Using GMOD tools to create, edit, and browse genome features

Scott Cain(1), Lincoln Stein(1), Suzanna Lewis(2,3), Doreen Ware(1,8), Mark Gibson(2,3), Nomi Harris(2,3), Ken Youens-Clark(1), David Emmert(6), Chris Mungall(2,4), Ben Faga(1), Allen Day(7)

(1) Cold Spring Harbor Laboratory, (2) Berkeley Drosophila Genome Project, (3) University of California Berkeley, (4) Howard Hughes Medical Institute, (5) Harvard University, (6) UCLA, (7) USDA-ARS NAA Plant, Soil & Nutrition Laboratory Research Unit

Application Area: Bioinformatics, Model Organism Database Software.

Platform: Any (Linux, MacOS X, Windows)

Dependencies:
- PostgreSQL (7.3+)
- Apache (1.3+)
- Java 1.4
- BioPerl (> 1.5)
- Perl modules available from CPAN: CGI, GD, DBI, DBD::Pg, Digest::MD5, Text::Shellwords, Module::Build, Class::DBI, Class::DBI::Pg, Class::DBI::Pager, XML::Simple, LWP, Template, Log::Log4perl, Term::Progress


The Generic Model Organism Database (GMOD) Project is an open source project to develop a complete set of software for creating and administering a model organism database. Components of this project include genome visualization and editing tools, literature curation tools, a robust database schema, biological ontology tools, and a set of standard operating procedures. This project is funded by the NIH and the USDA Agricultural Research Service, with participation from members of several database projects, including WormBase, FlyBase, Mouse Genome Informatics, Gramene, the Rat Genome Database, TAIR, EcoCyc, and the Saccharomyces Genome Database. Released modules include Chado, a flexible modular relational schema for genome information and tools to load and dump data (ChadoXML and XORT); Apollo, a genome feature editor and curator's tool; GBrowse, a flexible web-based genome browser; Textpresso, a paper indexing and search tool; the PubSearch/PubFetch literature curation tools, and Caryoscope, a gene expression visualization tool. Over the next year we will be releasing more components, ultimately creating a model organism database construction set.

The 'gmod integrated suite' currently consists of the chado schema(1), tools for loading and dumping data, GBrowse(2) for browsing genomic feature data via a web interface, Apollo(3), a Java application for editing genomic features, and CMap(4), a tool for browsing comparative map data. The demonstration will focus on what is currently possible with the suite, including:
- Using Apollo to create and edit gene models
- Using GBrowse to view the genome features, including the edits by Apollo in real time
- Using CMap to view comparative map data and moving back and forth between the CMap and GBrowse view of genomic features.

References:
(1) chado: http://www.gmod.org/schema/index.shtml; Mungall, C., Emmert, D., Cain, S.
(4) CMap: http://www.gmod.org/cmap/index.shtml; Youens-Clark, K., Faga, B., Ware, D., Stein, L.