BCM Server Core

Daniel Davison and Randall Smith
Baylor College of Medicine; Houston, TX 77030
713/798-3738, Fax: -3759, *davison@bcm.tmc.edu*
http://www.bcm.tmc.edu

We are providing a variety of molecular biology-related search and analysis services to Genome Program investigators to improve the identification of new genes and their functions. These services are available via the BCM Search Launcher World Wide Web (WWW) pages which are organized by function and provide a single point-of-entry for related searches. Pages are included for 1) protein sequence searches, 2) nucleic acid sequence searches, 3) multiple sequence alignments, 4) pairwise sequence alignments, 5) gene feature searches, 6) sequence utilities, and 7) protein secondary structure prediction. The Protein Sequence Search Page, for example, provides a single form for submitting sequences to WWW servers that provide remote access to a variety of different protein sequence search tools, including BLAST, FASTA, Smith-Waterman, BEAUTY, BLASTPAT, FASTAPAT, PROSITE, and BLOCKS searches. The BCM Search Launcher extends the functionality of other WWW services by adding additional hypertext links to results returned by remote servers. For example, links to the NCBI's Entrez database and to the Sequence Retrieval System (SRS) are added to search results returned by the NCBI's WWW BLAST server. These links provide easy access to Medline abstracts, links to related sequences, and additional information which can be extremely helpful when analyzing database search results. For novice or infrequent users of sequence database search tools, we have pre-set the parameter values to provide the most informative first-pass sequence analysis possible.

A batch client interface to the BCM Search Launcher for Unix and Macintosh computers has also been developed to allow multiple input sequences to be automatically searched as a background task, with the results returned as individual HTML documents directly on the user's system. The BCM Search Launcher as well as the batch client are available on the WWW at URL http://gc.bcm.tmc.edu:8088/search-launcher/launcher.html.

The BCM/UH Server Core provides the necessary computational resources and continuing support infrastructure for the BCM Search Launcher. The BCM/UH Server Core is composed of three network servers and currently supports electronic mail and WWW-based access; ultimately, specialized client-server access will also be provided. The hardware used includes a 2048-processor MasPar massively parallel MIMD computer, a DEC Alpha AXP/OSF1, a Sun 2-processor SparcCenter 1000 server, and several Sun Sparc workstations.

In addition to grouping services available elsewhere on the WWW and providing access to services developed at BCM and UH, the BCM/UH Server Core will also provide access to services from developers who are unwilling or unable to provide their own Internet network servers.

Grant Nos.: DOE, DE-FG03-9SER62097/A000; National Library of Medicine, R01-LM05792; National Science Foundation, BIR 91-11695; National Research Service Award, F32-HG00133-01; NIH, P30-HG00210 and R01-HG00973-01.

A Freely Sharable Database-Management System Designed for Use in Component-Based, Modular Genome Informatics Systems†

Steve Rozen,[1] Lincoln Stein,[1] and **Nathan Goodman**
The Jackson Laboratory; Bar Harbor, ME 04609
Goodman: 207/288-6158, Fax: -6078, *nat@jax.org*
[1]Whitehead Institute for Biomedical Research; Cambridge, MA 02139
http://goodman.jax.org
http://www.genome.wi.mit.edu/informatics/workflow

We are constructing a data-management component, built on top of commercial data-management products, tuned to the requirements of genome applications. The core of this genome data manager is designed to:

- support the semantic and object-oriented data models that have been widely embraced for representing genome data,
- provide domain-specific built-in types and operations for storing and querying bimolecular sequences,
- provide built-in support for tracking laboratory work flows, and admit further extensions for other special-purpose types,
- allow core facilities to be readily extended to meet the diverse needs of biological applications

The core data manager is being constructed on top of Sybase, Oracle, and Informix Universal Server. The software is available free of charge and is freely redistributable.

We will be reporting progress on the core data manager's architecture and interface at the URLs above, and we solicit comments on its design.

DOE Grant No. DE-FG02-95ER62101.

†Originally called Database Management Research for the Human Genome Project, this project was initiated in 1995 at the Massachusetts Institute of Technology–Whitehead Institute.

*Projects designated by an asterisk received small emergency grants following December 1992 site reviews by David Galas (formerly DOE Office of Health and Environmental Research, which was renamed Office of Biological and Environmental Research in 1997), Raymond Gesteland (University of Utah), and Elbert Branscomb (Lawrence Livermore National Laboratory).