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

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 the fill “all” of the needs of a new MOD.
Software

GMOD components can be categorized as

V Visualization
D Data Management
A Annotation
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/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

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
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](http://jbrowse.org)

Web Sites:
- JBrowse: [http://jbrowse.org](http://jbrowse.org)

Mailing List:
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.
<table>
<thead>
<tr>
<th><strong>Data Management</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chado</strong></td>
</tr>
<tr>
<td><strong>Tripal</strong></td>
</tr>
<tr>
<td><strong>TableEdit</strong></td>
</tr>
<tr>
<td><strong>BioMart</strong></td>
</tr>
<tr>
<td><strong>InterMine</strong></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>MAKER</th>
<th>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.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIYA</td>
<td></td>
</tr>
<tr>
<td>Galaxy</td>
<td>Try it out: <a href="http://usegalaxy.org/">http://usegalaxy.org/</a> Install it for yourself: <a href="http://getgalaxy.org/">http://getgalaxy.org/</a></td>
</tr>
<tr>
<td>Ergatis</td>
<td></td>
</tr>
<tr>
<td>Apollo</td>
<td></td>
</tr>
</tbody>
</table>

Google Summer of Code

“Genome Informatics” Project:
Reactome
GMOD/JBrowse/GBrowse
Galaxy
WormBase

Initial proposals due TOMORROW!
Final decisions on April 20th
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)
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.
NCBI GFF3

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

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!)
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

- VectorBase
- GRAMENE
- NESCent
- GMOD
- GDR
- sol genomics network
A simplified API
Use case: insecticide resistance assay

- **Protocol**: Deltametrin resistance assay
- **Phenotype**: uniquename = 44% mortality
- **Experiment**: type = phenotype_assay
- **Experiment ↔ Stock**: type = used_part
- **Stock**: name = Mbankomo 2010-06b
Use case: mosquito field collection

Geo-location
village = Mbankomo

Protocol
collection method

Experiment
type = field_collection

Experiment ↔ Stock
type = created_whole

Stock
name = Mbankomo 2010-06b
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