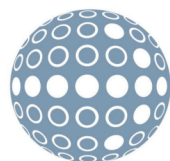




# GMOD Project Update

Scott Cain, PhD.  
GMOD Project Coordinator  
Ontario Institute for Cancer Research  
[scott@scottcain.net](mailto:scott@scottcain.net)

Dave Clements  
GMOD Help Desk  
US National Evolutionary Synthesis Center (NESCent)  
[clements@nescent.org](mailto:clements@nescent.org)



Ontario Institute  
for Cancer Research



Plant and Animal Genome (PAG) XVIII  
11 January 2010

# GMOD @ Plant and Animal Genome XVIII

## GMOD Users

Gramene  
RAP-DB  
ConiferGDB  
TAIR  
MaizeGDB  
SGN  
TreeGenes  
SoyBase  
Birdbase  
Marine Genomics  
BeeBase  
GnplS  
+ hundreds more

## GMOD Tools

Chado  
Apollo  
GBrowse  
CMap  
Tripal  
GBrowse\_syn  
Ergatis  
Textpresso  
BioMart  
JBrowse  
MAKER  
Galaxy  
InterMine  
...



## Animal Genomics with Galaxy: Analyze, Publish & Visualize

Saturday 1:30-2:30pm, California Room, Dan Blankenberg

## MAKER: An easy to use genome annotation pipeline

Saturday, 3:50-6:00pm, California Room, Carson Holt

## Database Resources at the EBI: Ensembl (including BioMart)

Sunday, 8:15-9:15am, California Room, Bert Overduin

## The Generic Genome Browser: A Hands on Workshop

Sunday, 1:30-3:40pm, California Room, Scott Cain

## Using Gramene (including CMap)

Tuesday, 3:50-6:00pm, California Room, Ken Youens-Clark

## Comparative Genomics with GBrowse\_syn

Wednesday, 10:20am-12:30pm, Pacific Salon 3, Sheldon McKay

Workshops

## GMOD Project Update

Monday, 12:50-1:10pm, California Room, Dave Clements, Scott Cain

## Using the JBrowse Genome Browser with Large Amounts of Data

Monday, 1:10-1:30pm, California Room, Mitch Skinner (also see poster P919)

## Tripal: A Construction Toolkit for Online Genomics Databases

Wednesday, 2:30-2:50pm, California Room, Stephen Ficklin

Talks

## Monday 10-11:30am

P858: GBrowse and Next Generation Sequencing Data

## Monday 3-4:30pm

P859: Visualizing and Comparing Genomes

P861: Comparative Genomics Tools in GMOD

Posters

at least 15 other talks and 19 more posters by GMOD users (see [http://gmod.org/wiki/PAG\\_2010](http://gmod.org/wiki/PAG_2010)) +

[http://gmod.org/wiki/PAG\\_2010](http://gmod.org/wiki/PAG_2010)

Got genome?

# 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?



Plus hundreds of others



# Agenda

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



# Software

GMOD components can be categorized as

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



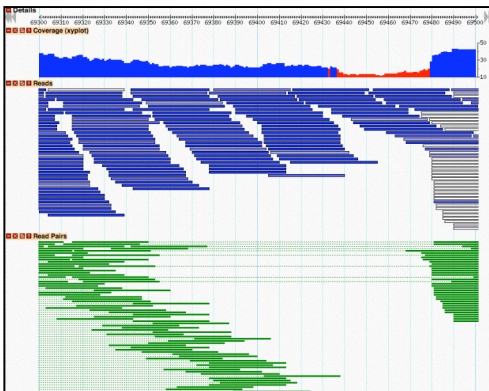
# Visualization: GBrowse

GBrowse

JBrowse

GBrowse\_syn

CMap



Releases

1.70 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/genotype 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 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](http://gmod.org/wiki/GBrowse_Tutorial)):

GBrowse User Tutorial at OpenHelix.com

GBrowse Admin Tutorial

NGS in GBrowse and SAMtools Tutorial

Web Sites:

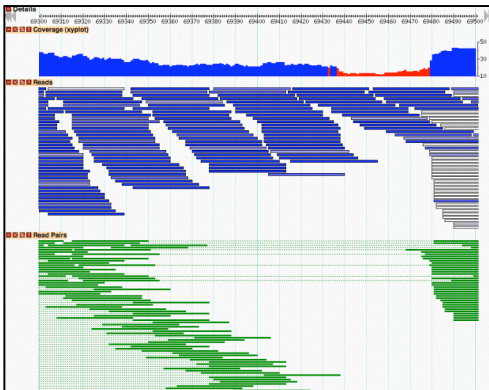
GMOD <http://gmod.org/wiki/GBrowse>

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

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

Mailing List:

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



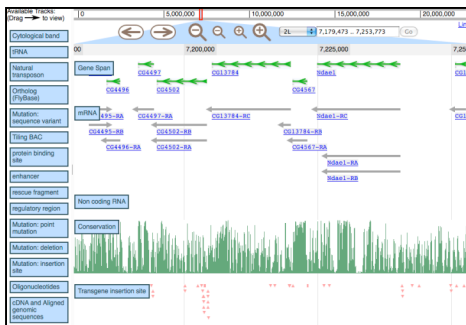
# Visualization

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

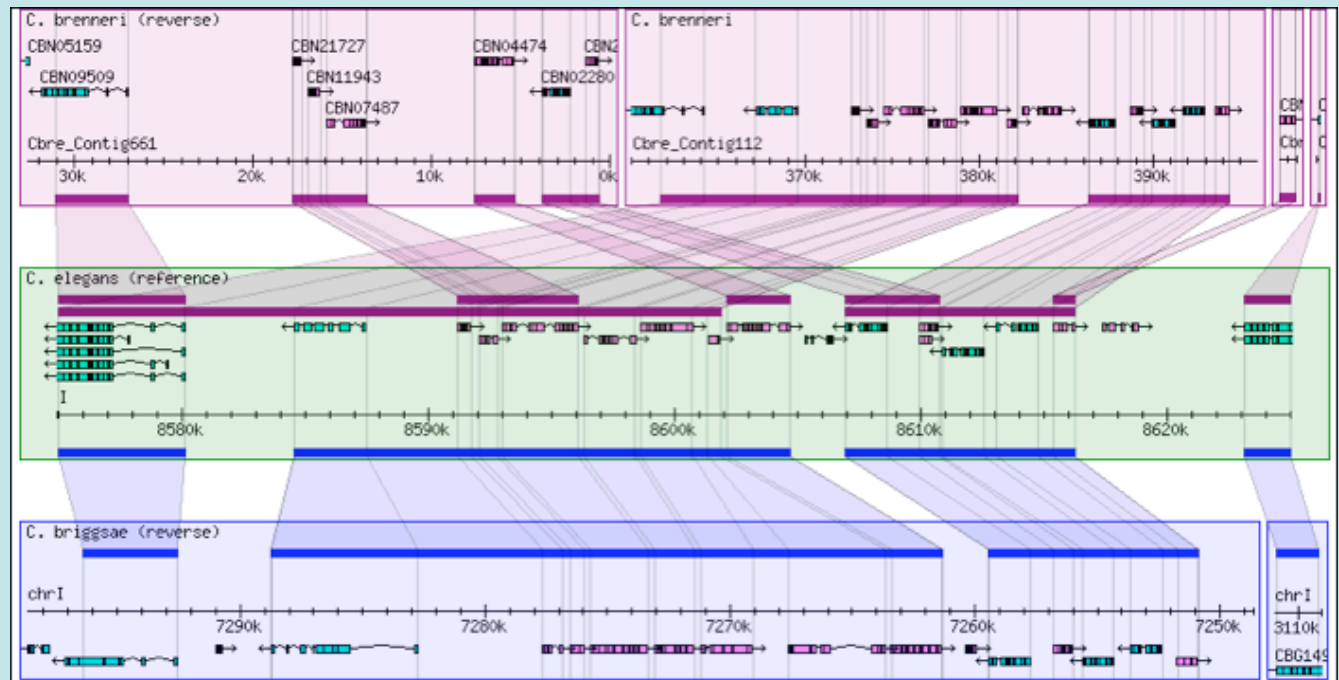
GBrowse

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

[http://gmod.org/wiki/GBrowse\\_syn](http://gmod.org/wiki/GBrowse_syn)



Sheldon McKay, Cold Spring Harbor Laboratory





# Visualization

GBrowse

JBrowse

GBrowse\_syn

CMap

Web based comparative map viewer

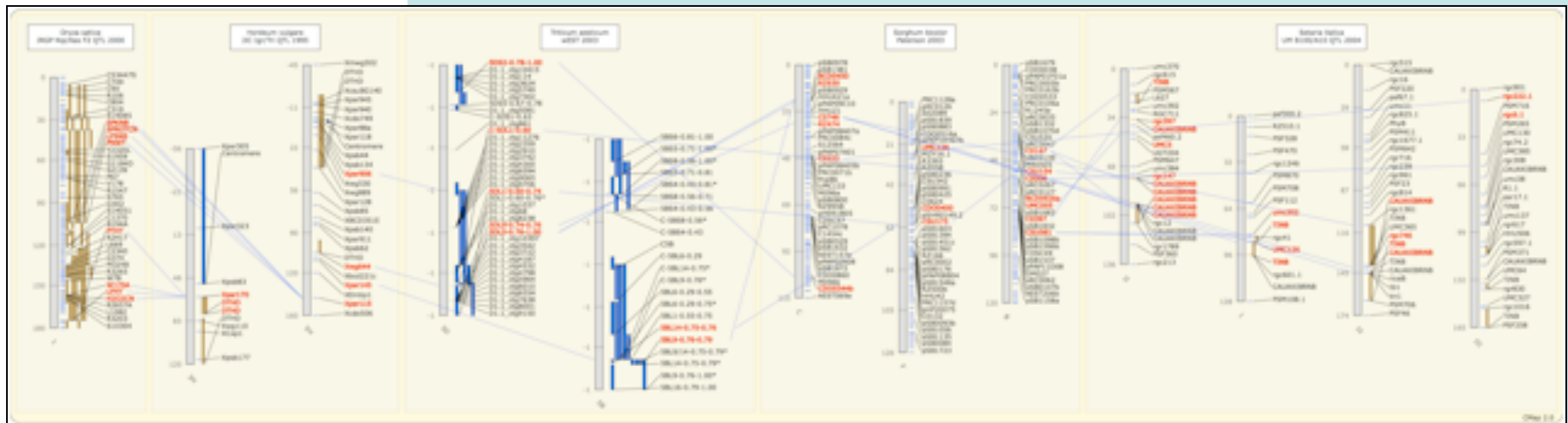
CMap is data type agnostic:

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





# 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

The screenshot shows the homepage of the Genome Database for Rosaceae (GDR). The header features the GDR logo, a search bar, and a 'Login' link. A navigation menu includes links for Home, General Info, Species, Projects, Maps, Search, Tools, Community, and Contact Us. The main content area is titled 'Genomic, Genetic, and Breeding Resources for Rosaceae Crop Improvement'. Below this, there is a 'Welcome to the Genome Database for Rosaceae' section with a paragraph of text. To the right, there is a 'Community News (7)' section with a list of upcoming meetings and workshops.

**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 shows a MediaWiki page for the gene *Danio rerio* (zebrafish) with the gene name *gscc*. The page displays a table of annotations for the ZFIN ID ZDB-GENE-980528-2060. The table includes columns for Qualifier, GO ID, GO term name, Reference(s), Evidence Code, with/from, Aspect, Notes, and Status. The annotations are as follows:

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:IPR012287	F	From ZFIN	
	GO:0003700	transcription factor activity	ZFIN:ZDB-PUB-020724-1	IEA: Inferred from Electronic Annotation	InterPro:IPR001356	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:0006355	regulation of transcription, DNA-dependent	ZFIN:ZDB-PUB-020724-1	IEA: Inferred from Electronic Annotation	InterPro:IPR001356	P	From ZFIN	
	GO:0006355	regulation of transcription,	ZFIN:ZDB-PUB-020724-1	IEA: Inferred from	InterPro:IPR017970	P	From ZFIN	



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

The screenshot shows the BioMart web interface. At the top, there are navigation tabs: HOME, MARTVIEW, MARTSERVICE, DOCS, CONTACT, and NEWS. Below the navigation is a search bar and a toolbar with buttons for New, Count, Results, URL, XML, Peri, and Help. The main content area displays a query result for Dataset 835 / 11624 Entries. The filters section shows "Trait category : Abiotic stress" and "Species : Oryza sativa". The attributes section lists: Qtl accession ID, Published symbol, Species, Trait name, Trait category, Start position, and Stop position. The results table is displayed with 10 rows and 7 columns. The table data is as follows:

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

biomart version 0.7



# 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 v 16.0 An integrated database for *Drosophila* and *Anopheles* genomics

help | FAQ | about | cite | software

Home Templates Lists QueryBuilder Data MyMine

Take a tour | Log in

What is FlyMine? Search Identifiers for e.g zen, Q9V4E1 Go

FlyMine > Home ?

FlyMine is back! As a supplement to another grant the NHGRI have said they will provide support to keep FlyMine going until 2011.

**Data Categories**  
Select a category to see more information about the data sets included. Each category includes associated templates and lists.

- Genomics
- Comparative Genomics
- Proteins
- Protein Structure
- Interactions
- Gene Ontology
- Gene Expression
- Transcriptional Regulation

**Templates**  
Templates are predefined queries, each has a simple form and a description. You can edit templates in the QueryBuilder, if you log in you can create new templates yourself.

Example templates (196 total):

- Chromosomal location [D. melanogaster] --> Regulatory elements.
- Gene [D. melanogaster] --> FlyAtlas data.
- Gene --> Orthologues.

Templates >

**Lists**  
You can run queries on whole lists of data. Create lists from the results of a query or by uploading identifiers. Click on a list to view graphs and summaries in a list analysis page, if you log in you can save lists permanently.



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 shows the Galaxy web interface. The main panel displays 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 visible. Below the configuration, a 'What it does' section explains that the tool relates coverage information by interval for each species. The right-hand 'History' panel shows a list of jobs, with the most recent one being '2: CEBPe sites', which has a table of genomic coordinates.

1	2	3
chr7	26931348	26931580
chr7	26935351	26935655
chr7	26943983	26944116
chr7	26952663	26953560
chr7	26954691	26955126
chr7	26982888	26983140



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

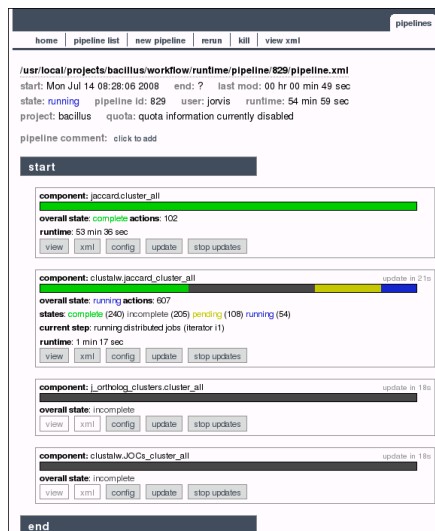
MAKER

DIYA

Galaxy

Ergatis

Apollo



The screenshot displays the Ergatis pipeline management interface. At the top, there are navigation links: home, pipeline list, new pipeline, rerun, kill, and view xml. Below this, the pipeline details are shown for a pipeline named 'pipeline.xml' with ID 829. The pipeline is currently in a 'running' state, started on Monday, July 14, 2008, at 08:28:06. The user is 'jovis' and the runtime is 54 minutes and 59 seconds. The project is 'bacillus' and the quota information is currently disabled. A pipeline comment 'click to add' is visible. The pipeline is divided into several components, each with a progress bar and detailed status information:

- component: jaccardcluster\_all**: overall state: complete, actions: 102, runtime: 53 min 36 sec. Buttons: view, xml, config, update, stop updates.
- component: clusterw\_jaccard\_cluster\_all**: overall state: running, actions: 607, state: complete (240), incomplete (206), pending (108), running (54), current step: running distributed jobs (iterator 11), runtime: 1 min 17 sec. Buttons: view, xml, config, update, stop updates.
- component: |\_ortholog\_clusters\_cluster\_all**: overall state: incomplete. Buttons: view, xml, config, update, stop updates.
- component: clusterw\_JOCs\_cluster\_all**: overall state: incomplete. Buttons: view, xml, config, update, stop updates.

The pipeline ends with an 'end' button.

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

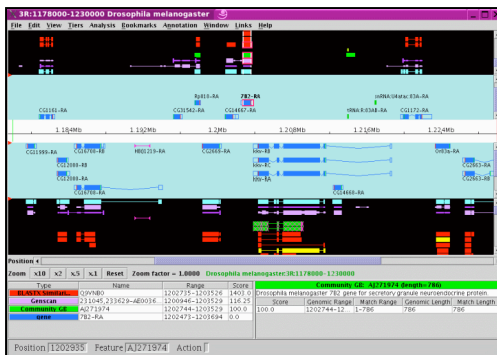
sgr, wiggle

Remote analysis to NCBI services

BLAST, Primer-BLAST

Undo support

More robust Java Web Start support



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



# Agenda

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



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

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



A screenshot of the GMOD.org website homepage. The page has a dark blue header with the GMOD logo and navigation tabs for 'page', 'discussion', 'view source', and 'history'. A 'Log in / create account' link is in the top right. 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?', 'How do I Get Support?', 'How do I Get Involved?', and 'Contributing Organizations'. A left sidebar contains navigation, documentation, community, developers, search, and toolbox sections. A right sidebar features logos of partner organizations, a 'GMOD News' RSS feed, and a list of recent news items. The bottom of the page has a footer with the text 'Launched by Brian Osborne, 2007' and a small circular logo for the 'GENOME BROWSE CONFERENCE'.

Launched by Brian Osborne, 2007



# Training and Outreach

- GMOD Summer Schools
  - 2009
    - July, NESCent, North Carolina, US
    - August, Oxford, UK
  - 2010
    - May, NESCent, North Carolina, US
    - ??, Asia / Pacific, maybe
- Outreach, Training
  - BA, SMBE, PAG, Arthropod Genomics, ...
- Semi-annual community meetings
  - Next Meeting:
    - January 2010, San Diego, after PAG



[http://gmod.org/wiki/Training\\_and\\_Outreach](http://gmod.org/wiki/Training_and_Outreach)

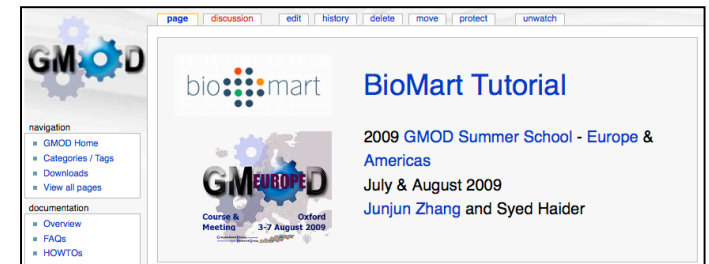


# Tutorials

- 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 NGS



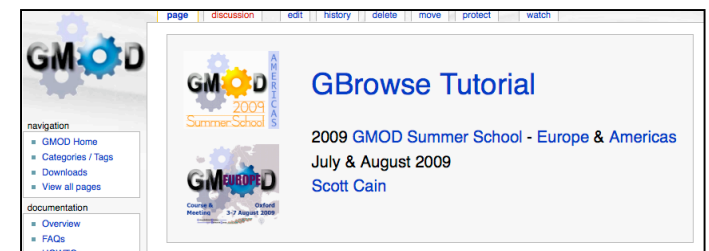
The screenshot shows the Apollo Tutorial page on the GMOD wiki. The page title is "Apollo Tutorial" and it is part of the "2009 GMOD Summer School - Europe & Americas" held in July & August 2009. The author is Ed Lee. The page includes a navigation menu with links to GMOD Home, Categories / Tags, Downloads, and View all pages. There is also a documentation section with links to Overview and FAQs.



The screenshot shows the BioMart Tutorial page on the GMOD wiki. The page title is "BioMart Tutorial" and it is part of the "2009 GMOD Summer School - Europe & Americas" held in July & August 2009. The authors are Junjun Zhang and Syed Haider. The page includes a navigation menu with links to GMOD Home, Categories / Tags, Downloads, and View all pages. There is also a documentation section with links to Overview, FAQs, and HOWTOs.



The screenshot shows the Chado Tutorial page on the GMOD wiki. The page title is "Chado Tutorial" and it is part of the "2009 GMOD Summer School - Europe & Americas" held in July & August 2009. The authors are Scott Cain, Joshua Orvis, and Dave Clements. The page includes a navigation menu with links to GMOD Home, Categories / Tags, Downloads, and View all pages. There is also a documentation section with links to Overview and FAQs.



The screenshot shows the GBrowse Tutorial page on the GMOD wiki. The page title is "GBrowse Tutorial" and it is part of the "2009 GMOD Summer School - Europe & Americas" held in July & August 2009. The author is Scott Cain. The page includes a navigation menu with links to GMOD Home, Categories / Tags, Downloads, and View all pages. There is also a documentation section with links to Overview and FAQs.

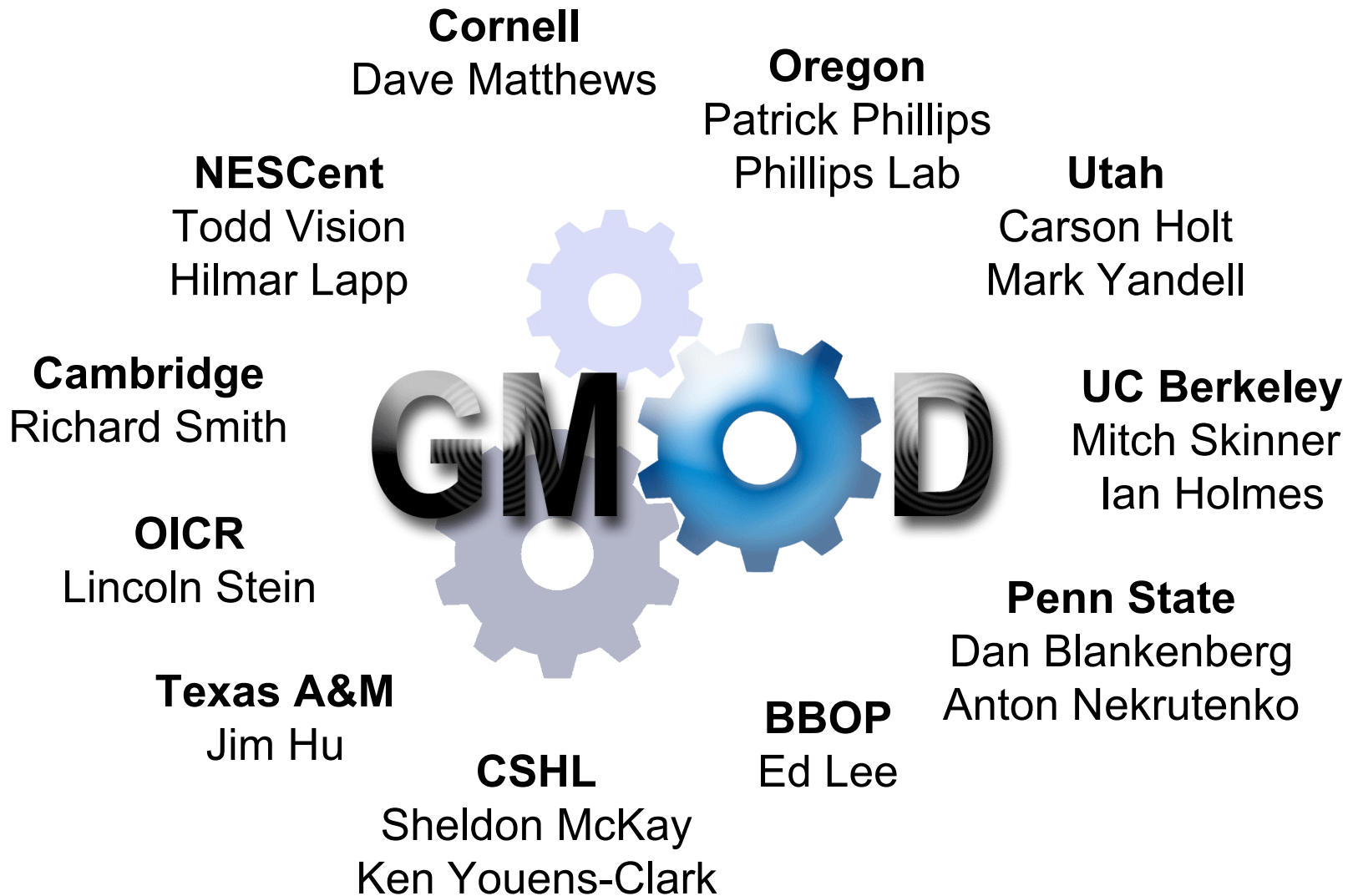
⋮



[http://gmod.org/wiki/Training\\_and\\_Outreach#Online\\_Tutorials](http://gmod.org/wiki/Training_and_Outreach#Online_Tutorials)



# Acknowledgements





# Thank You!



**Scott Cain**

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

**Dave Clements**

GMOD Help Desk  
US National Evolutionary Synthesis Center (NESCent)  
clements@nescent.org



[http://gmod.org/wiki/PAG\\_2010](http://gmod.org/wiki/PAG_2010)

