

Informatics Support for Mapping in Mouse-Human Homology Regions

Edward Uberbacher, Richard Mural,¹ Manesh Shah, Loren Hauser,¹ and Sergey Petrov
Computer Science and Mathematics Division and ¹Biology Division; Oak Ridge National Laboratory; Oak Ridge, TN 37831
423/574-6134, Fax: -7860, ube@ornl.gov

The purpose of this project is to develop databases and tools for the Oak Ridge National Laboratory (ORNL) Mouse-Human Mapping Project, including the construction of a mapping database for the project; tools for managing and archiving cDNAs and other probes used in the laboratory; and analysis tools for mapping, interspecific backcross, and other needs. Our initial effort involved installing and developing a relational SYBASE database for tracking samples and probes, experimental results, and analyses. Recent work has focused on a corresponding ACeDB implementation containing mouse mapping data and providing numerous graphical views of this data. The initial relational database was constructed with SYBASE using a schema modeled on one implemented at the Lawrence Livermore National Laboratory (LLNL) center; this was because of documentation available for the LLNL system and the opportunity to maximize compatibility with human chromosome 19 mapping. (Major homologies exist between human chromosome 19 and mouse chromosome 7, the initial focus of the ORNL work.)

With some modification, our ACeDB implementation was modeled somewhat on the Lawrence Berkeley National Laboratory (LBNL) chromosome 21 ACeDB system and designed to contain genetic and physical mouse map data as well as homologous human chromosome data. The usefulness of exchanging map information with LLNL (human chromosome 19) and potentially with other centers has led to the implementation of procedures for data export and the import of human mapping data into ORNL databases.

User access to the system is being provided by workstation forms-based data entry and ACeDB graphical data browsing. We have also implemented the LLNL database browser to view human chromosome 19 data maintained at LLNL, and arrangements are being made to incorporate mouse mapping information into the browser. Other applications such as the *Encyclopedia of the Mouse*, specific tools for archiving and tracking cDNAs and other mapping probes, and analysis of interspecific backcross data and YAC restriction mapping have been implemented.

We would like to acknowledge use of ideas from the LLNL and LBNL Human Genome Centers.

DOE Contract No. DE-AC05-84OR21400.

SubmitData: Data Submission to Public Genomic Databases

Manfred D. Zorn
Software Technologies and Applications Group;
Information and Computing Sciences Division; Lawrence Berkeley National Laboratory; University of California; Berkeley CA 94720
510/486-5041, Fax: -4004, mdzorn@lbl.gov
<http://www-hgc.lbl.gov/submitr.html>

Making information generated by the various genome projects available to the community is very important for the researcher submitting data and for the overall project to justify the expenses and resources. Public genome databases generally provide a protocol that defines the required data formats and details how they accept data, e.g., sequences, mapping information. These protocols have to strike a balance between ease of use for the user and operational considerations of the database provider, but are in most cases rather complex and subject to change to accommodate modifications in the database.

SubmitData is a user interface that formats data for submission to GSDB or GDB. The user interface serves data entry purposes, checking each field for data types, allowed ranges and controlled values, and gives the user feedback on any problems. Besides one-time submissions, templates can be created that can later be merged with TAB-delimited data files, e.g., as produced by common spreadsheet programs. Variables in the template are then replaced by values in defined columns of the input data file. Thus submitting large amounts of related data becomes as easy as selecting a format and supplying an input filename. This allows easy integration of data submission into the data generation process.

The interface is generated directly from the protocol specifications. A specific parser/compiler interprets the protocol definitions and creates internal objects that form the basis of the user interface. Thus a working user interface, i.e., static layout of buttons and fields, data validation, is automatically generated from the protocol definitions. Protocol modifications are propagated by simply regenerating the interface.

The program has been developed using ParcPlace VisualWorks and currently supports GSDB, GDB and RHdb data submissions. The program has been updated to use VisualWorks 2.0.

DOE Contract No. DE-AC03-76SF00098.