications) and 305 drugs and vaccines in clinical trial. This information was derived from numerous sources, including sales projections, market forecasts, estimated times of approval, and estimates of market penetration. Activities and affiliations are described for 71 sites of 41 pharmaceutical companies having biotechnology R&D programs and for biotechnology firms with products on the market. Specific markets and sales are identified. [$595, 1993. Institute for Biotechnology Information; P.O. Box 13547; Research Triangle Park, NC 27709-3547 (919/549-8880, Fax: /990-9521).]◊

Need more information on topics related to the Human Genome Project?

Contact *Human Genome News* staff at 615/576-6669, Fax: /574-9888.

Calendar of Genome-Related Events* (acronyms, p. 16)

September

28. †Francis Collins: Id. of Human Disease Genes by Positional Cloning; Bethesda, MD [NCHGR Lecture Series, E. Feingold, 301/496-7531, Fax: /480-2770]

October

13–15. *Natl. SBIR Conf.; Washington, DC [Hotline, 407/791-0720, Fax: -0098]

13–16. DNA: The Double Helix. Forty Years: Perspective and Prospective; Chicago (poster deadline: July 2) [NYAS, 212/838-0230, Fax: -5640]

21. †Glen Evans: Anal. of Human Chromosomes—Physical Mapping and Beyond; Bethesda, MD [see contact: Sept. 28]

21–22. Law and Science at the Crossroads: Biomedical Tech., Ethics, Public Policy, and the Law; Boston [SULS, 617/573-8627, Fax: /248-0648]

23–27. Genome Sequencing and Anal. Conf. V; Hilton Head, SC [S. Wallace, 301/216-9567, Fax: /977-7233]

28–30. ‡2nd Intl. Workshop on SBH; The Woodlands, TX [K. Beattie, 713/363-7947, Fax: -7931]

November

5–6. First Genetic Marker—Blood Group Res., "Race," and Disease: 1900–1950; Indianapolis [W. Schneider, 317/274-3811, Fax: -2347]

7–10. Electrophoresis '93; Charleston, SC (paper deadline: June 1) [J. Cunningham, 301/898-3772, Fax: -5596]

7–11. MGC 93; Hamamatsu, Japan [K. Moriwaki, + 81/559-75-0771, Fax: -6240]

7–11. Molecular Approaches to Cancer Immunotherapy Conf.; Asheville, NC (reg. deadline: Aug. 2) [AACR, 215/440-9300, Fax: -9313]

9–13. Interactions of Cancer Susceptibility Genes and Environmental Carcinogens; Lyon, France [AACR/IARC Conf., 215/440-9300, Fax: -9313 or + 33-7/273-8485, Fax: -8575]

10–12. *CCM 93; Tsukuba, Japan [N. Shimizu, + 81/333-53-1211, ext. 2721, Fax: -51-2370]

14–17. HGM 93; Kobe, Japan [HGM Secretariat, + 81/6-454-4811, Fax: -4711]

15–17. *Natl. SBIR Conf.; Seattle [see contact: Oct. 13–15]

15–19. Nanometer Scale Biotech.: DNA Nanoconstructions at the 1993 AVS Natl. Symposium; Orlando, FL [D. Manos, 804/221-3525, Fax: -3540]

17–19. Engineered Animal Models; San Francisco [CHI, 617/487-7989, Fax: -7937]

18. †Gerald Rubin: *Drosophila* Genome Center—A Progress Rpt.; Bethesda, MD [see contact: Sept. 28]

18–20. 1993 San Diego Conf.: Beyond DNA Probes; San Diego (abs. deadline: Sept. 1) [AACC/SDC, 800/892-1400, Fax: 202/887-5093]

20–21. Healthcare in Flux: How Will Families with Special Needs Fit In? Alexandria, VA [J. Weiss, 800/336-GENE]

December

2–3. Advances in Capillary Electrophoresis; CHI, Orlando, FL [see contact: Nov. 17–19]

8–9. Mathematical and Statistical Aspects of DNA and Protein Sequence Anal.; London [Sci. Mtgs. Sec., + 44/71-839-5561, ext. 278, Fax: -930-2170]

9–11. *‡Adjudication of Genetic Testing Evidence; Washington, DC [F. Zweig, 202/296-6922, Fax: /785-0114]

13–15. Genome Informatics Workshop IV; Yokohama, Japan [HGC, Fax: + 81/3-3440-6173, Internet: *workshop@ims.u-tokyo.ac.jp*]

16. †Nicholas Dracopoli: Genome Anal. of Melanoma; Bethesda, MD [see contact: Sept. 28]

January 1994

20. †Ellen Wright-Clayton: All in the Family—Special Problems Posed by Genetic Res. in Human Subjects; Bethesda, MD [see contact: Sept. 28]

24–25. *Natl. Advisory Council for Human Genome Res.; Bethesda, MD [J. Ades, 301/402-2205, Fax: -2218]

24–27. BioEast '94; Washington, DC [Bioconferences Intl., 301/652-3072, Fax: -4951]

24–27. Plant Genome II; San Diego [Scherago Intl., 212/643-1750, Fax: -1758]

31–February 5. Molecular Genetics of Progression and Metastasis; AACR, Big Sky, MT [see contact: Nov. 7–11]

February 1994

5–10. Advances in Gene Tech.: Molecular Biology and Human Genetic Disease at the 1994 Miami Bio/Technology Winter Symposia; Fort Lauderdale, FL [S. Black, 305/547-3597, Fax: /324-5665]

7–11. Intl. Conf. on Comparative Gene Mapping in Terrestrial and Aquatic Vertebrates; Oslo [H. Lewin, 217/333-5998, Fax: /244-5617 or O. Lie, + 47/22-96-47-82, Fax: -86]

17. †Neil Risch: Susceptibility Gene Localization and Exclusion for Complex Diseases; Bethesda, MD [see contact: Sept. 28]

28. Molecular Bioinformatics IEE 94 Colloquium; London (paper deadline: Oct. 31) [S. Schulze-Kremer,+ 49/30-463-3040, Fax: -464-4097, Internet: *steffen@kristall.chemie.fu-berlin.de*]

28–Mar. 2. Human Genome Project: Commercial Implications; CHI, San Francisco [see contact: Nov. 17–19]

March 1994

3–4. Gene Transcription-Based Therapeutics; CHI, San Francisco [see contact: Nov. 17–19]

17. †Jeffrey Trent: Genomic Applications of Chromosome Microdissection; Bethesda, MD [see contact: Sept. 28]

April 1994

21. †Ronald Davis: Technol. Development for High Throughput DNA Sequencing; Bethesda, MD [see contact: Sept. 28]

August 1994

28–Sept. 1. 10th World Congress on Medical Law; Jerusalem [A. Carmi, + 97/2-3-751-6422, Fax: -6635] ◊

Training Calendar**

October

2–4. Id. of Transcribed Sequences Workshop; New Orleans (abs. deadline: July 1) [K. Gardiner, 303/333-4515, Fax: -8423]

4–5. Future Technologies for DNA Anal.; Rockville, MD [AFIP, 301/427-5231, Fax: -5001]

4–8. Recombinant DNA Technologies I; LTI, Germantown, MD [L. Kerwin, 800/952-9166, Fax: 301/258-8212]

4–8. RNA Isolation and Characterization; Columbia, MD [Exon-Intron, Inc., 410/730-3984, Fax: -3983]

5–8. PCR Techniques; LTI, Burlington, VT [see contact: Oct. 4–8]

8–11. Recombinase Mediated Genome Reorganization; ASBMB, Keystone, CO [G. Goodenough, 301/530-7010, Fax: -7014]

11–12. Ancient Human DNA Anal.; AFIP, Alexandria, VA [see contact: Oct. 4–5]

11–14. PCR Techniques; Lake Tahoe, NV [CATCMB/CUA, 202/319-6161, Fax: -4467]

11–14. Recombinant DNA Methodology; CATCMB/CUA, Lake Tahoe, NV [see contact: Oct. 11–14, above]

➡

*Attendance at meetings listed with asterisk is either limited or restricted. Dates may change; check with contact person.
**Dates and course status may change, and courses may also be offered at other times and places; check with contact person.
†NCHGR-funded event.
‡DOE-funded event.

11–15. Essentials in Forensic Pathology; AFIP, Rockville, MD [see contact: Oct. 4–5]

11–16. Recombinant DNA Technologies II; LTI, Germantown, MD [see contact: Oct. 4–8]

12–15. DNA Fingerprinting; Rockville, MD [ATCC, 301/231-5566, Fax: /770-1805]

13–26. †Advanced In Situ Hybridization & Immunocytochemistry; Cold Spring Harbor, NY (application deadline: July 15) [CSHL, 516/367-8346, Fax: -8845]

14–27. †Anal. & Genetic Manipulation of YACs; CSHL (application deadline: July 15) [see contact: Oct. 13–26]

18–21. PCR Methodology; Exon-Intron, Inc., Columbia, MD [see contact: Oct. 4–8]

22–25. Epigenetic Modifications of the Genome; ASBMB, Keystone, CO [see contact: Oct. 8–11]

24–25. 4th Keck Symposium on Computational Biol.; Pittsburgh (reg. deadline: Sept. 7) [L. Jarzynka, 412/624-6978, Internet: *jarzynka@cs.pitt.edu*]

25–29. Recombinant DNA: Techniques & Applications; ATCC, Rockville, MD [see contact: Oct. 12–15]

November

2–5. PCR Applications; ATCC, Rockville, MD [see contact: Oct. 12–15]

2–15. Molecular Genetics, Cell Biol. & Cell Cycle of Fission Yeast; CSHL [see contact: Oct. 13–26]

3–12. †Essential Computational Genomics for Biologists; CSHL [see contact: Oct. 13–26]

4–5. Intro. to Molecular Cytogenetics; Gaithersburg, MD [Oncor, Inc., 800/776-6267, Fax: 301/926-6129]

11–12. *Advanced Course in Project Planning Strategies for Drug, Biologic, Biotech Development; Washington, D.C. [BioConferences Intl., 301/652-3072, Fax: -4951]

11–14. *RLGS Method Training Course; Tsukuba, Japan (application deadline: July 31) [Y. Hayashizaki, + 81/298-36-9145, Fax: -9098]

15–19. In Situ Hybridization Techniques; LTI, Germantown, MD [see contact: Oct. 4–8]

15–19. Recombinant DNA Methodology; Exon-Intron, Inc., Columbia, MD [see contact: Oct. 4–8]

18–19. Tissue In Situ Hybridization; Oncor, Inc., Gaithersburg, MD [see contact: Nov. 4–5]

29–Dec. 3. In Situ Hybridization & rDNA Technology; Exon-Intron, Inc., Columbia, MD [see contact: Oct. 4–8]

December

6–7. GDB/OMIM Training Courses [see schedule, p. 9]

13. Intro. to PCR; BTP, Gainesville, FL, and Houston [S. Chance, 800/821-4861, Fax: 603/659-4708]

14–15. Quantitative RNA-PCR; BTP, Gainesville, FL [see contact: Dec. 13]

16–17. DNA Sequencing Without Radioactivity; BTP, Gainesville, FL [see contact: Dec. 13]

20–21. Clin. Applications of PCR; BTP, Gainesville, FL [see contact: Dec. 13]

For Your Information

U.S. Genome Research Funding Guidelines

Note: Investigators wishing to apply for NIH and DOE funding are urged to discuss their projects with agency staff before submitting proposals.

NIH National Center for Human Genome Research (NCHGR)

Application receipt dates:

- R01, P01, R21, R29, P30, P50, K01,* and R13 grants – February 1, June 1, and October 1.
- Individual postdoctoral fellowships and institutional training grants – January 10, May 10, and September 10.
- Small Business Innovation Research Grants (SBIR: firms with 500 or fewer employees) – April 15, August 15, and December 15.
- Research supplements for under-represented minorities – applications are accepted on a continuing basis.
- Requests for Applications (RFAs) – receipt dates are independent of the above dates. Notices will appear in *HGN* and other publications.

*Expedited review possible. Check with NCHGR during application development phases.

Program announcements are listed in the weekly *NIH Guide for Grants and Contracts*,* which is available through

- Hard-copy subscription: call 301/496-7441.
- Electronic version (E-Guide): Access through one of the following methods.
  1. Institutional Hubs. A designee receives automatic updates and distributes them locally to researchers. To use this NIH-preferred method, send a message naming the responsible person to Rebecca Duvall (BITNET: *q2c@nihcu*, Internet: *q2c@cu.nih.gov*).
  2. NIH Grant Line (also known as DRGLINE). User reads electronic bulletin board for weekly updates. Connection is through a modem, and files can be transmitted rapidly via BITNET or Internet. For more information, contact John James (301/496-7554 or BITNET: *zns@nihcu*).

*Full text of RFAs listed in the NIH grants guide may be obtained from either of the two electronic sources or from NIH NCHGR in Bethesda, Maryland (301/496-0844).

DOE Human Genome Program

Solicitations for proposals are announced in the *Federal Register*, *Science*, and other publications. Proposals for FY 1994 were due July 15.

For funding information or general inquiries, contact the program office via

- 301/903-6488, Fax: -8521; or Internet: *#genome%er@mailgw.er.doe.gov* or *genome@oerv01.er.doe.gov*

1994 Human Genome Distinguished Postdoctoral Fellowships

Deadline: February 1, 1994. For further information, contact

- Linda Holmes, Oak Ridge Institute for Science and Education (615/576-4805).

SBIR Grants (Also see article, p. 6.)

DOE and NIH invite small business firms to submit grant applications addressing the human genome topic of SBIR programs, which are designed to strengthen innovative firms in research and development and contribute to the growth and strength of the nation's economy. For more information on human genome SBIR grants, contact

- Kay Etzler; c/o SBIR Program Manager, ER-16; DOE; Washington, DC 20585 (301/903-5867, Fax: -5488).
- Bettie Graham; Bldg. 38A, Rm. 610; NIH; 9000 Rockville Pike; Bethesda, MD 20892 (301/496-7531, Fax: /480-2770).

National SBIR conferences: Washington, DC (October 13–15,); Seattle, WA (November 15–17); Houston, TX (April 26–28, 1994). Conference Hotline: 407/791-0720.◊

January 1994

15–22. Molecular Biology of Human Genetic Disease; Copper Mountain, CO (abs. deadline: Sept. 1) [Keystone Symposia, 303/262-1230, Fax: -1525]

15–22. Gene Therapy; Copper Mountain, CO (abs. deadline: Sept. 1) [see contact: Jan. 15–22]

21–28. Transposition & Site-Specific Recombination: Mechanism & Biology; Park City, UT (abs. deadline: Sept. 22) [see contact: Jan. 15–22] ◊

Human Genome Management Information System Subscription/Document Request* (Vol. 5, No. 3)

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 _____

1. ___ *Human Genome News* ___New Subscriber ___Change of Name/Affiliation/Address (circle all that apply) ___Drop Subscription
2. ___ *Understanding Our Genetic Inheritance, The U.S. Human Genome Project: The First Five Years, FY 1991–1995* (Joint DOE-NIH 5-Year Plan)
3. ___ DOE *Human Genome 1991–92 Program Report* (includes "Primer on Molecular Genetics") ___*Primer* as a separate document
4. ___ *Meeting Report: DOE Informatics Summit*—DRAFT (April 26–27, 1993, Baltimore, Maryland)

*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.

CALENDAR ACRONYMS

AACC Am. Assoc. for Clin. Chem.
AACR Am. Assoc. for Cancer Res.
AFIP Armed Forces Inst. of Path.
ASBMB Am. Soc. for Biochem. & Mol. Biol.
ATCC Am. *Type Culture* Coll.
AVS Am. Vacuum Soc.
BTP Biotech. Training Programs
CATCMB/CUA Ctr. for Advanced Training in Cell and Mol. Biol./Catholic Univ. of Am.
CCM Chromosome Coordinating Meeting
CHI Cambridge Healthtech Inst.
CSHL Cold Spring Harbor Lab.
GDB/OMIM Genome Data Base/Online *Mendelian Inheritance in Man*
HGC Human Genome Center
HGM Human Genome Mapping
IARC Int. Agcy. for Res. on Cancer
IEE Inst. of Electrical Engineers
LTI Life Technologies, Inc.
MGC Mouse Genome Conf.
NCHGR Natl. Ctr. for Human Genome Res.
NYAS New York Acad. of Sci.
RLGS Restriction Landmark Genomic Scanning
SBH Sequencing by Hybridization
SBIR Small Business Innovation Res.
SDC San Diego Conf.
SULS Suffolk Univ. Law School

OAK RIDGE NATIONAL LABORATORY • MANAGED BY MARTIN MARIETTA ENERGY SYSTEMS, INC. • FOR THE U.S. DEPARTMENT OF ENERGY

HGMIS MAILING ADDRESS
Betty K. Mansfield
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831-6050
U.S.A.
Address Correction Requested
Return Postage Guaranteed

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

☆U. S. GOVERNMENT PRINTING OFFICE: 1992-750-229/80001