

BIOINFORMATICS ANALYSIS TOOLS FOR NGS DATA

ANNOTATIONS, VISUALISATION: INTEGRATIVE GENOME BROWSER

Philipp Rescheneder, Moritz Smolka

April 27, 2016

Center for Integrative Bioinformatics Vienna
Max F. Perutz Laboratories



CIBIV

Center for Integrative Bioinformatics Vienna

COMMONLY USED RESOURCES

Primary databases

- **NCBI GenBank:** an annotated collection of **all** publicly available DNA sequences.
- **NCBI RefSeq:** A collection of curated, **non-redundant** genomic DNA, transcript (RNA), and protein sequences produced by NCBI.
- **ENA: European Nucleotide Archive:** a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation

Commonly used meta databases

- **UCSC Genome Bioinformatics:** contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to the ENCODE and Neandertal projects
- **Ensembl Project** by the European Bioinformatics Institute (**EBI**), European Molecular Biology Laboratory (**EMBL**), and the Wellcome Trust Sanger Institute (**WTSI**).

- FlyBase (*Drosophila melanogaster*)
- Wormbase (*Caenorhabditis elegans*)
- SGD (*Saccharomyces Cerevisiae*)
- TAIR (*Arabidopsis thaliana*)
- Colibri (*Escherichia Coli*)

And many others – often you'll have to do your research to find the most appropriate resource

Note that each may provide a different user interface, data release methodology, data release update policy etc.

Surprisingly unregulated – and often non-transparent process.

SACCHAROMYCES: GENOME DATABASE

SGD *Saccharomyces* GENOME DATABASE

About Blog **Download** Site Map Help

Search

YeastMine: Batch Analysis or Advanced Search

SGD Analyze Sequence Function Literature Community

Natural isolates of *S. cerevisiae* form complex mats on low-agar media.
Image courtesy of Elyse A. Hope and Dr. Maitreya J. Dunham, University of Washington

About SGD

The Saccharomyces Genome Database (SGD) is a biological database for the yeast *Saccharomyces cerevisiae*. It provides a comprehensive set of biological data and tools for researchers studying this model organism.

Upcoming

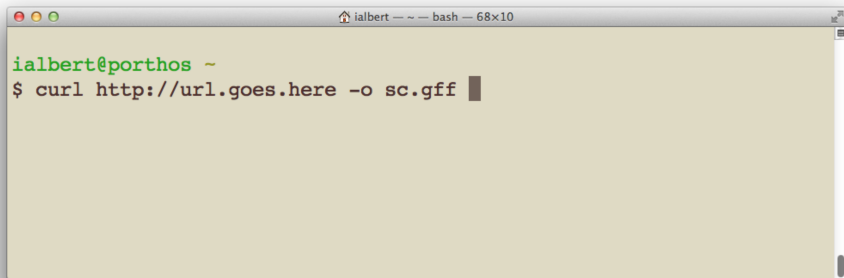
New & Noteworthy

New Sequence, Chromosome, and Contig pages
08/25/2014
New Sequence pages are now available in SGD for virtually every yeast gene (e.g., HMRA1 Sequence page), and include genomic sequence annotations for the Reference Strain S288C, as well as several Alternative Reference Genomes from strains such as CEN.PK, RM11-1a, Sigma1278b, and W303 (more Alternative References coming soon). Each page includes an Overview section containing descriptive information, maps depicting genomic context in Reference Strain S288C (as shown below) and Alternative Reference strains, as well... [Read...](#)

Special Delivery for Cytotoxic Proteins
08/21/2014
Say you want to send a letter to your friend on the other side of the country. First off

clone.tab	Yeast clones from Washington, St. Louis and the ATCC
dbxref.tab	Mappings between SGD and external identifiers
deleted_merged_features.tab	List of deleted or merged chromosomal features
genetic_map.tab	Genetic two-point data submitted to SGD
saccharomyces_cerevisiae.gff	Genomic sequence features, coordinates and annotations. Used to load GBrowse.
scerevisiae_2-micron.gff	2-micron plasmid sequence features, coordinates and annotations. Used to load GBrowse.

DOWNLOADING ON THE COMMAND LINE

A terminal window with a title bar that reads "ialbert ~ -- bash -- 68x10". The terminal content shows the prompt "ialbert@porthos ~" followed by the command "\$ curl http://url.goes.here -o sc.gff" with a cursor at the end of the line.

```
ialbert@porthos ~  
$ curl http://url.goes.here -o sc.gff
```

You can use the a browser or the command line:

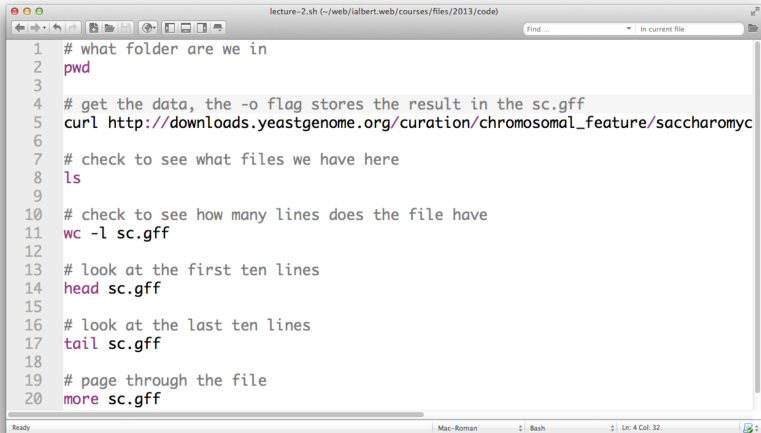
(The URL is on the course webpage)

http://downloads.yeastgenome.org/curation/chromosomal_feature/saccharomyces_cerevisiae.gff

WHAT IS THIS?

- Due to a historical limitation (20 years or so ago) and only on Windows → files ended up having a three character extension → **.txt**, **.exe** etc.
- This limitation also turned out to be a blessing and it stood the test of time. Makes it easy to see the file type.
- Note: the file extension can be incorrect → mind-bogglingly confusing errors may arise then.

UNIX COMMAND RECAP



```
lecture-2.sh (~/.web/ialbert.web/courses/files/2013/code)
1 # what folder are we in
2 pwd
3
4 # get the data, the -o flag stores the result in the sc.gff
5 curl http://downloads.yeastgenome.org/curation/chromosomal_feature/saccharomyc
6
7 # check to see what files we have here
8 ls
9
10 # check to see how many lines does the file have
11 wc -l sc.gff
12
13 # look at the first ten lines
14 head sc.gff
15
16 # look at the last ten lines
17 tail sc.gff
18
19 # page through the file
20 more sc.gff
```

Ready Mac-Roman Bash Ln: 4 Col: 32

- Many common bioinformatics data formats are column based and tab-separated
- Looks like the first format we have to deal with will be the

GFF3 – Generic Feature Format


(search for GFF3 to see the specification for version 3)

<http://www.sequenceontology.org/gff3.shtml>

Search for GFF3 → <http://www.sequenceontology.org/gff3.shtml>

Tab separated with 9 columns. Missing attributes may be replaced with a dot → .

1. **Seqid** (usually chromosome, **reference point!**)
2. **Source** (where is the data coming from)
3. **Type** (usually a term from the sequence ontology)
4. **Start** (interval start relative to the **seqid**)
5. **End** (interval end relative to the **seqid**)
6. **Score** (the score of the feature, a floating point number)
7. **Strand** (+/./.)
8. **Phase** (used to indicate reading frame for coding sequences)
9. **Attributes** (semicolon separated attributes → **Name=ABC;ID=1**)



These positions
may change on new
data releases

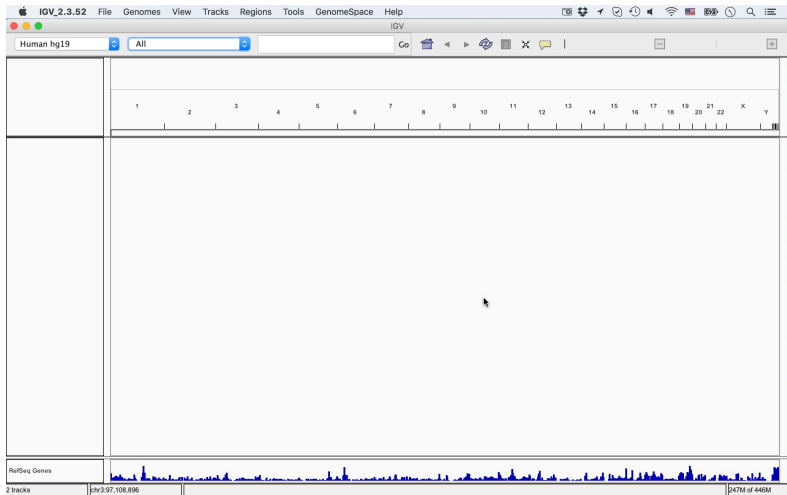
Example attribute specification: **name=REB1 ; id=YP33546**

GFF FORMAT

```
19 #
20 chrI   SGD chromosome 1 230218 . . . ID=chrI;dbxref=NCBI:NC
21 chrI   SGD telomeric_repeat 1 62 . - . ID=TEL01L-TR;Name=
22 chrI   SGD telomere 1 801 . - . ID=TEL01L;Name=TEL01L;Note
23 chrI   SGD X_element_combinatorial_repeat 63 336 . - . ID=TEL
24 chrI   SGD gene 335 649 . + . ID=YAL069W;Name=YAL069W;Ontolc
25 chrI   SGD CDS 335 649 . + 0 Parent=YAL069W_mRNA;Name=YAL069W_C
26 chrI   SGD mRNA 335 649 . + . ID=YAL069W_mRNA;Name=YAL069W_r
27 chrI   SGD X_element 337 801 . - . ID=TEL01L-XC;Name=TEL01L-X
28 chrI   SGD nucleotide_match 753 763 . - . ID=TEL01L-XC_nucle
29 chrI   SGD binding_site 532 544 . - . ID=TEL01L-XC_binding_s
30 chrI   SGD gene 538 792 . + . ID=YAL068W-A;Name=YAL068W-A;Or
31 chrI   SGD CDS 538 792 . + 0 Parent=YAL068W-A_mRNA;Name=YAL068W
32 chrI   SGD mRNA 538 792 . + . ID=YAL068W-A_mRNA;Name=YAL068W
33 chrI   SGD ARS 650 1791 . . . ID=ARS102;Name=ARS102;Alias=AR
34 chrI   SGD gene 1807 2169 . - . ID=YAL068C;Name=YAL068
35 chrI   SGD CDS 1807 2169 . - 0 Parent=YAL068C_mRNA;Name=Y
```

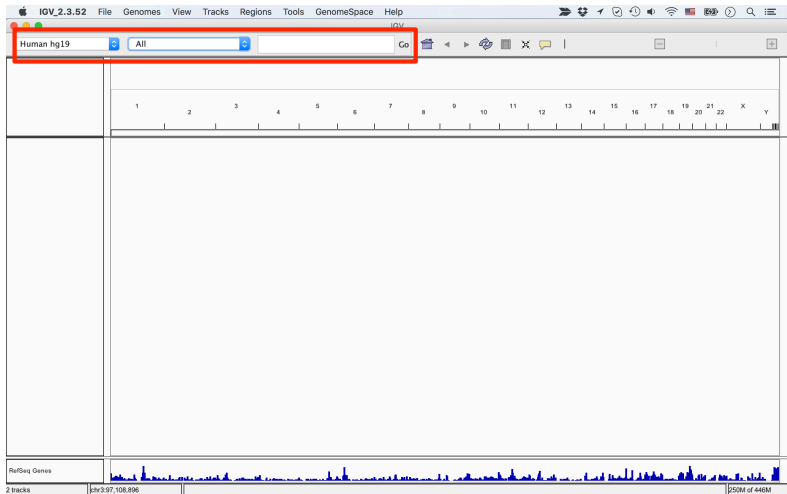
- ▶ Genome browser: one of the most important tools for data analysis
- ▶ "Always look at your data", true for statistics as well as for NGS data analysis
- ▶ Many genome browsers available: UCSC (web), IGB, Tablet, Artemis, MochiView (ChIP), Chip/SeqMonk (DNA methylation)
- ▶ Integrative Genomics Viewer
 - ▶ Developed by the Broad Institute
 - ▶ Awesome and extremely terrible at the same time
 - ▶ Works well and is fast enough
 - ▶ Data formats: BED, BEDGRAPH, VCF, GTF, BAM, WIG, BIGWIG, FASTA, etc.
 - ▶ Usability not perfect

IGV: BASICS



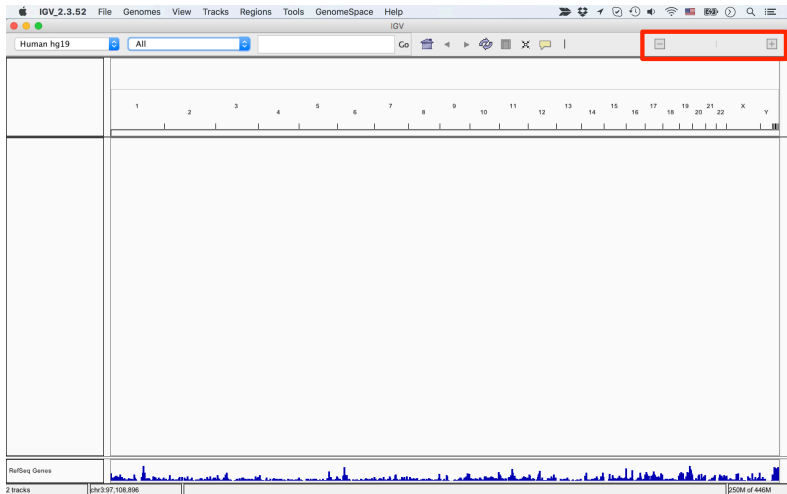
► Overview

IGV: BASICS



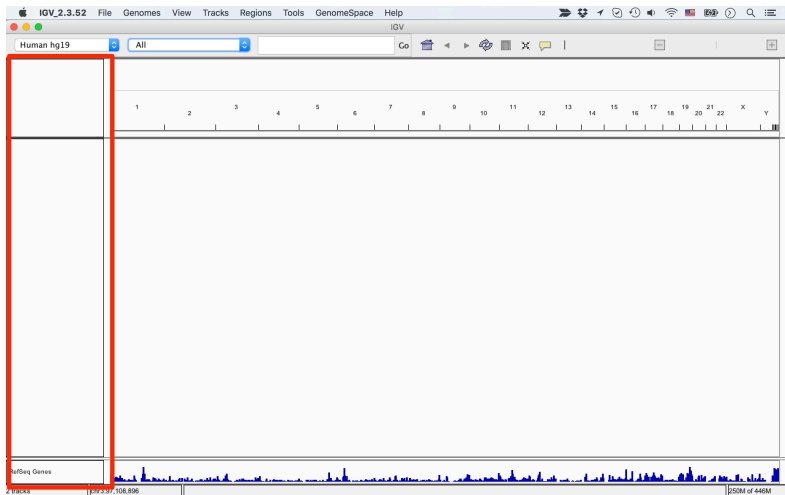
- ▶ Navigation area

IGV: BASICS



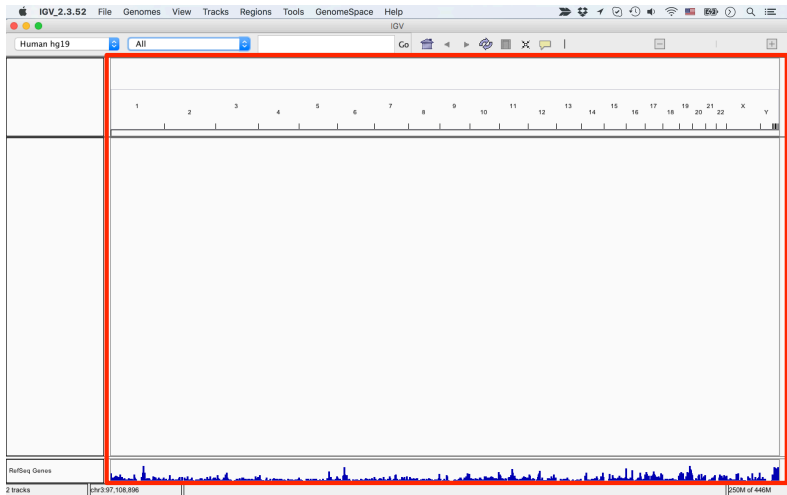
► Zoom

IGV: BASICS



► Name panel

IGV: BASICS



► Main area

IGV: GENOMES

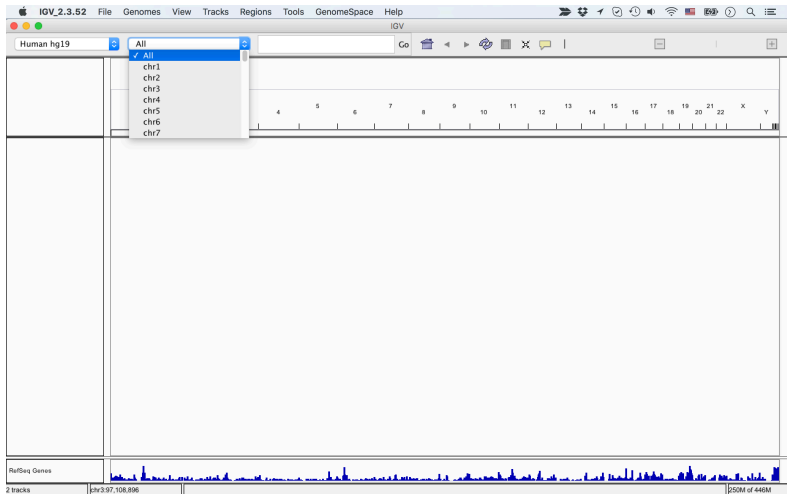
The screenshot shows the IGV application window. The title bar reads "IGV_2.3.52" and includes standard macOS window controls. The menu bar contains "File", "Genomes", "View", "Tracks", "Regions", "Tools", "GenomeSpace", and "Help". The main interface has a search bar with "Human hg19" selected and a "Go" button. A dropdown menu is open, listing the following genomes:

- Human hg19
- Mouse (mm9)
- A. thaliana (TAIR 10)
- Human hg13
- Mouse (mm10)
- B31_Schutzer_reference-goodnames
- Human (hg38)
- Ecoli k12 chr (NC000913.3)
- dmel_r5.57

The main display area shows a chromosome map with chromosomes 2 through 22, X, and Y. At the bottom, there is a "RefSeq Genes" track showing a blue signal across the chromosomes. The status bar at the bottom left indicates "2 tracks" and "chr3:97,106,896". The status bar at the bottom right indicates "250M of 446M".

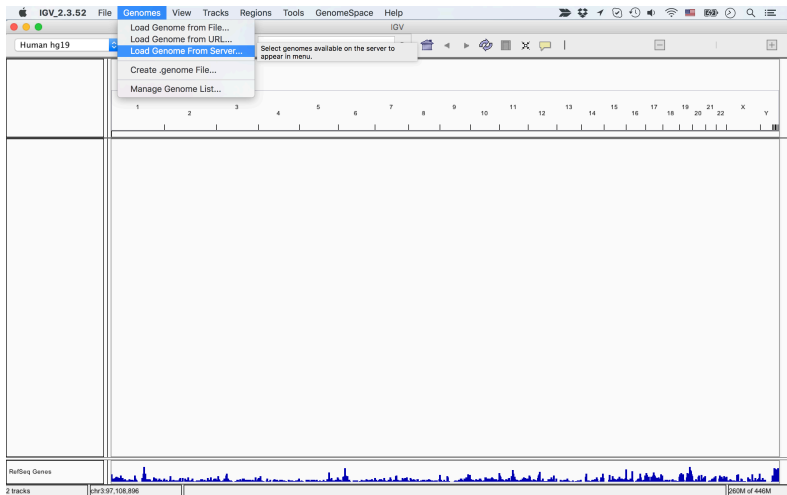
- List of available genomes

IGV: GENOMES



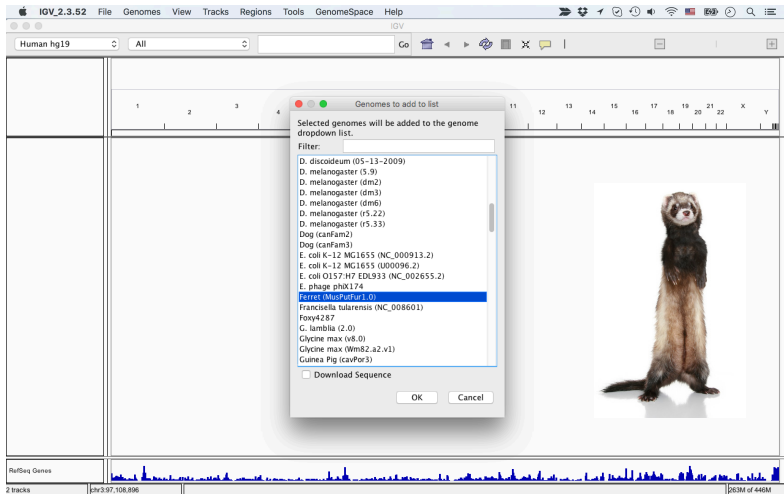
- ▶ Jump to chromosomes

IGV: GENOMES



- ▶ Load genomes from Server

IGV: GENOMES



The screenshot shows the IGV (Integrative Genomics Viewer) interface. The main window displays a genomic track for Human hg19. A dialog box titled "Genomes to add to list" is open, showing a list of genomes to be added to the genome dropdown list. The list includes:

- D. discoideum (05-13-2009)
- D. melanogaster (5.9)
- D. melanogaster (dm2)
- D. melanogaster (dm3)
- D. melanogaster (dm6)
- D. melanogaster (r5.22)
- D. melanogaster (r5.33)
- Dog (canFam2)
- Dog (canFam3)
- E. coli K-12 MG1655 (NC_000913.2)
- E. coli K-12 MG1655 (U00096.2)
- E. coli O157:H7 EDL933 (NC_002655.2)
- E. phage phiX174
- Ferret (MusPurFur 1.0)**
- Francisella tularensis (NC_008601)
- Foxy4287
- G. lamblia (2.0)
- Glycine max (v4.0)
- Glycine max (Wm82.a2.v1)
- Guinea Pig (cavPor3)

There is a checkbox for "Download Sequence" which is currently unchecked. The dialog has "OK" and "Cancel" buttons. In the background, a photograph of a ferret is visible on the right side of the IGV window.

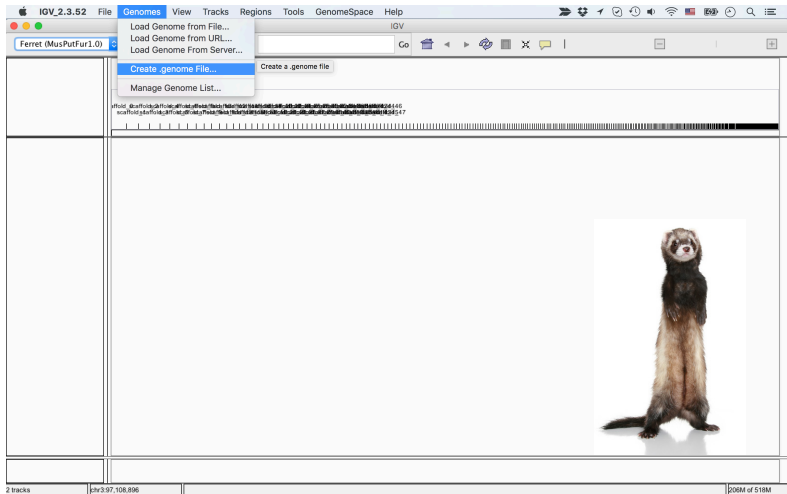
► Ferret it is

IGV: GENOMES

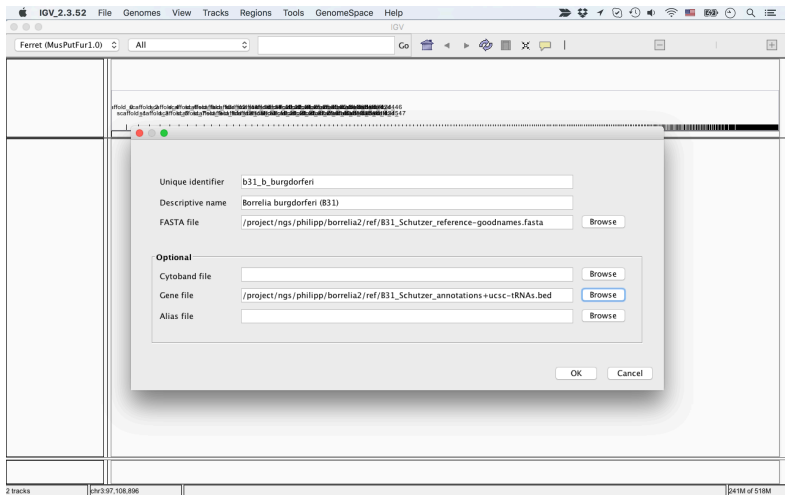
The screenshot shows the IGV application window titled "IGV_2.3.52". The menu bar includes "File", "Genomes", "View", "Tracks", "Regions", "Tools", "GenomeSpace", and "Help". The main window displays the "Ferret (MusPutFur1.0)" genome. A dropdown menu is open, showing a list of scaffolds: "All", "scaffold_0", "scaffold_1", "scaffold_2", "scaffold_3", "scaffold_4", "scaffold_5", and "scaffold_6". The "All" option is selected. The main display area shows a genomic track with a black bar representing a region of interest. On the right side of the viewer, there is a photograph of a ferret standing upright. The status bar at the bottom indicates "2 tracks", "chr3:97,106,896", and "204M of 518M".

► Ferret it is

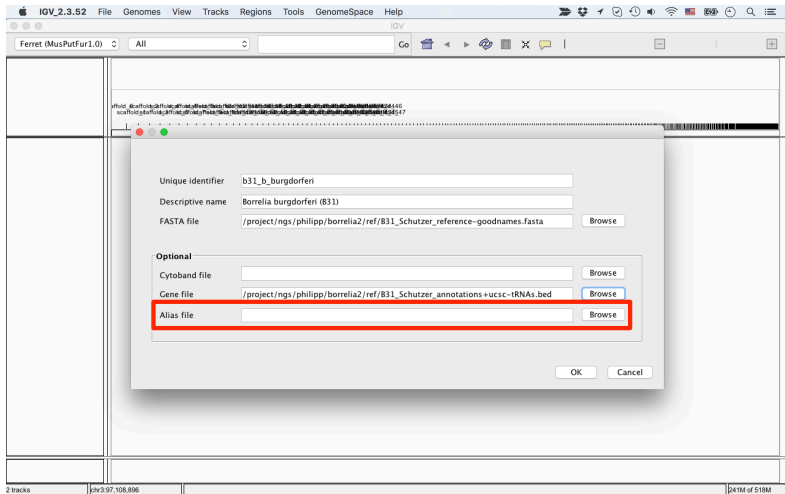
IGV: GENOMES



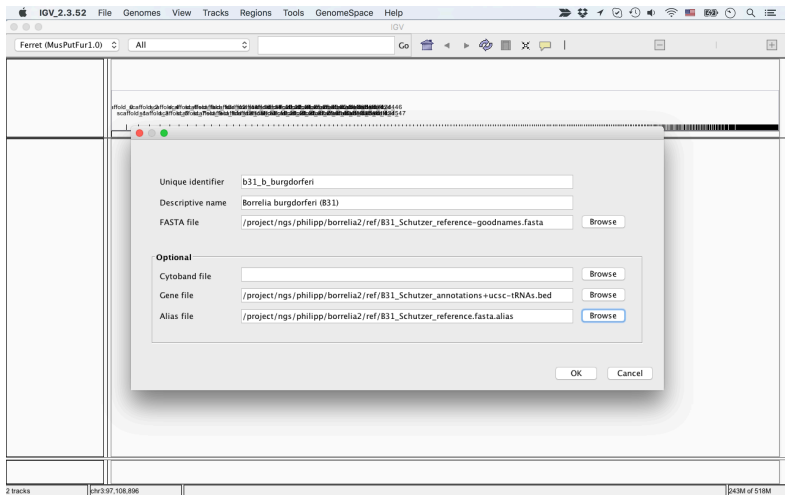
- ▶ We don't have Ferret data, so we need something else: loading a custom genome



- ▶ Loading the Borrelia Burgdorferi genome

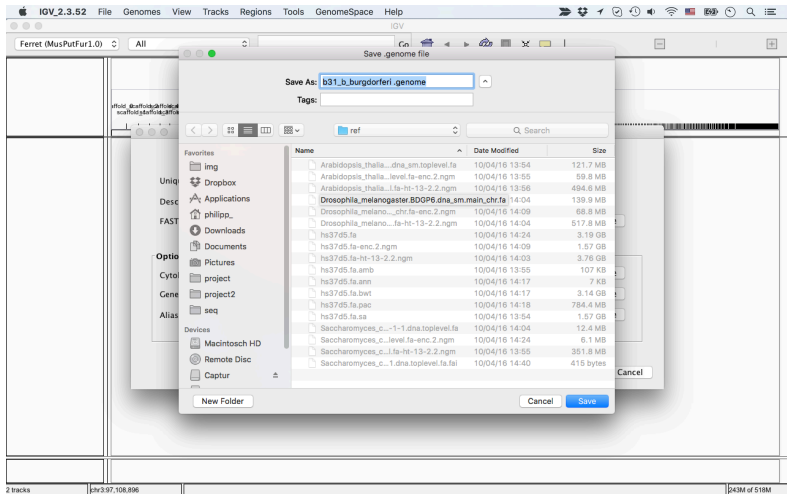


- **Important:** the alias file (I'll come back to that later)



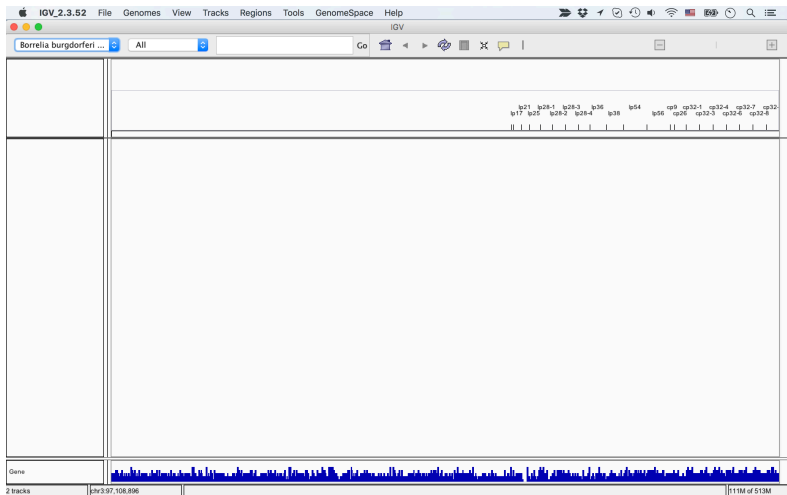
- **Important:** the alias file (I'll come back to that later)

IGV: GENOMES



- Save .genome file (contains all the information)

IGV: GENOMES



- ▶ Custom genome is now part of the genome list

- ▶ The single most problem with IGV I have encountered so far
- ▶ What happens? Sometimes when loading a BAM/BED/VCF/... file, the file appears to be empty.
- ▶ Problem: **different naming of chromosomes/scaffolds**
- ▶ Example:

IGV: ALIAS FILES

- ▶ The single most problem with IGV I have encountered so far
- ▶ What happens? Sometimes when loading a BAM/BED/VCF/... file, the file appears to be empty.
- ▶ Problem: **different naming of chromosomes/scaffolds**
- ▶ Example:

```
oort:~ philipp$ head /project/ngs/philipp/borrelia2/ref/B31_Schutzer_reference-goodnames.f
asta
>chr_Borrelia_burgdorferi_B31, complete genome.
TAAATATAATTTAAGTAGTATAAAAAAATTAATCAAATTAATAAGTTTTAAAAAAGCTG
TTTGATATAATATAATATATATATAATTAAGCACTACTATGATACTAATGAAGT
ATAGTGCTATTTTAAATATGTAGCGTTAATTTATTTGTTTTCAAAAATAAATACTA
CTTCTCGATGGGAATCCCTAAGAAGATTTAATTAAAAAAAAATAAAAAAGGCATAA
TTTACCAATATACATAAATCTATCTTTTACAATGAAAAATATAAATACATTGCCCTTA
TCGGAAATTTGACACTCTATAATGAATGGATTAATAACAAATTTAGCCCCATAAATTTT
TTACTATCCCAACAATAMAGATTTATTTCAAATCTATTTCAATTTAGCTTTCCTA
TTTTACATACCAAGTATTCAAATTTTAACTGATACACTGCTATAAAATTTTTATTGGAA
CCCAATCGATTTAACTCTGAGCACTACTATTTTACAGGAAAAACAATCATGCATTC
oort:~ philipp$
```

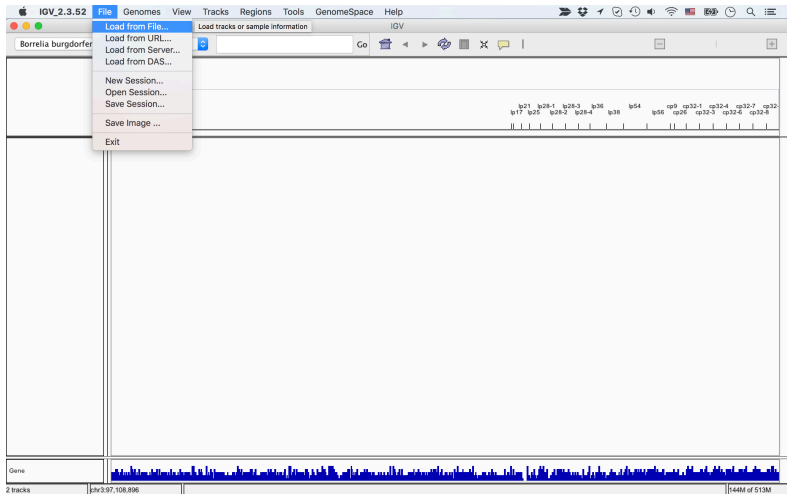

IGV: ALIAS FILES

- ▶ The single most problem with IGV I have encountered so far
- ▶ What happens? Sometimes when loading a BAM/BED/VCF/... file, the file appears to be empty.
- ▶ Problem: **different naming of chromosomes/scaffolds**
- ▶ Example:

```
t.bash
oort:- philipp_$ head /project/ngs/philipp/borrelia2/ref/B31_Schutzer_reference-goodnames.f
asta
s>chr Borrelia burgdorferi B31, complete genome.
TAAATATAATTTAAGTATAAAAAAATAAATCAAAATTAATAAGTTTAAAAAACTG
TTTGATAATAATATATATATATATAATTAAGCAACTACTATGACTACATGAAGT
ATAGTGCATTTTAAATATGAGCGTTAAATTTTGTTCAAAAATAAATAACTA
CTTCTCGATGGGAATCCCTAAGAAGATTAAATTAATAAAAAAATAAAAAAGGCATA
TTTACCAATAATACAAATCTACTCTTTACAATGAAAAATATAAATACATGCCCTTA
TCGGAATATTGACACTTATAAGTAATGGATTGAATACAAATTAAGCCCATAAATTTT
TTACTATCCCAAAATAMAGATTTATTTCAAACTACTTTTCATTTAGCTTCCACTA
TTTACATACC
CCCAATCGA
oort:- philipp_$ samtools view /project/ngs/philipp/borrelia2/reads-ngm/runise50_second/203
90_ACTTGA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam | cut -f 1,2,3 | head
HWI-ST1253F_0165:6:2312:17614:17340#20390_ACTTGA 16 AE000783.1
HWI-ST1253F_0165:6:1213:19639:3553#20390_ACTTGA 16 AE000783.1
HWI-ST1253F_0165:6:1205:9633:92581#20390_ACTTGA 16 AE000783.1
HWI-ST1253F_0165:6:1108:4538:84682#20390_ACTTGA 0 AE000783.1
HWI-ST1253F_0165:6:2184:16451:18209#20390_ACTTGA 16 AE000783.1
HWI-ST1253F_0165:6:1312:8912:35418#20390_ACTTGA 0 AE000783.1
HWI-ST1253F_0165:6:2305:11877:88207#20390_ACTTGA 0 AE000783.1
HWI-ST1253F_0165:6:1105:6746:89913#20390_ACTTGA 0 AE000783.1
HWI-ST1253F_0165:6:1108:7441:85549#20390_ACTTGA 0 AE000783.1
HWI-ST1253F_0165:6:1202:7530:80880#20390_ACTTGA 0 AE000783.1
oort:- philipp_$
```

```
B31_Schutzer_reference.fasta.alias
AE000783.1 chr
AE000790.2 lp54
AE001584.1 lp56
AE001583.1 lp5
AE000787.1 lp38
AE000793.2 lp17
AE000788.1 lp36
AE001581.1 cp32-9
AE000785.1 lp25
AE001580.1 cp32-8
AE001579.1 cp32-7
AE001578.1 cp32-6
AE001582.2 lp21
AE001577.1 cp32-4
AE000791.1 cp9
AE001576.1 cp32-3
AE000786.1 lp28-2
AE000794.2 lp28-1
AE000792.1 cp26
AE001575.1 cp32-1
AE000789.1 lp28-4
AE000784.1 lp28-3
```


IGV: LOADING BAM FILES



- ▶ BAM file has to be sorted and indexed

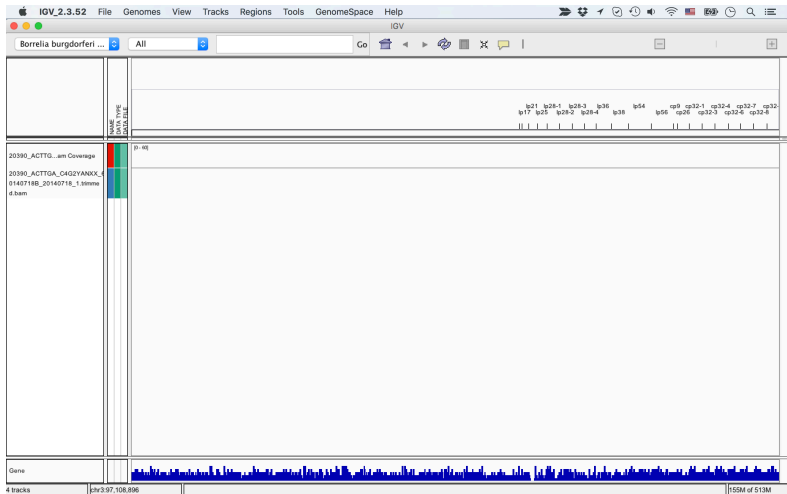
IGV: LOADING BAM FILES

The screenshot shows the IGV application window with a file browser overlay. The browser displays a list of files in a table with columns for Name, Date Modified, and Size. The files are organized by name, showing various BAM files and their associated codoc and wig files. The selected file is 20390_ACTTGA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam, which is 562.4 MB and was modified on 12/01/15 at 14:12.

Name	Date Modified	Size
20390_ACTTGA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam	12/01/15 14:12	562.4 MB
20390_ACTTGA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.bai	27/01/15 12:10	6 KB
20390_ACTTGA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.MINUS.codoc	27/03/15 14:05	987 KB
20390_ACTTGA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.MINUS.codoc-norm.wig	27/03/15 14:05	18.7 MB
20390_ACTTGA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.PLUS.codoc	27/03/15 14:03	993 KB
20390_ACTTGA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.PLUS.codoc-norm.wig	27/03/15 14:05	19 MB
20391_ATCACG_C4G2YANXX_6_201407188_20140718_1.trimmed.bam	08/09/14 17:31	563.9 MB
20391_ATCACG_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.bai	27/01/15 12:11	6 KB
20391_ATCACG_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.MINUS.codoc	27/03/15 14:43	994 KB
20391_ATCACG_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.MINUS.codoc-norm.wig	27/03/15 14:44	19 MB
20391_ATCACG_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.PLUS.codoc	27/03/15 14:42	999 KB
20391_ATCACG_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.PLUS.codoc-norm.wig	27/03/15 14:44	19.3 MB
20392_CGATGT_C4G2YANXX_6_201407188_20140718_1.trimmed.bam	08/09/14 17:51	550.5 MB
20392_CGATGT_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.bai	27/01/15 12:11	6 KB
20392_CGATGT_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.MINUS.codoc	27/03/15 14:41	1.5 MB
20392_CGATGT_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.MINUS.codoc-norm.wig	27/03/15 14:41	21.7 MB
20392_CGATGT_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.PLUS.codoc	27/03/15 14:40	1.5 MB
20392_CGATGT_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.PLUS.codoc-norm.wig	27/03/15 14:41	21.9 MB
20393_TTAGGC_C4G2YANXX_6_201407188_20140718_1.trimmed.bam	08/09/14 18:10	475.3 MB
20393_TTAGGC_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.bai	27/01/15 12:11	6 KB
20393_TTAGGC_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.MINUS.codoc	27/03/15 14:39	984 KB
20393_TTAGGC_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.MINUS.codoc-norm.wig	27/03/15 14:39	18.2 MB
20393_TTAGGC_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.PLUS.codoc	27/03/15 14:38	994 KB
20393_TTAGGC_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.PLUS.codoc-norm.wig	27/03/15 14:39	18.6 MB
20394_TGACCA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam	08/09/14 18:43	460.6 MB
20394_TGACCA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.bai	27/01/15 12:11	5 KB
20394_TGACCA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.MINUS.codoc	27/03/15 14:35	940 KB
20394_TGACCA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.MINUS.codoc-norm.wig	27/03/15 14:35	19.4 MB
20394_TGACCA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.PLUS.codoc	27/03/15 14:34	948 KB
20394_TGACCA_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.PLUS.codoc-norm.wig	27/03/15 14:35	19.8 MB
20395_CAGATC_C4G2YANXX_6_201407188_20140718_1.trimmed.bam	08/09/14 17:12	451.2 MB
20395_CAGATC_C4G2YANXX_6_201407188_20140718_1.trimmed.bam.bai	27/01/15 12:11	6 KB

- ▶ BAM file has to be sorted and indexed

IGV: LOADING BAM FILES



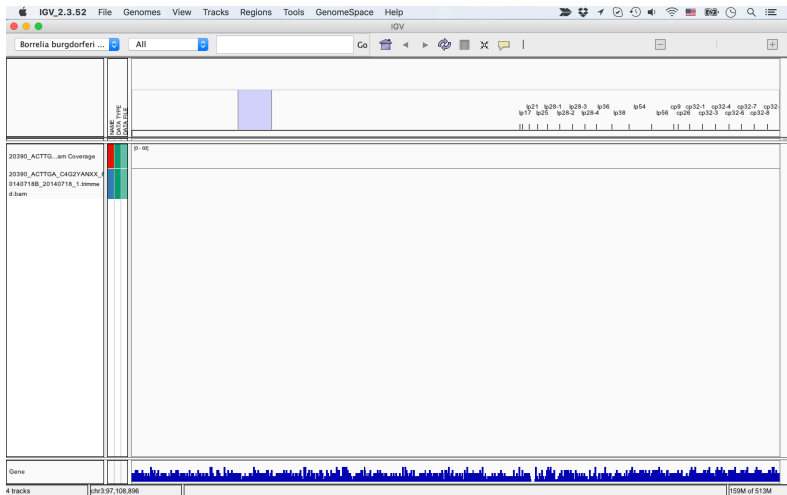
- ▶ Nothing visible when looking at the full genome

IGV: LOADING BAM FILES

The screenshot displays the IGV application window. The title bar reads 'IGV_2.3.52'. The menu bar includes 'File', 'Genomes', 'View', 'Tracks', 'Regions', 'Tools', 'GenomeSpace', and 'Help'. Below the menu bar, there is a search bar containing 'Borrelia burgdorferi ...' and a dropdown menu set to 'All'. To the right of the search bar are navigation icons, including a zoom-in icon (magnifying glass) and a zoom-out icon (minus sign), which are highlighted with a red rectangular box. The main display area is divided into several tracks. The top track is a reference genome track showing chromosome labels (lp21, lp25, lp28-1, lp28-2, lp28-3, lp28-4, lp38, lp54, cp8, cp32-1, cp32-4, cp32-7, cp32-8, cp17, cp25, cp28-2, cp28-4, cp38, cp56, cp26, cp32-3, cp32-6, cp32-8). Below this is a track labeled 'p- W'. The next track is a coverage track with a legend on the left showing colored bars for '20390_ACTTG...am Coverage' (red), '20390_ACTTGA_C4G2YANDXX_0140718B_20140718_1.stname.d.bam' (green), and 'Gene' (blue). The bottom track is a gene track showing a blue bar representing gene expression. The status bar at the bottom indicates '4 tracks', 'chr3:97,106,896', and '155M of 513M'.

- ▶ Zoom: either use the zoom panel

IGV: LOADING BAM FILES



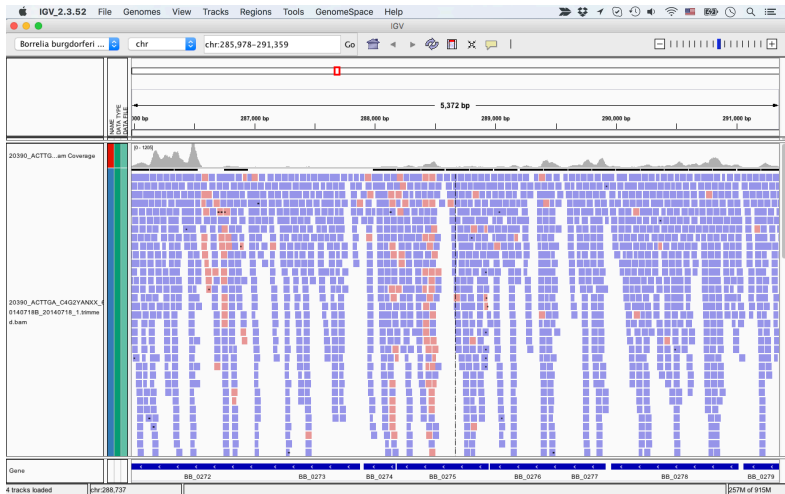
- ▶ Zoom: or use the mouse to mark region you want to look at

IGV: LOADING BAM FILES

The screenshot displays the IGV application window. The title bar shows 'IGV_2.3.52' and various system icons. The menu bar includes 'File', 'Genomes', 'View', 'Tracks', 'Regions', 'Tools', 'GenomeSpace', and 'Help'. The address bar shows 'Borrelia burgdorferi ...', 'chr', and 'chr:248,001-327,000'. The main view area is currently empty, displaying the text 'Zoom in to see alignments.' The track list on the left includes '20390_ACTTG...am Coverage', '20390_ACTTGA_C4G2YANDXX_...', and '0140718B_20140718_1_s1name.d.bam'. The gene track at the bottom shows 'Gene' and 'chr:278.586' with a scale of '164M of 513M'.

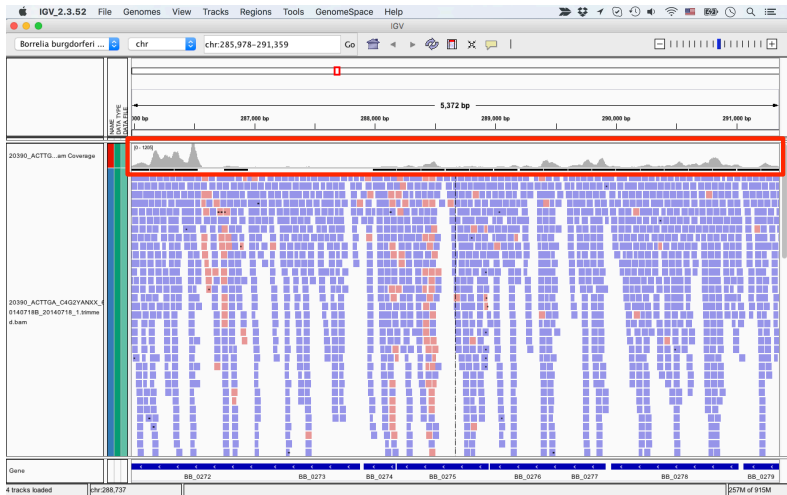
- ▶ Max zoom level for showing alignments can be adjusted on the Preference screen

IGV: WORKING WITH BAM FILES



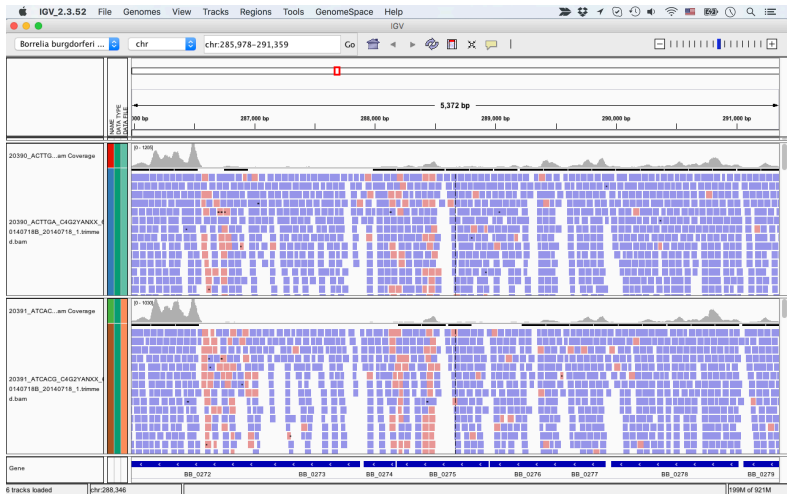
- Reads are displayed as red/blue or grey bars

IGV: WORKING WITH BAM FILES



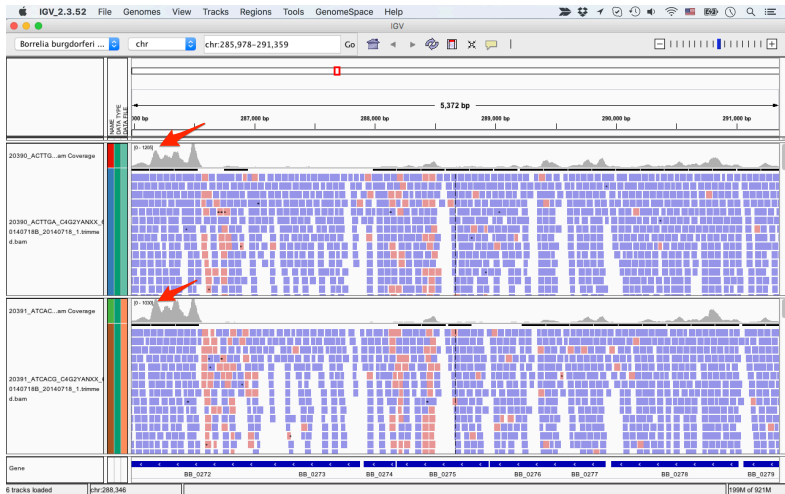
- Coverage: how many reads overlap each position of the genome

IGV: WORKING WITH BAM FILES



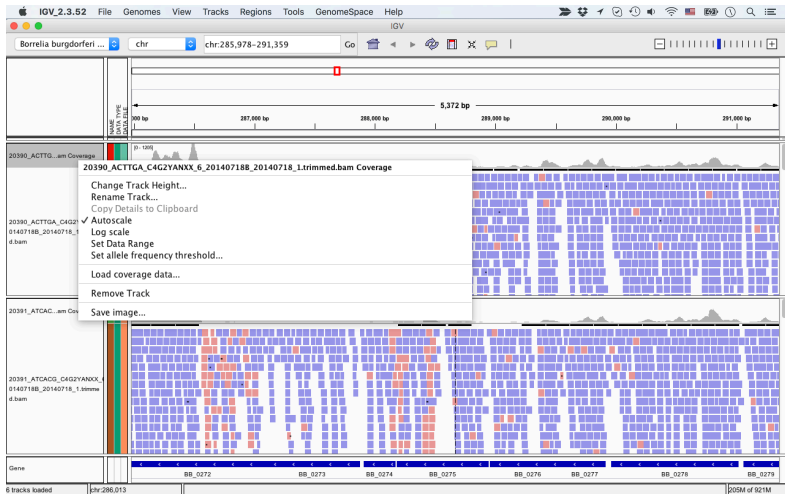
► Comparing two BAM files

IGV: WORKING WITH BAM FILES



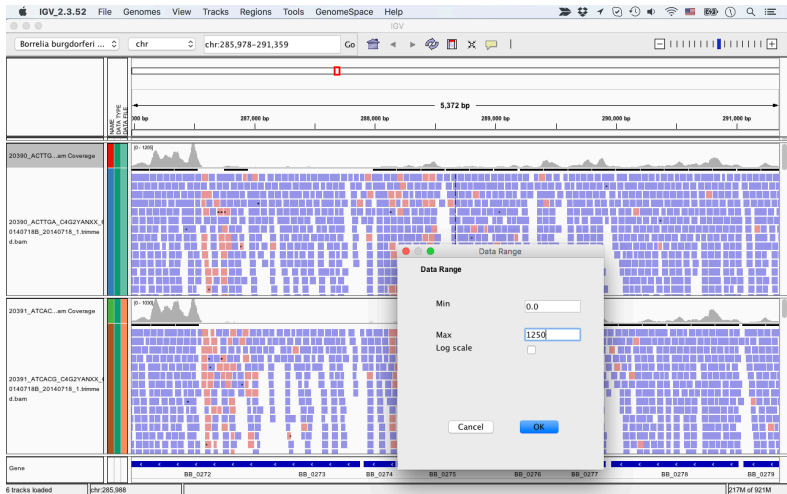
- ▶ Different ranges makes it hard to compare coverage

IGV: WORKING WITH BAM FILES



► Auto Scale vs. Set Data Range (+log scale)

IGV: WORKING WITH BAM FILES



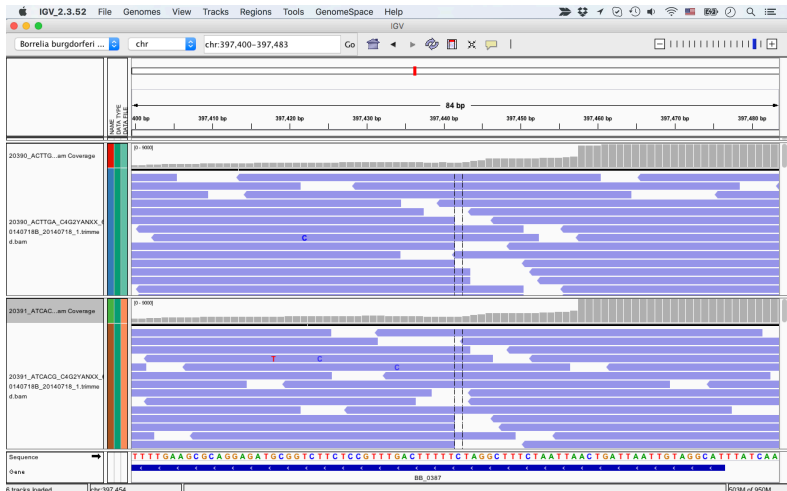
- ▶ Auto Scale vs. Set Data Range (+log scale)

IGV: WORKING WITH BAM FILES



- ▶ Auto Scale vs. Set Data Range (+log scale)

IGV: READ VISUALISATION



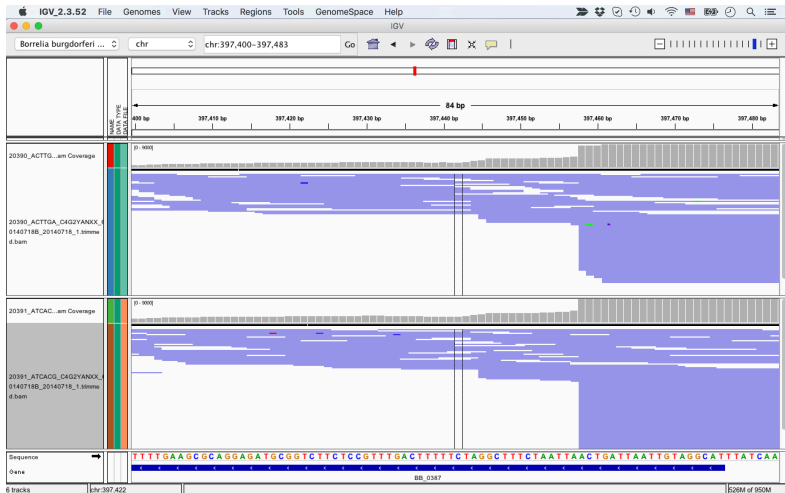
- Mismatches are shown with coloured letters. No letter means match to reference

IGV: READ VISUALISATION

The screenshot displays the IGV application window. The top menu bar includes 'IGV_2.3.52', 'File', 'Genomes', 'View', 'Tracks', 'Regions', 'Tools', 'GenomeSpace', and 'Help'. The address bar shows 'Borrelia burgdorferi ...', 'chr', and 'chr:397,400-397,483'. The main view shows two tracks: '20390_ACTTG...am Coverage' and '20391_ATCAC...am Coverage'. The tracks contain blue horizontal bars representing reads. A context menu is open over the tracks, listing various actions such as 'Rename Track...', 'Copy read details to clipboard', 'Group alignments by', 'Sort alignments by', 'Color alignments by', 'Shade base by quality', 'Show mismatched bases', 'View as pairs', 'Go to mate', 'View mate region in split screen', 'Set insert size options ...', 'Re-pack alignments', 'Show coverage track', 'Load coverage data...', 'Collapsed', 'Expanded', 'Squished', 'Select by name...', 'Clear selections', 'Copy read sequence', 'Blat read sequence', 'Copy consensus sequence', 'Sashimi Plot', 'Remove Track', 'Save image...', and 'Export Alignments...'. A small sub-menu is also visible over the tracks, listing 'no color', 'read strand', 'read group', 'sample', 'tag', and 'bisulfite mode'. The bottom status bar shows '6 tracks loaded' and 'chr:397,452'.

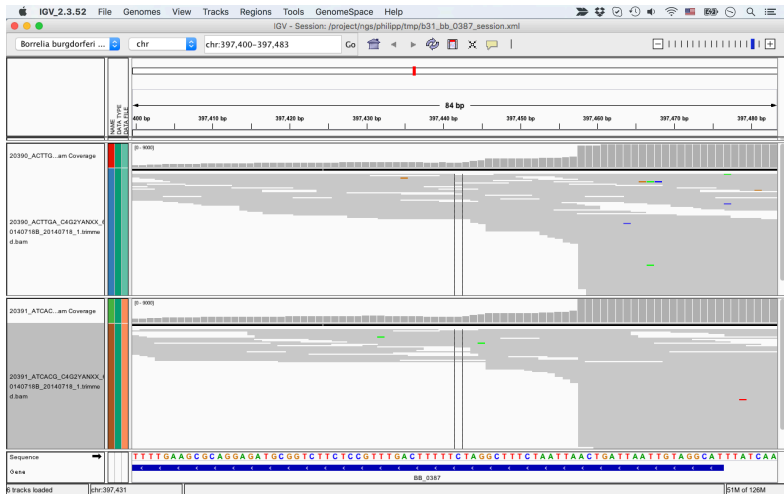
- Several options for colouring, sorting, grouping

IGV: READ VISUALISATION



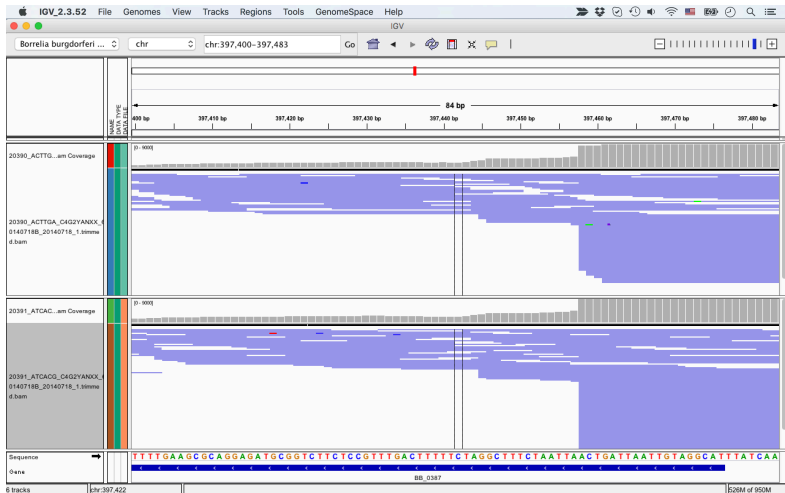
► Example: squished reads

IGV: READ VISUALISATION



► Example: squished reads + no color

IGV: SESSIONS



- ▶ Save current state of IGV

IGV: SESSIONS

The screenshot displays the IGV application window. The title bar reads "IGV_2.3.52" and the menu bar includes "File", "Genomes", "View", "Tracks", "Regions", "Tools", "GenomeSpace", and "Help". The main window shows a genomic track for "chr:397,400-397,483". A red vertical line indicates the current position. The "File" menu is open, with "Save Session..." highlighted in blue. Other menu items include "Load from File...", "Load from URL...", "Load from Server...", "Load from DAS...", "New Session...", "Open Session...", "Save Image...", and "Exit".

The tracks shown are:

- 20390_ACTTG...am Coverage
- 20390_ACTTGA_C4G2YANXX_0140718B_20140718_1_slimma.d.bam
- 20391_ATCAC...am Coverage
- 20391_ATCAC0_C4G2YANXX_0140718B_20140718_1_slimma.d.bam

The sequence track at the bottom shows the following sequence: `TTTTGAAGCGCAGGAGATGCGGTCTTCTCCGTTTGACTTTTTCTAGGCCTTCTAATTAACCTGATTAATTGATGGCATTATCAA`. The current position is marked as "BB_0387".

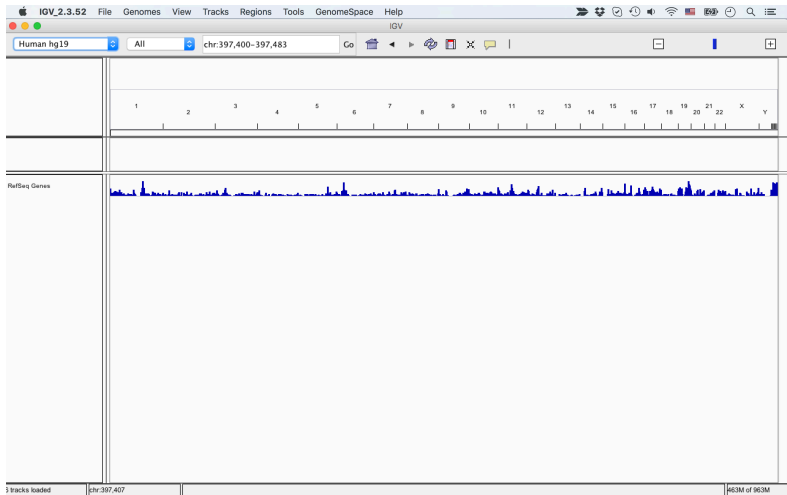
► Save session to file

IGV: SESSIONS

The screenshot displays the IGV application window with a 'Save Session' dialog box open. The dialog box is titled 'Save Session' and has a 'Save As' field containing 'b31_bb_0387_session.xml'. Below the 'Save As' field is a 'Tags' field. The dialog box also shows a file browser view of the 'tmp' directory, listing various files including genome files, BAM files, and BED files. The background IGV window shows tracks for '20390_ACTTG...am Coverage', '20391_ATCAC...am Coverage', and 'Sequence'. The sequence track at the bottom shows the DNA sequence: TTTTGAAGCGCAGGAGATGCGGGTCTTCTCGTTTGACTTTTTCTAGGCTTCTAATTAACGTGATTAATTGATGGCATTATCA.

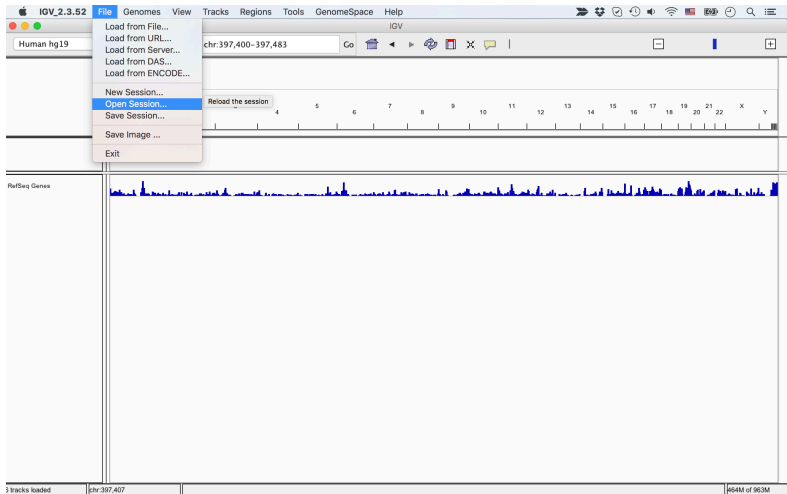
► Save session to file

IGV: SESSIONS



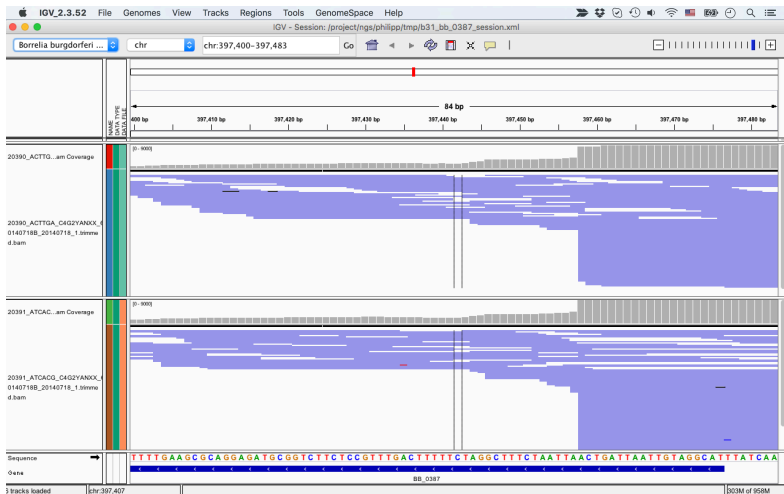
► New session

IGV: SESSIONS



► Open session

IGV: SESSIONS



► Open session

IGV: SESSIONS

The screenshot displays the IGV application window. The title bar reads "IGV_2.3.52" and includes standard macOS window controls. The menu bar contains "File", "Genomes", "View", "Tracks", "Regions", "Tools", "GenomeSpace", and "Help". The "View" menu is currently open, showing the following options:

- Preferences...
- Color Legends ...
- Show Name Panel
- Set Name Panel Width...
- Show Attribute Display
- Select Attributes to Show...
- Show Header Panel
- Reorder Panels...
- Go to

The main interface shows a genomic track for "Borrelia burgdorferi ... chr". The track is zoomed in to a region from 397,436 bp to 397,480 bp. A scale bar at the top indicates an 84 bp zoom. Below the scale bar, there are two coverage tracks for samples 20390 and 20391. The tracks show read coverage with blue bars. At the bottom, a sequence track displays the DNA sequence: TTTTG AAGCGCAGGAGATGCGGCTTCTCGGTTTGACTTTTCTAGGGTTTCTAATTAAGTGAATTGTAGGCATTTATCAA. The status bar at the bottom indicates "3 tracks loaded", "chr:397,426", and "305M of 958M".

► Change settings to **relative** paths

IGV: SESSIONS

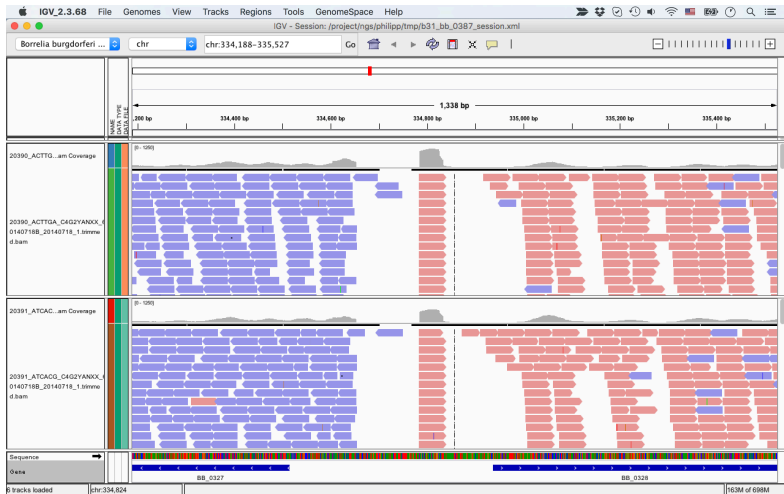
The screenshot displays the IGV application window with the 'General' settings panel open. A red arrow points to the 'General' tab. The settings are as follows:

- Use relative paths in sessions
- Distinguish missing data *Distinguish regions with value of zero from regions with no data.*
- Display all tracks in a single panel
- Show attribute panel
- Show default track attributes *(NAME, DATA_TYPE, and DATA_FILE).*
- Show region boundaries
- Zoom to features
- Enable Google access *Enable loading from Google apis.*
- Save Google credentials *Save authorization credentials across sessions*
- Feature flanking region (bp or %): *< 0 is interpreted as a percentage.*
- Default visibility window (kilobases): *< 0 disables visibility window.*
- Sequence resolution threshold (bp/pixel):
- Default font:
- Scale fonts *Scale fonts for high resolution screens. Requires restart.*
- Background color click to change:

The background shows a genomic track for 'Borrelia burgdorferi' on chromosome 'chr'. The track displays coverage for '20390_ACTTG...am Coverage' and '20391_ATCAC...am Coverage'. The sequence at the bottom is 'TTTTGAAGC'.

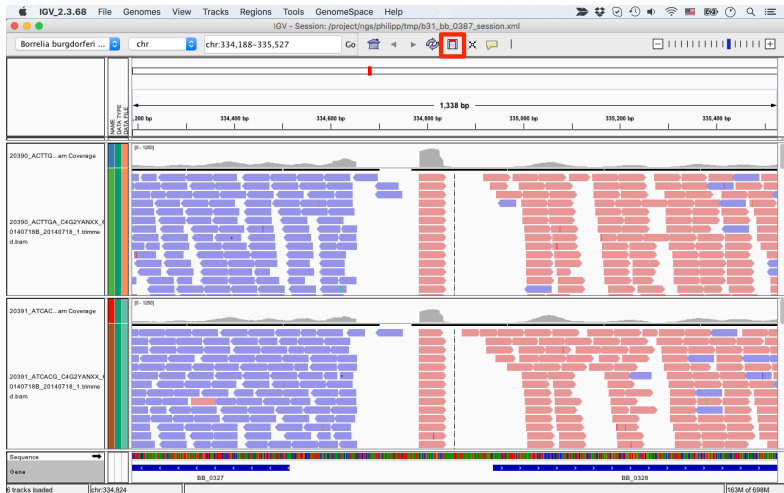
- Change settings to **relative** paths

IGV: REGIONS NAVIGATOR



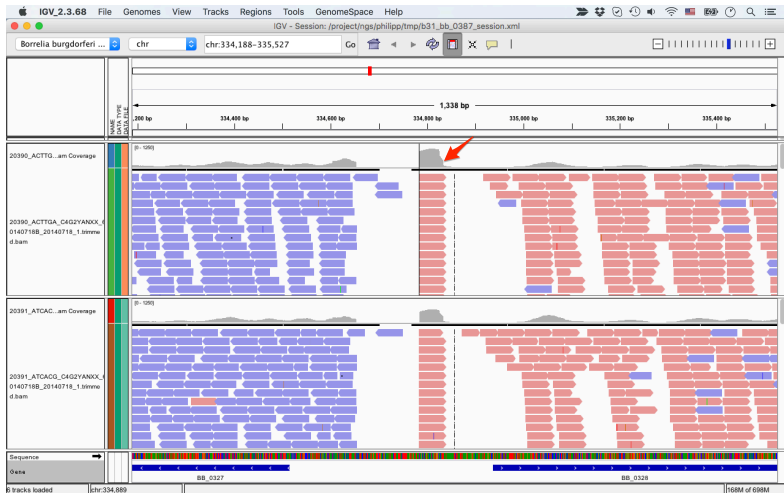
- ▶ Allows to define regions of interest

IGV: REGIONS NAVIGATOR



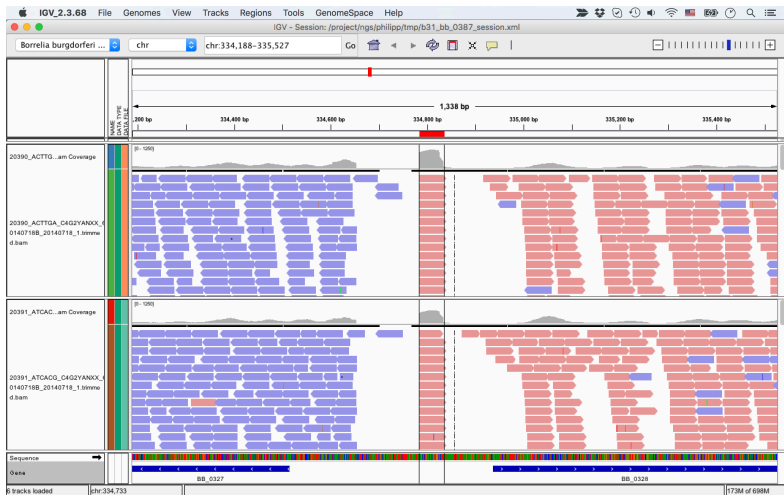
- ▶ Allows to define regions of interest

IGV: REGIONS NAVIGATOR



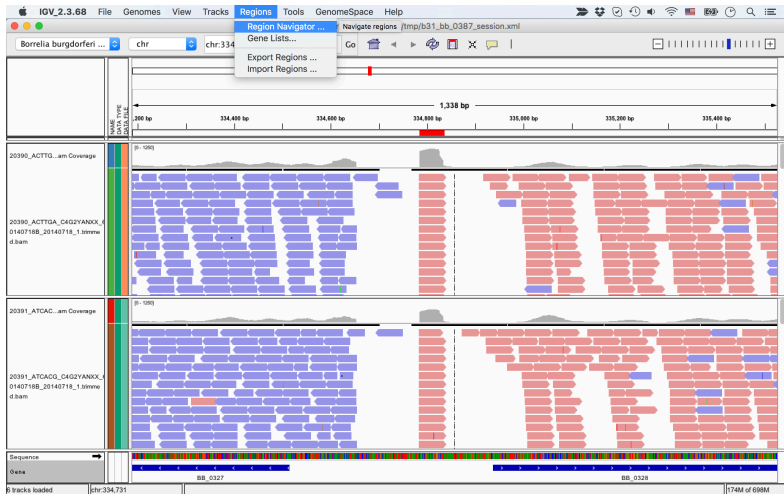
- ▶ Click twice to define region

IGV: REGIONS NAVIGATOR



► Click twice to define region

IGV: REGIONS NAVIGATOR



- ▶ Region navigator: allows to name regions and jump to their position

IGV: REGIONS NAVIGATOR

The screenshot displays the IGV interface with a 'Regions of Interest' dialog box. The dialog box has a 'Show All Chrs' checkbox checked and buttons for 'Add', 'Remove', 'View', and 'Zoom to Region'. Below the dialog box is a search input field and a 'Clear Search' button.

Chr	Start	End	Description
chr	334765	334830	
chr	144304	146669	Region 1
lp21	16149	16819	Region 2

Background tracks include:

- 20390_ACTTG...am Coverage
- 20390_ACTTGA_C4G2YANXX_0140718B_20140718_1_sinnme.d.bam
- 20391_ATCAC...am Coverage
- 20391_ATCAC_C4G2YANXX_0140718B_20140718_1_sinnme.d.bam
- Gene track with labels: BB_0142, BB_0143, BB_0144, BB_0145, BB_0146

Bottom status bar: 8 tracks loaded | chr:145,816 | 513M of 702M

- ▶ Region navigator: allows to name regions and jump to their position

IGV: REGIONS NAVIGATOR

The screenshot shows the IGV interface with a 'Regions of Interest' dialog box open. The dialog box contains the following table:

Chr	Start	End	Description
chr	334785	334830	Potential sRNA
chr	144304	146669	Region 1
lp21	16149	16819	Region 2

Below the table, there are buttons for 'View' and 'Zoom to Region'. A search bar is also present with a 'Clear Search' button.

The background shows a genomic track for chromosome 144, with tracks for coverage and gene models (BB_0142, BB_0143, BB_0144, BB_0145, BB_0146). The current view is centered on the region chr:144,129-146,808.

- ▶ Region navigator: allows to name regions and jump to their position

IGV: REGIONS NAVIGATOR



- ▶ Right click read bar to copy sequence to clipboard

IGV: REGIONS NAVIGATOR



- ▶ Export regions to BED file

Integrative Genomics Viewer

Home

Downloads

Documents

Hosted Genomes

FAQ

IGV User Guide

- User Interface
- Starting IGV
- Navigating
- Loading a Genome
- External Control of IGV
- Viewing the Reference Genome
- Loading Data and Attributes
- Viewing Data
- Viewing Alignments
- Viewing Variants
- Gene List View
- Regions of Interest
- Sample Attributes
- Sorting, Grouping, and Filtering
- Saving and Restoring Sessions
- Server Configuration
- igvtools
- Motif Finder
- BLAT search
- File Formats
- Release Notes
- IGV for iPad
- Credits

Contact

Home » IGV User Guide

IGV User Guide

This guide describes the Integrative Genomics Viewer (IGV).

- To start IGV, go to the IGV downloads page: <http://www.broadinstitute.org/igv/download>.

[Look at a printer-friendly HTML version of the whole User Guide.](#)

- [User Interface](#)
- [Starting IGV](#)
- [Navigating](#)
- [Loading a Genome](#)
- [External Control of IGV](#)
- [Viewing the Reference Genome](#)
- [Loading Data and Attributes](#)
- [Viewing Data](#)
- [Viewing Alignments](#)
- [Viewing Variants](#)
- [Gene List View](#)
- [Regions of Interest](#)
- [Sample Attributes](#)
- [Sorting, Grouping, and Filtering](#)
- [Saving and Restoring Sessions](#)
- [Server Configuration](#)
- [Motif Finder](#)
- [igvtools](#)
- [BLAT search](#)

[User Interface »](#)

► www.broadinstitute.org/software/igv/UserGuide