

GMOD/GBrowse



Sheldon McKay

Scientific Lead, iPlant Collaborative

University of Arizona

mckays@email.arizona.edu



Comparative Genomics Course Schedule

- Monday 7-10 PM
 - Introduction to GMOD
 - Gbrowse 2.0 part 1



1. [The Basics](#)
 1. [The Data File](#)
 2. [Defining Tracks](#)
 3. [Adding Descriptions to a Feature](#)
 4. [Adjusting GBrowse Name Searches](#)
 5. [Linking](#)
 6. [Adding Popup Balloons to Tracks](#)
 1. [Customizing Balloons](#)
2. [Displaying Common Types of Features](#)
 1. [Multi-segmented features](#)
 2. [Protein-Coding Genes](#)
 3. [Reading Frames](#)
 4. [Grouped Features](#)
 5. [Quantitative Data \(basic\)](#)
 6. [Quantitative Data \(advanced\)](#)
 7. [DNA and 3-frame translations](#)
 8. [ESTs and Other Alignments](#)
 1. [Adding DNA to Alignments](#)
3. [GBrowse Enhancements](#)
 1. [Semantic Zooming](#)
4. [Using Other Backends](#)
 1. [The MySQL Backend](#)
 2. [Other Backends](#)

Comparative Genomics Course Schedule

- Tuesday 9-12 AM
 - Gbrowse 2.0 part 2



1. [The Basics](#)
 1. [The Data File](#)
 2. [Defining Tracks](#)
 3. [Adding Descriptions to a Feature](#)
 4. [Adjusting GBrowse Name Searches](#)
 5. [Linking](#)
 6. [Adding Popup Balloons to Tracks](#)
 1. [Customizing Balloons](#)
2. [Displaying Common Types of Features](#)
 1. [Multi-segmented features](#)
 2. [Protein-Coding Genes](#)
 3. [Reading Frames](#)
 4. [Grouped Features](#)
 5. [Quantitative Data \(basic\)](#)
 6. [Quantitative Data \(advanced\)](#)
 7. [DNA and 3-frame translations](#)
 8. [ESTs and Other Alignments](#)
 1. [Adding DNA to Alignments](#)
3. [GBrowse Enhancements](#)
 1. [Semantic Zooming](#)
4. [Using Other Backends](#)
 1. [The MySQL Backend](#)
 2. [Other Backends](#)

Comparative Genomics Course Schedule

- Tuesday 2-5 PM
 - Comparative genomics with Gbrowse_syn
 - Gbrowse_syn tutorial



No data source configured for GBrowse_syn

Please consult [the documentation](#)

Generic Synteny Browser: Example Data

These are instructions for setting up the example databases provided for GBrowse_syn in the GBrowse distribution.

Data Files

The oryza data source contains genome annotations in GFF3 flat files for two rice species and blastz-derived whole genome alignment data between the two species.

- /var/www/gbrowse2/databases/gbrowse_syn/rice/rice.gff3
- /var/www/gbrowse2/databases/gbrowse_syn/wild_rice/wild_rice.gff3

⋮



GMOD Project Overview

Sheldon McKay

Core Gbrowse/Gbrowse_syn developer

mckays@email.arizona.edu

Agenda

- Introduction
- What's new: Software
- What's new: Project

Focusing on what's happened in the past 2 years
And what's happening in the coming years



<http://gmod.org/>



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?



Agenda

- Introduction
- What's new: Software
- What's new: Project

Software

GMOD components can be categorized as

- V** Visualization
- D** Data Management
- A** Annotation

Visualization: GBrowse

GBrowse

JBrowse

GBrowse_syn

CMap



Releases

1.71 (stable branch) released

2.17 released

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 (HTGS) adaptors

modENCODE Fly:

<http://modencode.oicr.on.ca/cgi-bin/gb2/gbrowse/fly/>



The generic genome browser: a building block for a model organism system database. Stein LD et al. (2002) *Genome Res* 12: 1599-610

Visualization

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:

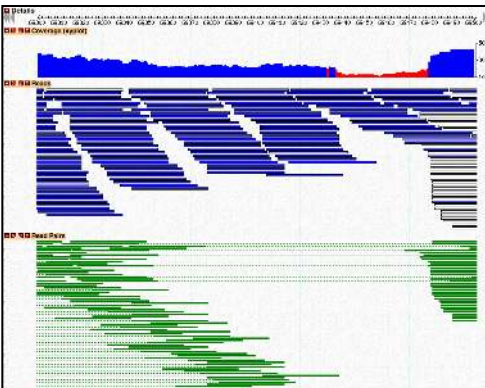
GMOD <http://gmod.org/wiki/GBrowse>

WebGBrowse <http://webgbrowse.cgb.indiana.edu/>

GBrowse.org <http://gbrowse.org>

Mailing List:

<https://lists.sourceforge.net/lists/listinfo/gmod-gbrowse>



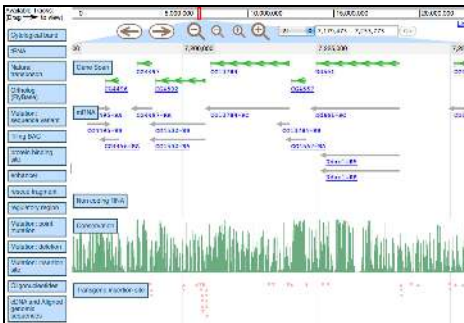
Visualization

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

JBrowse Fly:

<http://jbrowse.org/genomes/dmel/>

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



Visualization

GBrowse

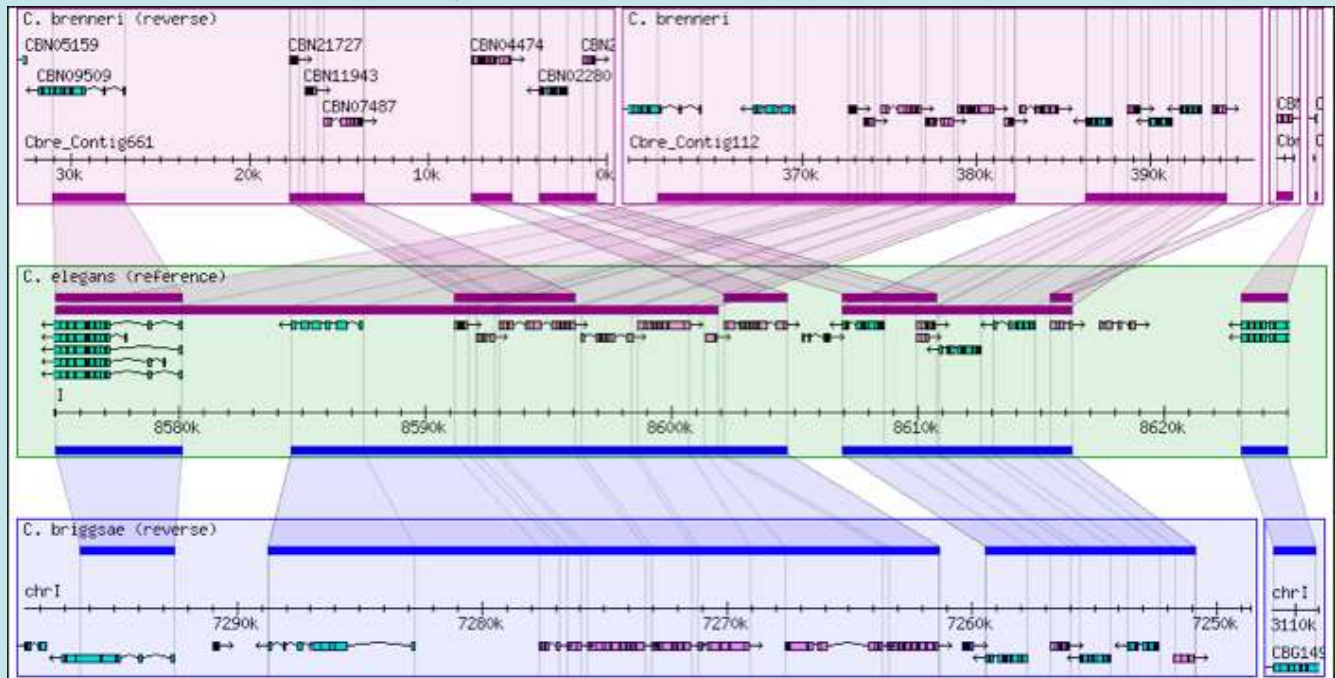
JBrowse

GBrowse_syn

CMap



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



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

wormbase.org

http://gmod.org/wiki/GBrowse_syn

McKay SJ, Vergara IA and Stajich, J. 2010. "Using the Generic Synteny Browser (Gbrowse_syn)" in Current Protocols in Bioinformatics (Wiley Interscience) doi: 10.1002/0471250953.bi0912s31



Visualization

GBrowse

JBrowse

GBrowse_syn

CMap

Web based comparative map viewer

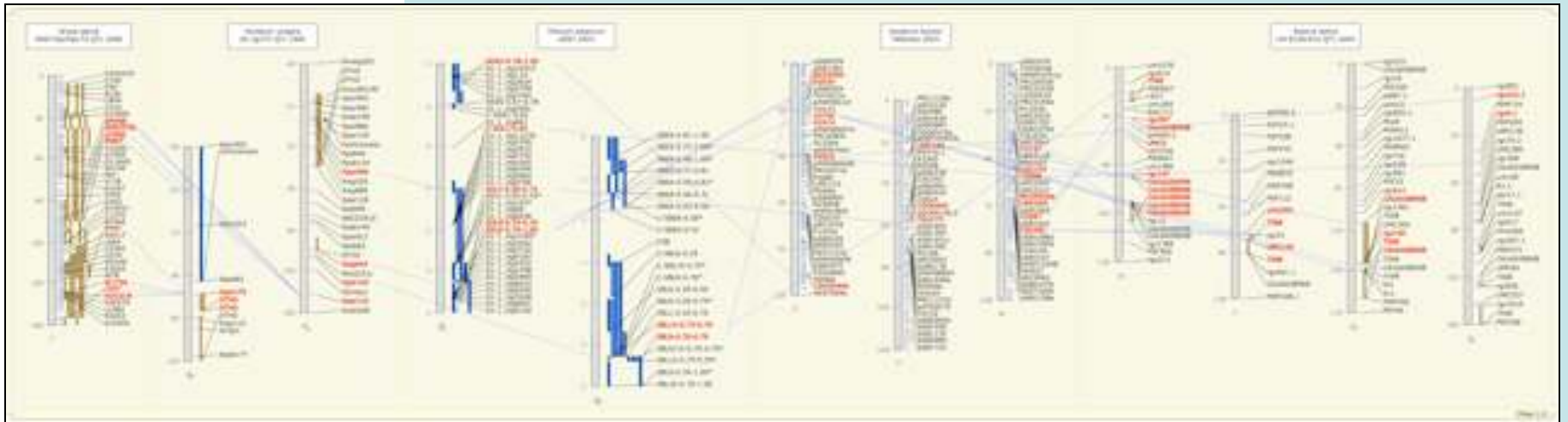
CMap is data type agnostic:

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



Data Management

Chado

A extensible, modular database schema for storing biological data

Tripal

1.0 release:

TableEdit

Stable schema

Tools for data in/out

BioMart

1.1 release (soon):

InterMine

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

Tripal

TableEdit

BioMart

InterMine

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



GDR | Genome Database for Rosaceae

Search Login

Home General Info Species Projects Maps Search Tools Community Contact Us

Genomic, Genetic, and Breeding Resources for Rosaceae Crop Improvement

Welcome to the Genome Database for Rosaceae

The Genome Database for Rosaceae (GDR) is a curated and integrated web-based relational database providing centralized access to Rosaceae genomics and genetics data and analysis tools to facilitate cross-species utilization of data. Supported by the NSF Plant Genome Program from 2003 to 2008, GDR is now funded until 2013 by the USDA Specialty Crop Research Initiative. As part of tree fruit Genome Database Resources (tfGDR), GDR will be expanded to include whole genome sequences and annotations for apple, peach, and strawberry, metacyc pathways, large scale phenotype and genotype data, breeding data, controlled vocabularies and new analysis tools, implemented within the open source infrastructure of and chado (tripal).

Whats new in GDR?

Community News (7)

- RosEXEC meeting to be held Sunday, January 10, 8am-10pm in the Towne Room at PAG 2010
- RosIGI meeting to be held Sunday January 10, 10am-12pm in the Towne Room at PAG 2010
- Fruit and Nut Workshop to be held Saturday, January 9, 8am-12.30pm at PAG 2010



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



Data Management

Chado

Tripal

TableEdit

BioMart

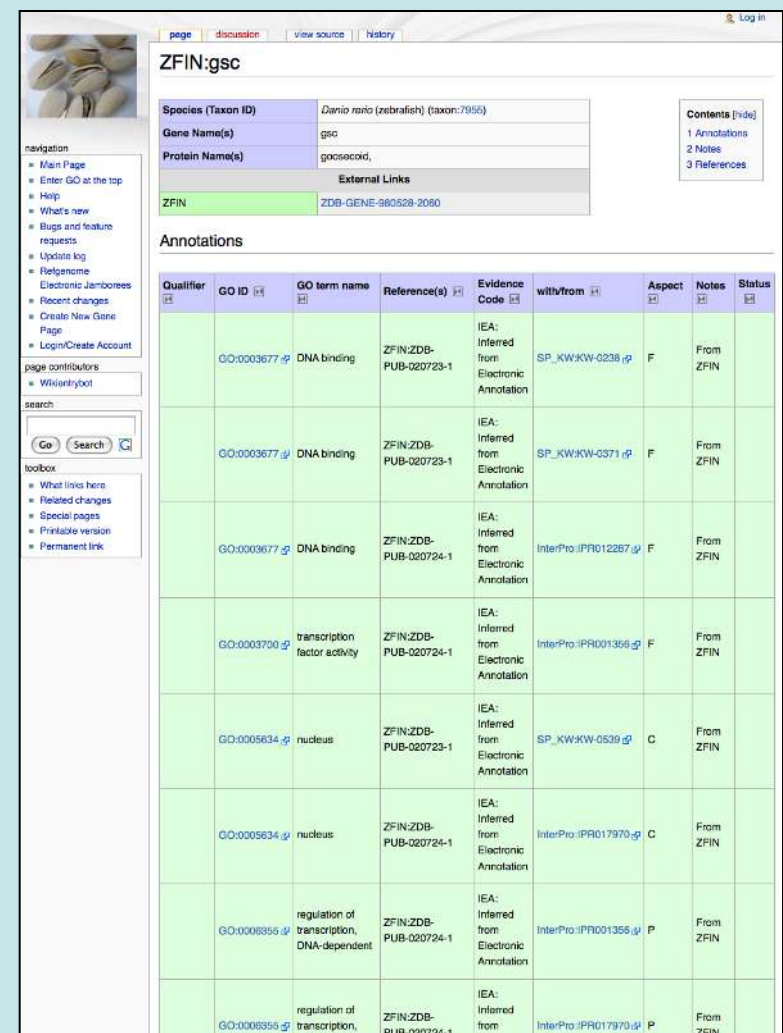
InterMine

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/>)



The screenshot displays the ZFIN: gsc page, which provides detailed information about the gene gsc (goosecoid) in Danio rerio (zebrafish). The page includes a navigation sidebar on the left with links to Main Page, Enter GO at the top, Help, What's new, Bugs and feature requests, Update log, RefGenome, Electronic Jamborees, Recent changes, Create New Gene Page, and Login/Create Account. The main content area shows the gene's taxonomic information, protein name(s), and external links. Below this, a table of annotations is presented, detailing various GO terms and their associated evidence and references.

Qualifier	GO ID	GO term name	Reference(s)	Evidence Code	with/from	Aspect	Notes	Status
	GO:0003677	DNA binding	ZFIN:ZDB-PUB-020723-1	IEA; Inferred from Electronic Annotation	SP_KW:KW-0238	F	From ZFIN	
	GO:0003677	DNA binding	ZFIN:ZDB-PUB-020723-1	IEA; Inferred from Electronic Annotation	SP_KW:KW-0371	F	From ZFIN	
	GO:0003677	DNA binding	ZFIN:ZDB-PUB-020724-1	IEA; Inferred from Electronic Annotation	InterPro:IPR012267	F	From ZFIN	
	GO:0003700	transcription factor activity	ZFIN:ZDB-PUB-020724-1	IEA; Inferred from Electronic Annotation	InterPro:IPR01356	F	From ZFIN	
	GO:0005634	nucleus	ZFIN:ZDB-PUB-020723-1	IEA; Inferred from Electronic Annotation	SP_KW:KW-0539	C	From ZFIN	
	GO:0005634	nucleus	ZFIN:ZDB-PUB-020724-1	IEA; Inferred from Electronic Annotation	InterPro:IPR017970	C	From ZFIN	
	GO:0008395	regulation of transcription, DNA-dependent	ZFIN:ZDB-PUB-020724-1	IEA; Inferred from Electronic Annotation	InterPro:IPR01355	P	From ZFIN	
	GO:0008395	regulation of transcription	ZFIN:ZDB-PUB-020724-1	IEA; Inferred from	InterPro:IPR017970	P	From ZFIN	

Data Management

Chado

Tripal

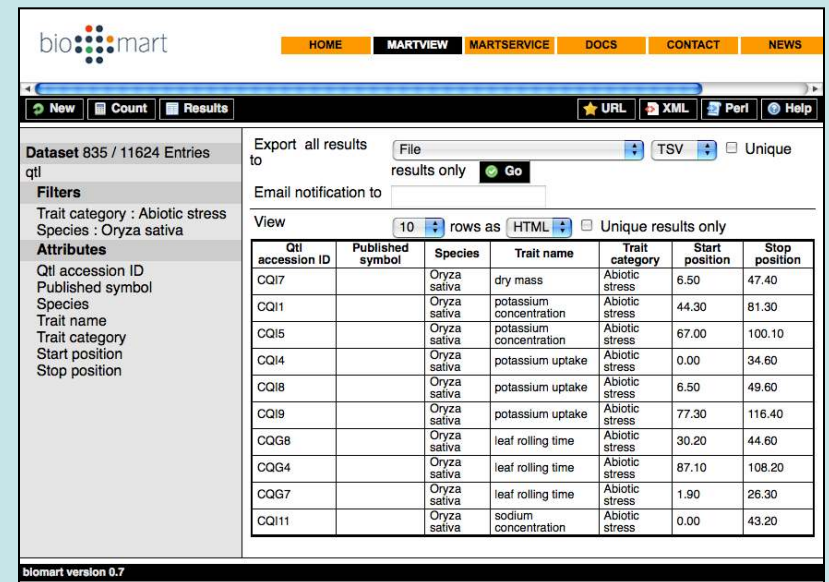
TableEdit

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

A screenshot of the BioMart web application. The top navigation bar includes links for HOME, MARTVIEW, MARTSERVICE, DOCS, CONTACT, and NEWS. Below this is a toolbar with buttons for New, Count, Results, and a search bar. The main content area shows a dataset of 835 / 11624 entries for "Oryza sativa". It includes filters for Trait category (Abiotic stress) and Species (Oryza sativa). A table of attributes is listed on the left. The main table displays 10 rows of data with columns for Qtl accession ID, Published symbol, Species, Trait name, Trait category, Start position, and Stop position. The bottom of the page shows the version "biomart version 0.7".

Qtl accession ID	Published symbol	Species	Trait name	Trait category	Start position	Stop position
CQI7		Oryza sativa	dry mass	Abiotic stress	6.50	47.40
CQI1		Oryza sativa	potassium concentration	Abiotic stress	44.30	81.30
CQI5		Oryza sativa	potassium concentration	Abiotic stress	67.00	100.10
CQI4		Oryza sativa	potassium uptake	Abiotic stress	0.00	34.60
CQI8		Oryza sativa	potassium uptake	Abiotic stress	6.50	49.60
CQI9		Oryza sativa	potassium uptake	Abiotic stress	77.30	116.40
CQG8		Oryza sativa	leaf rolling time	Abiotic stress	30.20	44.60
CQG4		Oryza sativa	leaf rolling time	Abiotic stress	87.10	108.20
CQG7		Oryza sativa	leaf rolling time	Abiotic stress	1.90	26.30
CQI11		Oryza sativa	sodium concentration	Abiotic stress	0.00	43.20



<http://www.biomart.org/>



Data Management

Chado

Tripal

TableEdit

BioMart

InterMine

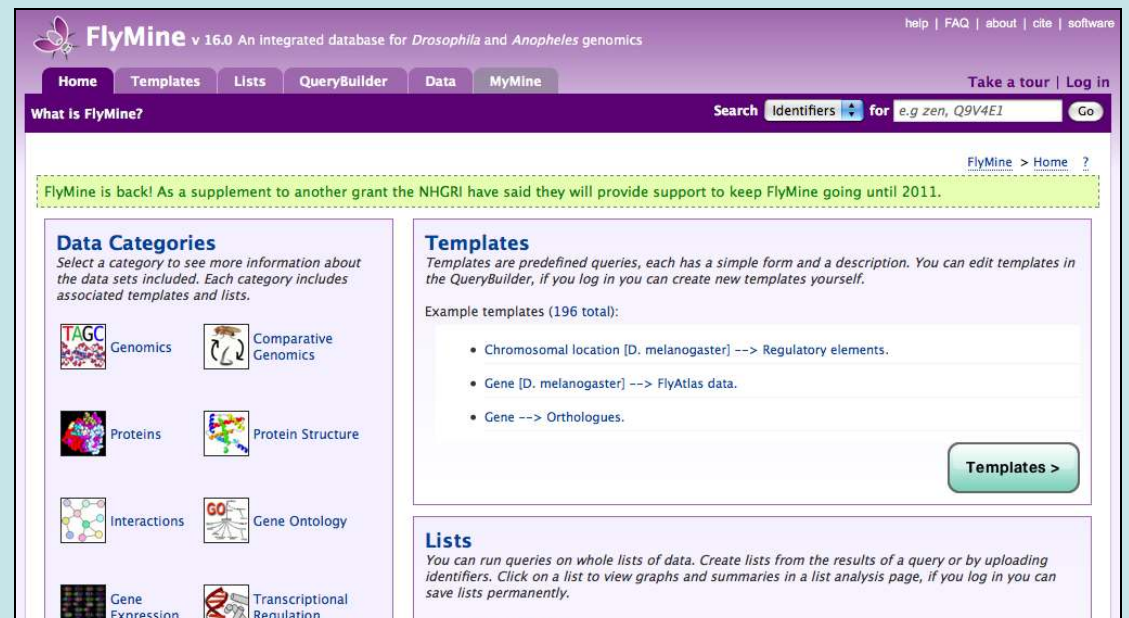
Data integration and web-based query package

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, Rachel Lyne, *et al.*, *Genome Biol.* 2007; 8(7): R129.



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 Web Annotation Service - MAKER online



MAKER: An Easy-to-use Annotation Pipeline Designed for Emerging Model
Organism Genomes. Cantarel B, *et al.*, *Genome Res.* 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

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

The screenshot displays the Galaxy web interface. On the left is a 'Tools' sidebar with categories like 'Get Data', 'Text Manipulation', and 'Fetch Alignments'. The main panel shows the 'MAF Coverage Stats' tool configuration. The 'Interval File' is set to '2: CEBPe sites', 'MAF Source' is 'Locally Cached Alignments', 'MAF Type' is '28-way multiZ (hg18)', and 'Type of Output' is 'Coverage by Region'. An 'Execute' button is at the bottom. Below the form is a 'What it does' section. On the right, a 'History (options)' panel shows a list of jobs, with '2: CEBPe sites' selected and its details expanded, showing 2,126 regions in bed format. Below that, '1: UCSC Main on Human: encodeAffyChIPHi60Sites (encode)' is visible.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A.. 2010. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.*15:1451-5.



Annotation

MAKER

DIYA

Galaxy

Ergatis

Apollo

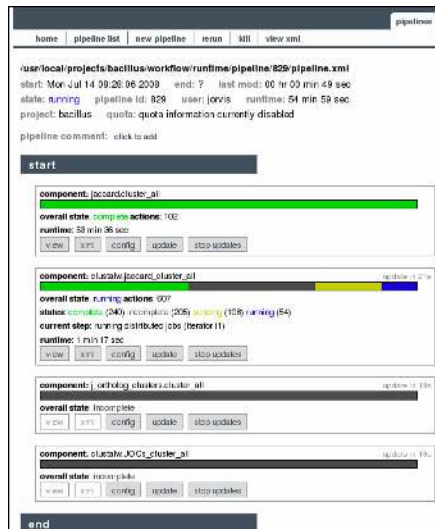
Currently up 162 analysis tool / components for use in pipelines

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/>

Annotation

MAKER

DIYA

Galaxy

Ergatis

Apollo

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



Apollo: a sequence annotation editor. Lewis SE, Searle SMJ, Harris N, Gibson M, Iyer V, Richter J, Wiel C, Bayraktaroglu L, Birney E, Crosby MA, Kaminker JS, Matthews B, Prochnik SE, Smith CD, Tupy JL, Rubin GM, Misra S, Mungall CJ, Clamp ME (2002). Genome Biology 2002, 3(12):research0082



Agenda

- Introduction
- What's new: Software
- What's new: Project

GMOD.org

A wiki, of course. GMOD.org is the hub for all things related to the project:

- Documentation
- News
- Links
- Calendar
- Tutorials
- HOWTOs
- Glossary
- Overview
- ...

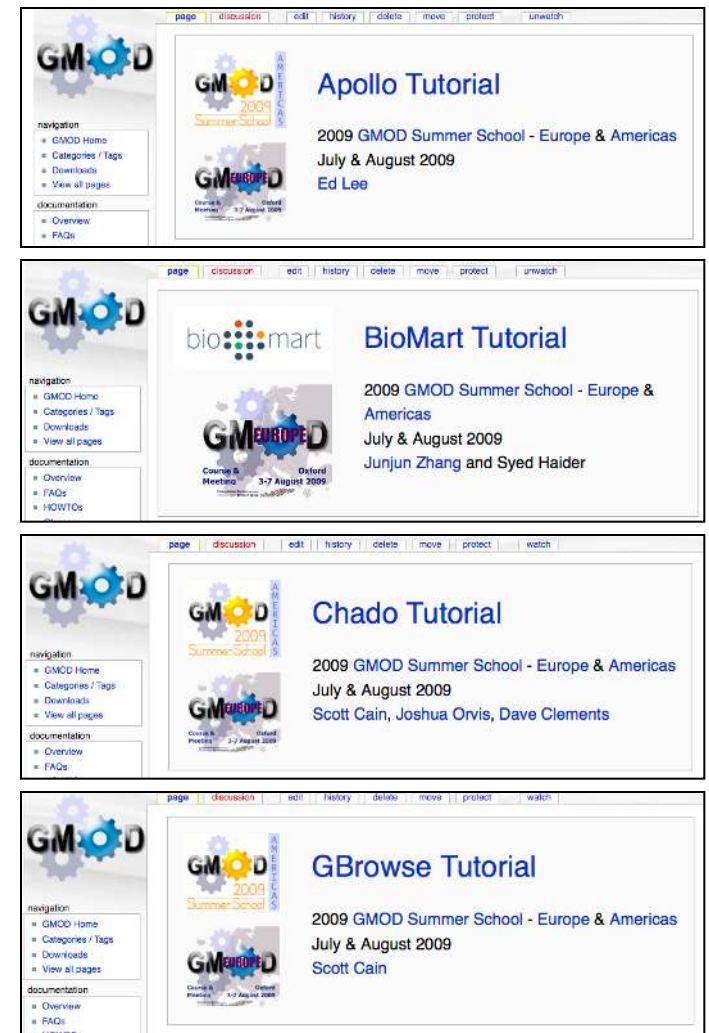


The screenshot shows the GMOD.org homepage with a dark blue header. The main content area is titled 'Welcome to GMOD' and contains several sections: 'GMOD is the Generic Model Organism Database project...', 'How do I Get Started?' (with links to Overview, GMOD Components, and community meeting), 'How do I Get Support?' (with links to Support, Mailing Lists, Training and Outreach, and Help Desk), 'How do I Get Involved?' (with links to Existing users, potential users, feedback, and project meetings), and 'Contributing Organizations'. The left sidebar includes navigation links (GMOD Home, Categories / Tags, Downloads, View all pages), documentation links (Overview, FAQs, HOWTOs, Glossary), community links (GMOD News, Support, Calendar, About this site), developers links (SVN, SourceForge Site), a search box, and a toolbox (What links here, Related changes, Upload file, Special pages, Printable version). The right sidebar features logos of partner organizations (e.g., University of Cambridge, ASHG), a section for 'GMOD Tutorials Now Available' (listing workshops and releases), 'GMOD News' (RSS feed), 'New & Revised Pages' (RSS feed), and 'Popular GMOD Tools' (listing Genome Browsing and Editing tools like GBrowse and Apollo).

Launched by Brian Osborne, 2007

Tutorials

- Summer school sessions become online tutorials with
 - Starting VMware images
 - Step by step instructions
 - Example datasets
 - Ending VMware images
- Topics:
 - Apollo, Artemis-Chado Integration, BioMart, Chado, CMap, GBrowse, GBrowse_syn, JBrowse, MAKER, Tripal, GBrowse HTGS



⋮



http://gmod.org/wiki/Training_and_Outreach#Online_Tutorials



Acknowledgements



NESCent
Todd Vision
Hilmar Lapp
Dave Clements

Cambridge
Richard Smith

OICR
Lincoln Stein
Scott Cain

Texas A&M
Jim Hu

CSHL
Ken Youens-Clark

Cornell
Dave Matthews

Oregon
Patrick Phillips
Phillips Lab

Utah
Carson Holt
Mark Yandell

UC Berkeley
Mitch Skinner
Ian Holmes

Penn State
Dan Blankenberg
Anton Nekrutenko

Arizona
Sheldon McKay

BBOP
Ed Lee

