AN ELECTRONIC LABORATORY NOTEBOOK BASED ON THE WORLD WIDE WEB

J. E. MARSTALLER, M. D. ZORN

Ernest Orlando Lawrence Berkeley National Laboratory
University of California
1 Cyclotron Rd., Berkeley, CA 94720

The Resource for Molecular Cytogenetics is a joint project at the Ernest Orlando Lawrence Berkeley National Laboratory and University of California, San Francisco. Funded by the U.S. Department of Energy, it has been created to facilitate the application of results in genomics and molecular cytogenetics to clinical applications in, for example, cancer biology. One of the goals of the Resource is the development of an evenly spaced set of probes along each chromosome that are mapped cytogenetically and tie in with genetic and physical maps that are being generated by the Human Genome project. The electronic laboratory notebook described in this paper handles the data management for this probe mapping effort.

The probe mapping effort starts with the selection of an appropriate target locus that is well defined on genetic and physical maps. For each target locus a set of primers is defined to select a clone out of a human P1 library and to develop this clone into a FISH probe. The clone is mapped onto metaphase chromosomes to determine the position along the axis in fractional length from the p-terminus of the chromosome.

A number of reasons, the volume of data that needs to be tracked, the physical separation between the two main participating laboratories, and not the least, to make the results available to the community as quickly as possible, led to the decision to develop an electronic laboratory notebook based on HTML browsers distributing Hypertext documents over the World Wide Web. We identified four key reagents in the lab protocol: target, primer pair, picked probe, and mapped probe. Each element has a one to many relationship with the one following in the protocol, i.e., a target may have many associated primer pairs. Each of these elements is represented in the database. The database has been implemented using the Unix dbm libraries which support simple dictionary- type databases. The tag is an automatically generated identifier, all attributes for the reagent are concatenated to form the value portion. Relationships among the reagents are stored in additional dbm files.

The HTML forms have been designed in close collaboration and over a series of iterations with the biologists. Data entry forms prompt the user to enter all the relevant information. The forms are sent to the Web server where they are processed by Perl scripts and the data stored in the corresponding dbm files. All entries are time-stamped and have administrative information added to each entry. For Example, particular stages corresponding to processing steps in the lab protocol are maintained as states within each reagent and allow for management control over the lab process. In some cases, the processing includes connecting to outside web servers to retrieve relevant background information, e.g., GDB identifiers for targets that have GDB Locus names. Various browser interfaces show the contents of the database and allow retrieval of stored information. The HTML documents are generated automatically by Perl scripts that access the database and build either document tables or text documents. External identifiers are translated into hypertext links to the respective database. Data access security is maintained through the web server access selection.

While it is quite common now to distribute information via the World-Wide Web, we used the technology to create a group-internal communication and information management tool. We developed an electronic laboratory notebook that tracts the progress of the probe mapping effort and its key reagents within the Resource for Molecular Cytogenetics. In using the web as the basis for the lab notebook we utilize not only the growing familiarity of biologists with this type of user interface, but also the hardware platform independence of the interface. By separating the user interface from the underlying database both ends may evolve more independently from each other.

Acknowledgements

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, UC-408, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.