

# The Future of GMOD and Chado

PAG 2015

Scott Cain  
GMOD Coordinator  
[scott@scottcain.net](mailto:scott@scottcain.net)

# What is GMOD?

- 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 more organizations.

# GMOD's status

- Funding is problematic (like I'm telling you something you didn't know)
- Went from 2 FTEs to 0.33 FTE
- Several popular software projects have dedicated funding (JBrowse, Tripal, Galaxy, WebApollo)
- Nevertheless...

# “Business” is Booming

- See <http://gmod.org/wiki/PAG> for a sampling of how widely used GMOD software is.
- Or see <http://bit.ly/1u4GXkR> (flickr) for a collection of 51 posters covering a wide range of taxa using GMOD software.

# What does the lack of funding mean

- Lack of “coordination”:
  - Difficult to arrange GMOD meetings and courses
  - Chado development has slowed (though we’re working on it!)
  - Building “combined” tools (like GMOD in the Cloud/Box) has stalled
  - Getting emails answered is slow too.

# Looking for meeting/ course venue

- We are looking for a venue where we could hold a two day GMOD meeting and 4-5 day course
- Naturally, it needs to be a pretty inexpensive place and accessible
- Can be any time of year
- Ideas? Want to volunteer?

# Chado

- A organism-agnostic relational database schema for storing many different biological data types (sequence features, ontologies, dbxrefs, genotypes/phenotypes, etc)
- Currently in use by many organizations



# Chado recent history

- Last release was December 2013
- Last schema change was September 2011
- Slow, non-destructive schema evolution is a feature, not a bug
- Nevertheless...

# Things we'd like to get into Chado

- Handling subspecies in the organism table
- Several prop and linking tables (currently some created by Tripal to make it “work”), things like feature\_stock and contactprop.
- Organism\_relationship (but not for phylogeny)
- Make feature.seqlen a bigint (darn pine trees!) (2 billion to 9 quintillion)
- Bug fixes for associated tools

# Other things that may make it in

- Adding types (cvterms) to linking tables?
- Not a schema change, but reorganizing database related tables (db, dbxref, etc) into it's own module and out of "general".
- The "Group" module for generic grouping of "items" (features, stocks, etc)
- Migrating the code repository to GitHub

# On tap for the rest of the day

- KnowPulse: A Breeder-Focused Web Portal That Integrates Genetics and Genomics of Pulse Crops with Model Genomes
- Phytozome Population Diversity Visualization in JBrowse: A very large Dataset Case Study
- JBrowse within the Arabidopsis Information Portal
- Apollo + i5K: Collaborative Curation and Interactive Analysis of Genomes
- The Teosinte (*Zea mays* ssp. *parviglumis*) de novo Genome Assembly
- “Panel discussion” after lunch

# Big Thanks

There are so many people involved in the GMOD project that I couldn't possibly list them here (at least, not without inadvertently missing somebody), so I won't try.

That said, I really do want to thank:

Dave Clements

Stephen Ficklin

Lacey Sanderson

Lincoln Stein

# 8th International Biocuration Conference

- Beijing, China
- April 23-26, 2015
- <http://biocuration2015.big.ac.cn>
- International Society for Biocuration:
  - <http://biocurator.org/>

