BV-BRC SARS-CoV-2 Genome Browser

User Guide

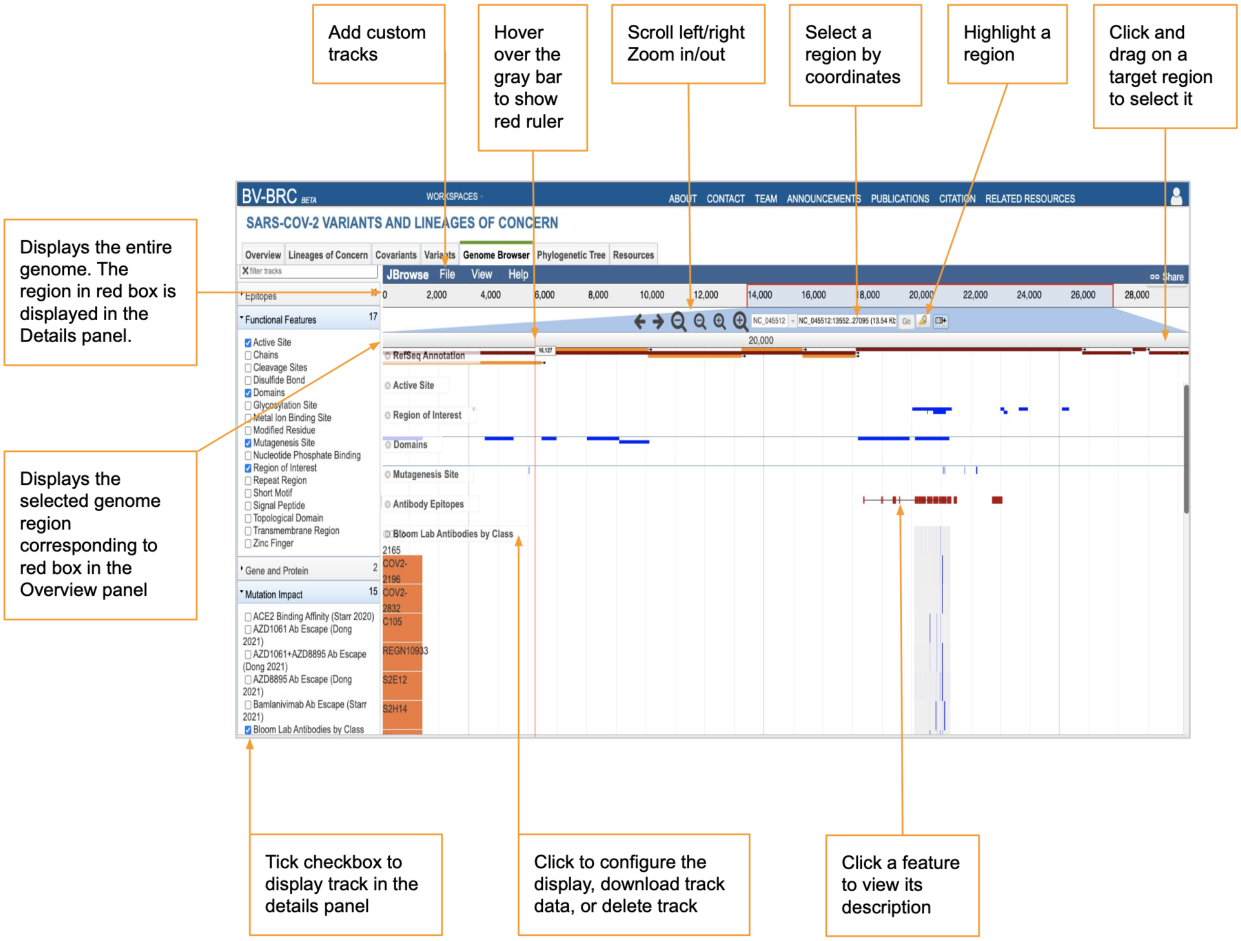
Last updated: April 29, 2021

Questions, comments: zwallace@jcvi.org

This document briefly describes how to use the genome browser on: https://www.bv-brc.org/view/VariantLineage/#view_tab=jbrowse

Methods on data curation and computation are described here: https://docs.google.com/document/d/1hsbcL0-7nDMJ6FiXaCTigf0gDUFIZcfp0h3A4K2kIeg/edit?usp=sharing

I. Genome Browser Navigation



Add custom tracks

Hover over the gray bar to show red ruler

Scroll left/right Zoom in/out

Select a region by coordinates

Highlight a region

Click and drag on a target region to select it

Displays the entire genome. The region in red box is displayed in the Details panel.

Displays the selected genome region corresponding to red box in the Overview panel

Tick checkbox to display track in the details panel

Click to configure the display, download track data, or delete track

Click a feature to view its description

II. Initial View

Upon accessing the “Genome Browser” tab within the resource, one will see the default view of the genome browser (Figure 1) which is built using Jbrowse (https://jbrowse.org/docs/installation.html). The genome browser displays a series of SARS-CoV-2 sequence features tracks for initial viewing by default: antibody epitope regions, active sites, domains, mutagenesis sites, regions of interest (sequence features like the receptor binding domain), and the reference annotation. Note that all the sequence features are mapped in the genetic background of the Wuhan-Hu-1 reference strain, which is the track labeled as Reference Sequence.

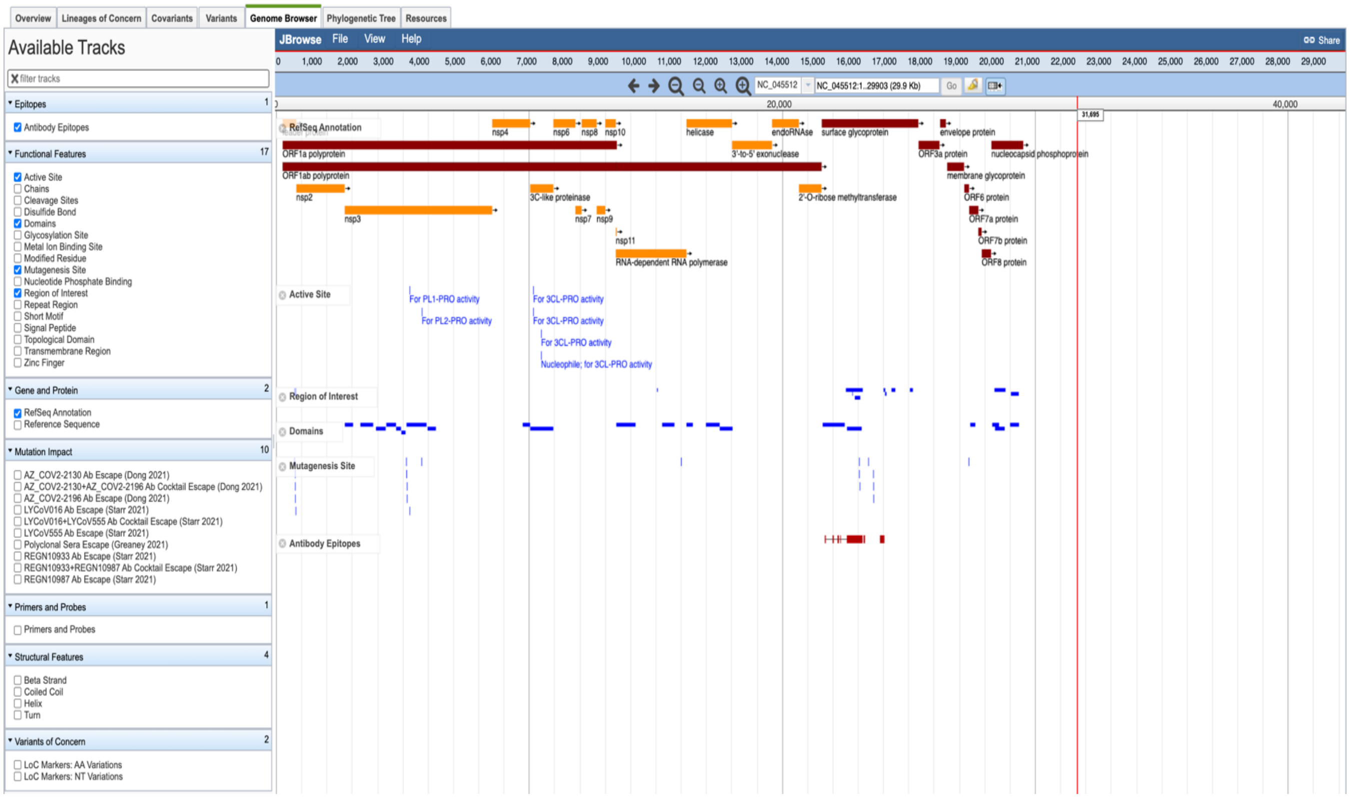
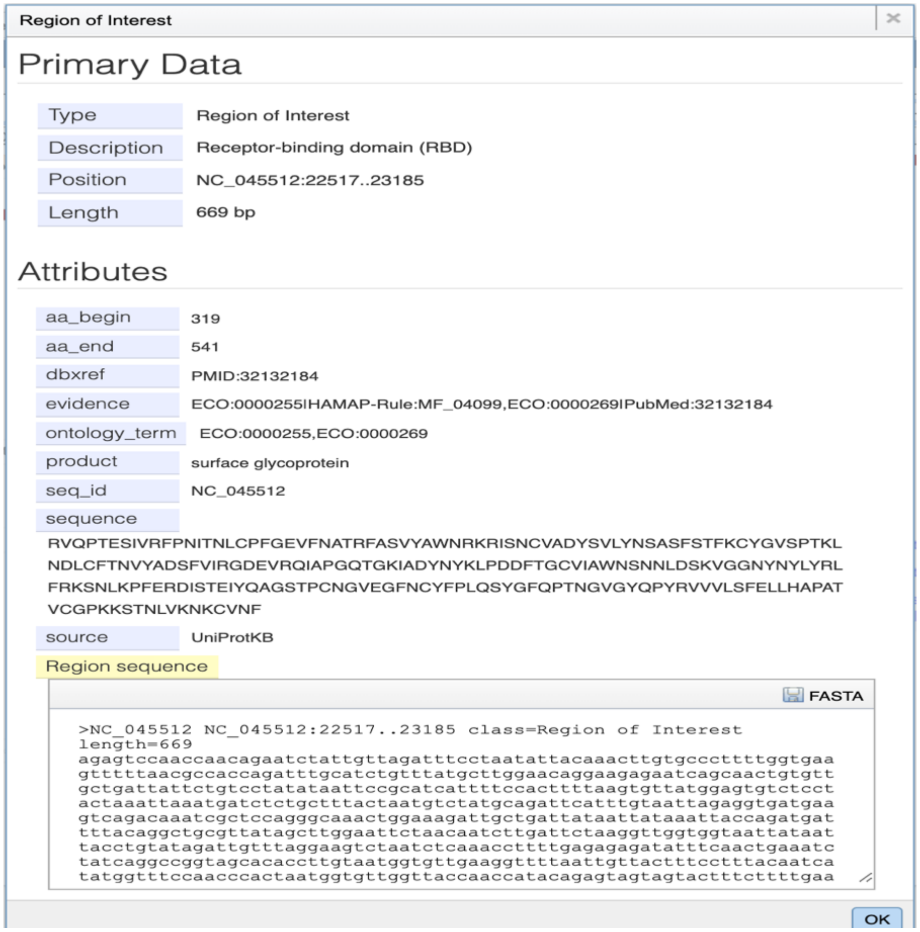


Figure 1: Initial view of the SARS-CoV-2 genome browser built with Jbrowse

III. Sequence Feature Tracks

By mousing over and clicking on any of the tracks, users can access the information that makes up the track. For example, hover over the RBD region within the Region of Interest tracks, click on it, and the user can retrieve an information box regarding that track (Figure 2A). Note, while the sequence features are denoting protein features, they’re being mapped to a genome; hence, we provide both the amino acid sequence and the genomic region sequence.



Primary Data

Type	Region of Interest
Description	Receptor-binding domain (RBD)
Position	NC_045512:22517..23185
Length	669 bp

Attributes

aa_begin	319
aa_end	541
dbxref	PMID:32132184
evidence	ECO:0000255|HAMAP-Rule:MF_04099,ECO:0000269|PubMed:32132184
ontology_term	ECO:0000255,ECO:0000269
product	surface glycoprotein
seq_id	NC_045512
sequence	

```
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKL
NDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRL
FRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPAT
VCGPKKSTNLVKNKCVNF
```

source	UniProtKB
Region sequence	

```
>NC_045512 NC_045512:22517..23185 class=Region of Interest
length=669
agagtccaaccaacagaatctattgttagatttcctaatattacaaacttgtgcccttttggtgaa
gtttttaacgccaccagatttgcatctgtttatgcttggaacaggaagagaatcagcaactgtgtt
gctgattattctgtcctatataattccgcatcattttccacttttaagtgttatggagtgtctcct
actaaattaaatgatctctgctttactaatgtctatgcagattcatttgtaattagaggtgatgaa
gtcagacaaatcgctccagggcaaactggaaagattgctgattataattataaattaccagatgat
tttacaggctgcgttatagcttggaattctaacaatcttgattctaaggttggtggtaattataat
tacctgtatagattgtttaggaagtctaatctcaaaccttttgagagagatatttcaactgaaatc
tatcaggccggtagcacaccttgtaatggtgttgaaggttttaattgttactttcctttacaatca
tatggtttccaacccactaatggtgttggttaccaaccatacagagtagtagtactttcttttgaa
```

Figure 2A: Information box for the RBD

Some of the tracks, such as Antibody Epitopes, are left in what's called a "collapsed" mode to prevent the track from taking up too much space in the genome browser. To expand the antibody epitopes track, click on the arrow next to the track name and select Display mode > normal, as shown in Figure 2B.

A range of sequence features track, including 17 functional features, 4 structure features, a track for primers and probes region, and the antibody epitopes. These tracks can be checked on or off depending on what the user is interested in exploring.

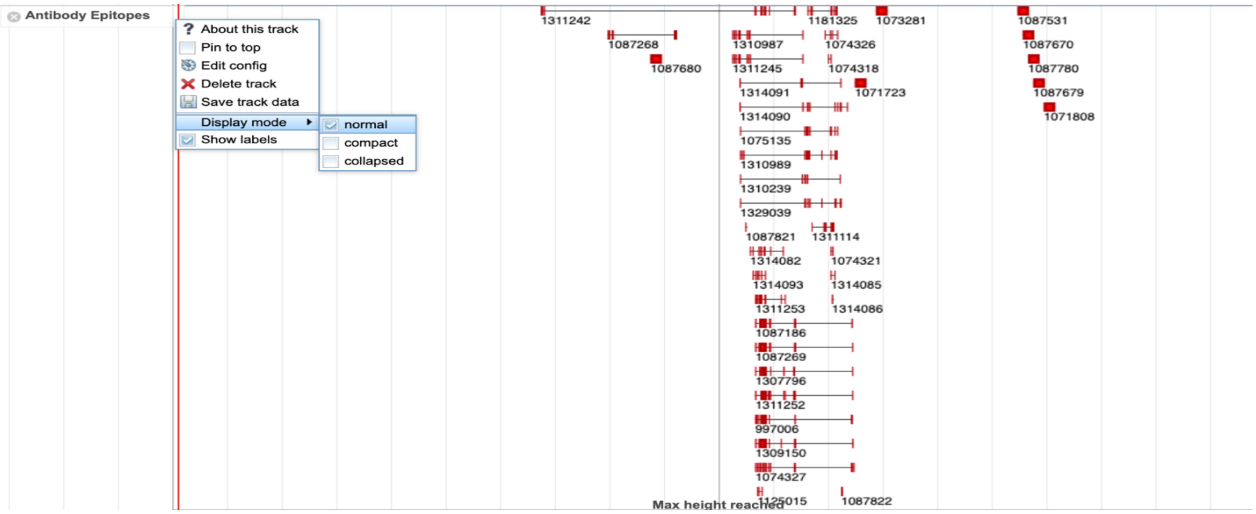


Figure 2B: Method for expanding certain tracks, such as Antibody Epitopes

IV. Variants of Concern Tracks

A list of variants with concerning properties has been compiled that exist within the circulating SARS-CoV-2 lineages typically defined by PANGO and the substitutions carried by these variants mapped to their positions on the genome for viewing in the genome browser. We refer to these as “LoC Markers” in the genome browser under the Variants of Concern category and offer the amino acid variants and nucleotide variants as two separate tracks. Each LoC Marker is meant to display a particular mutation and the lineage it is part of. Figure 3 shows the set of variants that exist in the spike glycoprotein and the various sequence features they overlap with.

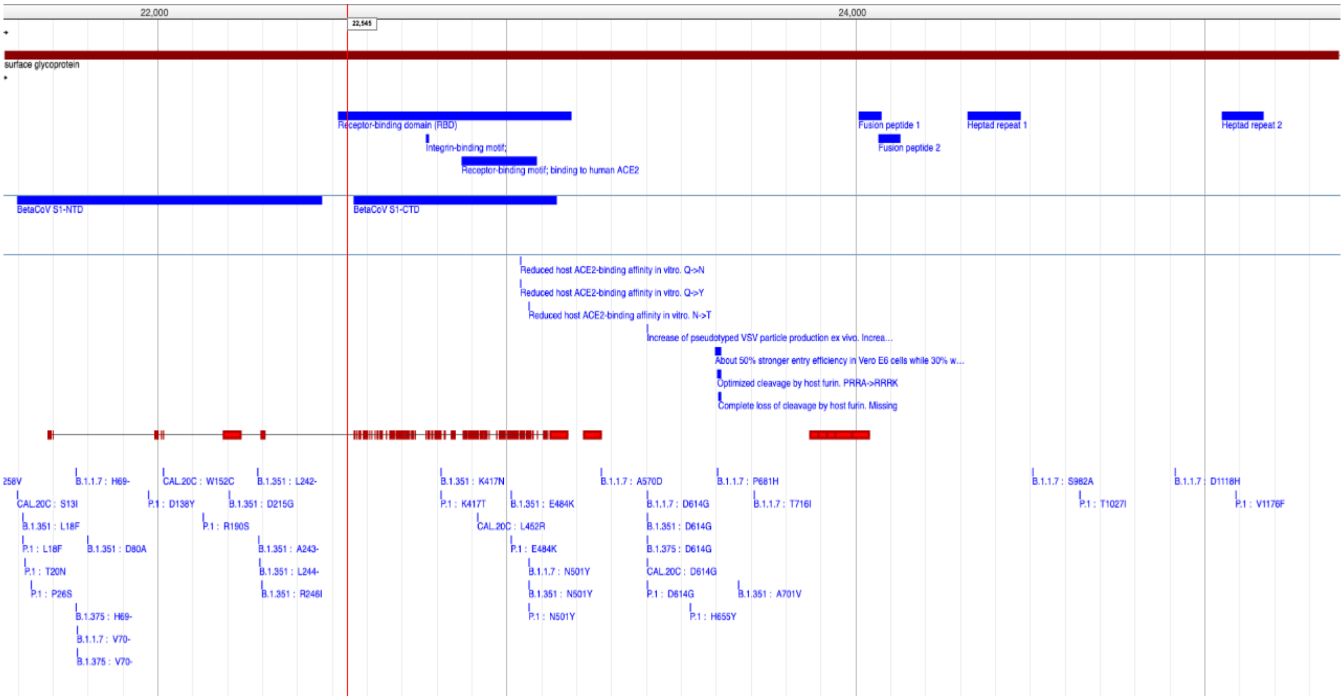


Figure 3: Display of the genome browser with the Variants of Concern track

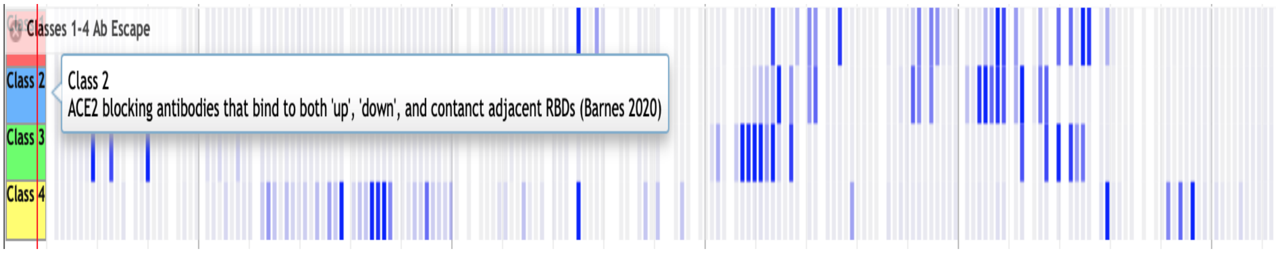
V. Mutation Impact Tracks

Jesse Bloom's Lab at the Fred Hutchinson Cancer Research Center have published a series of experiments quantifying how mutations across the RBD impact ACE2 and antibody binding to the spike protein. The methodology behind how these data were generated and then organized into tracks is described in the SOP document **here**. A series of overlaid bar graphs and heatmaps are provided to display which sites in the RBD may be most impacted by mutation. The default mutation impact track left for initial viewing, titled Classes 1-4 Ab Escape, is meant to summarize how mutations impact antibodies within a certain class, where a class is defined by the structure of the antibody epitopes (Barnes et al, 2020). The second heatmap mutation impact track, titled Bloom Lab Antibodies by Class, offers an expanded view of the Class 1-4 Ab Escape track by providing all the antibodies from the Bloom lab experiments that define each class. Note that tracks such as these will be continuously updated as more data become available.

The remaining mutation impact tracks display overlaid bar graphs that analyze the mutational impact towards antibody therapeutics, polyclonal sera, and Moderna vaccine elicited antibodies. The height of each bar denotes the "escape fraction", which, as defined by Bloom, is the quantity used to define antibody escape. Since each RBD site has data for all possible amino acid substitutions, we use maximum and median escape fraction values for constructing the overlaid bar graphs, where blue represents the maximum escape fraction and orange the median escape fraction for all substitutions. Figure 4A displays a few of these tracks for the therapeutic antibodies.

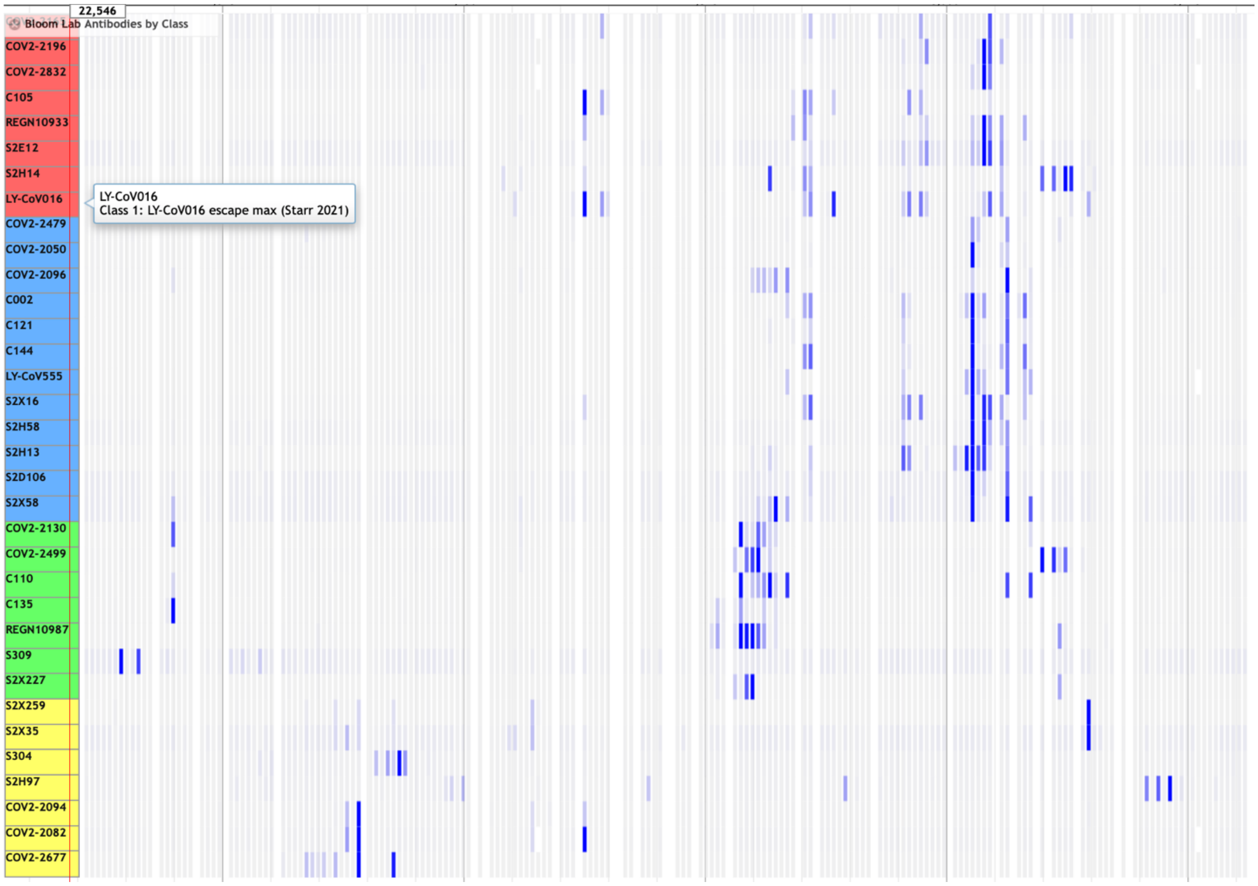
Figure 4A: Bloom mutation impact tracks overlapping with variants and sequence features

There are several ways to explore the mutation impact tracks further with the option of some information boxes and mouse hovering. As for the heatmap tracks that analyze the mutation impact by antibody class, hovering over the row names of the Classes 1-4 Ab Escape track allows users to identify the definition of the class according to Barnes et al., 2020 (Figure 4B). Likewise, hovering over the row names of the Bloom Lab Antibodies by Class track allows users to see the class assignment to the monoclonal antibodies as well as a reference to the first author of the antibody study (Figure 4C). Finally, to gain an information box about these tracks requires mousing over the arrow by the track name and clicking “About this track” (Figure 4D).



Class 2
ACE2 blocking antibodies that bind to both 'up', 'down', and contanct adjacent RBDs (Barnes 2020)

Figure 4B: Hovering over the row names of Classes 1-4 Ab Escape to see the class definition



LY-CoV016
Class 1: LY-CoV016 escape max (Starr 2021)

Figure 4C: Hovering over the row names of Bloom Lab Antibodies by Class

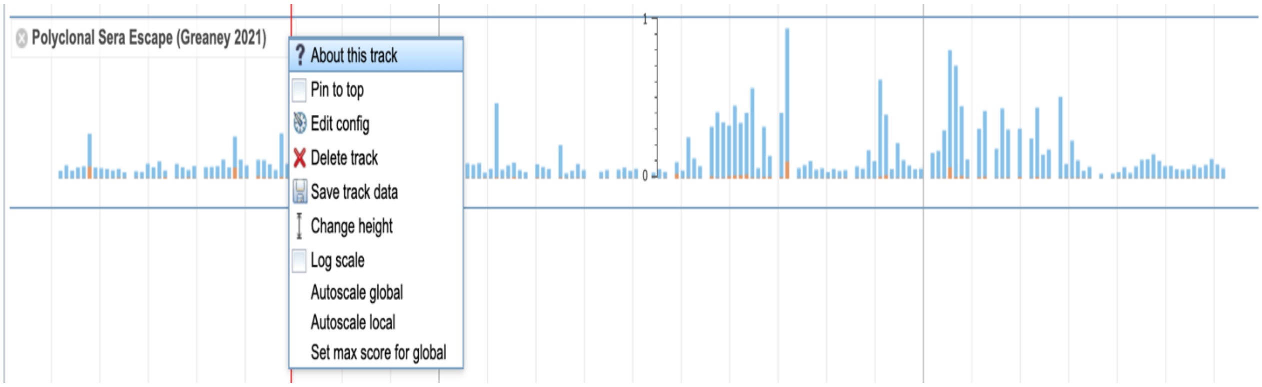


Figure 4C: Accessing the information box for the mutation impact tracks

VI. Additional Jbrowse Features

Navigating to genomic regions:

To navigate to a specific region in the genome, use the coordinates box in the top center of the display (Figure 5A). Note, these are genomic coordinates. To navigate to amino acid coordinates of proteins, the user needs to go about the genomic conversion computation themselves. We hope to soon implement a method to allow users to navigate to the genomic region with protein coordinates.

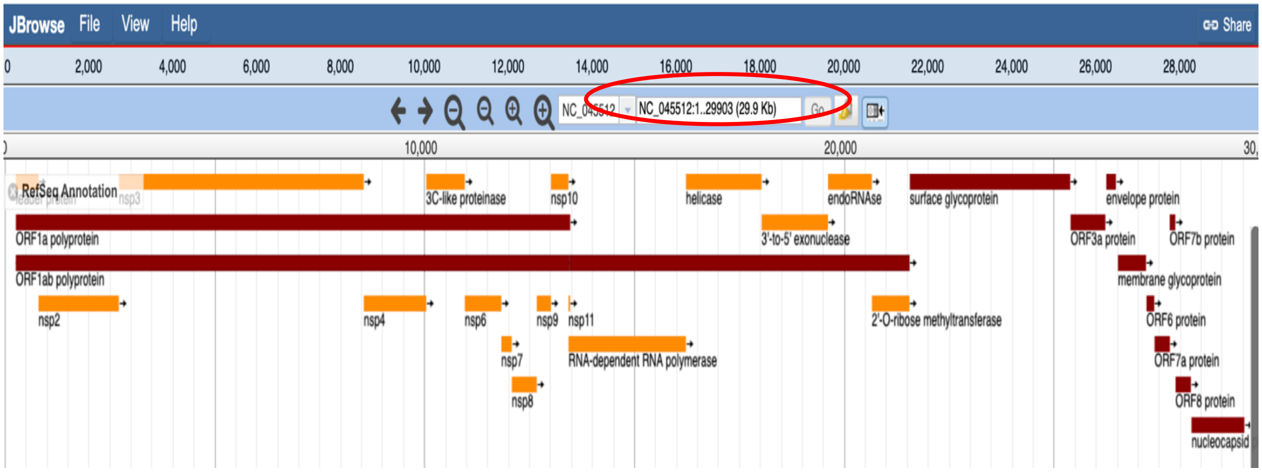


Figure 5A: Navigating to genomic regions

Removing track labels:

Each track in the genome browser has its name listed in the top left-hand side. The option to remove those track names for better visualization is accomplished by clicking the button shown in Figure 5B.

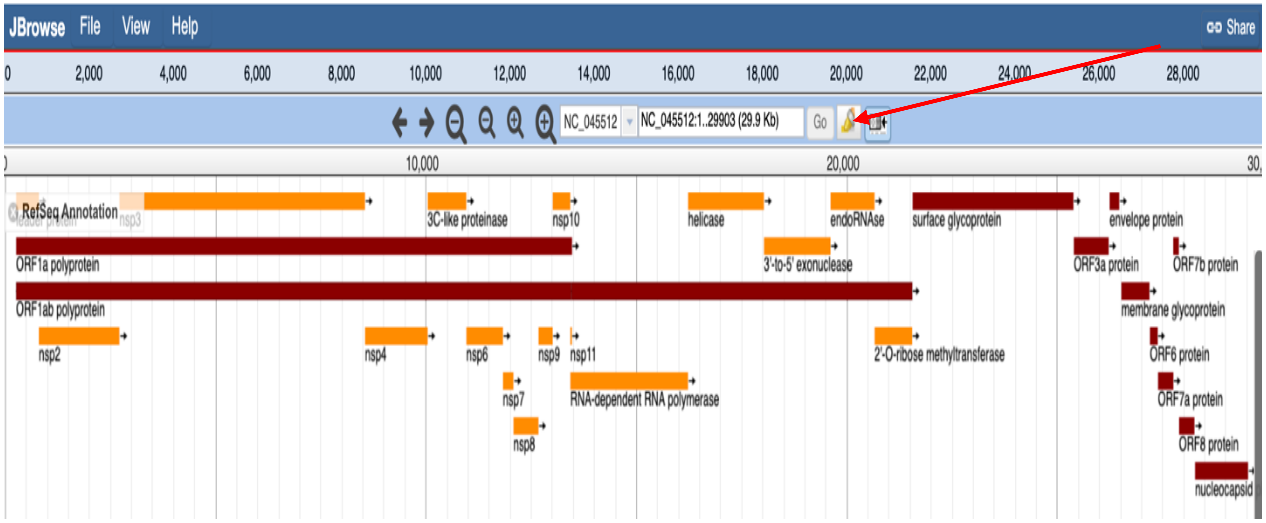


Figure 5B: Removing the track label in the genome browser

Loading in custom tracks:

Users may want to load in their own tracks to the genome browser in addition to what is already available. Select File then Open, and then load in tracks in various allowed formats, including GFF3 files, VCF files, FASTA files, BigWig, etc. (Figure 5C and 5D).

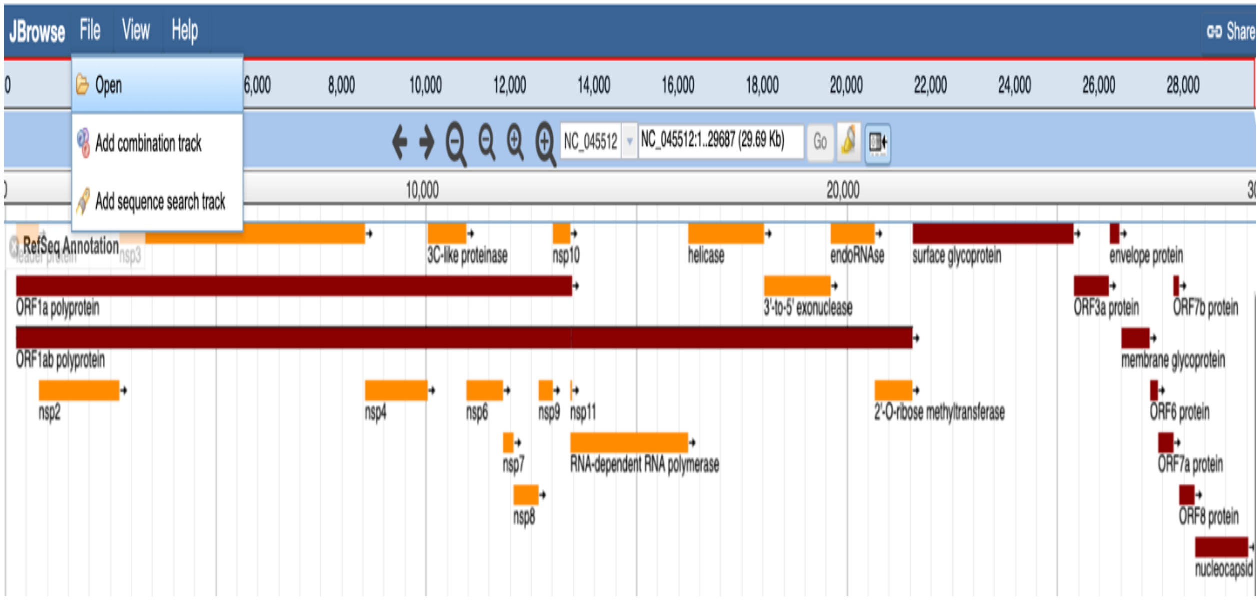
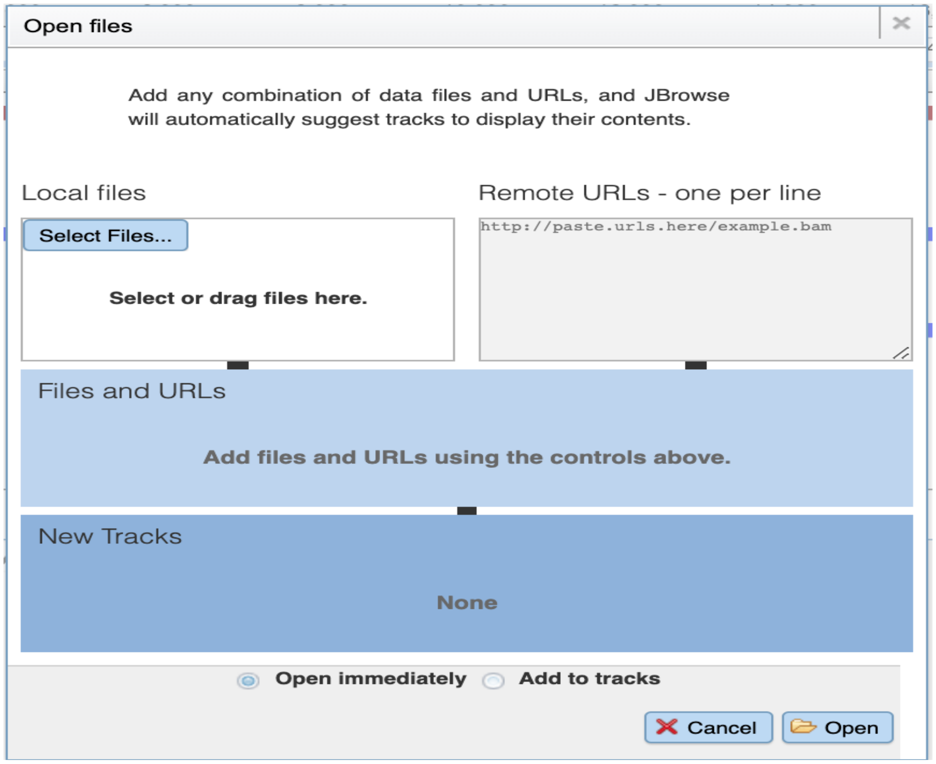


Figure 5C: Loading in custom tracks, part 1



Add any combination of data files and URLs, and JBrowse will automatically suggest tracks to display their contents.

Figure 5D: Loading in custom tracks, part 2

Downloading tracks from the genome browser:

Users may want to download the data that makes up a track for their own exploration and usage. Users have the option to download the entire track or just a portion of the track. To do so, click the arrow next to the track name > Save track data, then check off the appropriate “Region to save” (usually whole reference sequence), then check off the format (usually bedgraph for mutation impact tracks and GFF3 otherwise), choose the appropriate filename, then click Save (Figure 5E and 5F). Note, all files will need to be opened in a text editor.

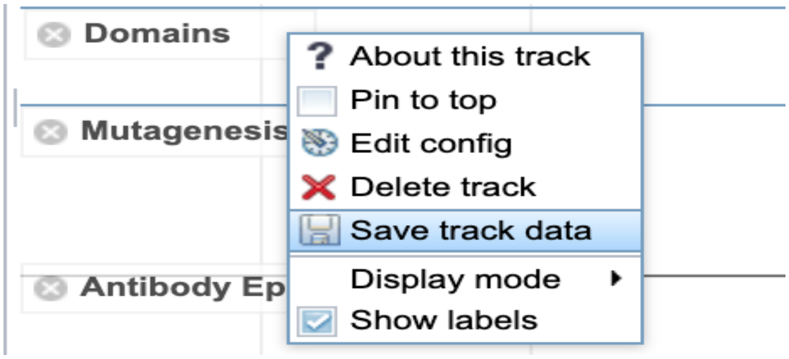
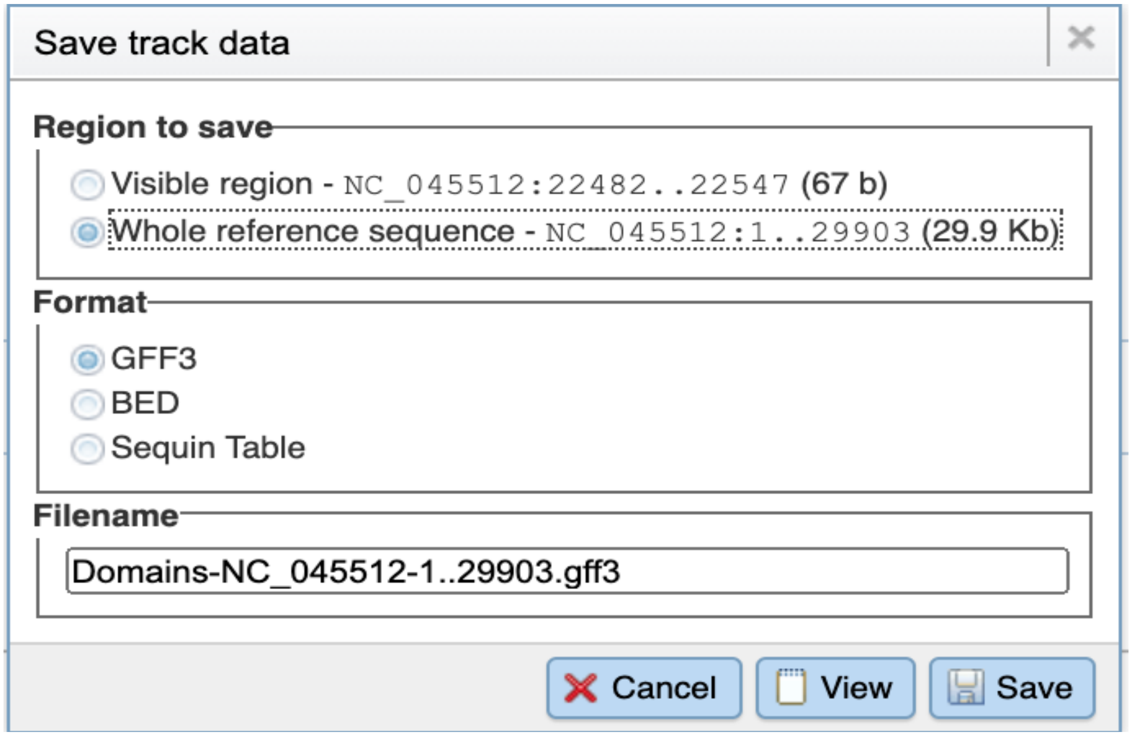


Figure 5E: Downloading tracks from the genome browser, part 1



Save track data

Region to save

○ Visible region - `NC_045512:22482..22547` (67 b)

● Whole reference sequence - `NC_045512:1..29903` (29.9 Kb)

Format

● GFF3

○ BED

○ Sequin Table

Filename

Domains-NC_045512-1..29903.gff3

Figure 5F: Downloading tracks from the genome browser, part 2