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 Here six GMOD tools are used to create an partially realized by the latest release of the Turnkey example comparative genome model GMOD project, in which several applications Turkey is a generic web front end for any database. organism database. Data obtained from In this case, it has been tuned to behave as one share genome feature data via a standard would expect for a MOD with gene pages. (Allen database schema called Chado. Chado is a the Saccharomyces Genome Database Day and Brian O'Connor) modular schema designed by developers at (SGD) was used to populate a Chado FlyBase to contain all of the data a model schema to power several views of the data. organism database would need. Applications GM:OD that currently are 'Chado aware' are Apollo for These applications are hosted at feature editing, CMap for comparative map Gene or Probe ▼ search Home Help http://gmod.cshl.edu/ displays, the Generic Genome Browser (GBrowse) for browsing genome features, Feature Turnkey for customized views of data in the database and BioMart (via a GFF3 to BioMart is analysis Gene or Probe ▼ search Home Help conversion script). md5checksum Sbay_Contig name Feature_Dbxref Feature residues SGD (GFF_source) Apollo gene:YJL010C timeaccessioned 2006-05-02 1 Apollo is a Java-based sequence feature Synonyms Feature Cvterm annotation editor. Like GBrowse, Apollo can timelastmodified 2006-05-02 NOP9 (Search in Textpresso) use several data sources for display and editing ♦ 33099 Sbay_Contig YJL010C (Search in Textpresso) uniquename of annotations. In this case, it can use a direct connection to the Chado database to edit dbxref id Database Links **GBrowse** features 'live.' (Mark Gibson and Nomi Harris) Featureloc organism id Genomic Location The Generic Genome Browser ○ chrX: 417477 .. 41947 (GBrowse) is a web-based genome ORF (SO:000 chrX, strand -1 [view in GBrowse] type_id _ _ X chrX:400000-500000 Saccharomyces cerevisiae feature browser. It can use several File Edit View Tiers Analysis Bookmarks Annotation Window Links Help Edit... data sources for displaying features; Feature_Relationship gene 417477..4/19478 here it is using a Chado adaptor that YJR022W.mRNA allows it to show the contents of the YJL010C.mRNA (mRNA Comparative Mapping database 'live'--that is, as edits are YJLWtau4 YJR020W.mRNA done on the database, they are View YJL010C (n/CMap YJL009W.mRNA YJL014W.mRNA YJR002W.mRNA YJRWTy1-2 Featureprop reflected in GBrowse. (Lincoln Stein, Transcript Neighborhood Scott Cain and Ben Faga) YJL016W.mRNA YJR001W.mRNA YJLWdelta9 YJR007W.mRNA YJRWTy1-1 Essential nucleolar pro S. bayanus contigs in chado YJL019W.mRNA YJLWdelta8 YIL001W.mRNA 415k 416k 417k 418k 419k 420k 421k 422k 423k 424k synthesis YJL020W-A.mRNA YJR010W.mRNA YJK011C YJL009W Showing 91.49 kbp from c646, positions 1 to 91,494 YJL010C Feature_Synonym 0.46Mb 0.48Mb 0.42Mb 0.44Mb YJL008C YJL006C ◇ NOP9 YJR030C.mRNA 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 YJL020C.mRNA YJL010C.mRNA YJL009W.mRNA YJL007C.mRNA Scroll/Zoom buttons to change magnification and position. YJR031 C.mRNA YJR010C-A.mRNA YJR023C.mRNA YJL011C.mRNA YJL006C.mRNA Examples: c660, c500. YJR033C.mRN/ YJR005C-A.mRNA YJR012C.mRNA YJR024C.mRNA [Hide banner] [Bookmark this] [Link to Image] [High-res Image] [Help] [Reset YJR011C.mRNA YJR021C.mRNA YJL011 C.mRNA YJL006C.mRNA YJR017C.mRNA Landmark or Region: Reports & Analysis: Highlight Selected Properties 🔀 Configure... Go c646:-8505..91494 Search YJR019C.mRNA Data Source Scroll/Zoom: Show 91.49 kbp T Flip YJR025C.mRNA S. bayanus contigs in chada S. cerevisiae chromosomes in chado Position **∢** Showing 20 kbp from chrX, positions 408,478 to 428,477 Zoom x10 x2 x.5 x.1 Reset Zoom factor = 1.0000 Saccharomyces cerevisiae:chrx:400000-50000 □ Details YJL010C \leftarrow YJL010C.mRNA 419478-417478 0.0 /JL010C.mRNA 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. Genomic Range Genomic Length 419478-417478 2001 View in C Examples: YAL031C, YCR02*, ChrV:100000-120000. [Hide banner] [Bookmark this] [Link to Image] [High-res Image] [Help] <mark>[Reset]</mark> Position 434133 Feature Action ☐ Tracks Reports & Analysis: Landmark or Region: Highlight Selected Properties 💌 Configure... chrX:408478..428477 Data Source Scroll/Zoom: Khow 20 kbp S. cerevisiae chromosomes in chado □ Display Settings Search term(s) should occur in: a sentence whole abstract/article Image Width Overview of chrX € 450 € 640 € 8 Words to search for: YJL010C Exact match Highlight feature(s) ☐ Details Categories to search for (optional): none ▼ none Sbay_Contig646.20@ `409k 410k 411k 412k 413k 414k 415k 416k 417k 418k 4<mark>1</mark>9k 420k 421k 422k <mark>4</mark>23k 424k 425k 426k 427k 428k Search terms(s) should occur in these fields: Title Abstract Named gene YJL011C YJL007C YJL005W Add your own tract Add your own Search! YJL013C YJL008C YJL0060 Upload your own a View in CHap YJL007C <u>YJL</u>0110 YJL009W YJL005W 7 matches in 5 publication(s) found. YJL008C YJL006C YJL013C Chado Go to page 1 ▼ of 1 E-mail results to including matches. YJL009W.mRNA YJL007C.mRNA YJL005W.mRNA YJL014W.mRNA YJL011C.mRNA YJL010C.MRNA [view all matching sentences] [all citations in pdf] [all citations in endnote] [all citations in endnote (including abstracts YJL013C.mRNA YJL008C.mRNA YJL006C.mRNA Search Results tM(CAU)J2 tL(UAA)J tk(CUU)J tW(CCA)J Matching publications Links/Downloads tM(CAU)J2 tL(UAA)J tk(CUU)J Hazbun TR, Malmstrom L, Anderson S, Graczyk BJ, Fox B, Riffle M, Sundin BA, tW(CCA)J Chado is a modular, generic genomics schema. Aranda JD, McDonald WH, Chiu CH, Snydsman BE, Bradley P, Muller EG, Fields S, [view [PubMed] tW(CCA)J It is intended to model all data that might be Baker D, Yates JR 3rd and Davis TN (2003) Assigning function to yeast proteins by sentences] required of model organism database. Current Transposable elements integration of technologies. Mol Cell, 12, 1353-1365 14690591 modules include ones for sequence features, Update Image expression data, genotypes, phenotypes and Clear highlighting De Las Penas A, Pan SJ, Castano I, Alder J, Cregg R and Cormack maps. (Chris Mungall, David Emmert, Scott BP (2003) Virulence-related surface glycoproteins in the yeast pathogen Candida [view Cain, Allen Day) glabrata are encoded in subtelomeric clusters and subject to RAP1 and SIR-dependent transcriptional silencing. *Genes Dev*, 17, 2245-2258 12952896 Fay JC, McCullough HL, Sniegowski PD and Eisen MB (2004) Population genetic [PubMed] variation in gene expression is associated with phenotypic variation in sentences] Saccharomyces cerevisiae. Genome Biol, 5, P: R26 15059259 CMap Home | Maps | Map Search | Feature Search | Matrix | Map Sets | 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 [PubMed] Hsp90 chaperone system by two-hybrid analysis : potential uses and limitations of Reference Comparative this approach. Cell Stress Chaperones, 9, 359-368 15633294 S.cerevisia:_chromosom .bayanus_contig Lai LC, Kosorukoff AL, Burke PV and Kwast KE (2005) Dynamical remodeling of the CMap chr). [43] **Textpresso** CMap is a comparative Textpresso is a literature search tool; in this case, it is being map viewer. Here it is YJL033W hosted at SGD and linked via URL (SGD developers and Eimear being used to show YJL 03 0M synteny between two YJL0290 related yeast species. YJL 026H (Ben Faga and Ken Clark) YJL 025H YJL 0230 YJL 02 0 ___Sbay_Contig646.2 YJL019H --Sbay_Contig646.4 Database: Scott bio::::mart ~Sbay_Contig646.5 YJL013 Sbay_Contig646.6 Dataset: Scotts scer dset \Sbay_Contig646.7 Query summary FILTERS Sbay_Contig646.9 DATASETS Sbay_Contig646.10 □ FILTER by FEATURE: □ Scotts scer dset YJL006C The Generic Model Organism YJL005M Sbay_Contig646.14 2,441 entries total Gene SGD ✓ Feature type YJL004C Sbay_Contig646.15 YJL003M Sbay_Contig646.16 Feature type YJL0020 ☐ FILTER GENOME REGION: √Sbay_Contig646.18 Database Project YJL001H~ Chromosome Sbay_Contig646.19 YJR001H-Chromosome Feature LIST Sbay_Contig646.20 Attributes chrlX ▼ Chromosome Sbay_Contig646.22 http://www.gmod.org/ Sbay_Contig646.23 Region Start >= 360,001-440,001 bp Sbay_Contig646.24 Region End <= <u>Add dataset</u> ∭Sbay_Contig646.25 Creating a "MOD in a box" Refresh Chrom. Size >= Sbay_Contig646.26 Sbay_Contig646.27 Chrom. Size <= Coordinator: Scott Cain <cain@cshl.edu> bay_Contig646.28 bay_Contig646.29 □ FILTER REGIONS WITH FEATURES: (include,exclude) A collaboration of: 1-91,494 bp Only EcoCyc FlyBase GRAMENE ✓ Feature LIST Has CDS SGD Feature Types: Excluded Features in red have correspondences ■ FILTER REGIONS WITH FEATURES: (include,exclude) Evidence Types:

Menu Symbols:

i Map Set Info

Map Details

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.

BioMart

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