



LEARN PYTHON & R FOR BIOINFORMATICS

Introduction:

To retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track, Table Browser is the program for that which we will use for the retrieval and annotation of SARS-CoV-2 Viral Genome. We should know what is genome annotation; Genome Annotation is the process to identify all of the locations of genes and coding regions in a genome.

Steps:

Retrieval of the Genome

- First go to the Table Browser which you can access from [here](#).
- Select the parameters according to the given table for SARS-CoV-2 Genome Retrieval:

Parameters	Options
Clade	Viruses
Genome	SARS-CoV-2
Assembly	Select the Recent Entry
Group	Genes and Gene Predictions
Track	NCBI genes
Table	ncbiGene
Region	Genome
Output Format	GTF or BED
File Type Returned	Plain Text

NOTE :

These chosen options are specifically for SARS-CoV-2 Viral Genome, but you can tweak the choices to get the data based upon your need.

- Now click on the 'Get Output'.
- The text file will be downloaded and represented which you can simply analyze.

Annotation of the Genome

(A)

- The downloaded file can be open on any editor.
- For this, first go back to the Table Browser's parameter page and provide in the 'Output File' box, the format name next to the address of your output file.
- Type '.gtf' next to your output file and click on the 'Get Output'.
- Drag your downloaded file to the editor *e.g. Visual Studio Code*.
- The Genetic structure of the genome will be displayed on the editor.
- You can Annotate the locations of the genes and coding regions of the SARS-CoV-2 Viral Genome.

(B)

- Same as before, you can go back to Table Browser's parameter page and change the format type.
- This time, select BED (Browser Extensible Data).
- In 'Output File', next to the file address, add '.bed'.
- Now click on the 'Get Output'.
- A parameter page will open up, where you have to make sure to select 'Whole Gene' option.
- Click on the 'Get BED' and the BED file will be created.
- You can visualize this file just like GTF file.

Statistics/Summary of Genes

- You must have noticed an option along with 'Get Output' on the Table Browser; the 'Summary/Statistics' option.
- You can select it to get the statistics of your gene data which is a summary of your gene.

Summary:

In this video, we learnt about the retrieval of the genome data by taking the example of the retrieval and the annotation of SARS-CoV-2. We got to learn about different annotation files (GFF, GFF3, GTF) and what is the major difference between them.

Learn Bioinformatics tools & bioinformatics programming.

<https://www.biocode.ltd/plans-pricing>