Other logistical details

Bathrooms
Twitter (#gmod2014)
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

GMOD Meeting
January 16-17, 2014
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 the fill “all” of the needs of a new MOD.
GMOD components can be categorized as

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

JBrowse

GBrowse_syn

CMap

Releases

1.70
2.55 (slow development in the last year or so)

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

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

JBrowse

GBrowse_syn

CMap

Chado

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.
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

Version 1.0 just released:
- Improved loaders
- New bulk loader
- Complete integration with Drupal Views
- Other good stuff
MAKER
DIYA
Galaxy
Ergatis
WebApollo

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

Web interface for creating one off or reproducible analysis pipelines for your genomic data. MANY applications are already wrappered in. Clearly defined API for adding more.

Try it out: http://usegalaxy.org/
Install it for yourself: http://getgalaxy.org/

Annotation

Annotation editor based on JBrowse
Uses servlet engine, berkeley db
Supports BAM/BigWig
Initial release was December 2012
What are GitC/GiaB

Virtual Ubuntu (Linux) servers with preconfigured GMOD software already installed, either:

- VirtualBox appliances that you can download and use locally.
- Amazon Web Services images (AMIs) that you can clone.

GMOD in the Cloud/in a Box

Installed and configured:

- Chado 1.23
- GBrowse 2.55 (with SAMtools, BigWig)
- JBrowse 1.10.12
- Tripal 1.1
- WebApollo
- Sample data
  - *S. cerevisiae*
  - *P. ultimum*
Potential future additions for GitC/GiaB

- GMOD in iPlant (Atmosphere)
- MAKER and/or Galaxy
- A nice web frontend for uploading/configuring
  - Tripal already has some
  - WebApollo has a command line script
  - WebGBrowse
- All contingent on future funding
apt-get GMOD

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

```bash
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.
Some usage statistics

- When the user runs ./Build install, it prompts the user to send a registration email to me.
- Asks for email address, site, organism, organization, submits to a CGI that collects IP address too.
Some usage statistics

- 9151 (8047) registrations
- 3458 (2868) distinct IP address
- 1437 (1233) distinct “organisms” (though this is free text, and has entries like “labuser”, “Scientists” and “Not Specified”)
- 1669 (1424) distinct organizations (though again free text—every continent represented)
Distinct IP addresses per month since Jan, 2010
When we compared new installs of JBrowse and GBrowse over the same time period, they were just about equal (over the last year or so).
Total Cloud registrations: 112
Total Box registrations: 3

Of course, GitC has been out for a year, GiaB for a few months.
Chado 1.3 or 2.0 release

Some prop and joining tables suggested by Tripal applications.

Possibly the “health” module suggested by FlyBase

– Longer term: recasting Chado as a triple store?
GMOD funding

Funding for my position is drying up this summer.

Need options for funding continued operations (ie, my salary, some AWS, some travel)

I'm open to ideas, but particularly like Rob's idea of a ad track in JBrowse.
Outreach

GMOD Help desk (Yay Amelia!)

Presence at many meetings/conferences
  Genome Informatics, ISMB, PAG, Biology of Genomes, Biocurators

Courses/Workshops
  GMOD summer school, Programming for Biologists, PAG, Malaysia

Wiki, mailing lists
Acknowledgements

- You! (Remember that part about GMOD being a community?)
- There are literally too many people associated with GMOD to thank them all, but Lincoln Stein and is a big part of why GMOD is successful. Amelia did a lot of work setting up this meeting.
- Funding agencies: NIH and USDA ARS