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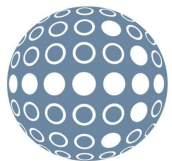
Scott Cain
GMOD Project Coordinator
Ontario Institute for Cancer Research
scott@scottcain.net

2010 GMOD Europe
13-16 September 2010

Ontario Institute
for Cancer Research

Agenda

- Introduction
- What's new

GMOD Europe 2010 Overview

- ## Main meeting today and tomorrow
  - ### Agenda at
    http://gmod.org/wiki/September_2010_GMOD_Meeting#Agenda

- ## Satellites and InterMine Wednesday

- ## BioMart Thursday

- 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

Agenda

- Introduction
- What's new: Software

Software

GMOD components can be categorized as

V Visualization

D Data Management

A Annotation

V Visualization

GBrowse

JBrowse

GBrowse_syn

CMap

Releases
  1.70
  2.14 (rapid development starting with 2.0 in Jan '10)

Features
  Rubberband region selection
  Drag and drop track ordering
  Collapsible tracks
  Popup balloons
  Allele/gentotype frequency, LD glyphs
  Geolocation popups
  Circular genome support (1.71)
  Asynchronous updates (2.0)
  User authentication
  Multiple server support (2.0)
  SQLite, SAMtools (NGS) adaptors

Demo: modENCODE Fly

Visualization

V

GBrowse

JBrowse

GBrowse_syn

CMap

Resources

Tutorials (http://gmod.org/wiki/GBrowse_Tutorial):
  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

GMOD's 2nd Generation Genome Browser
It's *fast*

Completely new genome browser implementation:
 Client side rendering
 Heavy use of AJAX
 Uses JSON and Nested Containment Lists

Demo: JBrowse Fly

Web Sites:
 GMOD   http://gmod.org/wiki/JBrowse
 JBrowse   http://jbrowse.org

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

JBrowse: A next-generation genome browser, Mitchell E. Skinner, Andrew V. Uzilov, Lincoln D. Stein, Christopher J. Mungall and Ian H. Holmes, *Genome Res.* 2009. 19: 1630-1638
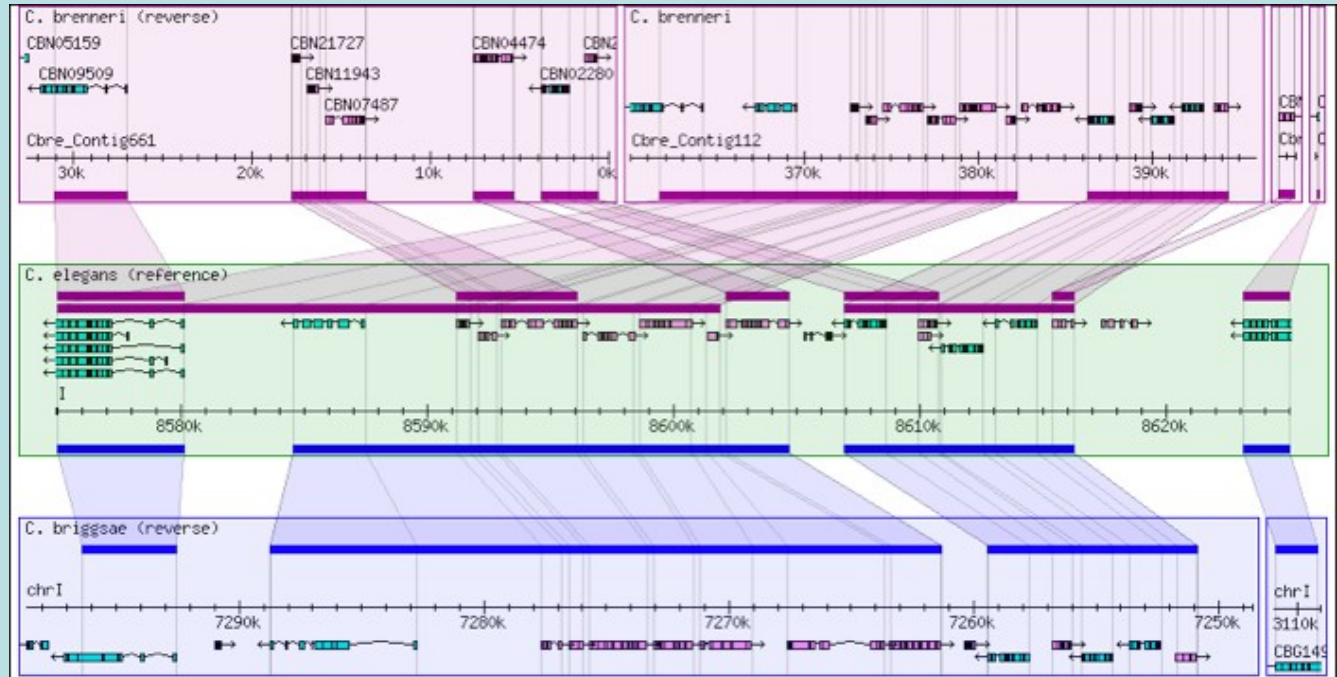
GBrowse

JBrowse

GBrowse_syn

CMap

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

wormbase.org

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

Sheldon McKay, Arizona State University

V Visualization

GBrowse

JBrowse

GBrowse_syn

CMap

Resources

Documentation:
  Home Page   http://gmod.org/wiki/GBrowse_syn
  Tutorial
            http://gmod.org/wiki/GBrowse_syn_Tutorial
  Example   http://www.wormbase.org/cgi-bin/gbrowse_syn/

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

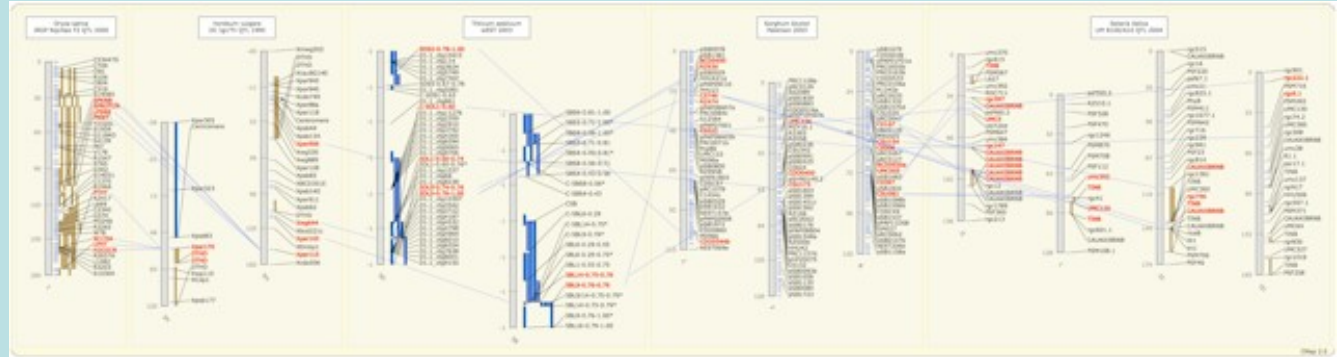Sheldon McKay, Arizona State University

GBrowse

JBrowse

GBrowse_syn

CMap

Web based comparative map viewer
CMap is data type agnostic: Can link sequence,
     genetic, physical, QTL, deletion, optical, …
Particularly popular in plant community

Data Management

Chado

Tripal

TableEdit

BioMart

InterMine

Chado is the GMOD schema; it is modular and extensible, allowing the addition of new data types "easily." Covered data types in ontologies, organisms, sequence features, genotypes, phenotypes, libraries, stocks, microarrays, with natural diversity recently being rolled into the schema (but not yet released).

1.0 Release solidified the Chado that most people were already using from source.

1.1 Introduced support for GBrowse to use full text searching and "summary statistics" (ie, feature density plots). Version 0.30 of Bio::DB::Das::Chado is needed for these functions.

Data Management

Chado

Tripal

TableEdit

BioMart

InterMine

New (2009) 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

Chado

Tripal

TableEdit

BioMart

InterMine

A MediaWiki extension (MediaWiki software used at Wikipedia, GMOD.org)

Provides graphical user interface (GUI) to wiki tables

Can also provide GUI to database tables

Work in progress to use this with Chado

Potential to give wiki access to a Chado database

See http://ecoliwiki.net

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

Data Management

Chado

Tripal

TableEdit

BioMart

InterMine

BioMart is a query-oriented data management system

Provides a web based query interface

Strong data federation

Workshop on Thursday

Syed Haider, Junjun Zhang, Arek Kasprzyk, *et al*.

Data Management

Chado

Tripal

TableEdit

BioMart

InterMine

InterMine is a query-oriented data management system

Provides a web based query interface

Very flexible queries and query optimization

Workshop on Wednesday

Richard Smith, Gos Micklem, et al, Cambridge University

Annotation

MAKER

DIYA

Galaxy

Ergatis

Apollo

Genome annotation pipeline for creating gene models
Output can be loaded into GBrowse, Apollo, Chado, …
Incorporates
   SNAP, RepeatMasker, exonerate, BLAST,
   Augustus, FGENESH, GeneMark, MPI
Other capabilities
   Map existing annotation onto new assemblies
   Merge multiple legacy annotation sets into a consensus set
   Update existing annotations with new evidence
   Integrate raw InterProScan results
Maker Online in beta

Cantarel B. L., et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008 Jan;18(1):188-96

Annotation

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

DIYA: a bacterial annotation pipeline for any genomics lab, Andrew C. Stewart, Brian Osborne and Timothy D. Read, *Bioinformatics* 2009 25(7):962-963

Annotation

MAKER

DIYA

Galaxy

Ergatis

Apollo

Java-based GUI application for browsing and annotating genomic sequences

Can be installed via WebStart (ie, by clicking on a link)

Can read/write to Chado, GFF3, GenBank, GAME XML

Lewis, S. E. et al. Apollo: a sequence annotation editor.
Genome Biol 3(12):RESEARCH0082

Next Spring Sometime:

ABRF: Association of Biomolecular Resource Facilities
        Feb. 19-22, San Antonio, TX

Biology of Genomes
        May 10-14, Cold Spring Harbor Lab, NY

Suggestions?