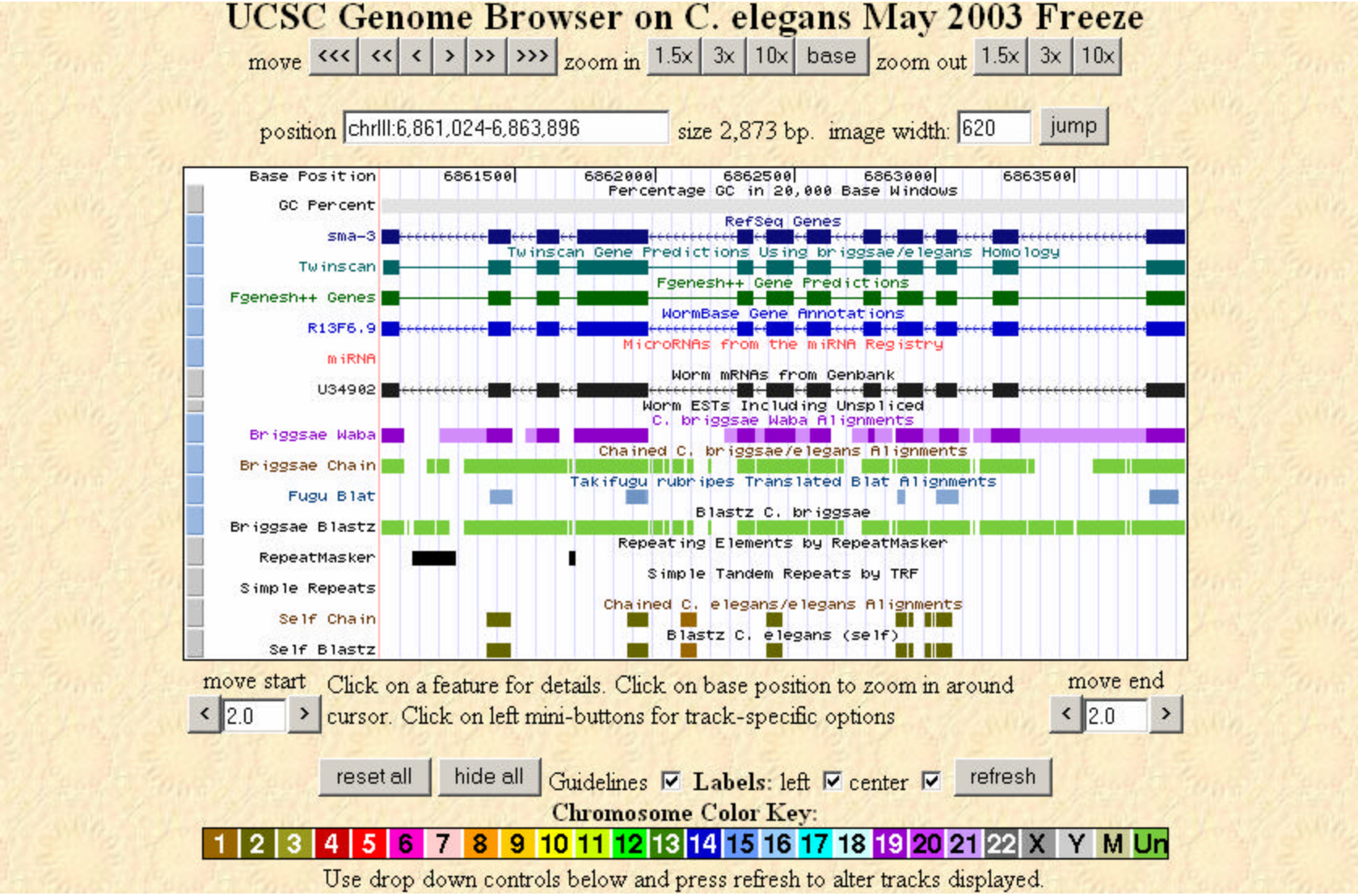
Generic Genome Browser

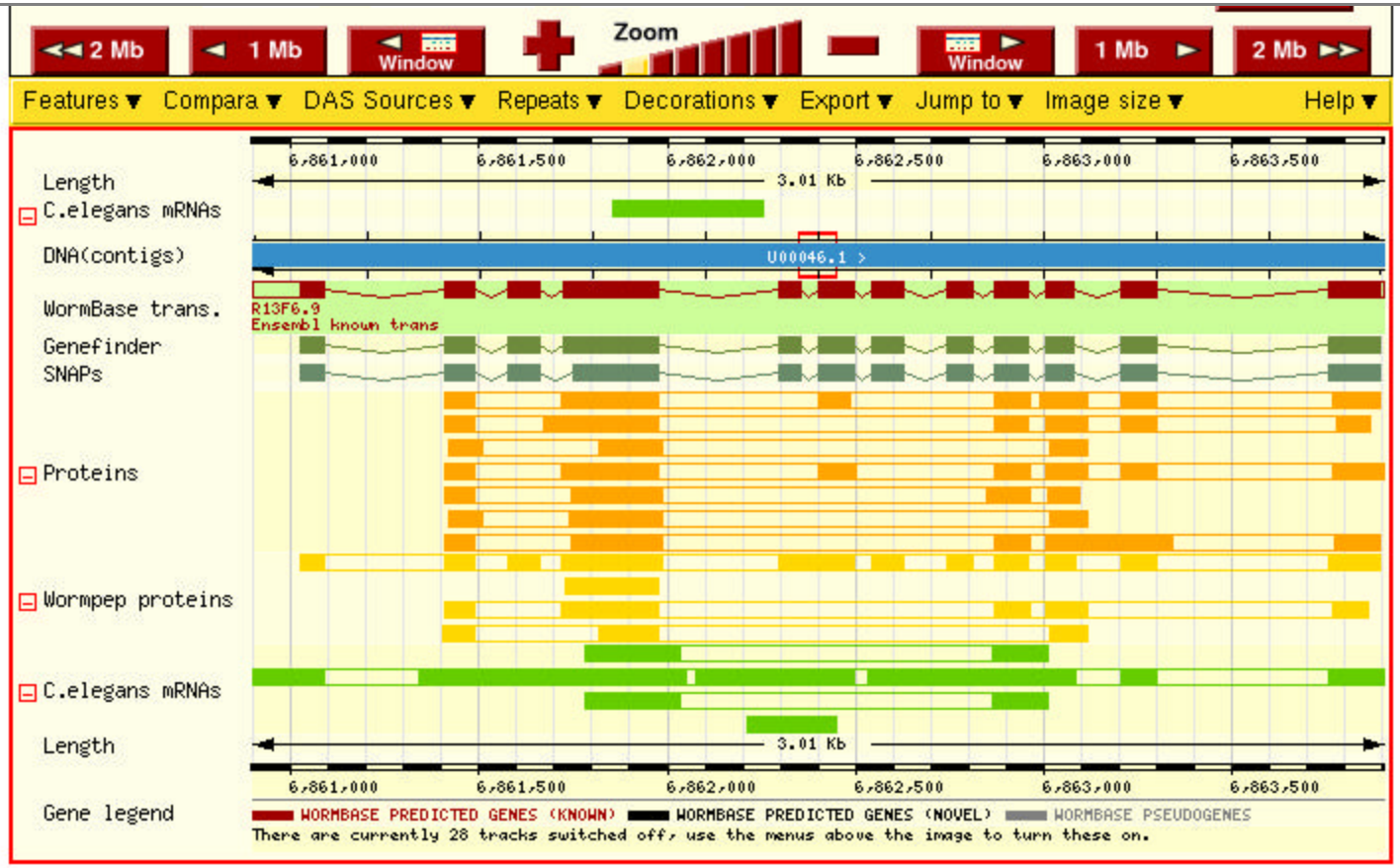
Lincoln Stein
Cold Spring Harbor Laboratory

Genome Browsers

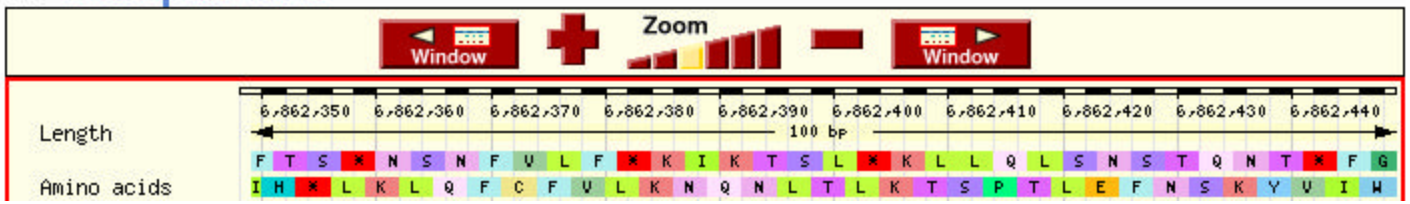
UCSC Genome Browser



EnsEMBL



Basepair view



NCBI Map Viewer

Map Viewer Home

Map Viewer Help
C.elegans Maps Help
Data As Table View
Maps & Options

Compress Map
Region Shown:
6,858K
6,866K
Go

out
zoom
in

Clone

default
master

Caenorhabditis elegans BLAST The Genome

Chromosome: I II [III] IV V X

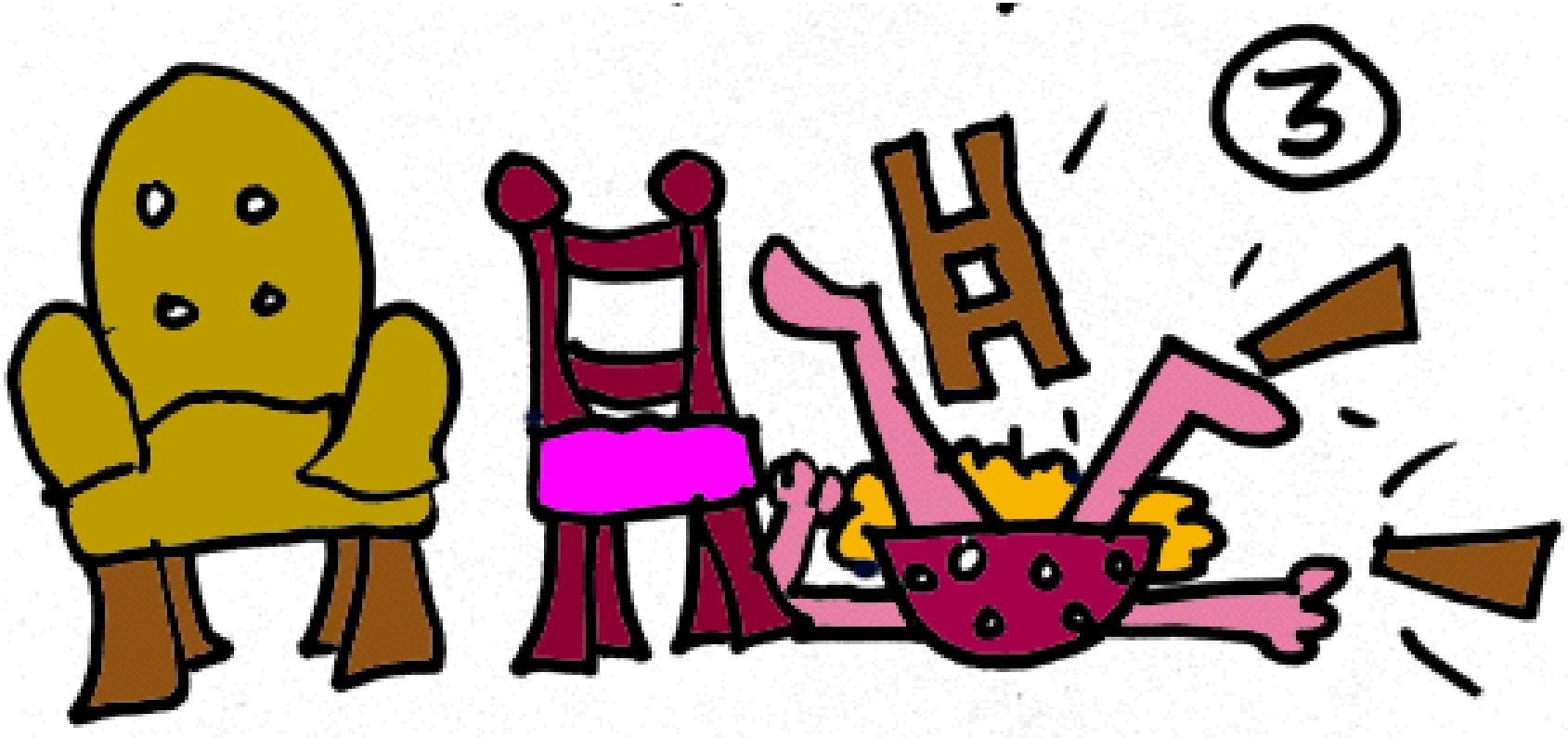
Query: sma-3 [clear]

Master Map: Gene View Summary Maps & Options

Region Displayed: **6,858K-6,866K bp** Download/View Sequence/Evidence

Gene	Gene	O	LinkOut	E	cM	Description
	3H948 ↑	WG	-	sv pr	-0.97	putative secreted or extracellular protein family member precursor (3H948
	3H949 ↓	WG WB		sv pr	-0.97	predicted CDS, putative secreted or extracellular protein family member p
	sma-3 ↑	WG	-	sv pr	-0.91	SMAll body size SMA-3, dwarfin family member SMA-3 (44.5 kD) (sma

Oops!

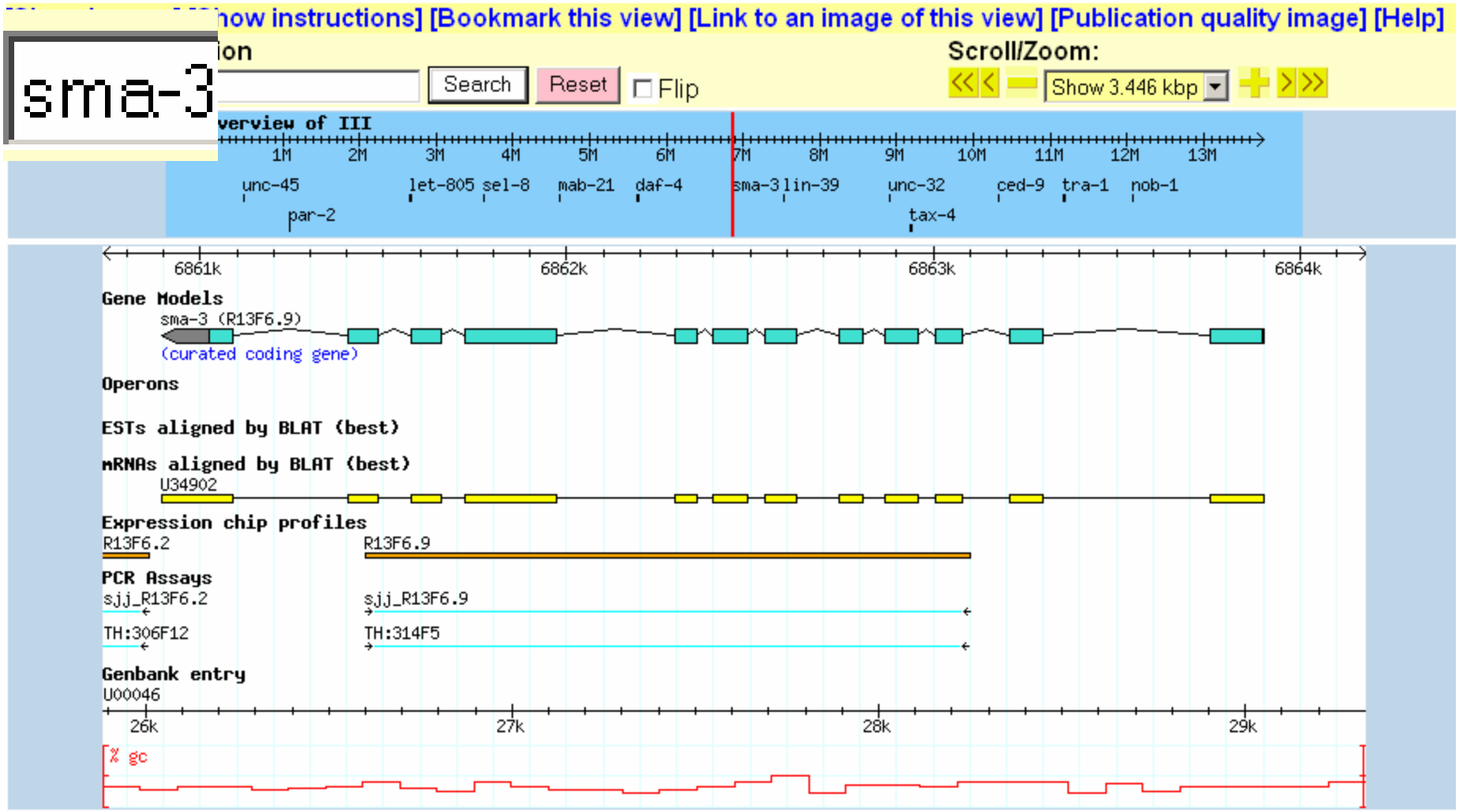


Gbrowse: Genome Browsing that's "Just Right"

- Part of Generic Model Organism Database Project
- Lightweight & portable
- Highly configurable
- Clean user interface
- Promotes data sharing

Find your favorite gene

Showing 3.446 kbp from III, positions 6,860,736 to 6,864,181

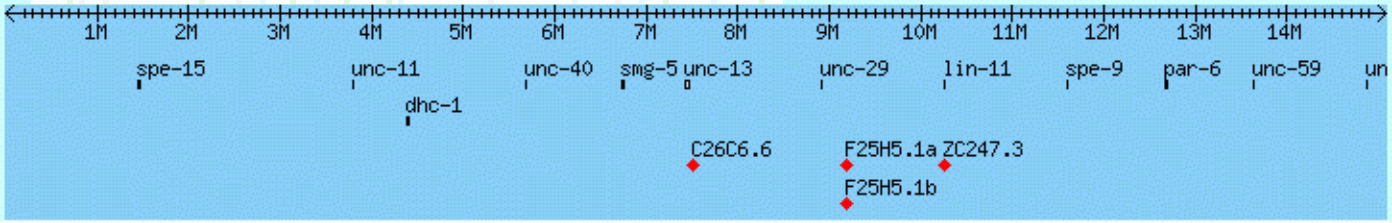


Find Your Favorite Domain

lim domain

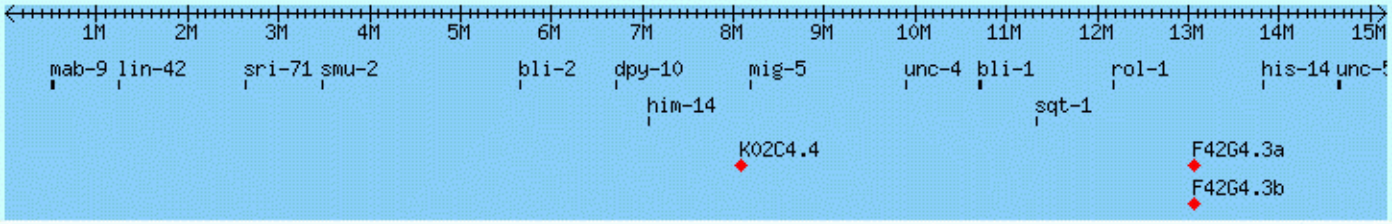
Landmark or F

n your request.



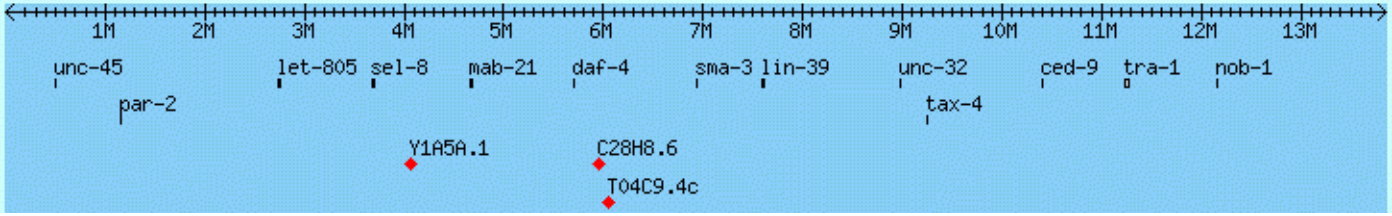
ZC247.3	**LIM domain** containing proteins (2 **domain**s), homeobox protein (**lim** subfamily)	I:10,256,394..10,263,125 (6732 bp)	score=18.00
F25H5.1a	**LIM domain** containing proteins (6 **domain**s)	I:9,190,482..9,200,886 (10405 bp)	score=14.44
F25H5.1b	**LIM domain** containing proteins (5 **domain**s)	I:9,191,132..9,200,886 (9755 bp)	score=14.44
C26C6.6	**lim domain**s	I:7,514,034..7,514,708 (675 bp)	score=8.85

Matches on II

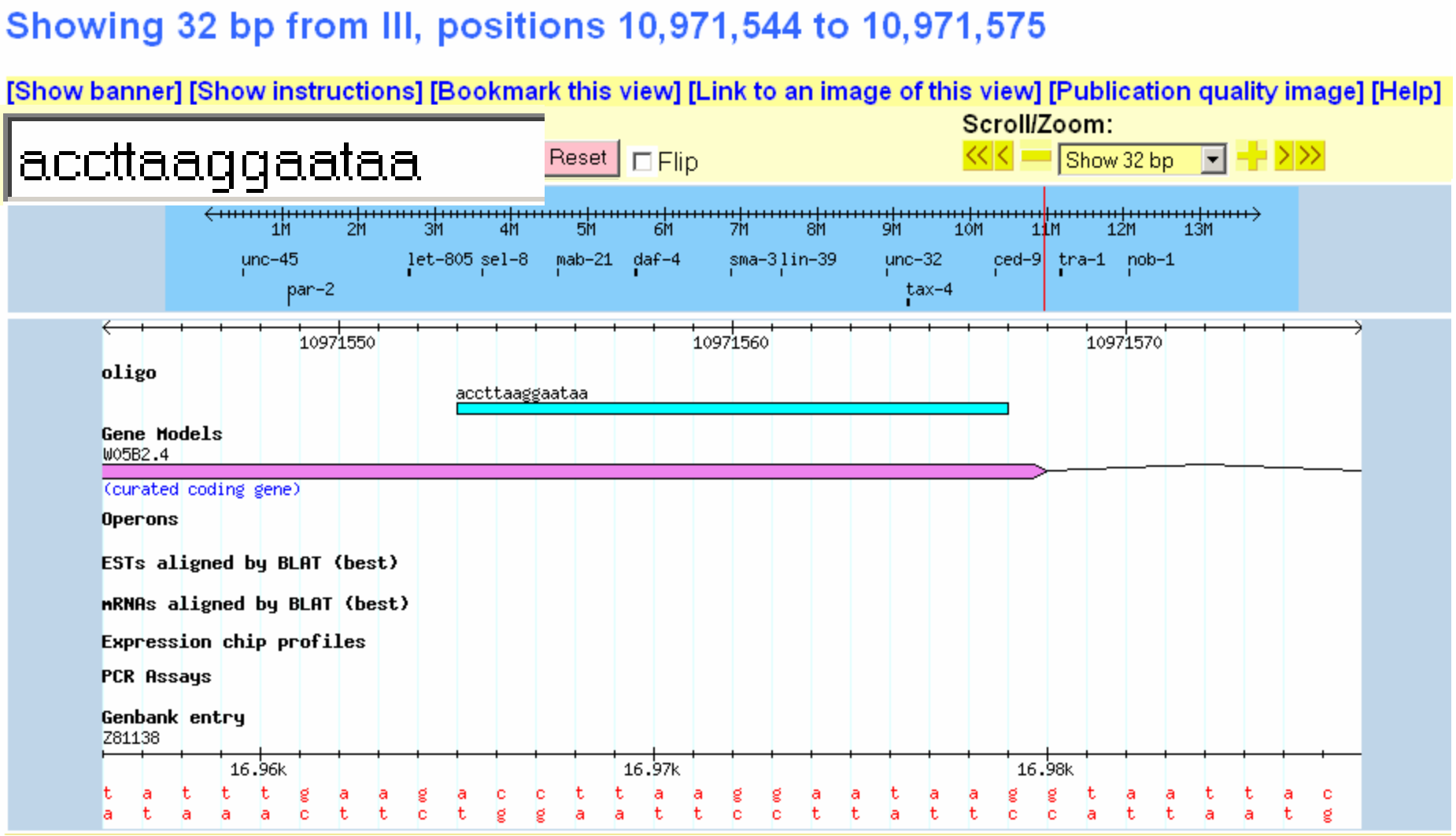


K02C4.4	**LIM domain**	II:8,087,722..8,092,611 (4890 bp)	score=14.76
F42G4.3a	**LIM domain** containing proteins	II:13,057,045..13,070,542 (13498 bp)	score=14.60
F42G4.3b	**LIM domain** containing proteins	II:13,068,689..13,070,542 (1854 bp)	score=14.60

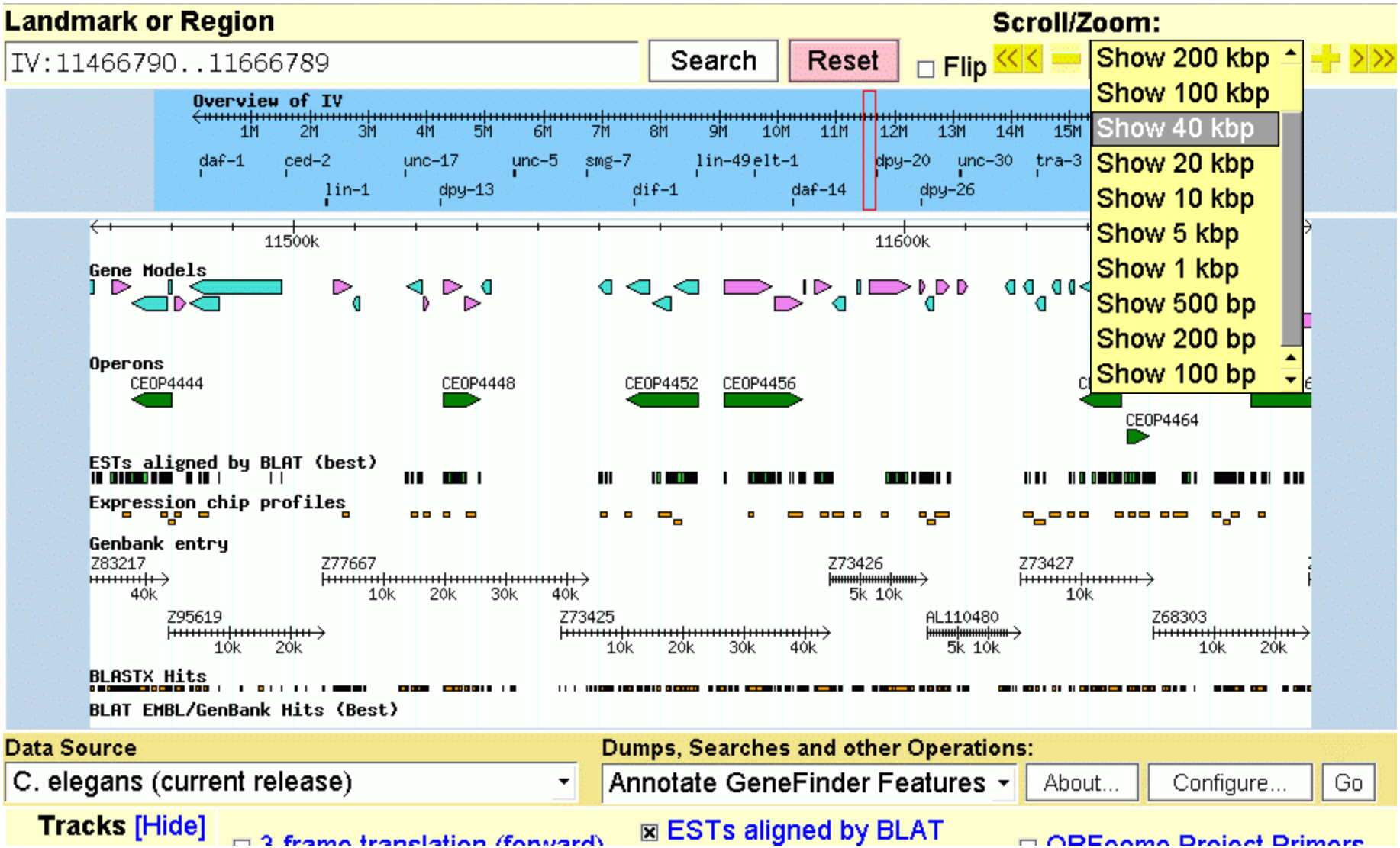
Matches on III



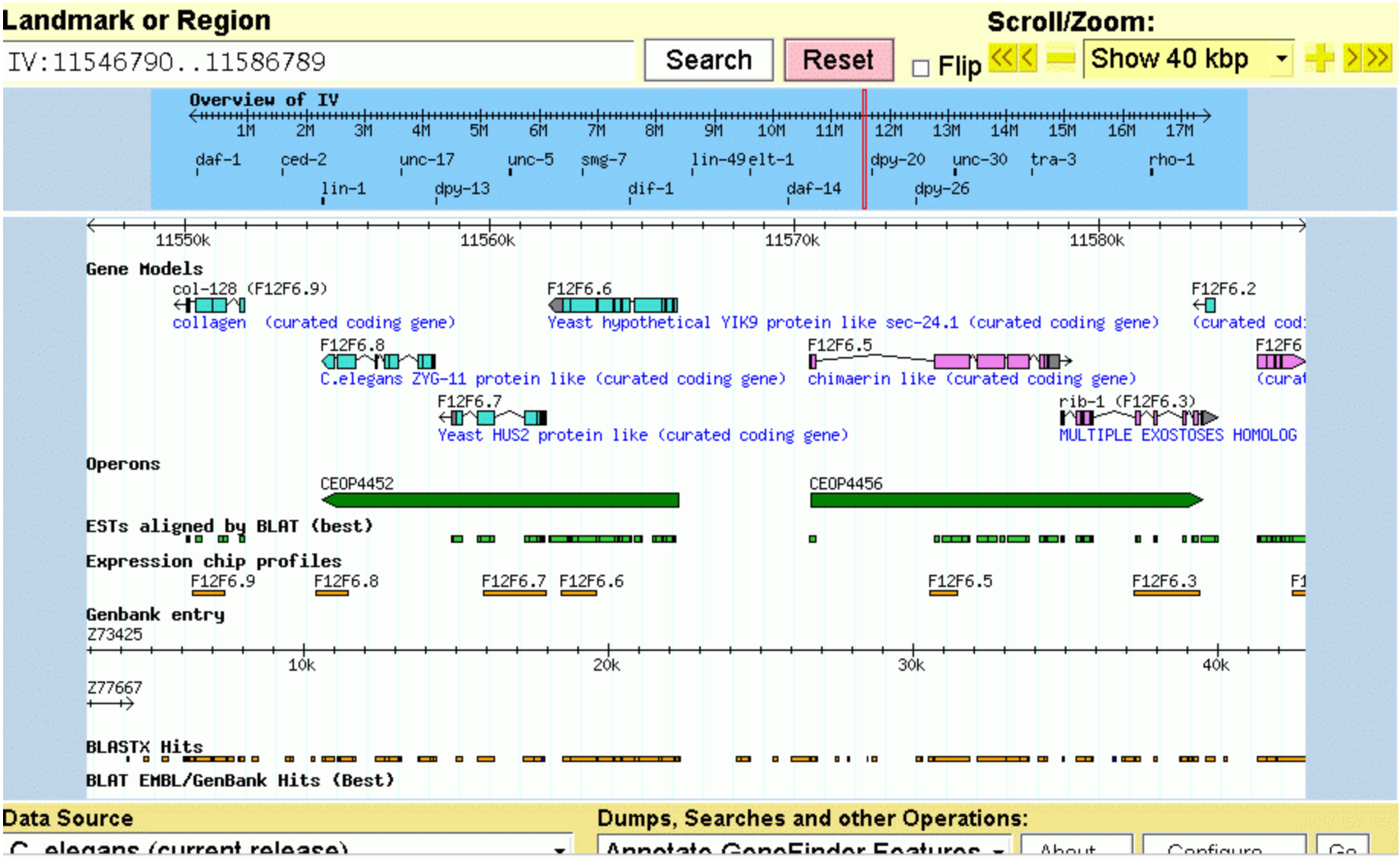
Find Your Favorite Oligo



Zoom – Low Magnification

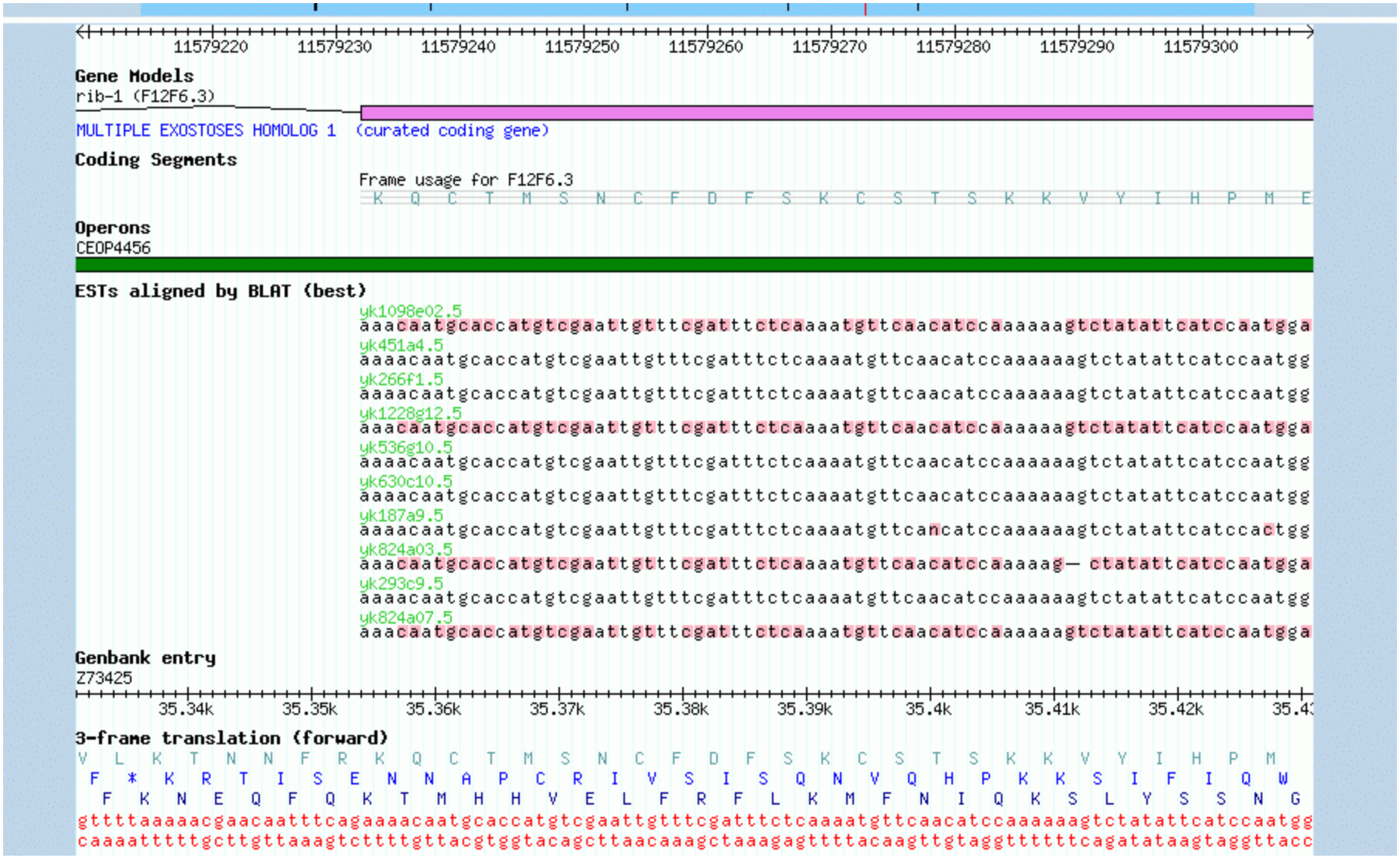


Zoom – Medium Magnification



Zoom – High Magnification

Zoom – Highest Magnification



Take-Home Reports

Alignments for X:6889692,6896579

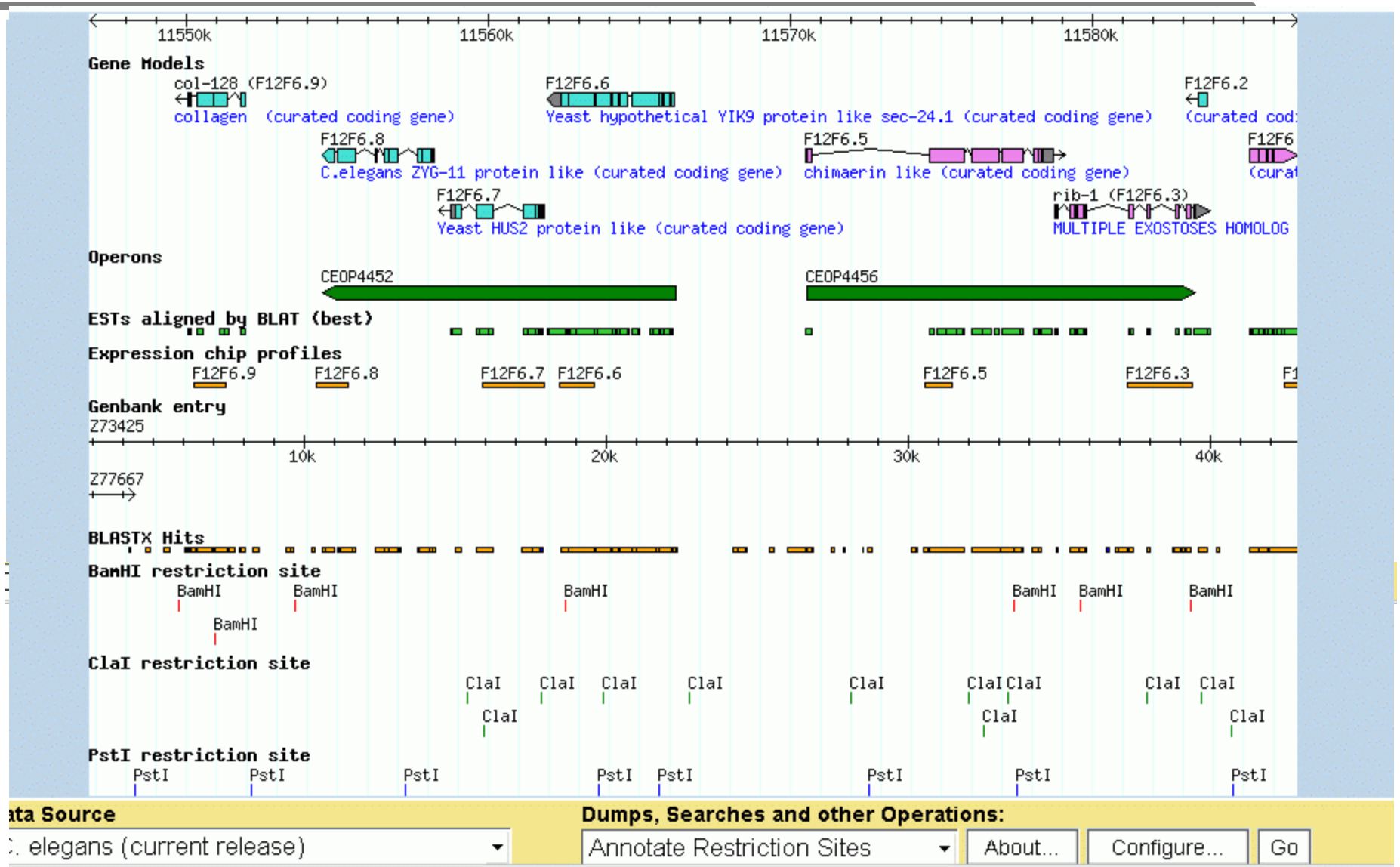
```
         X 6889692 .......... .......... .....ATGAT CACATCAGTA TTGCGCTATG TGCTAGCGCT CTAC
 yk449d8.5       1 ...ctttaca ccctttcaca caaacatgat cacatcagta ttgcgctatg tgctagcgct ctac
 yk523f4.5      14 ccactttaca ccctttcaca caaacatgat cacatcagta ttgcgctatg tgctagcgct ctac
 yk254a4.5      22 ccactttaca ccctttcaca caaacatgat cacatcagta ttgcgctatg tgctagcgct ctac
 yk492g9.5      27 ccactttaca ccctttcaca caaacatgat cacatcagta ttgcgctatg tgctagcgct ctac

         X 6889747 CTCATGGAGC ATACTTTTCA CAGgtagcac ctagggtgtt tttatgagta aacatttgtt tttt
 yk449d8.5      78 ctcatggagc atacttttca cag------- ---------- ---------- ---------- ----
 yk523f4.5      94 ctcatggagc atacttttca cag------- ---------- ---------- ---------- ----
 yk254a4.5     102 ctcatggagc atacttttca cag------- ---------- ---------- ---------- ----
 yk492g9.5     107 ctcatggagc atacttttca cag------- ---------- ---------- ---------- ----

         X 6889827 GAGCCCCAGA CCATGATCCT TGCCATGATC ATACTGGTCG ACCAGTTCGA TGTGTTCCCG AGTT
 yk398e4.5       1 .......... .......... .......... .......... .......... .......... ....
 yk449d8.5     111 gagccccaga ccatgatcct tgccatgatc atactggtcg accagttcga tgtgttcccg agtt
 yk523f4.5     127 gagccccaga ccatgatcct tgccatgatc atactggtcg accagttcga tgtgttcccg agtt
 yk254a4.5     135 gagccccaga ccatgatcct tgccatgatc atactggtcg accagttcga tgtgttcccg agtt
 yk492g9.5     140 gagccccaga ccatgatcct tgccatgatc atactggtcg accagttcga tgtgttcccg agtt

         X 6889907 GGAAAACCTG TTATTGCTAG TGATACATGC GG-AACAAAC CGACCAGACA Agtaagttta tttt
yk584g12.5       1 .......... .......... .......... ......aaac cgaccagaca a--------- ----
```

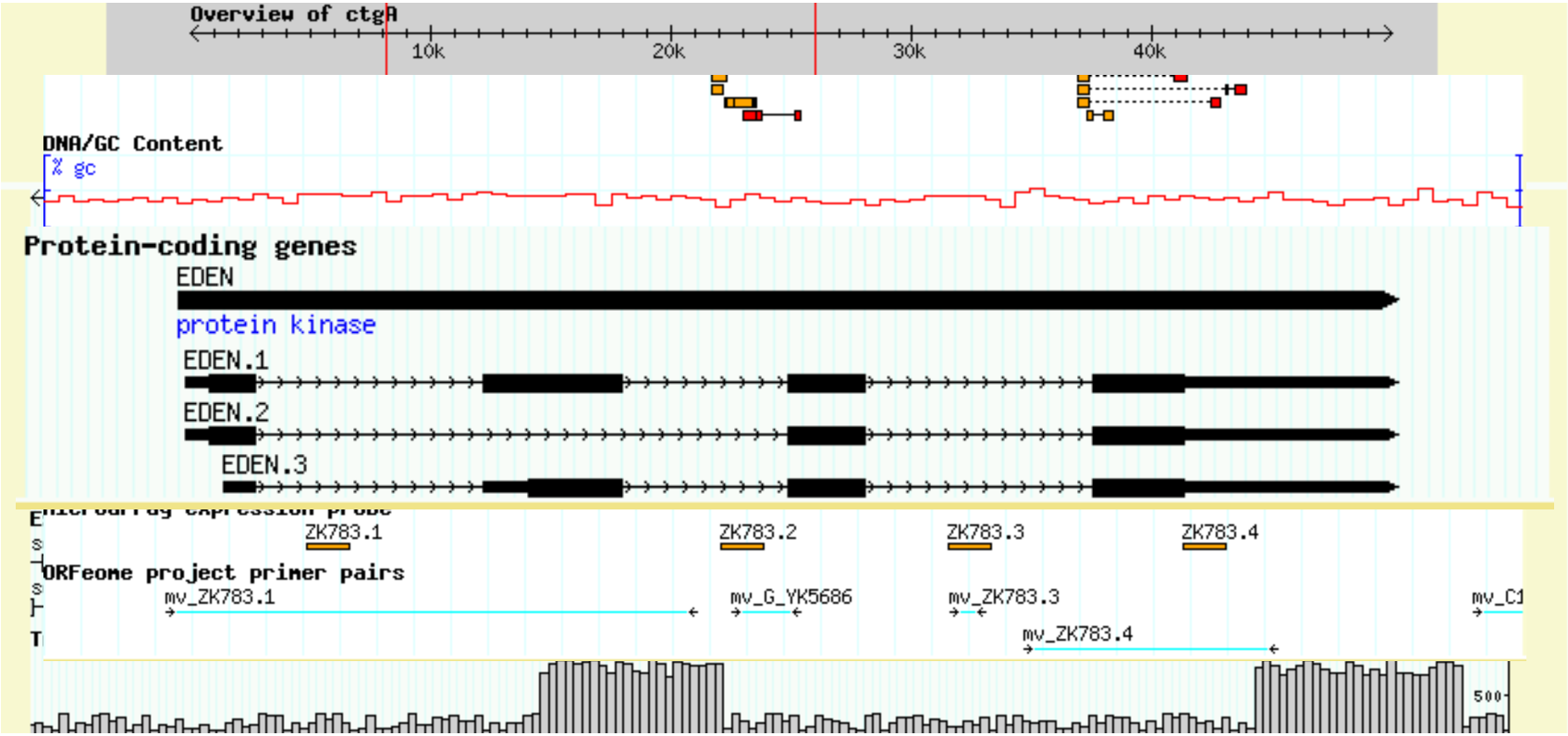
Custom Tracks



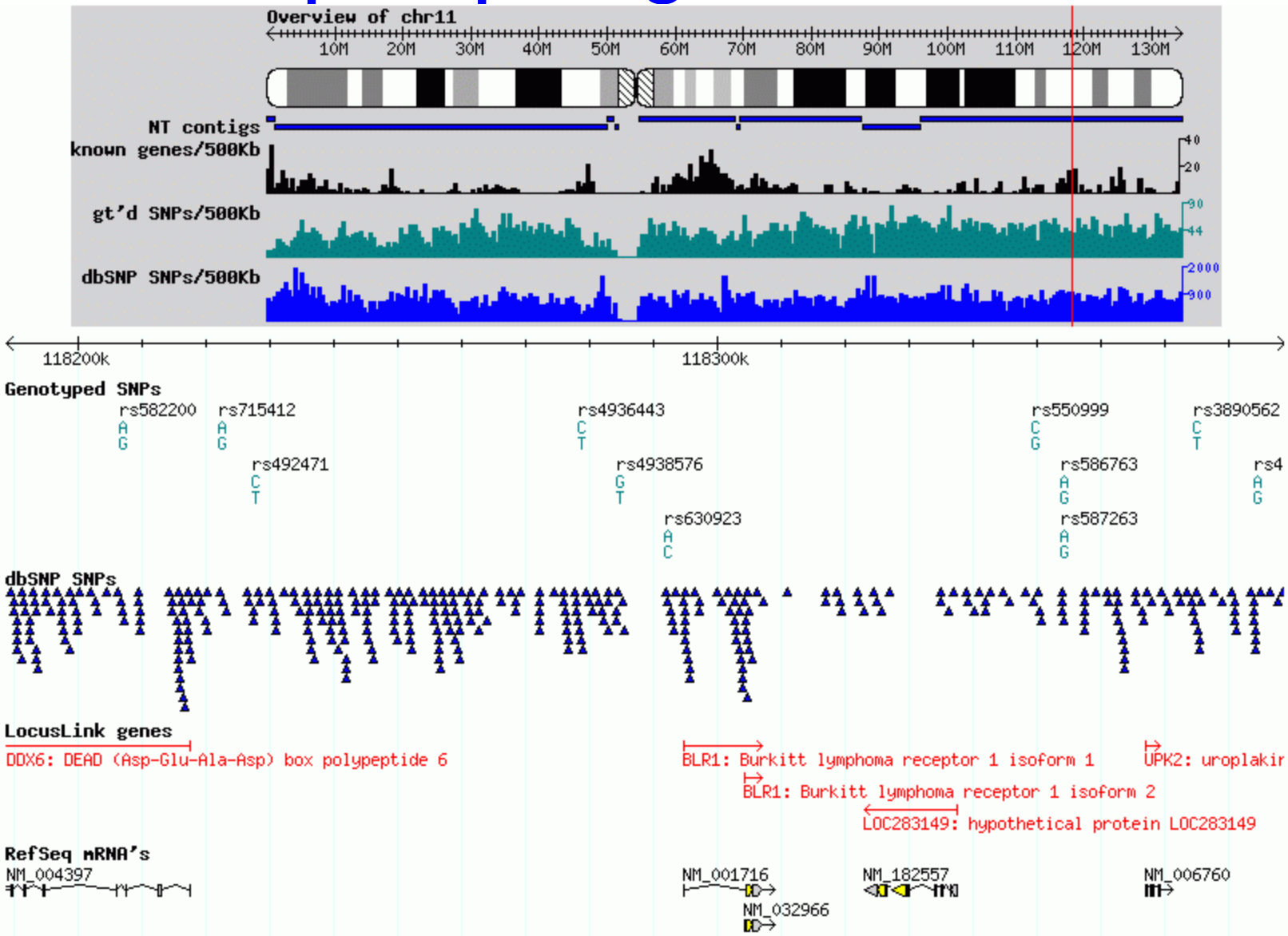
Extensively Customizable

- End-user
  - Turn tracks on and off, change order, change packing & labeling attributes (stored in cookie)
- Data provider
  - Change fonts, colors, text.
  - Change overview – genetic map, contigs, coverage, karyotype.
  - Define new tracks using simple config file.
  - Tinker with track appearance to hearts content.

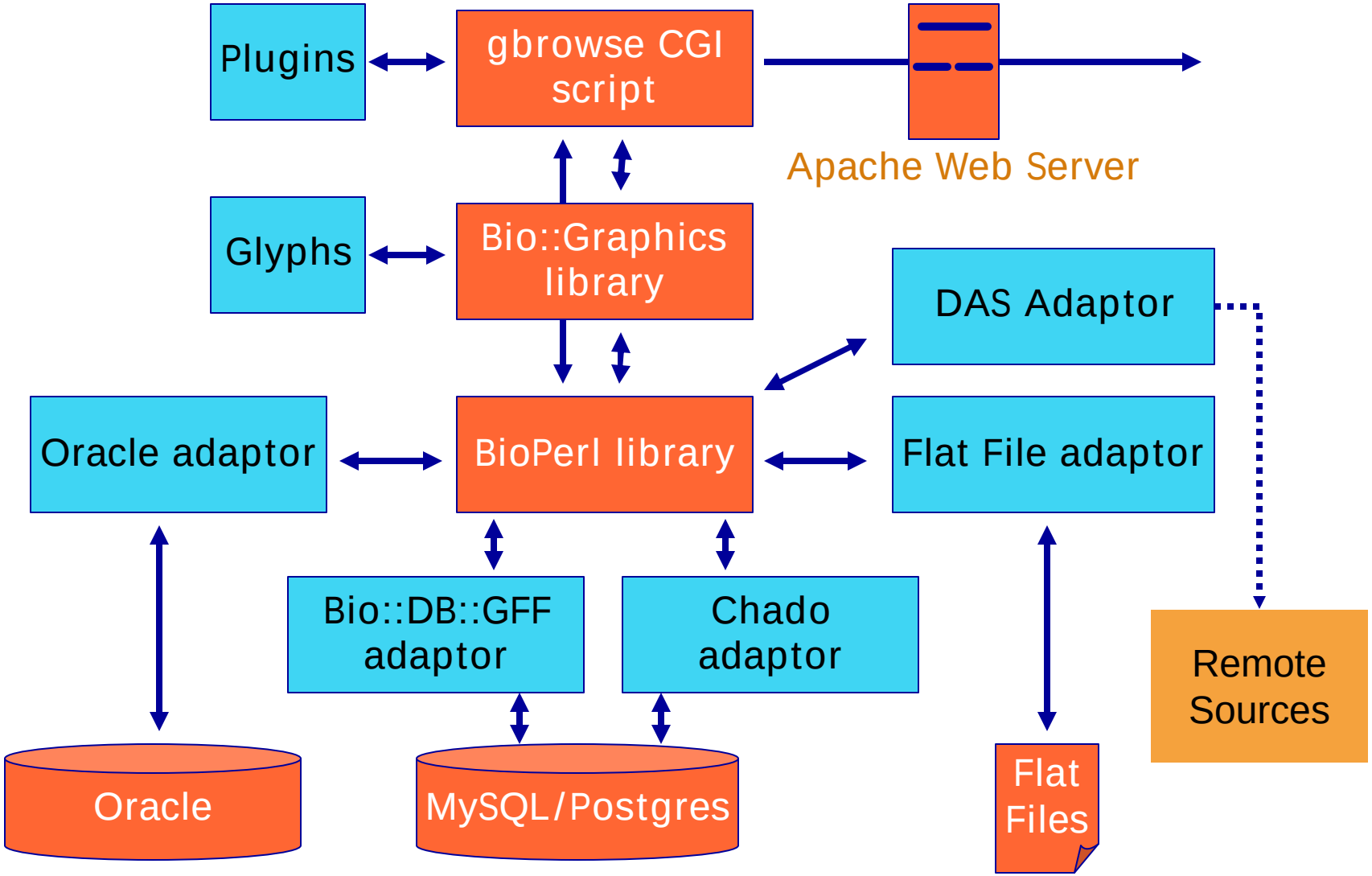
Config Examples from Tutorial



www.hapmap.org



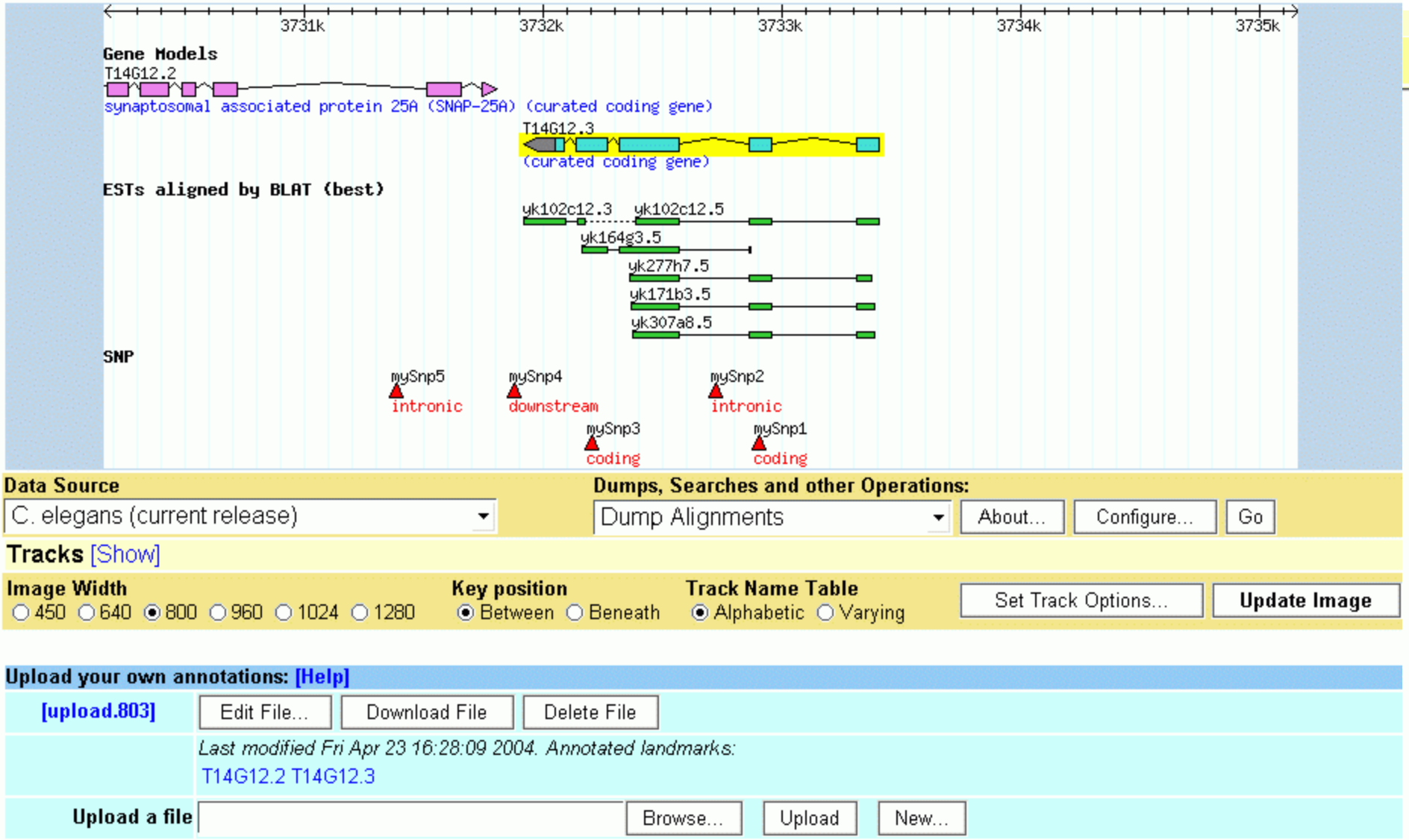
Extensively Extensible



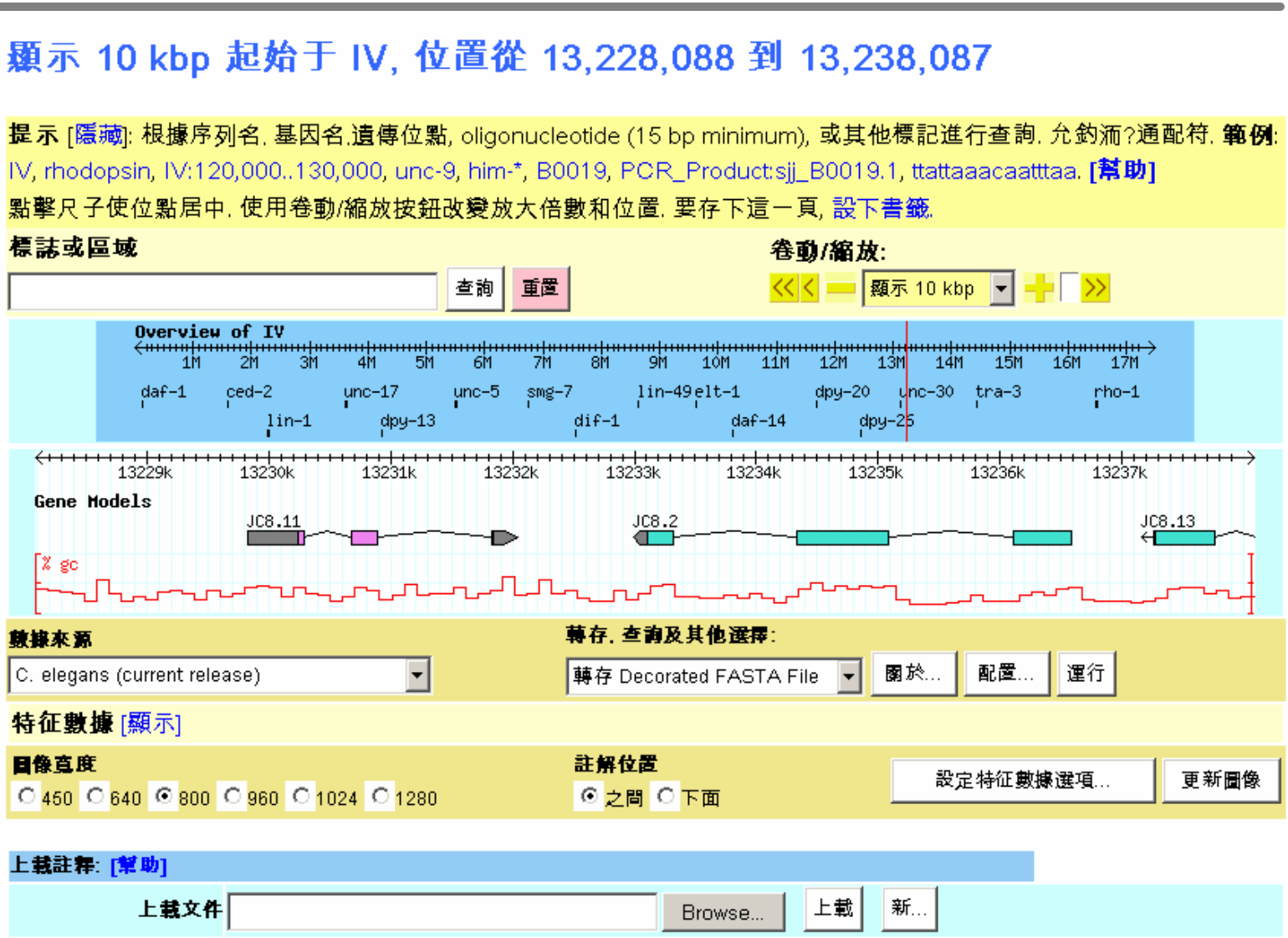
Data Sharing

- Add your own tracks
- Add 3d party tracks
- Distributed annotation system support

Adding your Own Tracks



Internationalization



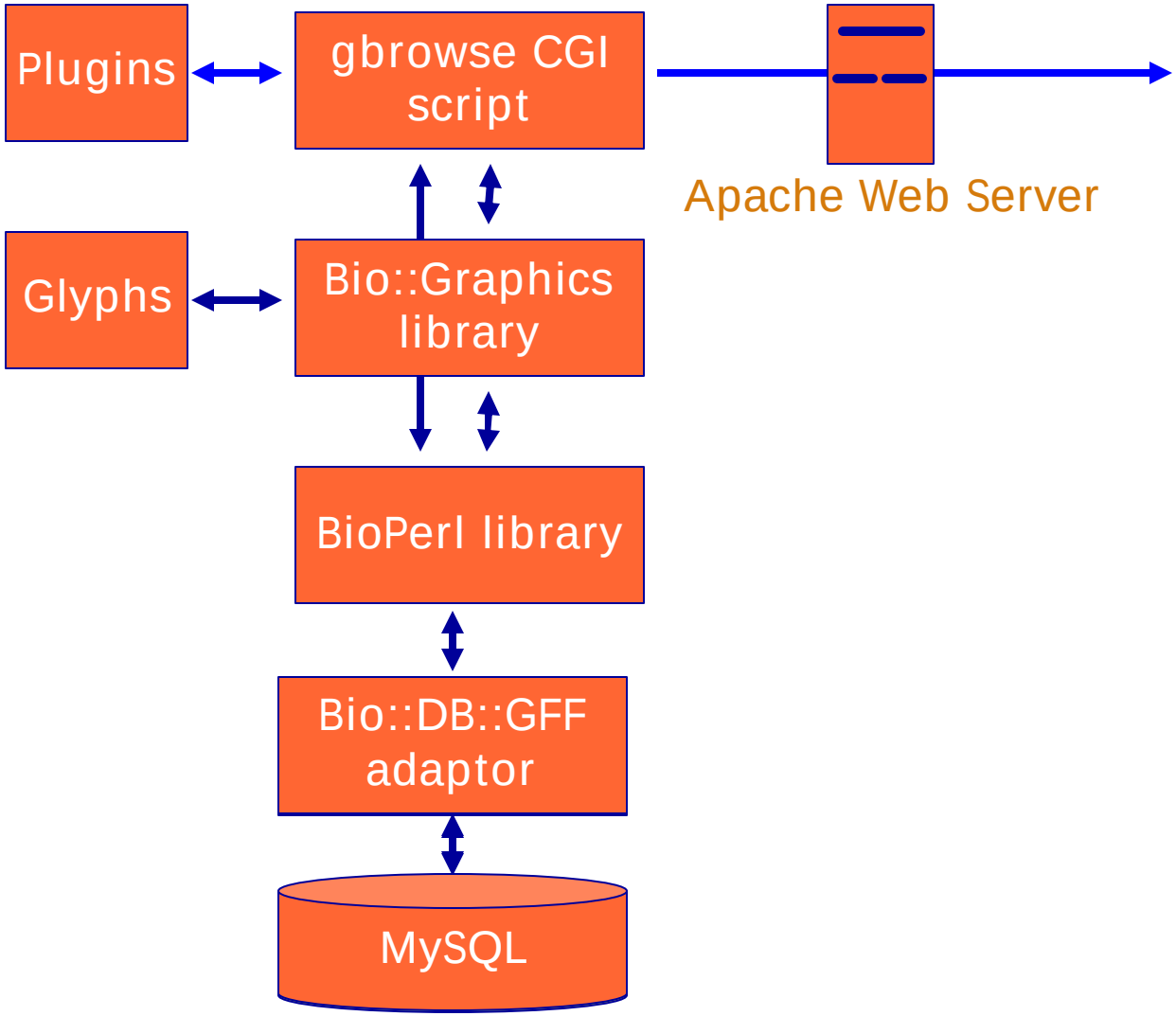
Adding 3d Party Tracks

C. elegans gene knockout consortium

Deletion allele gk109

gk109	Gene	T14G12.3
	Strain	VC195
	Locus	tag-18
	PCR	KO386-2P
	Type	Deletion_with_insertion
	Library	29
	Mutagen	EMS
	Reference_sequence	T14G12
	Left_Breakpoint	4756
	Right_Breakpoint	5214
	Deletion_Size	457
	Flanking_Sequences	gttgagagatttgggtaggatgtgactctt
		tttgtcgcgactgattttaccgctggctcc
	Insertion_Sequence	ttt

GBrowse on GenBank!



B. burgdorferi via GenBank proxy

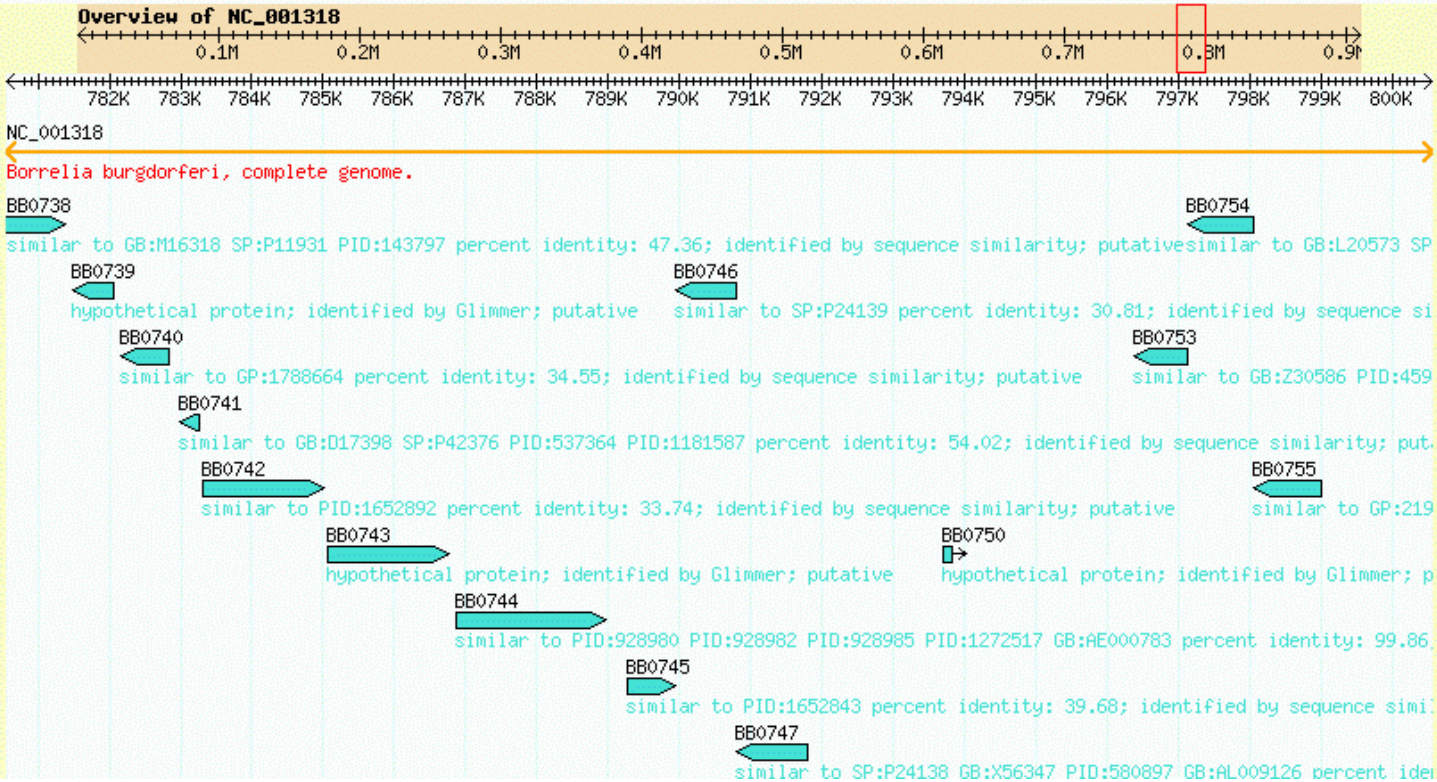
Genbank/EMBL proxy

Showing 20 Kbp from NC_001318, positions 780,559 to 800,558

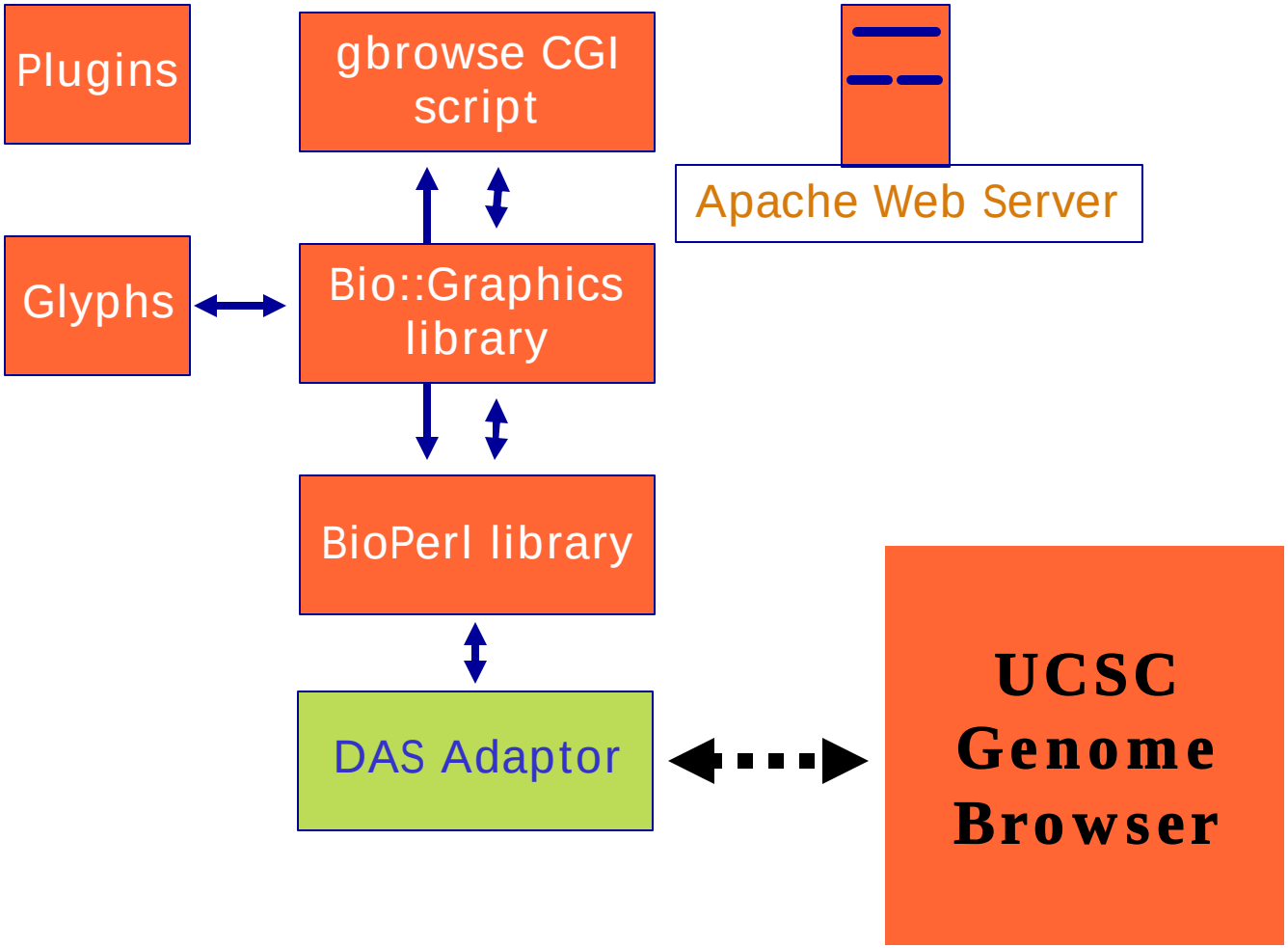
Search using a sequence name, gene name, locus, oligonucleotide (15 bp minimum) or other landmark. The wildcard character "*" is allowed. **Examples:** CEF58D5, NC_001320, AP003256. **[Help]**

To center on a location, click the ruler. User the Scroll/Zoom buttons to change magnification and position. To save this view, bookmark this link.

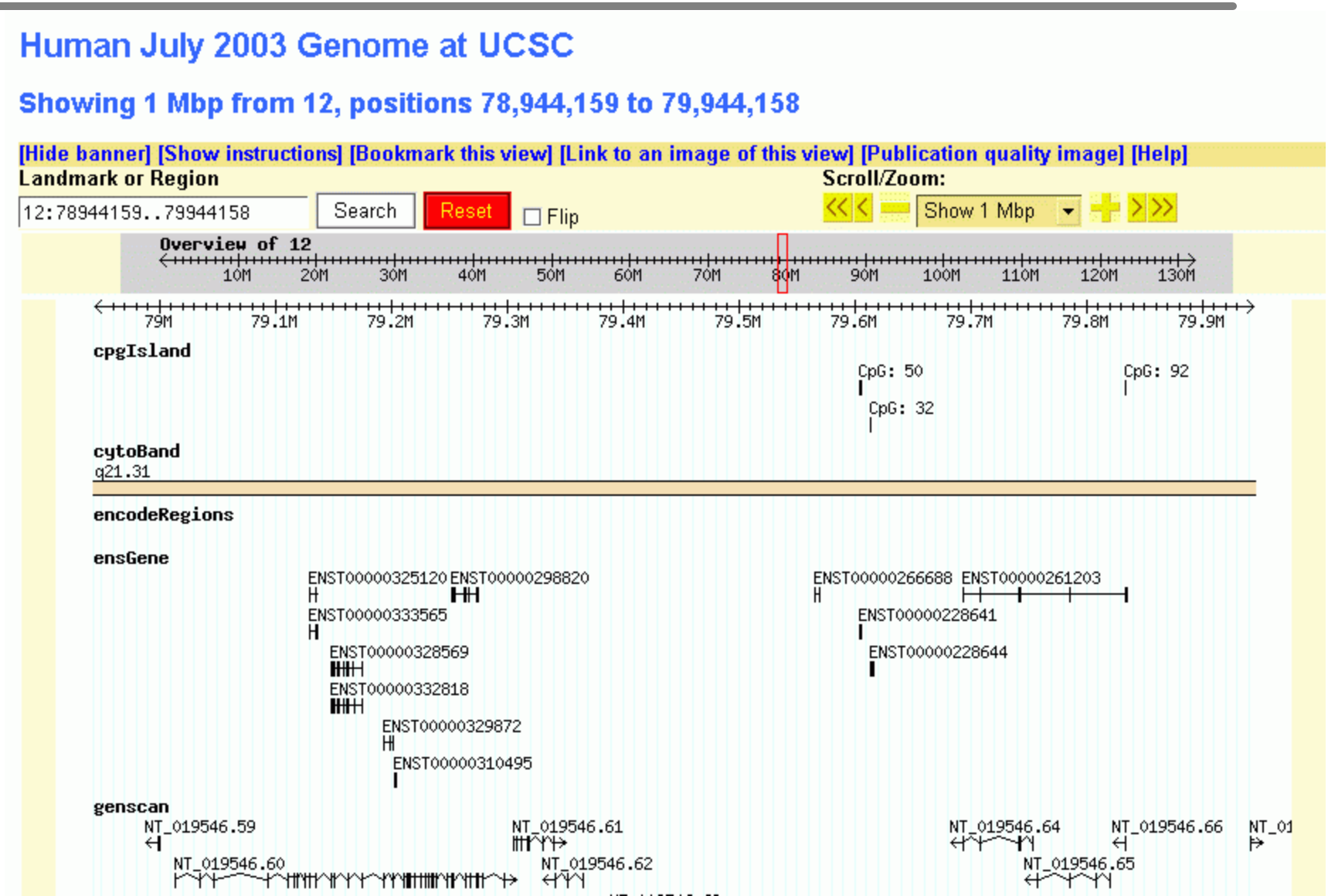
Landmark or Region NC_001318:780559..800558 Go Scroll/Zoom: Show 20 Kbp



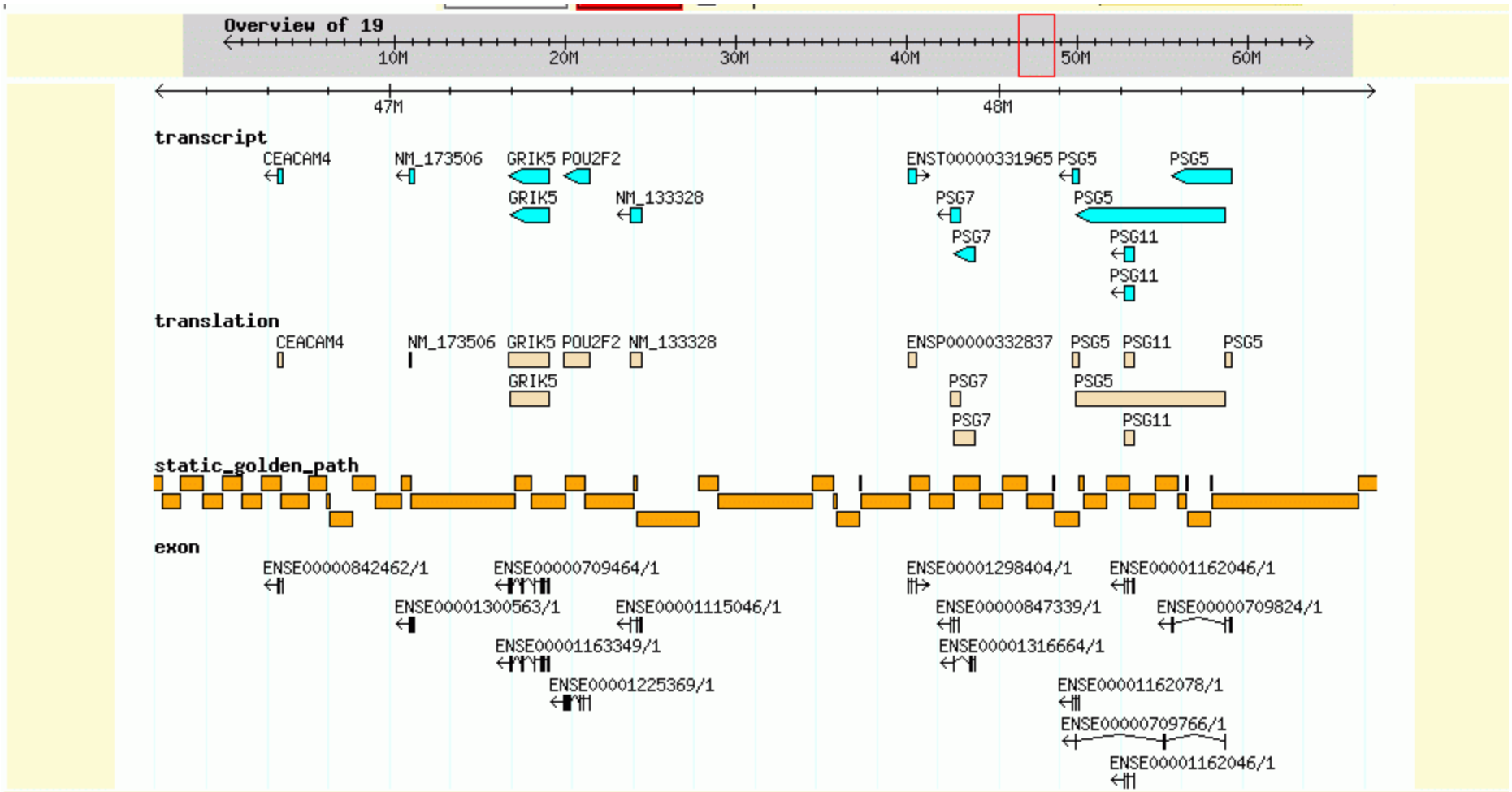
Gbrowse on Remote Sources



DAS: Gbrowse on UCSC



DAS: Gbrowse on EnsEMBL



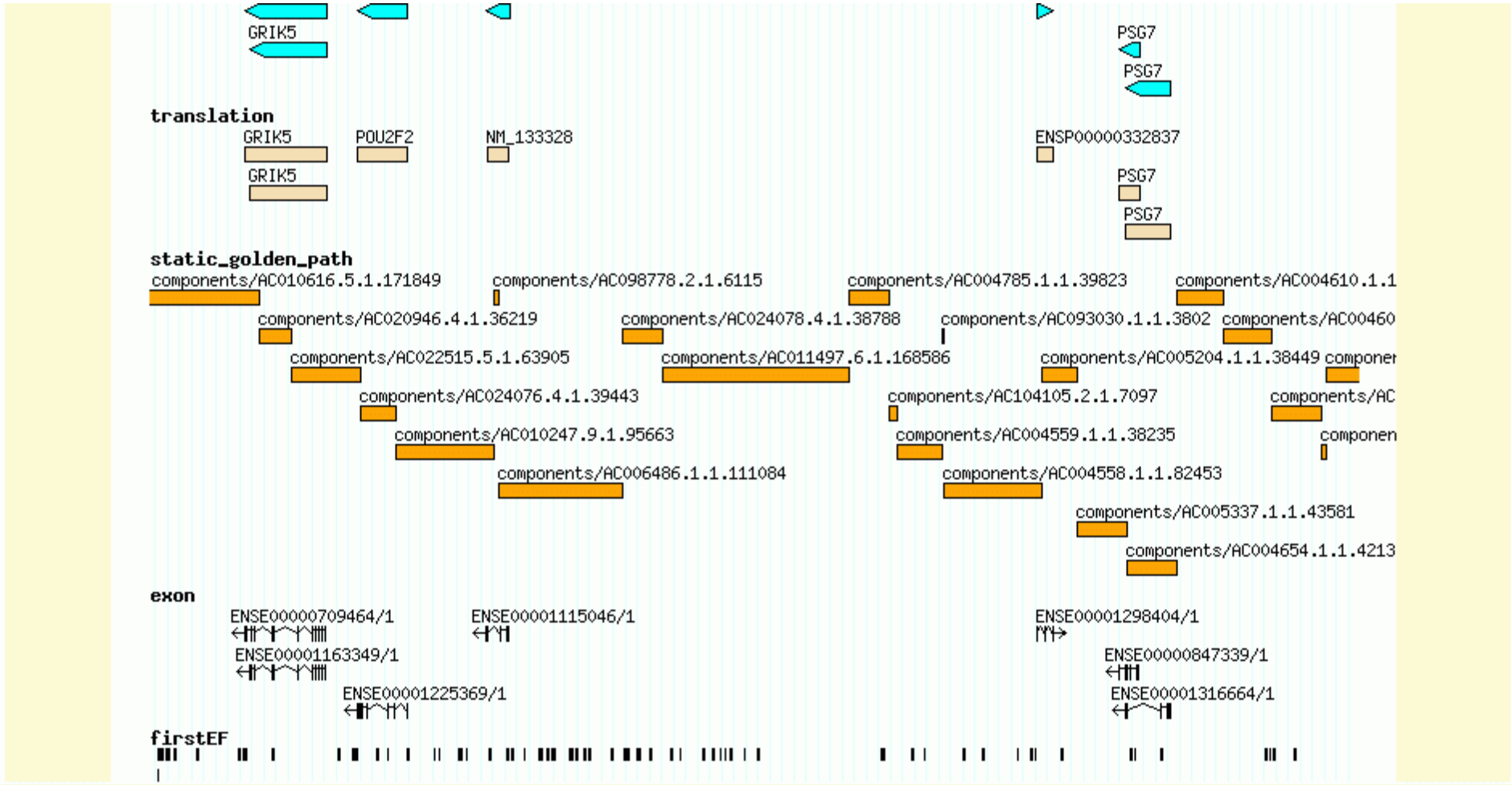
Data Source

The human genome from Ensembl

Tracks [Hide]

- ☐ annotation
- ☐ gene
- ☐ static_golden_path_clone
- ☐ cpg_island
- ☐ mapping
- ☐ sts

DAS: Gbrowse on both



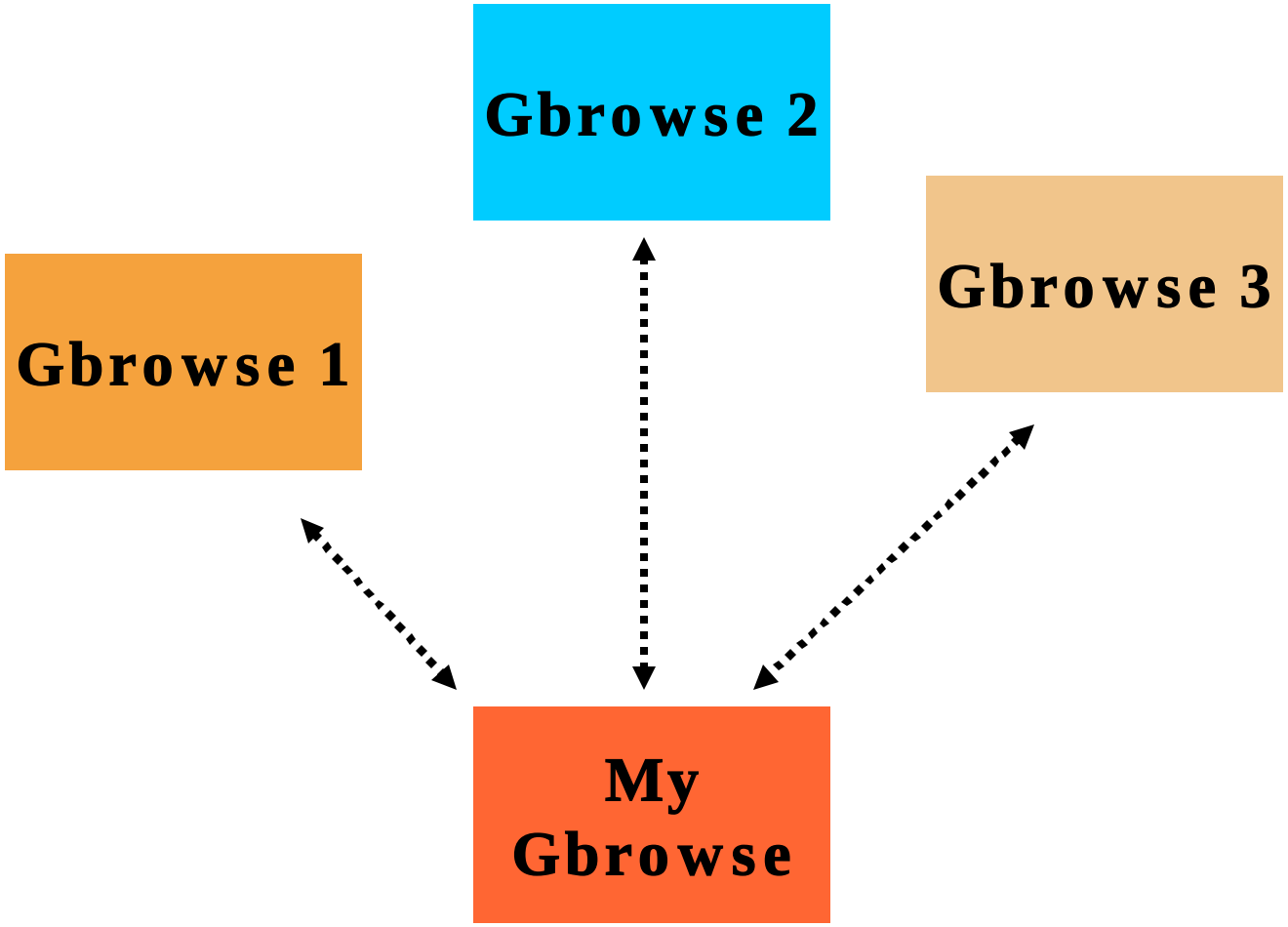
Data Source

The human genome from Ensembl

Tracks [Hide]

- ☐ annotation
- ☒ *http://genome.ucsc.edu/cgi-bin/das/hg16?type=firstEF*
- ☐ sts
- ☐ cpg_island
- ☐ mapping
- ☐ tandem_repeat

Distributed GBrowse



Installation

- “One Click” install on Macintosh OSX
- “Three Click” install on Windows XP
- “Tippity-tappity-tap” install on Linux/Solaris/BSD Unix

Known Gbrowse Users

Agricultural Biotechnology Center, Hungary
BAWI, S. Korea
Baylor College of Medicine
Biocrates GmbH, Innsbruck
Brandeis University
Bristol-Meyers Squibb
British Columbia Centre for Diseaes Control
CIRAD, France
CSIRO, Australia
Cambridge University (multiple labs)
Center for Genomics & Bioinformatics, Stockholm
Centre de Genetique Moleculaire, CNRS
Cold Spring Harbor Laboratory (multiple labs)
Compugen
Concordia University, Canada
Cornell Medical School
Cornell University
DNA Landmarks, Inc.
Donald Danforth Plant Sciences Center
Duke University (multiple labs)
EMBL, Heidelberg
Faculdade de Medicina de Ribeirão Preto
FlyBase
Foundation for Research and Technology, Crete
Fundação Hemocentro, Sao Paolo
Genoscope, France
GrainGenes
Harvard University
Hospital for Sick Kids, Toronto
Illinois Institute of Technology
Incyte Corporation
Inpharmatica, Ltd.
Institute for Systems Biology, Seattle
Institute of Molecular and Cell Biology, Singapore
International Rice Research Institute, Phillipines
John Innes Centre
KEGG
Kansas State University
Karolinska Institute
Kennedy Krieger Institute
Lawrence Berkeley Laboratories
Marine Biological Laboratories, Woods Hole
Massachusetts Institute of Technology (multiple labs)
Mayo Institute
McGill University
Meat Animal Research Center, University of Nebraska

Known Gbrowse Users

Open Informatics Consulting Corp.
Oxagen Corp.
Pasteur Institute, Paris
Pioneer Corporation
QIAGEN Operon Corp.
RIKEN (multiple labs)
RatDB
Regulome, Inc.
Rhobio (Bayer CropScience SA & Biogemma joint venture)
Rigshospitalet, Copenhagen
Rockefeller University
Roslin Institute, Edinburgh
Russian Academy Medical Sciences
Serono International Corp, Geneva
Simon Frasier University
South Africa National Bioinformatics Institute
Southern Illinois University
St. Jude Children's Research Hospital, Memphis
Stowers Institute for Medical Research
Texas A&M (multiple labs)
The Institute for Genome Research
Tulane University
University California Davis
University of Arizona (multiple labs)
University of British Columbia
University of California Santa Barbara
University of Georgia (multiple labs)
University of Minnesota
University of Muenster
University of New South Wales, Australia
University of Oklahoma (multiple labs)
University of Pennsylvania (multiple labs)
University of Southern California
University of Texas
University of Toronto
University of Virginia
University of Washington
Universität Giessen
Université de Liège, Belgium
Wageningen Universiteit & Researchcentrum, Netherlands
Washington University at St. Louis (multiple labs)
WormBase
deVGen, Belgium

Just Right