

An Example Comparative Genome Database for Yeasts Using GMOD Tools

The goal of the GMOD project is to create a suite of applications that interoperate, sharing data between them. That goal has been partially realized by the latest release of the GMOD project, in which several applications share genome feature data via a standard database schema called Chado. Chado is a modular schema designed by developers at FlyBase to contain all of the data a model organism database would need. Applications that currently are 'Chado aware' are Apollo for feature editing, CMap for comparative map displays, the Generic Genome Browser (GBrowse) for browsing genome features, Turnkey for customized views of data in the database and BioMart (via a GFF3 to BioMart conversion script).

GBrowse

The Generic Genome Browser (GBrowse) is a web-based genome feature browser. It can use several data sources for displaying features; here it is using a Chado adaptor that allows it to show the contents of the database 'live'--that is, as edits are done on the database, they are reflected in GBrowse. (Lincoln Stein, Scott Cain and Ben Faga)

S. bayanus contigs in chado

Showing 91.49 kbp from c646, positions 1 to 91,494

Instructions
Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed. To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position.

Examples: c660, c500.

Search
Landmark or Region: c646-8505..91494 Search
Data Source: S. bayanus contigs in chado
Reports & Analysis: Highlight Selected Properties Configure... Go
Scroll/Zoom: Show 91.49 kbp Flip

S. cerevisiae chromosomes in chado

Showing 20 kbp from chrX, positions 408,478 to 428,477

Instructions
Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed. To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position.

Examples: YAL031C, YCR02, ChrV:100000-120000.

Search
Landmark or Region: chrX:408478..428477 Search
Data Source: S. cerevisiae chromosomes in chado
Reports & Analysis: Highlight Selected Properties Configure... Go
Scroll/Zoom: Show 20 kbp Flip

Overview of chrX

Named gene
YJL014W, YJL012C, YJL011C, YJL009W, YJL007C, YJL005W, YJL008C, YJL006C

View in CMap
YJL014W, YJL012C, YJL011C, YJL009W, YJL007C, YJL005W, YJL008C, YJL006C

CDS
YJL014W.mRNA, YJL012C.mRNA, YJL011C.mRNA, YJL009W.mRNA, YJL007C.mRNA, YJL005W.mRNA, YJL008C.mRNA, YJL006C.mRNA

RRN
YJL013C.mRNA, YJL008C.mRNA, YJL006C.mRNA

non coding RNAs
YJL013C.mRNA, YJL008C.mRNA, YJL006C.mRNA

Transposable elements

Turnkey
Turnkey is a generic web front end for any database. In this case, it has been tuned to behave as one would expect for a MOD with gene pages. (Allen Day and Brian O'Connor)

GMOD Home Help Gene or Probe search

Feature
is_analysis: 0
md5checksum: Sbay_Contig
name: Sbay_Contig
residues: Sbay_Contig
seqlen: Sbay_Contig
timeaccessioned: 2006-05-02 1
timelastmodified: 2006-05-02 1
uniqueid: Sbay_Contig
dbxref_id: Sbay_Contig
organism_id: 11
type_id: ORF (SO:0000000)

Feature
gene: YJL010C
Synonyms: NOP9 (Search in Textpresso), YJL010C (Search in Textpresso)
Database Links: YJL010C (Search in Textpresso)
Genomic Location: chrX, strand -1 (View in GBrowse)
gene 417477..419478

Comparative Mapping
View YJL010C in CMap
Transcript Neighborhood

Feature_Dbxref
SGD (GFF_source)

Feature_Cvterm
33099

Featureloc
chrX: 417477..419478

Feature_Relationship
YJL010C.mRNA (mRNA)

Featureprop
Essential nuclear protein synthesis

Feature_Synonym
NOP9

Apollo

Apollo is a Java-based sequence feature annotation editor. Like GBrowse, Apollo can use several data sources for display and editing of annotations. In this case, it can use a direct connection to the Chado database to edit features 'live.' (Mark Gibson and Nomi Harris)

chrX:400000-500000 Saccharomyces cerevisiae

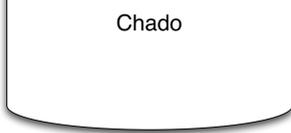
File Edit View Tiers Analysis Bookmarks Annotation Window Links Help

4Mb 0.42Mb 0.44Mb 0.46Mb 0.48Mb 0.5Mb

Position: 417477-419478

| Type | Name | Range | Score |
|------|----------------|---------------|-------|
| gene | YJL010C.mRNA | 419478-417478 | 0.0 |
| | Genomic Range | 419478-417478 | 2001 |
| | Genomic Length | | |

Position: 434133 Feature: Action



Chado is a modular, generic genomics schema. It is intended to model all data that might be required of model organism database. Current modules include ones for sequence features, expression data, genotypes, phenotypes and maps. (Chris Mungall, David Emmert, Scott Cain, Allen Day)

Search term(s) should occur in: whole abstract/article
Words to search for: YJL010C Exact match
Categories to search for (optional): none none
Search terms(s) should occur in these fields: Title Abstract Author Year Full Text
[customize] Search!

7 matches in 5 publication(s) found.
E-mail results to including matches. Go to page 1 of 1

Search Results

| Matching publications | Number of matches | View Sentences | Links/Downloads |
|--|-------------------|------------------|-----------------|
| Hazbun TR, Malmstrom L, Anderson S, Graczyk BJ, Fox B, Riffle M, Sundin BA, Aranda JD, McDonald WH, Chiu CH, Snysman BE, Bradley P, Muller EG, Fields S, Baker D, Yates JR 3rd and Davis TN (2003) Assigning function to yeast proteins by integration of technologies. <i>Mol Cell</i> , 12, 1353-1365 14690591 | 3 | [view sentences] | [PubMed] |
| De Las Penas A, Pan SJ, Castano I, Alder J, Cregg R and Cormack BP (2003) Virulence-related surface glycoproteins in the yeast pathogen <i>Candida glabrata</i> are encoded in subtelomeric clusters and subject to RAP1 and SIR-dependent transcriptional silencing. <i>Genes Dev</i> , 17, 2245-2258 12952896 | 1 | [view sentences] | [PubMed] |
| Fay JC, McCullough HL, Sniegowski PD and Eisen MB (2004) Population genetic variation in gene expression is associated with phenotypic variation in <i>Saccharomyces cerevisiae</i> . <i>Genome Biol</i> , 5, P: R26 15059259 | 1 | [view sentences] | [PubMed] |
| Millson SH, Truman AW, Wolfram F, King V, Panaretou B, Prodromou C, Pearl LH and Piper PW (2004) Investigating the protein-protein interactions of the yeast Hsp90 chaperone system by two-hybrid analysis: potential uses and limitations of this approach. <i>Cell Stress Chaperones</i> , 9, 359-368 15633294 | 1 | [view sentences] | [PubMed] |
| Lai LC, Kosoroff AL, Burke PV and Kwast KE (2005) Dynamical remodeling of the | | | |

Textpresso

Textpresso is a literature search tool; in this case, it is being hosted at SGD and linked via URL (SGD developers and Eimear Kenny)

bioMart Query summary

DATASETS
Scotts_scer_dset
2,441 entries total

FILTERS
FILTER BY FEATURE:
Feature type: Gene SGD

FILTER GENOME REGION:
Chromosome: chrIX
Region Start >=, Region End <=, Chrom. Size >=, Chrom. Size <=

FILTER REGIONS WITH FEATURES: (include, exclude)
Feature LIST: Has CDS SGD Only Excluded

BioMart

BioMart is a powerful query-oriented data management system. (Gudmundur Thirrisson) Here it interacts with Chado via a GFF3 dumper and GFF3 to BioMart script that is part of GMODTools (Don Gilbert)

CMap Home | Maps | Map Search | Feature Search | Matrix | Map Sets | Tutorial

Reference yeast
S. cerevisiae chromosome

Comparative
S. bayanus contig

chrX: 408478-428477

YJL014W, YJL012C, YJL011C, YJL009W, YJL007C, YJL005W, YJL008C, YJL006C

Repeat Map: 360-891-449,891 bp

Feature Types
Name: Features in red have correspondences
Evidence Types: Lightblue lines denote from a file I got from Ensembl

Menu Symbolist
Map Set Info, Map Details

CMap

CMap is a comparative map viewer. Here it is being used to show synteny between two related yeast species. (Ben Faga and Ken Clark)



The Generic Model Organism Database Project
http://www.gmod.org/

Creating a "MOD in a box"
Coordinator: Scott Cain <cain@cshl.edu>

A collaboration of:



GMOD is supported by a specific cooperative agreement from the USDA Agricultural Research Service, and by NIH grants co-funded from the National Human Genome Research Institute and the National Institute of General Medical Sciences.