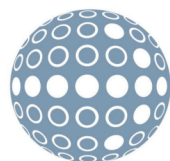# GMOD Project Update

Scott Cain, PhD.
GMOD Project Coordinator
Ontario Institute for Cancer Research
scott@scottcain.net

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

Plant and Animal Genome (PAG) XVIII

11 January 2010

GMOD @ Plant and Animal Genome XVIII

GMOD Users
Gramene
RAP-DB
ConiferGDB
TAIR
MaizeGDB
SGN
TreeGenes
SoyBase
Birdbase
Marine Genomics
BeeBase
GnpIS
+ hundreds more

GMOD Tools
Chado
Apollo
GBrowse
CMap
Tripal
GBrowse_syn
Ergatis
Textpresso
BioMart
JBrowse
MAKER
Galaxy
InterMine
…

Workshops

Animal Genomics with Galaxy: Analyze, Publish & Visualize
Saturday 1:30-2:30pm, California Room, Dan Blankenberg

MAKER: An easy to use genome annotation pipeline
Saturday, 3:50-6:00pm, California Room, Carson Holt

Database Resources at the EBI: Ensembl (including BioMart)
Sunday, 8:15-9:15am, California Room, Bert Overduin

The Generic Genome Browser: A Hands on Workshop
Sunday, 1:30-3:40pm, California Room, Scott Cain

Using Gramene (including CMap)
Tuesday, 3:50-6:00pm, California Room, Ken Youens-Clark

Comparative Genomics with GBrowse_syn
Wednesday, 10:20am-12:30pm, Pacific Salon 3, Sheldon McKay

Talks

GMOD Project Update
Monday, 12:50-1:10pm, California Room, Dave Clements, Scott Cain

Using the JBrowse Genome Browser with Large Amounts of Data
Monday, 1:10-1:30pm, California Room, Mitch Skinner (also see poster P919)

Tripal: A Construction Toolkit for Online Genomics Databases
Wednesday, 2:30-2:50pm, California Room, Stephen Ficklin

Posters

Monday 10-11:30am
P858: GBrowse and Next Generation Sequencing Data

Monday 3-4:30pm
P859: Visualizing and Comparing Genomes
P861: Comparative Genomics Tools in GMOD

+ *at least 15 other talks and 19 more posters by GMOD users (see http://gmod.org/wiki/PAG_2010)*

http://gmod.org/wiki/PAG_2010

Got genome?

Agenda

- Introduction
- What's new: Software
- What's new: Project

Focusing on what's happened in the past 2 years

And what's happening in the coming years

http://gmod.org/

Introduction: 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.

Who uses GMOD?

Plus hundreds of others

Agenda

- Introduction
- What's new: Software
- What's new: Project

Software

GMOD components can be categorized as

V	Visualization
D	Data Management
A	Annotation

Visualization: GBrowse

GBrowse

JBrowse

GBrowse_syn

CMap
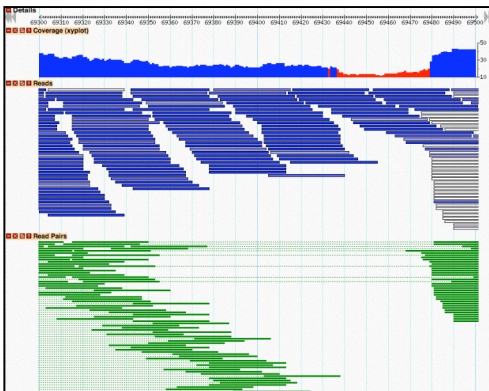
Releases
  1.70 released
  2.0, 1.71 in the pipe

AJAX/Interface:
  Rubberband region selection, drag and drop
  track ordering, collapsible tracks, popup balloons,
  asynchronous rendering (2.0)

Biology:
  Allele/gentotype frequency, LD glyphs, geolocation
  popups, circular genome support (1.71)

Infrastructure
  User logins, server multiplexing (2.0), SQLite and
  SAMtools (NGS) adaptors

modENCODE Fly:
  http://modencode.oicr.on.ca/cgi-bin/gb2/gbrowse/fly/

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

Visualization

GBrowse

JBrowse

GBrowse_syn

CMap

Resources

Tutorials (http://gmod.org/wiki/GBrowse_Tutorial):
  GBrowse User Tutorial at OpenHelix.com
  GBrowse Admin Tutorial
  NGS in GBrowse and SAMtools Tutorial

Web Sites:
  GMOD   http://gmod.org/wiki/GBrowse
  WebGBrowse   http://webgbrowse.cgb.indiana.edu/
  GBrowse.org   http://gbrowse.org

Mailing List:
  https://lists.sourceforge.net/lists/listinfo/gmod-gbrowse
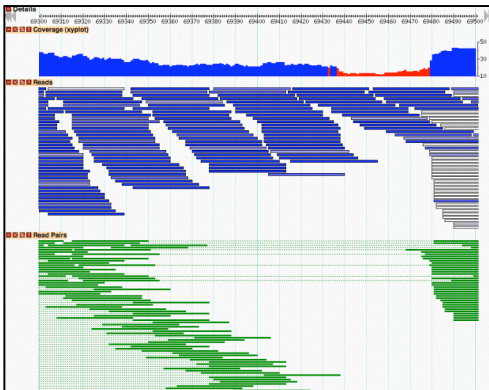
Visualization

GBrowse

JBrowse

GBrowse_syn

CMap
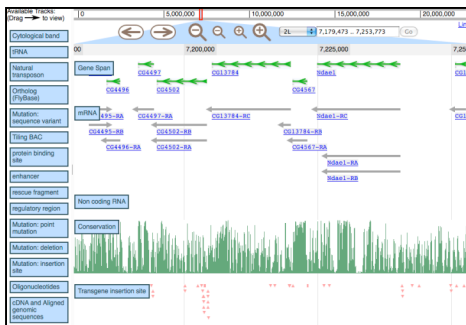
GMOD's 2nd Generation Genome Browser
It's *fast*

Completely new genome browser implementation:
  Client side rendering
  Heavy use of AJAX
  Uses JSON and Nested Containment Lists

JBrowse Fly:
  http://jbrowse.org/genomes/dmel/

Web Sites:
  GMOD   http://gmod.org/wiki/JBrowse
  JBrowse   http://jbrowse.org

Mailing List:
  https://lists.sourceforge.net/lists/listinfo/gmod-ajax

Visualization

GBrowse

JBrowse

GBrowse_syn

CMap

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

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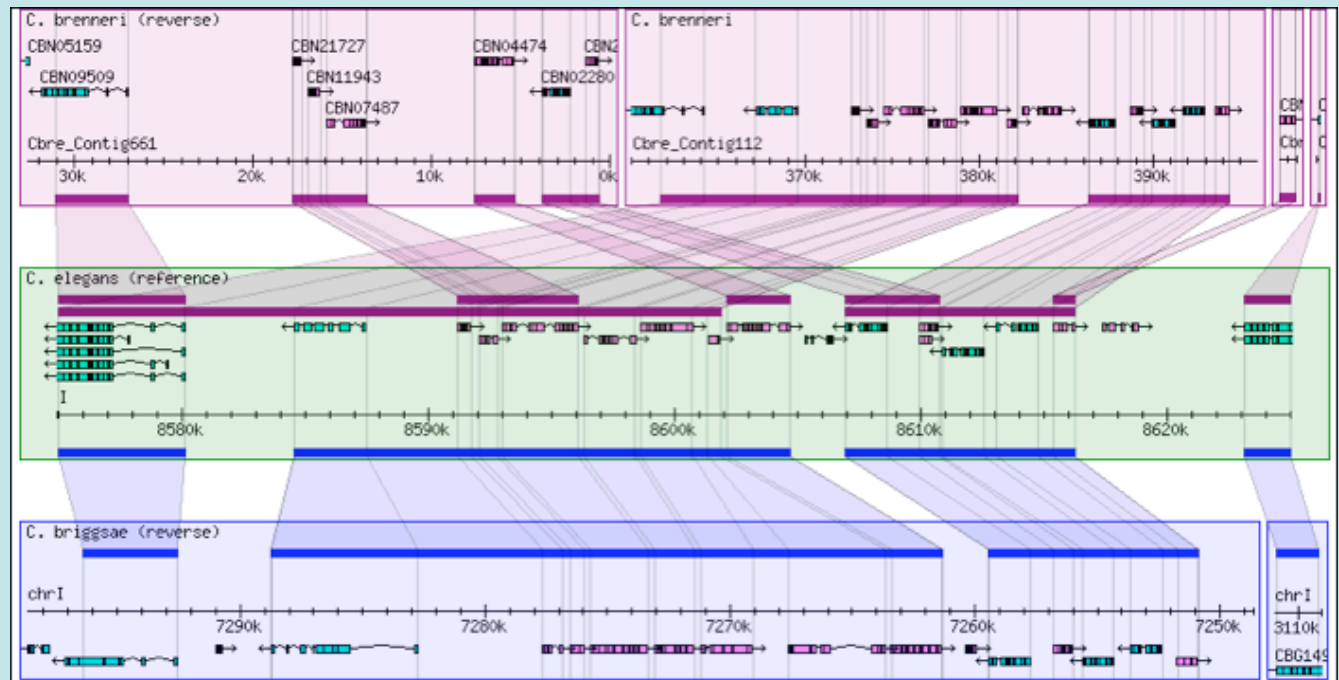
Syntenic blocks do not have to be colinear
Can also show duplications

http://gmod.org/wiki/GBrowse_syn

Sheldon McKay, Cold Spring Harbor Laboratory

Visualization

GBrowse

JBrowse

GBrowse_syn

CMap

Web based comparative map viewer
CMap is data type agnostic:
  Can link sequence, genetic, physical, QTL, deletion, optical, …
CMap 2.0 coming
  Faster, internals cleanup
  Circos export
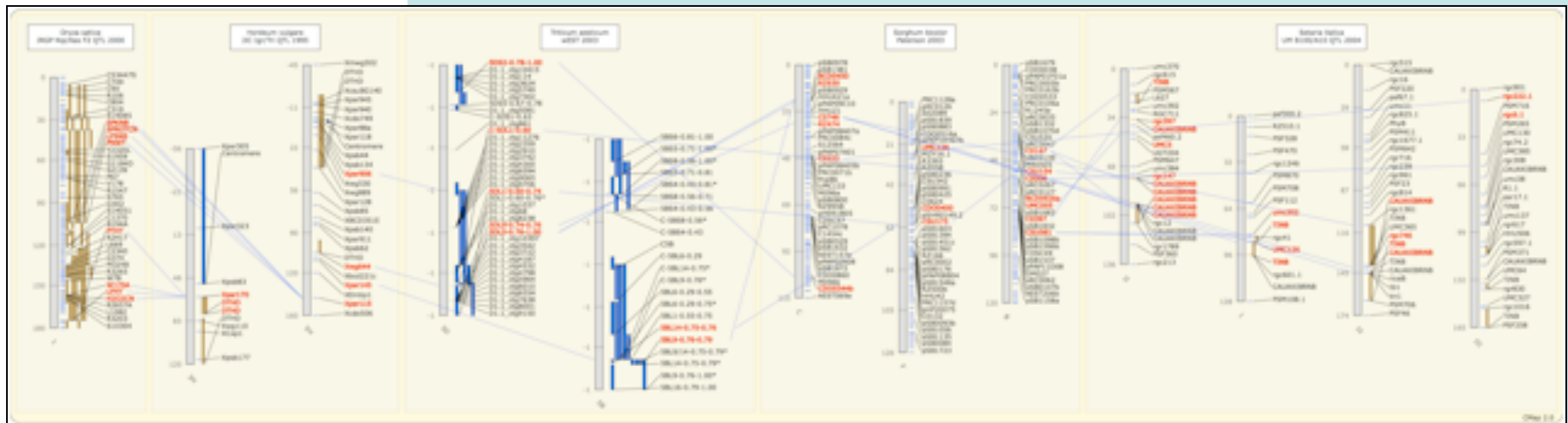
Data Management

Chado

Tripal

TableEdit

BioMart

InterMine

A extensible, modular database schema for storing
    biological data

1.0 release:
    Stable schema
    Tools for data in/out

1.1 release (soon):
    Stable schema (minor, nondestructive changes)
    Improvements to data loading scripts
    Additional modules: cell line, natural diversity
    Tool for managing materialized views
    Tool for creating ontology-based views

Data Management

Chado

Tripal

TableEdit

BioMart

InterMine

New web front end for Chado databases

Set of Drupal modules

Modules approximately correspond to Chado modules

Easy to create new modules

Includes user authentication, job management,
        curation support

Stephen Ficklin, Meg Staton, Chun-Huai Cheng, …
Clemson University Genomics Institute

Data Management

Chado

Tripal

TableEdit

BioMart

InterMine

MediaWiki extension
  MediaWiki software
  used at Wikipedia,
  GMOD.org, …

GUI to wiki tables
Also a GUI to
  database tables
Work in progress to
  use this with Chado

**Potential to give
wiki access to
Chado databases**

Example: GONUTS (http://gowiki.tamu.edu/)

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

Data Management

Chado

Tripal

TableEdit

BioMart

InterMine

New GUIs - more configurable and easier to use
Virtual Marts - marts running off source schema
  without materializing
Improved scalability
Security and
  access control
Improved federation
New configuration
  tool

Data Management

Chado

Tripal

TableEdit

BioMart

InterMine

Data integration and web-based query package

Now supports ~20 common data formats:
GFF3, Chado, GO annotation, biopax, BioGrid,
TreeFam, PubMed, Ensembl, …

Interfaces: RESTful web service, Java & Perl APIs

Upload & analyse gene lists with graphical and
statistical widgets

FlyMine: an integrated database for *Drosophila* and *Anopheles* genomics,
Rachel Lyne, *et al., Genome Biol.* 2007; 8(7): R129.

Annotation

MAKER

DIYA

Galaxy

Ergatis

Apollo

MAKER — Annotate this!

Genome annotation pipeline for creating gene models

Output can be loaded into GBrowse, Apollo, Chado, …

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 Web Annotation Service - MAKER online

Annotation

MAKER

DIYA

Galaxy

Ergatis

Apollo

Lightweight, modular, and configurable Perl-based
   pipeline framework

Initial application is gene prediction for prokaryotes

Working on integration of Amos assembly tools

Annotation

MAKER

DIYA

Galaxy

Ergatis

Apollo

NGS tools support
  QC and Manipulation: FASTQ, 454, SOLiD support
  Mapping: Bowtie or BWA, Megablast
  SAMtools: Web interface to SAMtools scripts

LIMS system in beta.
  Import data from your sequencer into Galaxy

DIYA: a bacterial annotation pipeline for any genomics lab, Andrew C. Stewart, Brian Osborne and Timothy D. Read, *Bioinformatics* 2009 25(7):962-963

Annotation

MAKER

DIYA

Galaxy

Ergatis

Apollo

Currently up 162 analysis tool / components for use in pipelines

Updated prokaryotic annotation pipeline template

Updated comparative annotation pipeline template

Lots of work for use on Amazon EC2

Now the engine behind the new CloVR cloud computing project (http://clovr.igs.umaryland.edu/)

http://ergatis.sourceforge.net/

Annotation

MAKER

DIYA

Galaxy

Ergatis

Apollo

Better Chado support

  including DBMS independent support)

GFF3 support

GUI based configurations

Multiple alignment transcript viewer and editor
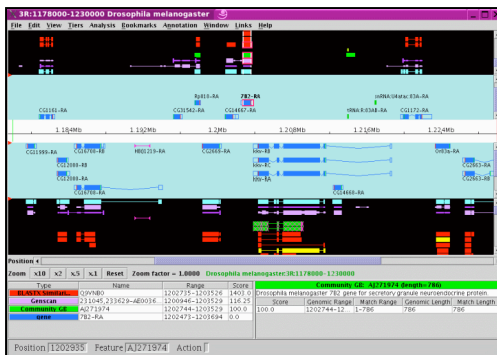
Continuous data display

  sgr, wiggle

Remote analysis to NCBI services

  BLAST, Primer-BLAST

Undo support

More robust Java Web Start support

Agenda

- Introduction
- What's new: Software
- What's new: Project

GMOD.org

A wiki, of course. GMOD.org is the hub for all things related to the project:

- – Documentation
- – News
- – Links
- – Calendar
- – Tutorials
- – HOWTOs
- – Glossary
- – Overview
- – …

Launched by Brian Osborne, 2007

Training and Outreach

- ## GMOD Summer Schools
  - 2009
    - July, NESCent, North Carolina, US
    - August, Oxford, UK
  - 2010
    - May, NESCent, North Carolina, US
    - ??, Asia / Pacific, maybe

- ## Outreach, Training
  - BA, SMBE, PAG, Arthropod Genomics, …

- ## Semi-annual community meetings
  - Next Meeting:
    - January 2010, San Diego, after PAG

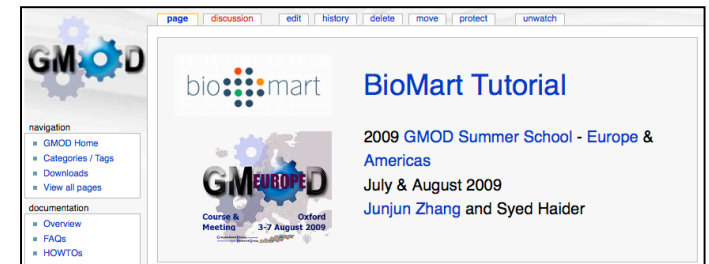http://gmod.org/wiki/Training_and_Outreach

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
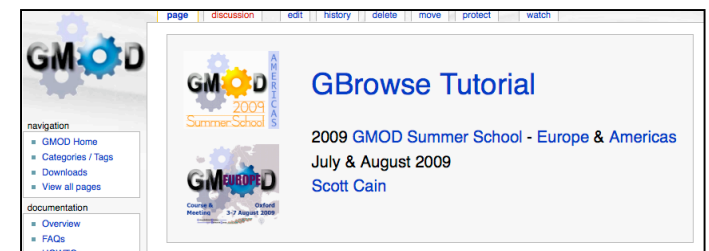

http://gmod.org/wiki/Training_and_Outreach#Online_Tutorials

Acknowledgements

Cornell
Dave Matthews

Oregon
Patrick Phillips
Phillips Lab

Utah
Carson Holt
Mark Yandell

NESCent
Todd Vision
Hilmar Lapp

Cambridge
Richard Smith

UC Berkeley
Mitch Skinner
Ian Holmes

OICR
Lincoln Stein

Penn State
Dan Blankenberg
Anton Nekrutenko

Texas A&M
Jim Hu

BBOP
Ed Lee

CSHL
Sheldon McKay
Ken Youens-Clark
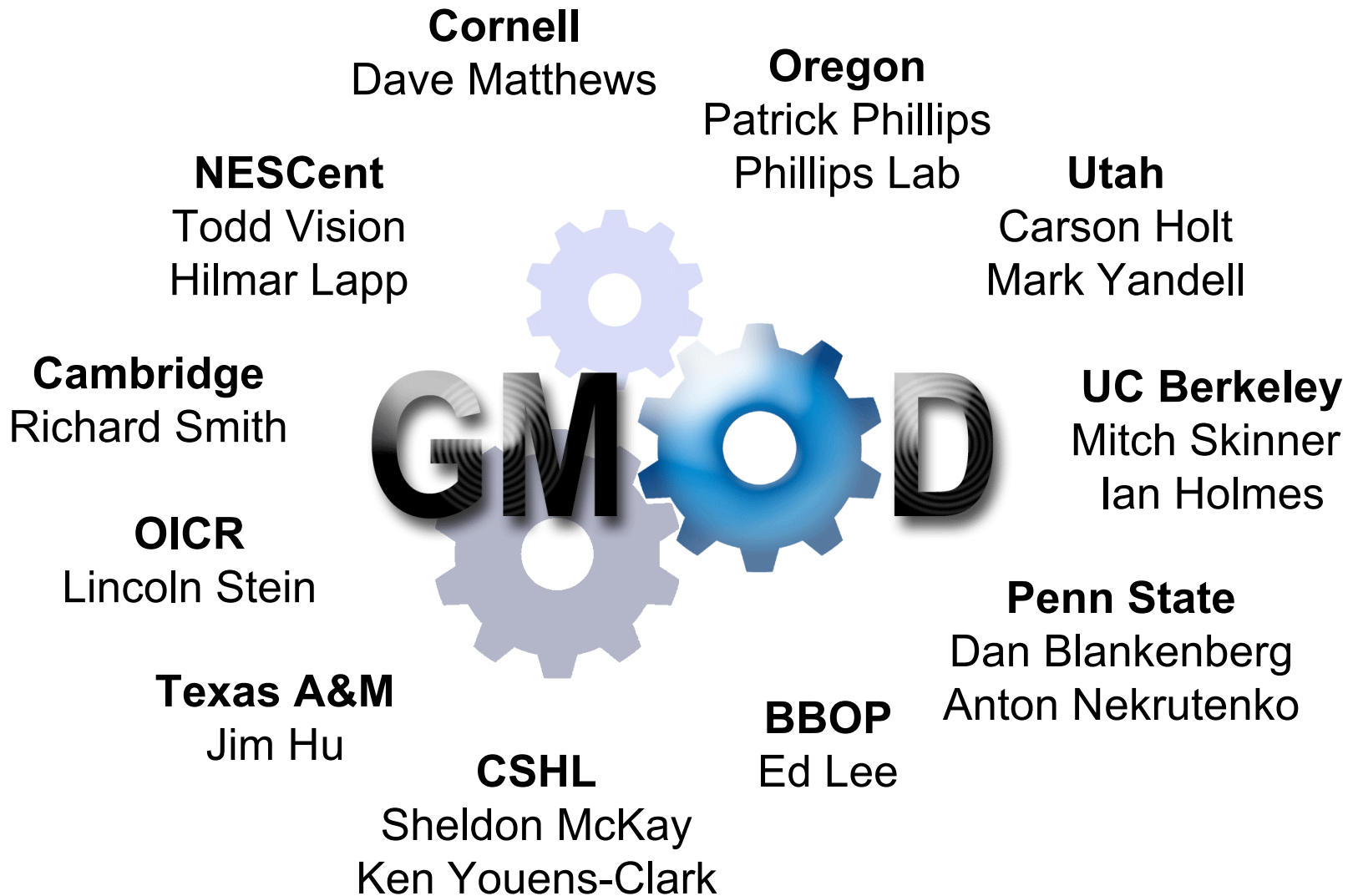
Thank You!

Scott Cain
GMOD Project Coordinator
Ontario Institute for Cancer Research
scott@scottcain.net

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

http://gmod.org/wiki/PAG_2010