Generic Model Organism Database Project

Lincoln Stein

Cold Spring Harbor Laboratory

Model Organism Databases

- Curated resources for a research community
  - Genomes, maps, stocks, phenotypes, anatomies, pathways, interactions, expression patterns, and much more.
- Many, many organisms
- Same problems
- Wheel reinvented many times
- This madness must end!

The GMOD Project

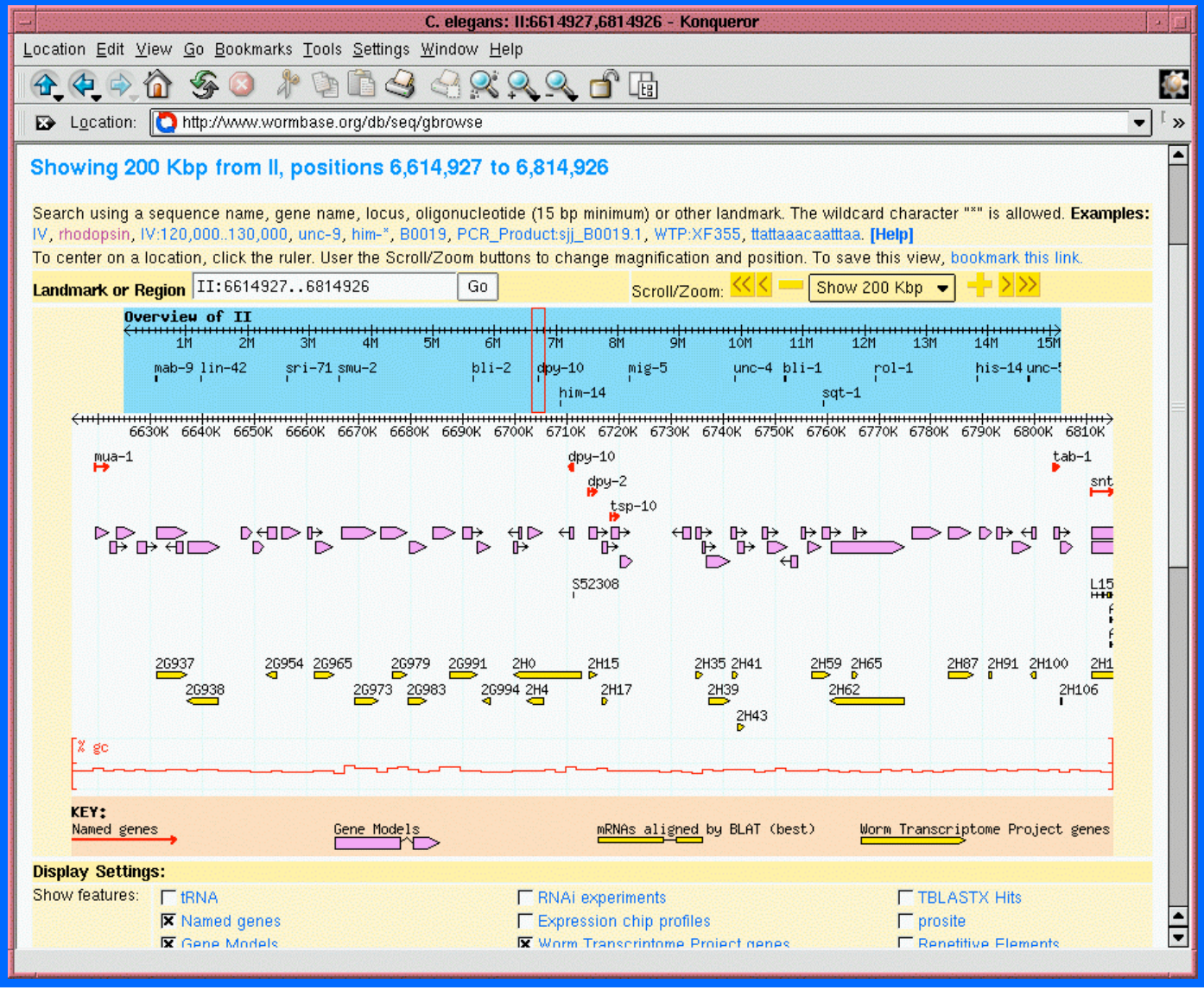
- Portable, open source software to support model organism databases
- Application Layer on Top of Bio* projects
- Five MODs involved
  - Worm, fly, yeast, mouse, arabidopsis, rat
- Funded by NIH as of June 2002.

http://www.gmod.org

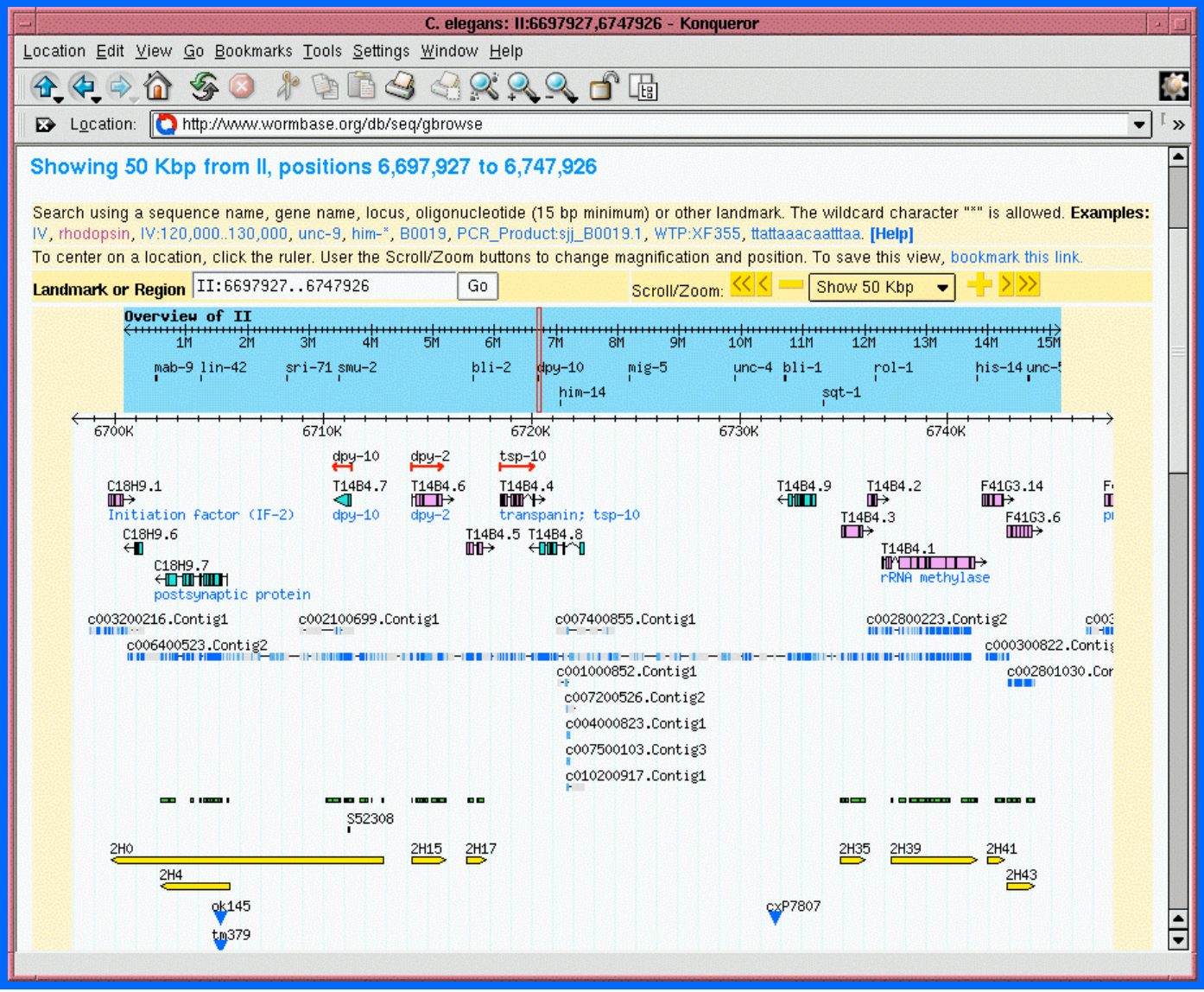
GMOD Modules

- Generic genome browser (gbrowse)
- Apollo genome annotation editor
- TAIR literature curation tools
- GOET
- GO tools
- LabDoc

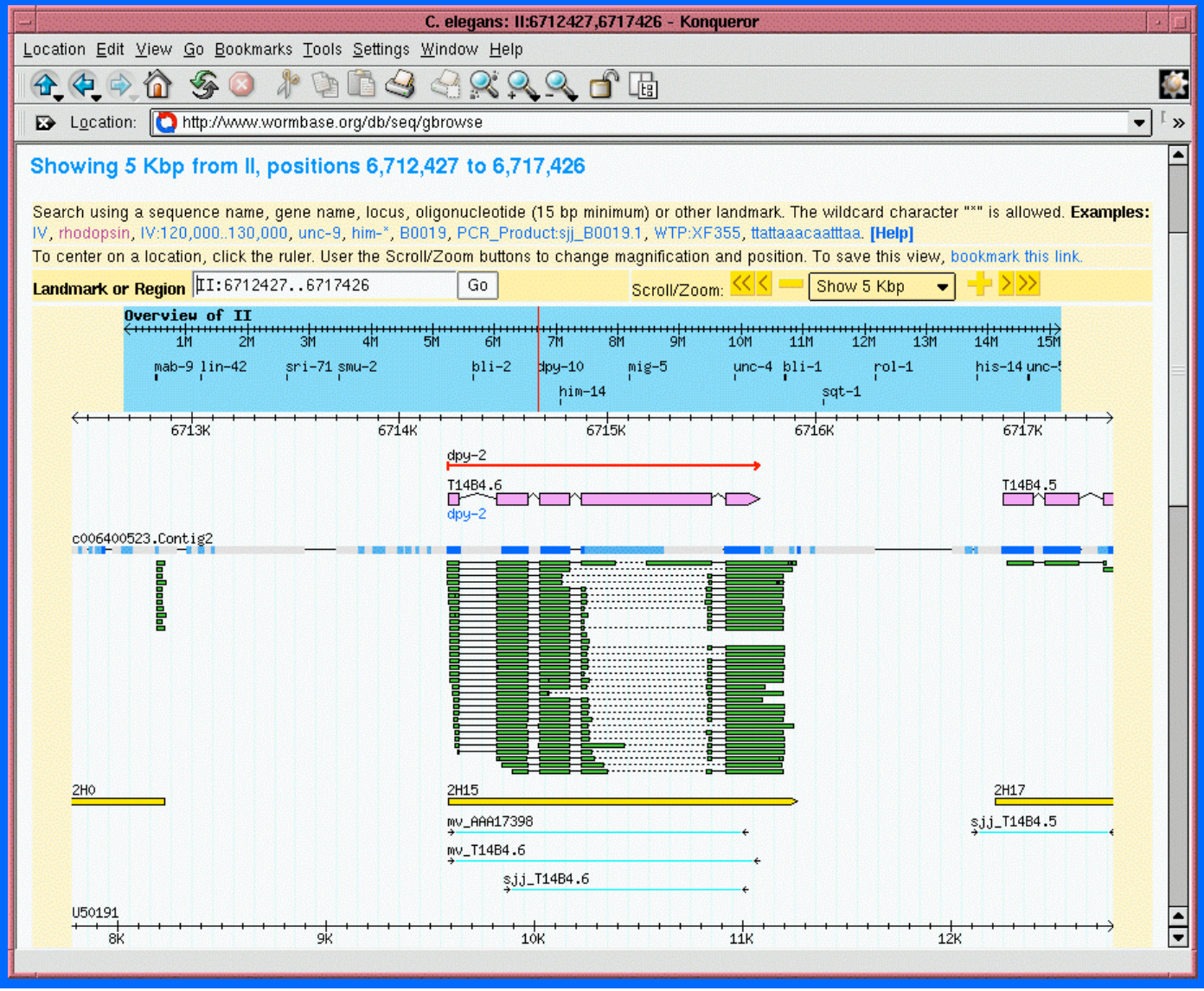
GBrowse: 200 kb of *C. elegans*



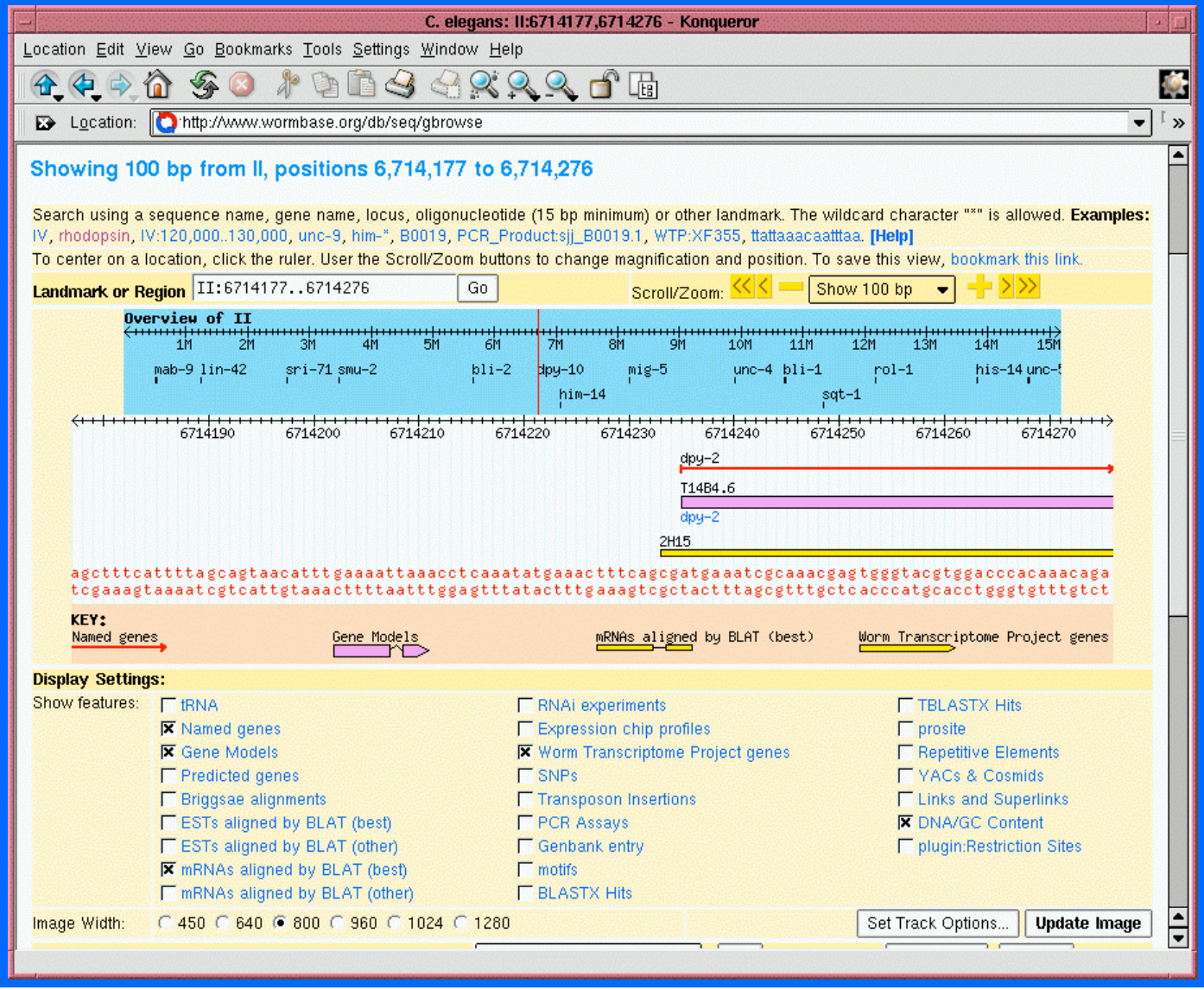
Zoomed In


C. elegans: II:6697927,6747926 - Konqueror
Location Edit View Go Bookmarks Tools Settings Window Help
Location: http://www.wormbase.org/db/seq/gbrowse
Showing 50 Kbp from II, positions 6,697,927 to 6,747,926
Search using a sequence name, gene name, locus, oligonucleotide (15 bp minimum) or other landmark. The wildcard character "*" is allowed. Examples:
IV, rhodopsin, IV:120,000..130,000, unc-9, him-*, B0019, PCR_Product:sjj_B0019.1, WTP:XF355, ttattaaacaatttaa. [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 II:6697927..6747926 Go
Scroll/Zoom: Show 50 Kbp
Overview of II
1M 2M 3M 4M 5M 6M 7M 8M 9M 10M 11M 12M 13M 14M 15M
mab-9 lin-42 sri-71 smu-2 bli-2 dpy-10 mig-5 unc-4 bli-1 rol-1 his-14
him-14 sqt-1
6700K 6710K 6720K 6730K 6740K
dpy-10 dpy-2 tsp-10
C18H9.1 T14B4.7 T14B4.6 T14B4.4 T14B4.9 T14B4.2 F41G3.14
Initiation factor (IF-2) dpy-10 dpy-2 transpanin; tsp-10
T14B4.3 F41G3.6
C18H9.6 T14B4.5 T14B4.8
T14B4.1
rRNA methylase
C18H9.7
postsynaptic protein
c003200216.Contig1 c002100699.Contig1 c007400855.Contig1 c002800223.Contig2
c006400523.Contig2 c000300822.Contig
c001000852.Contig1 c002801030.Cor
c007200526.Contig2
c004000823.Contig1
c007500103.Contig3
c010200917.Contig1
S52308
2H0 2H15 2H17 2H35 2H39 2H41
2H4 2H43
ok145 cxP7807
tm379

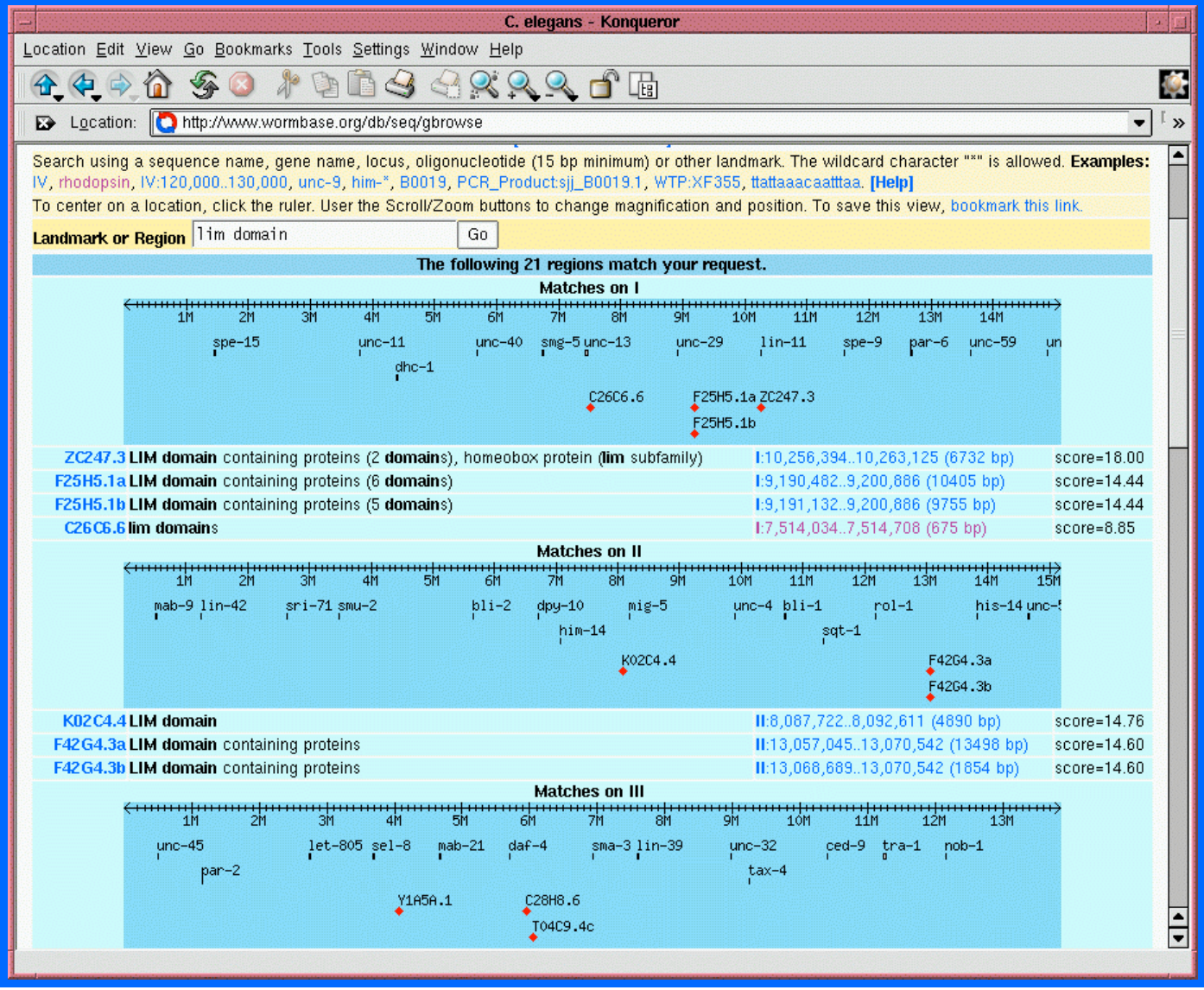
Zoomed Way In



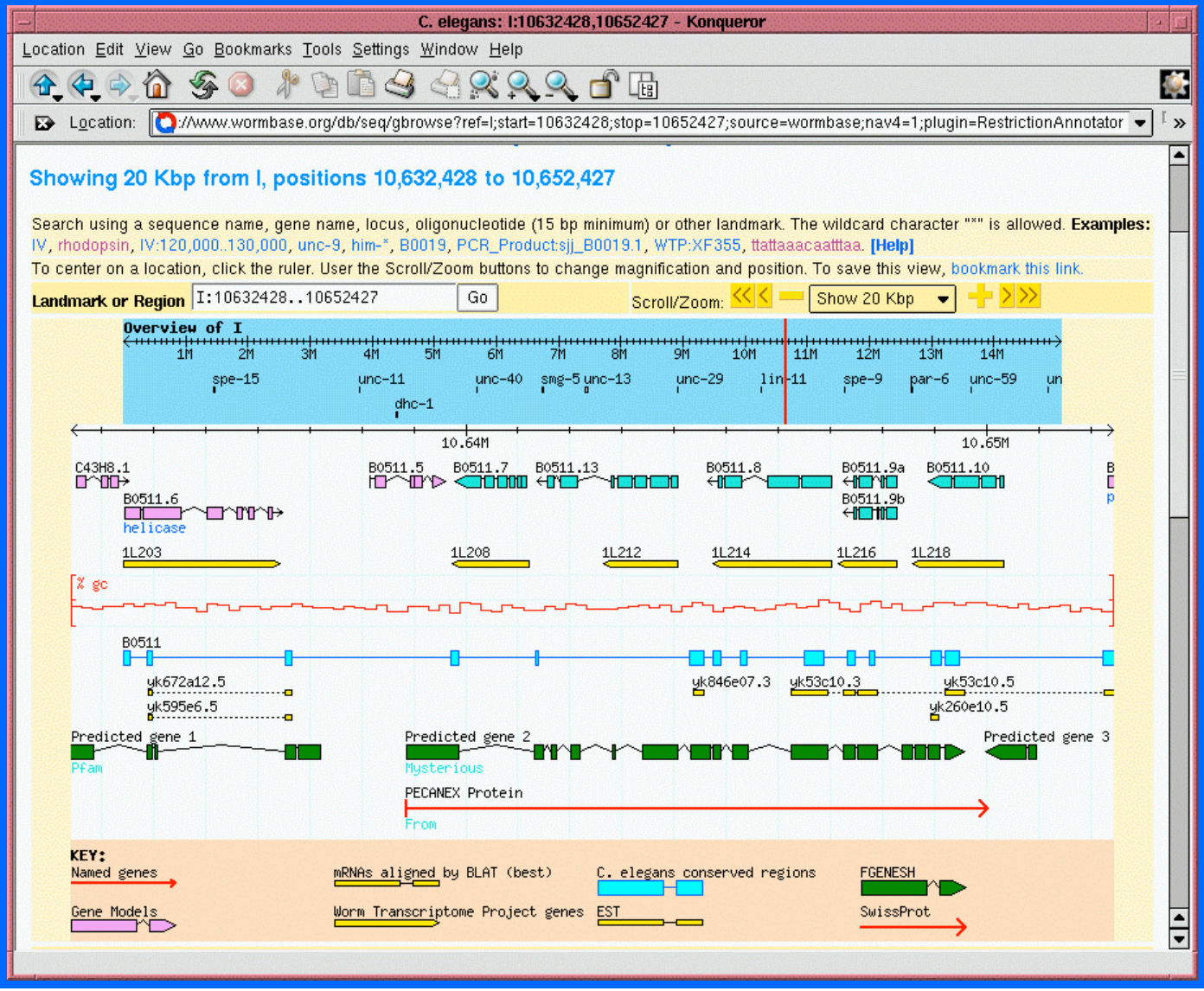
Zoomed Way Way In



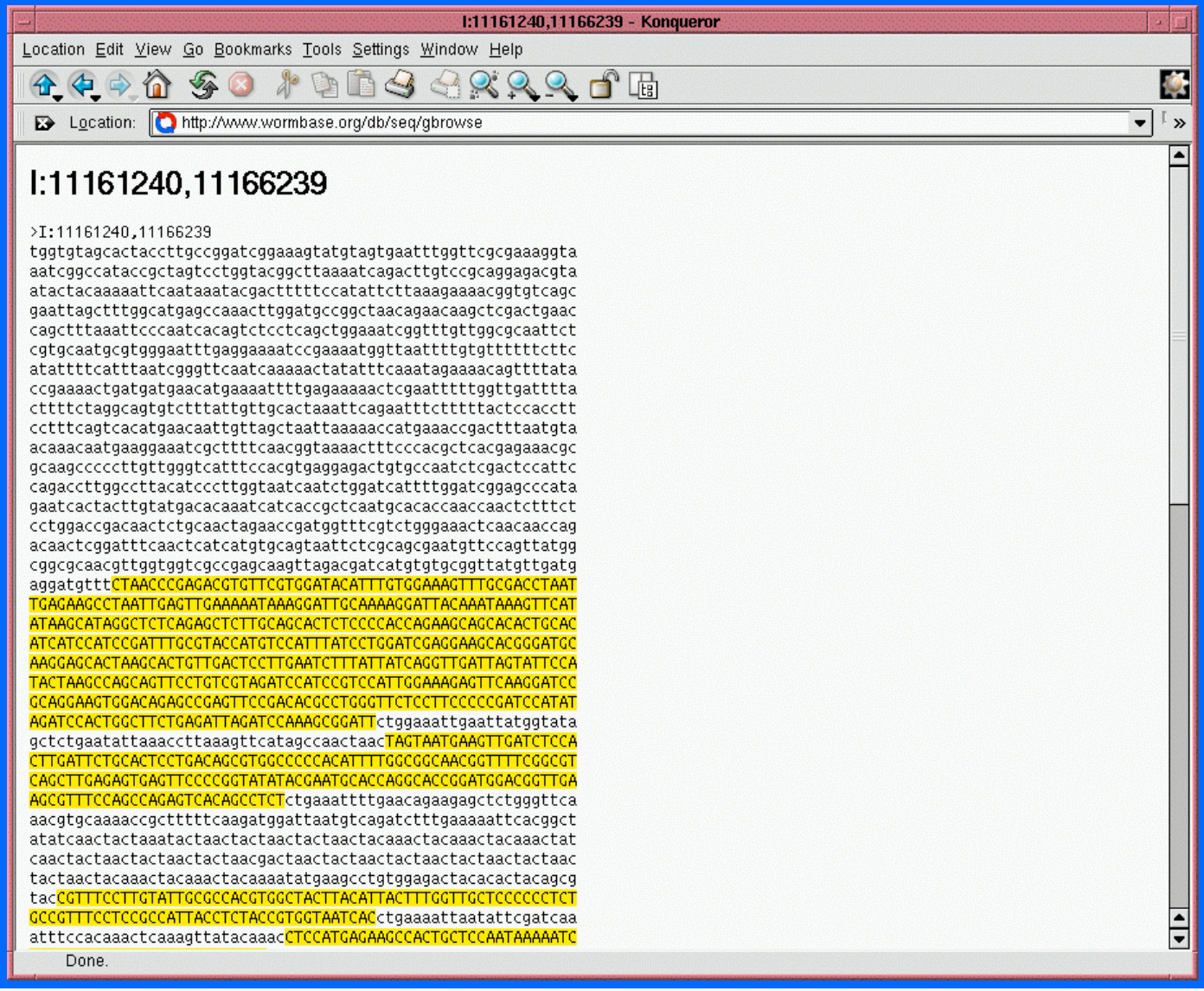
Keyword Search



Uploaded Annotations



Sequence dumps & other reports



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.

Adding a New Track

(a) Create a GFF file named “deletions.gff”

```
Chr1  targeted  deletion  1293224 1294901 . . . Deletion d101k2
Chr1  targeted  deletion  8239811 8241116 . . . Deletion d680k2
Chr2  targeted  deletion  5866382 5866500 . . . Deletion d007k2
```

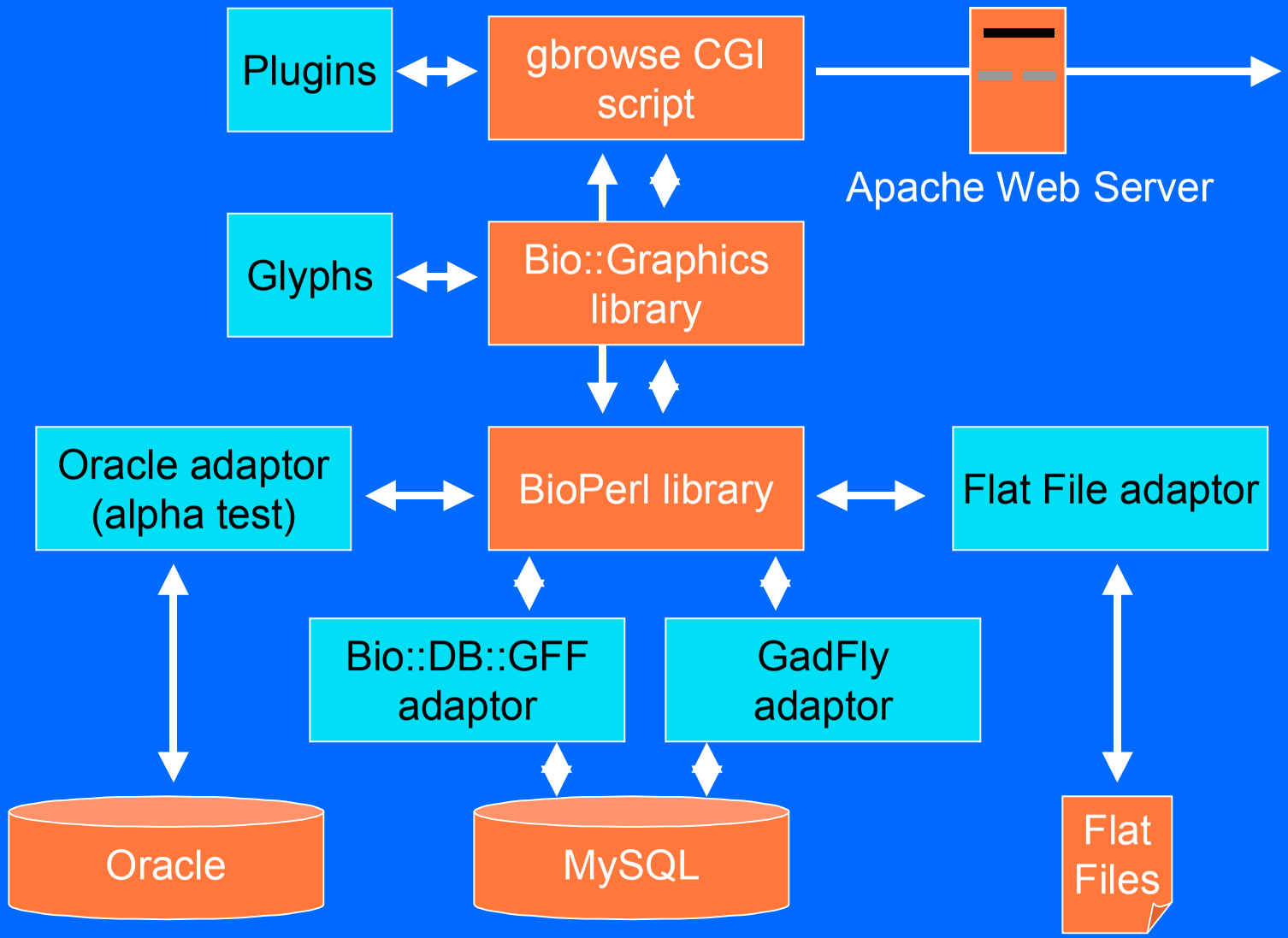
(b) Run the load_gff.pl script

```
> load_gff.pl -d example_database deletions.gff
Loading features…
Done.  3 features loaded.
```

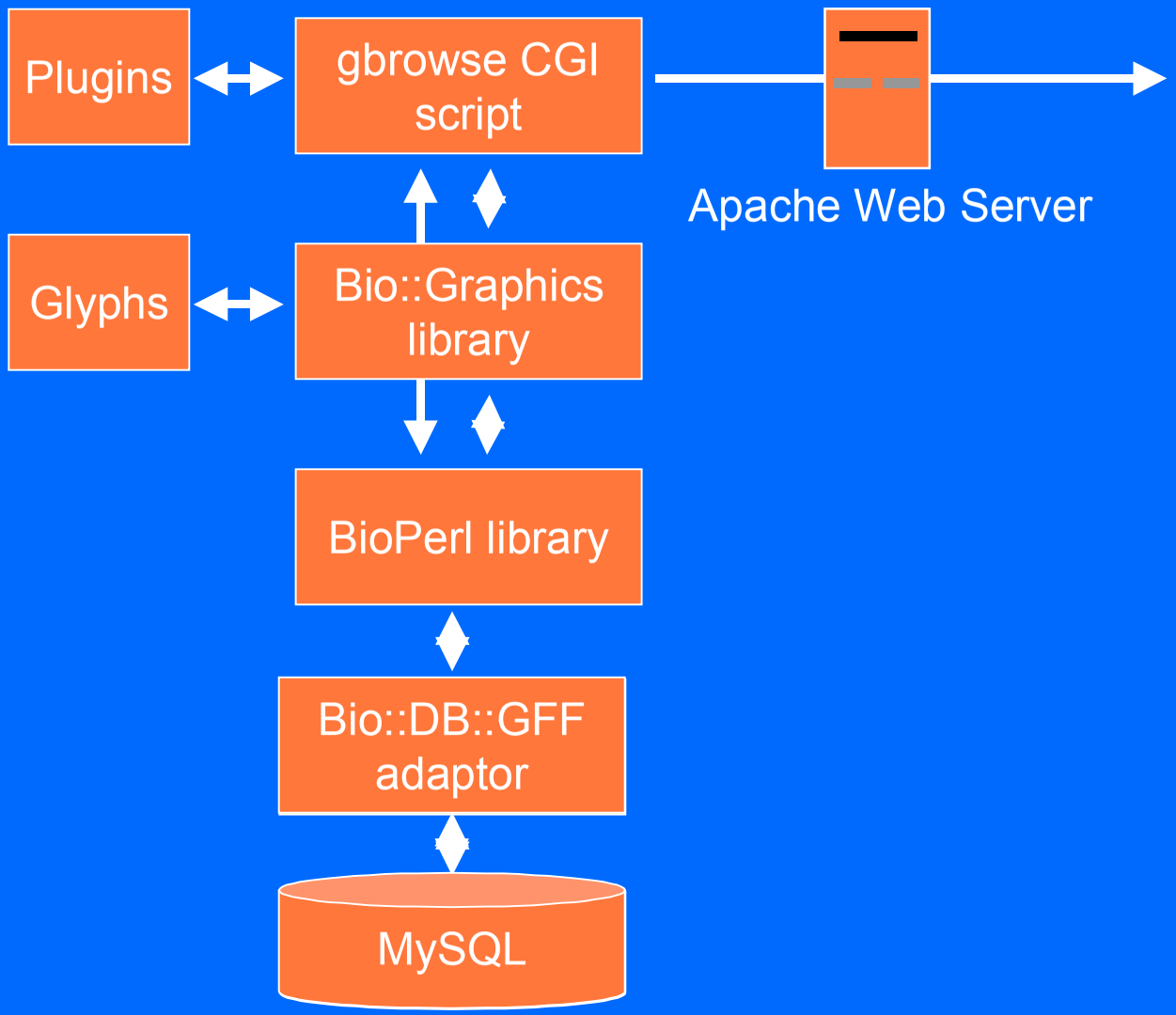
(c) Add a new track “stanza” to the gbrowse configuration file

```
[Knockout]
feature  = deletion
glyph    = span
fgcolor  = red
key      = Knockouts
link     = http://example.org/cgi-bin/knockout_details?$name
citation   = These are deletion knockouts produced by the
    example knockout consortium (http://example.org/knockouts.html)
```

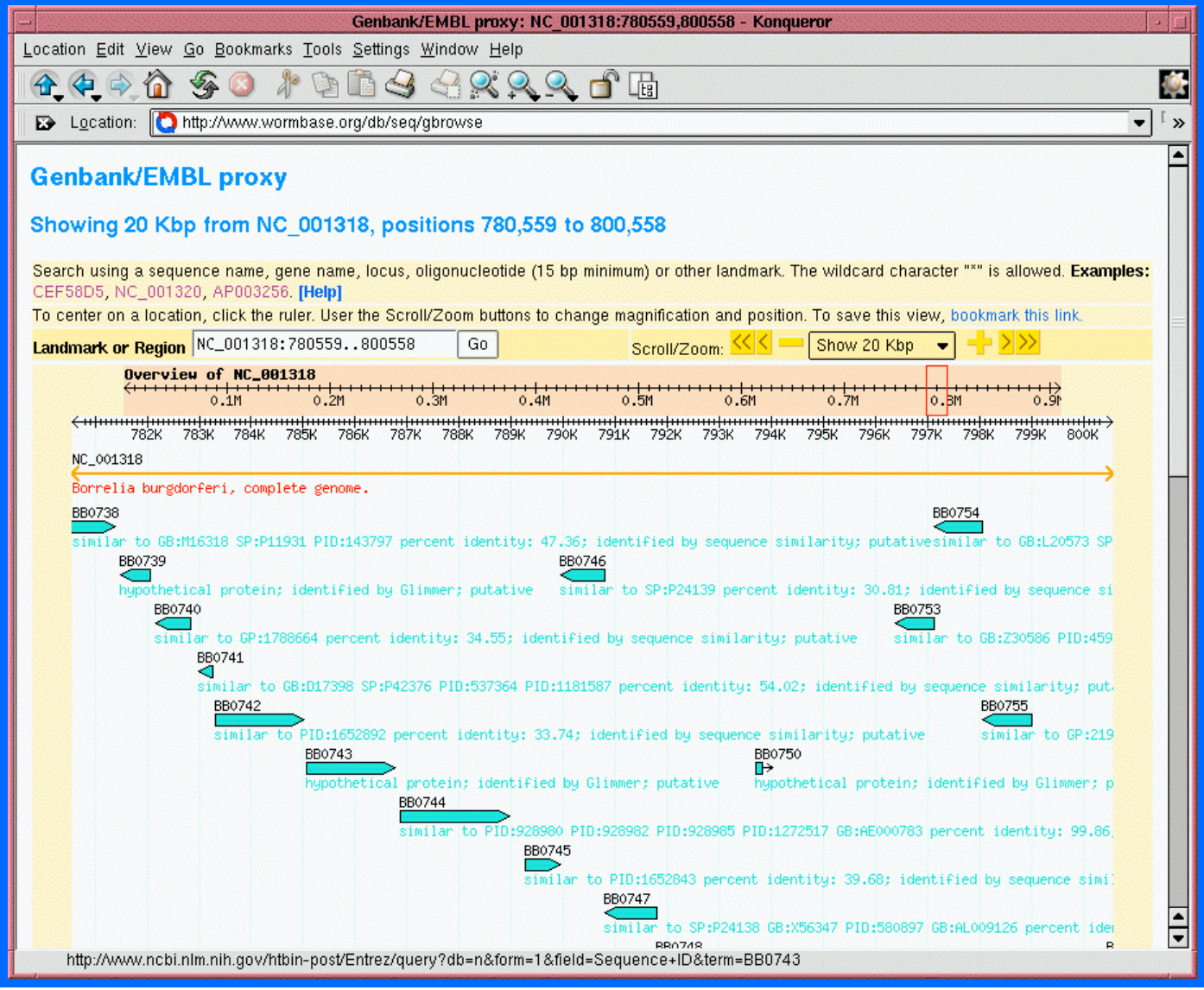
Extensively Extensible



GBrowse on GenBank?



B. burgdorferi via GenBank proxy



Coming Soon to www.gmod.org

- Comparative map viewer
- Insertional mutagenesis db & pipeline
- Schema repository
- Standard operating procedure repository
- Generic MOD web site design

Joining GMOD

- Go to www.gmod.org
- Examine software matrix
- Find a project or suggest new one
- Send mail to gmod-dev@lists.sourceforge.net

NB: We're Hiring!

Credits

CSHL

Adrian Arva
Allen Day
Todd Harris
Marco Mangone

BDGP

Suzanna Lewis
Chris Mungall
ShengQiang Shu

http://www.gmod.org

GMOD & DAS BOF
TONIGHT AFTER
LIGHTNING TALKS