



# Other logistical details

Bathrooms

Twitter (#gmod2012 #isb2012)

GMOD future discussion tomorrow

Lightning talks tomorrow

Where for dinner?

# Introductions

Who you are

Where you're from

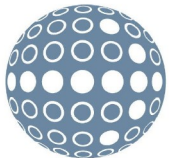
What sorts of things you work on

Anything else you might want to add



# The State of GMOD

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



Ontario Institute  
for Cancer Research

GMOD Meeting  
October 13-14, 2011

# 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

# Where did GMOD come from?

- Established in 2001 by NIH and USDA-ARS.
- Goal: cutting down on costs for new and existing MODs by making reusable software components.
- Initially there was just an FTP site where you could get software.
- Over the past 10 years, produced interoperable components that fill “all” of the needs of a new MOD.



GMOD components can be categorized as

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



# V Visualization

GBrowse

JBrowse

GBrowse\_syn

CMap

Releases

1.70

2.48 (rapid development starting with 2.0 in Jan '10)

Features

Rubberband region selection

Drag and drop track ordering

Collapsible tracks

Popup balloons

Allele/genotype 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](#)



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

# V 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

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>



# 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.

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.

1.2 Natural diversity and more.



# Data Management

Chado

New (2009) web front end for Chado databases  
Set of Drupal modules

Tripal

Modules approximately correspond to Chado modules  
Easy to create new modules

TableEdit

Includes user authentication, job management,  
curation support

BioMart

InterMine



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

# 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

Web interface for creating one off or reproducible analysis pipelines for your genomic data.

MANY applications are already wrapped in.  
Clearly defined API for adding more.

Try it out: <http://usegalaxy.org/>

Install it for yourself: <http://getgalaxy.org/>

Goecks, J. et al. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 2010;11(8):R86. Epub 2010 Aug 25.



# Google Summer of Code

“Genome Informatics” Project:

Reactome

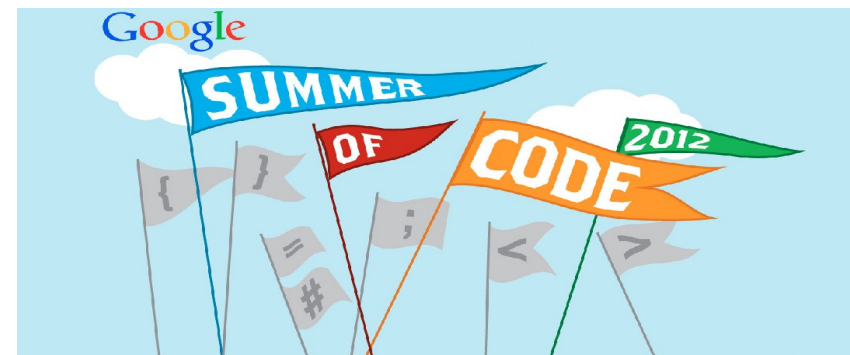
GMOD/JBrowse/GBrowse

Galaxy

WormBase

Initial proposals due TOMORROW!

Final decisions on April 20th



# GMOD in the Cloud

## Two cloud efforts:

### GBrowse2

- Goal to allow easy addition of data, rapid scaling of rendering servers

### GMOD

- Goal to have a “reference implementation” for several GMOD tools that people can “try out” and continue to use if desired.

(There are also Galaxy and Ergatis cloud efforts)





# apt-get GMOD

Thanks to the efforts of debian-med (Olivier Sallou in particular), you can now do this:

```
sudo apt-get install gbrowse
```

```
sudo apt-get install libchado-perl
```

Gets all prereqs (libgd, BioPerl, etc).

Chado not perfect yet; still working with Olivier to get it right.



# NCBI GFF3

NCBI now provides GFF3 for some reference genomes (more on the way I think). From the NCBI Facebook page:

NCBI now offers Reference Sequence (RefSeq) genome annotation files in the latest Generic Feature Format (GFF3) specification (1.20). RefSeq genome data can be downloaded from the genomes area of the NCBI FTP site (<http://bit.ly/gers8b>). GFF3 files are in the GFF directory within each organism directory. Currently GFF3 files are available for the NCBI annotations of the latest assemblies for human (<http://bit.ly/GEoisT>), cow (<http://bit.ly/GIs96D>), dog (<http://bit.ly/GE0yBu>), chicken (<http://bit.ly/GFd6WA>), and many others.



# GMOD in Publications

Potential for a GMOD “Virtual Issue” in the journal “Database”  
Papers accepted can be GMOD software papers, schema papers, user tutorials, organism databases.

Already some candidates in the pipeline (published fall '11):

- Chado and Natural diversity
- Tripal

Other Potential papers:

- GMOD and Chado (needs to be written, but isn't)
- A MOD looking for a pub
- GBrowse and/or Gbrowse\_syn
- Other suggestions?



# GBrowse 2.40-2.49

Lots of bug fixes

Track transparency (allows overlaying multiple plots)

Adding track metadata (allows faceted track selection—  
not released yet)

Improved FastCGI performance, removed mod\_perl  
support

Added “hard limit” for number of features per track

Added config file caching (better performance for really  
big configs)

Can save and restore saved snapshots



## Faceted track management (soon!)

Internet Explorer users: If this page is too slow for you, consider Firefox, Safari or Chrome. Please do not abort script execution.

844: H3K27ac;Y cn bw sp;Adult Female;ChIP-seq (*D. melanogaster*)

Search:

[Browse Worm Tracks](#) [Browse Fly Tracks](#) [Download](#) [View in ModMine](#) [Clear All](#)

**Search Filters**

- Organism** 1 
  - 2 *D. melanogaster*
- Project Category**
- Genomic Target Element**
- Technique**
- Principal Investigator**
- Assay Factor** 2 
  - 1 cad
  - 1 H3K27ac
  - 1 H3K27me3
  - 2 H3K4me1
  - 2 H3K4me3
  - 3 H3K9ac
  - 4 H3K9me3
- Developmental Stage** 1 
  - 2 **Adult Female**
  - 5 Adult Male
  - 1 Embryos 0-12 hr
  - 4 Embryos 0-4 hr
  - 4 Embryos 12-16 hr
  - 4 Embryos 16-20 hr

2 data sets filtered from 1119 originally ([Reset All Filters](#))

Showing items 1-2, Max items:

<input type="checkbox"/>	Dataset	Organism	Technique	Target Element	Assay Factor	Conditions	PI	ID
<input checked="" type="checkbox"/>	H3K27ac;Y cn bw sp;Adult Female;ChIP-seq	<i>D. melanogaster</i>	ChIP-seq	Histone Modification	H3K27ac	Adult Female	White, K.	844
<input type="checkbox"/>	H3K27me3;Adult Female;ChIP-chip	<i>D. melanogaster</i>	ChIP-chip	Histone Modification	H3K27me3	Adult Female	White, K.	346

# Chado 1.20 release

Fixes to several scripts:

`gmod_fasta2gff.pl`

`gmod_bulk_load_gff3.pl`

`load_ncbi_taxonomies.pl`

Fixes so that GFF3 dumping from GBrowse works as expected

Added the ability for the database to “introspect” (chadoprop)



# Chado 1.21-1.23 releases

## 1.21 (October 2011)

- Added rudimentary support for non-public schemas (to better support Tripal)
- Bug fixes related to schema versioning

## 1.22 (November 2011)

- Fixes to make automated schema version migration easier

## 1.23 (Not yet)

- Made perl scripts more “modern” (circa 2004)
- Bug fix to GFF3 preprocessor (splitter/sorter)



# Chado versioning

Addition to the schema: chadoprop table

Additional cv: chado\_properties

SQL diffs (ie, CREATE and ALTER statements)

Makefile target: “make update”

New scripts:

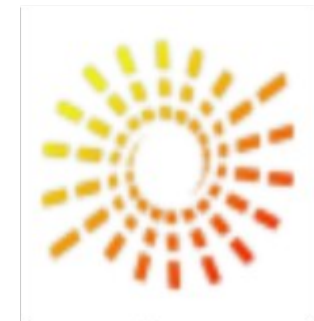
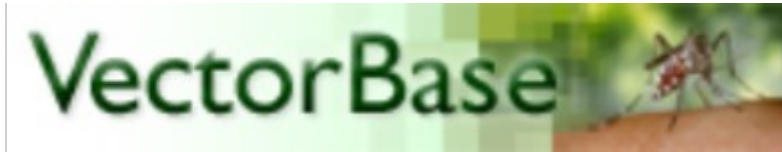
- gmod\_chado\_properties.pl to gather info about schema (specifically version)

- gmod\_update\_chado.pl to update from an older version of Chado (1.0+) to the current schema





# Natural Diversity module: Collaboration



sol genomics network

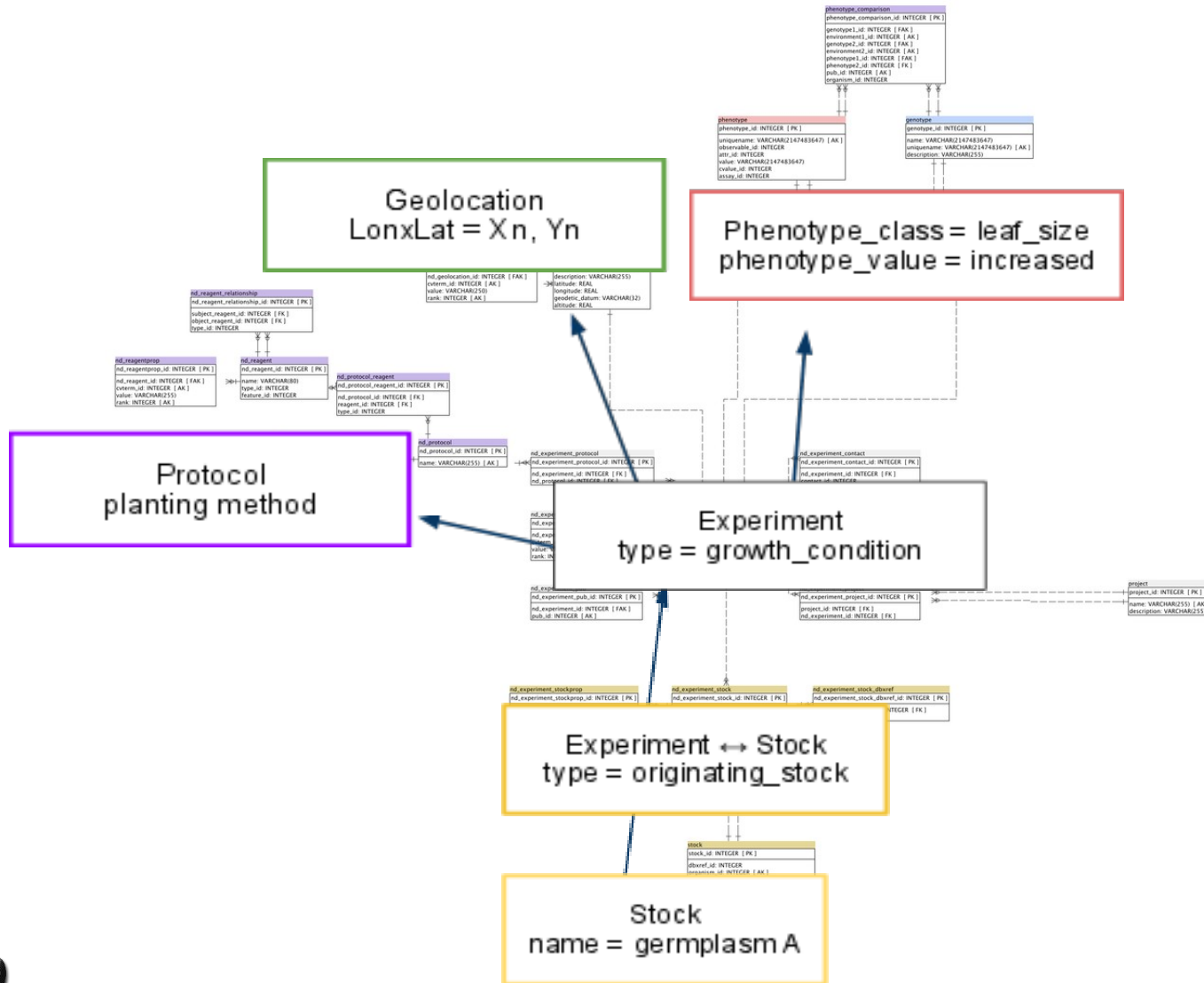








# Use case: rice growth condition



# Outreach

GMOD Help desk (Hiring)

Presence at many meetings/conferences

Genome Informatics, ISMB, PAG, Biology of Genomes

Courses/Workshops

GMOD summer school, Comparative Genomics Workshop, Programming for Biologists, PAG

Wiki, mailing lists



# Acknowledgements

- You! (Remember that part about GMOD being a community?)
- Chris Elsik and Monica Munoz-Torres for their considerable help putting this meeting together
- There are literally too many people associated with GMOD to thank them all, but Lincoln Stein and Dave Clements are a big part of why GMOD is successful.
- Eventbrite
- Funding agencies: NIH and USDA ARS

