



Comparative Genomics with GMOD

Dave Clements

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

Information Systems for Insect Pests (ISyIP)
16-17 November 2009

Outline

- **GMOD Project Overview**
- **GMOD for Comparative Genomics**
 - Multiple Sequence Alignment
 - Visualisation: GBrowse_syn
 - Capabilities and future plans
 - Generating & Analyzing: Ergatis & Galaxy
 - Comparative Mapping Visualisation
 - CMap
 - Population Genetics & Natural Diversity
- **Resources**



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.



GMOD Software: A whirlwind tour

Apollo

GBrowse & JBrowse

CMap

MAKER & DIYA

GBrowse_syn, Sybil, ...

Chado, GMODWeb, Tripal

BioMart, InterMine

Pathway Tools

Galaxy, Ergatis

...













The image displays a stack of overlapping screenshots from various GMOD software interfaces. The top-most screenshot is the Galaxy workflow monitoring interface for 'ergatis', showing a pipeline with two components: 'jaccard.cluster_all' (complete) and 'clustalw.jaccard_cluster_all' (running). Other visible screenshots include a genome browser for Drosophila melanogaster, a comparative genomics tool, and a FlyMine database interface.



Who uses GMOD?



Who uses GMOD at ISyIP?

-  *GMOD for Comparative Genomics*
-  *From AphidBase and Lepido-DB to an Information System for Insect Plant pest genomics studies*
-  *Towards assembly of the Helicoverpa armigera genome after 454 sequencing*
-  *The Diabrotica Genetics Consortium, Sequencing the Rootworm Genome and Creating WCRbase*
-  *CycADS: an annotation management system for the development and update of BioCyc metabolic network databases*
- Whole genome sequencing of Tetranychus urticae: novel genomic tools in chelicerate important for agriculture*
-  *InsectaCentral - Facilitating transcriptome work through a GMOD platform*
-  *The URGI bioinformatic platform: an original information system to bridge genetic and genomic plant and fungal data*
-  *The Hymenoptera Database*
-  *Ensembl and Insect Genomes*
-  *FlyBase as a Model for Insect Genome Information Systems*
-  *Enabling, cost-effective bioinformatics analysis for arthropod genomics*
-  *Arthropod Genomics Initiative at K-State*

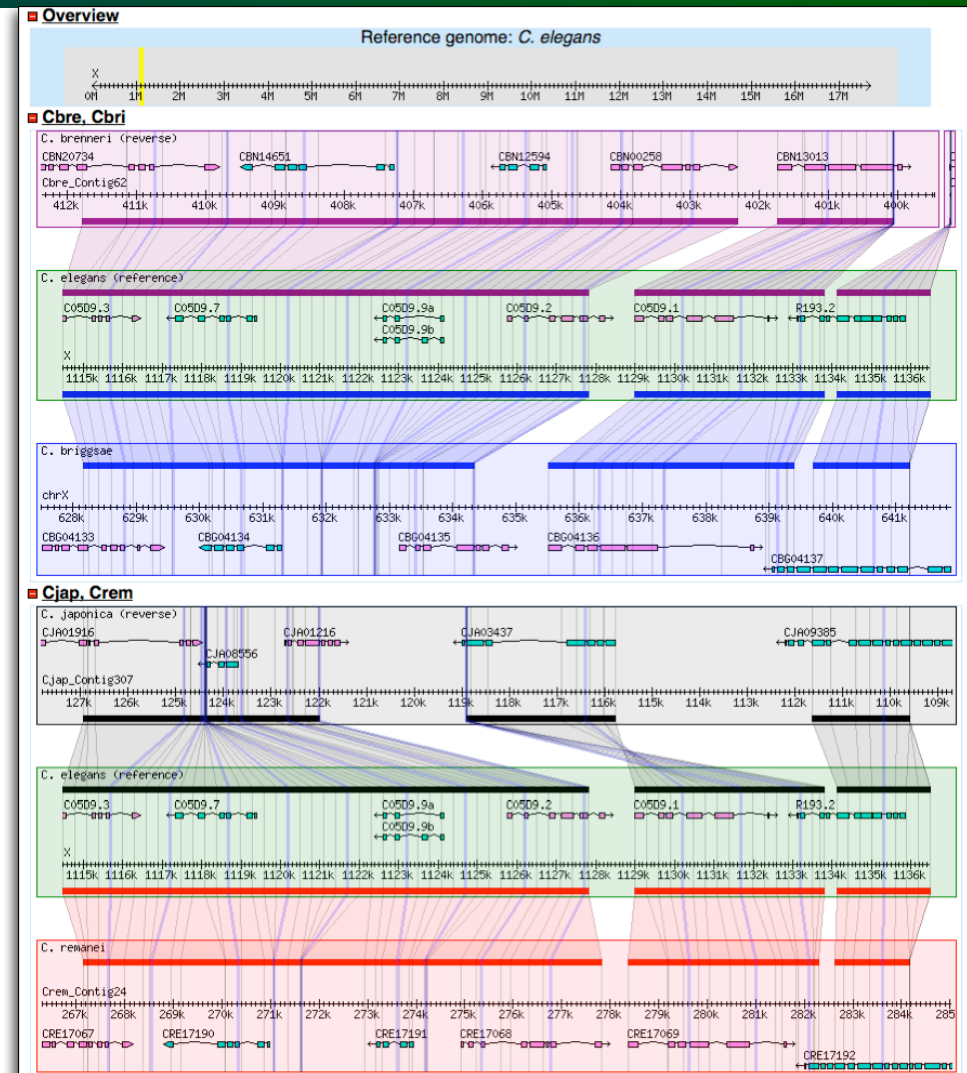
Outline

- GMOD Project Overview
- GMOD for Comparative Genomics
 - Multiple Sequence Alignment
 - Visualisation: GBrowse_syn
 - Capabilities and future plans
 - Generating & Analyzing: Ergatis & Galaxy
 - Comparative Mapping Visualisation
 - CMap
 - Population Genetics & Natural Diversity
- Resources



GBrowse_syn: Visualisation

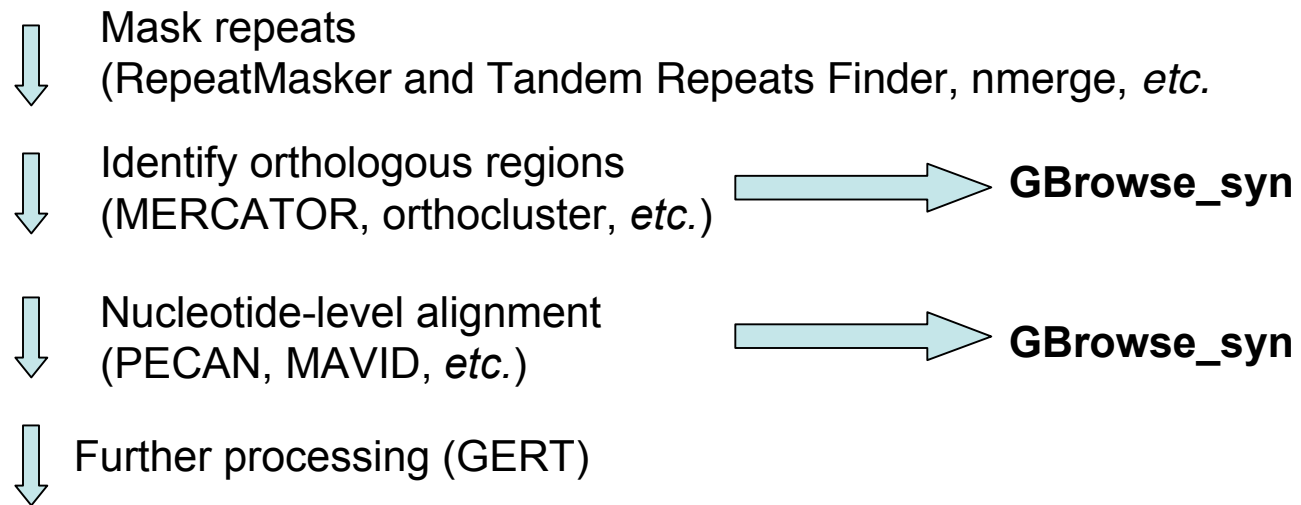
- GBrowse based synteny viewer
- Shows a reference sequence compared to 2 or more others
- Can also show any GBrowse-based annotations



C. elegans compared to 4 other species at WormBase

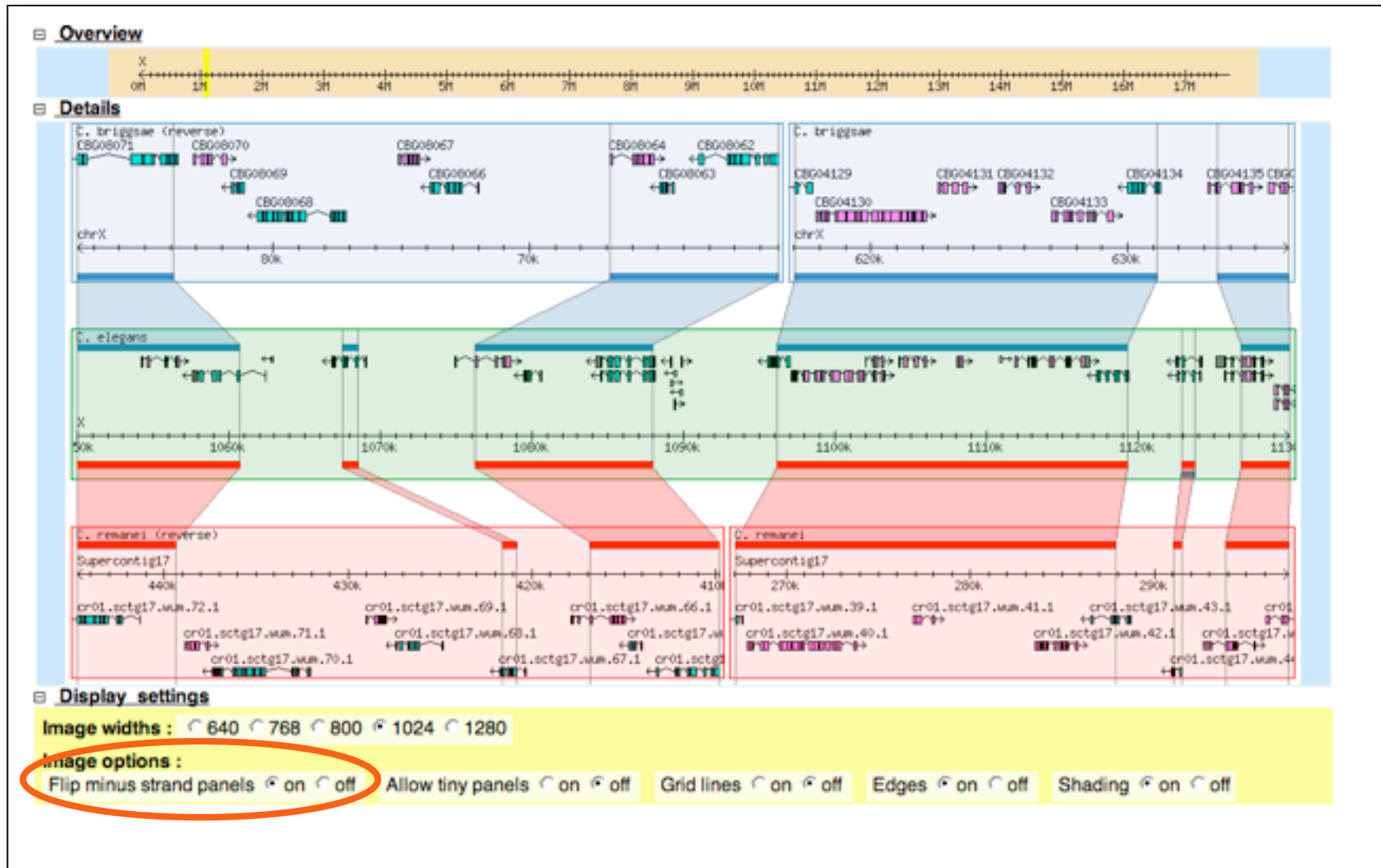
GBrowse_syn in the Big Picture

Raw genomic sequences

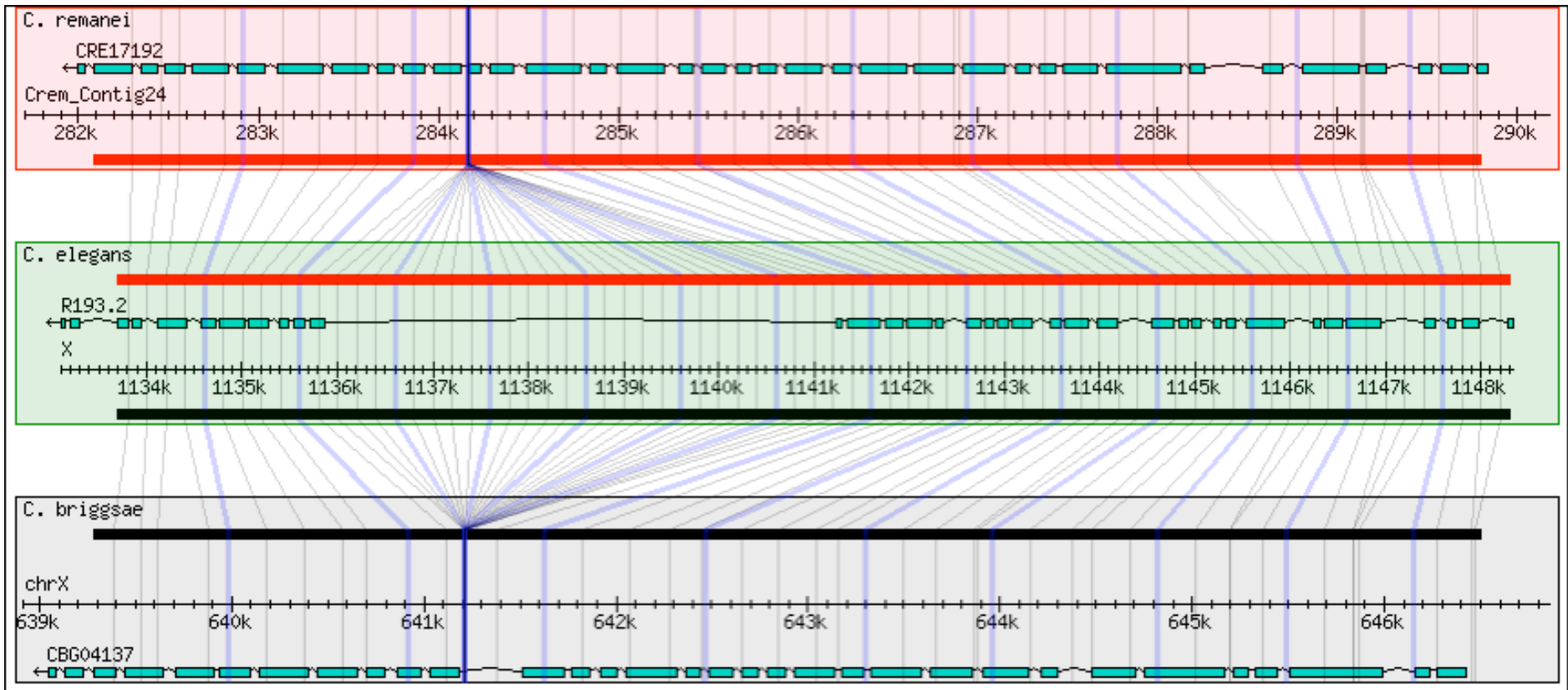


GBrowse

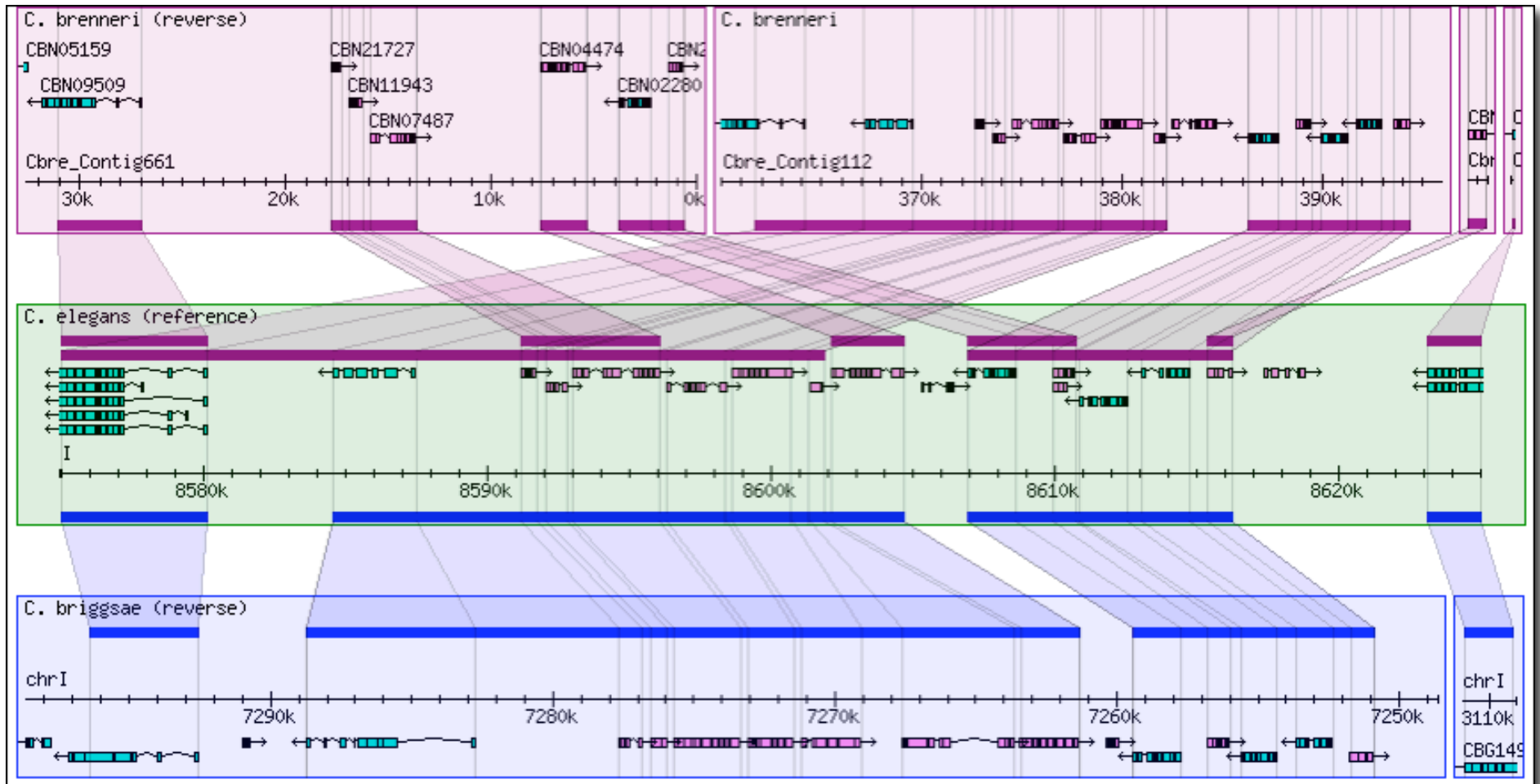
Strand correction, Chaining



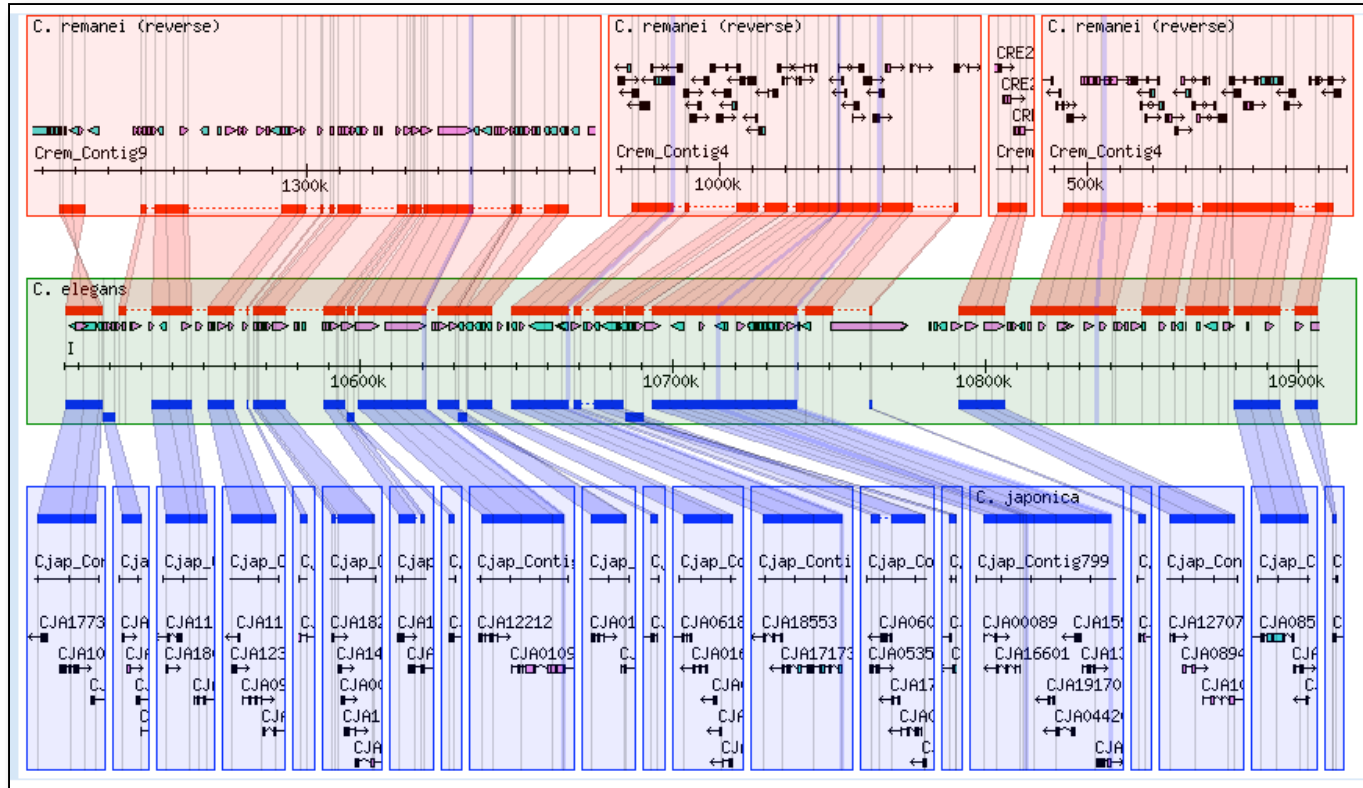
Showing Base Level Alignments



Duplications



Visual inspection of an assembly



Promising

Done

Not so much

GBrowse_syn Future Work

- Integration with GBrowse 2
- High-level graphical overview
- ...

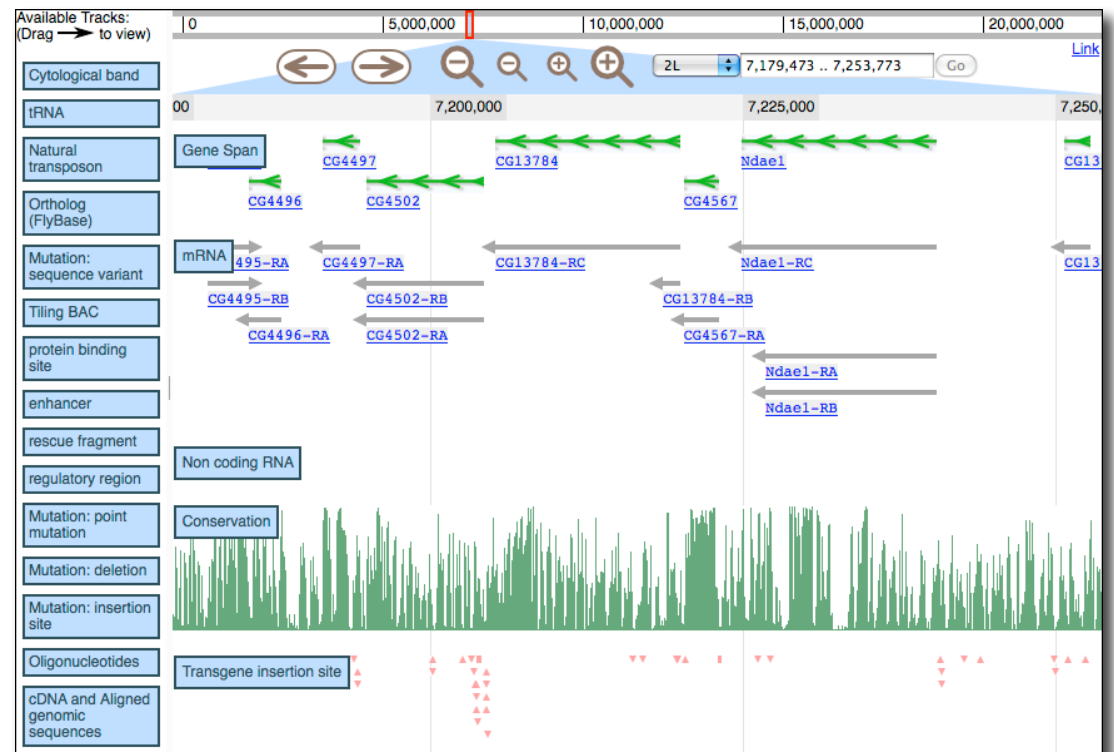
GBrowse_syn + JBrowse = JBrowse_syn

Add GBrowse_syn functionality to JBrowse genome browser

browser

- JBrowse

- Complete rewrite of GBrowse
- Uses AJAX, JSON, client side rendering, NCList
- Very fast



JBrowse.org

- Grant proposal submitted November 2009



JBrowse: A next-generation genome browser, Mitchell E. Skinner, *et al.*, Genome Res. 2009. 19: 1630-1638

GBrowse_syn Resources

Home Page	http://gmod.org/wiki/GBrowse_syn
Tutorial	http://gmod.org/wiki/GBrowse_syn_Tutorial
User Help	http://gmod.org/wiki/GBrowse_syn_Help
Configuration	http://gmod.org/wiki/GBrowse_syn_Configuration
Example	http://www.wormbase.org/cgi-bin/gbrowse_syn/
Mailing List	https://lists.sourceforge.net/lists/listinfo/gmod-gbrowse
JBrowse	http://gmod.org/wiki/JBrowse

Ergatis

- Web interface to the TIGR-Workflow engine
- Create, run and monitor reusable computational analysis pipelines
- Manage compute clusters or single machines
- Several pre-configured pipelines
 - including pipelines for prokaryotic and eukaryotic alignments

The screenshot displays the Ergatis web interface for a pipeline. At the top, there is a navigation bar with links: home, pipeline list, new pipeline, rerun, kill, and view xml. Below this, the pipeline path is shown as `/usr/local/projects/bacillus/workflow/runtime/pipeline/829/pipeline.xml`. The pipeline status is **running**, with a start time of Mon Jul 14 08:28:06 2008, an end time of ?, and a last modification of 00 hr 00 min 49 sec. The pipeline ID is 829, the user is jorvis, and the runtime is 54 min 59 sec. The project is bacillus, and the quota information is currently disabled. A link to add a pipeline comment is provided.

The main content area is titled **start** and lists four components:

- component: jaccard.cluster_all**: overall state: **complete** actions: 102, runtime: 53 min 36 sec. Buttons: view, xml, config, update, stop updates.
- component: clustalw.jaccard_cluster_all**: overall state: **running** actions: 607, runtime: 1 min 17 sec. States: complete (240), incomplete (205), pending (108), running (54). Current step: running distributed jobs (iterator i1). Buttons: view, xml, config, update, stop updates.
- component: j_ortholog_clusters.cluster_all**: overall state: **incomplete**. Buttons: view, xml, config, update, stop updates.
- component: clustalw.JOCs_cluster_all**: overall state: **incomplete**. Buttons: view, xml, config, update, stop updates.

The main content area is titled **end**.

Galaxy

- Web portal
 - Search remote resources, combine data from independent queries and visualize results
- Supports several tools for manipulating alignment data
- Queries / workflows can be saved & referenced in papers or rerun later
- Set-theory operations on results
- Links to outside tools, including GBrowse
- Central server or install locally



Using galaxy to perform large-scale interactive data analyses.
Taylor J, Schenck I, Blankenberg D, Nekrutenko A.
Curr Protoc Bioinformatics. 2007 Sep;Chapter 10:Unit 10.5.

Galaxy

Tools

Extract Features

Fetch Sequences

Fetch Alignments

- [Extract Pairwise MAF blocks](#) given a set of genomic intervals
- [Extract MAF blocks](#) given a set of genomic intervals
- [Stitch MAF blocks](#) given a set of genomic intervals
- [Stitch Gene blocks](#) given a set of coding exon intervals
- [MAF Coverage Stats](#) Alignment coverage information
- [Join MAF blocks](#) by Species
- [Filter MAF blocks](#) by Species
- [Filter MAF blocks](#) by Size
- [Extract MAF by block number](#) given a set of block numbers and a MAF file
- [Filter MAF](#) by specified attributes
- [Reverse Complement](#) a MAF file

Get Genomic Scores

Operate on Genomic Intervals

Statistics

Graph/Display Data

Regional Variation

Multiple regression

Outline

- GMOD Project Overview
- GMOD for Comparative Genomics
 - Multiple Sequence Alignment
 - Visualisation: GBrowse_syn
 - Capabilities and future plans
 - Generating & Analyzing: Ergatis & Galaxy
 - Comparative Mapping Visualisation
 - CMap
 - Population Genetics & Natural Diversity
- Resources



CMap: Comparative maps

Web based comparative map viewer

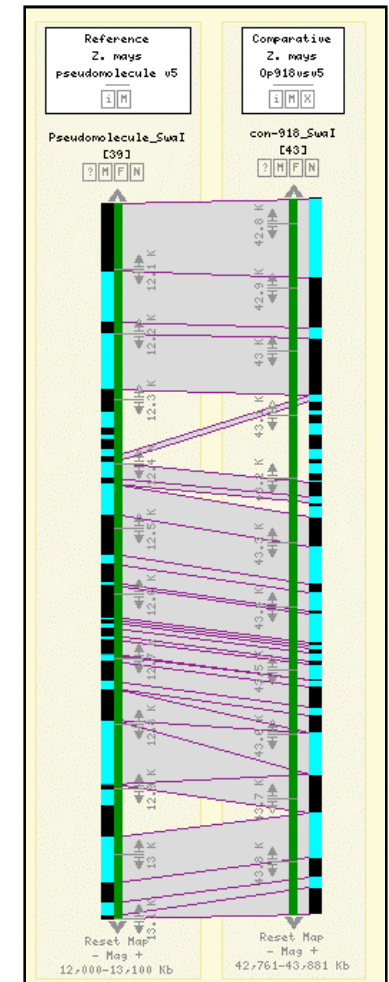
Shows maps and correlations between them

CMap is data type agnostic: Can link sequence, genetic, physical, QTL, deletion, optical, ...

Visualise an arbitrary number of maps

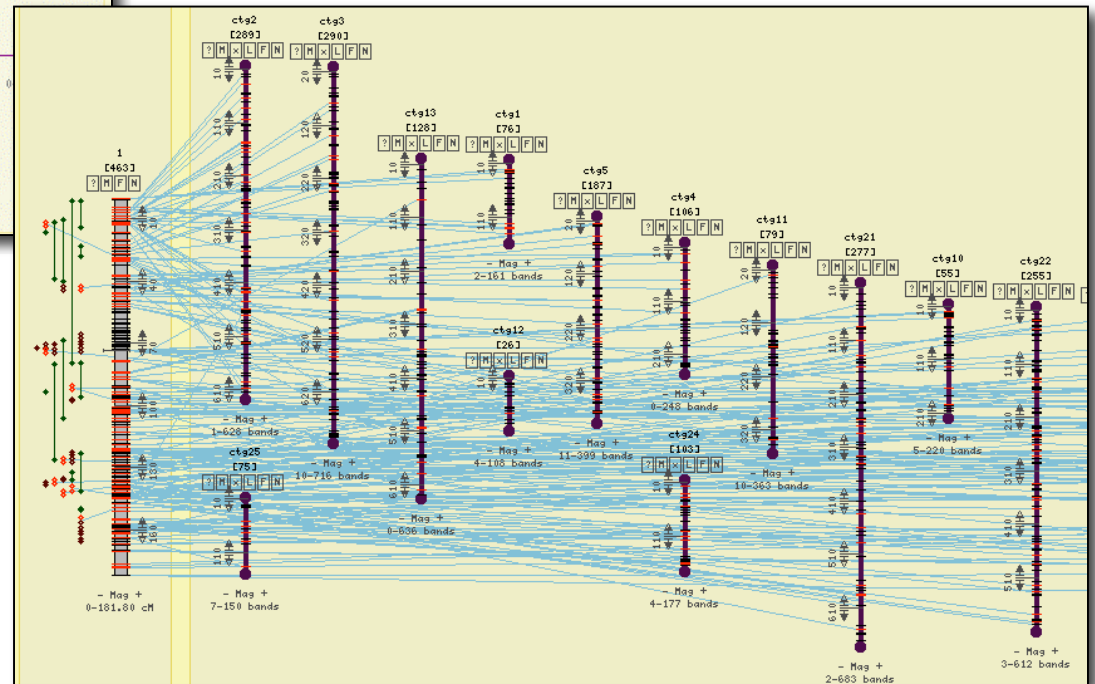
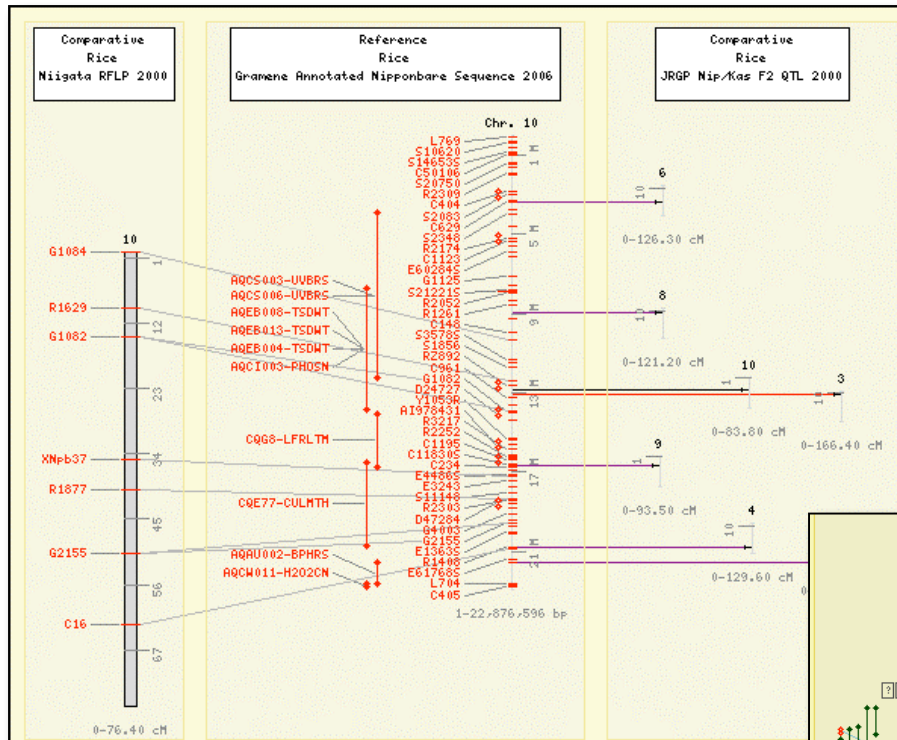
Anything you can put on a line can be shown in CMap

Particularly popular in plant community



CMap

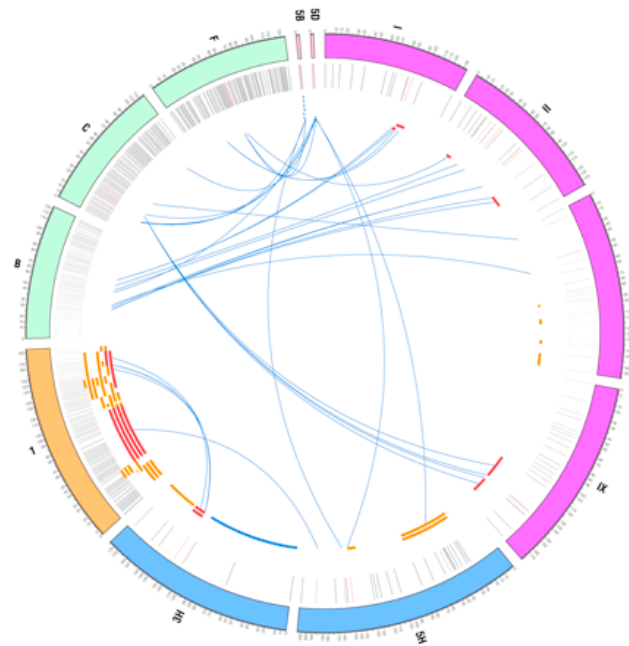
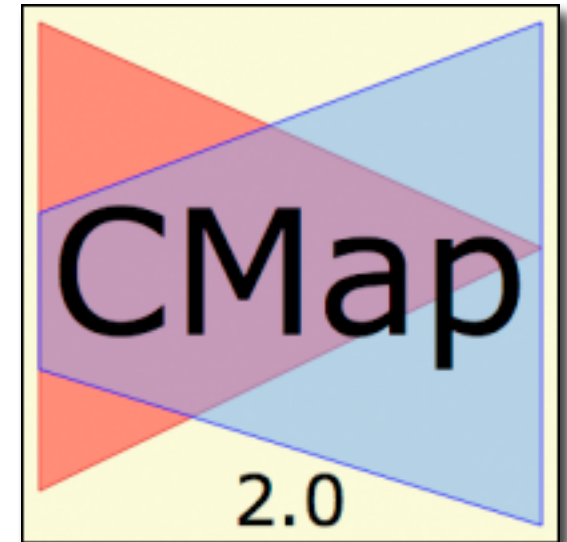
A rice sequence assembly compared against a genetic map (left side) and several QTL maps (right side).



Sequence assembly vs FPC maps

CMap Future Work

- Streamline the database
- Faster access
- Display in SVG
- Save in Circos / MizBee format



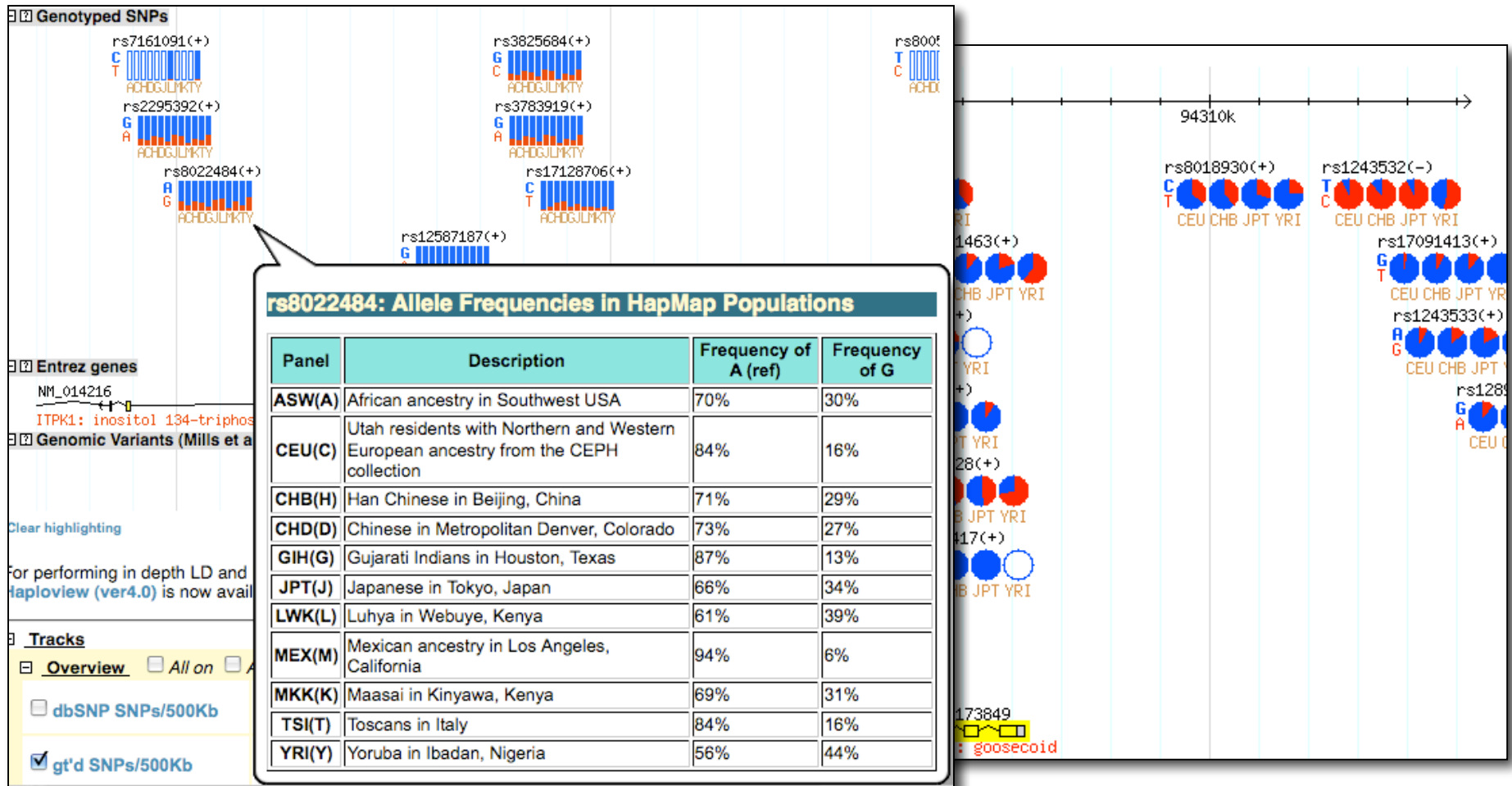
CMap Resources

Home Page	http://gmod.org/wiki/CMap
User Tutorial	http://www.gramene.org/tutorials/cmap.html
Admin Guide	http://gmod.svn.sourceforge.net/viewvc/gmod/cmap/trunk/docs/ADMINISTRATION.pod
Example	http://www.gramene.org/cmap/
Mailing List	https://lists.sourceforge.net/lists/listinfo/gmod-cmap

Outline

- GMOD Project Overview
- GMOD for Comparative Genomics
 - Multiple Sequence Alignment
 - Visualisation: GBrowse_syn
 - Capabilities and future plans
 - Generating & Analyzing: Ergatis & Galaxy
 - Comparative Mapping Visualisation
 - CMap
 - Population Genetics & Natural Diversity
- Resources

HapMap Allele Frequencies



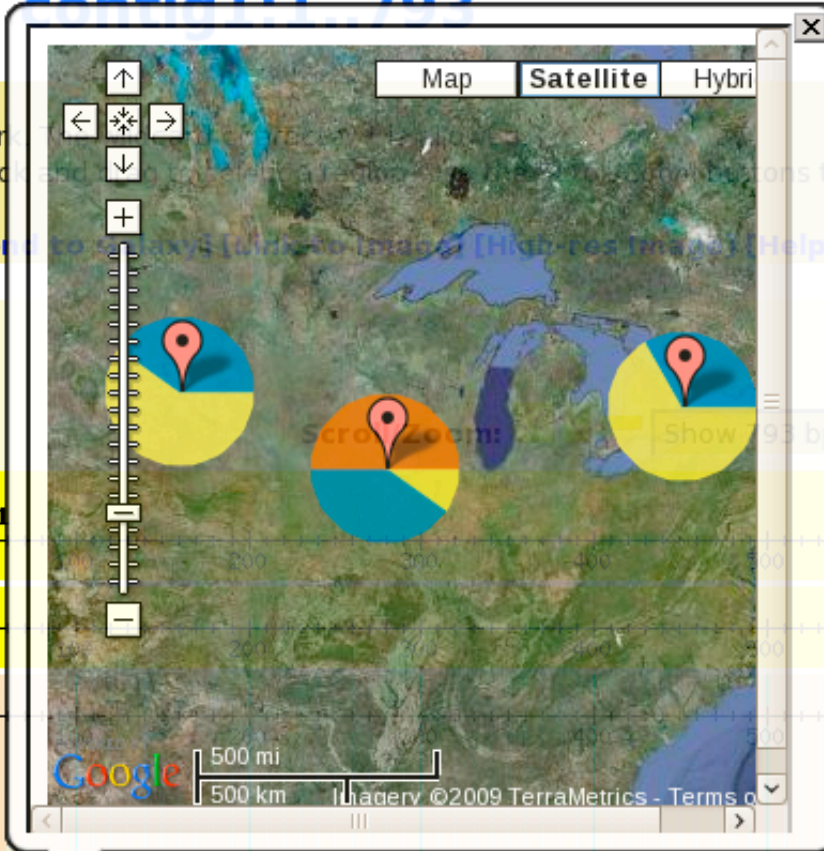
<http://hapmap.org>

Geolocation data

793 bp from contig1:1_793

Name, locus, or other landmark
to center on a location, or click

[Share these tracks] [Send to Galaxy] [Link to Image] [High-res Image] [Help] [Reset] Make an Error



Outline

- GMOD Project Overview
- GMOD for Comparative Genomics
 - Multiple Sequence Alignment
 - Visualisation: GBrowse_syn
 - Capabilities and future plans
 - Generating & Analyzing: Ergatis & Galaxy
 - Comparative Mapping Visualisation
 - CMap
 - Population Genetics & Natural Diversity
- Resources



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
- ...



Log in / create account

page discussion view source history

Welcome to GMOD

GMOD is the **Generic Model Organism Database** project, a collection of open source software tools for creating and managing genome-scale biological databases. You can use it to create a small laboratory database of genome annotations, or a large web-accessible community database. GMOD tools are in use at [many large and small community databases](#).

How do I Get Started?

See [Overview](#) for the big picture. For an introduction to specific GMOD components see the list of the most popular tools at the right, or visit [GMOD Components](#) for a comprehensive list of GMOD tools. If GMOD looks promising for your needs, consider attending the next [GMOD community meeting](#).

How do I Get Support?

GMOD support is available from several different sources. [Support](#) introduces each support option (this web site, [GMOD Mailing Lists](#), [Training and Outreach](#) activities, and the [GMOD Help Desk](#)) and offers guidance on which one is the most appropriate for your question.

How do I Get Involved?

As an open source project GMOD relies on the donation of time and software by groups and individuals. Contribution of new tools, adoption of existing ones, and improving the documentation are all welcome. [Existing](#) and potential users are encouraged to provide feedback via [mailing lists](#) or the [help desk](#). The [GMOD Project Page](#) lists projects in need of ideas and developers. You can also attend project [meetings](#).

Contributing Organizations

[FlyBase](#) [WormBase](#)

Logos: National Evolutionary Synthesis Center, University of Cambridge, American Society of Human Genetics, GMOD 2009 Summer School, Information Systems for Insect Pests

Logos: National Evolutionary Synthesis Center, University of Cambridge, American Society of Human Genetics, GMOD 2009 Summer School, Information Systems for Insect Pests

GMOD News RSS

Visualizing Biological Data Workshop
Openings at Syngenta
Opening at SGD
Apollo 1.11.2 Released
2009 GMOD Survey Results
GMOD Tutorials! GMOD Training!
modENCODE & Gramene Openings
GMOD @ Bioinformatics Australia
InterMine Data Warehouse Workshop
TriPal: A Web Front End for Chado

New & Revised Pages RSS

- Training and Outreach
- GMOD News
- InterMine
- FlyBase Field Mapping Tables
- GBrowse syn Configuration
- GBrowse syn Tutorial
- Meetings
- Sandbox
- GBrowse karyotype
- August 2009 GMOD Meeting

Popular GMOD Tools

Genome Browsing and Editing
GBrowse: Genome annotation viewer
Apollo: Genome annotation editor

Comparative Genomics
CMap: Comparative map viewer

Information Systems for Insect Pests

Mailing Lists

- Several project lists
- Many component-specific lists
- 3100 messages in last 12 months on the 7 lists managed by GMOD staff
- Up 69% from previous year
- Mailing lists are very active



http://gmod.org/wiki/GMOD_Mailing_Lists

GMOD Mailing Lists

This list contains most of the mailing lists relevant to GMOD. This page does not describe some of the lists that automatically deliver every change or *commit* to the GMOD CVS repositories, see [SourceForge](#) for all mailing lists.

Nabble has a GMOD mail mirror of all SourceForge GMOD lists. Nabble provides a much more useful search than does SourceForge.

[Contents](#) [\[hide\]](#)

- 1 Overview Lists
- 2 Component Lists
- 3 Inactive Lists
- 4 A Few Non GMOD Lists

Overview Lists

[\[edit\]](#)

Topic	List Link	Comment
Announcements	gmod-announce	Low volume GMOD announcements. Moderated.
Architecture	gmod-architecture	GMOD architecture working group list. Very low traffic.
GMOD Developers List	gmod-devel	General GMOD developer list.

Component Lists

[\[edit\]](#)

Mailing lists about specific GMOD Components.

Component	List Link(s)	Comment
Apollo	apollo	Apollo mailing list
	gmod-apollo-cmts	Apollo code updates.
Bio::Graphics	gmod-biographics-commits	Bio::Graphics code updates.
BioMart	mart-announce	BioMart announcements mailing list
	mart-dev	BioMart users, developers, code and installation
Chado	gmod-schema	All Chado issues
	gmod-schema-cmts	Chado code updates.
CMap	gmod-cmap	Discussion of CMap development, installation problems, etc.
	gmod-cmap-commits	Notification of SVN activity for CMap.
DIYA	diy-pub-l	DIYA general discussion and support.
	diy-dev-l	DIYA developer discussion.
Ergatis	ergatis-users	Ergatis users mailing list.
	announcement	Ergatis announcements.
	ergatis-devel	Ergatis developers.
Galaxy	galaxy-dev	Discussion and questions regarding local installations and development of Galaxy.
	galaxy-user	General questions and discussion regarding Galaxy usage, especially pertaining to the public sites hosted by the Galaxy Team. Also used for announcements relevant to the Galaxy user community.
GBrowse	gmod-gbrowse	GBrowse users and developers.
	gmod-gbrowse-cmts	GBrowse code updates.
InterMine	dev	InterMine support and development list.
Java TreeView	jtreeview-users	Java TreeView users mailing list.
	jtreeview-devel	Java TreeView developers.
	jtreeview-cvs	Java TreeView code updates.
JBrowse	gmod-ajax	AJAX based genome browser.
MAKER	maker-devel	MAKER developers and users list.
Modware	gmod-ware-users	Modware mailing list.
Pathway Tools	biocyc-users	Pathway Tools users mailing list.
Sybil	sybil-info	Sybil mailing list.
Textpresso		There is no mailing list for Textpresso. Use the Textpresso Feedback Form to ask questions. See the User Guide for help.
Tripal	gmod-tripal	Tripal-related announcements, questions, and requests for help from developers and the community.
Turnkey	turnkey-users	Turnkey user list.
	turnkey-devel	Turnkey development list.
	turnkey-cmts	Turnkey code updates.

Meetings, Training and Outreach

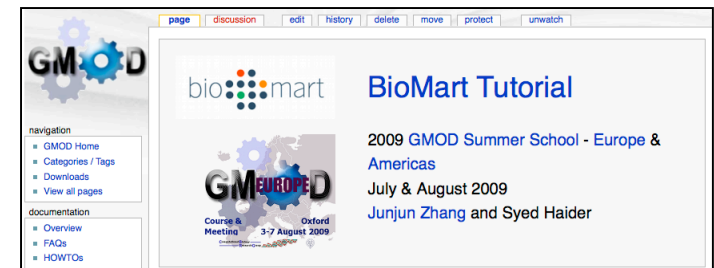
- Semi-annual community meetings
 - Next Meeting:
 - January 2010, San Diego, after PAG
- GMOD Summer Schools
 - 2009
 - July, NESCent, North Carolina, US
 - August, Oxford, UK
 - 2010
 - ??, NESCent, North Carolina, US
 - ??, Asia / Pacific, maybe
- Outreach
 - BA, SMBE, PAG, Arthropod Genomics, ...



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.



⋮



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

Information Systems for Insect Pests

Acknowledgements



Penn State

Anton Nekrutenko

Duke

Yi-Hsin Erica Tsai

Oregon

Patrick Phillips
Phillips Lab

IGS

Joshua Orvis

NESCent

Todd Vision
Hilmar Lapp



Iowa

Ben Faga

CSHL

Sheldon McKay
Ken Youens-Clark

OICR

Scott Cain
Lincoln Stein

ISyIP

Denis Tagu
Dominique Volland
Marie-Noëlle Georgeault



Thank You!



Dave Clements
GMOD Help Desk

US National Evolutionary
Synthesis Center
<http://nescent.org>

clements@nescent.org
help@gmod.org



http://gmod.org/wiki/GMOD_Help_Desk

Information Systems for Insect Pests