

GMOD Coordinator, software development manager at The Ontario Institute for Cancer Research.

Scott graduated from The Ohio State University in 1990 with a B. S. in Chemical Engineering and continued on to the University of California in San Diego for graduate school, where he studied improvements to monoclonal antibody production in industrial processes through cell culture and cell cycle modifications. Upon earning his Ph. D. in Chemical Engineering, he began work in the software industry, working for MSI/Accelrys, Pangea Systems/DoubleTwist, and Athersys, until finally coming to rest at Cold Spring Harbor Laboratory in Lincoln Stein's lab. He is the coordinator and a lead developer of the GMOD project, focusing mainly on core schema and tool development, but is also a developer for GBrowse as well as in the BioPerl project. Scott also teaches programming, web development and database design courses at the University of Phoenix. Scott is a happy husband and father of twelve year old boy-girl twins.

Scott can be reached at his email address: sco...@scottcain.net or Skype to username scottcain.



Georgia Tech Seminar

Title: The GMOD Project and Community Annotation

Abstract: Scott will introduce the Generic Model Organism Database (GMOD) project, both its philosophy and the practical results. With open source software projects covering genomics topics like database storage, visualization, feature editing and literature curation, the GMOD project tries to be many things to many people. Then Scott will discuss work tying together several GMOD projects (Chado, Apollo, ModWare, and GBrowse) along with the popular open source wiki engine, MediaWiki, to create a reusable community annotation system.

The GMOD Project

A talk to be given at the Ohio Collaborative Conference on Bioinformatics (OCCBIO), Miami University, Oxford, Ohio, July 9, 2007

Abstract: The Generic Model Organism Database (GMOD) project is a program to develop generic, reusable software components for building organism databases. The resulting collection of software tools can be used for creating and managing both small laboratory database of genome annotations and large web-accessible community database. GMOD tools are in use at FlyBase, WormBase, SGD, BeeBase and many other large and small community databases. I will introduce several of the GMOD components and discuss several options for configuring and using the software.

Publications

[\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#) [\[6\]](#) [\[7\]](#)

1. [?](#) O'Connor BD et al. GMODWeb: a web framework for the Generic Model Organism Database. *Genome Biol.* 2008 9:R102.
[DOI:10.1186/gb-2008-9-6-r102](#) [PMCID:PMC2481422](#) [PMID:18570664](#)
2. [?](#) Arnaiz O, Cain S, Cohen J, and Sperling L. ParameciumDB: a community resource that integrates the Paramecium tetraurelia genome sequence with genetic data. *Nucleic Acids Res.* 2007 35:D439-44.
[DOI:10.1093/nar/gkl777](#) [PMCID:PMC1669747](#) [PMID:17142227](#)
3. [?](#) Harrington JJ et al. Creation of genome-wide protein expression libraries using random activation of gene expression. *Nat. Biotechnol.* 2001 19:440-5.
[DOI:10.1038/88107](#) [PMID:11329013](#)
4. [?](#) Cain SJ and Chau PC. Transition probability cell cycle model with product formation. *Biotechnol. Bioeng.* 1998 58:387-99.
[PMID:10099273](#)
5. [?](#) Cain SJ and Chau PC. Transition probability cell cycle model. Part I--Balanced growth. *J. Theor. Biol.* 1997 185:55-67.
[DOI:10.1006/jtbi.1996.0289](#) [PMID:9093553](#)
6. [?](#) Cain SJ and Chau PC. Transition probability cell cycle model. Part II--Non-balanced growth. *J. Theor. Biol.* 1997 185:69-79.
[DOI:10.1006/jtbi.1996.0290](#) [PMID:9093554](#)
7. [?](#) Cain SJ and Chau PC. A transition probability cell cycle model simulation of bivariate DNA/bromodeoxyuridine distributions. *Cytometry* 1997 27:239-49.
[PMID:9041112](#)

Genome Informatics 2010 machine prep notes

Ubunut 9.10 Server

```
sudo apt-get install subversion libgd2-xpm-dev autoconf apache2 libcgi-session-perl libdbd-pg-p
```

```
sudo apt-get install php5 libapache2-mod-php5 php5-xsl php5-gd php-pear php5-pgsql php5-curl ph
```

User:Scott

```
sudo apt-get install libdbd-sqlite3-perl libproc-processtable-perl libio-all-perl libio-prompt-
```

Installed Java 5 after adding Ubuntu jaunty repos to /etc/apt/sources.list:

```
sudo apt-get install sun-java5-jdk
```

Installed via CPAN:

- Bio::Graphics::Browser2 (GBrowse version 2.13) (also installed Bio::Graphics, JSON, Data::Stag)
- Bio::DB::Das::Chado (Chado GBrowse adaptor version 0.30)
- GO::Parser
- Module::Load
- DBIx::DBSchema
- XML::Parser::PerlSAX

```
sudo a2enmod rewrite
```

Downloaded software in ~/Software:

- <http://ftp.drupal.org/files/projects/drupal-6.19.tar.gz>
- [maker.203.tar.gz](#)
- [chado-1.11.tar.gz](#)
- ant-1.8.1 (can't have Java5 and ant both from apt repos :-/)
- DBIx::DBStag (needs to be installed by hand)
- <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/release/LATEST/blast-2.2.23-ia32-linux.tar.gz>
- <http://www.repeatmasker.org/rmblast-1.2-ncbi-blast-2.2.23+-src.tar.gz> (used organism specific repeat library from MAKER sample data)
- <http://tandem.bu.edu/trf/trf.html>
- <http://www.repeatmasker.org/RMBlast.html>
- <http://www.ebi.ac.uk/~guy/exonerate/>
- <http://augustus.gobics.de/binaries/augustus.2.4.src.tar.gz>
- <http://homepage.mac.com/iankorf/snap-2010-07-28.tar.gz>

Downloaded software in ~/Updateable_repos

- bioperl-live
- Apollo
- jbrowse

PostgreSQL prep:

- create gmod and drupal users

- open up connections from local host (setting "trust" in pg_hba.conf)
- create test, chado and drupal databases

Drupal related stuff

- Added AllowOverride All to /var/www
- sudo a2enmod rewrite

GBrowse2 stuff:

- added a simple .htaccess file to override the drupal on in /var/www/gbrowse2. It just contains the line "Options +Indexes"
- also added index.html to the DirectoryIndex directive in the drupal .htaccess file.

Chado stuff:

- put environment variables (GMOD_ROOT and CHADO_DB_NAME) in .profile

Preping ING app

- What is the mission of your organization/project?

The Generic Model Organism Database project (GMOD), based at the Ontario Institute for Cancer Research, is devoted to creating open source software to facilitate biological research across a variety of disciplines, including cancer research, genetics, crop improvement and evolution. The GMOD community consists of scientists and software developers from around the world with contributors from academia, non-profits, and the biotech and pharmaceutical industries.

- What makes your organization/project one that provides value?

Software created by the GMOD project is in use by hundreds of research groups around the world. The cross-organization nature of the project both reduces waste due to duplicated efforts and improves the resulting software by getting multiple points of view.

- Give one example of how your organization gave back to the community in the last four months.

In March of this year, we held a week-long training course for biologists to learn how to work with their data in the context of GMOD software. Attendees were from North America, Asia and Europe, and were from a variety of disciplines, including crop science, biomedical research, vaccine development and insect/pest research. While this course was not free, the minimal tuition of \$265 was only to cover the travel expenses of some of the instructors. See http://gmod.org/wiki/2011_GMOD_Spring_Training for more information.

- What networks, collaborations or professional associations are you a part of?

The GMOD project is a large collaboration itself.

User:Scott

- Why do you want to be a member of Network Orange?

???

- What do you feel you can offer to the community at Network Orange?

???