



GBrowse and NextGen Sequencing Data

Scott Cain
GMOD Project Coordinator
Ontario Institute for Cancer Research
scott@scottcain.net



Ontario Institute
for Cancer Research



OCCBio
June 15, 2010

Introduction: GMOD is ...

- A set of interoperable open-source **software** components for visualizing, annotating, and managing biological data.
- An active **community** of developers and users asking diverse questions, and facing common challenges, with their biological data.



Who uses GMOD?



Plus hundreds of others

Visualization

GBrowse

JBrowse

GBrowse_syn

Releases

1.70 and 2.08 released

2.09 and 1.71 in the pipe

Features

Rubberband region selection

Drag and drop track ordering

Collapsible tracks

Popup balloons

Allele/genotype frequency, LD glyphs

Geolocation popups

Circular genome support (1.71)

Asynchronous updates (2.0)

User authentication

Multiple server support (2.0)

SQLite, BigWig, SAMtools (NGS) adaptors



The generic genome browser: a building block for a model organism system database. Stein LD et al. (2002) *Genome Res* 12: 1599-610



Visualization

GBrowse

JBrowse

GBrowse_syn

GMOD's 2nd Generation Genome Browser
It's *fast*

Completely new genome browser implementation:
Client side rendering
Heavy use of AJAX
Uses JSON and Nested Containment Lists

Demo: [JBrowse Fly](#)

<http://jbrowse.org/genomes/dmel/>



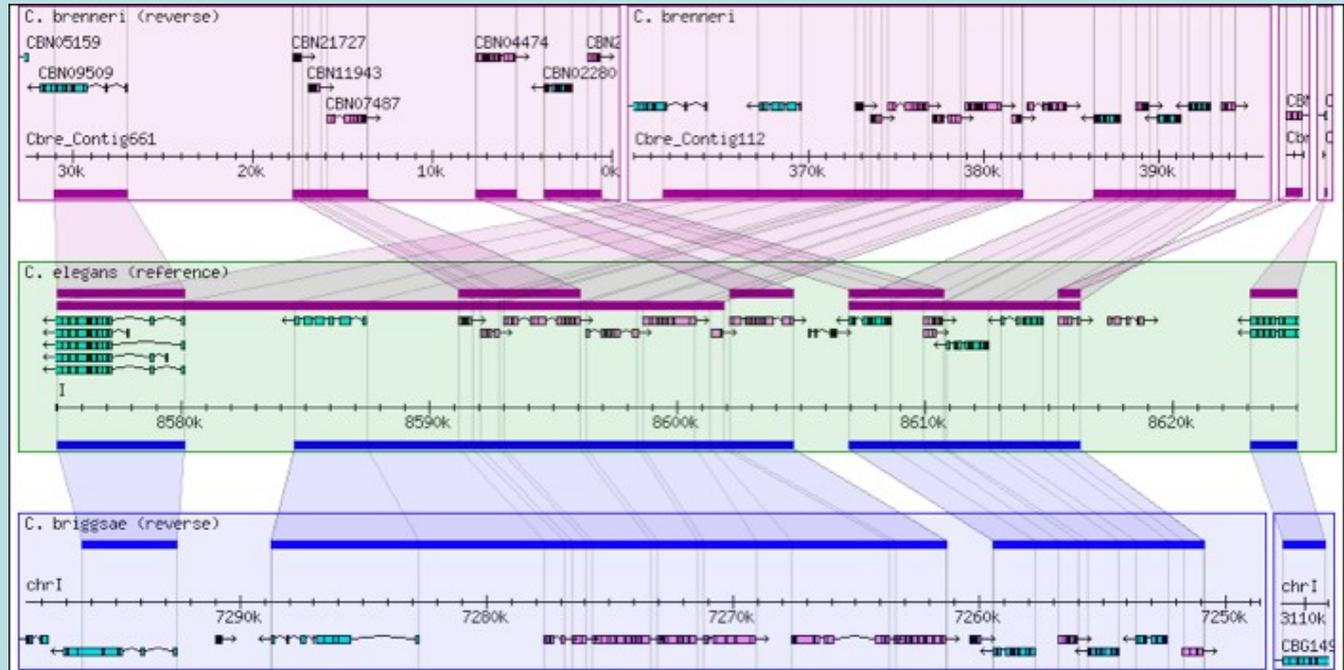
Visualization

GBrowse

JBrowse

GBrowse_syn

GBrowse based comparative genomics viewer
Shows a reference sequence compared to 2+ others
Can also show any GBrowse-based annotations



wormbase.org

Syntenic blocks do not have to be colinear
Can also show duplications



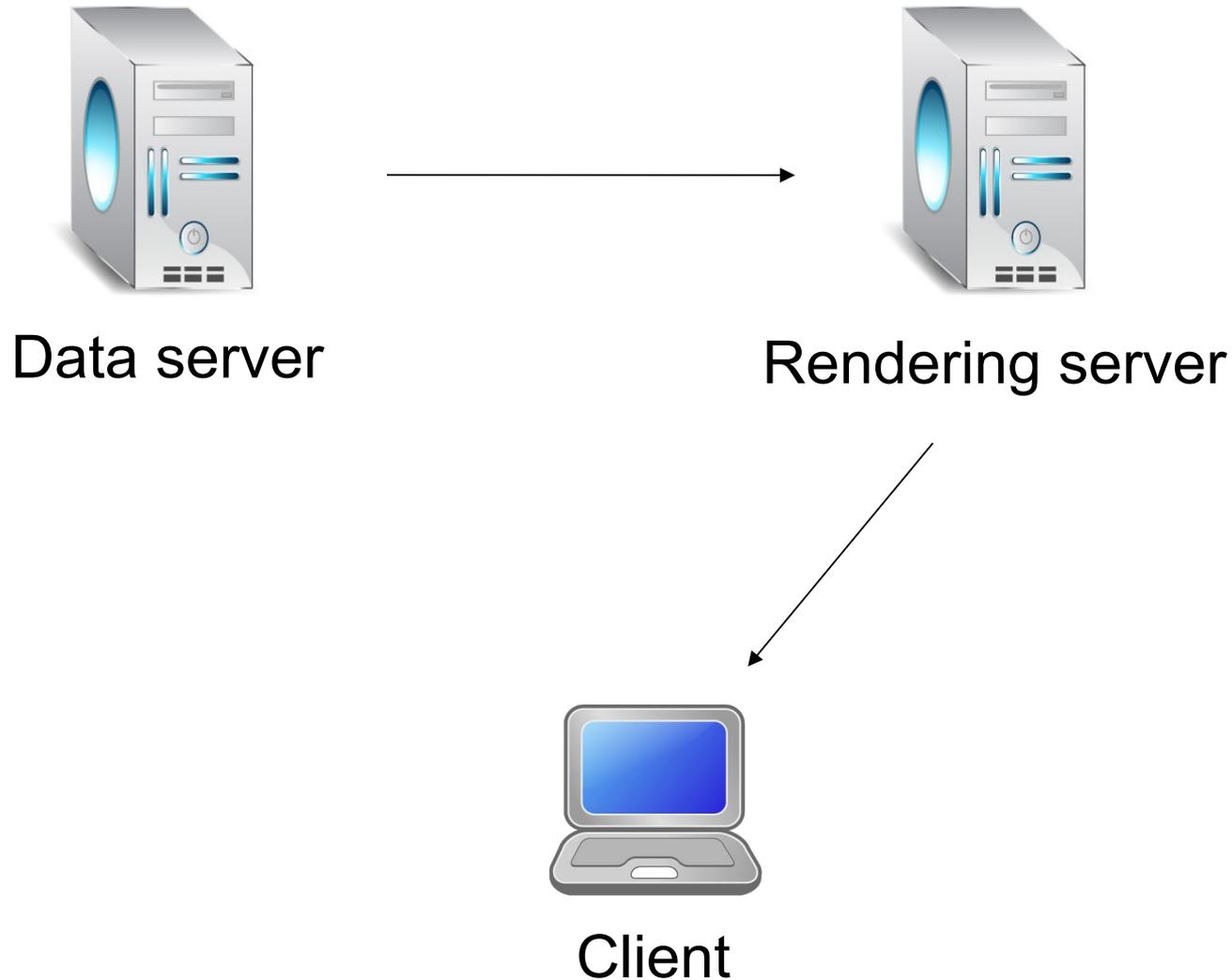
Displaying NGS in GBrowse

- GBrowse (2.0 or 1.70, but 2.0 is easier)
 - Install 2.0 from CPAN
 - Install 1.70 from installer script at gmod.org
- SAMtools
 - Install from Sourceforge:
<http://samtools.sourceforge.net/>
- Some NextGen data that can be aligned to a reference sequence

Why is GBrowse2 Better?

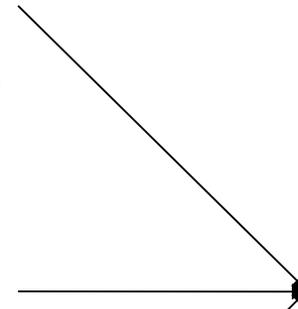
- Nicer user interface (AJAX to reload tracks, tabs instead of one long page)
- User accounts
- But, really important for us: support for multiple data and rendering servers

GBrowse 1.X architecture



GBrowse 2.0 architecture

Gene annotations
Bio::DB::GFF,
SeqFeature::Store



Client

Continuous or
conservation data
(wiggle, Big Wig)



NGS Bio::DB::Sam
data



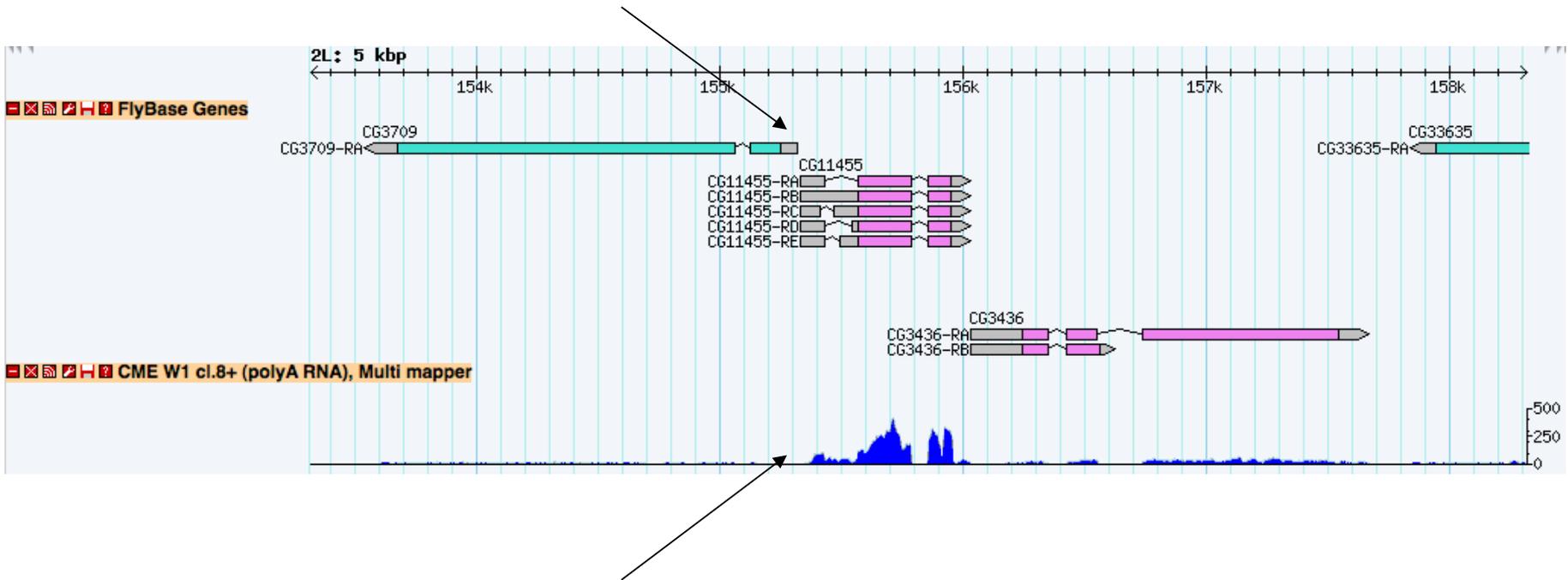
Data servers

Rendering servers

What does SAMtools get you?

- Speed!
- SAMtools takes alignment data from various NGS alignment tools (Bowtie, MAQ, etc.)
- Converts to a common text format (SAM, Sequence Alignment/Map)
- Then converts to an indexed binary file (BAM, Binary Alignment/Map), which provides for very fast data retrieval.

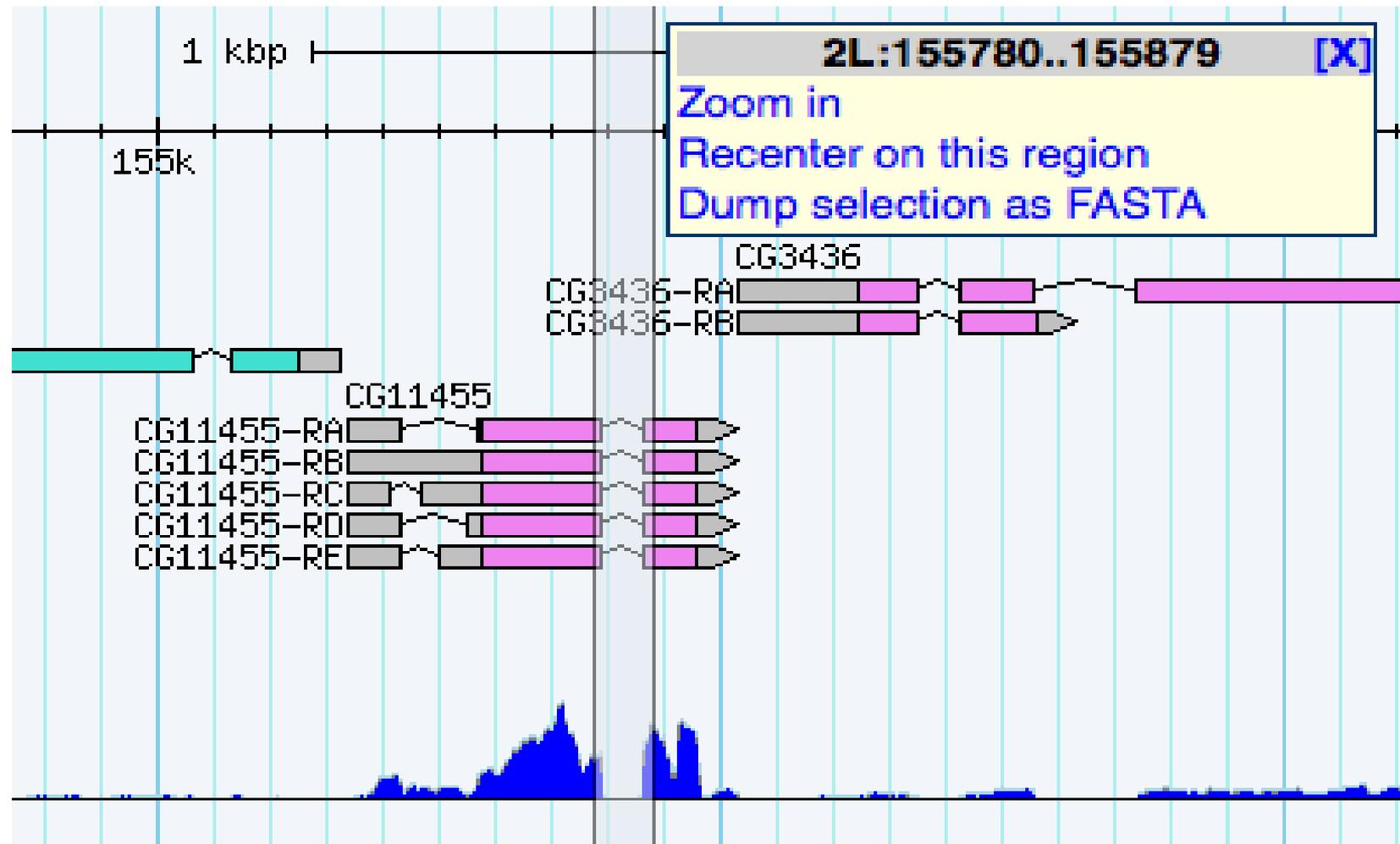
Gene annotations from a Bio::DB::SeqFeature::Store database



RNA-Seq data from a SAMtools BAM file

From ModEncode: <http://modencode.oicr.on.ca/gb2/gbrowse/fly/>

Zooming in...

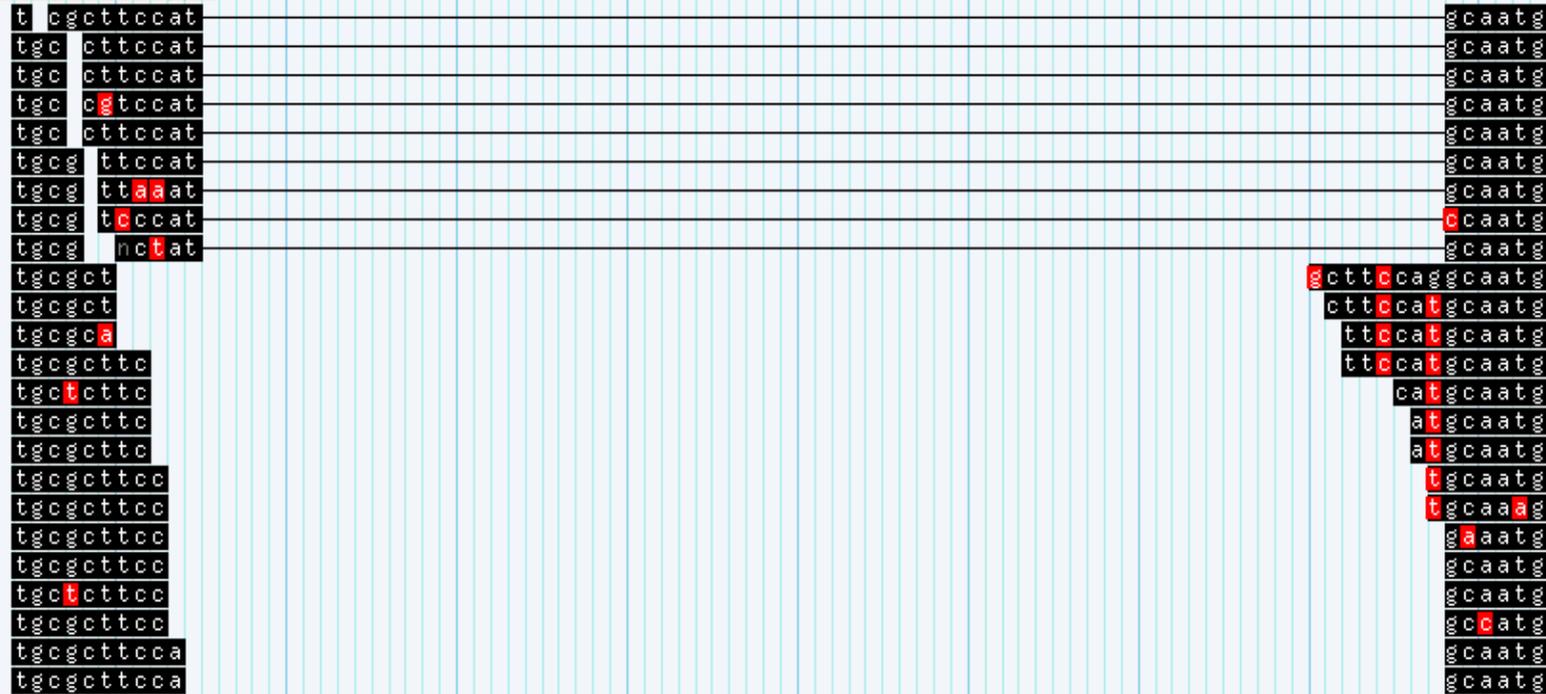


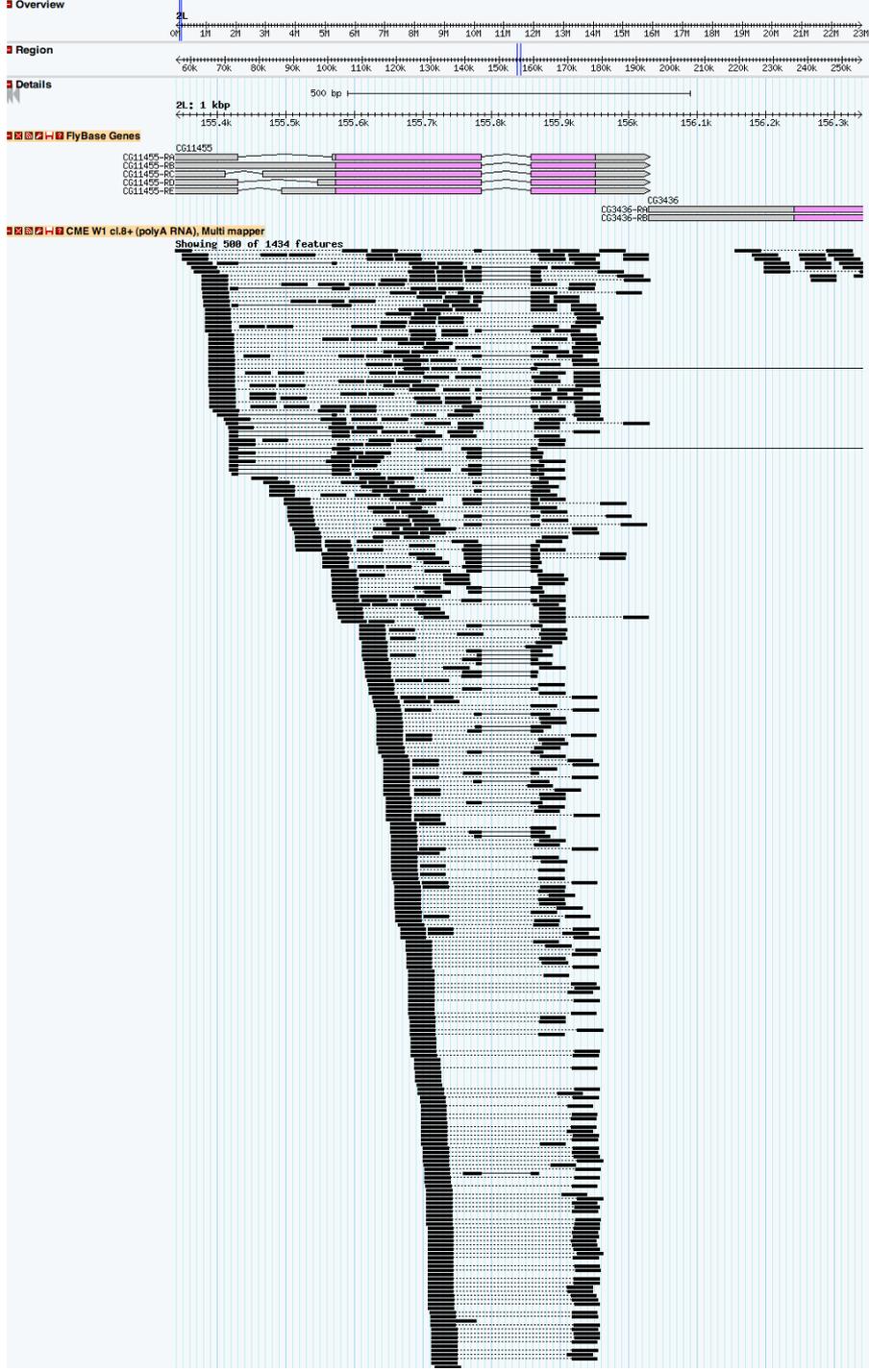
Base-level views with mismatch highlighting

FlyBase Genes

CG11455
CG11455-RA
CG11455-RB
CG11455-RC
CG11455-RD
CG11455-RE

CME W1 cl.8+ (polyA RNA), Multi mapper





Conclusion

GBrowse is a widely used, open source genome browser that has support for multiple data sources and the flexibility to display any type of annotation, along with continuous data and NextGen sequencing alignments.

Resources

- **GBrowse**
 - Home Page <http://gmod.org/wiki/GBrowse>
 - WebGBrowse <http://webgbrowse.cgb.indiana.edu/>
 - Mailing List <https://lists.sourceforge.net/lists/listinfo/gmod-gbrowse>
- **SAMtools**
 - Home Page <http://samtools.sourceforge.net/>
 - Mailing List <https://lists.sourceforge.net/lists/listinfo/samtools-help>
- **JBrowse**
 - Home Page <http://jbrowse.org>
 - Mailing List <https://lists.sourceforge.net/lists/listinfo/gmod-ajax>
- **Gbrowse_syn**
 - Home Page http://gmod.org/wiki/GBrowse_syn
 - Mailing List <https://lists.sourceforge.net/lists/listinfo/gmod-gbrowse>

Acknowledgements

GBrowse

Lincoln Stein – Ontario Institute for Cancer Research

Sheldon McKay – Arizona State

Ben Faga – University of Iowa

Ian Davies – University of Waterloo

SAMtools

Heng Li – Broad Institute