GMOD Project Update

Scott Cain
GMOD Project Coordinator
Ontario Institute for Cancer Research
scott@scottcain.net

GMOD Meeting
San Diego, CA
January 14-15, 2010
Introduction: GMOD is …

• A set of interoperable open-source software components for visualizing, annotating, and managing biological data.

• An active community of developers and users asking diverse questions, and facing common challenges, with their biological data.
Who uses GMOD?

Plus hundreds of others
Software

GMOD components can be categorized as

- V Visualization
- D Data Management
- A Annotation
Visualization: GBrowse

GBrowse

Releases
1.70 released
2.0, 1.71 in the pipe

AJAX/Interface:
Rubberband region selection, drag and drop
track ordering, collapsible tracks, popup balloons,
asynchronous rendering (2.0)

Biology:
Allele/gentotype frequency, LD glyphs, geolocation
popups, circular genome support (1.71)

Infrastructure
User logins, server multiplexing (2.0), SQLite and
SAMtools (NGS) adaptors

modENCODE Fly:
http://modencode.oicr.on.ca/cgi-bin/gb2/gbrowse/fly/

The generic genome browser: a building block for a model organism
Visualization

GBrowse

JBrowse

GBrowse_syn

CMap

Resources

GBrowse User Tutorial at OpenHelix.com
GBrowse Admin Tutorial
NGS in GBrowse and SAMtools Tutorial

Web Sites:
GMOD http://gmod.org/wiki/GBrowse
WebGBrowse http://webgbrowse.cgb.indiana.edu/
GBrowse.org http://gbrowse.org

Mailing List:
https://lists.sourceforge.net/lists/listinfo/gmod-gbrowse
GBrowse

JBrowse

GBrowse_syn

CMap

Visualization

GBrowse

GBrowse based comparative genomics viewer
Shows a reference sequence compared to 2+ others
Can also show any GBrowse-based annotations

Syntenic blocks do not have to be colinear
Can also show duplications

http://gmod.org/wiki/GBrowse_syn

Sheldon McKay, Cold Spring Harbor Laboratory
Visualization

GBrowse

JBrowse

GBrowse_syn

CMap

Web based comparative map viewer
CMap is data type agnostic:
  Can link sequence, genetic, physical, QTL, deletion, optical, ...
CMap 2.0 coming
  Faster, internals cleanup
  Circos export

CMap 1.01: A comparative mapping application for the Internet, Ken Youens-Clark, Ben Faga, Immanuel V. Yap, Lincoln Stein and Doreen Ware, *Bioinformatics*, doi:10.1093/bioinformatics/btp458
A extensible, modular database schema for storing biological data

1.0 release:
   - Stable schema
   - Tools for data in/out

1.1 release (soon):
   - Stable schema (minor, nondestructive changes)
   - Improvements to data loading scripts
   - Additional modules: cell line, natural diversity
   - Tool for managing materialized views
   - Tool for creating ontology-based views
Data Management

Chado

New web front end for Chado databases
Set of Drupal modules
Modules approximately correspond to Chado modules
Easy to create new modules
Includes user authentication, job management, curation support

Stephen Ficklin, Meg Staton, Chun-Huai Cheng, ...
Clemson University Genomics Institute
Data Management

Chado
Tripal
TableEdit
BioMart
InterMine

MediaWiki extension
MediaWiki software
used at Wikipedia,
GMOD.org, …

GUI to wiki tables
Also a GUI to
database tables
Work in progress to
use this with Chado

Potential to give
wiki access to
Chado databases

Example: GONUTS (http://gowiki.tamu.edu/)

Jim Hu, Daniel Renfro, et al., Texas A&M
Data Management

Chado
Tripal
TableEdit
BioMart
InterMine

New GUIs - more configurable and easier to use
Virtual Marts - marts running off source schema without materializing
Improved scalability
Security and access control
Improved federation
New configuration tool
Data Management

Chado
Tripal
TableEdit
BioMart
InterMine

Data integration and web-based query package
Now supports ~20 common data formats:
  GFF3, Chado, GO annotation, biopax, BioGrid, TreeFam, PubMed, Ensembl, …
Interfaces: RESTful web service, Java & Perl APIs
Upload & analyse gene lists with graphical and statistical widgets

FlyMine: an integrated database for Drosophila and Anopheles genomics,
MAKER
DIYA
Galaxy
Ergatis
Apollo

Lightweight, modular, and configurable Perl-based pipeline framework
Initial application is gene prediction for prokaryotes
Working on integration of Amos assembly tools

MAKER
DIYA
Galaxy
Ergatis
Apollo

NGS tools support
QC and Manipulation: FASTQ, 454, SOLiD support
Mapping: Bowtie or BWA, Megablast
SAMtools: Web interface to SAMtools scripts

LIMS system in beta.
Import data from your sequencer into Galaxy

- Currently up to 162 analysis tool / components for use in pipelines
- Updated prokaryotic annotation pipeline template
- Updated comparative annotation pipeline template
- Lots of work for use on Amazon EC2
- Now the engine behind the new CloVR cloud computing project (http://clovr.igs.umaryland.edu/)

http://ergatis.sourceforge.net/
Better Chado support
  including DBMS independent support)
GFF3 support
GUI based configurations
Multiple alignment transcript viewer and editor
Continuous data display
  sgr, wiggle
Remote analysis to NCBI services
  BLAST, Primer-BLAST
Undo support
More robust Java Web Start support