



# Generic Model Organism Database Project

---

Lincoln Stein

Cold Spring Harbor Laboratory

# Model Organism Databases

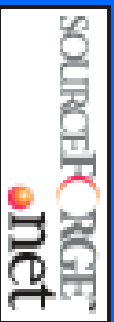
---

- ◆ 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





# Zoomed In

C. elegans: II:6697927.6747926 - Konqueror

Location Edit View Go Bookmarks Tools Settings Window Help

Location: <http://www.wormbase.org/dv/seq/browse>

### 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-5, B0019, PCR\_Product:slj\_B0019.1, WTP-XF355, thadaaacatthaa. [\[Help\]](#)

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

Landmark or Region [II:6697927..6747926](#)  Scroll/Zoom:   Show 50 Kbp

#### Overview of II

6700K 6710K 6720K 6730K 6740K

1M 2M 3M 4M 5M 6M 7M 8M 9M 10M 11M 12M 13M 14M 15M

mab-9 lin-42 sri-71 smu-2 p11-2 dpy-10 him-14 mig-5 unc-4 p11-1 pol-1 his-14 unc-1 p11-1 pqt-1

C18H9.1 T14B4.7 dpy-2 T14B4.6 T14B4.4 T14B4.5 T14B4.8 T14B4.9 T14B4.2 T14B4.3 T14B4.9 T14B4.10 T14B4.11 T14B4.12 T14B4.13 T14B4.14 T14B4.15 T14B4.16 T14B4.17 T14B4.18 T14B4.19 T14B4.20 T14B4.21 T14B4.22 T14B4.23 T14B4.24 T14B4.25 T14B4.26 T14B4.27 T14B4.28 T14B4.29 T14B4.30 T14B4.31 T14B4.32 T14B4.33 T14B4.34 T14B4.35 T14B4.36 T14B4.37 T14B4.38 T14B4.39 T14B4.40 T14B4.41 T14B4.42 T14B4.43 T14B4.44 T14B4.45 T14B4.46 T14B4.47 T14B4.48 T14B4.49 T14B4.50 T14B4.51 T14B4.52 T14B4.53 T14B4.54 T14B4.55 T14B4.56 T14B4.57 T14B4.58 T14B4.59 T14B4.60 T14B4.61 T14B4.62 T14B4.63 T14B4.64 T14B4.65 T14B4.66 T14B4.67 T14B4.68 T14B4.69 T14B4.70 T14B4.71 T14B4.72 T14B4.73 T14B4.74 T14B4.75 T14B4.76 T14B4.77 T14B4.78 T14B4.79 T14B4.80 T14B4.81 T14B4.82 T14B4.83 T14B4.84 T14B4.85 T14B4.86 T14B4.87 T14B4.88 T14B4.89 T14B4.90 T14B4.91 T14B4.92 T14B4.93 T14B4.94 T14B4.95 T14B4.96 T14B4.97 T14B4.98 T14B4.99 T14B4.100

Initiation factor (IF-2) dpy-10 dpy-2 tsp-10

C18H9.6 C18H9.7 postsynaptic protein

c002100699.Contig1 c007400855.Contig1 c002800223.Contig2 c000300822.Contig1 c002801030.Contig1 c007200526.Contig2 c004000823.Contig1 c007500103.Contig3 c010200917.Contig1

S52308

2H0 2H4 ok-145 h9379 2H15 2H17 2H35 2H39 2H41 2H43 exp7807

FRNH methylase

# Zoomed Way In

C. elegans: II:5712427..5717426 - Konqueror

Location Edit View Go Bookmarks Tools Settings Window Help

Location: <http://www.wormbase.org/db/seq/gbrowse>

### Showing 5 Kbp from II, positions 6,712,427 to 6,717,426

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-5, B0019, PCR\_Products|J\_B0019.1, WTP:XF355, tathaacaatthaa. [\[Help\]](#)

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

Landmark or Region   Scroll/Zoom:     Show 5 Kbp

#### Overview of II

The overview shows a genomic map of chromosome II with landmarks labeled 1M through 15M. A blue box highlights the region from 6713K to 6717K. Below the overview, a zoomed-in view shows a 5 Kbp region from 6713K to 6717K. The zoomed view includes a ruler at the top, a sequence alignment with green bars representing reads, and several landmarks: qpy-2 (red arrow), T14B4.6 (pink box), qpy-2 (pink box), T14B4.5 (pink box), 2H0 (yellow bar), 2H15 (yellow bar), mv\_099417398 (cyan bar), mv\_T14B4.6 (cyan bar), \$JJ\_T14B4.6 (cyan bar), 2H17 (yellow bar), and \$JJ\_T14B4.5 (cyan bar). A blue bar at the bottom indicates the sequence coordinates from 6713K to 6717K.



# Zoomed Way Way In

C. elegans: II:6714177..6714276 - Konqueror

Location Edit View Go Bookmarks Tools Settings Window Help

Location: <http://www.wormbase.org/dv/seq/browse>

### Showing 100 bp from II, positions 6,714,177 to 6,714,276

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-x, B0019, PCR\_Products|\_B0019.1, WTP-XF355, tattaacaathaa. [\[Help\]](#)

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:6714177..6714276  Scroll/Zoom:   Show 100 bp

**Overview of II**

The genomic map shows landmarks from 1M to 15M. A blue box highlights the region from 6714190 to 6714270. Below the map, a zoomed-in sequence is shown with three highlighted features: a red arrow pointing to a sequence, a purple bar labeled 'dpy-2', and a blue bar labeled '2H15'. The zoomed-in sequence is: agctttccatttttagcagtaaacatttggaaaatttaaacctcaaatatgaaaccttcaggcatgaatacgaaacaggatgggtactgtggaccaccaaacaga tcgaaagtaaaaatcgtaacttgaataactttaactttggatgattatacttgaagaagtcgctactttagcgttgctcaccatgcacctgggtgtttgtct

**KEY:**  
Named genes →

**Display Settings:**

Show Features:

- tRNA
- Named genes
- Gene Models
- Predicted genes
- Briggsae alignments
- ESTs aligned by BLAT (best)
- ESTs aligned by BLAT (other)
- mRNAs aligned by BLAT (best)
- mRNAs aligned by BLAT (other)
- RNAI experiments
- Expression chip profiles
- Worm Transcriptome Project genes
- SNPs
- Transposon Insertions
- PCR Assays
- Genbank entry
- motifs
- BLASTX Hits
- TBLASTX Hits
- prosite
- Repetitive Elements
- YACs & Cosmids
- Links and Superlinks
- DNA/GC Content
- plugin:Restriction Sites

Gene Models  mRNAs aligned by BLAT (best)  Worm Transcriptome Project genes

Image Width:  450  640  800  960  1024  1280



# Keyword Search

Location: <http://www.wormbase.org/dv/seq/gbrowse>

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\[s\]\\_B0019.1](#), [WTP:XFC35](#), [tadaaacacatthaa](#). [\[Help\]](#)

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:

The following 21 regions match your request.

Region	Score
<b>ZC247.3 LIM domain</b> containing proteins (2 domains), homeobox protein (lim subfamily)	110,256,394..10,263,125 (6732 bp) score=18.00
<b>F29H5.1a LIM domain</b> containing proteins (6 domains)	19,190,482..9,200,886 (10405 bp) score=14.44
<b>F29H5.1b LIM domain</b> containing proteins (5 domains)	19,191,132..9,200,886 (9755 bp) score=14.44
<b>C26C6.6 lim domains</b>	17,514,034..7,514,708 (675 bp) score=8.85

**Matches on I**

**Matches on II**

**Matches on III**

Region	Score
<b>K02C4.4 LIM domain</b>	118,087,722..8,092,611 (4890 bp) score=14.76
<b>F42G4.3a LIM domain</b> containing proteins	1113,057,045..13,070,542 (13498 bp) score=14.60
<b>F42G4.3b LIM domain</b> containing proteins	1113,068,689..13,070,542 (1854 bp) score=14.60

# Uploaded Annotations

C. elegans: 1:10632428..10652427 - Konqueror

Location Edit View Go Bookmarks Tools Settings Window Help

Location: <http://www.wormbase.org/db/seq/gbrowse?ref=1;start=10632428;stop=10652427;source=wormbase;nav4=1;plugin=RestrictionAnnotator>

### Showing 20 Kbp from I, positions 10,632,428 to 10,652,427

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:sl.B0019.1, WTP-XF355, tattaacaattaa. [\[Help\]](#)

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

Landmark or Region I:10632428..10652427  Scroll/Zoom:   Show 20 Kbp

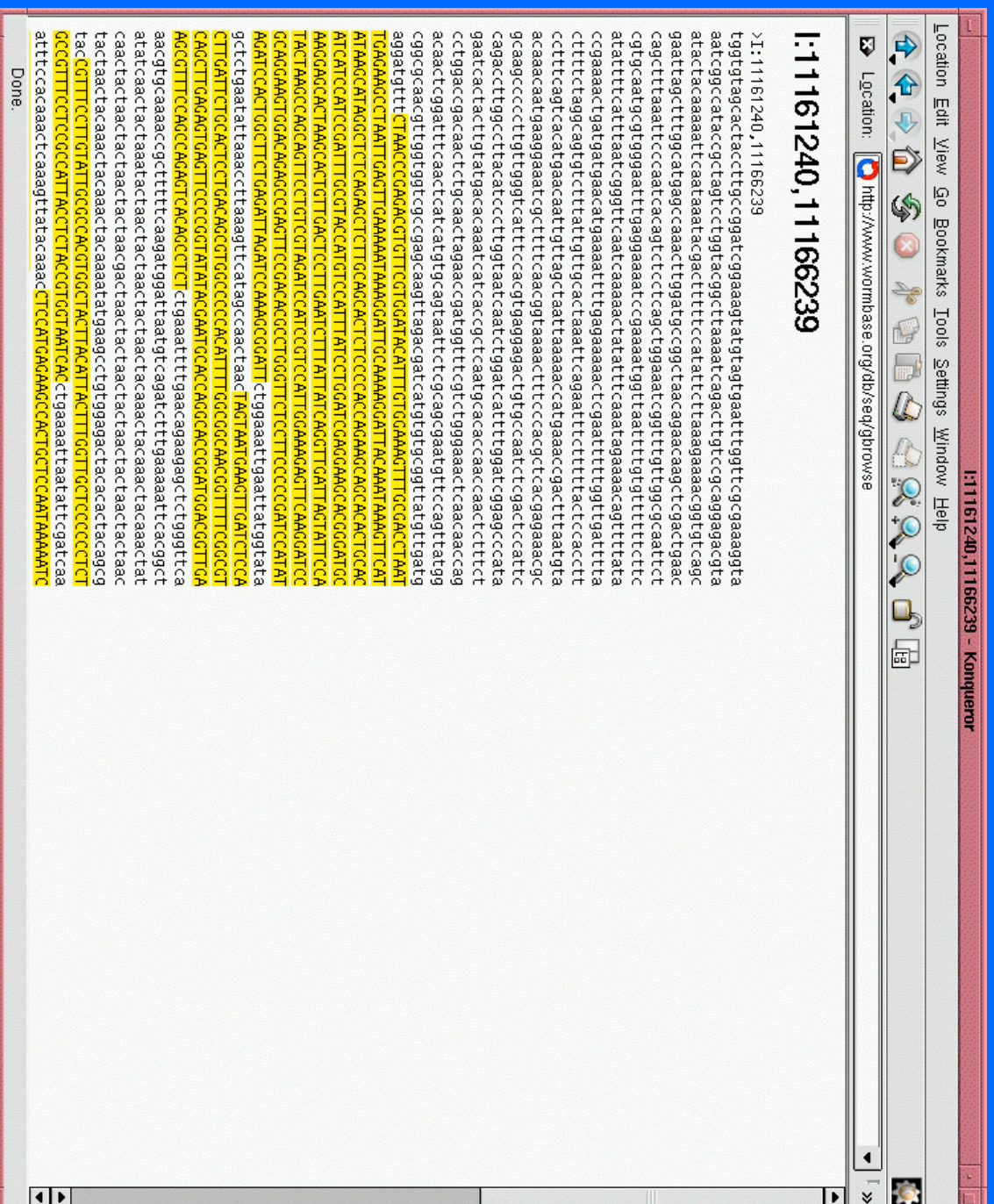
**Overview of I**

**KEY:**

- Named genes
- Gene Models
- mRNAs aligned by BLAT (best)
- Worm Transcriptome Project genes
- EST
- C. elegans conserved regions
- FGENESH
- SwissProt



# 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
```

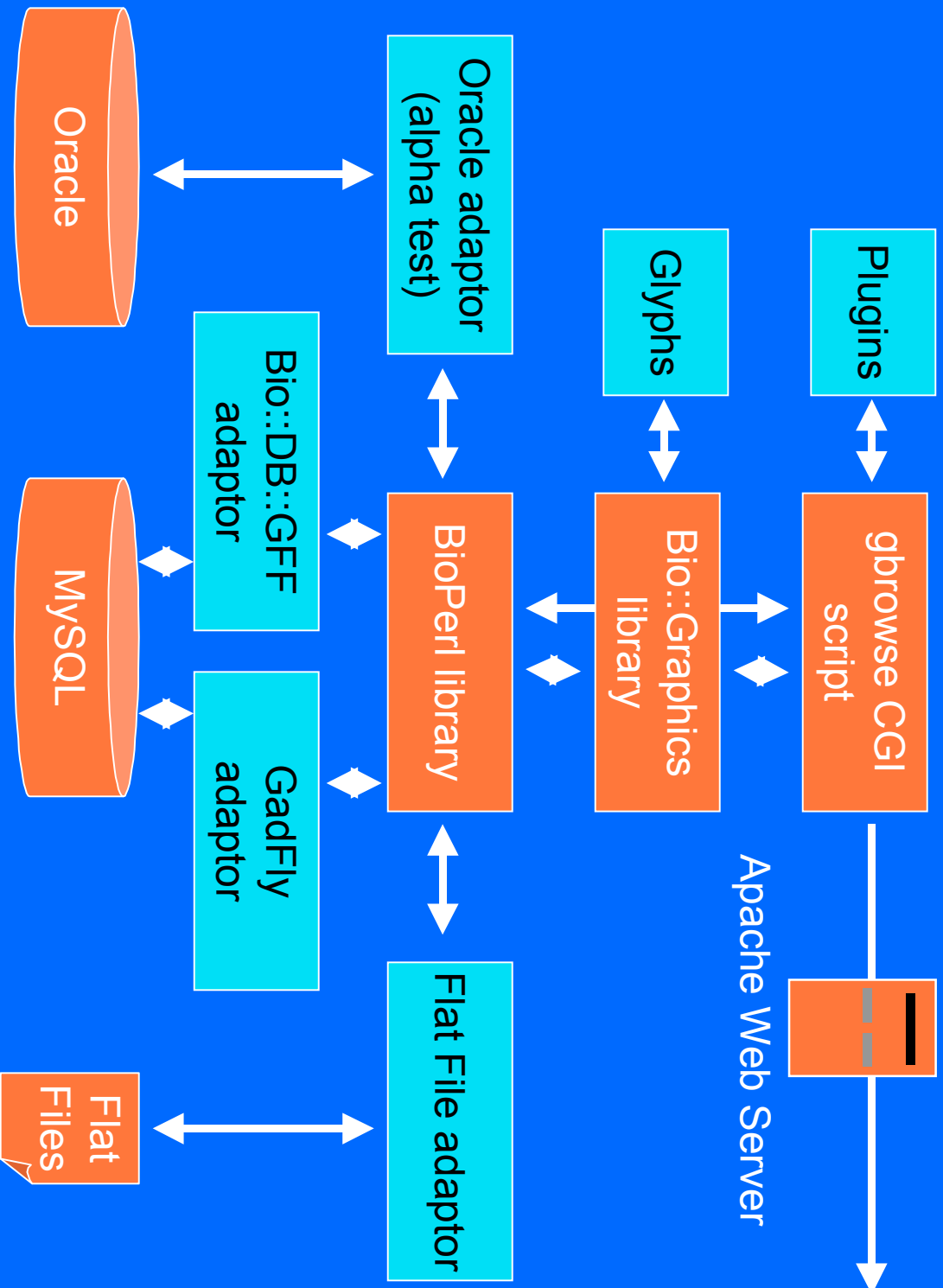
(b) Run the `load_gff.pl` script

```
> 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

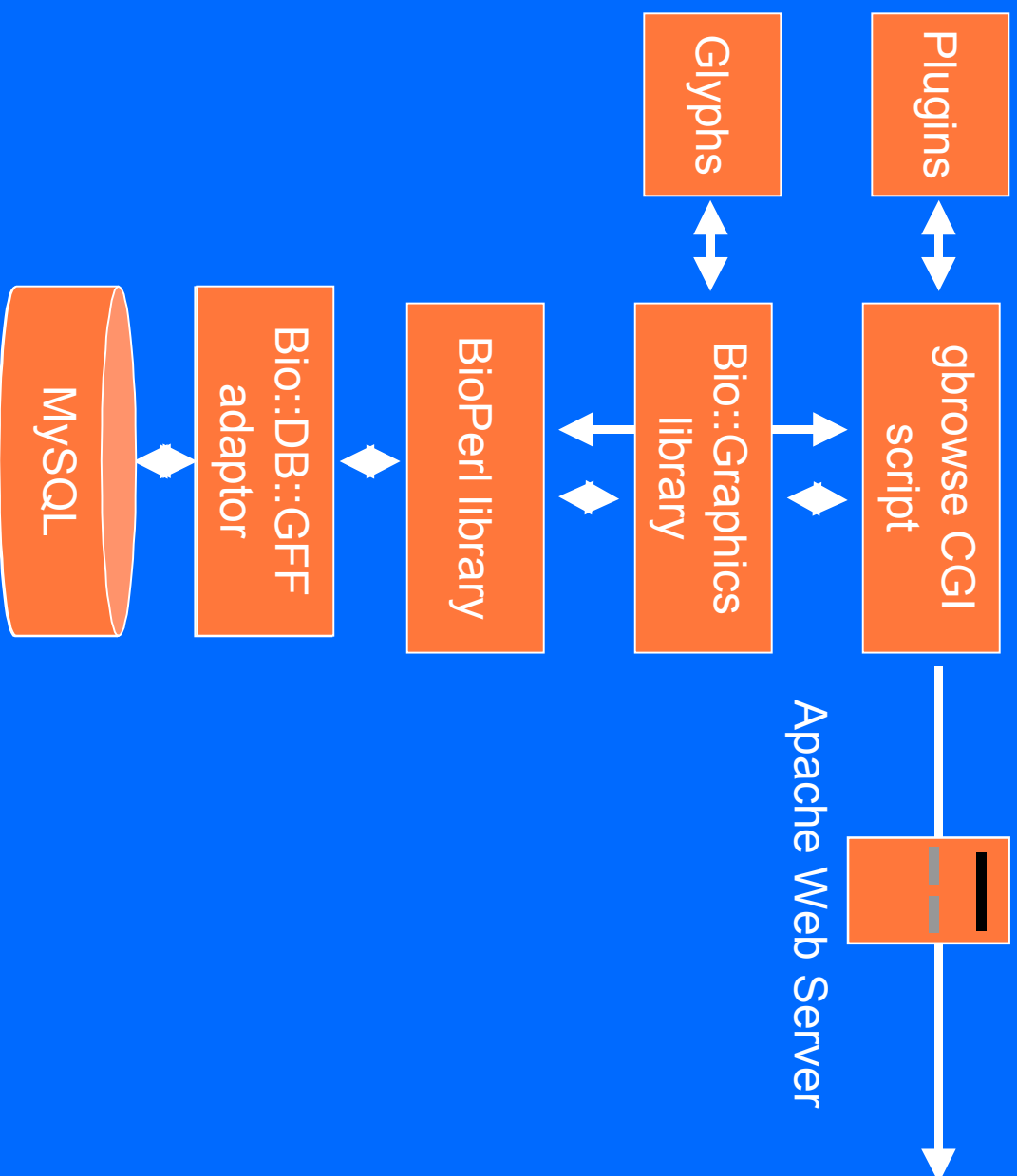
```
[Knockout]
feature = deletion
glyph   = span
fgcolor = red
key     = Knockouts
link    = http://example.org/cgi-bin/knockout_details?name
citation = These are deletion knockouts produced by the
         example knockout consortium (http://example.org/knockouts.html)
```

# Extensively Extensible





# GBrowse on GenBank?





# Coming Soon to [www.gmod.org](http://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](http://www.gmod.org)
- ◆ Examine software matrix
- ◆ Find a project or suggest new one
- ◆ Send mail to [gmod-dev@lists.sourceforge.net](mailto:gmod-dev@lists.sourceforge.net)

*YAB: We're Hiring!*

# Credits

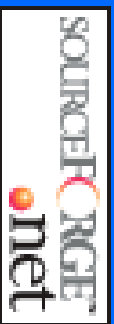
---

## 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