# Step-by-Step Guide for the SARS-CoV-2 Primer Mutation Analysis

## Prerequisite

The following software need to be installed on the computer:

1. Any distribution of Python 3.7 (e.g., Anaconda 3)

Installation guide: <https://docs.anaconda.com/anaconda/install/windows/>

1. Biopython 1.76

Installation guide: <https://biopython.org/wiki/Packages>

1. NCBI BLAST+ 2.10.0

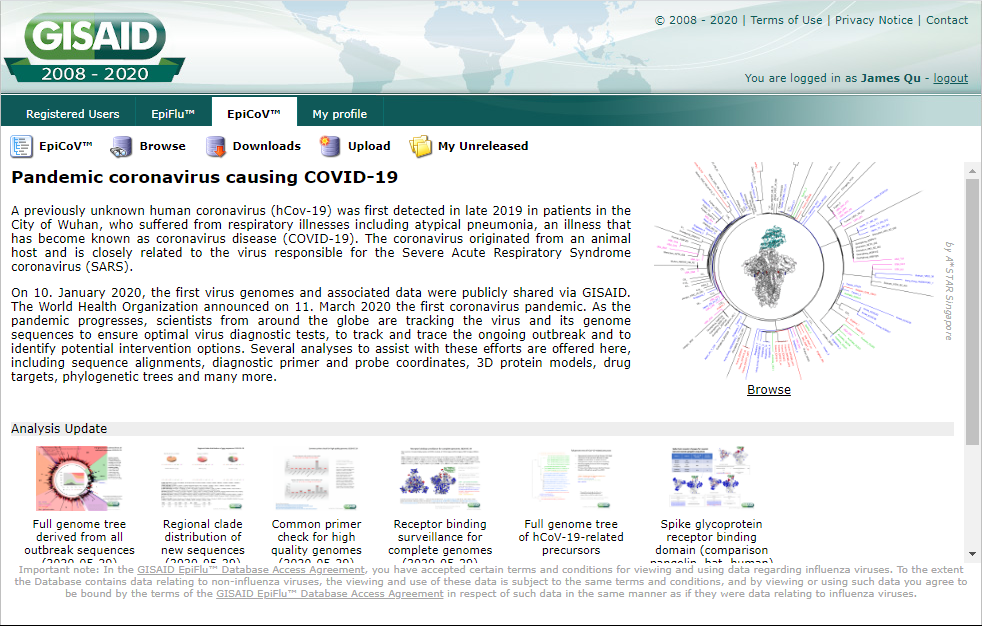
Installation guide: <https://www.ncbi.nlm.nih.gov/books/NBK52637/>

A user account needs to be registered from GISAID: <https://www.gisaid.org/>

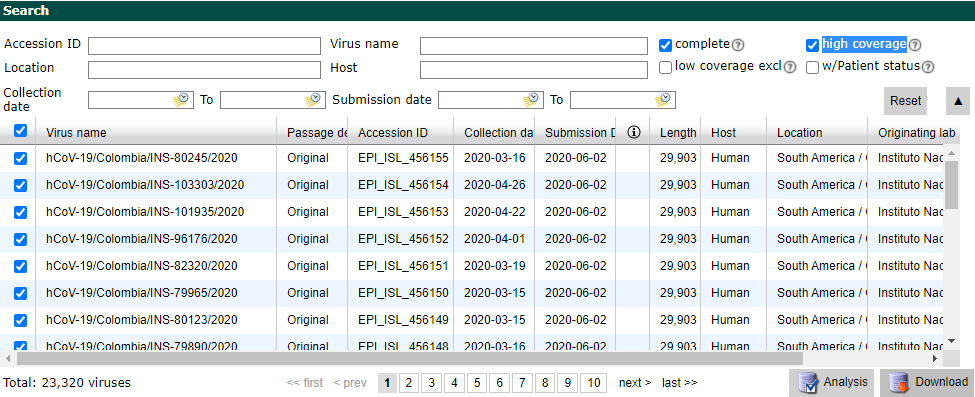
## Step-by-Step Guide

### Download SARS-CoV-2 sequencing data from GISAID

1. Login GISAID from <https://www.gisaid.org/>
2. Click “Browse”.



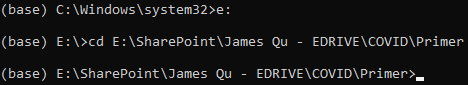
1. Tick “complete” and “high coverage”, tick all entries, and then click “Download”.



1. Click “Download” again, rename the file to “gisaid\_hcov-19.fasta”, save the file to your working folder.

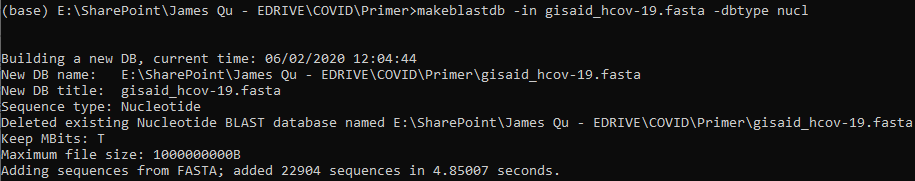
### Create SARS-CoV-2 Blast Database

1. Run any command-line interface (e.g., Anaconda Prompt, or Windows Commands) as administrator.
2. Using “cd” command to relocate to your working folder.



1. Create Blast Database with the following command:

*makeblastdb -in gisaid\_hcov-19.fasta -dbtype nucl*



### Blast primer sequences against SARS-CoV-2 Database

1. Primers sequences have to be saved in FASTA format in a file named “primers.fasta”. All spaces in the primer name have to be replaced with underscores.
2. Blast primer sequences against SARS-CoV-2 Database with the following command:

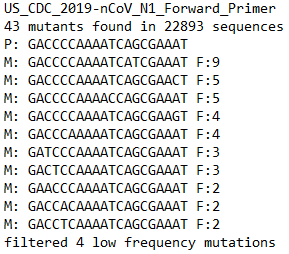
*blastn -task blastn-short -db gisaid\_hcov-19.fasta -query primers.fasta -out primerblast.xml -outfmt 5 -max\_target\_seqs 99999 -max\_hsps 1*

### Run the custom python script for primer mutation analysis

1. Run the custom python script with the following command:

*Python primer\_check.py>primer\_check\_results.txt*

1. The running time depends on the number of primers to be analyzed, and the number of mutants found. For instance, analysis for CDC N1, CDC N2, China CDC ORF1ab, Berlin E, and Berlin RdRp all five assays together will take about 18 minutes. Analysis for China CDC N gene assay will take about one hour since this assay has abundance of mutations.
2. Below is an example entry of the output file:



P stands for the original primer sequence. M stands for the mutated sequences. F stands for the number of times of the mutation observed in the GISAID database.

1. You can parse the output file to excel or any text operating tools. Recommended font will be “Courier New”, as this is a monospaced font and it will be much easy to align the bases and spot mutations.

## Annex: Source Code of primer\_check.py

