



Digesting the Genome Glut

Promoting the Use and Extension of GMOD To Emerging Model Organisms



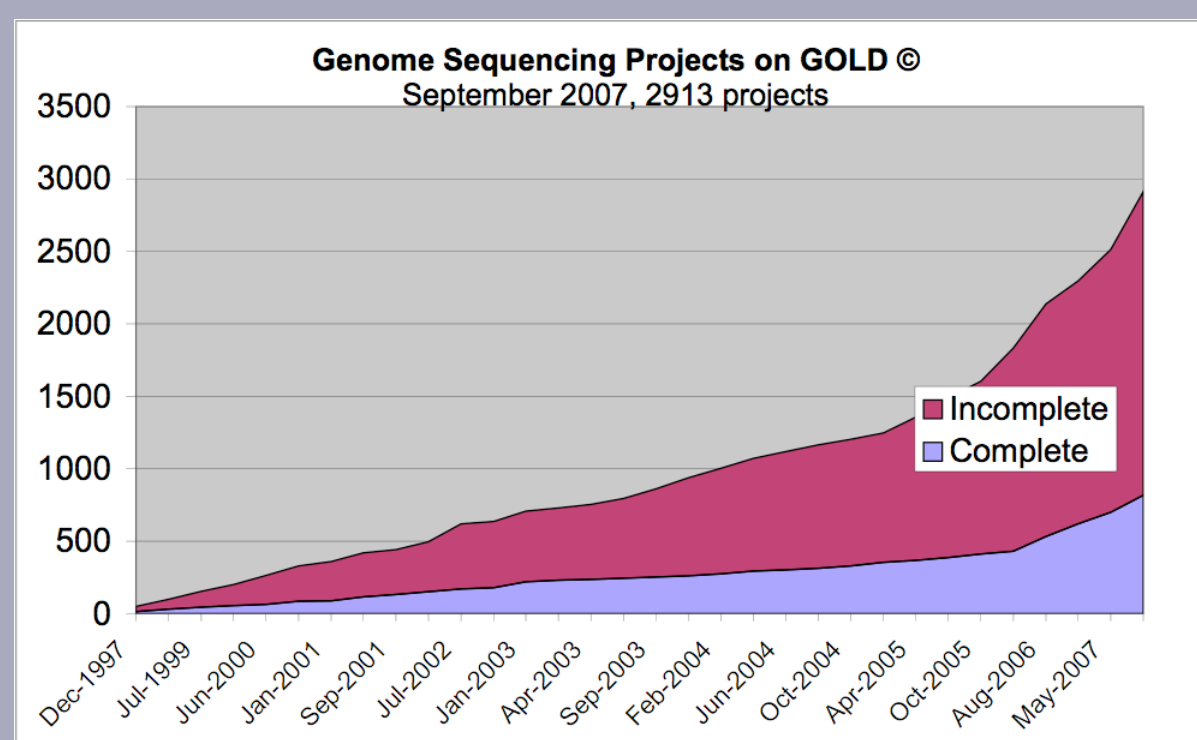
David Clements¹ Brian Osborne² Hilmar Lapp¹ Xianhua Liu¹
W. Owen McMillan³ Nassib Nassar⁴ Todd J. Vision¹

¹National Evolutionary Synthesis Center ²ZBio LLC ³North Carolina State University ⁴Renaissance Computing Institute

Abstract

An ever-increasing volume of genomic data is being generated from organisms well suited to comparative genomic, evolutionary and ecological studies. More data offers more potential for discovery, but it also makes it is harder to organize, visualize and annotate.

The Generic Model Organism Database (GMOD, <http://gmod.org>) is a toolkit of interconnected components that meet many of the data management, analysis, and visualization needs of these studies, and on a small informatics budget.



GMOD has recently added a helpdesk to promote use of GMOD more broadly in emerging model organism research communities. GMOD has also been extended to better support evolutionary research.

Getting Started with GMOD

- Start at GMOD.org
- Download software
- FAQs and HOWTOs
- GMOD for the Biologist
- GMOD Mailing Lists
- Feature requests & Bug Reports
- Contribute doc! GMOD.org is a wiki.
- The most popular GMOD components.
- GMOD Support Helpdesk: help@gmod.org and mailing lists for each GMOD Component
- Contribute code! GMOD is open source.

In the past year, the GMOD web site has been revamped and a GMOD Helpdesk (help@gmod.org) has been created.

What Can GMOD Do?

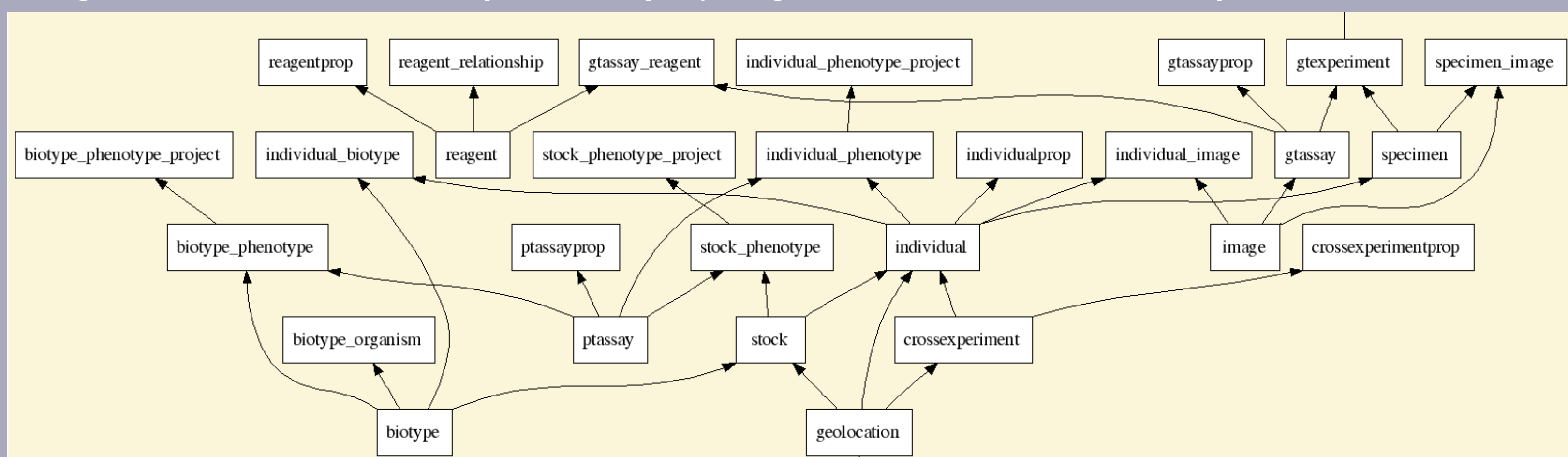
GMOD is a collection of software tools for creating and managing genome-scale biological databases. You can use it to create a small laboratory database of genome annotations, or a large web-accessible community database. GMOD includes a modular database schema called Chado that supports many common needs.

If you have	& you need to	then GMOD can help with
Genomic Sequence	View Annotate Organize	GBrowse Apollo Chado Sequence Module
Physical, Genetic & Comparative Maps	View Annotate Organize	CMap, SynView, SynBrowse, Sybil Sybil Chado Map Module
Phylogenetic Trees	Organize View	Chado Organism & Phylogeny modules GMODWeb
Phenotype, Genotype	Organize View	Chado Genetics Module GMODWeb
Microarray and Expression	View Organize	Java TreeView, Caryoscope, GeneXplorer Chado Mage and Expression Modules
Ontologies	Curate View	Chado CV module GO Term Viewer, GMODWeb
Phylogenetic Trees	Organize View	Chado Organism & Phylogeny modules GMODWeb
Publications	Curate, search Organize	TextPresso, PubSearch Chado Publication Module

GMOD for Natural Diversity Data

Heliconius butterflies are a favored model for studying mimicry, because geographically overlapping populations of different species frequently converge on similar wing color patterns. Complementing the long history of evolutionary genetic studies in the genus, an increasing amount of molecular marker and sequence data are being collected.

To capture the diversity data in a GMOD-compatible way, Chado (GMOD's relational data model) has been extended with the 'diversity' module (see <http://sf.net/projects/heliconiusdb>). The module supports stocks, individuals, pedigrees, cross experiments, geolocations, and phenotype and genotype experiments. In addition, some of Chado's core modules have been revised to consolidate how organism relationships and phylogenetic trees are represented.



Entity graph of the 'diversity' extension to Chado, and architecture and Web GUI for HeliconiusDB, an evolutionary model database.

The architecture diagram shows a stack of layers: Client program (AJAX-driven) and Web-GUI (AJAX-driven) at the top; Data Service API (REST-based); Middleware (Spring/J2EE); Hibernate (Object persistence); and Chado Database with Diversity extension (PostgreSQL) at the bottom.

The screenshot shows the HeliconiusBase web interface with a 'Broods' table and a pedigree chart.

Brood Number	Type of Cross	Father	Mother	Number of Eggs	Number of Adults
2	F2	Heliconius erato ssp. erato x Heliconius himera	Heliconius erato ssp. erato x Heliconius himera	0	
	F2	Heliconius erato ssp. erato x Heliconius himera	Heliconius erato ssp. erato x Heliconius himera	118	80
	BX	Heliconius erato ssp. erato	Heliconius himera	219	70
	BX	Heliconius erato ssp. erato	Heliconius himera	72	43
	F1	Heliconius erato ssp. erato	Heliconius himera		
	F2	Heliconius erato ssp. erato x Heliconius himera	Heliconius erato ssp. erato x Heliconius himera	175	120
	BX	Heliconius erato ssp. erato	Heliconius himera	27	12
	BX	Heliconius erato ssp. erato	Heliconius himera	111	55
	F2	Heliconius erato ssp. erato	Heliconius himera	161	118
	F1	Heliconius himera	Heliconius himera	23	7
	F1			66	
	F1			1	

The National Evolutionary Synthesis Center (NESCent) Sponsors Synthetic, Interdisciplinary and Transformative Research in Evolutionary Biology.

The center hosts working groups, catalysis meetings, sabbaticals, post-doctoral fellowships, and a short-term visitor program. NESCent's informatics branch undertakes initiatives in support of software interoperability, data sharing, and informatics training. The center also solicits whitepapers for projects that fall outside the scope of existing programs. More information on how to apply or become involved at: <http://www.nescent.org>

NESCent is located in Durham NC and is a collaborative effort of Duke University, North Carolina State University, and the University of North Carolina at Chapel Hill.



NESCent is supported by the National Science Foundation (Grant # EF-0423641)



The GMOD Helpdesk is hosted by NESCent and is funded by a National Institutes of Health grant to Ian Holmes and a USDA Agricultural Research Service grant to Doreen Ware.

