



Comparative Genomics with GBrowse_syn

Sheldon McKay



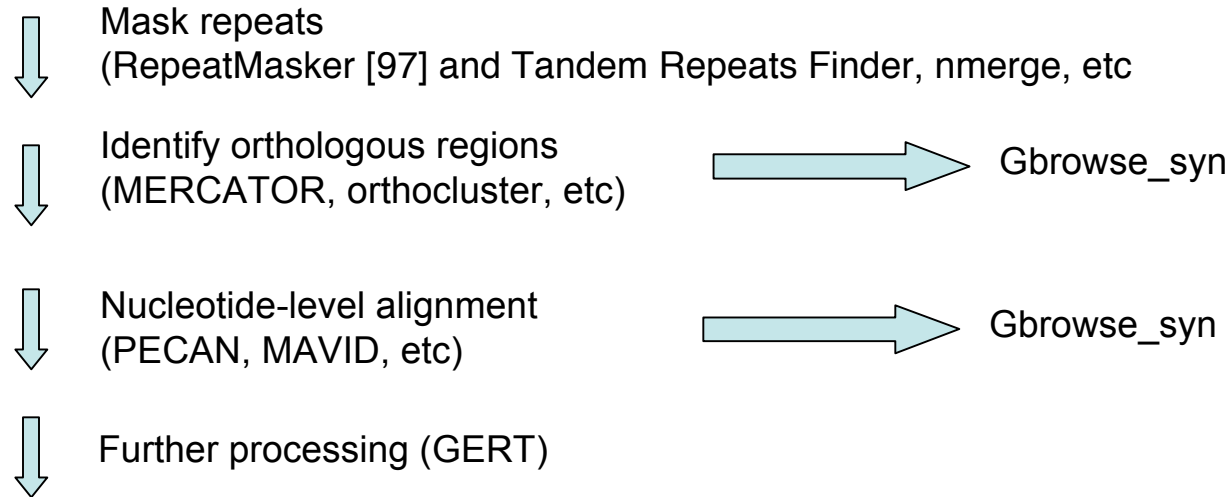


- How to get comparative genomics data
- How to view the data with GBrowse_syn
- Example uses



Hierarchical Genome Alignment Strategy

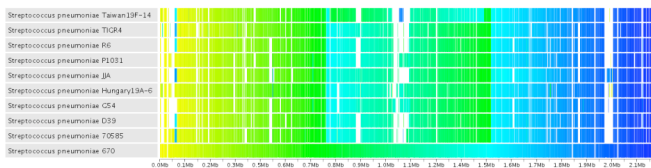
Raw genomic sequences



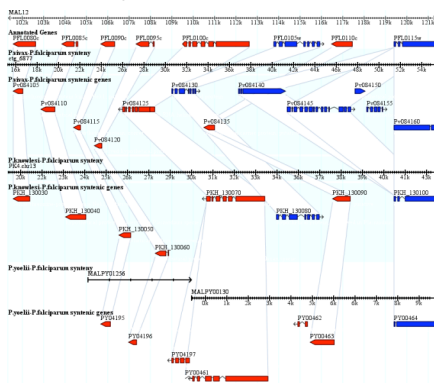
GBrowse



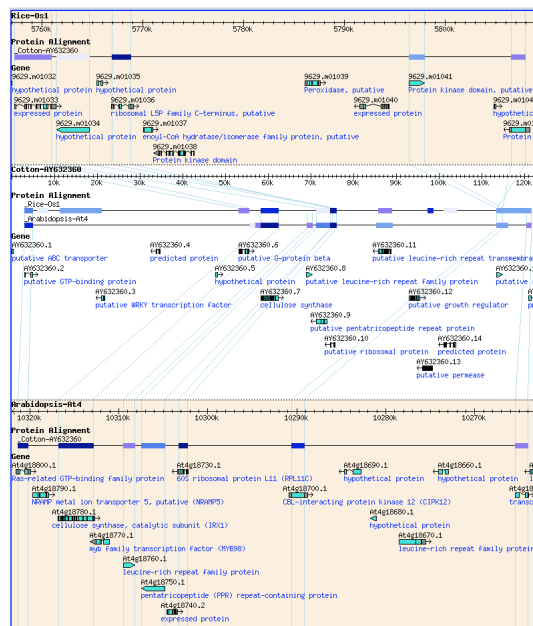
Sybil



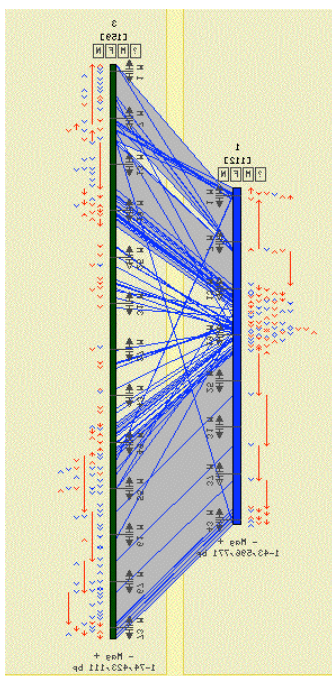
SynView



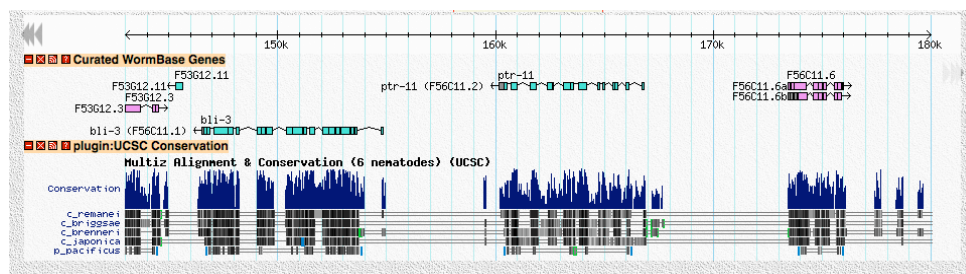
SynBrowse



CMap



GBrowse UCSC-style conservation tracks





What is GBrowse_syn?

- Part of the Generic Genome Browser Package (GBrowse)
- A graphical multiple sequence alignment viewer
- Superimposes sequence alignment data on genes and other sequence features
- Compares two or more species to a central reference species



```
Ce-CHROMOSOME_I(+)/5195-16585 TTCTTCAGATATTTTATAGAATTTACTGACTTTTCAGAAATAGATGTAGGACAAATTTTG
Cb-chr I(-)/4091935-4097143
Cr-Cont Iq8(+)/571998-577344 ATGGTTTTGGTTTTTGAGCTGTATTTTCGGGGTTTTTAAAGCGGAAAAACGAAATGTTT

Ce-CHROMOSOME_I(+)/5195-16585 TTGTTTTAAAAATTGAAATCTGAAATTTCCAAACAAAAAATAGTCCAAACCCACAAGT
Cb-chr I(-)/4091935-4097143
Cr-Cont Iq8(+)/571998-577344 TGCATTTTCTGACTTCTATATCTGAAATTAGGCACGGGACATTGGAACCTCCGACAT

Ce-CHROMOSOME_I(+)/5195-16585 TGGCAAAAAATTTTGCATTTGCCGTTTTCCGGTTTGGCCAAAAGCTAATTTCCGGTAA
Cb-chr I(-)/4091935-4097143
Cr-Cont Iq8(+)/571998-577344 TTGCAAAAC-----

Ce-CHROMOSOME_I(+)/5195-16585 TTGGGCCATTTTTCGAAATTTTGAGCCACATAAAAACTTTGAACCATTTTGTGAGAAGTA
Cb-chr I(-)/4091935-4097143
Cr-Cont Iq8(+)/571998-577344 -----AGAGAAATCTGAGATCTTGA-----
-----CAGAGAAACAGAAACAATTTTA-----
** * ** * **

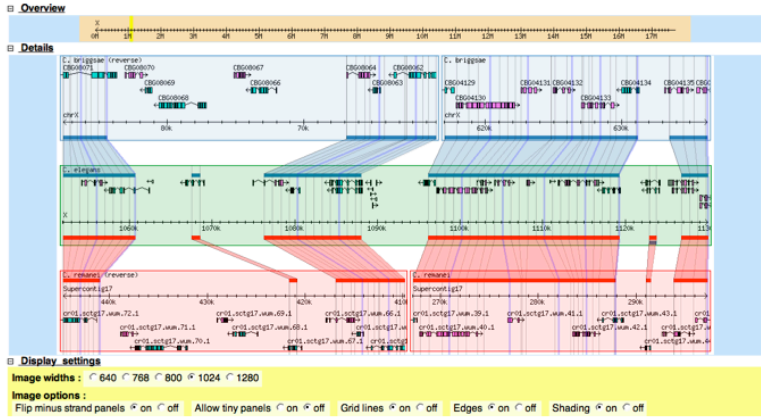
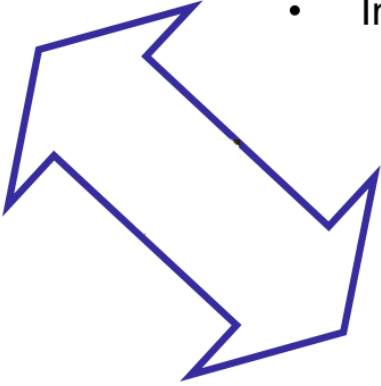
Ce-CHROMOSOME_I(+)/5195-16585 TTAATAGCAGATTCGTTTATTTGAGCCACAATTTGGGGCTACTATTTCAAAATCGGGGTTT
Cb-chr I(-)/4091935-4097143
Cr-Cont Iq8(+)/571998-577344 -----TTTCATGTGAA-----TCAT
-----TTTCTGAAAAACGGTAGTATTATGGTTCCGAGGGTCTAGGGTTCCAAAACCGGGCTAG
* * *

Ce-CHROMOSOME_I(+)/5195-16585 GAAAACCCCTATATGTTCCAGCCGAATGTTAATCTCATAAAAAATTTGATGAAAAATAAAT
Cb-chr I(-)/4091935-4097143
Cr-Cont Iq8(+)/571998-577344 CTAGTCTCCCAATAATGTCATCTCATAT-----ATT
CCAAACTTTTTTGCATCTCTCTTGGCT-----TTT
* * * * * **

Ce-CHROMOSOME_I(+)/5195-16585 TTCTACGGCTATAAAGGTATAGCCCGGTGAGTCTCAAAATTTATAGATAGACACTTT
Cb-chr I(-)/4091935-4097143
Cr-Cont Iq8(+)/571998-577344 TCGATCCAAC-----TTGTTACAAATCTAACAGCCAGGGTTCCGATCCCGACTGGTGGCCAAACTCTTTTATTTT
* * * *
```

Goals

- More than two species
- Nucleotide-level resolution (gapped alignments)
- High-level resolution (synteny)
- Intuitive graphical rendering





GBrowse-like interface

PECAN alignments for *Caenorhabditis* (WS197)

Instructions

Select a Region to Browse and a Reference species:

Examples: *c_elegans* X:1050001..1150000, *c_briggsae* chrX:620000..670000, *c_elegans* R193.2.

Search

Landmark:

X:1050001..1150000

Reference Species:

C. elegans ▾

<< < > >>

Aligned Species:

C. briggsae *C. remanei* *C. brenneri* *C. japonica*

Data Source :

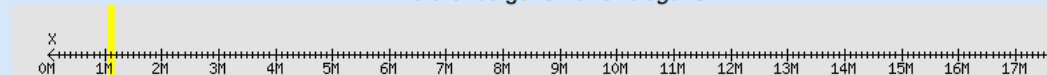
PECAN alignments for *Caenorhabditis* ▾

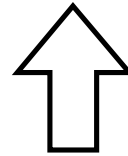
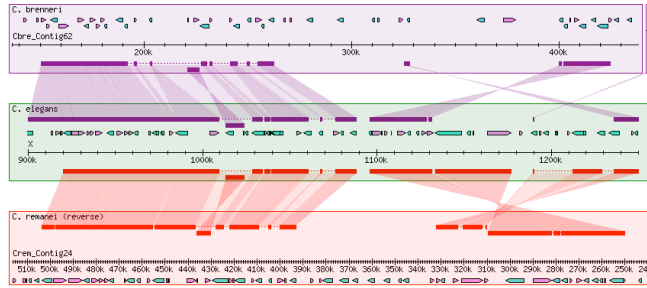
Display Mode :

Three species/panel [Click to show all species in one panel](#)

Overview

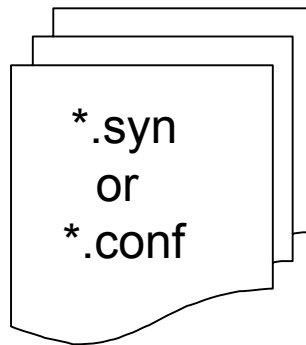
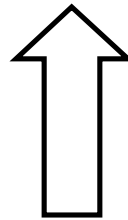
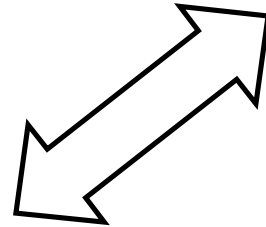
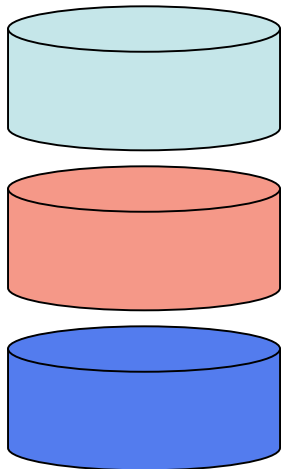
Reference genome: *C. elegans*



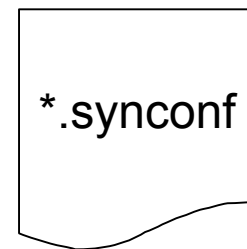
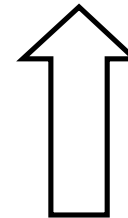


GBrowse_syn

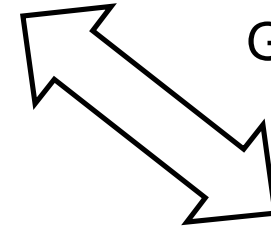
GBrowse
Databases*



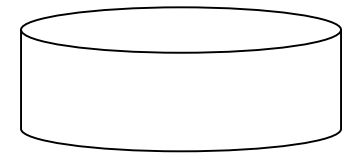
Species config.



Master config.



GBrowse_syn
alignment
database

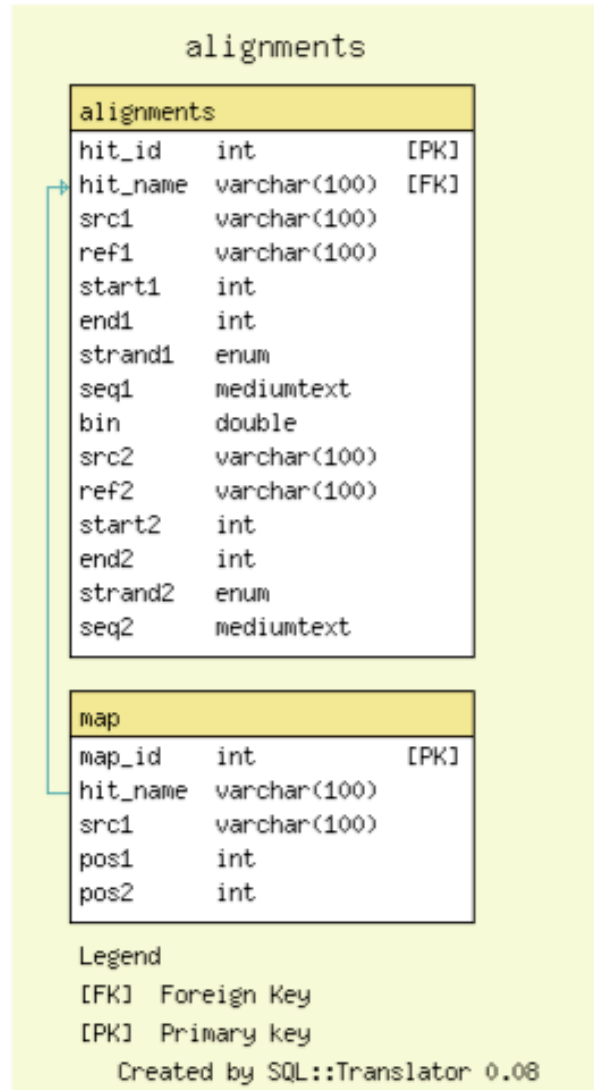




GBrowse_syn Architecture

[GBrowse]

Bio::DB::GFF
species1



[GBrowse]

Bio::DB::GFF
species3



Bio::DB::GFF
species2



Bio::DB::GFF
species4

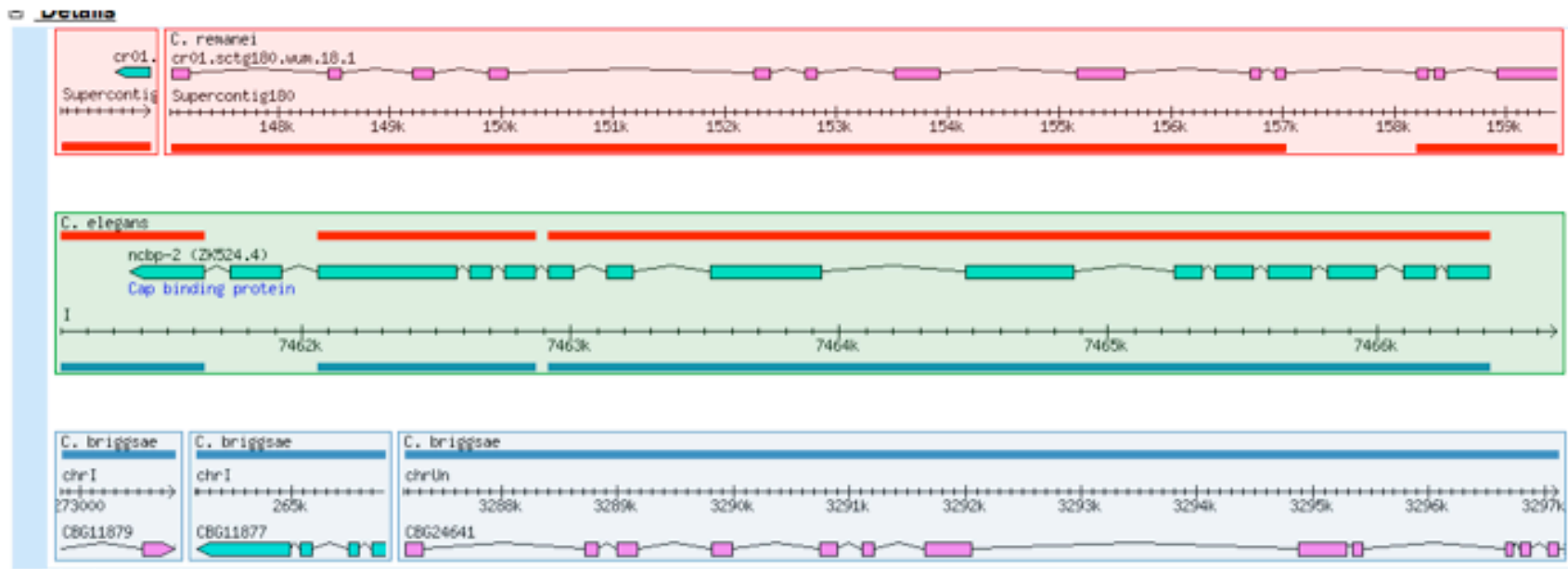


[GBrowse]

[GBrowse]



How to get the most information about the alignments?



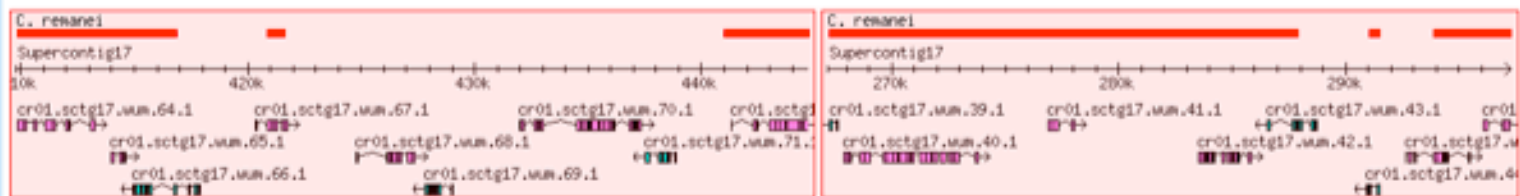
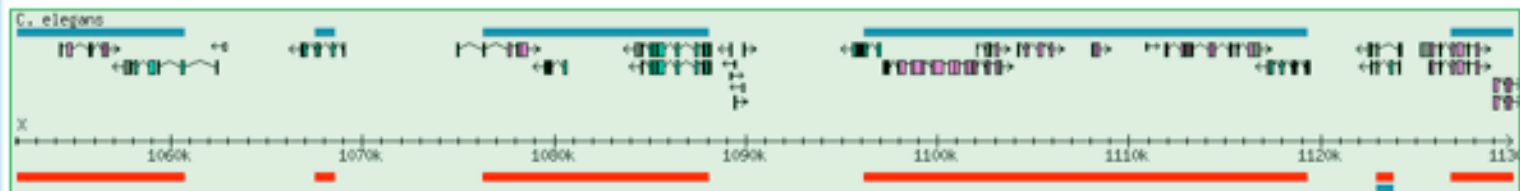


Gbrowse_syn: quick tour

Overview



Details



Display settings

Image widths : 640 768 800 1024 1280

Image options :

Flip minus strand panels on off

Allow tiny panels on off

Grid lines on off

Edges on off

Shading on off



Gbrowse_syn: quick tour (shaded alignments)

Overview

Details

C. briggsae
chrX
70k 80k

C. elegans
X
100k 107k 108k 109k 110k 111k 112k 113k

C. reanei
Supercontig17
10k 42k 43k 44k

C. reanei
Supercontig17
27k 28k 29k

Display settings

Image widths : 640 768 800 1024 1280

Image options :

Flip minus strand panels on off Allow tiny panels on off Grid lines on off Edges on off Shading on off

The image displays a genomic browser interface with four tracks. The top track is an overview of chromosome X from 0k to 170k. Below it are three detailed tracks: C. briggsae (chrX, 70k-80k), C. elegans (X, 100k-113k), and C. reanei (Supercontig17, 10k-44k and 27k-29k). Shaded regions connect homologous segments across the species. The C. reanei track shows two separate supercontigs. The bottom section contains display settings for image width and various image options.



Gbrowse_syn: quick tour (outlines)

Overview



Details



Display settings

Image widths : 640 768 800 1024 1280

Image options :

Flip minus strand panels on off Allow tiny panels on off Grid lines on off Edges on off Shading on off

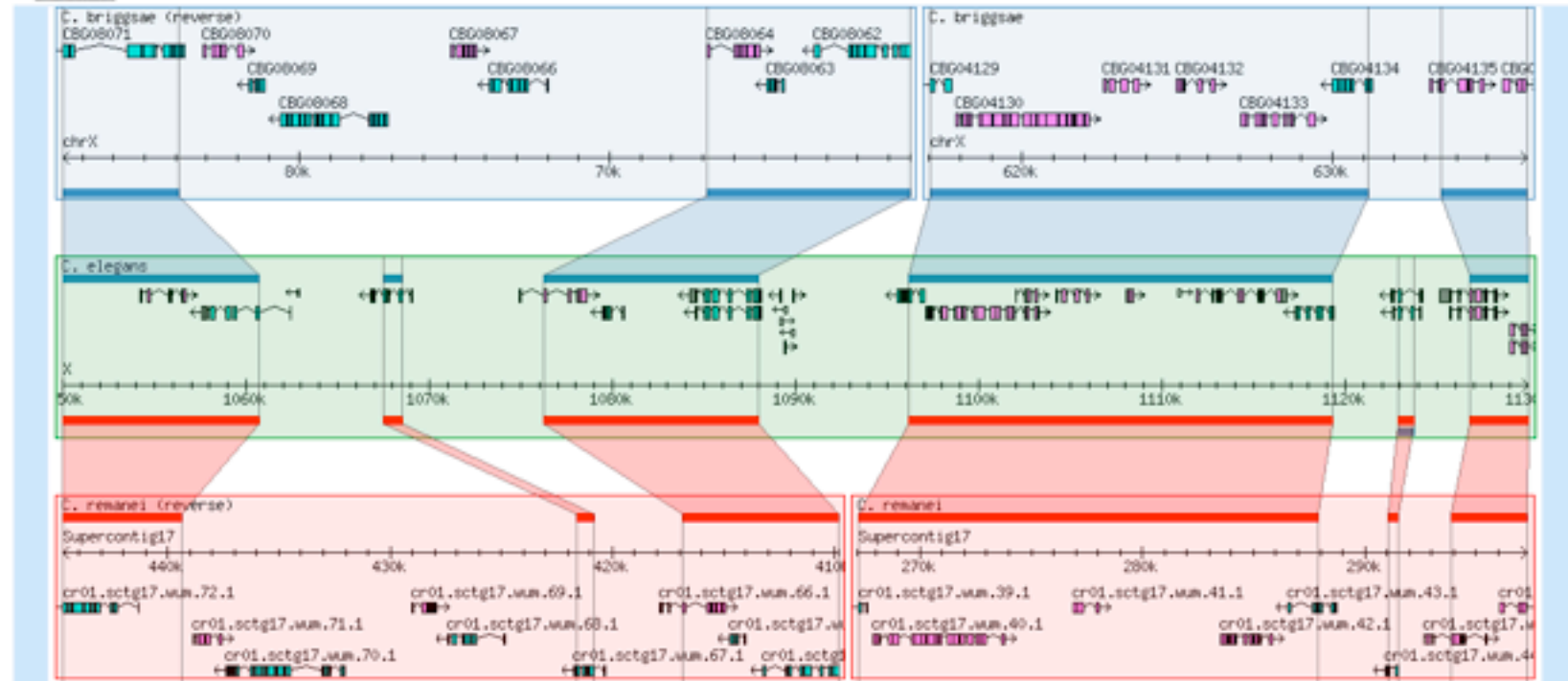


Gbrowse_syn: quick tour (strand correction)

Overview



Details



Display settings

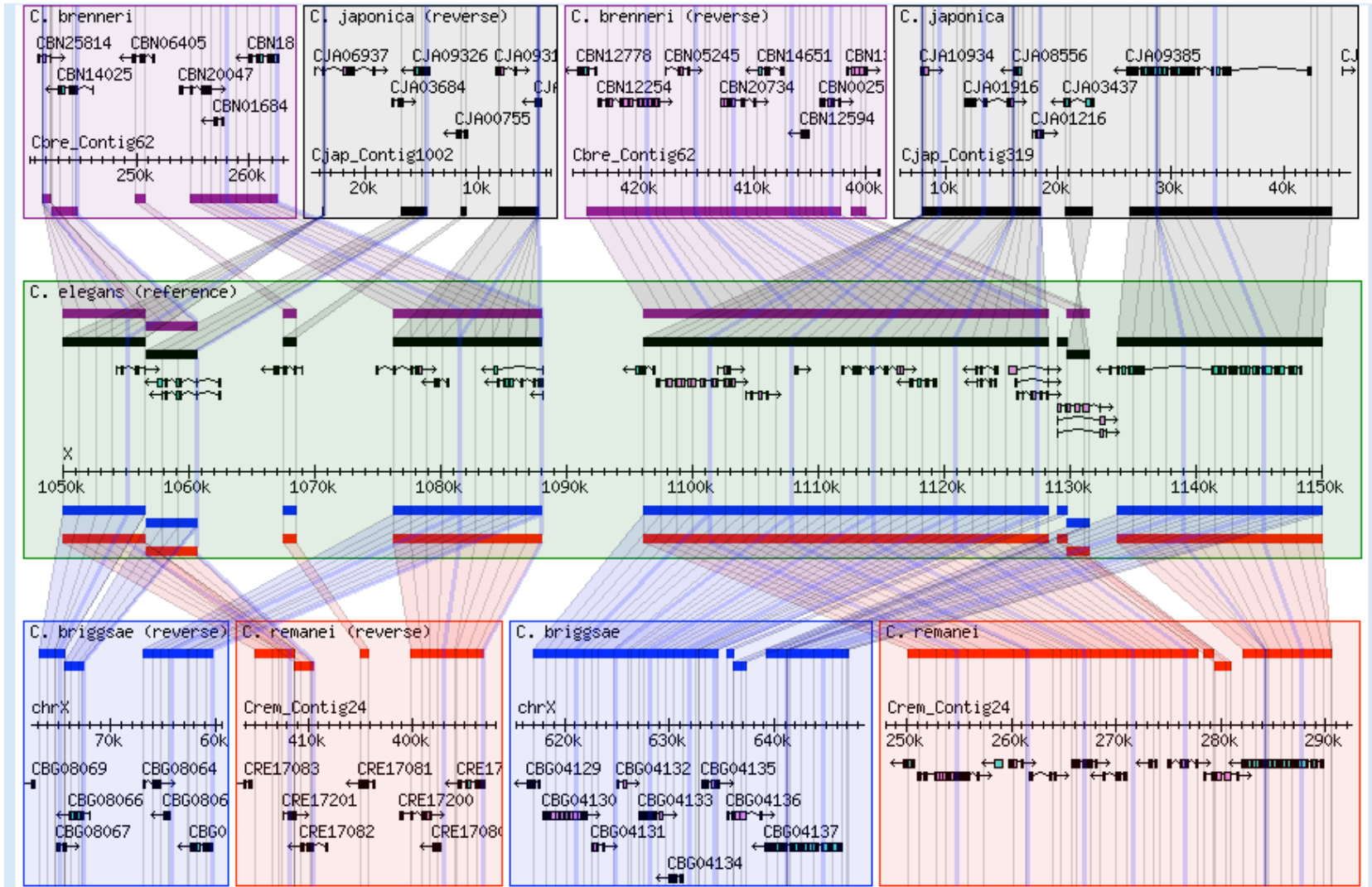
Image widths : 640 768 800 1024 1280

Image options :

Flip minus strand panels on off Allow tiny panels on off Grid lines on off Edges on off Shading on off



All in one view





How to use Insertions/Deletion data

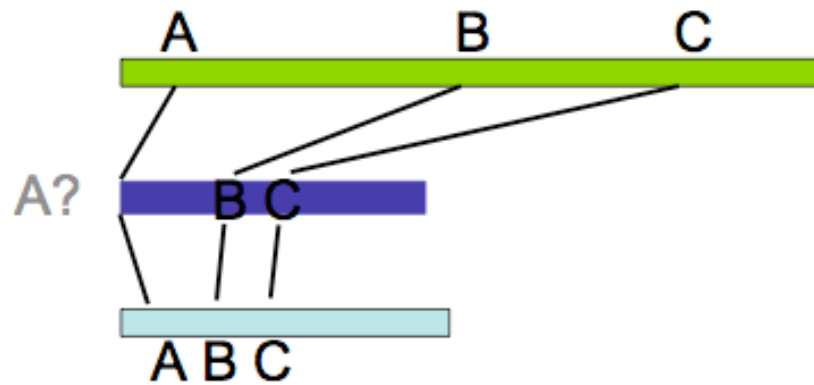
```

A
Ce-CHROMOSOME_I(+)/5195-16585 TGGCAAAAATATTTTGCATTTGCCGTTTTTCCCGTTTGCCGAAAAGTCTAATTTTCGGTAA
Cb-chrI(-)/4091935-4097143 -----
Cr-Contig8(+)/571990-577344 TTCGAAAC-----

B
Ce-CHROMOSOME_I(+)/5195-16585 TTGGGCCATTTTTCGAAATTTTGAGCCACATAAAAACTTTGAACCATTTTTGAGAAGTA
Cb-chrI(-)/4091935-4097143 -----AGAGAATGTGAAGATCTTCA-----
Cr-Contig8(+)/571990-577344 -----CAGAGAAACAGAAACAATTTTA-----
                               ** * ** * **

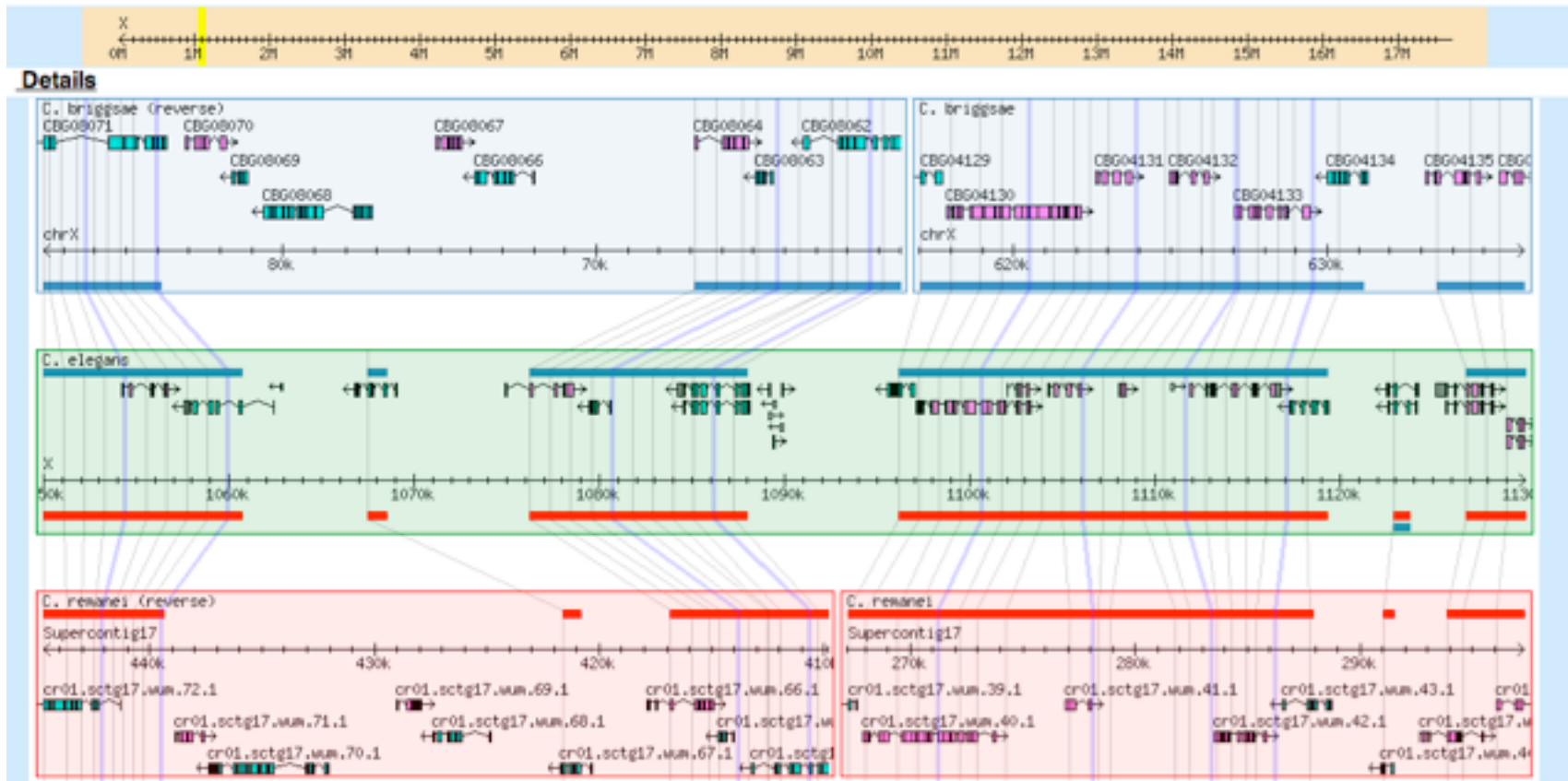
C
Ce-CHROMOSOME_I(+)/5195-16585 TTATTACGACATTCGTTTTTTGAGCACAATTTGGGCCTATACTTTCAAAATCGGGGTTT
Cb-chrI(-)/4091935-4097143 --TTCATGTCAA-----TCAT
Cr-Contig8(+)/571990-577344 --TTTCTGAAAACAGGTAGTATTATGGTTCCGAGGGTGTAGGGTTTCGAAACCGGGCCTAG
                               * * *

```



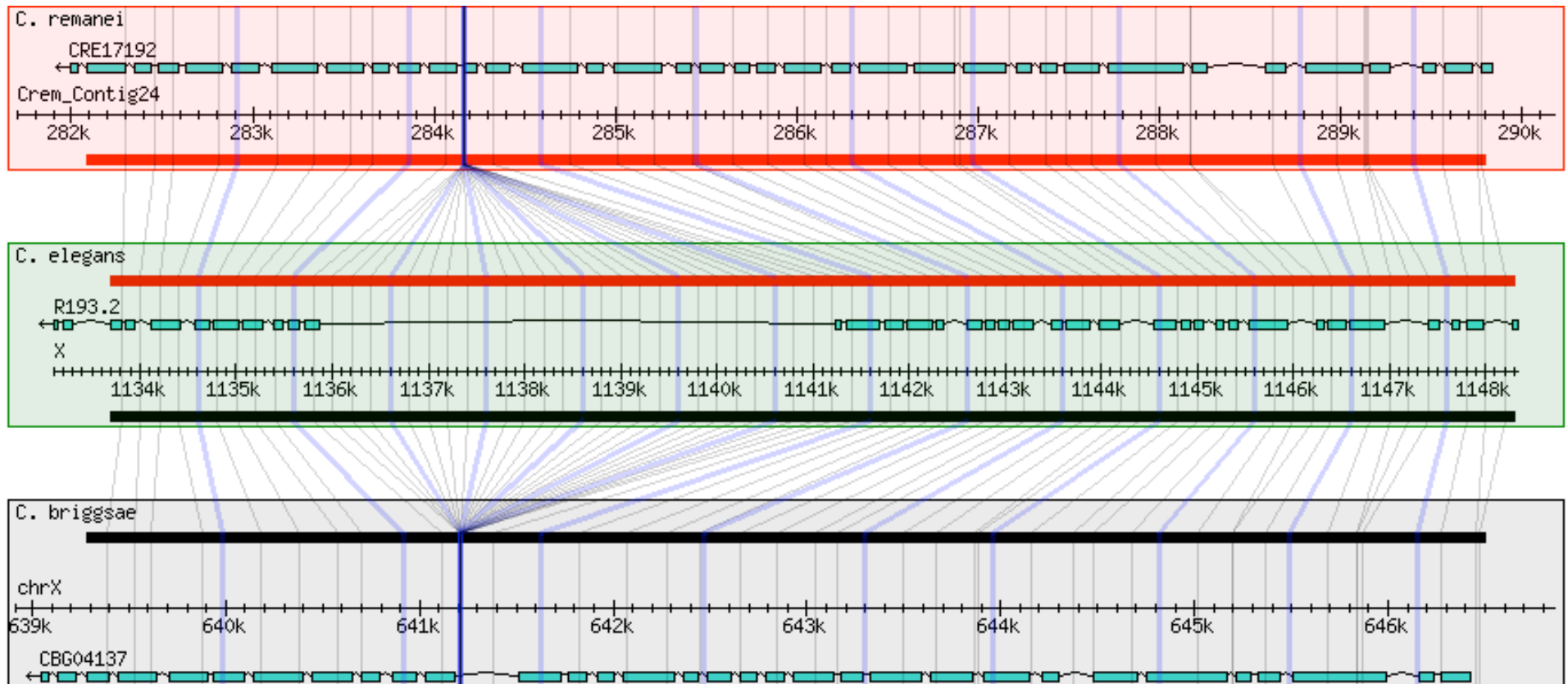


Tracking Indels with grid lines



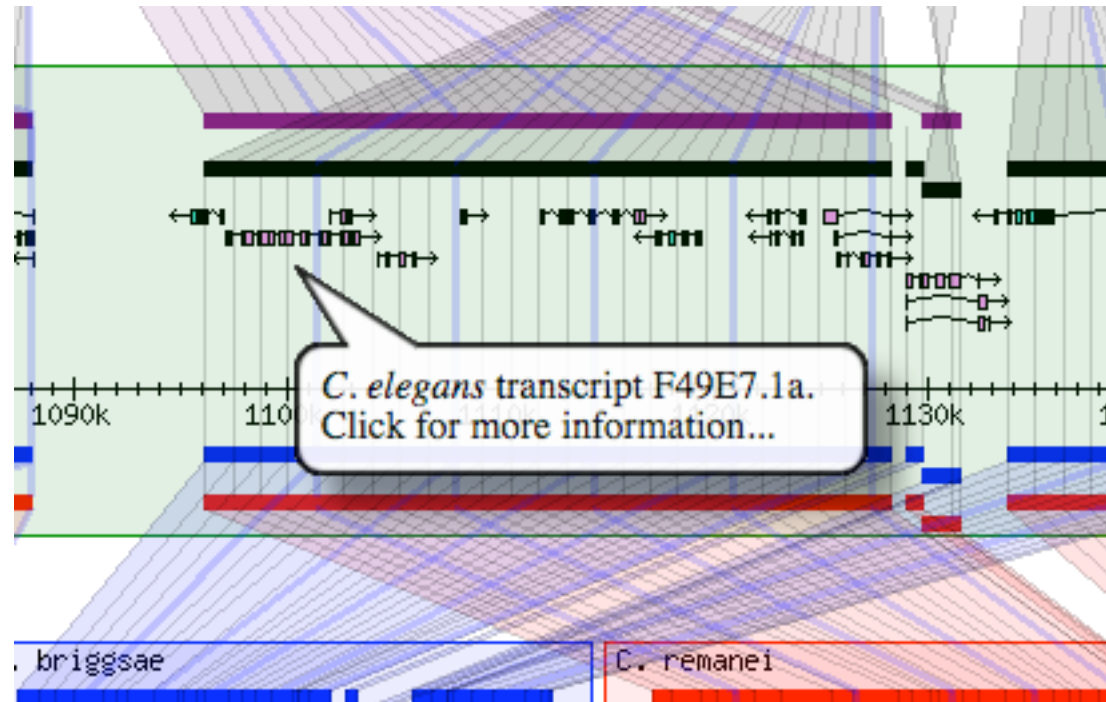


Evolution of Gene Structure



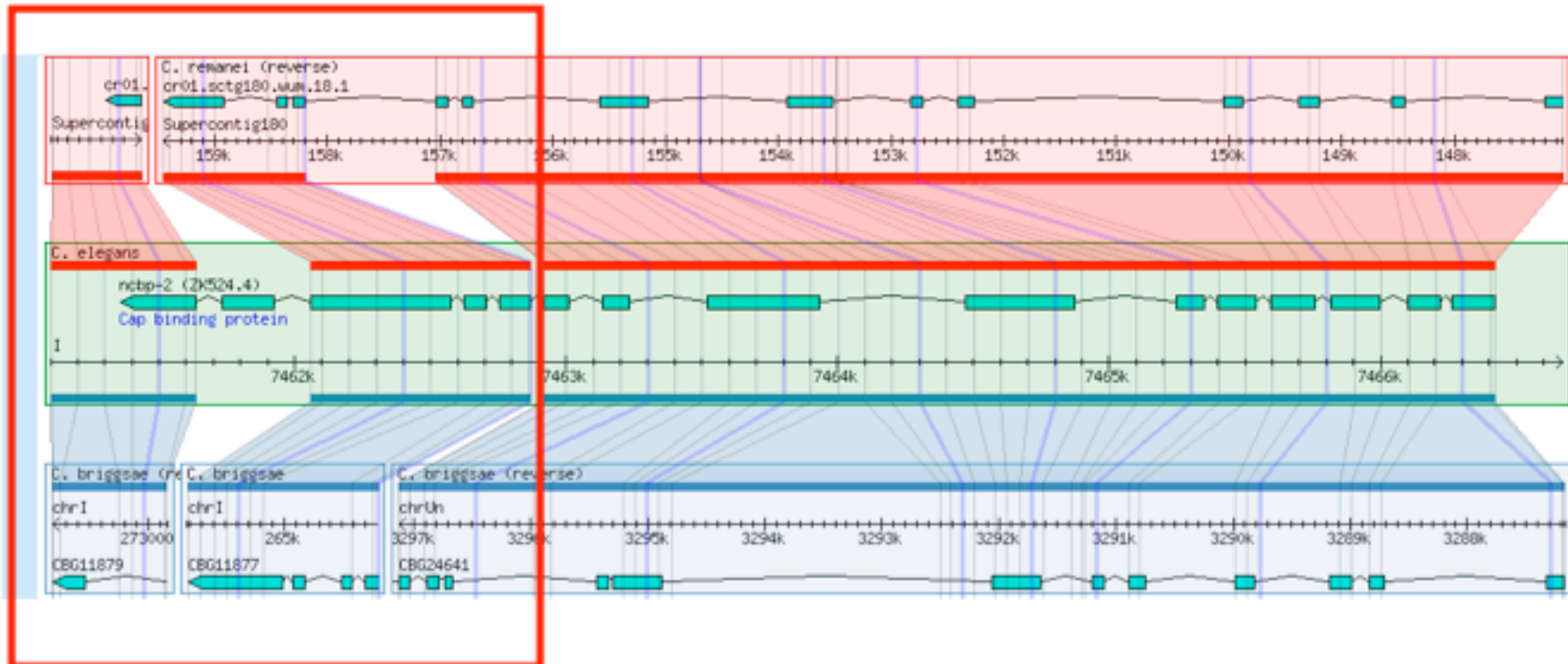


Very (un)popular popup balloons



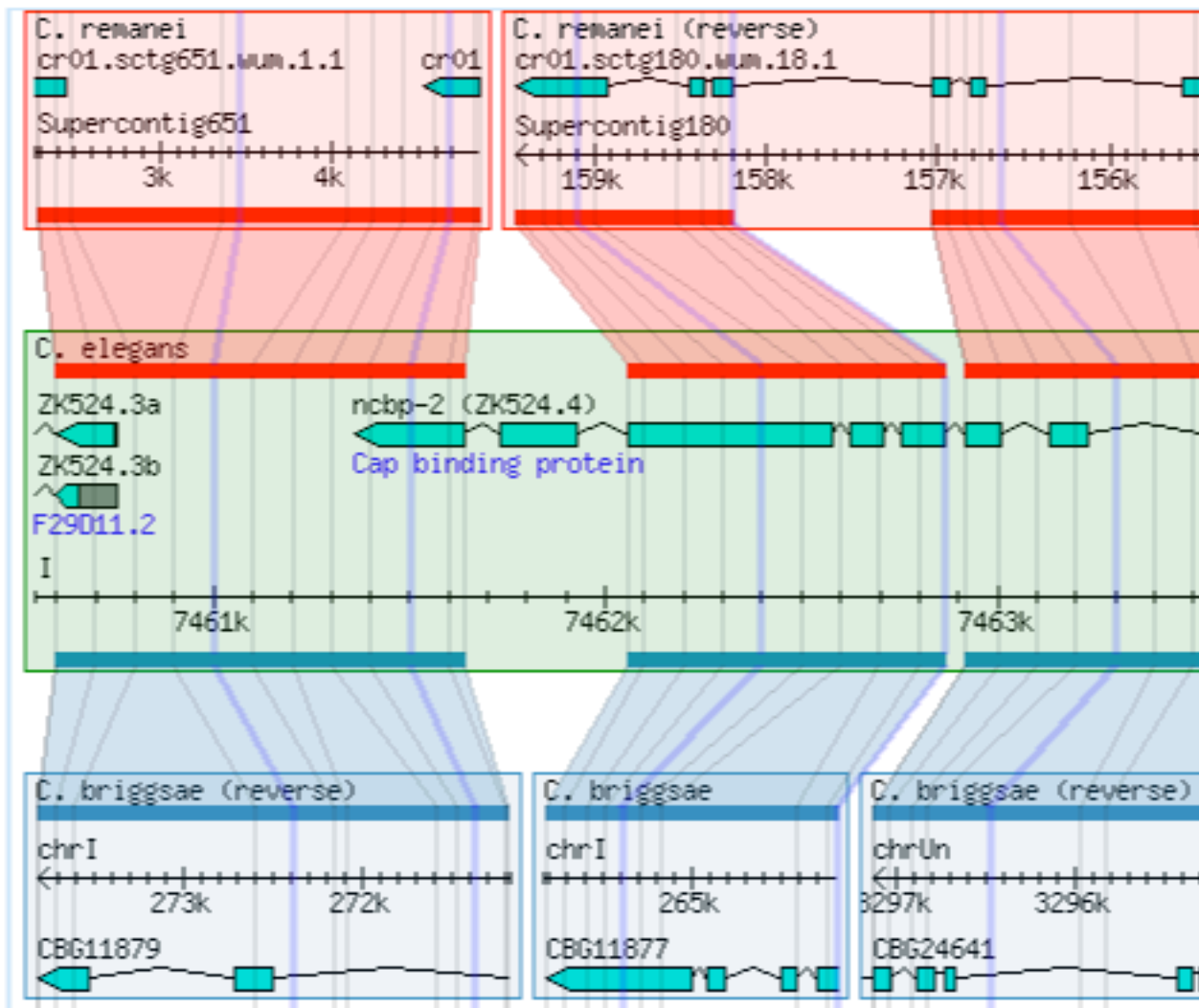


C. elegans ZK524.4 gene



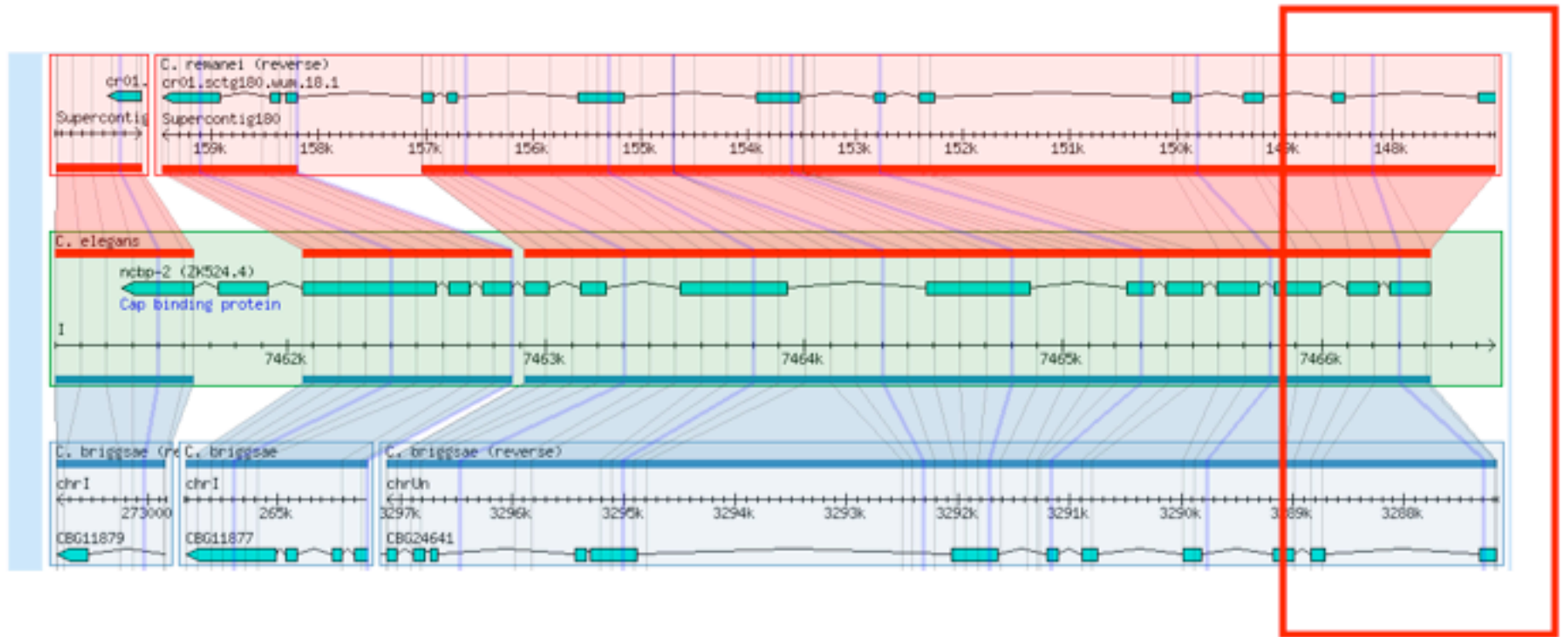


3' end



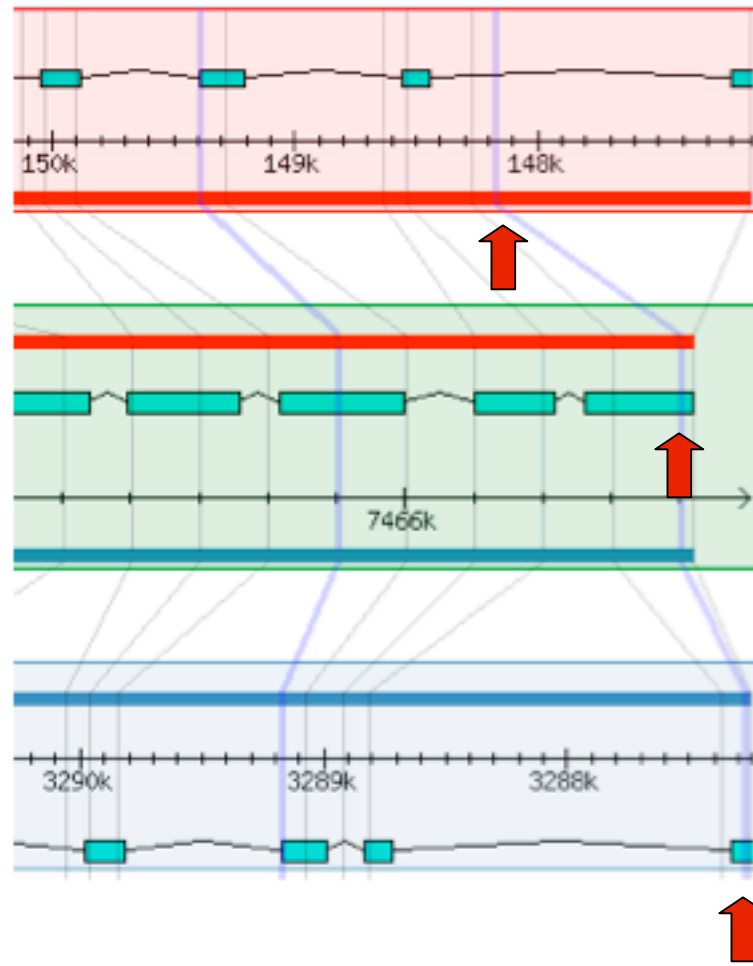


C. elegans ZK524.4 gene



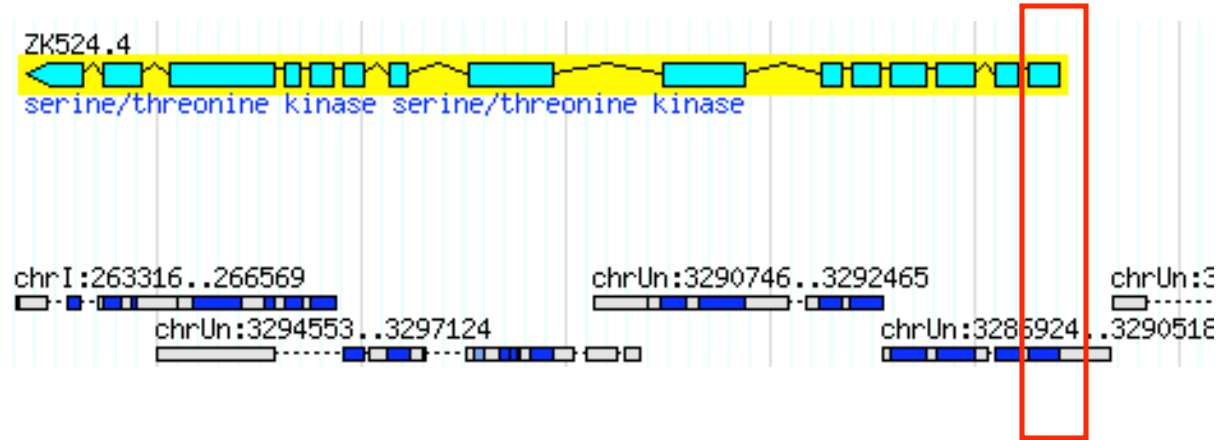


5' end

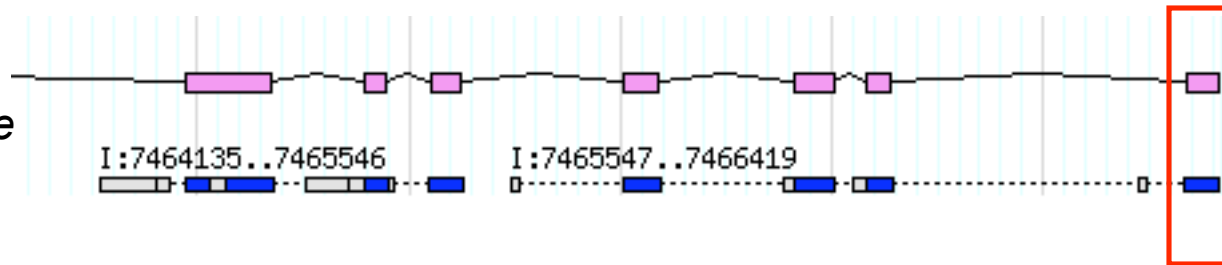




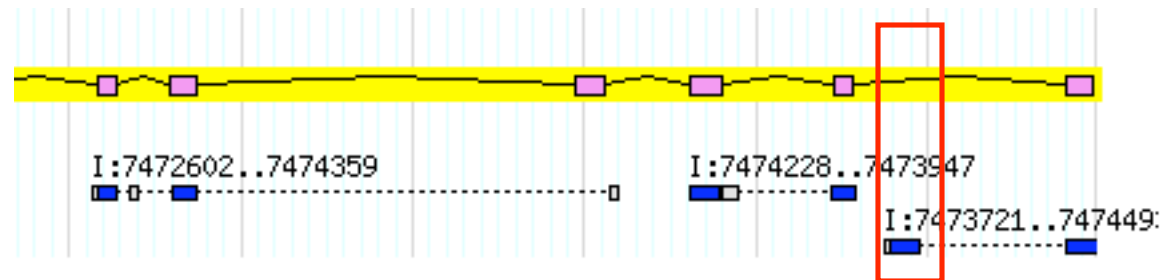
C. elegans



C. briggsae

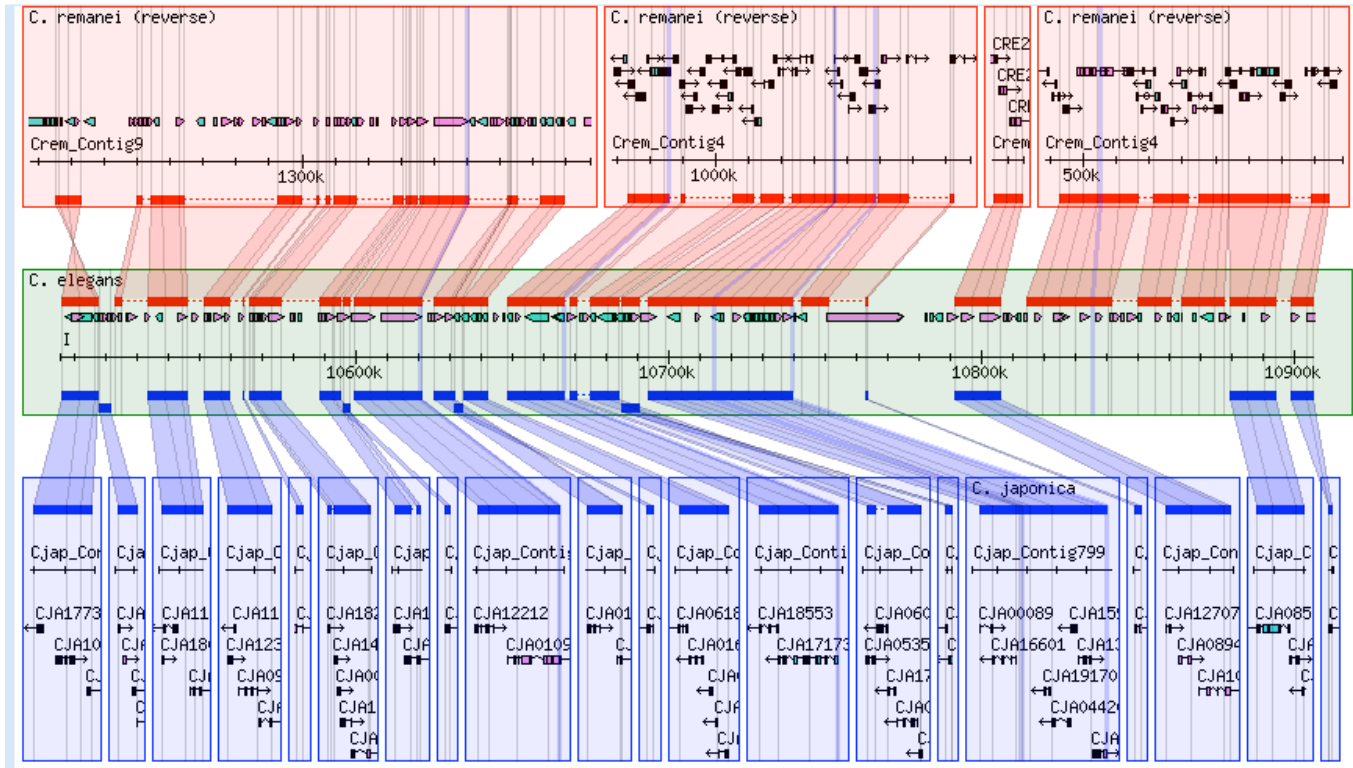


C. remanei





Visual inspection of an assembly



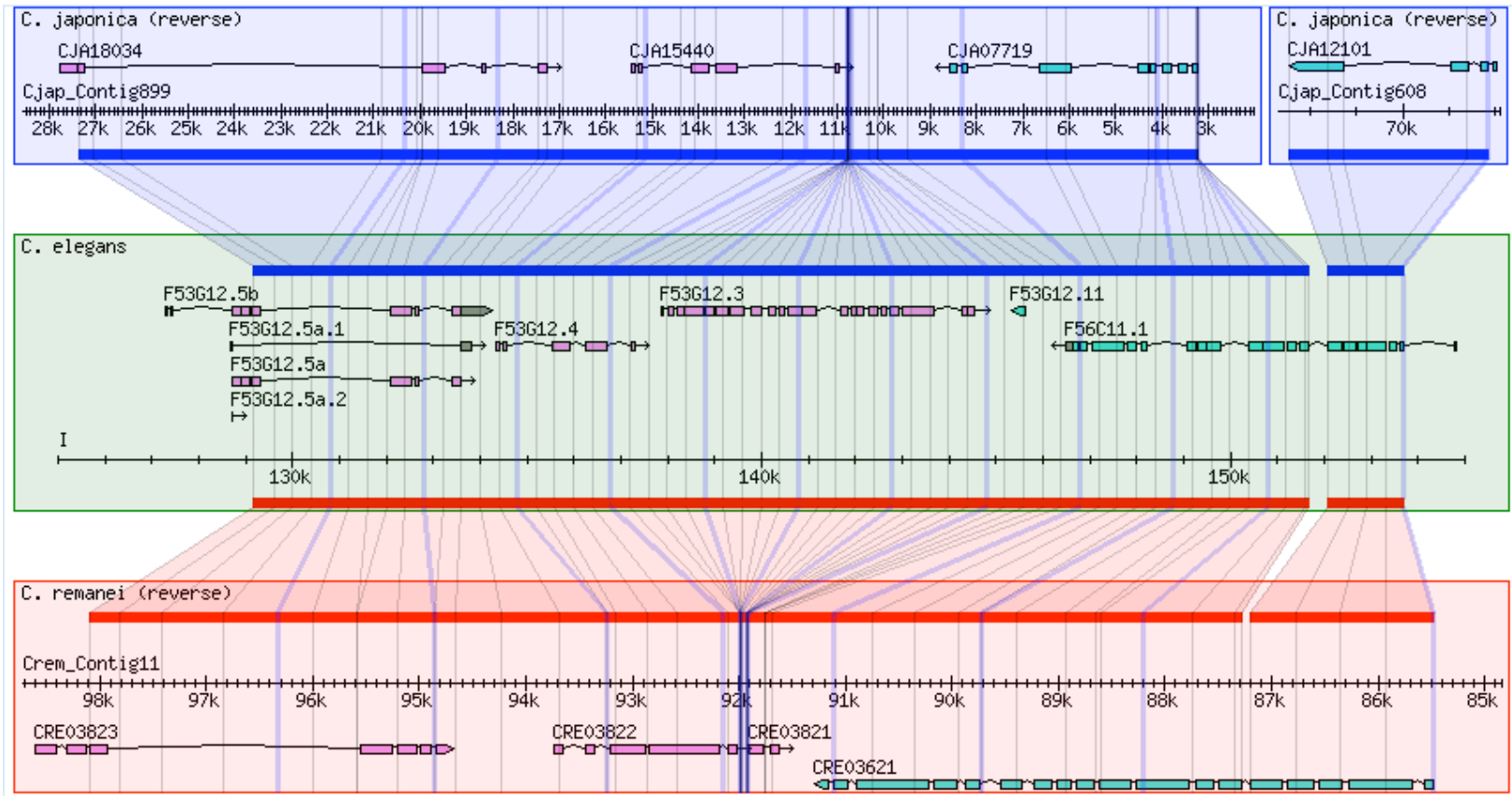
Not bad

Done

Needs work

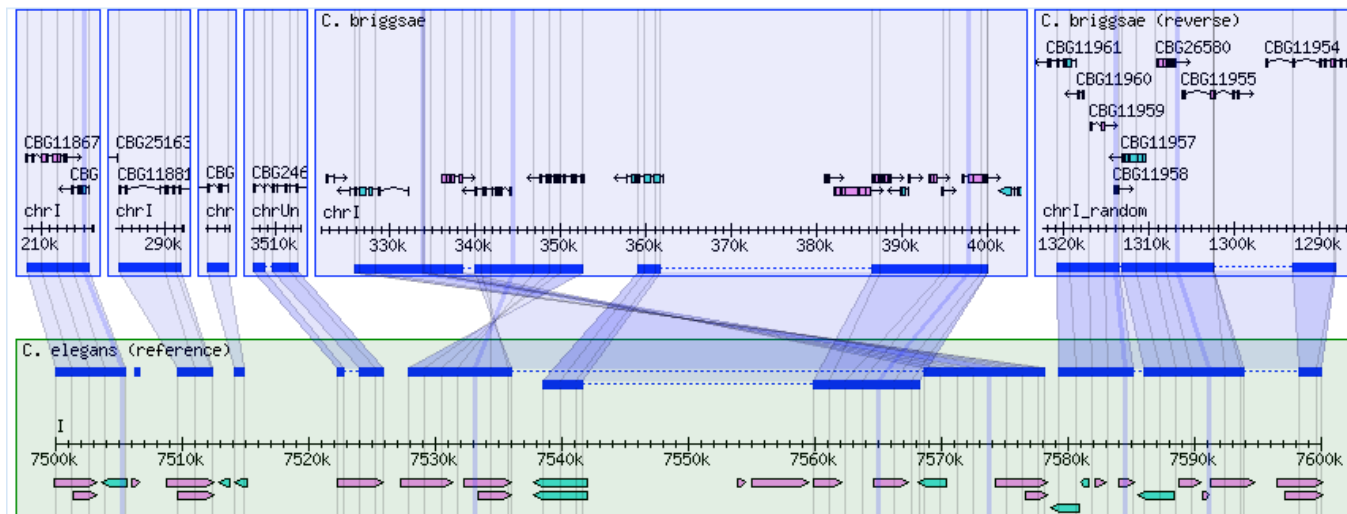
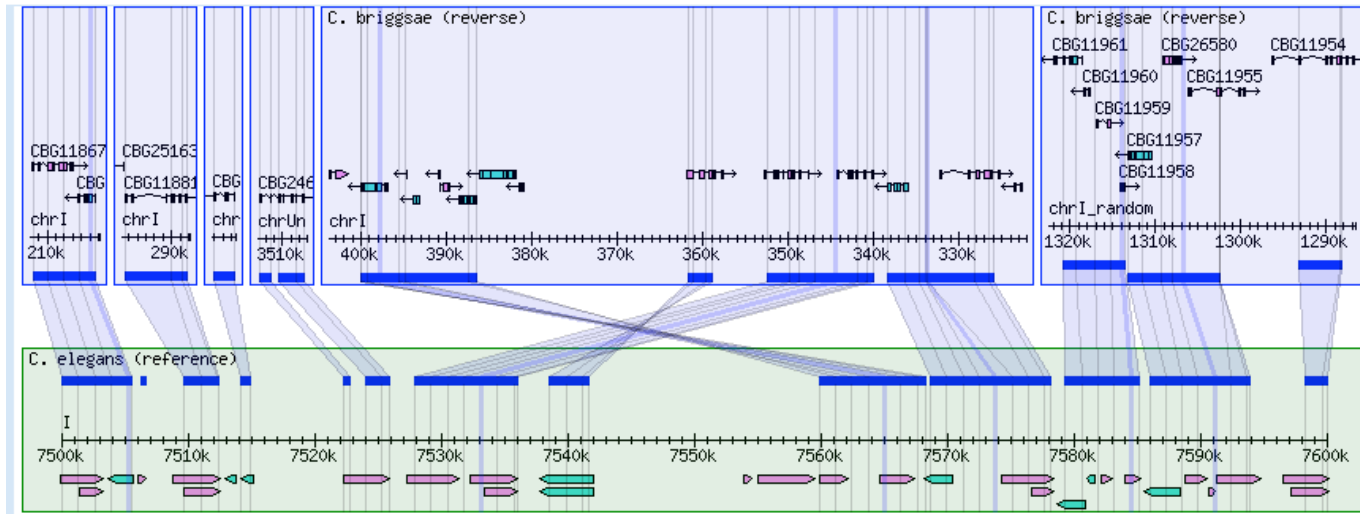


Putative gene loss



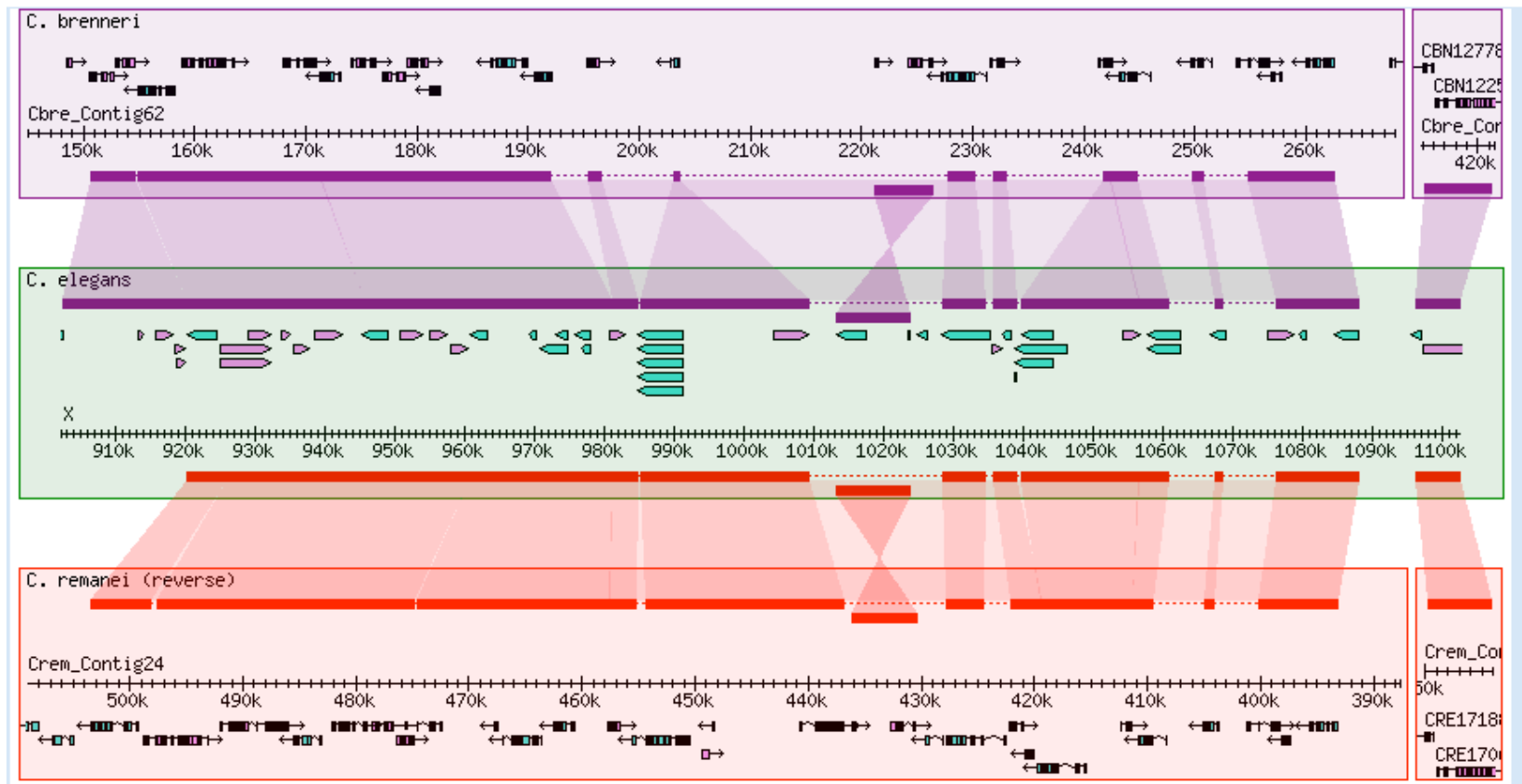


Chaining Alignments



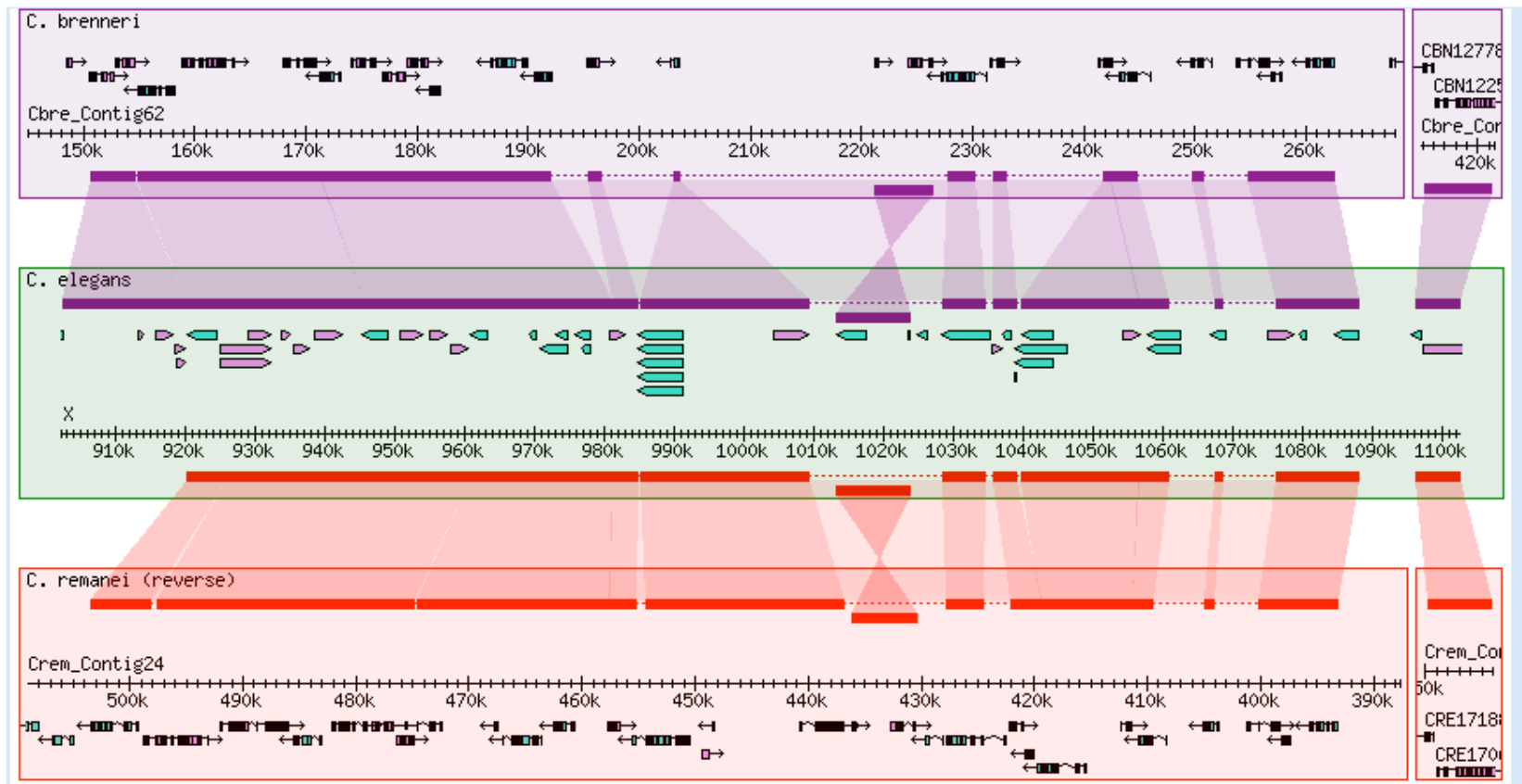


Small inversions





Small inversions





Future Improvements

- “On the fly” sequence alignment view (definitely)
- AJAX-based image configuration (probably)
- 3D image rendering? (maybe)
- Suggestions?



Documentation:

http://gmod.org/wiki/GBrowse_syn

The screenshot shows a web browser displaying the documentation for GBrowse_syn. At the top, there is a navigation bar with links for 'page', 'discussion', 'edit', 'history', 'delete', 'move', 'protect', and 'watch'. The main title is 'GBrowse syn'. Below the title, a paragraph describes GBrowse_syn as a GBrowse-based synteny browser. To the right of the text is a screenshot of the GBrowse_syn interface running on WormBase.org, showing a genomic track with various colored bars representing different data layers. Below the description is a 'Contents' table of contents with links to sections: 1 Documentation, 1.1 Installation, 1.2 Configuration, 1.3 The Alignment data, 1.4 User interface, and 1.5 Presentations and Workshops. The 'Documentation' section is currently selected and expanded, showing an 'Installation' section with three bullet points: GBrowse_syn uses much of the same infrastructure as GBrowse; it comes as part of the GBrowse distribution; and it differs in that databases and configuration for individual species are linked together via a central configuration file and a joining database that contains reciprocal alignments. The 'Installation' section also includes a recommendation to use the most up-to-date version of the application. On the left side of the page, there is a sidebar with navigation links for 'GMOD Home', 'Categories / Tags', 'Downloads', 'View all pages', 'documentation' (Overview, FAQs, HOWTOs, Glossary), 'community' (GMOD News, Support, Calendar, About this site), 'developers' (CVS, SourceForge Site), and 'search'. On the right side, there is a box titled 'Learn about GBrowse_syn at the SMBE 2009 GMOD Workshop' with a logo for the workshop.



Acknowledgements

Lincoln Stein
Todd Vision
Dave Clements
Scott Cain
Jason Stajich
Michael Han
WormBase Curators



Washington University Genome Sequencing Center

<http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-104.html>

