JBrowse — 1.3.0 and beyond

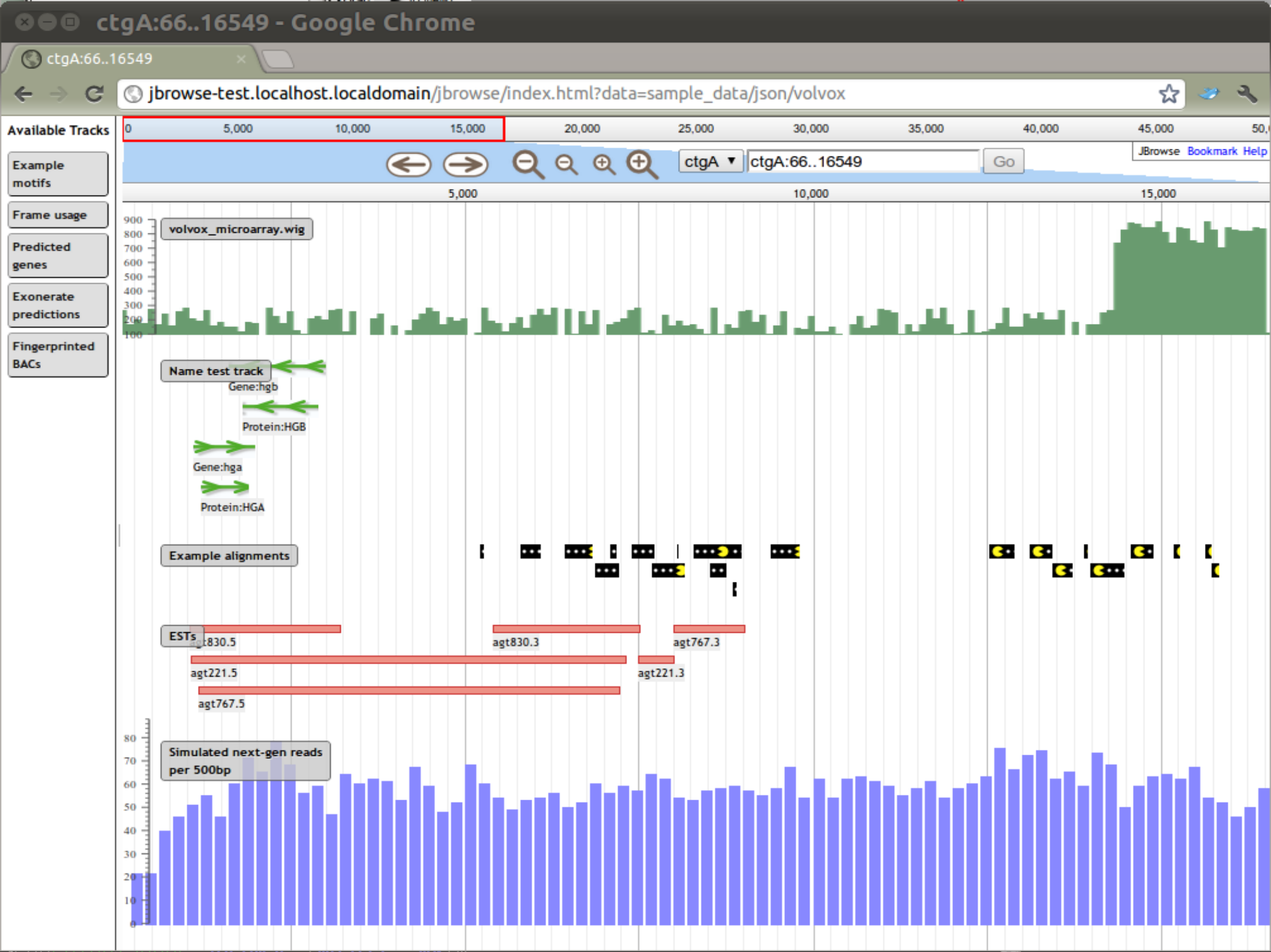
Robert Buels

GMOD Community Meeting

April 5, 2012

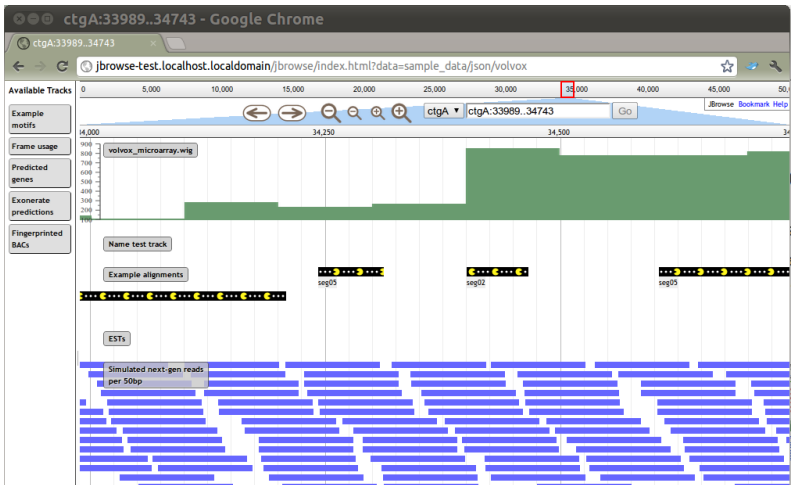
University of California, Berkeley

(JBrowse is currently at 1.2.2)


ctgA:66..16549 - Google Chrome
ctgA:66..16549
jbrowse-test.localhost.localdomain/jbrowse/index.html?data=sample_data/json/volvox
Available Tracks
Example motifs
Frame usage
Predicted genes
Exonerate predictions
Fingerprinted BACs
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
45,000
ctgA
ctgA:66..16549
Go
JBrowse Bookmark Help
volvox_microarray.wig
Name test track
Gene:hgb
Protein:HGB
Gene:hga
Protein:HGA
Example alignments
ESTs
agt830.5
agt830.3
agt767.3
agt221.5
agt221.3
agt767.5
Simulated next-gen reads per 500bp

JBrowse...

- Fast, smooth navigation (*think Google Maps for genomes)*
- Supports BED, GFF, Bio::DB::*, Chado, WIG, BAM, UCSC (*intron/exon structure*, *name lookups*, *quantitative plots*).
- Relies on pre-indexed static files, which is highly secure and performant. (*future versions will have an optional server-side component*)
- Is stably funded by NHGRI, many interesting innovations implemented and/or pending integration.
- Is open source, of course.
- Did I mention it's fast?



The JBrowse Project

- free and open source (license: GPL / Artistic)
- a GMOD project
  http://gmod.org

- developed using `git`, hosted on GitHub
  http://github.com/GMOD/jbrowse
- PIs most involved: Ian Holmes, Lincoln Stein, Suzi Lewis
- just got 3 more years of funding from NHGRI!

A Brief History of JBrowse

2006 First prototype (pre-rendered GBrowse).
NHGRI funds 3yr development of project.
Mitch Skinner joins the team.

2007 CSHL Biology of Genomes: scaling.

2008 JavaScript-only rendering.

2009 Genome Research paper.

2010 WebApollo funded to develop JBrowse.
NHGRI renews for 6mo.
JBrowse = “official” GBrowse successor.

2011 Mitch Skinner joins genomics startup.
US politicians play chicken with economy.
NHGRI renews for 3yr.

2012 (January) Robert Buels joins the team.

Release Roadmap

1.3.0 – Very Imminent™

1.3.N bugfix releases as needed

1.4.0 – next month

1.4.N bugfix releases as needed

2.0.0 – end of 2012?

Version numbers according to Semantic Versioning (semver.org) informal standard.

Course Set for 2.0

- JBrowse needs a lot more **Big Features:**
  - need to upload and share tracks
  - need data export
  - need large numbers of tracks with rich metadata; searchable
  - need powerful, modular extensions
    - data backends
    - client-side plugins

Yes, but ...

JBrowse 1.2.x Urgent Problems

- Flat-file formatting is horribly inefficient.
- Lack of Y-axis labels make quantitative tracks nearly meaningless.
- Lack of dynamic (rubber-band) zooming is very annoying.
- Unusable for large numbers of ref seqs (e.g. rough assemblies) – a major use case.
- No tests for anything: **major** problem for developers.

JBrowse Architecture

- Still has its baby teeth.
- Non-backward-compatible changes needed to for JBrowse to sustainably achieve the **Big New Features.**
- Separation of configuration from data.
- Modular data backends.
- Modular frontend (event publish/subscribe).

JBrowse 1.3.0 Objectives

- Fix the urgent problems that are not so hard.
- Fix minor problems that are easy.
- Add more polish.
- Preserve backward compatibility.

JBrowse 1.3.0 Highlights

- **“rubber-band” dynamic zooming**
- much better flat file formatting performance
- **support for JavaScript “hooks” to modify FeatureTracks**
- **Y-axis labels**
- sequence track revcom (thanks to Ian Davis)
- stock index.html supports a `data` URL param to use different data dirs
- **inline help dialog**
- more grid lines
- data format supports more than 2 levels of features
- **NCL and sequence data can be compressed on disk**
- compatibility layer for 1.2.x data
- **1-based coordinates in the user interface**
- Julie Moon - “embedded” mode (like `gbrowse_img`)
- improved documentation – POD, JSdoc, and tutorial
- (and lots of improvements behind the scenes)
  - Perl unit and integration tests, JS integration tests with Selenium, internal refactoring

FeatureTrack Hooks

```
“hooks”: {
    “modify”: “function(track, feat, elem) {
          var fType = feat.get('Type');
          if (fType) {
              elem.className = 'basic';
              switch (fType]) {
              case 'CDS':
              case 'thick':
                  elem.style.height = "10px';
                  elem.style.marginTop = '-3px';
                  break;
              case "UTR":
              case "thin":
                  elem.style.height = '6px';
                  elem.style.marginTop = '-1px';
                  break;
          }
          elem.style.backgroundColor = 'red';
     }”
}
```

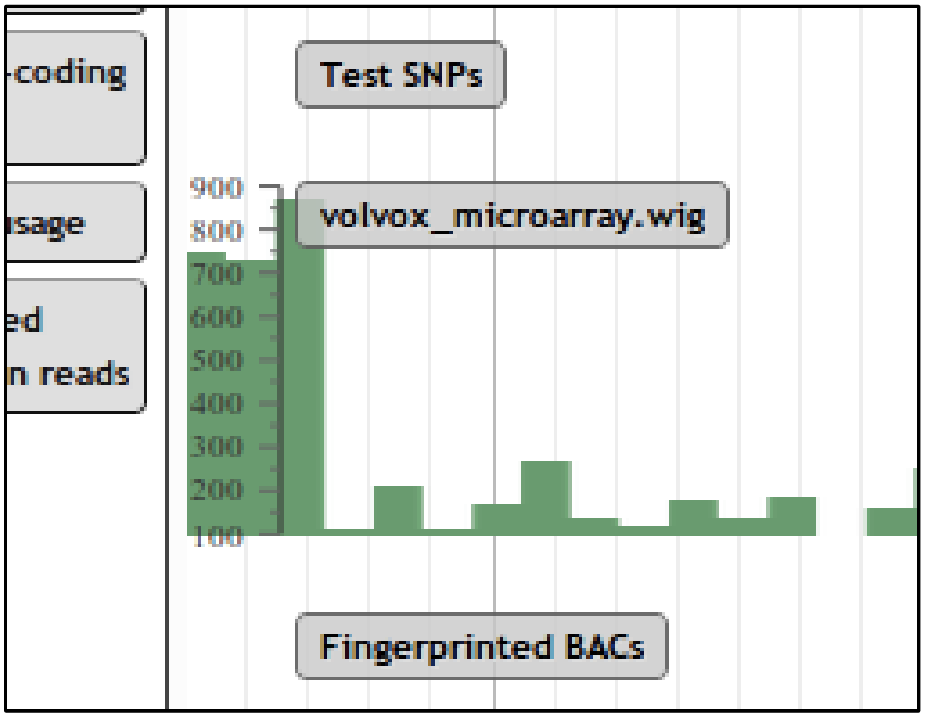
have to hack the `data/trackList.json` to use these >:-{

Embedded Mode

Julie Moon, OICR co-op student

http://localhost/jbrowse/index.html
?loc=ctgA:1386..6057
&tracks=<list of tracks>
&data=sample_data/json/volvox
&nav=0
&overview=0
&tracklist=0

Y-Axis Labels

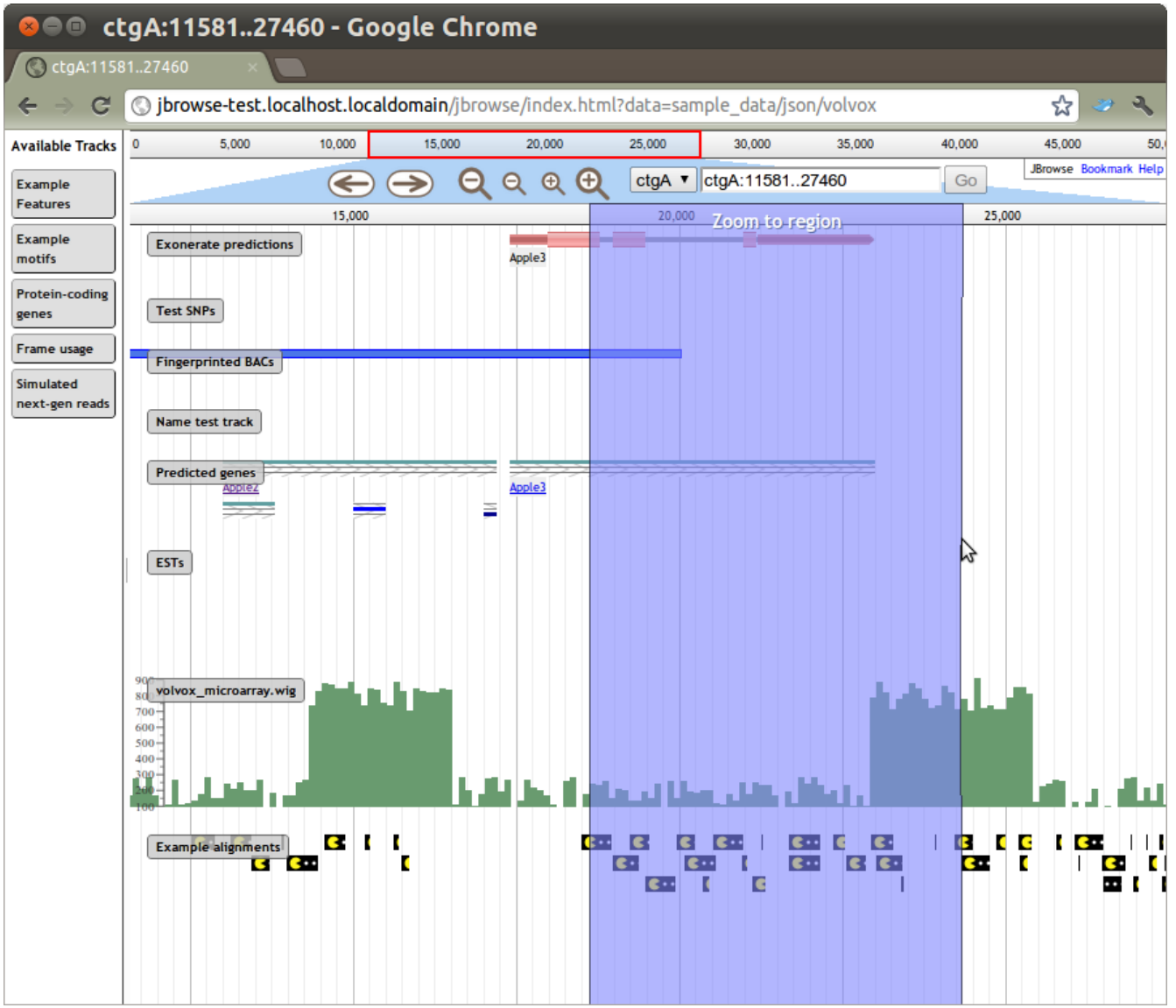


- dojox.charting (SVG/VML)
- fully dynamic, of course

Pre-Compression

- gzips static data files on disk
- shrinks data directories by **about 60%**
  - takes ~~hg19~~GRCh37 from 19GB to 7.6GB.
- needs some webserver configuration for headers
  - Apache: AllowOverride FileInfo (.htaccess files)
  - `nginx`: add a smallish config snippet

Rubber-band Zooming



Queued for 1.4.0

- performance for large numbers of reference sequences
- autocompletion for refseq and feature names
- probably a few more things
- maybe can hoist faceted track search and rich track metadata here

Low-hanging fruit, really-super-urgent stuff.

Aims for 2.0

- User-editable, direct config file.
- Load and search rich track metadata for thousands of tracks. (what formats?)
- Datasource plugin system, support direct byte-range access to BAM, BigWig, and BigBed files.
- New, documented pub/sub event API ties together both JBrowse components and **any other JS code.**

Config File

- what format?
- leaning toward GBrowse-native format
- will also support injection of configuration information dynamically

Publish / Subscribe Example

- JBrowse publishes “/jbrowse/evt/region/highlighted chrI:12345..67890”
- Some other code has subscribed to this, and
  - calls a web service to launch an analysis on that region
  - pops up a notification saying it has done so
- When the analysis is done, it then publishes: “/jbrowse/command/track/add <track config JSON>”
- Track with the analysis results pops up in the browser

Track Metadata

- What formats do people use to manage this?
  - Excel?
  - RDBMS schemas?
  - XML?
- What should JBrowse do for loading or directly reading it?

Big Thanks To:

Suzi Lewis (LBNL)

Gregg Helt

Ed Lee

Nomi Harris

Ian Holmes (Berkeley)

Mitch Skinner

Stephen Cummings

Lincoln Stein (OICR)

Julie Moon

Natalie Fox

Adam Wright

Lukas Mueller (BTI, SGN)

NHGRI

http://jbrowse.org

http://gmod.org/wiki/JBrowse

github

http://github.com/GMOD/jbrowse

(yes, jbrowse needs a logo)

The End

Robert Buels

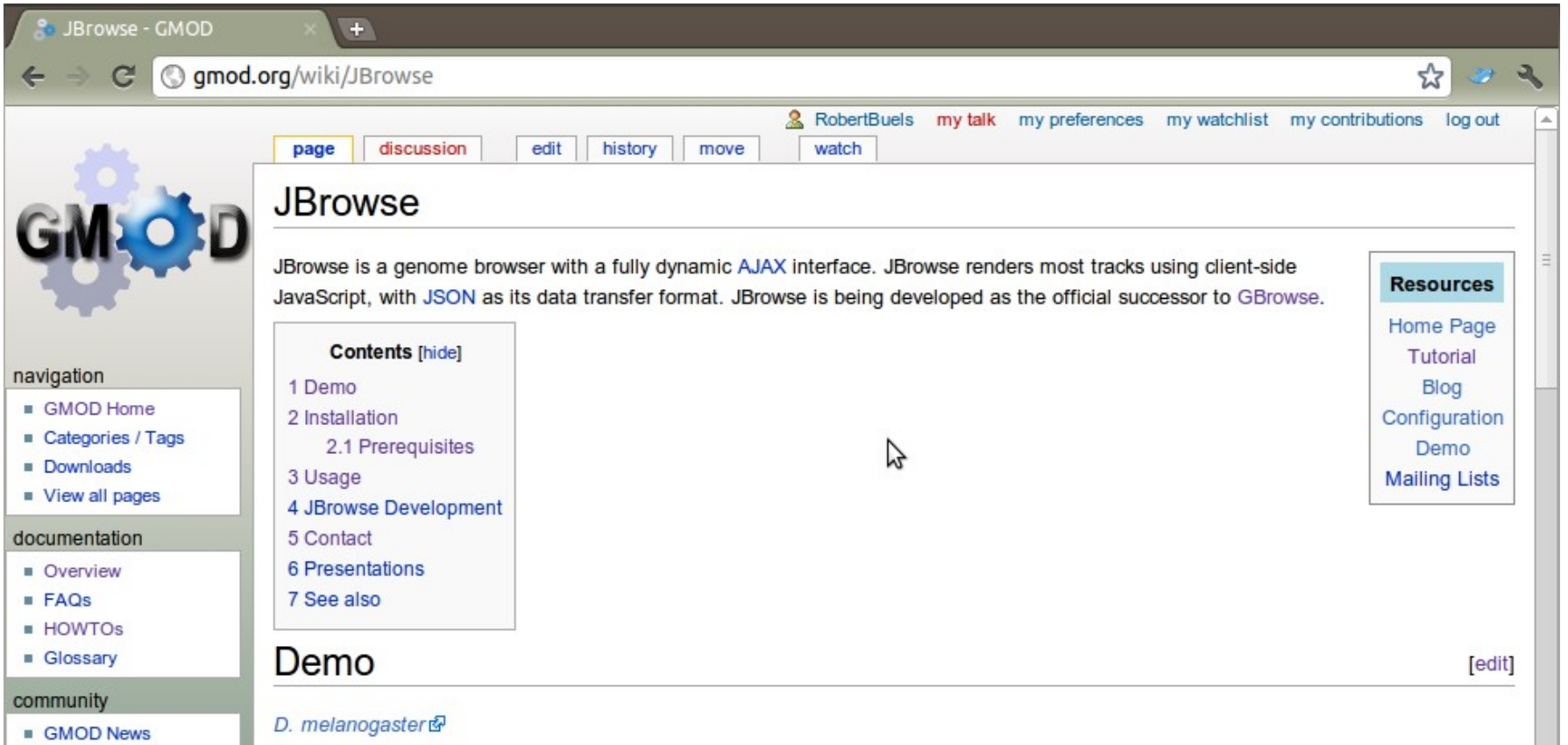
rbuels@gmail.com

rbuels on freenode IRC
(#bioperl, #gmod)

- Was Engineering Lead at Sol Genomics Network http://solgenomics.net, at BTI with Lukas Mueller
- Became new JBrowse Lead Developer as of **January 2012** at UC Berkeley with Ian Holmes

JBrowse Documentation

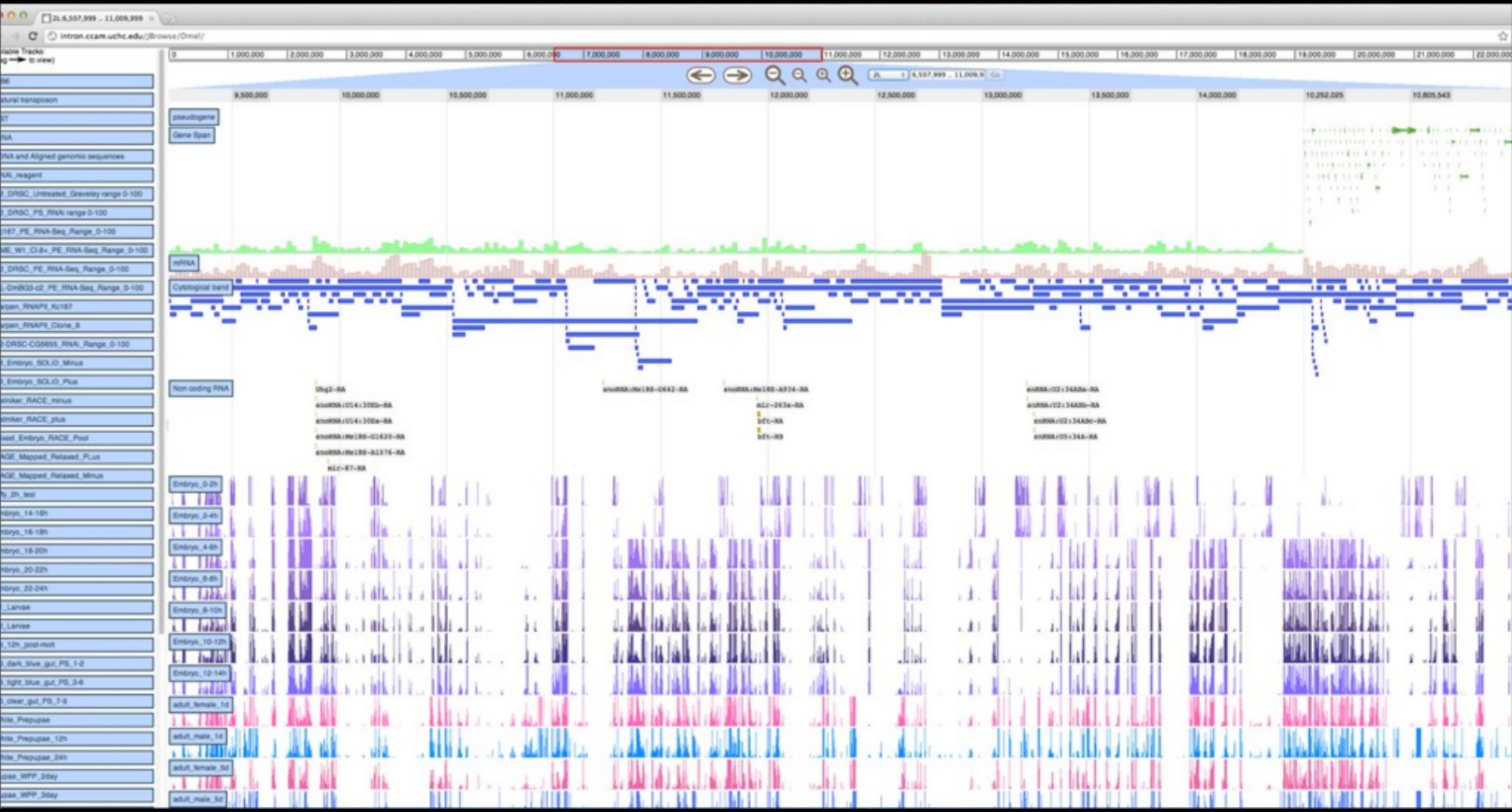
The GMOD wiki is the documentation hub
http://gmod.org/wiki/JBrowse



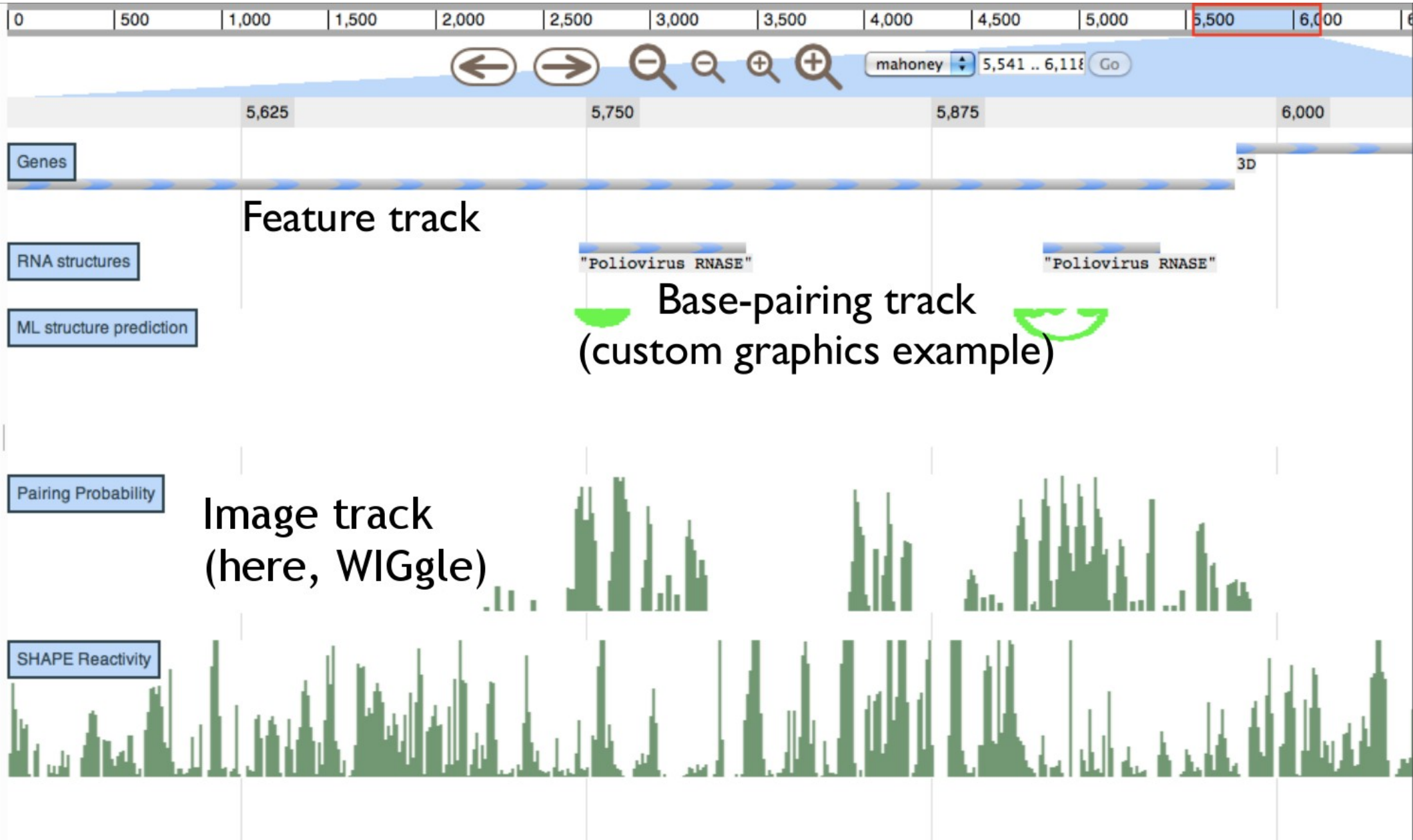
First look: Live Demo

- a couple of JBrowses around the web
  - http://intron.ccam.uchc.edu/JBrowse/Dmel/
  - http://jbrowse.org/ucsc/hg19/

- maybe install a demo JBrowse locally with the tutorial

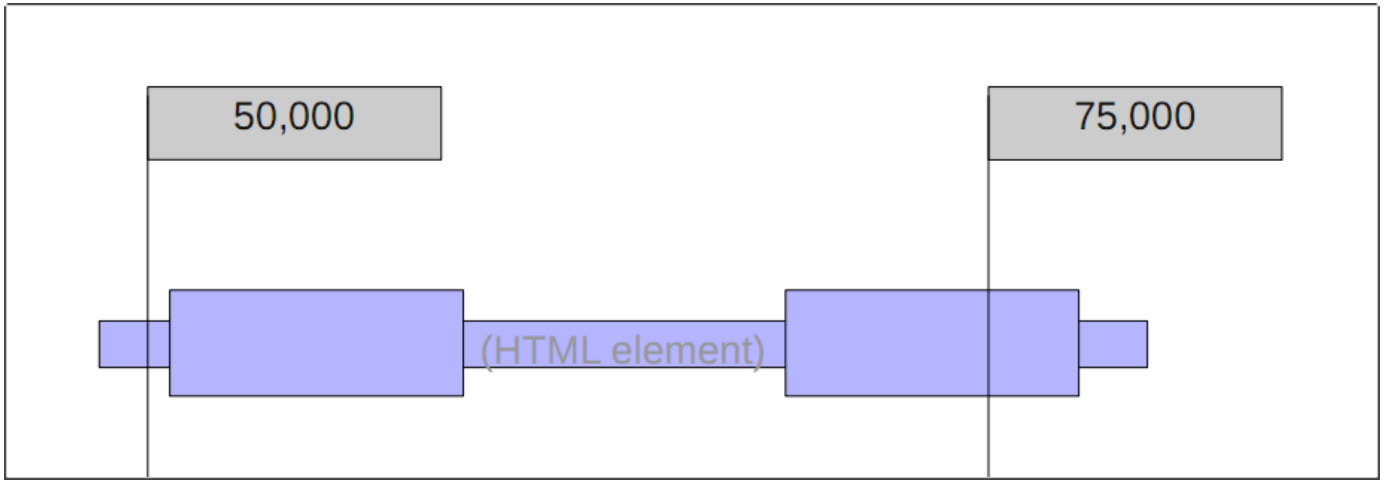

intron.ccam.uchc.edu/JBrowse/Dmel/
pseudogene
Gene Span
mRNA
Cytological band
Non coding RNA
Embryo_0-2h
Embryo_2-4h
Embryo_4-6h
Embryo_6-8h
Embryo_8-10h
Embryo_10-12h
Embryo_12-14h
adult_female_1d
adult_male_1d
adult_female_5d
adult_male_5d

Types of Tracks



Feature Tracks

- rectangles (`<div>`s) with various fills and dimensions to represent the features
- can do a surprising amount, but still limited



Feature Tracks: Density Plots

Zoomed out

Zoomed in

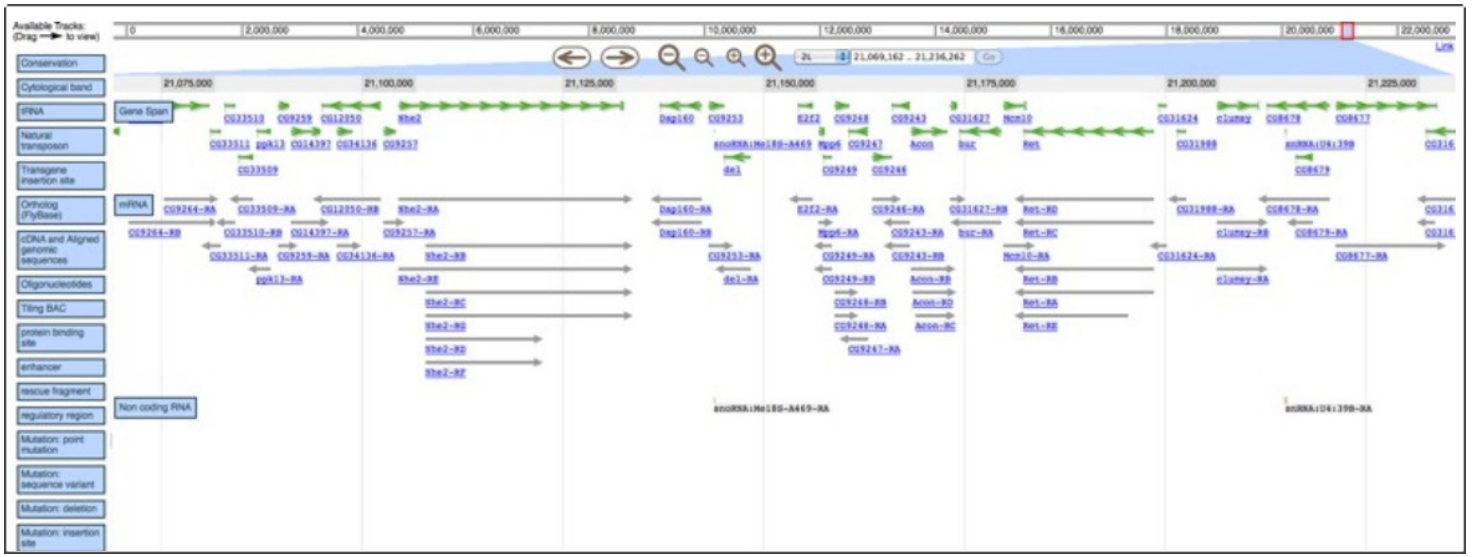
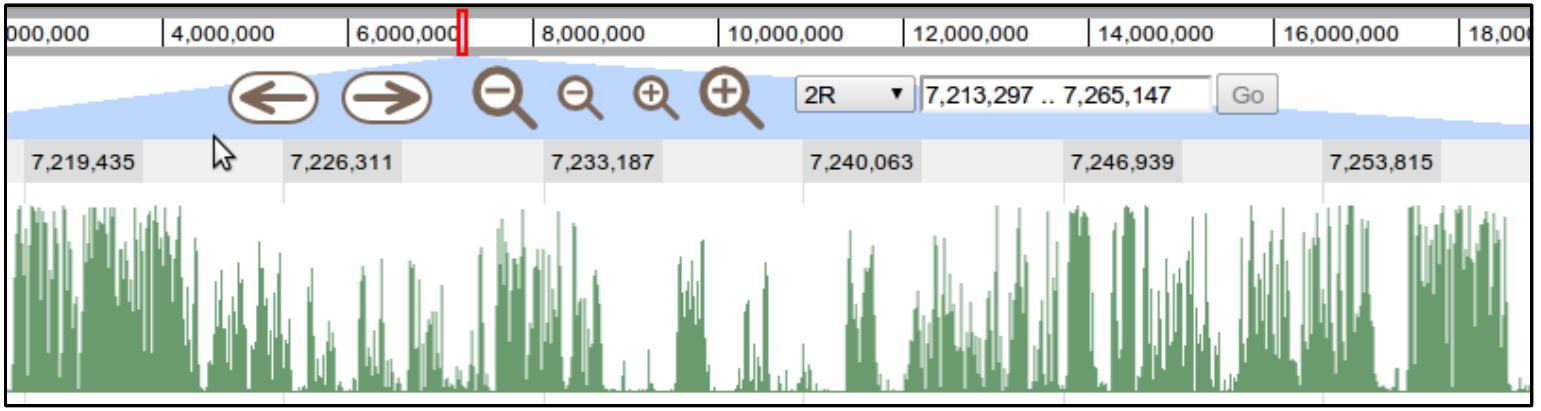


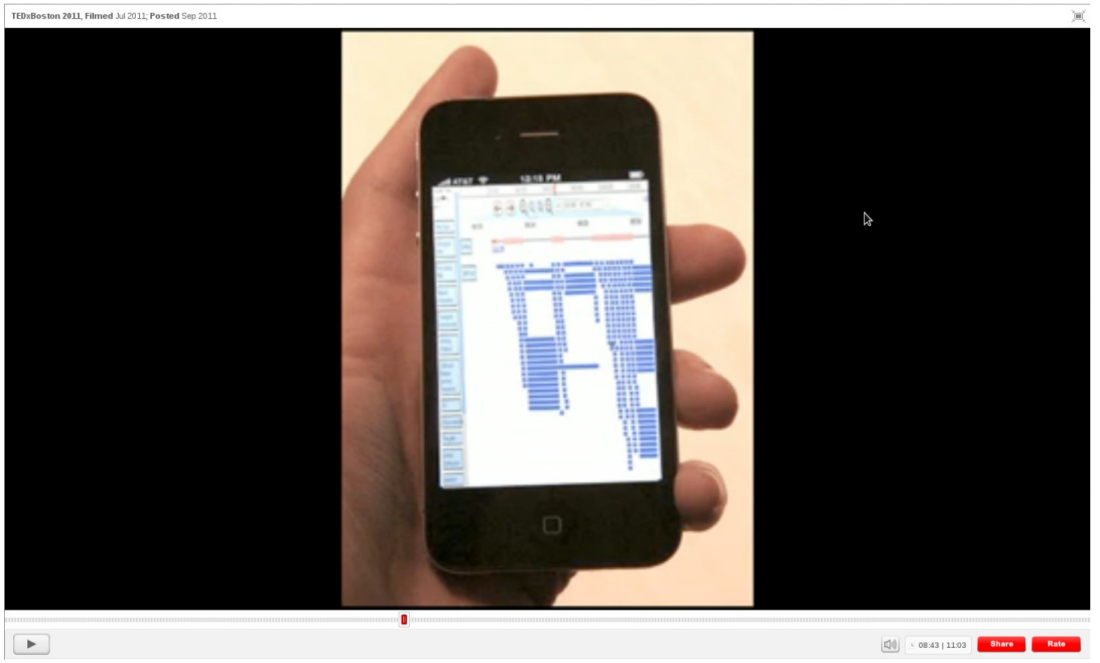
Image Tracks

- any pre-generated images that cover the genome
- quantitative data (wiggle/BigWig)
  - wig-to-json.pl
- RNA base-pairing



± Particular Strengths

- **very** fast and smooth!
- easy things are easy!
- works nicely on an iPad/iPhone



JBrowse on an iPhone
video frame from TEDx Boston talk by Richard Resnick

± Particular Weaknesses

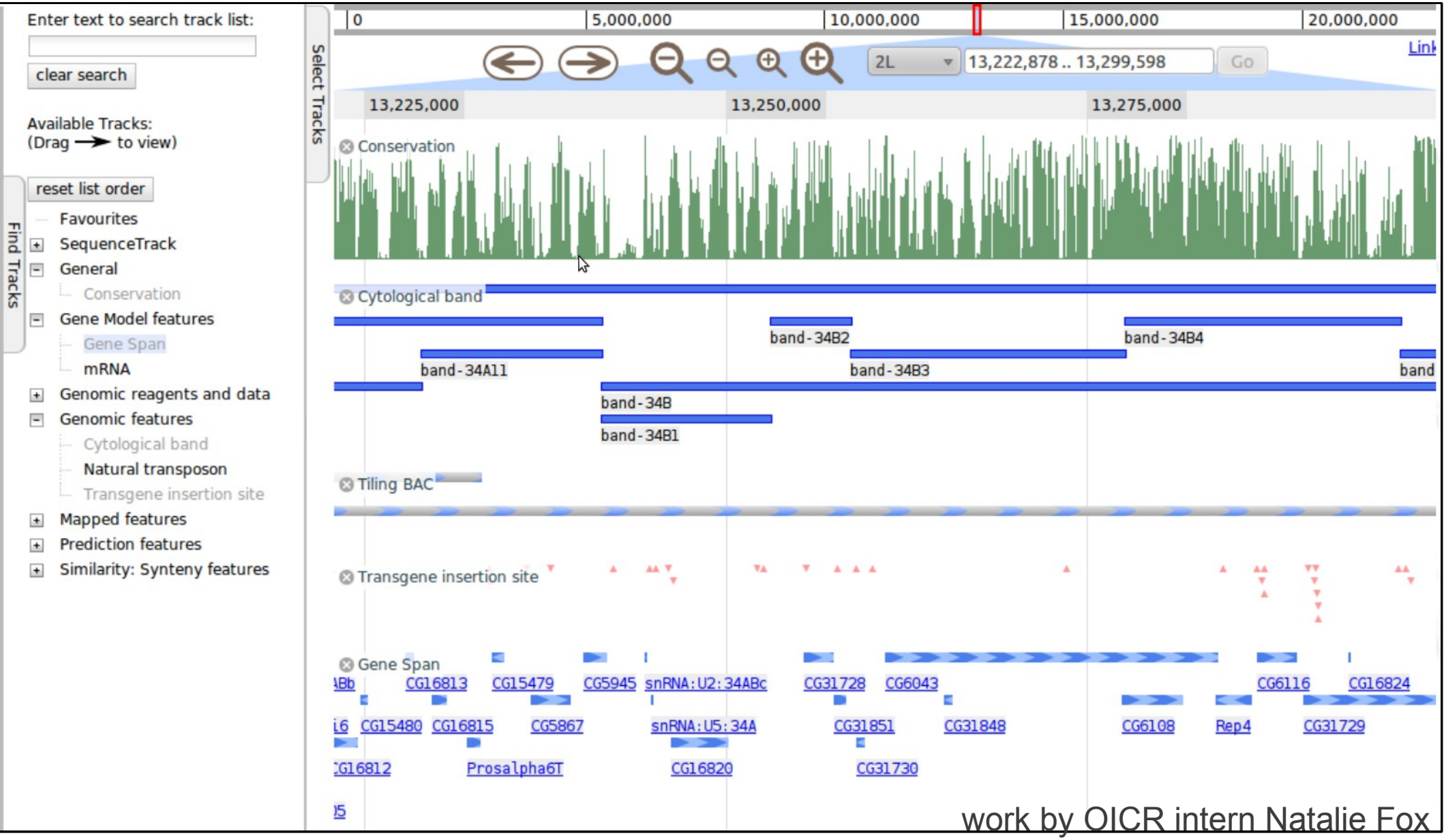
- no user-uploaded data
- no data export
- slow for big numbers of reference seqs (e.g. 5,000 annotated contigs)
- few glyph options, feature tracks are limited by the facts of `<div>`

These will all be improved in the 2.x series.

Coming Soon

- new track selection
- overview images
- context menus
- track customization
- drag to zoom!

Coming: Hierarchical Track Sel.

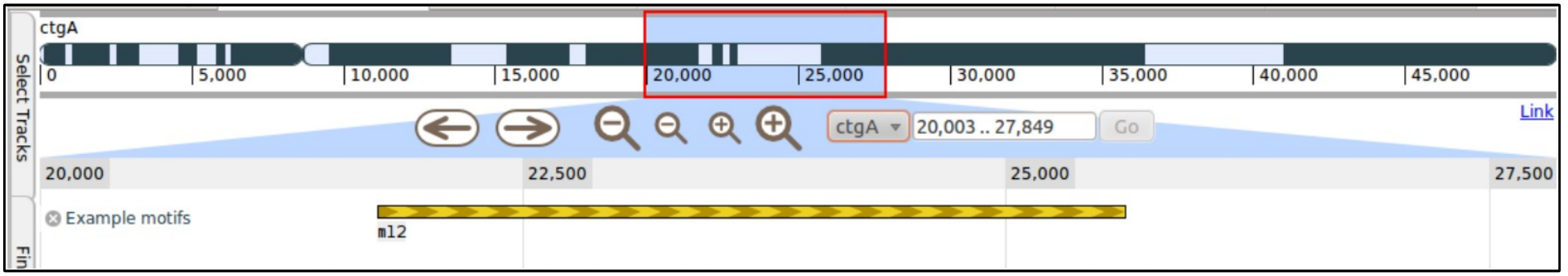


work by OICR intern Natalie Fox

Coming: Faceted Track Selection

work by OICR intern Natalie Fox

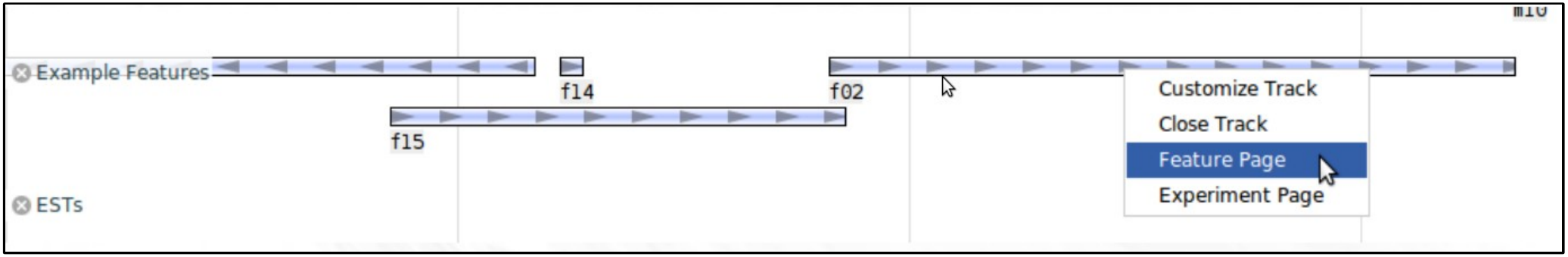
Coming: Overview Images



Can have a karyotype, or any other image.

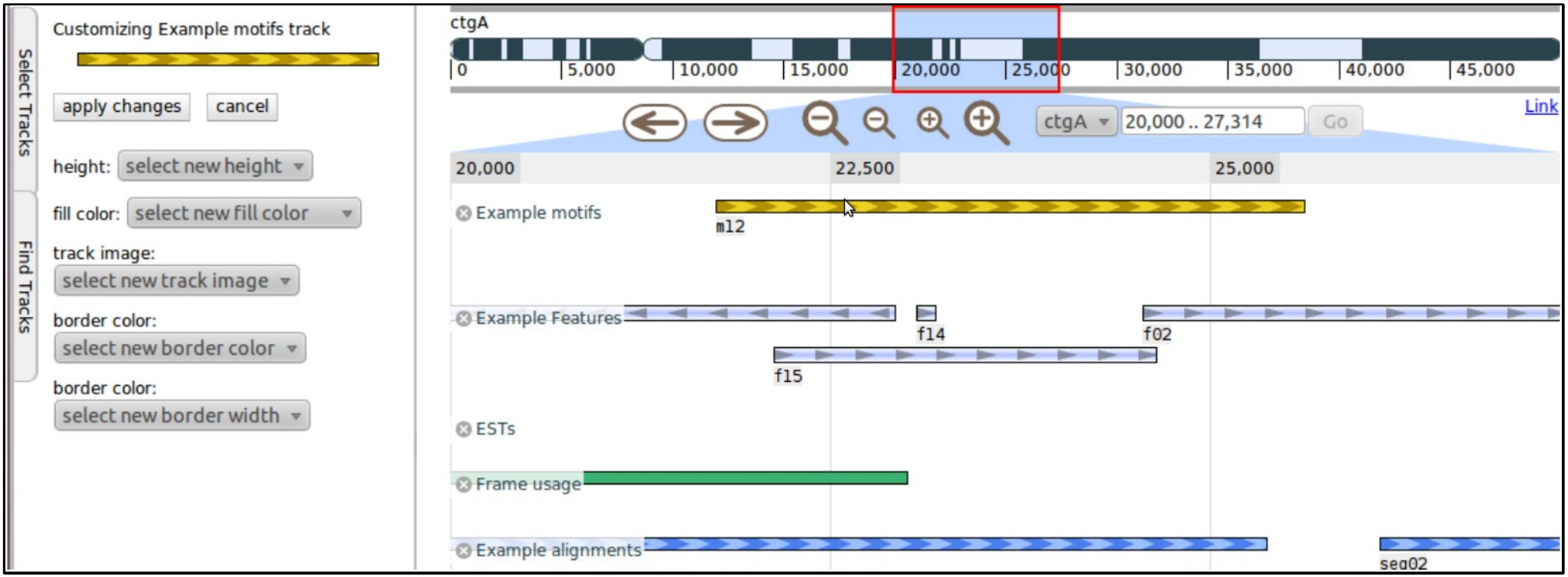
work by OICR intern Natalie Fox

Coming: Context Menus



work by OICR intern Natalie Fox

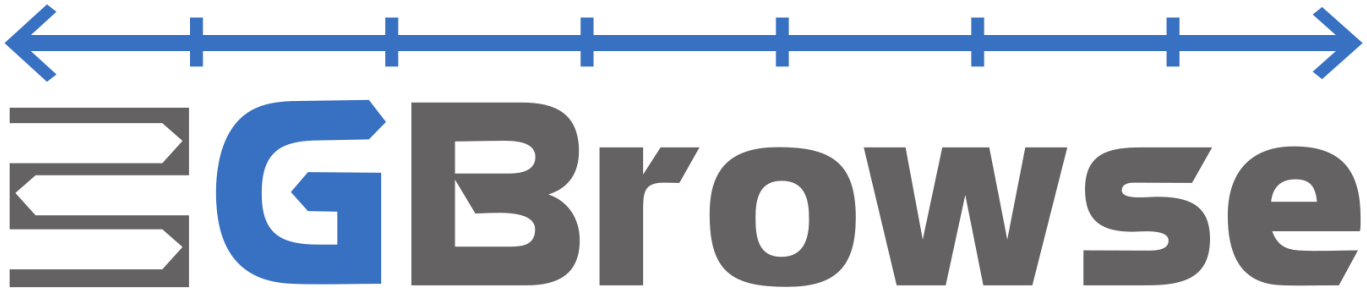
Coming: Track Customization



work by OICR intern Natalie Fox

Migrating: GBrowse to JBrowse

- JBrowse does not yet have all the features of GBrowse
- Configuration and data formats are different
- Adam Wright, intern at OICR, has worked on automated migration tools to help
- Another thing I need to merge into master ;-)



WebApollo

- next generation of the popular Apollo annotation editor
- genome editing in your browser
- based on JBrowse

WebApollo

- clients receive annotation updates **in real time** (like Google Docs)
- saves edits to a central database
- has produced several extensions that will be coming to core JBrowse
  - highlighting and selecting features
  - HTML5 Canvas tracks for wiggle data (opening the door for many more canvas-based techniques)

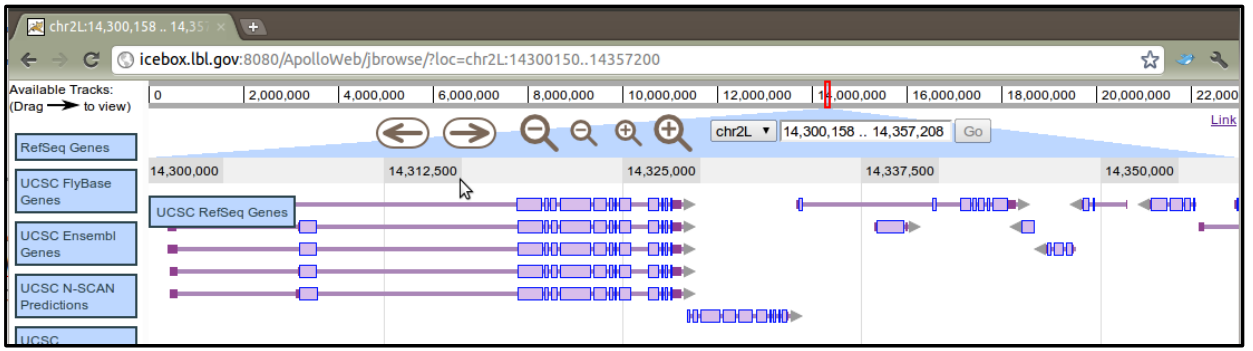
Apollo

¡WebApollo Live Demo!

Step 1

http://goo.gl/SHByG

User name: **write_1**
Password: **write_1**

Step 2

http://goo.gl/mSh3w