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

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**).

MODEL ORGANISMS

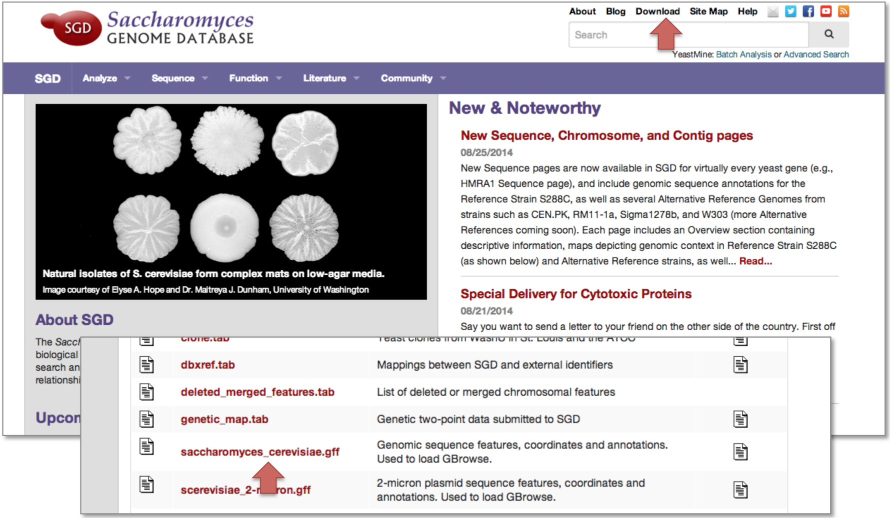
- FlyBase (Drosophila melanogaster)
- Wormbase (Caenorhabditis elegans)
- SGD (Saccharomyces Cerevisiae)
- TAIR (Arabidopsis thaliana)
- Colibri (Escherichia Coli)

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



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

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...**

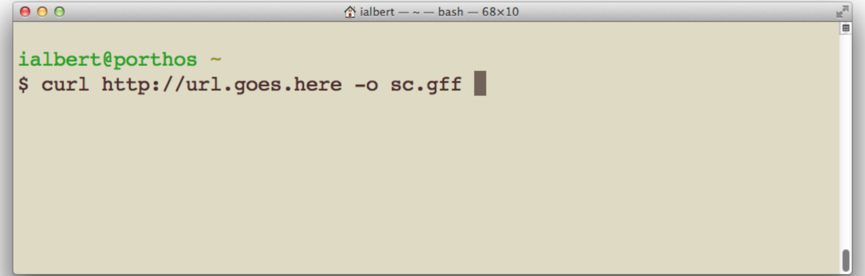
Special Delivery for Cytotoxic Proteins

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 WashU in 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-[illegible]ron.gff	2-micron plasmid sequence features, coordinates and annotations. Used to load GBrowse.	

DOWNLOADING ON THE COMMAND LINE



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

You can use the a browser or the command line:

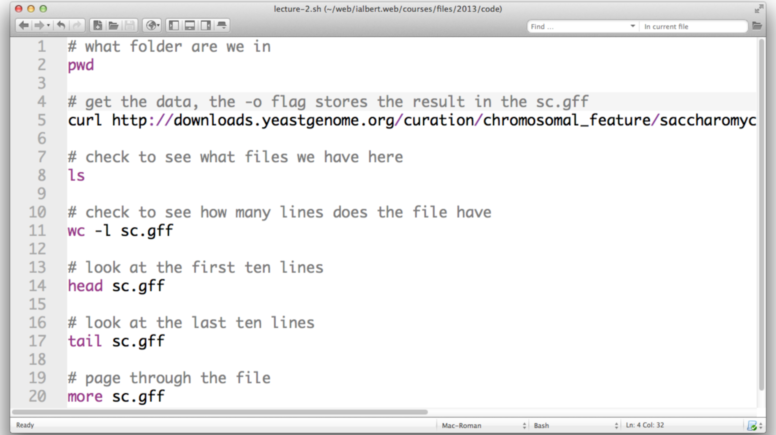
(The URL is on the course webpage)

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



```bash
# what folder are we in
pwd

# get the data, the -o flag stores the result in the sc.gff
curl http://downloads.yeastgenome.org/curation/chromosomal_feature/saccharomyc

# check to see what files we have here
ls

# check to see how many lines does the file have
wc -l sc.gff

# look at the first ten lines
head sc.gff

# look at the last ten lines
tail sc.gff

# page through the file
more sc.gff
```

TABLUAR FORMATS

- 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

GFF FORMAT

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

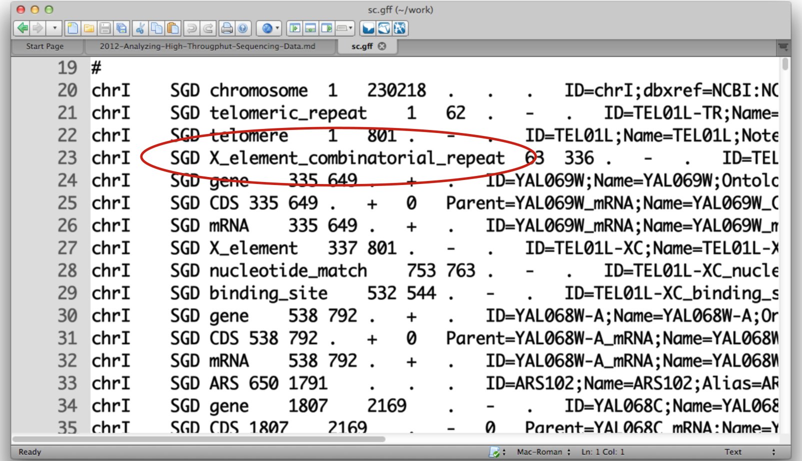
Tab separated with 9 columns. Missing attributes may be replaced with a dot → .

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

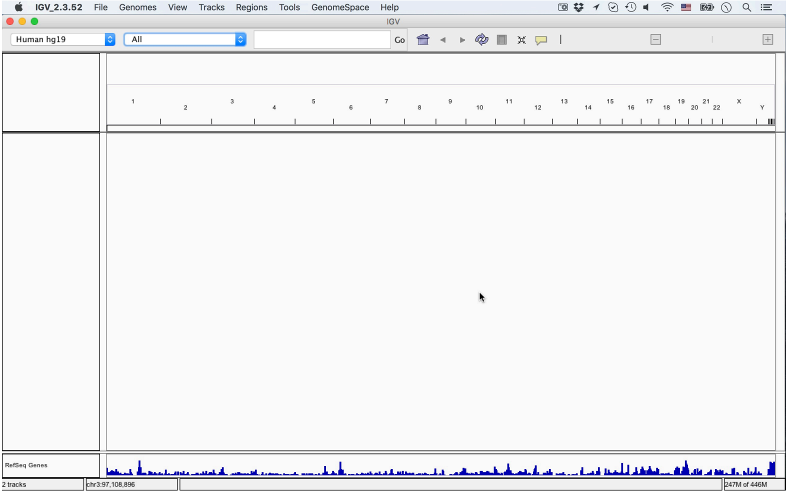


```
#
chrI    SGD chromosome  1   230218  .   .   .   ID=chrI;dbxref=NCBI:NC
chrI    SGD telomeric_repeat    1   62  .   -   .   ID=TEL01L-TR;Name=
chrI    SGD telomere    1   801 .   -   .   ID=TEL01L;Name=TEL01L;Note
chrI    SGD X_element_combinatorial_repeat  63  336 .   -   .   ID=TEL
chrI    SGD gene    335 649 .   +   .   ID=YAL069W;Name=YAL069W;Ontol
chrI    SGD CDS 335 649 .   +   0   Parent=YAL069W_mRNA;Name=YAL069W_
chrI    SGD mRNA    335 649 .   +   .   ID=YAL069W_mRNA;Name=YAL069W_m
chrI    SGD X_element   337 801 .   -   .   ID=TEL01L-XC;Name=TEL01L-X
chrI    SGD nucleotide_match    753 763 .   -   .   ID=TEL01L-XC_nucle
chrI    SGD binding_site    532 544 .   -   .   ID=TEL01L-XC_binding_s
chrI    SGD gene    538 792 .   +   .   ID=YAL068W-A;Name=YAL068W-A;Or
chrI    SGD CDS 538 792 .   +   0   Parent=YAL068W-A_mRNA;Name=YAL068W
chrI    SGD mRNA    538 792 .   +   .   ID=YAL068W-A_mRNA;Name=YAL068W
chrI    SGD ARS 650 1791    .   .   .   ID=ARS102;Name=ARS102;Alias=AR
chrI    SGD gene    1807    2169    .   -   .   ID=YAL068C;Name=YAL068
chrI    SGD CDS 1807    2169    .   -   0   Parent=YAL068C_mRNA;Name=Y
```

GENOME BROWSER

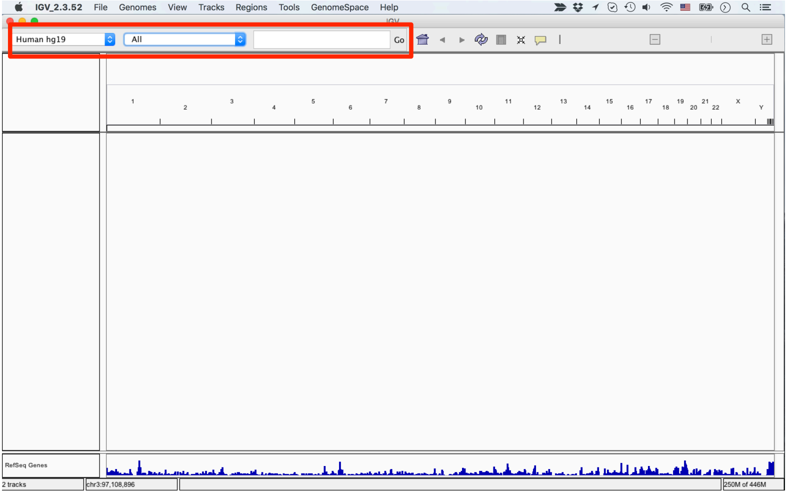
- Genome browser: one of the most important tools for data analysis
- ”Always look at your data”, true for statistics as well as for NGS data analysis
- Many genome browsers available: UCSC (web), IGB, Tablet, Artemis, MochiView (ChIP), Chip/SeqMonk (DNA methylation)
- Integrative Genomics Viewer
  - 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



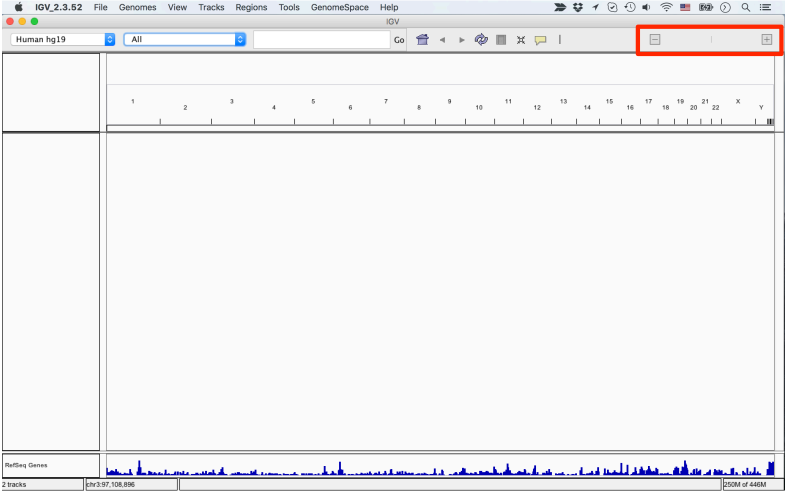
- Overview

IGV: BASICS



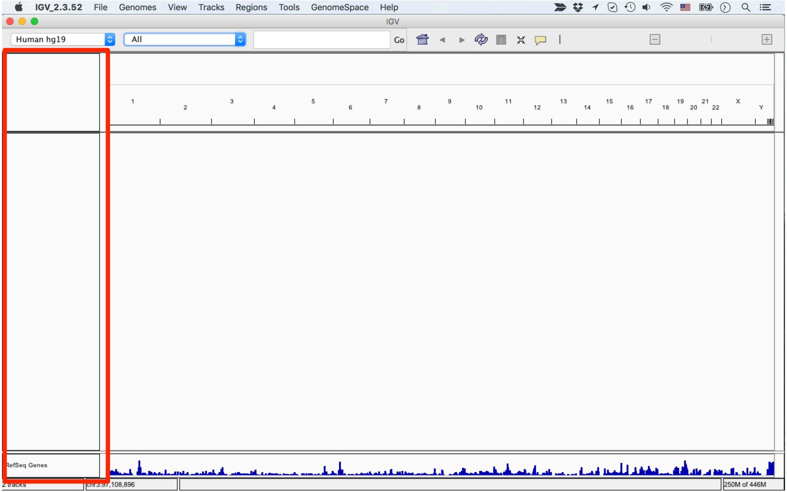
- Navigation area

IGV: BASICS



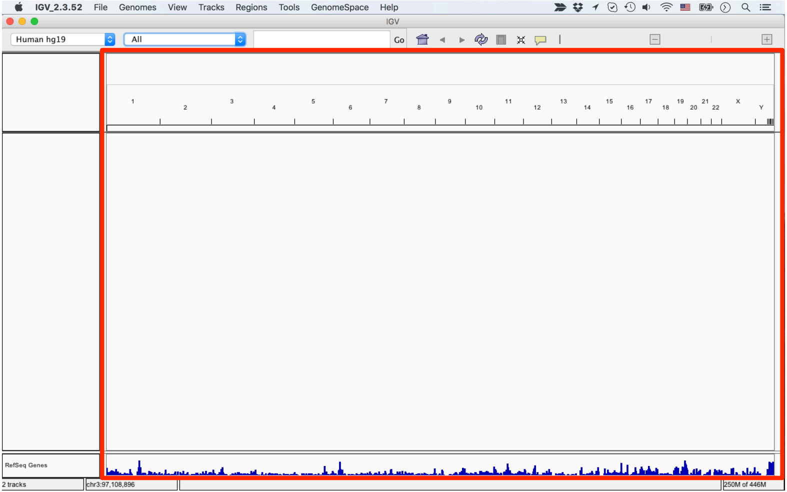
- Zoom

IGV: BASICS



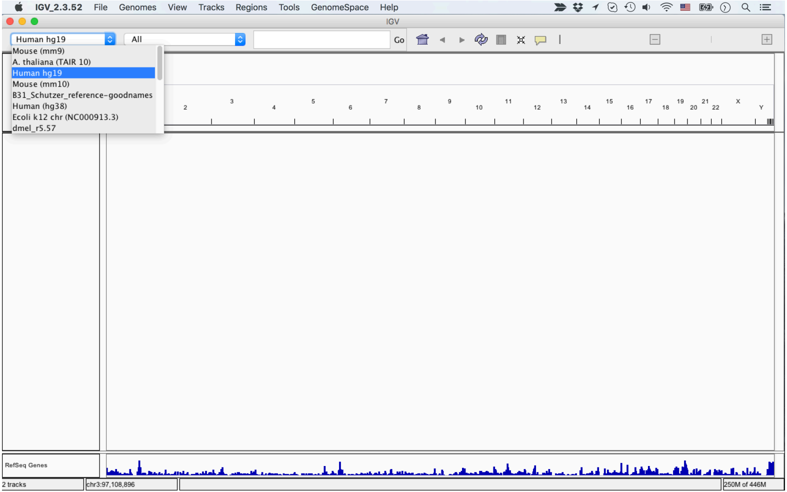
- Name panel

IGV: BASICS



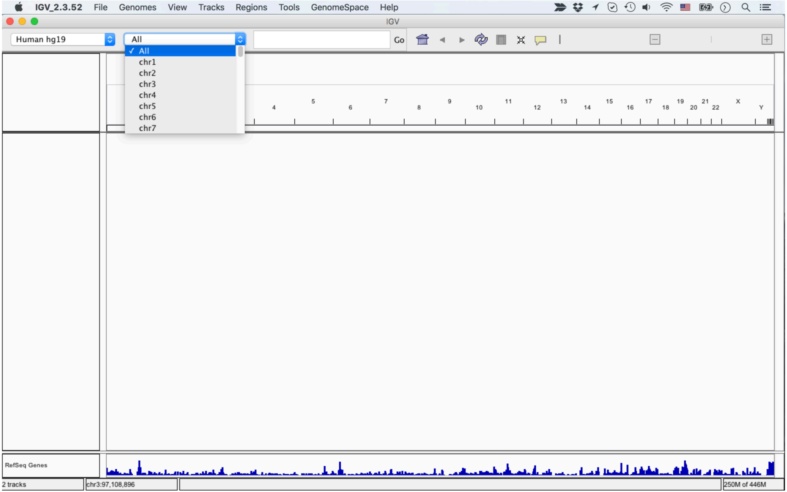
- Main area

IGV: GENOMES



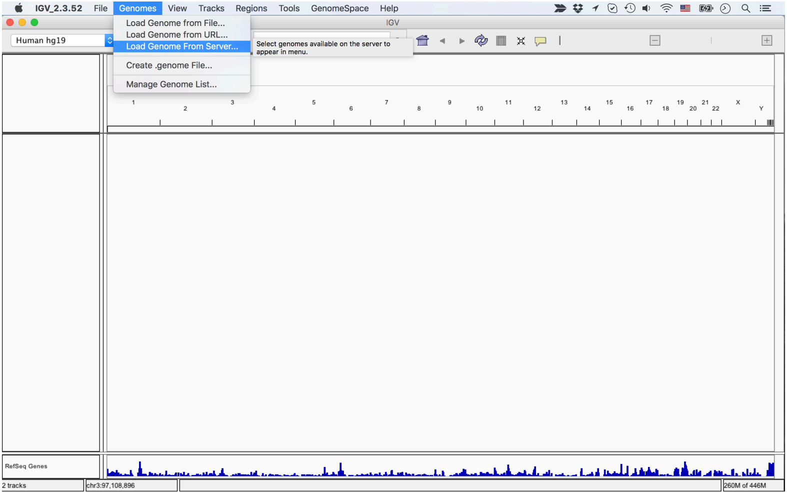
- List of available genomes

IGV: GENOMES



- Jump to chromosomes

IGV: GENOMES

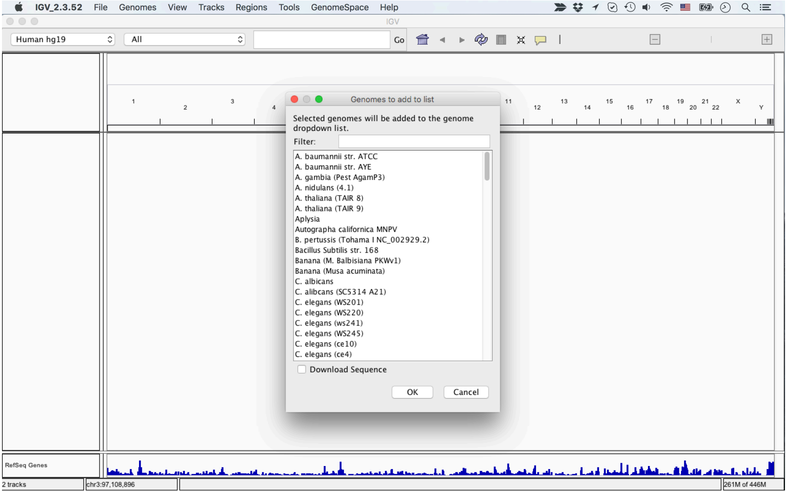


Select genomes available on the server to appear in menu.

2 tracks
chr3:97,108,896
260M of 446M

- Load genomes from Server

IGV: GENOMES



Genomes to add to list

Selected genomes will be added to the genome dropdown list.

Filter:

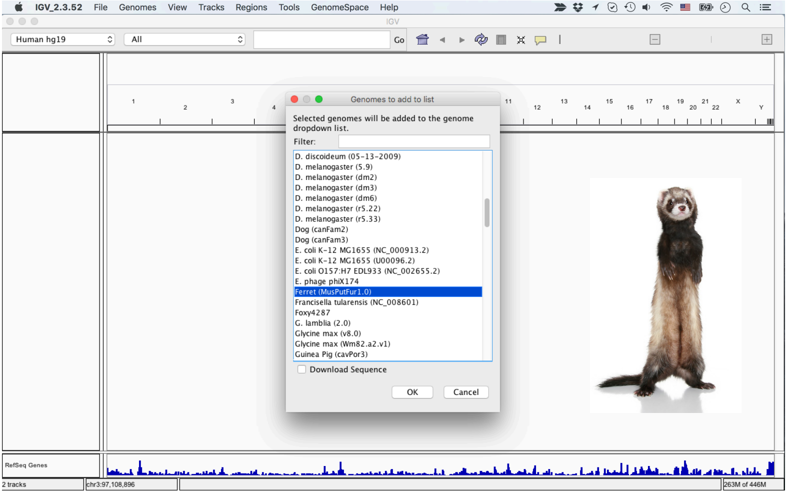
A. baumannii str. ATCC
A. baumannii str. AYE
A. gambia (Pest AgamP3)
A. nidulans (4.1)
A. thaliana (TAIR 8)
A. thaliana (TAIR 9)
Aplysia
Autographa californica MNPV
B. pertussis (Tohama I NC_002929.2)
Bacillus Subtilis str. 168
Banana (M. Balbisiana PKWv1)
Banana (Musa acuminata)
C. albicans
C. alibcans (SC5314 A21)
C. elegans (WS201)
C. elegans (WS220)
C. elegans (ws241)
C. elegans (WS245)
C. elegans (ce10)
C. elegans (ce4)

☐ Download Sequence

OK Cancel

- Load genomes from Server

IGV: GENOMES



Genomes to add to list

Selected genomes will be added to the genome dropdown list.

Filter:

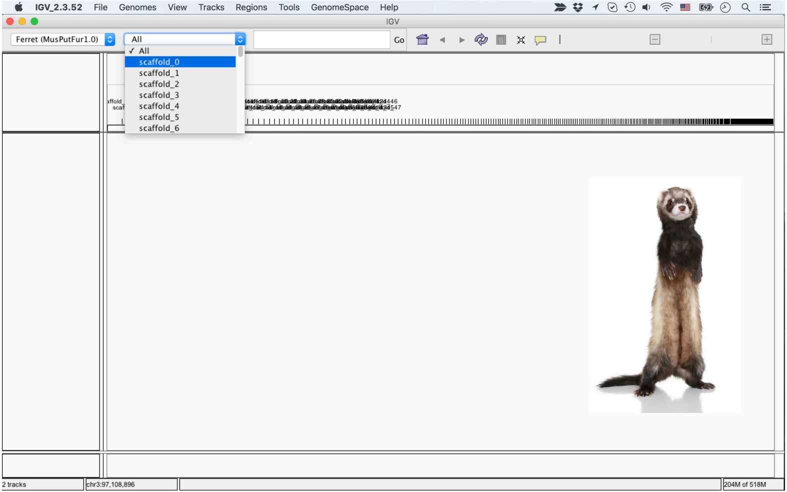
- D. discoideum (05-13-2009)
- D. melanogaster (5.9)
- D. melanogaster (dm2)
- D. melanogaster (dm3)
- D. melanogaster (dm6)
- D. melanogaster (r5.22)
- D. melanogaster (r5.33)
- Dog (canFam2)
- Dog (canFam3)
- E. coli K-12 MG1655 (NC_000913.2)
- E. coli K-12 MG1655 (U00096.2)
- E. coli O157:H7 EDL933 (NC_002655.2)
- E. phage phiX174
- Ferret (MusPutFur1.0)
- Francisella tularensis (NC_008601)
- Foxy4287
- G. lamblia (2.0)
- Glycine max (v8.0)
- Glycine max (Wm82.a2.v1)
- Guinea Pig (cavPor3)

☐ Download Sequence

OK Cancel

- Ferret it is

IGV: GENOMES



Ferret (MusPutFur1.0)

All

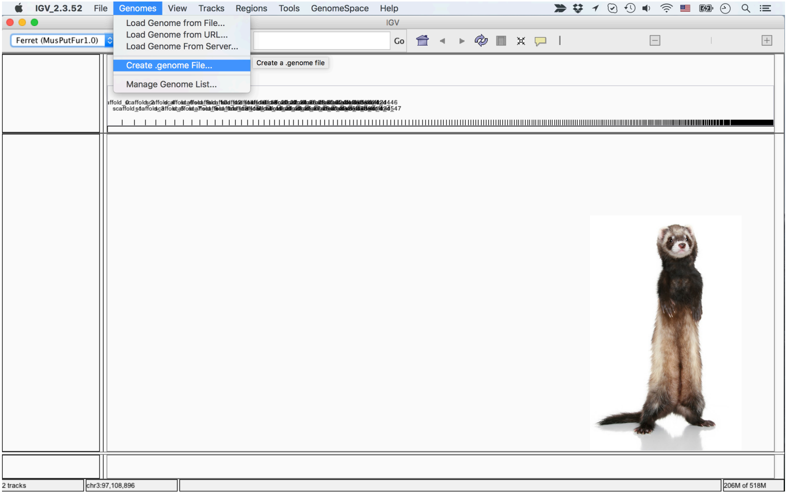
2 tracks

chr3:97,108,896

204M of 518M

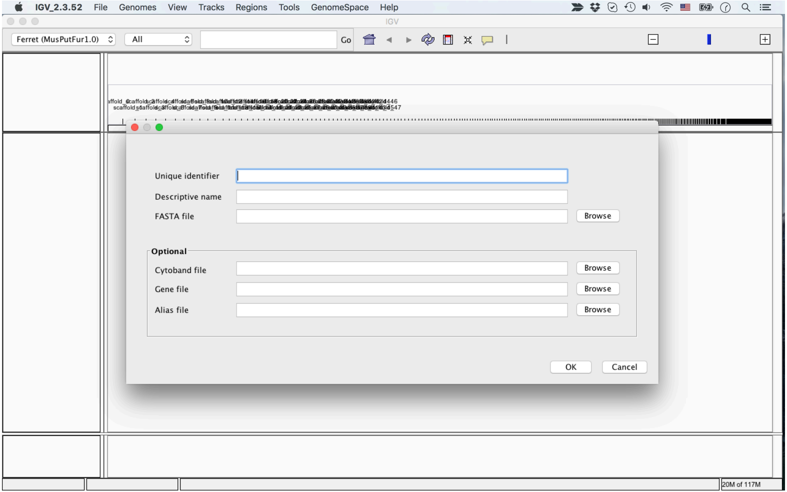
- Ferret it is

IGV: GENOMES



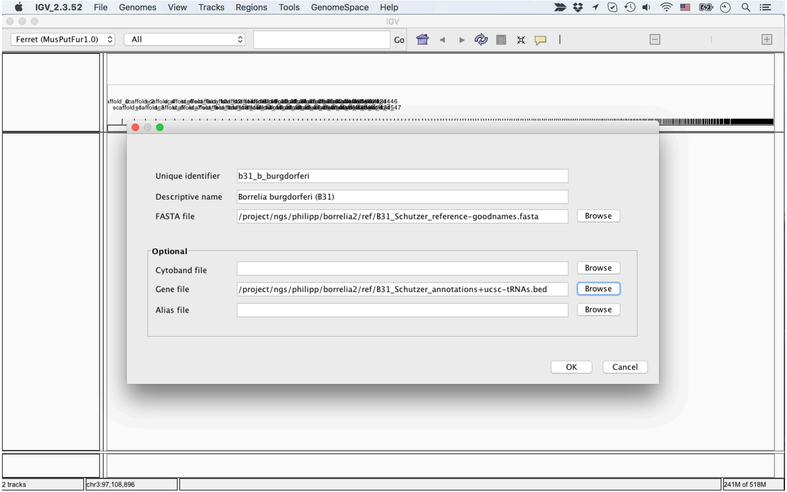
- We don’t have Ferret data, so we need something else: loding a custom genome

IGV: GENOMES



- Loading the Borrelia Burgdorferi genome

IGV: GENOMES



Unique identifier: b31_b_burgdorferi

Descriptive name: Borrelia burgdorferi (B31)

FASTA file: /project/ngs/philipp/borrelia2/ref/B31_Schutzer_reference-goodnames.fasta

Optional

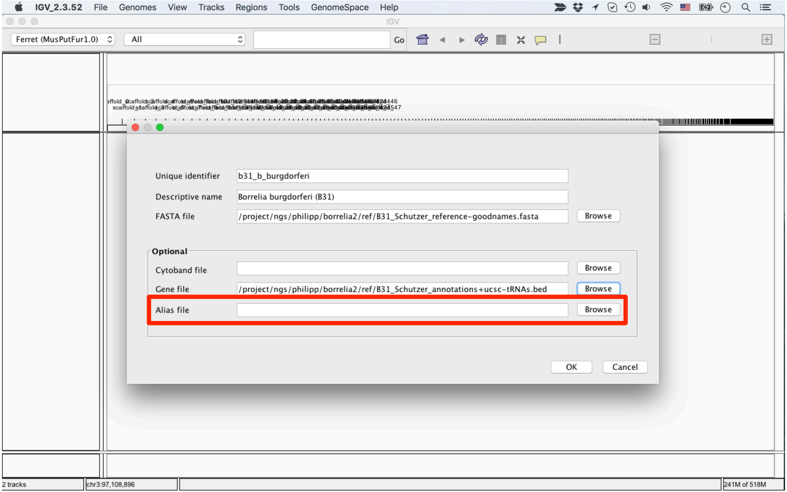
Cytoband file:

Gene file: /project/ngs/philipp/borrelia2/ref/B31_Schutzer_annotations+ucsc-tRNAs.bed

Alias file:

- Loading the Borrelia Burgdorferi genome

IGV: GENOMES



Unique identifier: b31_b_burgdorferi

Descriptive name: Borrelia burgdorferi (B31)

FASTA file: /project/ngs/philipp/borrelia2/ref/B31_Schutzer_reference-goodnames.fasta

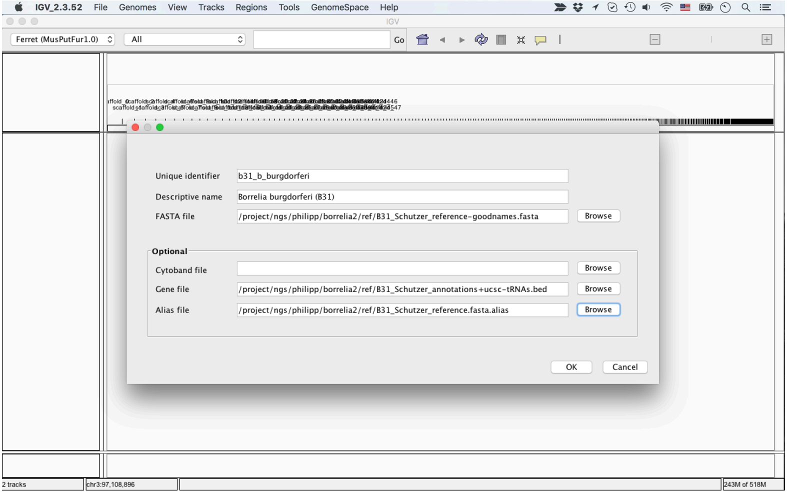
Cytoband file:

Gene file: /project/ngs/philipp/borrelia2/ref/B31_Schutzer_annotations+ucsc-tRNAs.bed

Alias file:

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

IGV: GENOMES



Unique identifier: b31_b_burgdorferi

Descriptive name: Borrelia burgdorferi (B31)

FASTA file: /project/ngs/philipp/borrelia2/ref/B31_Schutzer_reference-goodnames.fasta

Optional

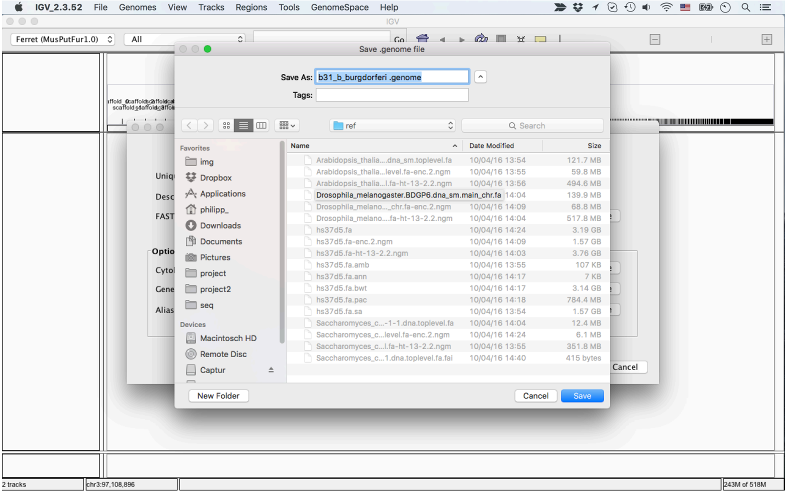
Cytoband file:

Gene file: /project/ngs/philipp/borrelia2/ref/B31_Schutzer_annotations+ucsc-tRNAs.bed

Alias file: /project/ngs/philipp/borrelia2/ref/B31_Schutzer_reference.fasta.alias

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

IGV: GENOMES



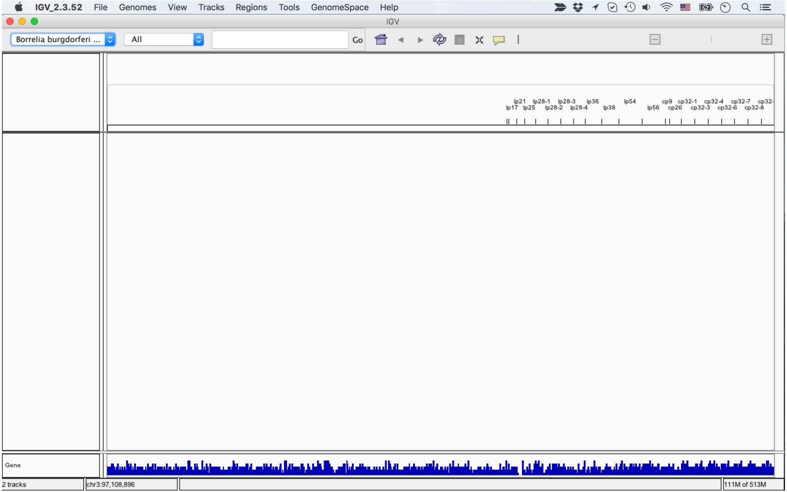
Save As: b31_b_burgdorferi .genome

Tags:

Name	Date Modified	Size
Arabidopsis_thalia....dna_sm.toplevel.fa	10/04/16 13:54	121.7 MB
Arabidopsis_thalia...level.fa-enc.2.ngm	10/04/16 13:55	59.8 MB
Arabidopsis_thalia...l.fa-ht-13-2.2.ngm	10/04/16 13:56	494.6 MB
Drosophila_melanogaster.BDGP6.dna_sm.main_chr.fa	14:04	139.9 MB
Drosophila_melano..._chr.fa-enc.2.ngm	10/04/16 14:09	68.8 MB
Drosophila_melano....fa-ht-13-2.2.ngm	10/04/16 14:04	517.8 MB
hs37d5.fa	10/04/16 14:24	3.19 GB
hs37d5.fa-enc.2.ngm	10/04/16 14:09	1.57 GB
hs37d5.fa-ht-13-2.2.ngm	10/04/16 14:03	3.76 GB
hs37d5.fa.amb	10/04/16 13:55	107 KB
hs37d5.fa.ann	10/04/16 14:17	7 KB
hs37d5.fa.bwt	10/04/16 14:17	3.14 GB
hs37d5.fa.pac	10/04/16 14:18	784.4 MB
hs37d5.fa.sa	10/04/16 13:54	1.57 GB
Saccharomyces_c...-1-1.dna.toplevel.fa	10/04/16 14:04	12.4 MB
Saccharomyces_c...level.fa-enc.2.ngm	10/04/16 14:24	6.1 MB
Saccharomyces_c...l.fa-ht-13-2.2.ngm	10/04/16 13:55	351.8 MB
Saccharomyces_c...1.dna.toplevel.fa.fai	10/04/16 14:40	415 bytes

- Save .genome file (contains all the information)

IGV: GENOMES



- Custom genome is now part of the genome list

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:

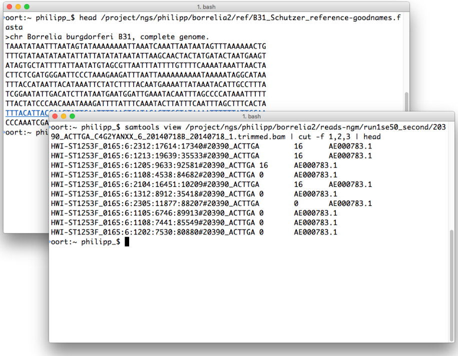
IGV: ALIAS FILES

- 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.
TAAATATAATTTAATAGTATAAAAAAAATTAAATCAAATTAATAATAGTTTAAAAAACTG
TTTGTATAATATAATATTATTATATATAATATTAAGCAACTACTATGATACTAATGAAGT
ATAGTGCTATTTTATTAATATGTAGCGTTAATTTATTTTGTTTTCAAAATAAATTAACTA
CTTCTCGATGGGAATTCCCTAAAGAAGATTTAATTAAAAAAAAAATAAAAATAGGCATAA
TTTACCATAATTACATAAATTCTATCTTTTACAATGAAAATTATAAATACATTGCCTTTA
TCGGAATATTGACATCTTATAATGAATGGATTGAAATACAATTTAGCCCCATAAATTTTT
TTACTATCCCAACAAATAAAGATTTTATTTCAAATACTTATTTCAATTTAGCTTTCACTA
TTTACATTACCAAGTATTCAATTTTAACTGATACACTTGCTATAAAATTTTTTATTGGAA
CCCAAATCGATTTAACTCTGAGAACTACTATATTTACAGGAAAAACAACTCATGCATTTC
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:

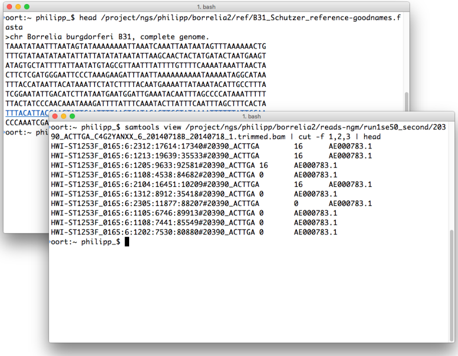


```
oort:~ philipp_$ head /project/ngs/philipp/borrelia2/ref/B31_Schutzer_reference-goodnames.f
asta
>chr Borrelia burgdorferi B31, complete genome.
TAAATATAATTTAATAGTATAAAAAAAATTAAATCAAATTAATAATAGTTTAAAAAACTG
TTTGTATAATATAATATTATTATATATAATATTAAGCAACTACTATGATACTAATGAAGT
ATAGTGCTATTTTATTAATATGTAGCGTTAATTTATTTTGTTTTCAAAATAAATTAACTA
CTTCTCGATGGGAATTCCCTAAAGAAGATTTAATTAAAAAAAAAATAAAAATAGGCATAA
TTTACCATAATTACATAAATTCTATCTTTTACAATGAAAATTATAAATACATTGCCTTTA
TCGGAATATTGACATCTTATAATGAATGGATTGAAATACAATTTAGCCCCATAAATTTTT
TTACTATCCCAACAAATAAAGATTTTATTTCAAATACTTATTTCAATTTAGCTTTCACTA
TTTACATTAC
CCCAAATCGA
oort:~ phi
```

```
oort:~ philipp_$ samtools view /project/ngs/philipp/borrelia2/reads-ngm/run1se50_second/203
90_ACTTGA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam | cut -f 1,2,3 | head
HWI-ST1253F_0165:6:2312:17614:17340#20390_ACTTGA        16      AE000783.1
HWI-ST1253F_0165:6:1213:19639:35533#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_0165:6:2104:16451:10209#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_$ 
```

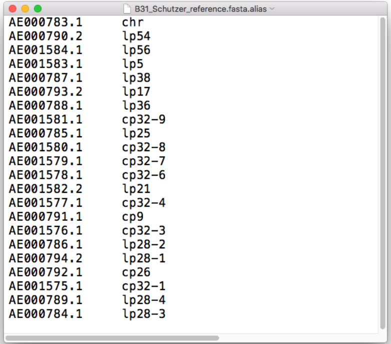
IGV: ALIAS FILES

- 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.
TAAATATAATTTAATAGTATAAAAAAAATTAAATCAAATTAATAATAGTTTAAAAAACTG
TTTGTATAATATAATATTATTATATATAATATTAAGCAACTACTATGATACTAATGAAGT
ATAGTGCTATTTTATTAATATGTAGCGTTAATTTATTTTGTTTTCAAAATAAATTAACTA
CTTCTCGATGGGAATTCCCTAAAGAAGATTTAATTAAAAAAAAAAATAAAAATAGGCATAA
TTTACCATAATTACATAAATTCTATCTTTTACAATGAAAATTATAAATACATTGCCTTTA
TCGGAATATTGACATCTTATAATGAATGGATTGAAATACAATTTAGCCCCATAAATTTTT
TTACTATCCCAACAAATAAAGATTTTATTTCAAATACTTATTTCAATTTAGCTTTCACTA
TTTACATTAC
CCCAAATCGA
oort:~ phi
```

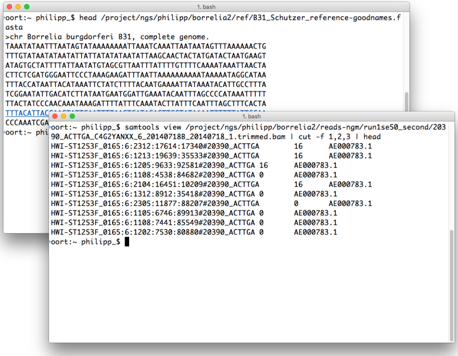
```
oort:~ philipp_$ samtools view /project/ngs/philipp/borrelia2/reads-ngm/run1se50_second/203
90_ACTTGA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam | cut -f 1,2,3 | head
HWI-ST1253F_0165:6:2312:17614:17340#20390_ACTTGA        16      AE000783.1
HWI-ST1253F_0165:6:1213:19639:35533#20390_ACTTGA        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_0165:6:2104:16451:10209#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_$ 
```



```
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: 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:



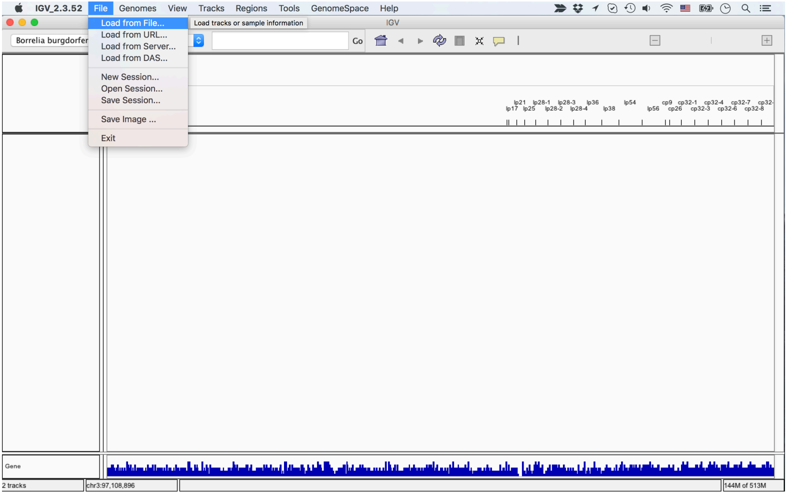
```
oort:~ philipp_$ head /project/ngs/philipp/borrelia2/ref/B31_Schutzer_reference-goodnames.f
asta
>chr Borrelia burgdorferi B31, complete genome.
TAAATATAATTTAATAGTATAAAAAAAATTAAATCAAATTAATAATAGTTTAAAAAACTG
TTTGTATAATATAATATTATTATATATAATATTAAGCAACTACTATGATACTAATGAAGT
ATAGTGCTATTTTATTAATATGTAGCGTTAATTTATTTTGTTTTCAAAATAAATTAACTA
CTTCTCGATGGGAATTCCCTAAAGAAGATTTAATTAAAAAAAAAATAAAAATAGGCATAA
TTTACCATAATTACATAAATTCTATCTTTTACAATGAAAATTATAAATACATTGCCTTTA
TCGGAATATTGACATCTTATAATGAATGGATTGAAATACAATTTAGCCCCATAAATTTTT
TTACTATCCCAACAAATAAAGATTTTATTTCAAATACTTATTTCAATTTAGCTTTCACTA
TTTACATTAC
CCCAAATCGA
oort:~ phi
```

```
oort:~ philipp_$ samtools view /project/ngs/philipp/borrelia2/reads-ngm/run1se50_second/203
90_ACTTGA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam | cut -f 1,2,3 | head
HWI-ST1253F_0165:6:2312:17614:17340#20390_ACTTGA        16      AE000783.1
HWI-ST1253F_0165:6:1213:19639:35533#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_0165:6:2104:16451:10209#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_$ 
```

```
AE000783.1      chr
```

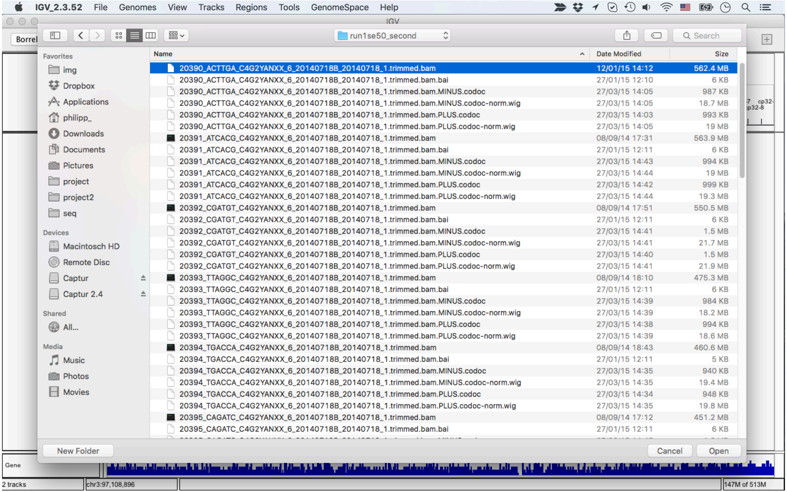
```
gi|387825439|ref|NC_012971.2|   chr
NC_012971.2     chr
```

IGV: LOADING BAM FILES



- BAM file has to be sorted and indexed

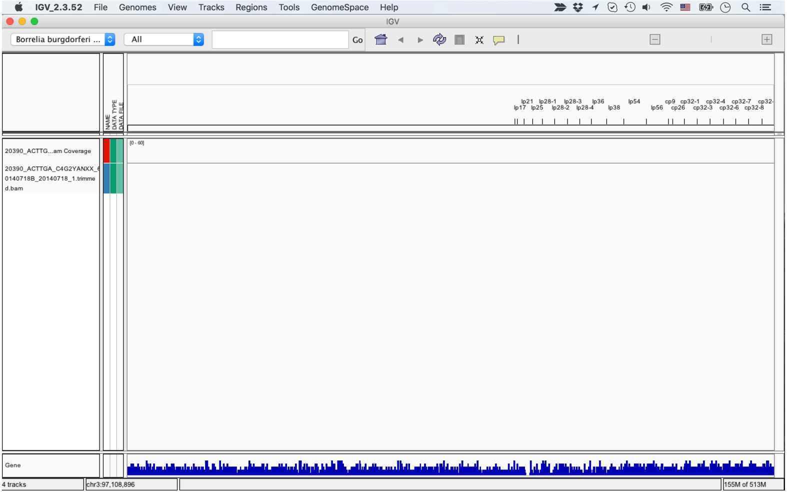
IGV: LOADING BAM FILES



Name	Date Modified	Size
20390_ACTTGA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam	12/01/15 14:12	562.4 MB
20390_ACTTGA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.bai	27/01/15 12:10	6 KB
20390_ACTTGA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.MINUS.codoc	27/03/15 14:05	987 KB
20390_ACTTGA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.MINUS.codoc-norm.wig	27/03/15 14:05	18.7 MB
20390_ACTTGA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.PLUS.codoc	27/03/15 14:03	993 KB
20390_ACTTGA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.PLUS.codoc-norm.wig	27/03/15 14:05	19 MB
20391_ATCACG_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam	08/09/14 17:31	563.9 MB
20391_ATCACG_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.bai	27/01/15 12:11	6 KB
20391_ATCACG_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.MINUS.codoc	27/03/15 14:43	994 KB
20391_ATCACG_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.MINUS.codoc-norm.wig	27/03/15 14:44	19 MB
20391_ATCACG_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.PLUS.codoc	27/03/15 14:42	999 KB
20391_ATCACG_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.PLUS.codoc-norm.wig	27/03/15 14:44	19.3 MB
20392_CGATGT_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam	08/09/14 17:51	550.5 MB
20392_CGATGT_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.bai	27/01/15 12:11	6 KB
20392_CGATGT_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.MINUS.codoc	27/03/15 14:41	1.5 MB
20392_CGATGT_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.MINUS.codoc-norm.wig	27/03/15 14:41	21.7 MB
20392_CGATGT_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.PLUS.codoc	27/03/15 14:40	1.5 MB
20392_CGATGT_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.PLUS.codoc-norm.wig	27/03/15 14:41	21.9 MB
20393_TTAGGC_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam	08/09/14 18:10	475.3 MB
20393_TTAGGC_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.bai	27/01/15 12:11	6 KB
20393_TTAGGC_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.MINUS.codoc	27/03/15 14:39	984 KB
20393_TTAGGC_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.MINUS.codoc-norm.wig	27/03/15 14:39	18.2 MB
20393_TTAGGC_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.PLUS.codoc	27/03/15 14:38	994 KB
20393_TTAGGC_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.PLUS.codoc-norm.wig	27/03/15 14:39	18.6 MB
20394_TGACCA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam	08/09/14 18:43	460.6 MB
20394_TGACCA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.bai	27/01/15 12:11	5 KB
20394_TGACCA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.MINUS.codoc	27/03/15 14:35	940 KB
20394_TGACCA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.MINUS.codoc-norm.wig	27/03/15 14:35	19.4 MB
20394_TGACCA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.PLUS.codoc	27/03/15 14:34	948 KB
20394_TGACCA_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.PLUS.codoc-norm.wig	27/03/15 14:35	19.8 MB
20395_CAGATC_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam	08/09/14 17:12	451.2 MB
20395_CAGATC_C4G2YANXX_6_20140718B_20140718_1.trimmed.bam.bai	27/01/15 12:11	6 KB

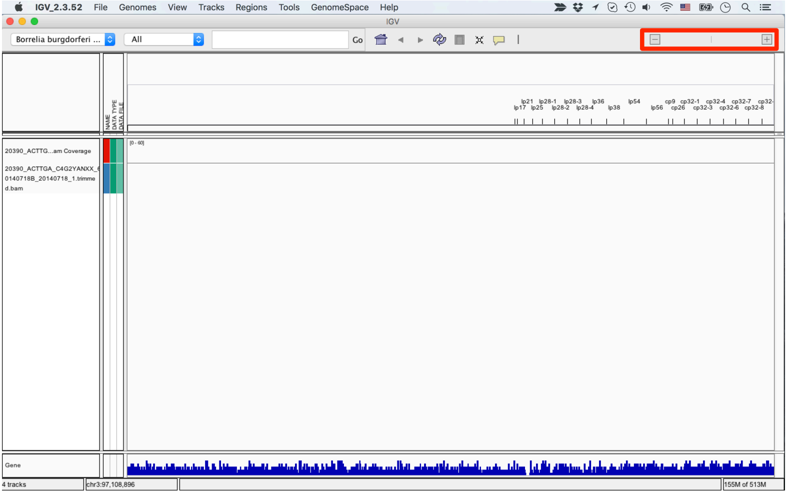
- BAM file has to be sorted and indexed

IGV: LOADING BAM FILES



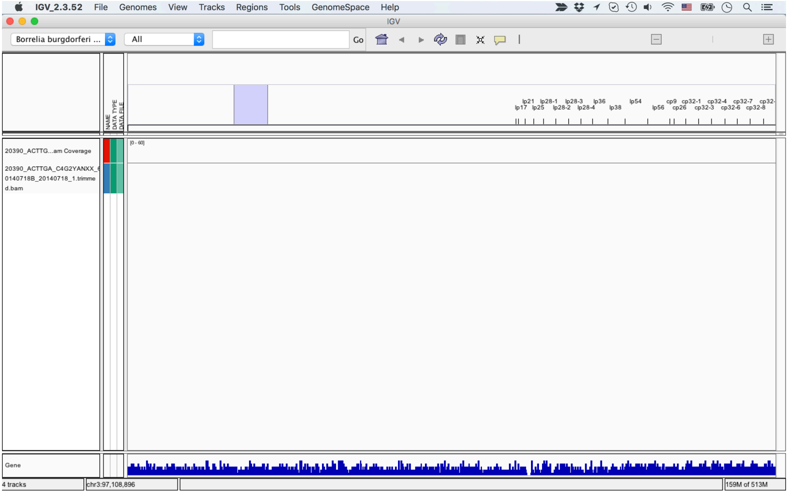
- Nothing visible when looking at the full genome

IGV: LOADING BAM FILES



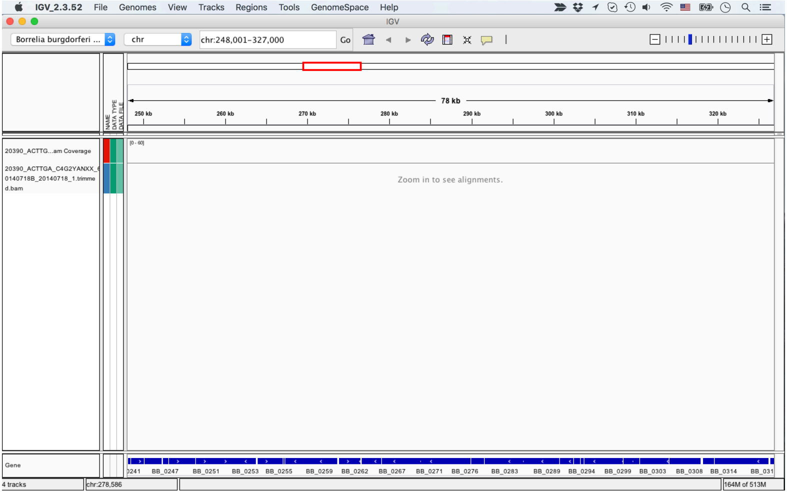
- Zoom: either use the zoom panel

IGV: LOADING BAM FILES



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

IGV: LOADING BAM FILES

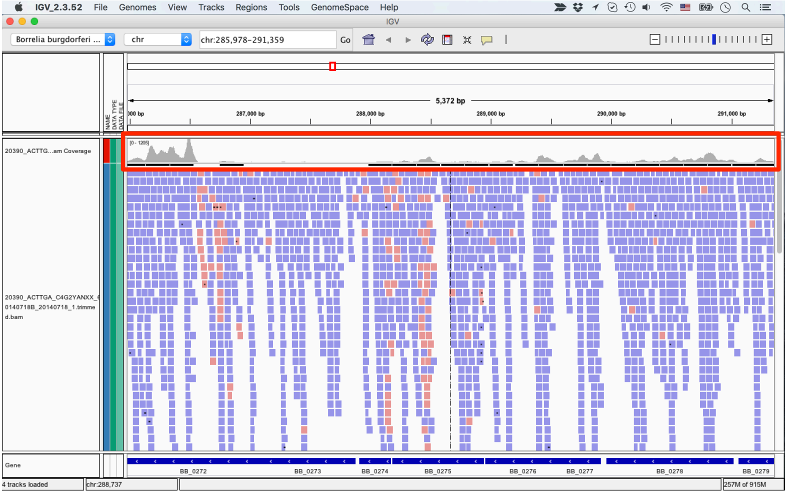


- Max zoom level for showing alignments can be adjusted on the Preference screen

IGV: WORKING WITH BAM FILES

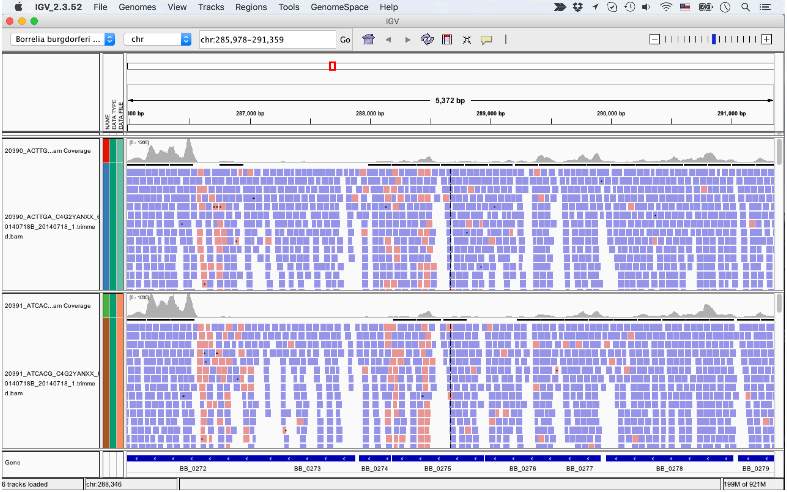
- Reads are displayed as red/blue or grey bars

IGV: WORKING WITH BAM FILES



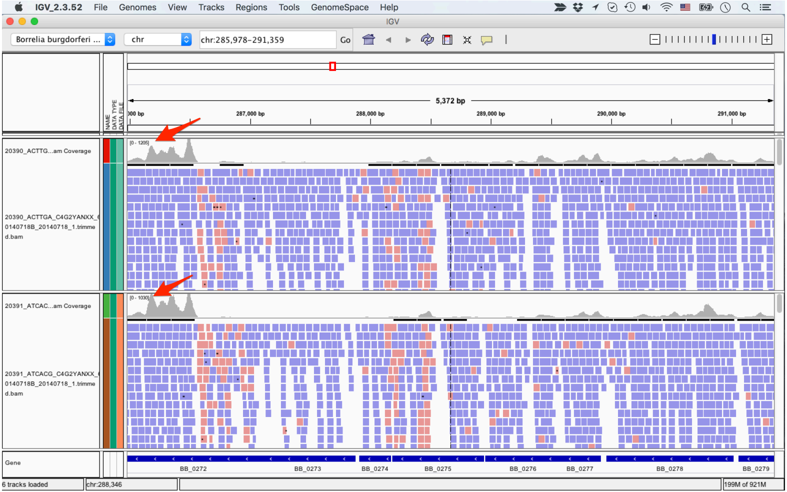
- Coverage: how man reads overlap each position of the genome

IGV: WORKING WITH BAM FILES



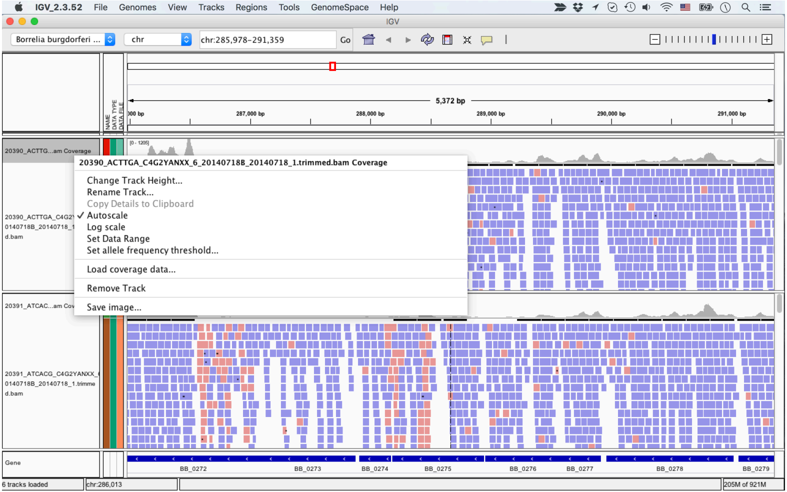
- Comparing two BAM files

IGV: WORKING WITH BAM FILES



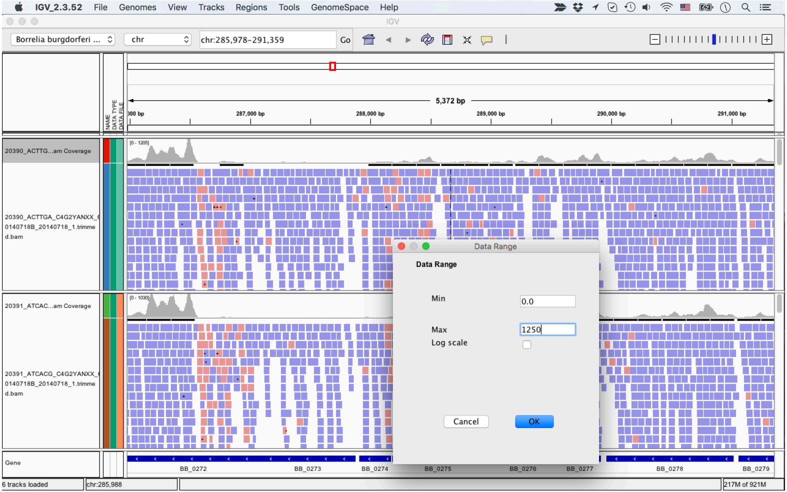
- Different ranges makes it hard to compare coverage

IGV: WORKING WITH BAM FILES



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

IGV: WORKING WITH BAM FILES



Data Range

Min: 0.0

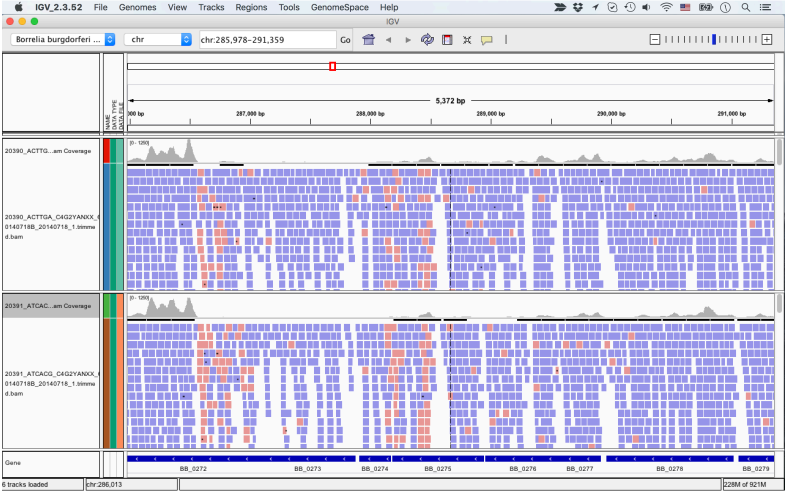
Max: 1250

Log scale ☐

Cancel   OK

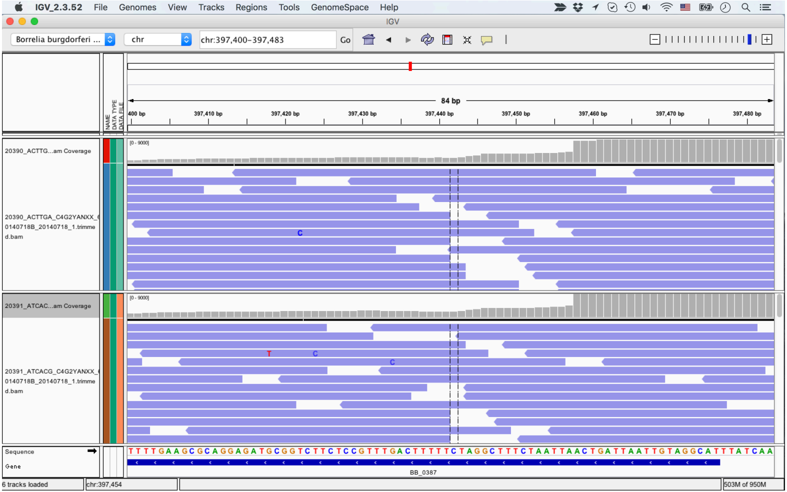
- Auto Scale vs. Set Data Range (+log scale)

IGV: WORKING WITH BAM FILES



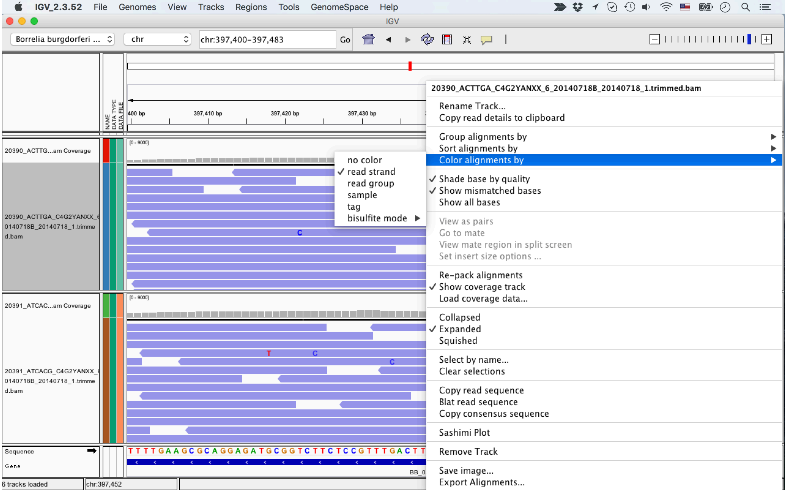
- Auto Scale vs. Set Data Range (+log scale)

IGV: READ VISUALISATION



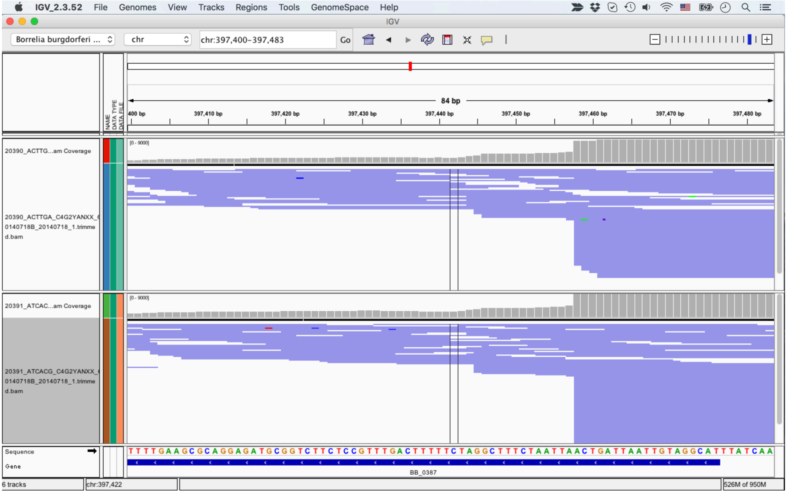
- Mismatches are shown with coloured letters. No letter means match to reference

IGV: READ VISUALISATION



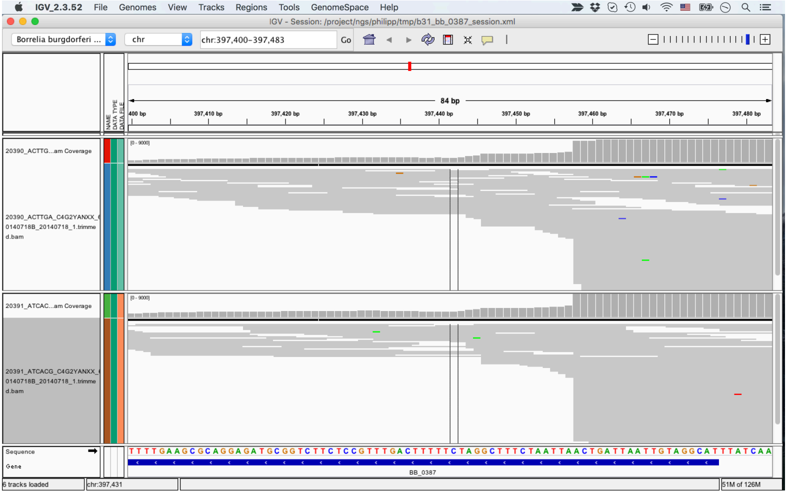
- Several options for colouring, sorting, grouping

IGV: READ VISUALISATION



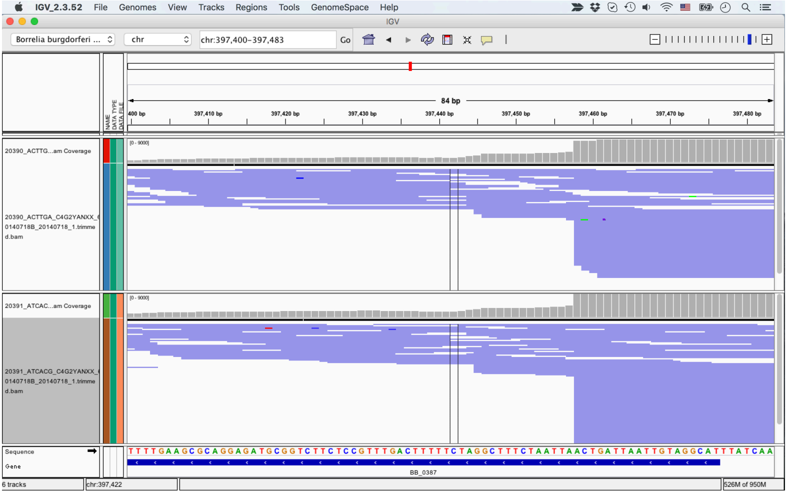
- Example: squished reads

IGV: READ VISUALISATION



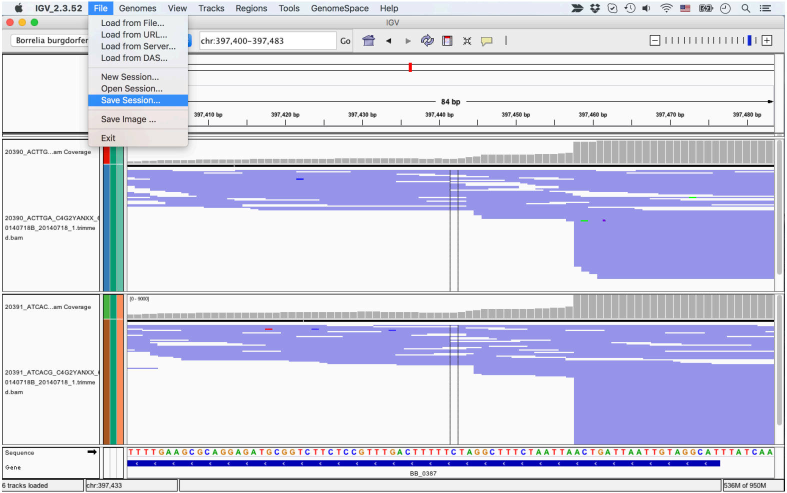
- Example: squished reads + no color

IGV: SESSIONS



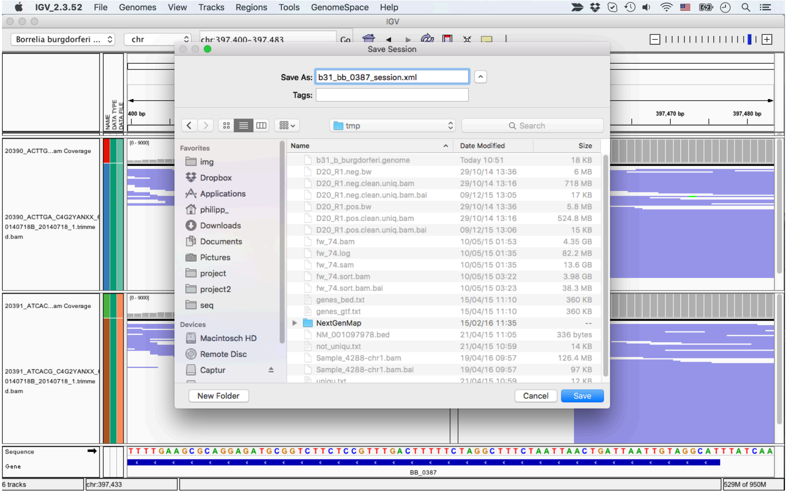
- Save current state of IGV

IGV: SESSIONS



- Save session to file

IGV: SESSIONS



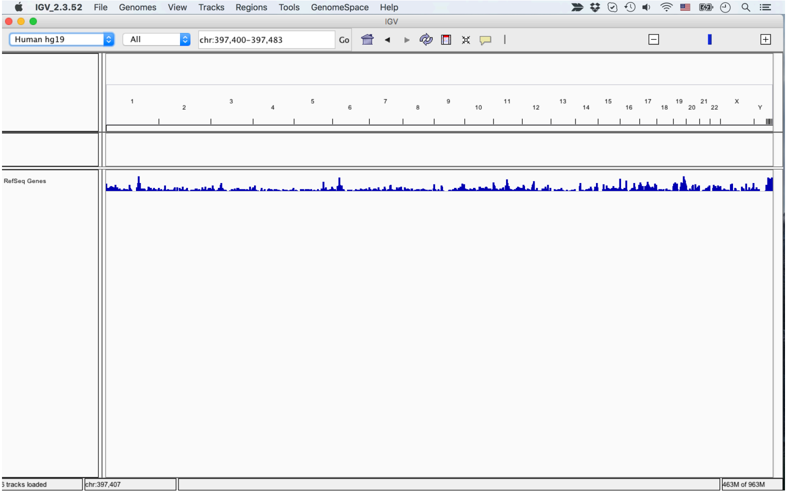
Save As: b31_bb_0387_session.xml

Tags:

Name	Date Modified	Size
b31_b_burgdorferi.genome	Today 10:51	18 KB
D20_R1.neg.bw	29/10/14 13:36	6 MB
D20_R1.neg.clean.uniq.bam	29/10/14 13:16	718 MB
D20_R1.neg.clean.uniq.bam.bai	09/12/15 13:05	17 KB
D20_R1.pos.bw	29/10/14 13:36	5.8 MB
D20_R1.pos.clean.uniq.bam	29/10/14 13:16	524.8 MB
D20_R1.pos.clean.uniq.bam.bai	09/12/15 13:06	15 KB
fw_74.bam	10/05/15 01:53	4.35 GB
fw_74.log	10/05/15 01:35	82.2 MB
fw_74.sam	10/05/15 01:35	13.6 GB
fw_74.sort.bam	10/05/15 03:22	3.98 GB
fw_74.sort.bam.bai	10/05/15 03:23	38.3 MB
genes_bed.txt	15/04/15 11:10	360 KB
genes_gtf.txt	15/04/15 11:10	360 KB
NextGenMap	15/02/16 11:35	--
NM_001097978.bed	21/04/15 11:05	336 bytes
not_uniqu.txt	21/04/15 10:59	14 KB
Sample_4288-chr1.bam	19/04/16 09:57	126.4 MB
Sample_4288-chr1.bam.bai	19/04/16 09:57	97 KB
uniqu.txt	21/04/15 10:59	12 KB

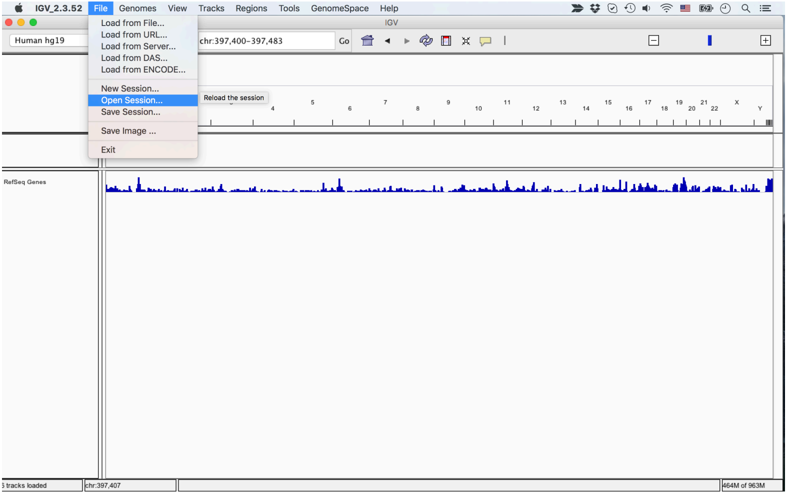
- Save session to file

IGV: SESSIONS



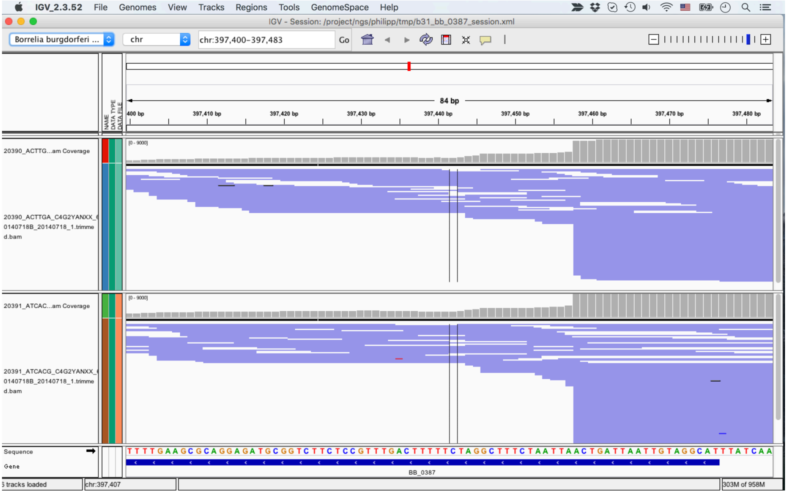
- New session

IGV: SESSIONS



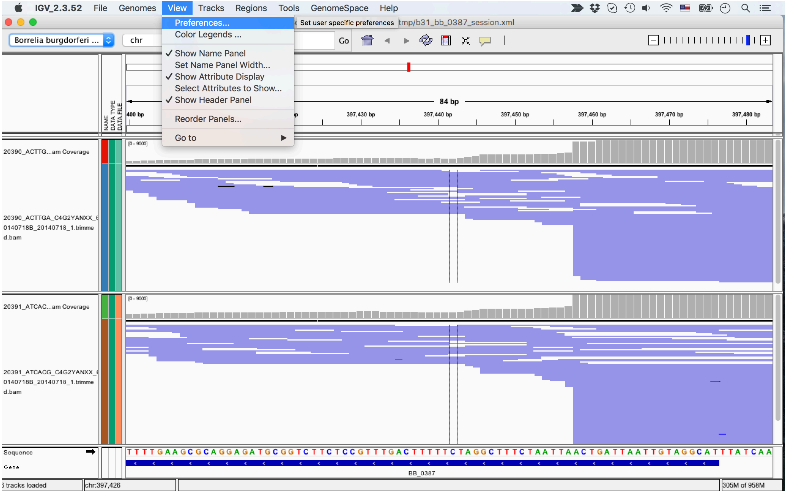
- Open session

IGV: SESSIONS



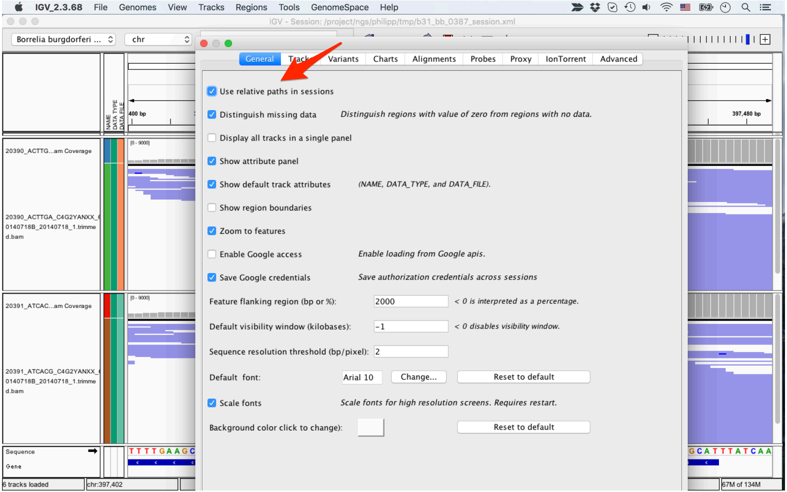
- Open session

IGV: SESSIONS



- Change settings to relative paths

IGV: SESSIONS



☑ 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 %): 2000 — *< 0 is interpreted as a percentage.*

Default visibility window (kilobases): -1 — *< 0 disables visibility window.*

Sequence resolution threshold (bp/pixel): 2

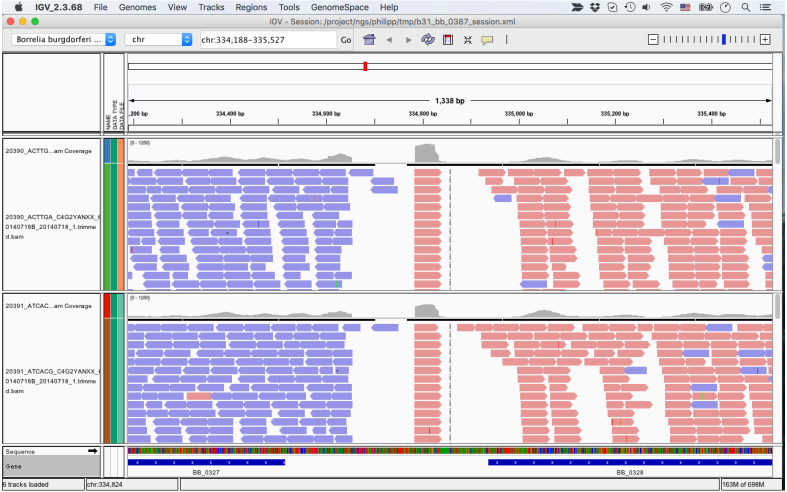
Default font: Arial 10 — Change... — Reset to default

☑ Scale fonts — *Scale fonts for high resolution screens. Requires restart.*

Background color click to change): — Reset to default

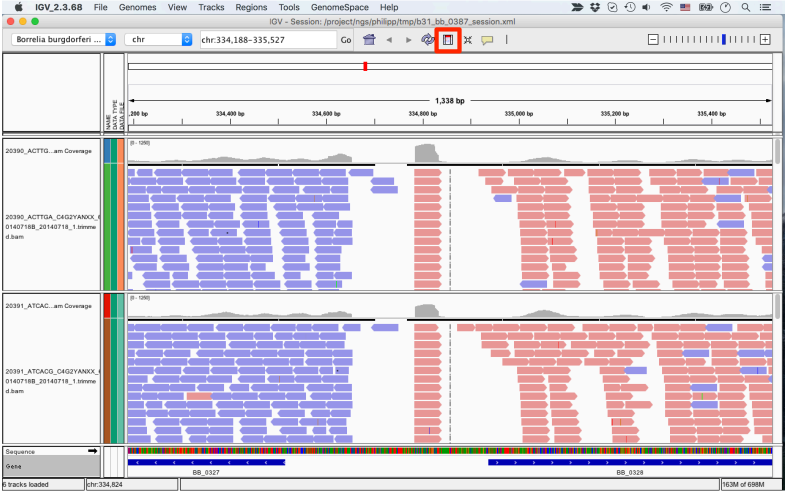
- Change settings to relative paths

IGV: REGIONS NAVIGATOR



- Allows to define regions of interest

IGV: REGIONS NAVIGATOR

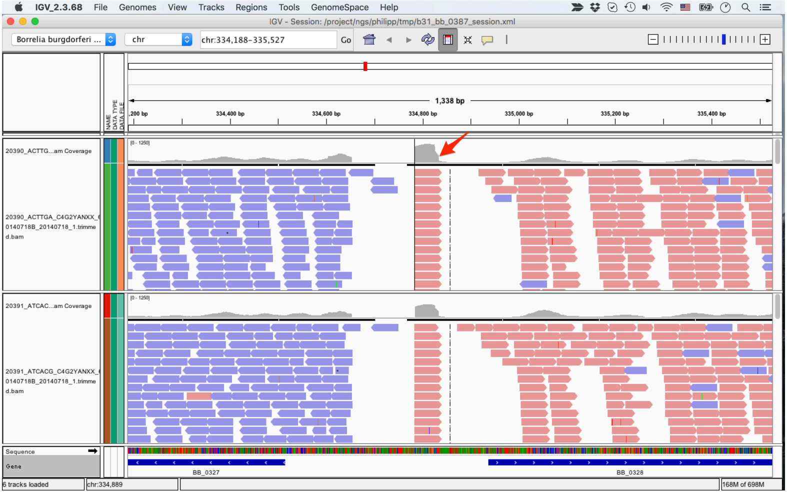


IGV - Session: /project/ngs/philipp/tmp/b31_bb_0387_session.xml

chr:334,188-335,527

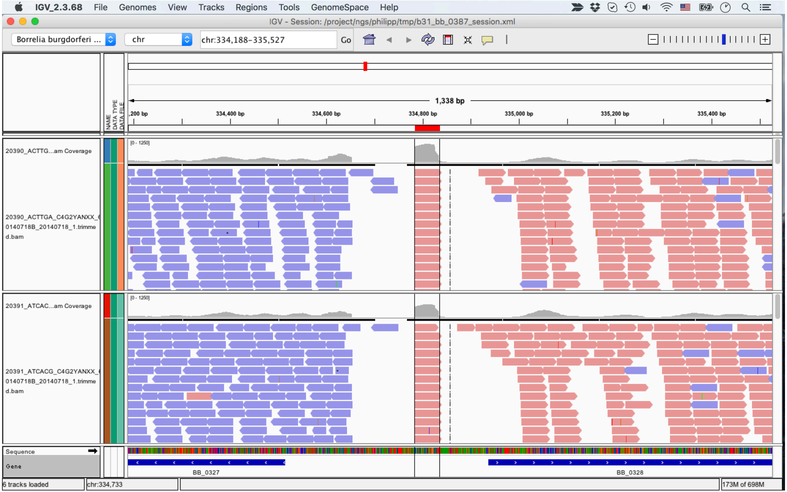
- Allows to define regions of interest

IGV: REGIONS NAVIGATOR



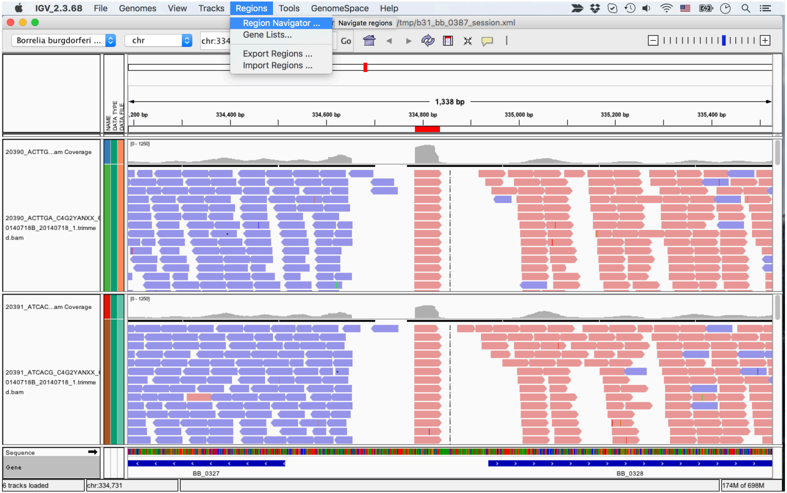
- Click twice to define region

IGV: REGIONS NAVIGATOR



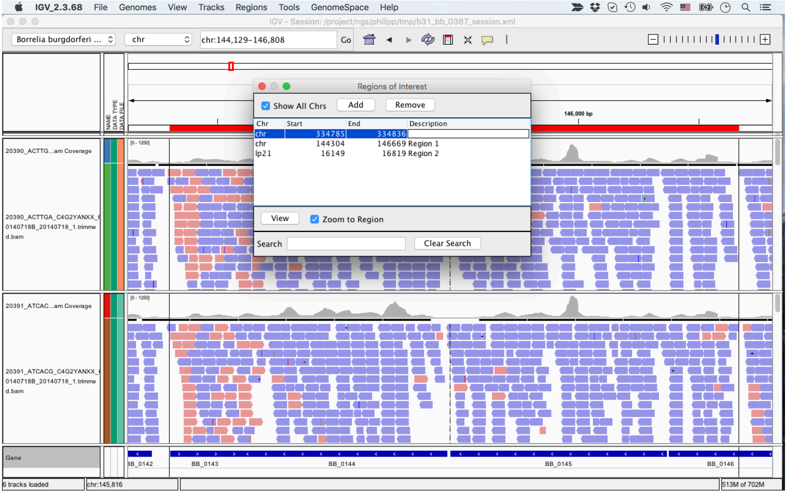
- Click twice to define region

IGV: REGIONS NAVIGATOR



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

IGV: REGIONS NAVIGATOR



IGV - Session: /project/ngs/philipp/tmp/b31_bb_0387_session.xml

Borrelia burgdorferi ... | chr | chr:144,129-146,808

20390_ACTTG...am Coverage

20390_ACTTGA_C4G2YANXX_0140718B_20140718_1.trimmed.bam

20391_ATCAC...am Coverage

20391_ATCACG_C4G2YANXX_0140718B_20140718_1.trimmed.bam

☑ Show All Chrs

Chr	Start	End	Description
chr	334785	334836	
chr	144304	146669	Region 1
lp21 | 16149 | 16819 | Region 2 |

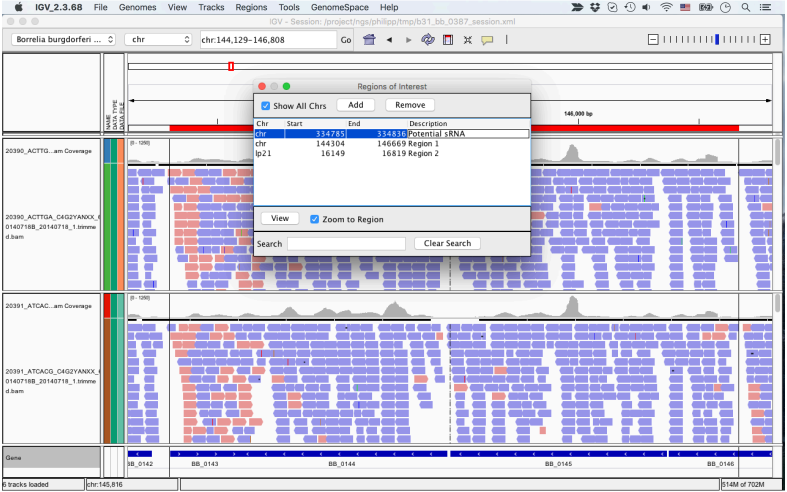
☑ Zoom to Region

Search

6 tracks loaded | chr:145,816 | 513M of 702M

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

IGV: REGIONS NAVIGATOR



Regions of Interest

☑ Show All Chrs Add Remove

Chr	Start	End	Description
chr	334785	334836	Potential sRNA
chr	144304	146669	Region 1
lp21	16149	16819	Region 2

View ☑ Zoom to Region

Search Clear Search

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

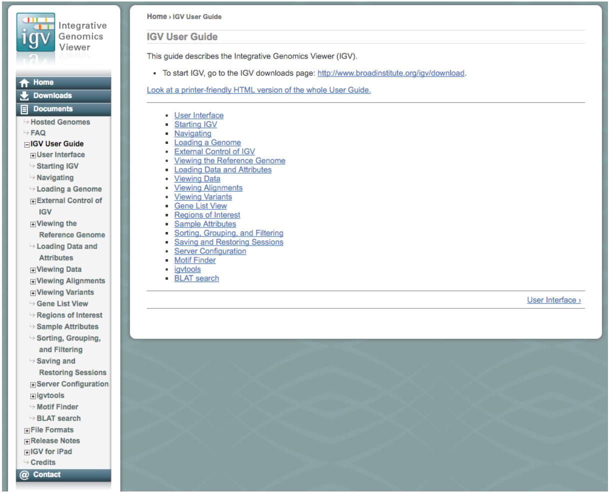
IGV: REGIONS NAVIGATOR

- Right click read bar to copy sequence to clipboard

IGV: REGIONS NAVIGATOR

- Export regions to BED file

IGV: DOCUMENTATION



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`