Database Tools for Biologists

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

Sponsored by
Alliance for Bioinformatics, Computational Biology, and Systems Biology
Laboratory for Genome Bioinformatics

Texas A&M University
Biochemistry and Biophysics
8 December 2009
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:00</td>
<td>Introduction</td>
</tr>
<tr>
<td>3:10</td>
<td>Software</td>
</tr>
<tr>
<td></td>
<td>Visualization</td>
</tr>
<tr>
<td></td>
<td>GBrowse</td>
</tr>
<tr>
<td></td>
<td>JBrowse</td>
</tr>
<tr>
<td></td>
<td>GBrowse_syn</td>
</tr>
<tr>
<td></td>
<td>Sybil &amp; SynView</td>
</tr>
<tr>
<td></td>
<td>CMap</td>
</tr>
<tr>
<td>4:00</td>
<td>Software, cont.</td>
</tr>
<tr>
<td></td>
<td>Data Management</td>
</tr>
<tr>
<td></td>
<td>Chado, Tripal, GMODWeb</td>
</tr>
<tr>
<td></td>
<td>BioMart and InterMine</td>
</tr>
<tr>
<td></td>
<td>GFF3</td>
</tr>
<tr>
<td></td>
<td>Annotation</td>
</tr>
<tr>
<td></td>
<td>MAKER &amp; DIYA</td>
</tr>
<tr>
<td></td>
<td>Apollo</td>
</tr>
<tr>
<td></td>
<td>Textpresso</td>
</tr>
<tr>
<td></td>
<td>Community Annotation</td>
</tr>
<tr>
<td></td>
<td>Pipelines &amp; Workflows</td>
</tr>
<tr>
<td>4:40</td>
<td>Community</td>
</tr>
<tr>
<td>5:00</td>
<td>Finish</td>
</tr>
</tbody>
</table>
**This Talk**

Algorithms? **NO!**

Plumbing!

---

**INEXACTSEARCH**($W, z$)

**CALCULATE D**($W$)

**return** **INEXTR**($W, |W| - 1, z, 1, |X| - 1$)

**CALCULATE D**($W$)

$k ← 1$

$l ← |X| - 1$

$z ← 0$

for $i = 0$ to $|W| - 1$ do

$k ← C(W[i]) + O^*(W[i], k - 1) + 1$

$l ← C(W[i]) + O^*(W[i], l)$

if $k > l$ then

$k ← 1$

$l ← |X| - 1$

$z ← z + 1$

$D(i) ← z$

**INEXTR**($W, i, z, k, l$)

if $z = D(i)$ then

**return** $\emptyset$

if $i < 0$ then

**return** $\{} [k, l] \}$

$I ← \emptyset$

$I ← I \cup \text{INEXTR}(W, i - 1, z - 1, k, l)$

for each $b \in \{A, C, G, T\}$ do

$k ← C(b) + O(b, k - 1) + 1$

$l ← C(b) + O(b, l)$

if $k ≤ l$ then

$I ← I \cup \text{INEXTR}(W, i, z - 1, k, l)$

if $b = W[i]$ then

$I ← I \cup \text{INEXTR}(W, i - 1, z, k, l)$

else

$I ← I \cup \text{INEXTR}(W, i - 1, z - 1, k, l)$

**return** $I$
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.
# Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:00</td>
<td>Introduction</td>
</tr>
<tr>
<td>3:10</td>
<td><strong>Software</strong></td>
</tr>
<tr>
<td></td>
<td>Visualization</td>
</tr>
<tr>
<td></td>
<td>GBrowse</td>
</tr>
<tr>
<td></td>
<td>JBrowse</td>
</tr>
<tr>
<td></td>
<td>GBrowse_syn</td>
</tr>
<tr>
<td></td>
<td>Sybil &amp; SynView</td>
</tr>
<tr>
<td></td>
<td>CMap</td>
</tr>
<tr>
<td>4:00</td>
<td><strong>Software, cont.</strong></td>
</tr>
<tr>
<td></td>
<td>Data Management</td>
</tr>
<tr>
<td></td>
<td>Chado, Tripal, GMODWeb</td>
</tr>
<tr>
<td></td>
<td>BioMart and InterMine</td>
</tr>
<tr>
<td></td>
<td>GFF3</td>
</tr>
<tr>
<td></td>
<td>Annotation</td>
</tr>
<tr>
<td></td>
<td>MAKER &amp; DIYA</td>
</tr>
<tr>
<td></td>
<td>Apollo</td>
</tr>
<tr>
<td></td>
<td>Textpresso</td>
</tr>
<tr>
<td></td>
<td>Community Annotation</td>
</tr>
<tr>
<td></td>
<td>Pipelines &amp; Workflows</td>
</tr>
<tr>
<td>4:40</td>
<td><strong>Community</strong></td>
</tr>
<tr>
<td>5:00</td>
<td><strong>Finish</strong></td>
</tr>
</tbody>
</table>
Software

GMOD components can be categorized as

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

<table>
<thead>
<tr>
<th>You have</th>
<th>GMOD Has</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence</td>
<td><img src="#" alt="V" /> GBrowse</td>
</tr>
<tr>
<td>Gene models</td>
<td><img src="#" alt="V" /> JBrowse</td>
</tr>
<tr>
<td>Mapping data</td>
<td><img src="#" alt="V" /> CMap</td>
</tr>
<tr>
<td>Alternative transcripts</td>
<td><img src="#" alt="V" /> GBrowse_syn</td>
</tr>
<tr>
<td>Expression</td>
<td><img src="#" alt="V" /> Sybil</td>
</tr>
<tr>
<td>SNP / variation</td>
<td><img src="#" alt="V" /> SynView</td>
</tr>
<tr>
<td>Methylation</td>
<td><img src="#" alt="D" /> Chado</td>
</tr>
<tr>
<td>GO terms</td>
<td><img src="#" alt="A" /> Tripal</td>
</tr>
<tr>
<td>Stocks / lines</td>
<td><img src="#" alt="V" /> GMODWeb</td>
</tr>
<tr>
<td>Publications / Attribution</td>
<td><img src="#" alt="D" /> BioMart</td>
</tr>
<tr>
<td>Orthology</td>
<td><img src="#" alt="D" /> InterMine</td>
</tr>
</tbody>
</table>

- **A**: Annotation
- **D**: Data Management
- **V**: Visualization
GMOD Requirements

• Server
  – Most use Linux

• GMOD Systems Administrator
  – Understands Linux package management, a scripting language, command line interfaces, relational databases, …
  – Grad/Undergrad, half time when starting up.

http://gmod.org/wiki/Computing_Requirements
Overview:
chromosome / contig wide

Region:
intermediate zoom

Details:
current area

Tracks:
current configuration

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

• Uses GBrowse 2

http://www.modencode.org/gb2/gbrowse/fly/
GBrowse Tutorials

• GBrowse User Tutorial at OpenHelix
  – Flash based, has handouts, very snazzy and thorough
  – Great resource for your users

• GBrowse Admin Tutorial
  – HTML based, written by Lincoln Stein, mostly
  – Excellent way to learn how to configure GBrowse

• GBrowse Admin Tutorial w/ VMware Image
  – From the 2009 GMOD Summer Schools
  – Gives you a system to start with

• NGS in GBrowse and SAMtools Tutorial
  – From Bioinformatics Australia 2009, October
  – Gives you a system to start with

http://gmod.org/wiki/GBrowse_Tutorial
GBrowse Future Plans

• Circular genome support
  – Work done by Nathan Liles at Texas A&M
• 2.0, Release in 2010
  – Database and rendering multiplexing
  – Asynchronous track loading
  – GBrowse in the cloud
  – User authentication
• 1.x has a few more maintenance releases left.
# GBrowse Resources

<table>
<thead>
<tr>
<th>Resource</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>WebGBrowse</td>
<td><a href="http://webgbrowse.cgb.indiana.edu/">http://webgbrowse.cgb.indiana.edu/</a></td>
</tr>
<tr>
<td>GBrowse.org</td>
<td><a href="http://gbrowse.org">http://gbrowse.org</a></td>
</tr>
</tbody>
</table>
JBrowse

- GMOD's 2nd generation genome browser
- It's fast
- Completely new
  - Client side rendering
  - Heavily AJAX
  - JSON, Nested Containment Lists

JBrowse Demo

http://jbrowse.org
JBrowse Future Plans

• Tools for migrating from GBrowse
• An ecosystem comparable to GBrowse
  – Glyph library, user defined glyphs, callbacks, track sharing, …
• Comparative genomics (more on that later)
• Community Annotation
  – User authentication
  – User uploadable and sharable tracks and annotation
## JBrowse Resources

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Home Page</strong></td>
<td><a href="http://jbrowse.org">http://jbrowse.org</a></td>
</tr>
<tr>
<td><strong>Getting Started</strong></td>
<td><a href="http://jbrowse.org/code/jbrowse-master/docs/tutorial/">http://jbrowse.org/code/jbrowse-master/docs/tutorial/</a></td>
</tr>
<tr>
<td><strong>Admin Tutorial</strong></td>
<td><a href="http://gmod.org/wiki/JBrowse_Tutorial">http://gmod.org/wiki/JBrowse_Tutorial</a></td>
</tr>
<tr>
<td><strong>Configuration</strong></td>
<td><a href="http://jbrowse.org/code/jbrowse-master/docs/config.html">http://jbrowse.org/code/jbrowse-master/docs/config.html</a></td>
</tr>
<tr>
<td><strong>Demo</strong></td>
<td><a href="http://jbrowse.org/genomes/dmel/">http://jbrowse.org/genomes/dmel/</a></td>
</tr>
<tr>
<td><strong>Mailing List</strong></td>
<td><a href="https://lists.sourceforge.net/lists/listinfo/gmod-ajax">https://lists.sourceforge.net/lists/listinfo/gmod-ajax</a></td>
</tr>
</tbody>
</table>
GBrowse or JBrowse

**GBrowse**
- Robust ecosystem
- Feature rich
- Large and growing user base
- Track sharing

**JBrowse**
- Very fast
- Rapidly growing user base
- Lots of future development
- Easy to configure
GBrowse_syn

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

Example comparing C. elegans to 4 other species at WormBase

Sheldon McKay, Cold Spring Harbor Laboratory
GBrowse_syn

Syntenic blocks do not have to be colinear
Can also show duplications

Sheldon McKay, Cold Spring Harbor Laboratory
GBrowse_syn Future Work

- Integration with GBrowse 2
- High-level graphical overview

- AJAX based user interface and navigation.
  - Submitted grant last month proposing implementing a JBrowse based synteny browser based on GBrowse_syn
# GBrowse_syn Resources

<table>
<thead>
<tr>
<th></th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Configuration</strong></td>
<td><a href="http://gmod.org/wiki/GBrowse_syn_Configuration">http://gmod.org/wiki/GBrowse_syn_Configuration</a></td>
</tr>
<tr>
<td><strong>Example</strong></td>
<td><a href="http://www.wormbase.org/cgi-bin/gbrowse_syn/">http://www.wormbase.org/cgi-bin/gbrowse_syn/</a></td>
</tr>
<tr>
<td><strong>Mailing List</strong></td>
<td><a href="https://lists.sourceforge.net/lists/listinfo/gmod-gbrowse">https://lists.sourceforge.net/lists/listinfo/gmod-gbrowse</a></td>
</tr>
</tbody>
</table>
SynView and Sybil


GBrowse_syn or Sybil or SynView?

**GBrowse_syn**
Most actively developed
Scalable
Familiar interface
Extensive documentation
Growing user community

**SynView**
Scalable
Runs inside GBrowse

**Sybil**
Scalable
Whole genome and
other unique visualizations
Built on Chado
CMap

Web based comparative map viewer
CMap is data type agnostic: Can link sequence, genetic, physical, QTL, deletion, optical, …
Particularly popular in plant community
CMap Future Work

- Streamline the database
- Faster access
- Display in SVG
- Save in Circos / MizBee format
- CMap3D?
### CMap Resources

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Home Page</strong></td>
<td><a href="http://gmod.org/wiki/CMap">http://gmod.org/wiki/CMap</a></td>
</tr>
<tr>
<td><strong>User Tutorial</strong></td>
<td><a href="http://www.gramene.org/tutorials/cmap.html">http://www.gramene.org/tutorials/cmap.html</a></td>
</tr>
<tr>
<td><strong>Example</strong></td>
<td><a href="http://www.gramene.org/cmap/">http://www.gramene.org/cmap/</a></td>
</tr>
<tr>
<td><strong>Mailing List</strong></td>
<td><a href="https://lists.sourceforge.net/lists/listinfo/gmod-cmap">https://lists.sourceforge.net/lists/listinfo/gmod-cmap</a></td>
</tr>
<tr>
<td>Time</td>
<td>Section</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>3:00</td>
<td>Introduction</td>
</tr>
<tr>
<td>3:10</td>
<td>Software</td>
</tr>
<tr>
<td>4:00</td>
<td>Software, cont.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>4:40</td>
<td>Community</td>
</tr>
<tr>
<td>5:00</td>
<td>Finish</td>
</tr>
</tbody>
</table>
Chado: A database schema for biological data

• A schema is a database design
  – Blueprint for a database, a way of organizing data

• Independent of specific data
  – Chado provides structure
  – You provide the hard work and data
Why use Chado?

- Very good at genomic data
- Widely used
  - AphidBase, BeetleBase, dictyBase, FlyBase, SGN, SpBase, VectorBase, wFleaBase, ...
- Integrates with other GMOD tools
- Community of support
- Modular, flexible and extensible
Chado Modules

organism

sequence

mage

generic

companionalysis

general

pub

CV
Controlled Vocabularies and Ontologies

Controlled Vocabulary (CV)
List of terms from which a value must come
Pull down menus are examples of CVs

Ontology
Ontology = CV + rules + relationships between terms
Gene Ontology, Sequence Ontology
Many standard ontologies available from OBO

FlyBase CV Term Viewer:
GO term “tissue regeneration”
CVs and Ontologies in Chado

• Controlled vocabularies and ontologies are key in Chado
• Maximally used for
  – Integrity
  – Interoperability
• Can create your own, but …
  – Please use standard ontologies when they exist
  – See OBO: http://www.obofoundry.org/
Chado Future Developments

Flexibility means core schema changes *slowly*
That's a feature.

• **Natural Diversity module**
  – Better support for phenotypes, crosses, individuals, geolocation, …
  – Based on GDPDM from Cornell University, Terry Casstevens, *et al.* (http://www.maizegenetics.net/gdpdm/)

• **Expression / Anatomy / Cell Fate Atlas support**
  – Aniseed (http://aniseed-ibdm.univ-mrs.fr/) converting to Chado and extending it to better support atlases
  – Will have a web front end for atlases
# Chado Resources

<table>
<thead>
<tr>
<th>Home Page</th>
<th><a href="http://gmod.org/wiki/Chado">http://gmod.org/wiki/Chado</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutorial</td>
<td><a href="http://gmod.org/wiki/Chado_Tutorial">http://gmod.org/wiki/Chado_Tutorial</a></td>
</tr>
<tr>
<td>Introduction</td>
<td><a href="http://gmod.org/wiki/Introduction_to_Chado">http://gmod.org/wiki/Introduction_to_Chado</a></td>
</tr>
<tr>
<td>Modules</td>
<td><a href="http://gmod.org/wiki/GBrowse_Modules">http://gmod.org/wiki/GBrowse_Modules</a></td>
</tr>
<tr>
<td>Mailing List</td>
<td><a href="https://lists.sourceforge.net/lists/listinfo/gmod-schema">https://lists.sourceforge.net/lists/listinfo/gmod-schema</a></td>
</tr>
</tbody>
</table>
Chado Web Front Ends

- Chado is a schema, a server side technology
- It is not a web front end or a desktop client

Options for Chado web front ends:
- Do it yourself
- GMODWeb
- Tripal
Do it yourself

• GMOD provides some support in form of libraries
• Perl
  – Chado::AutoDBI
  – Modware → Bio:Chado:Schema
• Java
  – Two projects under development
GMODWeb

• A Chado specific set of templates for the generic Turnkey web site generation system
• Written in Perl
• Lots of Perl module dependencies

ParameciumDB, a website built with GMODWeb
http://paramecium.cgm.cnrs-gif.fr/

Tripal

- Added to GMOD this year
- Set of Drupal modules
  - Feature, Organism, Library, Analysis
  - Modules roughly correspond to Chado modules
  - Easy to create new modules
- Includes user authentication, job management, and data entry support
- Developed by Clemson University Genomics Institute

MarineGenomics.org

Stephen Ficklin, Meg Staton, Chun-Huai Cheng, …
Clemson University Genomics Institute
<table>
<thead>
<tr>
<th>Tripal Resources</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Home Page</strong></td>
<td><a href="http://gmod.org/wiki/Tripal">http://gmod.org/wiki/Tripal</a></td>
</tr>
<tr>
<td><strong>Tutorial</strong></td>
<td><a href="http://gmod.org/wiki/Tripal_Tutorial">http://gmod.org/wiki/Tripal_Tutorial</a></td>
</tr>
<tr>
<td><strong>Example</strong></td>
<td><a href="http://marinegenomics.org">http://marinegenomics.org</a></td>
</tr>
<tr>
<td><strong>Mailing List</strong></td>
<td><a href="https://lists.sourceforge.net/lists/listinfo/gmod-tripal">https://lists.sourceforge.net/lists/listinfo/gmod-tripal</a></td>
</tr>
</tbody>
</table>
Chado Web: DIY or GMODWeb or Tripal?

GMODWeb
- Complete
- Requires some tuning
- Perl

Tripal
- User authentication
- Data entry
- Actively developed
- Well documented
- Easy to extend
- Drupal

Do It Yourself
- More work
- Get exactly what you want

What really made us decide to switch over to Drupal was that we needed authentication mechanisms, customized data entry mechanisms, and the ability to add social networking features and other non-biological components to our sites. Drupal supported all of this and was widely used, well documented, and well supported.

Stephen Ficklin, CUGI
BioMart and InterMine

• Chado well-suited for setting up organism databases that have
  – Easy to use query interface to support common types of questions
  – Unified, coherent presentation of information

• BioMart and InterMine
  – Allow users to ask complex queries on all data
  – At the expense of having to do more work
GFF3

• The common file format of GMOD for genomic annotation
• Supported by Chado, GBrowse, JBrowse, CMap, Apollo, ....
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:00</td>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>3:10</td>
<td>Software</td>
<td>Visualization: GBrowser, JBrowser, GBrowser_syn, Sybil &amp; SynView, CMap</td>
</tr>
<tr>
<td>4:00</td>
<td>Software, cont.</td>
<td>Data Management: Chado, Tripal, GMODWeb, BioMart and InterMine, GFF3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annotation: MAKER &amp; DIYA, Apollo, Textpresso, Community Annotation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pipelines &amp; Workflows</td>
</tr>
<tr>
<td>4:40</td>
<td>Community</td>
<td></td>
</tr>
<tr>
<td>5:00</td>
<td>Finish</td>
<td></td>
</tr>
</tbody>
</table>
• Genome annotation pipeline for creating gene predictions

• 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

MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes, Brandi L. Cantarel, et al., Genome Res. 2008. 18: 188-196
# MAKER Resources

<table>
<thead>
<tr>
<th>Home Page</th>
<th><a href="http://www.yandell-lab.org/software/maker.html">http://www.yandell-lab.org/software/maker.html</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutorial</td>
<td><a href="http://gmod.org/wiki/MAKER_Tutorial">http://gmod.org/wiki/MAKER_Tutorial</a></td>
</tr>
<tr>
<td>PAG Workshop</td>
<td><a href="http://gmod.org/wiki/MAKER_PAG_2010_Workshop">http://gmod.org/wiki/MAKER_PAG_2010_Workshop</a></td>
</tr>
<tr>
<td>Mailing List</td>
<td><a href="http://yandell-lab.org/mailman/listinfo/maker-devel_yandell-lab.org">http://yandell-lab.org/mailman/listinfo/maker-devel_yandell-lab.org</a></td>
</tr>
</tbody>
</table>
DIYA

• Lightweight, modular, and configurable Perl-based pipeline framework.

• Initial application is gene prediction pipeline for prokaryotes

• Working on integration of Amos assembly tools.

Ergatis

- Web interface to the TIGR-Workflow engine
- Create, run and monitor reusable computational analysis pipelines
- Manage compute clusters or single machines
- Comes with several pre-configured pipelines
Galaxy

- Web portal
  - Search remote resources, combine data from independent queries and visualize results
- Queries / pipelines can be saved and referenced in papers or rerun later.
- Supports set-theory operations on results
- Links to outside tools, including GBrowse
- Can use central server or install locally
Apollo

- GMOD's genome annotation editor
- Add and refine annotations.
- Java desktop client
- Widely used
- Read/write in multiple formats
- Keep track of evidence, curator
- Used in several community annotation efforts
Apollo Future Work

• Berkeley Bioinformatics Open-source Projects (BBOP)
  – Current developers of Apollo
  – Submitted a grant proposal for
    • Apollo on the web
    • Using same underlying tools as JBrowse

• Meanwhile, CCG/ABF
  – Is using Apollo (and Chado) for genome annotation
  – ABF is exploring the possibility of developing a web-based application to complement Apollo
  – NCRIS 5.1 funding for a 6 month project

• These two groups are talking to each other
## Apollo Resources

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutorial</td>
<td><a href="http://gmod.org/wiki/Apollo_Tutorial">http://gmod.org/wiki/Apollo_Tutorial</a></td>
</tr>
<tr>
<td>Mailing List</td>
<td><a href="http://mail.fruitfly.org/mailman/listinfo/apollo">http://mail.fruitfly.org/mailman/listinfo/apollo</a></td>
</tr>
</tbody>
</table>
Textpresso

- Text mining system for scientific papers
- Analyzes full article text
- Indexes articles by keywords and by category tags
- Stand alone search engine w/ web interface
- Curation tool


Textpresso for *E. coli*, extensions by Nathan Liles, Hu Lab
Community Annotation

• How do you get others to contribute?

• Social:
  – Sticks
    • Work well if your database is already the authority on your topic/organism and you have curators and a huge community
  – Carrots
    • Give people credit
    • Give people ownership
    • Seek mutually beneficial relationships
  – Comfort Level
    • Recent popularity of social computing
Community Annotation

• Technological
  – Make it easier to fix something then it is to be irritated by its error or absence.
    • The Wikipedia model.
  – Make it relatively easy for people who really care to contribute significant content
    • Also the Wikipedia model.
Community Annotation: GMOD Technology

• Apollo
  – Several projects use Apollo to distribute genome annotation efforts
  – Apollo infrastructure supports:
    • Read from Chado → Save to XML → Review → Upload to Chado
  – But
    • Java application; Infrequent Apollo users forget a lot.
    • Web Apollo will help some, maybe a lot

• Tripal
  – Supports update interfaces for data in Chado databases.
  – Has access to all of Drupal's social networking.
Community Annotation: GMOD Technology

- **Table Edit**
  - A MediaWiki extension that provides a GUI interface to updating MediaWiki tables.
  - MediaWiki software used at Wikipedia.
  - Has been extended to update and render database tables through a MediaWiki interface.
  - Work is in progress to apply it to Chado.
  - See http://ecoliwiki.net
  - Has potential to turn Chado into a wiki.

Jim Hu, Daniel Renfro, et al., Texas A&M
## Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Section</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:00</td>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>3:10</td>
<td>Software</td>
<td>Visualization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GBrowser</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JBrowser</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GBrowser_syn</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sybil &amp; SynView</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CMap</td>
</tr>
<tr>
<td>4:00</td>
<td>Software, cont.</td>
<td>Data Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chado, Tripal, GMODWeb</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BioMart and InterMine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GFF3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annotation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MAKER &amp; DIYA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Apollo</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Textpresso</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Community Annotation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pipelines &amp; Workflows</td>
</tr>
<tr>
<td>4:40</td>
<td>Community</td>
<td></td>
</tr>
<tr>
<td>5:00</td>
<td>Finish</td>
<td></td>
</tr>
</tbody>
</table>
Who uses GMOD?

Plus hundreds of others
GMOD Project

• Open Source

• Two full time project staff:
  – Project Coordinator: Scott Cain
  – Help Desk: Dave Clements

• Components
  – Some have dedicated funding
  – Others are contributed
  – New components must have:
    • An open source license
    • Interoperability with other GMOD components
    • A good faith commitment of at least 2 years of support
A wiki, of course. GMOD.org is the hub for all things related to the project:

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

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

...
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
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
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:00</td>
<td>Introduction</td>
</tr>
<tr>
<td>3:10</td>
<td>Software</td>
</tr>
<tr>
<td></td>
<td>Visualization</td>
</tr>
<tr>
<td></td>
<td>GBrowser</td>
</tr>
<tr>
<td></td>
<td>JBrowser</td>
</tr>
<tr>
<td></td>
<td>GBrowser_syn</td>
</tr>
<tr>
<td></td>
<td>Sybil &amp; SynView</td>
</tr>
<tr>
<td></td>
<td>CMap</td>
</tr>
<tr>
<td>4:00</td>
<td>Software, cont.</td>
</tr>
<tr>
<td></td>
<td>Data Management</td>
</tr>
<tr>
<td></td>
<td>Chado, Tripal, GMODWeb</td>
</tr>
<tr>
<td></td>
<td>BioMart and InterMine</td>
</tr>
<tr>
<td></td>
<td>GFF3</td>
</tr>
<tr>
<td></td>
<td>Annotation</td>
</tr>
<tr>
<td></td>
<td>MAKER &amp; DIYA</td>
</tr>
<tr>
<td></td>
<td>Apollo</td>
</tr>
<tr>
<td></td>
<td>Textpresso</td>
</tr>
<tr>
<td></td>
<td>Community Annotation</td>
</tr>
<tr>
<td></td>
<td>Pipelines &amp; Workflows</td>
</tr>
<tr>
<td>4:40</td>
<td>Community</td>
</tr>
<tr>
<td>5:00</td>
<td>Finish</td>
</tr>
</tbody>
</table>
Acknowledgements

NESCent
Todd Vision
Hilmar Lapp

BBOP
Ed Lee

OICR
Scott Cain
Lincoln Stein

Oregon
Patrick Phillips
Phillips Lab

CBRG, Oxford
Simon McGowan

CUGI
Stephen Ficklin

CSHL
Sheldon McKay
Ken Youens-Clark

Texas A&M
Jim Hu
Daniel Renfro
Nathan Liles
Brenley McIntosh

Broad
Heng Li
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
SAMtools

Platform neutral set of programs and file formats specifically for short reads.

GBrowse for Population Genetics

Shows
Where we looked
Allele & genotype frequencies
By population
Individual genotypes

Could also show:
• Frequency by phenotype or any other characteristic
• Sliding window stats
HapMap Allele Frequencies

http://hapmap.org
Geolocation data