DOE/ER/61230--1

DE92 040626

PROGRESS REPORT OF THE GDB FOR YEAR 1

The report covers the period from Sept 1, 1991, the commencement of joint DOE and NIH funding, up to the present. This was the third year of operation of GDB in Baltimore, the previous ones having been funded by the Howard Hughes Medical Research Institute (HHMI), and was generally one of consolidation and extension of the policies initiated during the first two years. Only some of the application and policy highlights are discussed. However, a complete outline listing of the year's activities is provided in attachment 1.

SOFTWARE DEVELOPMENT: In mid August 1991, GDB was used as the central database for collating the information at HGM11. In a joint effort with the ICRF, simultaneous online editing facilities were provided for up to 120 editors on a local area network. By comparison to the performance of GDB one year earlier at HGM10.5 under similar circumstances, the general impression amongst editors was one of "robustness". Accordingly, with the exception of the map manager, nearly all of the development carried out during the past year has been directed towards developing an application programmable interface, locally called the "DAT". Currently, the database comprises approximately 370 relational tables and 80 dictionaries. This high level of complexity is difficult to track and makes development and maintenance of front-end software extremely time consuming. The "DAT" provides a stable intermediate layer between the data structures and front-end software. It eliminates the need for the front-end software from having to directly interpret the complexities of the data structures and minimizes having to modify the software every time changes are made to the schemas. The "DAT" will permit new interfaces to be developed much more rapidly than in the past and in particular for other groups to develop their own interfaces without having to get involved in interpreting the relational schemas. The "DAT" will be used for the first time in the next production version of the database to be released this summer. Further, this version will embody X-windows and open up the use of the mouse to a much wider range of users than at present.

USER AND DATA STATISTICS: GDB has approx. 4500 registered users in Baltimore, 700 at Northwick Park in London and 80 at the DKFZ in Heidelberg. The number of registered users is increasing by about 150 per month. Over the last year there have been about 2400 logons per month in Baltimore with an average connect time of 50 minutes. Of the 4500 registered users, 1300 are active and 300 hyperactive with several logons per week. The number of mapped genes and anonymous DNA segments has increased by 1189, the probes by 5850, the polymorphisms by 1020 and the citations by 1514. The most significant increase is in the number of PCR based probes which has increased from 886 to 1975.

INTERNATIONAL CONSIDERATIONS: The read-only nodes established at Northwick Park, UK and the DKFZ, Germany have now been extended with nodes at Sydney, Australia, Tokyo, Japan and Uppsala, Sweden. The Australian node is already active and the other 2 will become so this summer. A 16 member International Scientific Advisory Council (ISAC) has been established and met for the first time in mid-May.. Amongst the most pressing issues considered by the ISAC were: the need for developing graphics interfaces, the role of editors, direct electronic submission of data, the representation of mapping data and the number and distribution of remote nodes,. The Director and Informatics PI have attended several meetings as observers of an Interagency Coordinating Committee (IACC), a consortium of governmental funding agencies whose purpose is to coordinate the joint funding of GDB activities on an international scale. The channeling of such funding is a complex matter and it seems likely the GDB will have to

adopt different procedures with different agencies to acquire such funding.

DATA ACQUISITION AND MAINTENANCE: A new set of chromosome editors was elected last autumn, approximately one third being new to the task. Many have subsequently been involved in single chromosome workshops which were also attended by GDB staff members who took the opportunity of discussing the editorial activities with editors. It is obvious that many editors do not feel capable of keeping pace with the enormous amount of new information. Recent discussions within the ISAC and Human Genome Mapping Committee (HGMC) have centered on this problem and the opinion has been expressed that the editors should not be concerned with data entry per se but of validating data already organized and assembled in GDB by chromosome curators in conjunction with GDB central support staff. Such curators would have full time appointments and need not necessarily be appointed with GDB but could be stationed elsewhere as long as they had excellent network connections with GDB in Baltimore. Three curators would be necessary to cover all chromosomes at present.

Administrative, Informatics and Data Acquisition Cores

Attachment 1

Outline Summary of GDB activities, Year 1

Administrative Core (Matheson)

I. Operation of Governance Structure

A) Inter-Agency Coordinating Committee (IACC)
1) PI participated in first meeting in Paris, 12/91
2) second meeting organized by GDB and NIH staff, and hosted at the Johns Hopkins University School of Medicine, 2/92
3) 3rd meeting hosted by JHU in Baltimore, in conjunction with the inaugural meeting of the GDB International Scientific Advisory Council, 5/92

B) International Scientific Advisory Council (ISAC)
1) Inaugural meeting hosted by JHU in conjunction with meeting of the IACC, 5/92
a) final committee recommendations expected during late summer/ early fall 1992
2) Second meeting to be planned during Chromosome Coordinating Meeting 1992 (CCM 92), 11/92

C) Ad Hoc Committees
1) Currently recruiting advisory committee on Comparative Mapping, and plann first meeting to be held in Baltimore, 10/92
2) Meeting of an advisory group to discuss the representation of D-segments planned for 6/92
2) Anticipated that other committees will be organized between no / and the e the first year of funding

D) Ongoing Interaction with HUGO Americas
Organization of Single Chromosome Workshops and maintenance of international contacts

II. Manage Licensing and Distribution

A) Remote Nodes
1) Node Agreement is in final stages of approval

2) Cooperating with the ISAC to define the approval process for nodes
3) Creating mechanism to prioritize node requests, and to act on them as quickly as possible
4) During year 1 of current funding, negotiated the establishment of several new nodes
   a) GDB continued to work with first GDB node, at the MRC (Medical Research Council) in London, established 9/90
   b) DKFZ (German Cancer Research Center) in Heidelberg, Germany, began offering service, 9/91
   c) ANGIS (Australian National Genomic Information Service) in Sydney, Australia, began offering local service 4/92
   d) University of Uppsala in Uppsala, Sweden, plan to offer local service b late summer '92
   e) JICST (Japan Information Center for Science and Technology) in Tokyo, Japan, plan to offer local service by fall '92

B) Front-End Software
1) Signed VAR (Value Added Remarketer) Agreement with Sybase, to act as a vendor for the license needed for GDB users to operate GDB front-end software on individual Sun workstations
2) Currently finalizing agreements to be signed by front-end users, to ensur rights of GDB and Sybase
3) Currently developing distribution mechanisms and proposal for cost-recovery to GDB for postage, shipping, labor involved with providing initial software and subsequent updates

C) Recently drafted the first agreement for commercial distribution of the database, to be presented for negotiation with a firm seeking to develop OMIM as part of the reference package they distribute.

III. Administer the project

A) In response to the need created by the new funding to the GDB and the growth in the size of the staff, the administrative staff was enlarged a
1) Recruited Matheson, 9/91
2) Promoted Ruch
3) Hiring of additional clerical staff continues.

IV. Support additional efforts to seek funding and to increase access to GDB

A) CCM 92
1) Being co-organized by GDB PI Peter Pearson and Kay E. Davies, Ph.D. of the University of Oxford, at the request of and in cooperation with HUGO
2) Application pending with the NIH and DOE, with award letters expected in the next few weeks (during the time that these renewal applications a review)
3) To be hosted at JHU, and fully supported by GDB staff and equipment using on-site resources
4) All aspects of meeting logistics, correspondence, etc. being handled by G staff
5) Several satellite meetings are anticipated, including the GDB ISAC, the IACC, and HUGO Council

B) Additional grant proposals under consideration
1) Development of GUI with Marr to DOE
2) Application to EEC for collaboration with European center for editing interface

Informatics Core (P.I. Robbins)

I. Product Development

A) GDB
1) Release 4.0, 9/91
2) Release 4.2, 4/92
- includes ability for nodes to register local users, independently of JHU
3) Release 5.0, targeted 9/92
- will allow users access to submitted data in several new categories, using the standard user interface, including:
a) sites of genetic crossovers, defined as map objects
b) recombiration frequency, linkage distance, and likelihood data
c) sequences of regions containing STSs and the relevant PCR primer conditions
d) long-range restriction maps
4) Projection of future releases
a) 5.1, targeted 11/92
b) 5.2, targeted 3/93
c) 6.0, targeted 7/93

B) DAT
1) Projection of release schedule
a) 1.0 Release, targeted 9/92
b) 1.1 Release, targeted 11/92
c) 1.2 Release, targeted 3/93
d) 1.3 Release, targeted 7/93
e) 2.0 Release, targeted 9/93

C) Sybase
1) 4.8 upgrade, targeted 9/92

D) SunOS
1) 4.1.2 (open window 3.0) upgrade, targeted 9/92

E) Development of Graphical User Interface (GUI)
1) development, summer '92
2) testing, targeted 9/92
3) release, targeted 11/92

F) Interactions with other databases
1) To develop methods of direct entry of data
a) Lawrence Livermore National Laboratory
b) Washington University
c) GDB in discussion of potential collaborative efforts with Los Alamos National Laboratory and the University of Utah
2) To develop access tools
a) several potential collaborations in discussion
3) To develop databasing of comparative mapping information
a) GDB is currently collaborating with the Jackson Laboratories to integrate Gbase (database of mouse mapping data) with the GDB

II. Product Services

A) User Support
1) Current User statistics, including comparison with earlier similar information whenever possible
2) Maintenance of Help Line (telephone and e-mail)

B) User Documentation
1) manuals updated, targeted to coincide with each GDB release (?)
2) plans for cost-recovery in years 2 and 3
3) documentation and data dictionary have been made available on ftp server

C) User access
1) ISQL access, 2/92
2) ftp server, 2/92
3) data available in text files, 2/92

D) Training
1) courses offered for editors and general users
2) scheduled year-round, both at JHU and off-site. When possible, training courses have been scheduled to coincide with relevant conferences, for t convenience of participants

III. Faculty Recruitment

A) Kenneth H. Fasman, Ph.D., 5/92

Data Acquisition and Maintenance Core (P.I. Pearson)

I. Interaction with scientific community

A) Chromosome Workshop Participation
1) GDB attempts to have a representative in attendance at every SCW
2) at the X Chromosome Workshop in Amalfi, Italy, 5/92, GDB provided on-site editorial access to the database for the first time
B) GDB representatives attended major conferences and meetings, including
1) ICHG, 10/91
2) CSHL Genome Mapping and Sequencing, 5/92
3) AAAS Annual Meeting, 2/92
4) FASEB Annual Meeting, 3/92
C) personal meetings between GDB and Genome Center Directors
1) Washington University, 1/92

II. Editing and validation of data

A) Editors recruited and, in cooperation with Administrative Core staff, consulting contracts signed by each, early spring '92
B) Literature awareness, 3/92, 5/92, then monthly thereafter

III. Improve methods of data submission

A) develop electronic forms
1) design, 12/91
2) production, 3/92

IV. Development of Map Manager

A) Introduced at HGM 11, 8/91

V. Faculty recruitment

A) A. Jamie Cuticchia, Ph.D., 3/92

END

DATE
FILMED
11/24/92