



GMOD Community Meeting

July 16-17, 2008

University of Toronto
Toronto, Ontario, Canada

The July 2008 GMOD community meeting was held on July 16-17, 2008 at the University of Toronto, immediately before BOSC and ISMB 2008 (also in Toronto), and just a few days after the 2008 GMOD Summer School. The meeting was attended by over 30 people representing more than 20 different groups.

Agenda

July 16

| Time | Topic | Who | Presentation |
|-------|---|-----------------------------|-----------------|
| 9:00 | Introductions | <u>Scott Cain</u> | |
| 9:30 | The State of GMOD | Scott Cain | <u>PPT</u> |
| 10:10 | break | | |
| 10:40 | MediaWiki/ <u>TableEdit</u> Roundtripping | <u>Jim Hu</u> | |
| 11:00 | More MediaWiki enhancements | <u>Sheldon McKay</u> | <u>Links...</u> |
| | <u>SGN</u> Community Annotation | Lukas Mueller | <u>PDF</u> |
| | WikiMods & Chado API | Brad Arshinoff | <u>PDF</u> |
| 11:30 | Lunch | | |
| 1:30 | <u>GMOD Help Desk</u> | <u>Dave Clements</u> | <u>PDF PPT</u> |
| 2:15 | Rearchitecting <u>Apollo</u> and the need for a database independent Biological API layer | Ed Lee | |
| 2:50 | break | | |
| 3:20 | <u>InterMine</u> and <u>Chado</u> | Richard Smith | <u>PDF</u> |
| 3:50 | Show and Tell | "What I did with my Summer" | |
| | <u>CMap</u> | <u>Ben Faga</u> | |

July 17

| Time | Topic | Who | Presentation |
|-------|--|---------------------------|--------------|
| 9:00 | New things for <u>GBrowse</u> 1.69 | <u>Sheldon McKay</u> | |
| | GBrowse 2.0 and Roadmap | <u>Lincoln Stein</u> | |
| 9:30 | New things for GBrowse 3.0 | Ian Holmes | <u>PDF</u> |
| 10:00 | break | | |
| 10:30 | The need for a computable common gene page (<u>Don Gilbert's proposal</u>) | Scott Cain, Lincoln Stein | <u>PPT</u> |
| 11:30 | Lunch | | |
| | More Show and Tell or a mini hackathon or go see Toronto | | |
| | Traits at SGN | Lukas Mueller | |
| 1:30 | CellFrame | Yunchen Gong | |
| | Matching Gene Names to Articles at Xenbase | Jeff Bowes | |
| | <u>Django and Chado</u> - A user interface exploration | Victor de Jager | |

Attendees



1. David Arcoleo - BeeSpace, University of Illinois
2. Brad Arshinoff - XanthusBase
3. Jeff Bowes - Xenbase
4. Robert Buels - Sol Genomics Network (SGN)
5. Scott Cain - GMOD
6. Dave Clements - NESCent, GMOD
7. Sean Davey - BirdBase, U of Arizona
8. Victor de Jager University of Nijmegen, The Netherlands & Centre for Molecular and Biomolecular Informatics
9. Mary E Dolan Mouse Genome Informatics, The Jackson Laboratory
10. Ben Faga - CSHL
11. Yunchen Gong - University of Toronto
12. Josh Goodman - FlyBase
13. Todd Harris - WormBase
14. Chris Hemmerich - Center for Genomics and Bioinformatics, Indiana U.
15. Ian Holmes - UC Berkeley
16. Jim Hu - EcoliWiki, Texas A&M
17. Thomas Keane - Wellcome Trust Sanger Institute

18. Ed Lee - BBOP and Apollo
19. Suzi Lewis - BBOP
20. Margie Manker - The Centre for Applied Genomics, Toronto
21. Sheldon McKay - modENCODE, WormBase
22. Lukas Mueller Sol Genomics Network (SGN)
23. Brian O'Connor - UCLA
24. Joshua Orvis - Institute for Genome Sciences, University of Maryland
25. Barry Sanders - BeeSpace, University of Illinois
26. Stéphanie Sidibe Bocs - CIRAD
27. Richard Smith - InterMine (and InterMine)
28. Kevin Snyder - Xenbase; University Of Calgary
29. Jason Stajich - UC Berkeley
30. Haiyan Zhang - FlyBase
31. Junjun Zhang - The Hospital for Sick Children, Toronto



GMOD Components

This section covers discussion about the software components in GMOD. For a summary of talks and discussion on how those components are used at particular databases, see the GMOD User Community section.

Chado

Scott Cain spoke on Chado.

The GMOD 1.1 release is in the works. There are no schema changes yet.

Companalysis Module

Joshua Orvis requested better typing / the use of controlled vocabularies in the Chado Companalysis Module to better represent scores that are currently in the analysisfeature table. Without it there is no way to keep track of what the scores mean. This issue was also raised by Brett Whitty at the 2008 GMOD Summer School the week before.

Also, Joshua (again) proposed the addition of a type_id field to the analysisfeature table. The use case for this is to allow the distinction between types of features involved in an analysis. The most direct examples are 'input_of' and 'created_by' which allow the user to perform queries of a features role in the analysis. This has been brought up in previous meetings and in the GMOD mailing list and seems to have had general approval.

Action Items

- get Chris Mungall's input on these issues.

Natural Diversity Module

Dave Clements discussed the Chado Natural Diversity Module. It was developed at NESCent and NCSU to enable Chado to better support natural diversity studies. This has been laying dormant for a while and it would be nice to get it in use by more groups so that we can better generalize it and make it an official Chado module. Lincoln pointed out there was grant money to do exactly this.

Action Items:

- Dave will follow up with Lincoln.

Community Annotation System

Scott Cain also spoke about his work on the Community Annotation System (CAS). The next release of CAS, 1.1, will feature

- Switch from Ubuntu to CentOS.
- GMODWeb will be included.
- cas-utils 0.1

cas-utils 0.1

cas-utils is a set of tools that tie together GBrowse, Apollo and Chado. This includes

- A CGI for selecting a region in GBrowse, extracting the data for that region from Apollo and creating an XML file and a jnlp (webstart) file for Apollo.
- A CGI for accepting uploads of edited XML files to either be immediately loaded into Chado or to be held for validation.
- A configuration Perl module to make modifying the CGI's behavior easy.
- A Module::Build based installer that queries the user for needed setup data.

cas-utils is now available for download.

TableEdit

Jim Hu spoke about progress on TableEdit, currently at release 0.8.

- Now refuses edits until user entered HTML tags are closed, thus avoiding nasty side effects.
- Round trip between MediaWiki and Chado is not yet done.
 - ◆ Chado to MediaWiki is done, vice versa is not.

Apollo

Ed Lee, lead developer for Apollo spoke about enhancements to Apollo that have happened since he started working on it last September:

- undo function
- preferences editor
- Chado adapter enhancements
- Improved graph and GFF3 support

InterMine

Richard Smith spoke about InterMine, a query optimized data warehouse system for biological data. Has the ability to create precomputed tables (a la materialized views) at any time (and do this from the GUI) in response to popular query patterns. Also supports query templates, which are fill-in-the blank versions of popular queries.

InterMine is written in Java. It has one class per Sequence Ontology (SO) term, and use Java class inheritance for *is_a* relationships. *part_of* relationships are implemented with Java references and collections.

CMap

Ben Faga gave a talk on what's new in CMap. Some highlights:

- CMap 1.0 came out in March 2008.
- ribbon displays for syntenic blocks
- dotplot displays
- new feature glyphs
- embeddable image generation
- directory guessing for easier installation.

GBrowse

Three talks gave us the GBrowse roadmap. Talks covered the next incremental release (1.69), and the next two major releases (2 and 3).

GBrowse 1.69

Sheldon McKay and Lincoln Stein spoke about recent enhancements to GBrowse. These features are available in the current development version ("stable") of GBrowse and will be included in the upcoming (some would say imminent) 1.69 release of GBrowse.

- Wiggle - Dense quantitative tracks, density can have colored peaks, and go below 0.
- Quantitative (BP resolution) data.
- Inline track configuration.
- Design Primers
- Rubberbanding
- Popup windows (with a nice example showing WormBase anatomy cartoons)
- Draggable tracks

- Easy-share tracks
 - ◆ DAS server is inside GBrowse. Also a web service.
 - ◆ Can now have one GBrowse server share a track with another GBrowse server.
 - ◆ Data is transferred on the fly, as the user navigates the genome.
 - ◆ Can form chains of sharing.
- Galaxy Integration.
 - ◆ Within Galaxy click on the get data link.
 - ◆ Lists data sources including BioMart and WormBase GBrowse.
- Multiple Alignment Format (MAF) and conservation tracks.

GBrowse 2

Lincoln Stein talked about GBrowse 2, the next major release of GBrowse. This release focuses on performance and stability. GBrowse 2 will be cluster aware:

- Tracks can be assigned to read data from specific data servers, and render tracks using specific render servers.
- Assignment of machines as data and/or render servers is configurable.
 - ◆ A server can be a data server or a render server or both,
 - ◆ A track may have multiple data and render servers.
 - ◆ A single node can serve data and rendering for one or more tracks.
- Tracks loaded with AJAX. Grayed out until they load.
- Turning tracks on and off no longer requires a reload.

Our experience is that the database is usually the bottleneck with existing GBrowse installations.

- Can also enable editing of feature comments.

GBrowse 3

GBrowse 3 was renamed JBrowse after this meeting.

Ian Holmes presented his group's work on GBrowse 3, a complete rewrite of GBrowse using a Web 2.0 style interface. Mitch Skinner has done most of the coding work on this.

Most tracks are now rendered in client using JavaScript. Tracks such as wiggle tracks can also still be rendered on the server.

GBrowse 3 uses nested containment lists to quickly determine what features to display. These are 5 to 500x faster than R-trees. The group is using the modENCODE project as a target test audience.

Ian made the observation that when you are asking for guidance on GUIs, you need large sample sizes. Small sample sizes lead to a large set of suggestions with very little overlap between users. Large sample sizes enables you to identify a core set of requests.

Genome Wiki

Ian would like to move GBrowse 3 in the direction of being a *genome wiki*

- Upload tracks and track sharing
- Ability to add comments, ratings, ...
- Requires user management

Genome Wiki is about people sharing tracks, not so much about individual genes.

Approximate Schedule

- 2008?: A Lightweight AJAX Genome Browser
- 2009?: An AJAX Genome Wiki

They are not currently working on a Chado adaptor. They hope to do that, but probably not soon.

GBrowse Glyphs Page

At the 2008 GMOD Summer School there were several requests for a GBrowse glyphs page that

- shows what the glyphs look like,
- what track you might use it with,
- links to any other documentation on the glyph.

Lincoln believes that there is already similar documentation in the GBrowse distribution.

Action Items:

- Dave will investigate further.

Common Gene Page

Scott Cain eerily yet effectively channeled Don Gilbert on the topic of a Common Gene Page.

This not the gene page that people see when they come to your web site. Rather, it is some minimal set of information about a gene in your organism, stored in XML format, that can be easily accessed and parsed by other organizations. It is meant to enable easy sharing of information about genes between GMOD users.

If you've been around GMOD for a while you know that the concept of a common gene page is almost as old as GMOD itself. We *might* have actually moved forward on this at this meeting.

There was discussion on what should be included in the gene page. The consensus was to keep track of only the minimal amount of information, See Scott's presentation for the list we settled on.

Uniprot XML may be suitable for this.

Now What?

Lincoln proposed a CGI script that has a set of predefined hooks for populating the XML. This could be a Perl program with methods for fetching data and then passing it to another routine for placing the data into an XML format. Each organization would write the classes called by the hooks to get the data from wherever they keep it. Provides a framework that can be used across multiple organizations and that will always produce structurally identical XML, no matter how it is originally stored.

Rob Buells from [SGN](#) produced a prototype of this program while at the meeting.

Action Items:

- ?

Gene Wiki

We also discussed the [Gene Wiki project](#). This project has created around 7,000 human gene pages in Wikipedia. Wikipedia asked

- Only *interesting* genes have pages. Interesting was defined as any gene with at least one PubMed reference.
- The pages be easy to edit. Moved some nasty tables to the end of the page.

Someone might eventually be able to create a MODGeneWiki from GMOD Common Web Pages.

MediaWiki Enhancements

[Sheldon McKay](#) spoke about MediaWiki related work he's been doing for the [modENCODE project](#).

FCKEditor

[FCKEditor](#) is a WYSIWIG editor for MediaWiki, but if you use it off the shelf it becomes hard for your users to use any other editor, including the default MediaWiki editor, which they may already be familiar with. Sheldon has extended FCKEditor to make it optional. Users now see "edit" "rich edit" links and tabs.

Action Items:

- Dave will investigate FCKEditor and the modified version for use in the GMOD web site.

Popup Balloons

Sheldon has also created an extension for creating popup balloons in a MediaWiki Web Site. See [Popup Balloons](#) for details. This extension is installed on the GMOD web site.

Collapsible Sections Extension

Does what it says - enables users to collapse and expand sections on pages in MediaWiki.

Predefined Page Creation Extensions

A set of extensions were created to

- automatically populate pages based on what type of page is being created
- generate forms to help users fully populate pages with required fields.

These use the Yahoo autocompletion library.

APIs

Perl based Schema Abstraction Layer for Chado

Brad Arshinoff from [XanthusBase](#), (soon to be [WikiMods](#), see below) gave a talk titled *Perl based Schema Abstraction Layer for Chado*. Brad's [talk](#) gave an overview of a [Perl middleware](#) package for [Chado](#) that was developed at XanthusBase.

Discussion

Q: [Modware](#) is a Perl-based Chado API that already exists. Why not use it?

A: Thought this would be less work and a lot less [SQL](#) than Modware. May or may not have worked out that way.

Eric Just, the developer of Modware, is no longer at [DictyBase](#). Someone has replaced him, but we don't know if that person is supporting Modware.

It seems that we have a lot of Perl and Java APIs to Chado, perhaps too many. What should we do about that? [Lincoln Stein](#) suggested that we document them all and provide a list of pros and cons for each. That will allow new users to make the best informed choice about what they want to do.

Action Items:

- Dave will create a Chado APIs page.
- Dave will work with Brad to make the middleware available and documented.
- Dave will Contact Eric and/or DictyBase about the status of Modware.
 - ◆ 2008/08 - Done. Modware is actively being worked on by DictyBase staff.

Chado Java API

Ed Lee presented a talk on the need for a Java interface to to the Chado schema. He's going to be rewriting the Apollo data model to clearly define biological concepts and to map well to any of Apollo's potential data sources, including Chado.

This could be a way to enforce/encourage [Chado Best Practices](#). A current problem for tool developers (such as the Apollo team) is writing code to work with Chado, when not all Chado users represent the same biological concepts in similar ways.

Having a cleanly designed, biological level (as opposed to DBMS table level) API for Java would help organizations follow best practices when using Chado. It also would make tool development much easier.

GMOD User Community

SGN



Lukas Mueller from SGN spoke about community annotation at the Sol Genomics Network.

- SGN does annotation on genotype and phenotype.
- Have about 60 community annotators.
- ~130 loci have been edited at least once by community members.
- Easy to use interface. Updates go directly to main database.
- Have assigned some entire gne families to people.
- Lukas actively recruits volunteer editors at meetings.

Lukas also talked about SGN's traits (phenotypes) database. SGN uses a custom database design for their phenotypic data. (They do not use the Chado Phenotype Module. Suzi Lewis indicated that her group is working on a new phenotype module for Chado which will address issues with the current design.)

WikiMods.org

Brad Arshinoff from XanthusBase, introduced the WikiMods web site, a collection of MODS for prokaryotes with small research communities. This will replace the existing XanthusBase site and add an additional organism in the process. It is scheduled to launch on July 30 2008 with these sites:

- <http://xanthus.wikimods.org>
- <http://xeno.wikimods.org>
- <http://demo.wikimods.org>

They have migrated Chado from Oracle to MySQL.

CellFrame

Yunchen Gong gave a talk about CellFrame, a web site about cell biology and construction of cell perturbation networks

Xenbase

Jeff Bowes of [Xenbase](#) talked about automatic loading, linking, and indexing publication abstracts. Xenbase downloads information for every *Xenopus* related publication. The abstract is then scanned for gene names/symbols and other controlled vocabulary terms. The publication is then associated with those terms and genes in Xenbase.

Xenbase has extended the schema to support this indexing scheme and uses DB2 Net Extender for indexing (but any indexing tool could be used). Xenbase also scrapes images from each journal they have an agreement with. They use a Java class for journals, and every journal has its own subclass.

Centre for Molecular and Biomolecular Informatics

Victor de Jager of the University of Nijmegen and the Centre for Molecular and Biomolecular Informatics, gave a talk on using the Django web framework with Chado (see [Chado Django HOWTO](#) for more). A Django based web site could be layered on top of the BioObjects proposed by Ed Lee in his talk.

GMOD Project

Google Summer of Code

Last year a Google Summer of Code student worked with Lincoln, and Hilmar Lapp (at NESCent) on a Google Summer of Code project to add phylogenetic information to GBrowse. Lincoln and Hilmar liked it enough that they recommend the program. Lincoln cautions that it is a lot of work to be a mentor in the program.

Action Items:

- Dave will investigate further and encourage the GMOD community to participate in the program during the summer of 2009.

Packages

At the end of the [GMOD Help Desk talk \(see below\)](#), Dave asked for what else he should be working on. The number one response was creating GMOD packages that could be installed with Linux package installers.

Everyone agreed this was an excellent idea, and that it was hard to do, particularly to keep the packages up to date for all the distributions you want to support. BioPackages.net would be the place to put them, if we did this.

Lincoln mentioned that there are 1 year infrastructure grants for this sort of thing. That would get us where we want to be for a year, but not after that.

Action Items:

- No solution or action item was settled on.

GMOD Help Desk

Dave Clements gave a talk on his first 10 months at the GMOD Help Desk, and what he is planning doing in the coming months.

What's Been Done

- Web Site
- Outreach - Posters, talks, representation, and promotion
- Education
 - ◆ Chado Workshop with Scott @ Arthropod Genomics, ~35 people
 - ◆ 2008 GMOD Summer School with Scott, Ed, Ben, 25 students for 2 1/2 days
- Components
 - ◆ MAKER now part of GMOD.
 - ◆ Phenote is on its way.

What's Planned

GMOD User Directory

Planning to TableEdit to make parts of the GMOD web site be database driven. Plan on having the same core set of data and a web page for each user. The core data set will describe what components they use and how, and be implemented in TableEdit tables. We'll then be able to use that information to also show which users use a component on each component page, as well as a complete list of users.

This is a continuation of the *community portal* idea that was started in the past 10 months. This will help new and existing users get a handle on who is using which components for what kind of biology.

User Experience Logs

We can't possibly describe or maintain HOWTO pages for all possible combinations of operating system (in all their versions), external software (BioPerl, Java, libgd,... - in all their versions), and GMOD Components (in all their possible versions and combinations).

However, if we made it easy for GMOD users to record their experiences installing whatever combination they are using then that might be a useful approximation. New users would then be able to find several possible workarounds when they, for example, can't get libgd to work. Maybe one of the workarounds will even be for there Linux distribution.

We already have several such logs on the web site.

Dave will create a plan for

- organizing user logs in the web site,
- making it easy to do so, and
- encourage users to do this.

Documentation

- Components
 - ◆ Chado Documentation Reorganization
 - ◇ Chado API doc
 - ◆ GBrowse doc, including cookbook and glyphs page.
 - ◆ Community Annotation System
 - ◆ TableEdit
 - ◆ Better document (and encourage tighter) integration between Galaxy, InterMine, BioMart and the other GMOD components.
 - ◆ Tutorials - screencasts for sophisticated user interfaces, perhaps GBrowse, Apollo and CMap
- Web site upgrade
 - ◆ New MediaWiki
 - ◆ TableEdit
 - ◆ Better searching
 - ◆ New skin.

GMOD Logo Service

ZFIN's current logo was designed several years ago by Kari Pape, a student in a University of Oregon design class. Judy Sprague, ZFIN's manager, worked with the professor and the students to communicate what ZFIN was all about and at the end of the quarter we had about 20 designs to pick from, and most of them were spectacularly good.

Many GMOD user databases, web sites, and GMOD components don't have snazzy logos. Dave offered to contact the same department and the local community college as well, and ask if they would be interested in doing something similar GMOD community. This time around I would propose that each student or team get a different database/web site/component.

This was clearly the most popular idea Dave has ever had during his time at GMOD. I'll investigate ASAP. (*See GMOD Logo Program.*)

Education and Outreach

- 2009 Summer School
- GMOD Course in Europe?
- Conferences
 - ◆ PAG, Arthropod Genomics, IPlant
- Special emphasis on comparative genomics and community annotation.

Grants

The Help Desk now offers to review grant proposal prior to submission to help them fully state how much they can use GMOD components, and thus avoid reinventing the wheel for their project.

We will also start suggesting that grants that propose using GMOD components also include a limited amount of funding for GMOD in the grant. This could either be core project funding, funding for existing components or funding for new components to become part of GMOD.

Agenda Proposals

If you have something you want to be on the agenda at this meeting please add it below.

- The AJAX-GBrowse project, as demo'd at <http://genome.biowiki.org/> - Ian Holmes
- [InterMine](#) and [Chado](#) Richard Smith (Scott 14:04, 6 June 2008 (EDT))
- Advances in the [Common Gene Page](#) effort (see also here an old page at blog.gmod.org: http://blog.gmod.org/common_gene_pages , as well as Don Gilbert's page on the topic on his server: <http://eugen.es.org/gmod/gene-report-examples/>) (Scott 14:04, 6 June 2008 (EDT))
- TableEdit round trip/integration progress and plans --[JimHu](#) 18:35, 6 June 2008 (EDT)
- Java Chado data model API with higher level, user friendly "Biological" layer [Ed](#)
- GBrowse 1.69 show and tell [Sheldon McKay](#)
- MediaWiki enhancements [Sheldon McKay](#)

- [GMOD Help Desk](#) - [Dave Clements](#)
 - ◆ priorities for 2008-2009
 - ◆ Evaluation - Has the Help Desk been helpful? - [Dave](#) and [Don Gilbert](#)
 - ◆ Grant review service
 - ◆ Funding GMOD
 - ◆ GMOD and the Google Summer of Code in 2009?
 - ◆ The Chado Natural Diversity module
 - ◆ [Galaxy](#) Integration. [Galaxy](#) already integrates with [BioMart](#), and the current (May 2008) development version of [Galaxy](#) integrates with the current (May 2008, 1.69 Beta, e.g. "stable") development version of [GBrowse](#). Once this goes to production in both Galaxy and GBrowse, should Galaxy work on integrating with other GMOD components such as [Chado](#) or [InterMine](#)?