Why this workshop?

- Learning how to download data from public databases
- Understanding how to make use of this data
- Streamlining analysis
- Sharing interactive reports with collaborators
- Add-on training resource for classes

Overview

Research Question

- Data driven
- Hypothesis driven
- Product driven

Data collection

- Experiments
- Field Observation
- Museums
- Surveys
- Literature
- Database

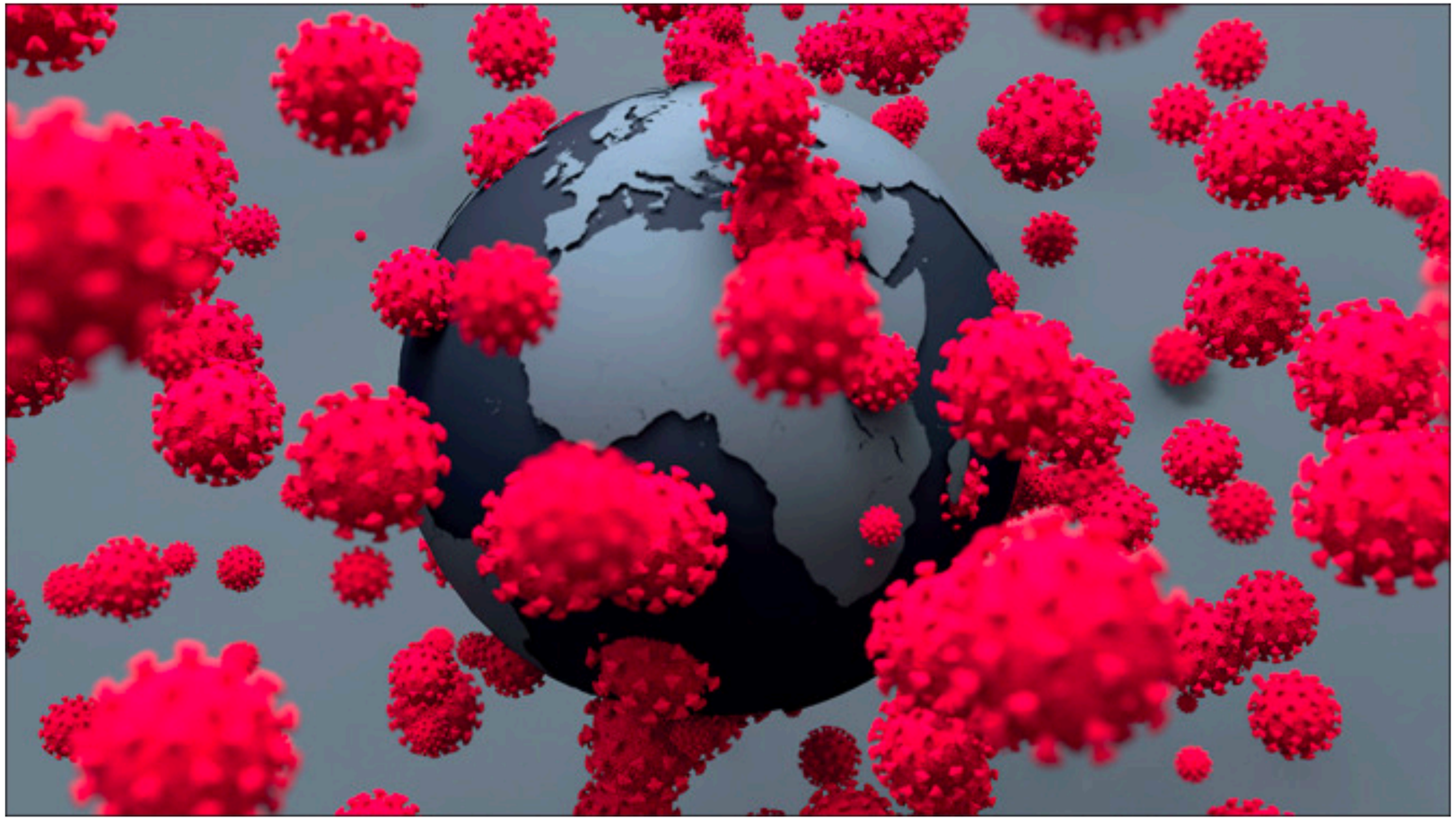
Analysis

Reports

- Scientific Paper
- Presentation
- Interactive reports

Research Question

How do we track SARS-CoV2 variants?



Workshop Plan

This is a three part series with tutorials on:

Workshop 1: Data: How to download data using public database

- NCBI tools
- Example dataset: SARS-CoV2 nucleotide sequences
- Fasta format

Workshop 2: **Analyses:** Using this data for tracking viral evolution

- Transferring sequences to HiPerGator
- Using CLUSTAL for sequence alignment
- Using RAxML to build a tree

Workshop 3: Reports: Make an interactive data report

- Using RShiny and FlexDashboard
- Loading and formatting data in RMarkdown
- Creating an interactive Visualization Dashboard

Material and Data Availability

All workshop material can be downloaded from GitHub:

- Workshop slides
- Tutorials
- Dataset
- RMarkdown and scripts
- Link to GitHub Page: http://NatyaHans.github.io/Workshops

Source: NatyaHans/Workshops

SCAN ME

March 1st, 2022

Workshop 1

Downloading SARS-COV2 genomic sequences from NCBI

Instructor: Natya Hans

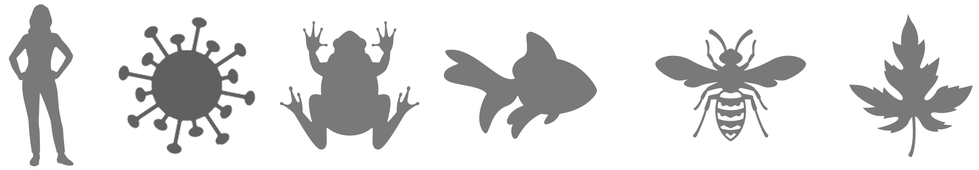
Workshop 1 plan

NCBI

- National Center for Biotechnology Information
- Part of the National Library of Medicine (NLM) and National institute of health (NIH)
- Develops automated systems for storage, retrieval, and analysis of genetic and bio-molecular information
- Develops tools and softwares for studying genetic data, molecules, structure and function
- Hosts GenBank since 1992

What is GenBank

- Genetic sequence database
- Multiple organisms
- Publicly Available
- Annotated internationally



Genome Record for SARS-CoV2

Accession Number: NC_045512

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

FASTA Graphics

Go to:

```
LOCUS       NC_045512              29903 bp ss-RNA     linear   VRL 18-JUL-2020
DEFINITION  Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1,
            complete genome.
ACCESSION   NC_045512
VERSION     NC_045512.2
DBLINK      BioProject: PRJNA485481
KEYWORDS    RefSeq.
SOURCE      Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
  ORGANISM  Severe acute respiratory syndrome coronavirus 2
            Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
            Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae;
            Betacoronavirus; Sarbecovirus.
REFERENCE   1  (bases 1 to 29903)
  AUTHORS   Wu,F., Zhao,S., Yu,B., Chen,Y.M., Wang,W., Song,Z.G., Hu,Y.,
            Tao,Z.W., Tian,J.H., Pei,Y.Y., Yuan,M.L., Zhang,Y.L., Dai,F.H.,
            Liu,Y., Wang,Q.M., Zheng,J.J., Xu,L., Holmes,E.C. and Zhang,Y.Z.
  TITLE     A new coronavirus associated with human respiratory disease in
            China
  JOURNAL   Nature 579 (7798), 265-269 (2020)
   PUBMED   32015508
  REMARK    Erratum:[Nature. 2020 Apr;580(7803):E7. PMID: 32296181]
REFERENCE   2  (bases 13476 to 13503)
  AUTHORS   Baranov,P.V., Henderson,C.M., Anderson,C.B., Gesteland,R.F.,
            Atkins,J.F. and Howard,M.T.
  TITLE     Programmed ribosomal frameshifting in decoding the SARS-CoV genome
  JOURNAL   Virology 332 (2), 498-510 (2005)
   PUBMED   15680415
REFERENCE   3  (bases 29728 to 29768)
  AUTHORS   Robertson,M.P., Igel,H., Baertsch,R., Haussler,D., Ares,M. Jr. and
            Scott,W.G.
  TITLE     The structure of a rigorously conserved RNA element within the SARS
            virus genome
  JOURNAL   PLoS Biol. 3 (1), e5 (2005)
   PUBMED   15630477
REFERENCE   4  (bases 29609 to 29657)
  AUTHORS   Williams,G.D., Chang,R.Y. and Brian,D.A.
  TITLE     A phylogenetically conserved hairpin-type 3' untranslated region
            pseudoknot functions in coronavirus RNA replication
```

MORE FEATURES

5’ Untranslated region

Coding DNA Sequence

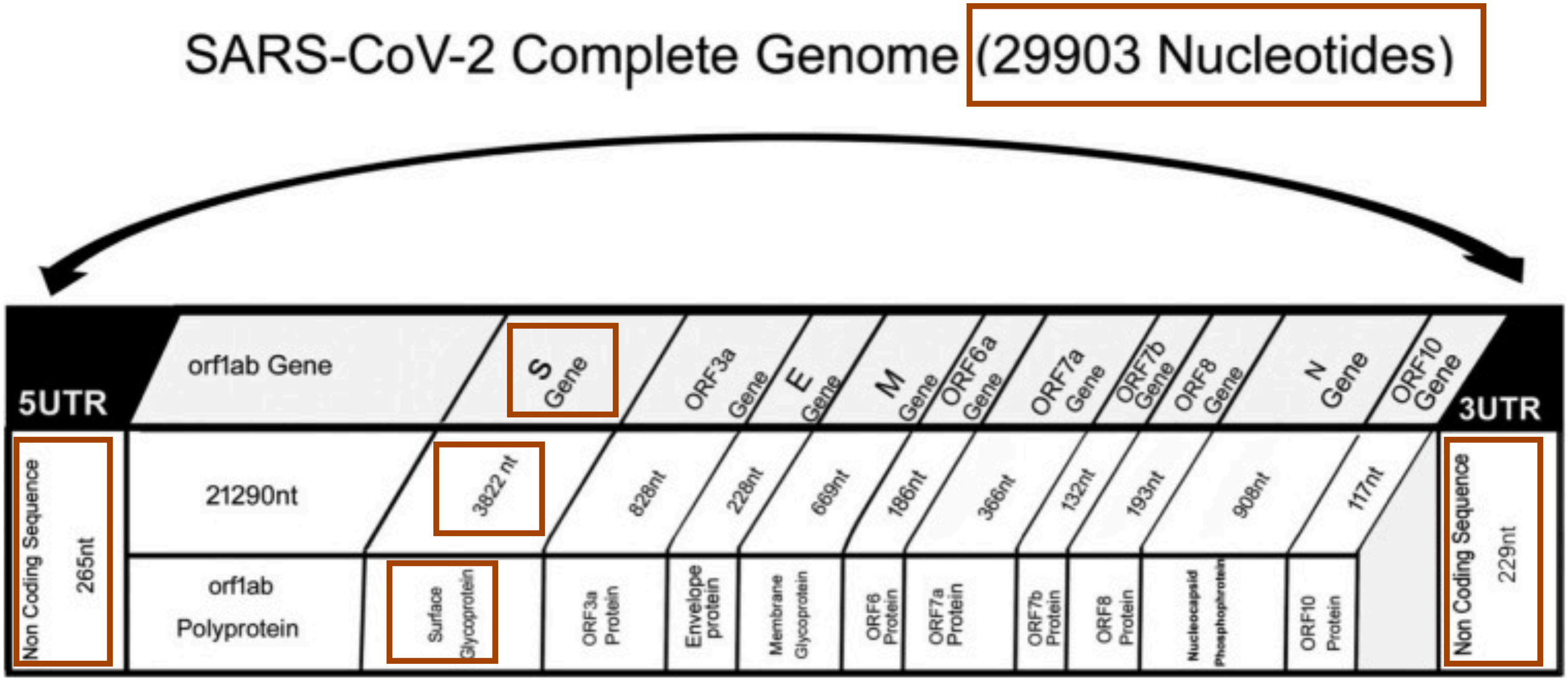
Start .. End

```
                     COMPLETENESS: full length.
     FEATURES             Location/Qualifiers
          source          1..29903
                          /organism="Severe acute respiratory syndrome coronavirus
                          2"
                          /mol_type="genomic RNA"
                          /isolate="Wuhan-Hu-1"
                          /host="Homo sapiens"
                          /db_xref="taxon:2697049"
                          /country="China"
                          /collection_date="Dec-2019"
          5'UTR           1..265
          gene            266..21555
                          /gene="ORF1ab"
                          /locus_tag="GU280_gp01"
                          /db_xref="GeneID:43740578"
          CDS             join(266..13468,13468..21555)
                          /gene="ORF1ab"
                          /locus_tag="GU280_gp01"
                          /ribosomal_slippage
                          /note="pplab; translated by -1 ribosomal frameshift"
                          /codon_start=1
                          /product="ORF1ab polyprotein"
                          /protein_id="YP_009724389.1"
                          /db_xref="GeneID:43740578"
                          /translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQ
                          HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE
                          TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN
                          WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ
                          LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP
                          LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG
                          DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG
                          LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL
                          LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN
```

```
ORIGIN
        1 attaaaggtt tataccttcc caggtaacaa accaaccaac tttcgatctc ttgtagatct
       61 gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcatgc ttagtgcact
      121 cacgcagtat aattaataac taattactgt cgttgacagg acacgagtaa ctcgtctatc
      181 ttctgcaggc tgcttacggt ttcgtccgtg ttgcagccga tcatcagcac atctaggttt
      241 cgtccgggtg tgaccgaaag gtaagatgga gagccttgtc cctggtttca acgagaaaac
      301 acacgtccaa ctcagtttgc ctgttttaca ggttcgcgac gtgctcgtac gtggctttgg
      361 agactccgtg gaggaggtct tatcagaggc acgtcaacat cttaaagatg gcacttgtgg
      421 cttagtagaa gttgaaaaag gcgttttgcc tcaacttgaa cagccctatg tgttcatcaa
      481 acgttcggat gctcgaactg cacctcatgg tcatgttatg gttgagctgg tagcagaact
      541 cgaaggcatt cagtacggtc gtagtggtga gacacttggt gtccttgtcc ctcatgtggg
      601 cgaaatacca gtggcttacc gcaaggttct tcttcgtaag aacggtaata aaggagctgg
      661 tggccatagt tacggcgccg atctaaagtc atttgactta ggcgacgagc ttggcactga
      721 tccttatgaa gattttcaag aaaactggaa cactaaacat agcagtggtg ttacccgtga
      781 actcatgcgt gagcttaacg gaggggcata cactcgctat gtcgataaca acttctgtgg
      841 ccctgatggc taccctcttg agtgcattaa agaccttcta gcacgtgctg gtaaagcttc
      901 atgcactttg tccgaacaac tggactttat tgacactaag aggggtgtat actgctgccg
      961 tgaacatgag catgaaattg cttggtacac ggaacgttct gaaaagagct atgaattgca
     1021 gacacctttt gaaattaaat tggcaaagaa atttgacacc ttcaatgggg aatgtccaaa
     1081 ttttgtattt cccttaaatt ccataatcaa gactattcaa ccaagggttg aaaagaaaaa
     1141 gcttgatggc tttatgggta gaattcgatc tgtctatcca gttgcgtcac caaatgaatg
```

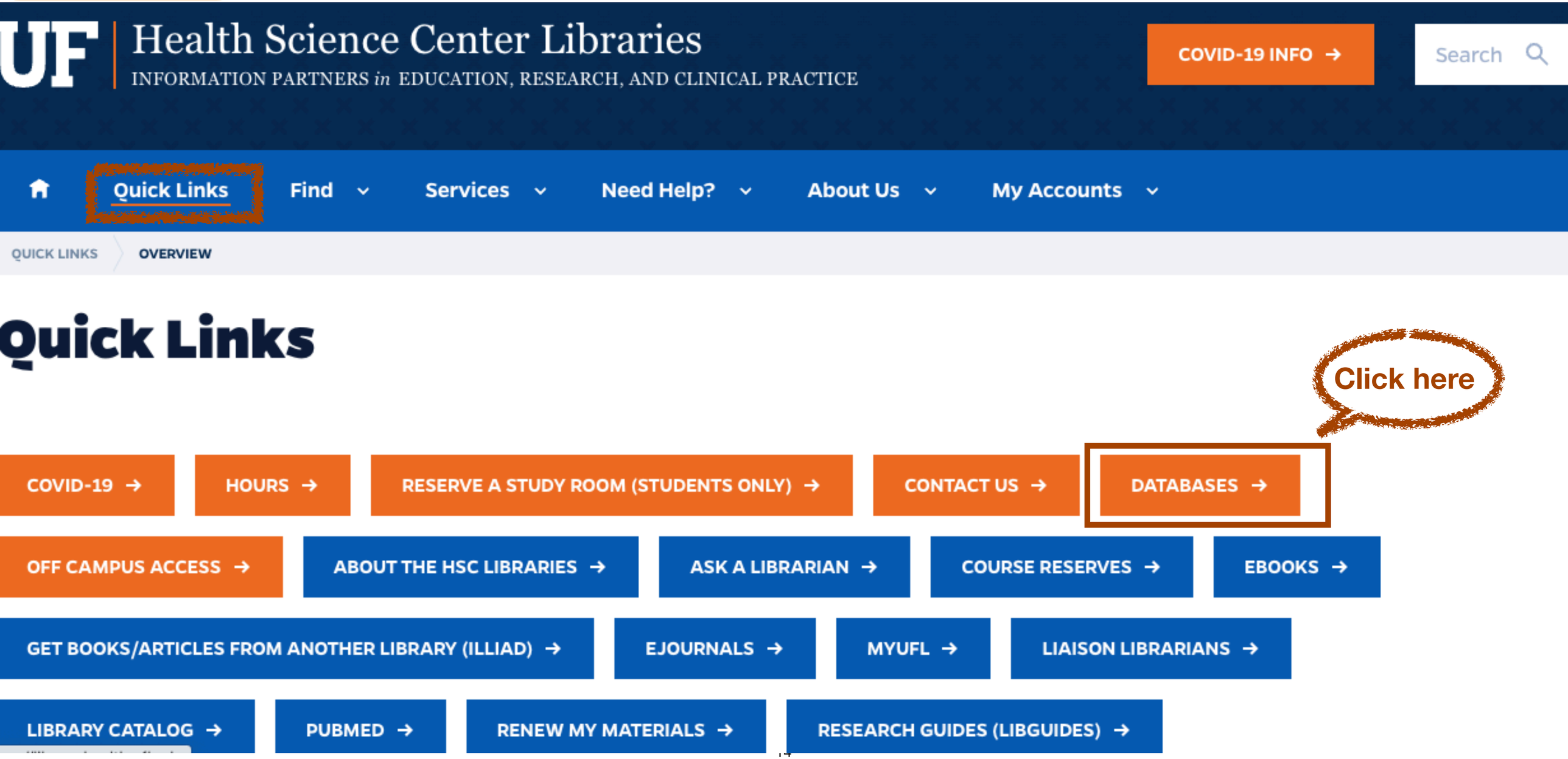
Accession Number: NC_045512

For full Genomic Record: https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2



Structure of the SARS-CoV-2 genome. (*Khailany et al., 2020*)

What is NCBI


UF
Health Science Center Libraries
INFORMATION PARTNERS in EDUCATION, RESEARCH, AND CLINICAL PRACTICE
COVID-19 INFO →
Search
Quick Links
Find
Services
Need Help?
About Us
My Accounts
QUICK LINKS
OVERVIEW
Quick Links
Click here
COVID-19 →
HOURS →
RESERVE A STUDY ROOM (STUDENTS ONLY) →
CONTACT US →
DATABASES →
OFF CAMPUS ACCESS →
ABOUT THE HSC LIBRARIES →
ASK A LIBRARIAN →
COURSE RESERVES →
EBOOKS →
GET BOOKS/ARTICLES FROM ANOTHER LIBRARY (ILLIAD) →
EJOURNALS →
MYUFL →
LIAISON LIBRARIANS →
LIBRARY CATALOG →
PUBMED →
RENEW MY MATERIALS →
RESEARCH GUIDES (LIBGUIDES) →

UF | Health Science Center Libraries
INFORMATION PARTNERS *in* EDUCATION, RESEARCH, AND CLINICAL PRACTICE

COVID-19 INFO →

Search

Quick Links | Find | Services | Need Help? | About Us | My Accounts

FIND > DATABASES

Find Overview
- Library Catalog
- Databases
  - HSCL databases
  - All library databases (A-Z list)
- eBooks
- eJournals
- Research Guides (LibGuides)

Databases

Quick Picks from Your Librarians

Access Medicine	Access Pharmacy	BIOSIS	CABI/Global Health	CINAHL
Clinical Pharmacology	Cochrane Library	Dissertations & Theses Global (ProQuest)	EBSCOhost Web Databases	Embase
ERIC (ProQuest)	HaPI	InCites Journal Citation Reports	Natural Medicines	NCBI

Click here

NCBI Resources How To Sign in to NCBI

NCBI
National Center for Biotechnology Information

All Databases Search

COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

UNITE
A new NIH initiative to end structural racism and achieve racial equity in the biomedical research enterprise.

LEARN MORE

- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

Research
Explore NCBI research and collaborative projects

Popular Resources

PubMed
Bookshelf
PubMed Central
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

NCBI News & Blog

GenBank Release 248.0
25 Feb 2022

GenBank release 248.0 (2/18/2022) is now available on the NCBI FTP site. This

Click here

COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

GenBank Overview

What is GenBank?

GenBank ® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2013 Jan;41(D1):D36-42). GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

A GenBank release occurs every two months and is available from the ftp site. The release notes for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank. Release notes for previous GenBank releases are also available. GenBank growth statistics for both the traditional GenBank divisions and the WGS division are available from each release.

An annotated sample GenBank record for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file format.

Access to GenBank

There are several ways to search and retrieve data from GenBank.

- Search GenBank for sequence identifiers and annotations with Entrez Nucleotide.
- Search and align GenBank sequences to a query sequence using BLAST (Basic Local Alignment Search Tool). See BLAST info for more information about the numerous BLAST databases.
- Search, link, and download sequences programatically using NCBI e-utilities.
- The ASN.1 and flatfile formats are available at NCBI's anonymous FTP server: ftp://ftp.ncbi.nlm.nih.gov/ncbi-asn1 and ftp://ftp.ncbi.nlm.nih.gov/genbank.

GenBank Resources

GenBank Home

Submission Types

Submission Tools

Search GenBank

Update GenBank Records

National Library of Medicine
National Center for Biotechnology Information

Search NCBI | Search

COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

Click here

Quick Navigation Guide

- Sequence Submission
- Literature
- Sequence-Related Resources
- Clinical Resources
- Other Websites

SARS-CoV-2 Data

3,099,725	4,070,899	3,215
SRA runs	Nucleotide records	ClinicalTrials.gov

231,749	288,388
PubMed	PMC

NCBI Virus
Sequences for discovery

About Us | Find Data | Help | How to Participate | Submit Sequences | Contact Us

SARS-CoV-2 Data Hub

Download

Click here

Quick Links: Betacoronavirus BLAST | CDC Outbreak Information | SARS-CoV-2 Articles in PubMed | SRA Data | NCBI SARS-CoV-2 Resources | Datasets command line

Tabular View | Dashboard Visualizations | Mutations in SRA | Complete Tree

Selected Results: 0

Align | Build Phylogenetic Tree

Refine Results | Reset

Virus +

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049 ×

Accession +

Sequence Length +

Expand Table

Nucleotide (4,083,187) | Protein (24,012,480) | RefSeq Genome (1)

Select Columns

Accession	Submitters	Release Date	Pangolin	Isolate	Species
NC_045512 RefSeq	Wu,F., et al.	2020-01-13	B	Wuhan-Hu-1	Severe acute respirat
OM840138	Andrews,K.…	2022-02-27	B.1.2	ID-UI-IIDS-U0847	Severe acute respirat
OM840139	Andrews,K.…	2022-02-27	B.1.2	ID-UI-IIDS-U0850	Severe acute respirat
OM840140	Andrews,K.…	2022-02-27	B.1	ID-UI-IIDS-U0852	Severe acute respirat

SARS-CoV-2 Data Hub

Filters (0) | View results and download

Quick Links

CDC Outbreak Information | SRA Data | Datasets command line

Tabular View | Dashboard Visualizations | Mutations in SRA | Complete Tree

Statistics

1	23,865,769	4,070,899	38	892,702
RefSeq Genomes	All Proteins	All Nucleotides	RefSeq Proteins	Complete Nucleotides

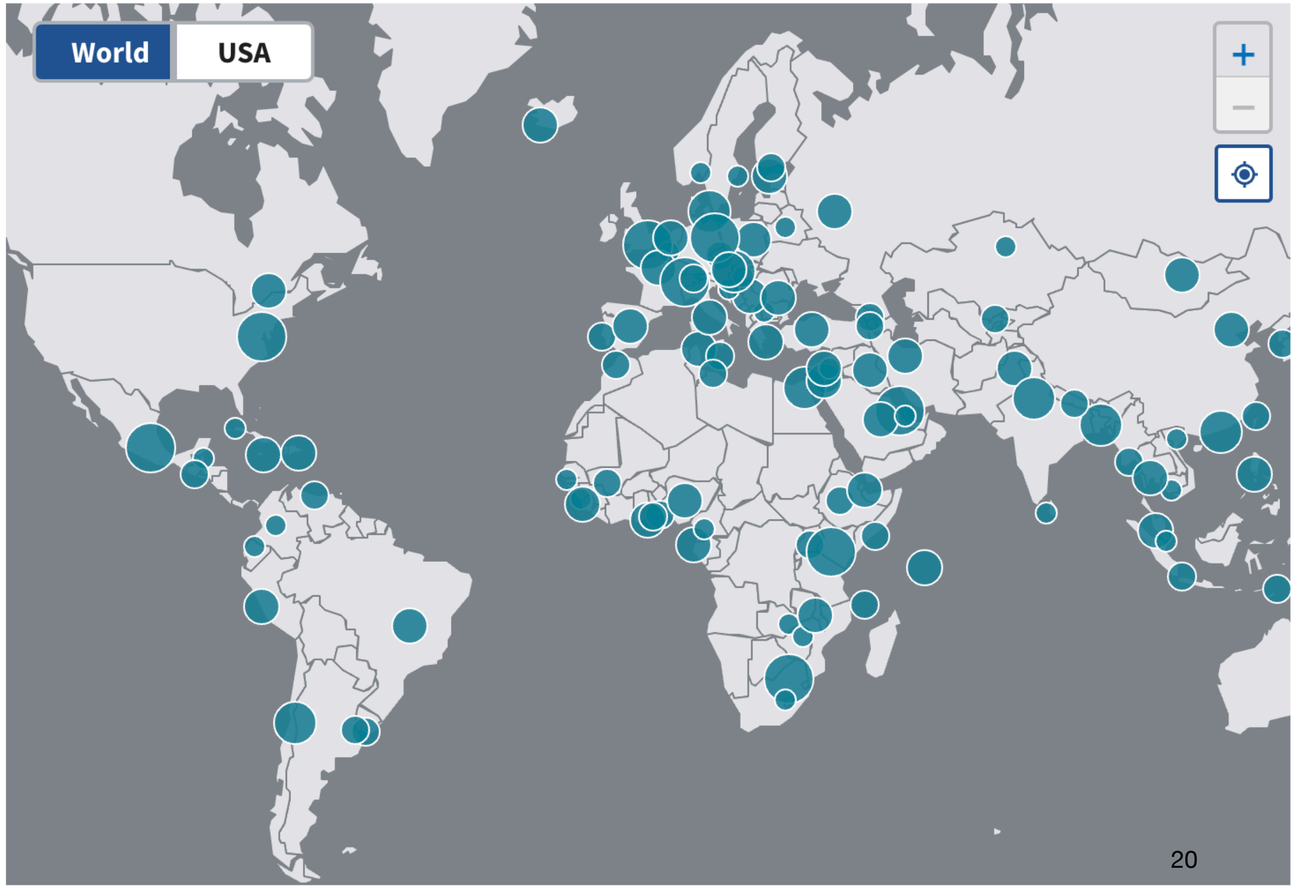
We are going to learn how to make a similar dashboard in Workshop 3.

Geographic and Time Distribution

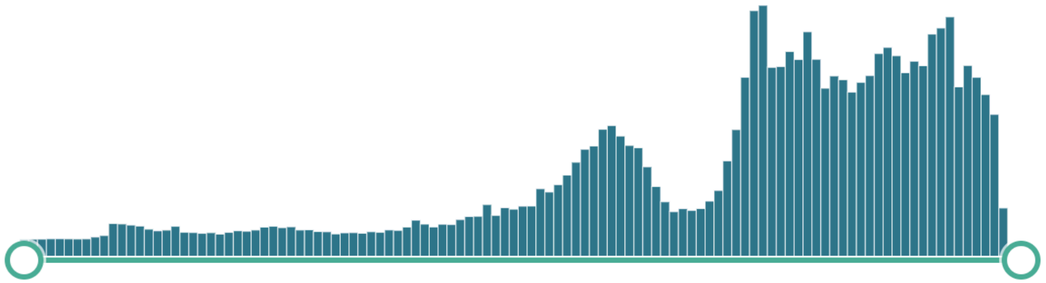
Choose locations to select SARS-CoV-2 sequence records with that collection location. Use the sliders or click date columns to select SARS-CoV-2 records by their sample collection date and/or their GenBank release date.

Geographic Distribution

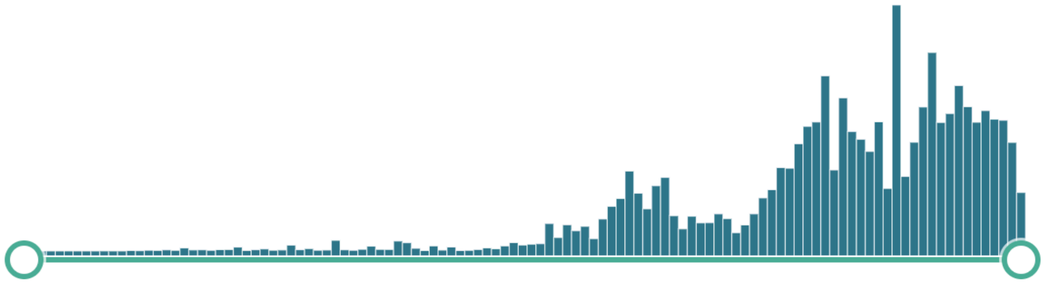
Search for a country



Collection Date Weekly



Release Date Weekly

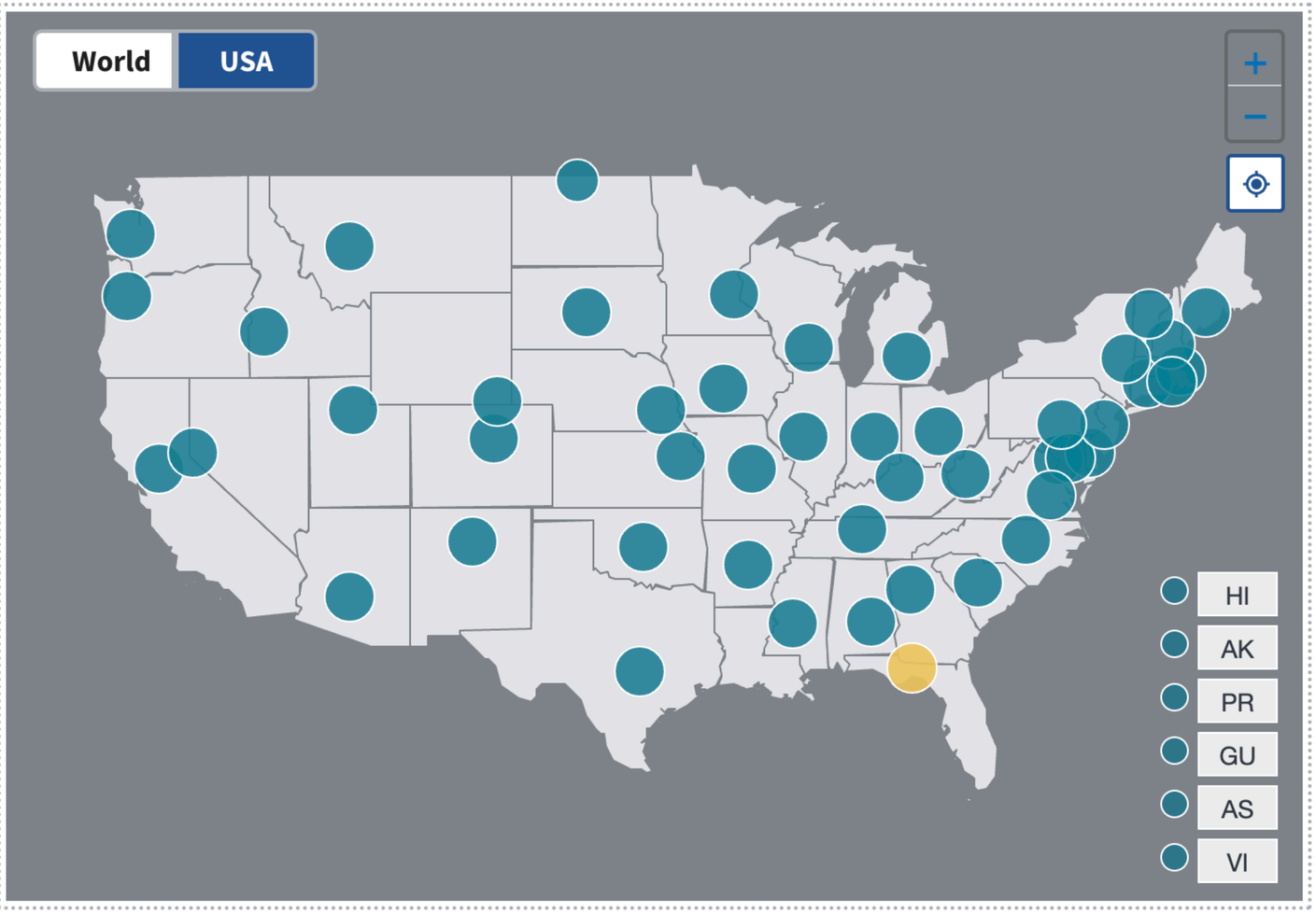


Geographic and Time Distribution

Choose locations to select SARS-CoV-2 sequence records with that collection location. Use the sliders or click date columns to select SARS-CoV-2 records by their sample collection date and/or their GenBank release date.

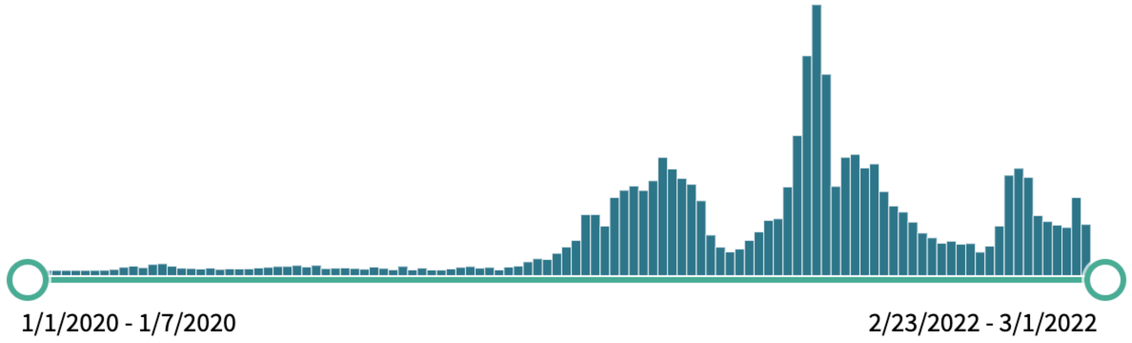
Geographic Distribution

Search for US state or territory



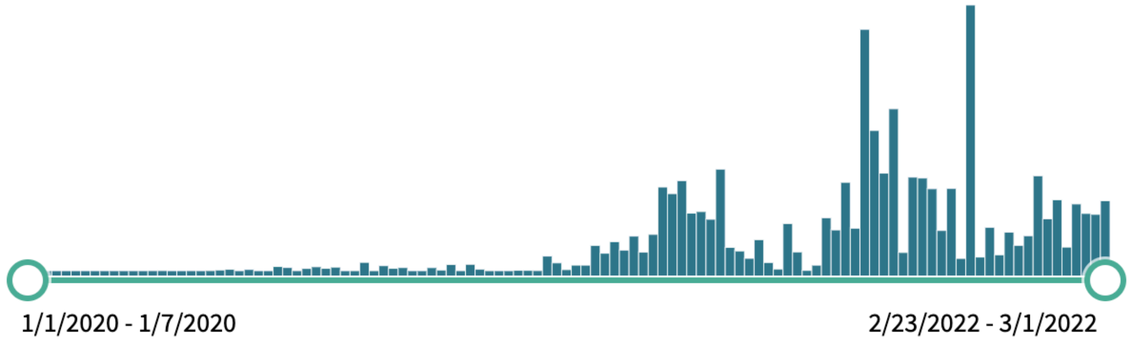
Collection Date

Weekly



Release Date

Weekly



Filtering data

NCBI Virus
Sequences for discovery

About Us | Find Data | Help | How to Participate | Submit Sequences | Contact Us

SARS-CoV-2 Data Hub

Download

Quick Links
Betacoronavirus BLAST
CDC Outbreak Information
SARS-CoV-2 Articles in PubMed
SRA Data
NCBI SARS-CoV-2 Resources
Datasets command line

Tabular View | Dashboard Visualizations | Mutations in SRA | Complete Tree

Selected Results: 0

Align | Build Phylogenetic Tree

Click here

Refine Results | Reset

Virus +

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049 ×

Accession +

Sequence Length +

Nucleotide (4,083,187) | Protein (24,012,480) | RefSeq Genome (1)

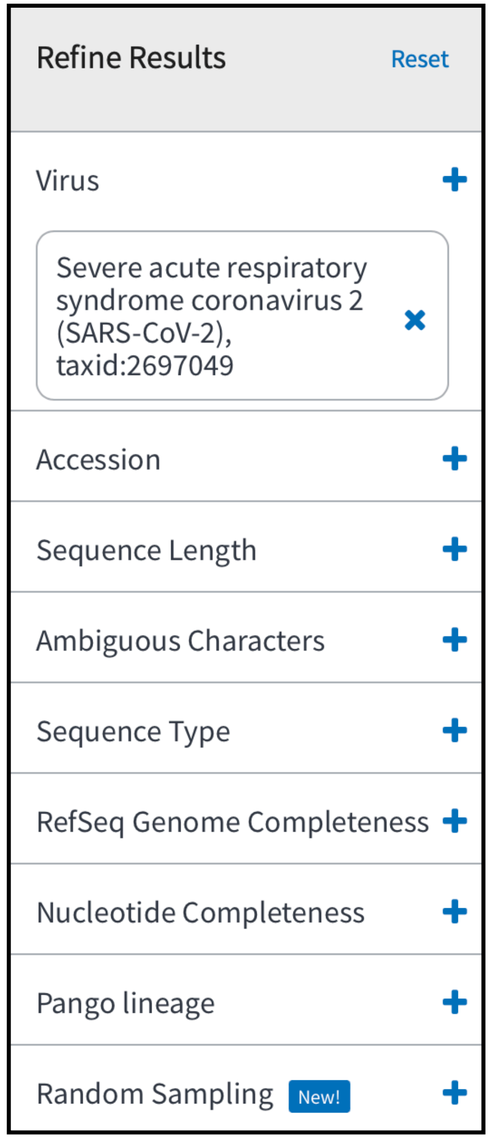
Select Columns

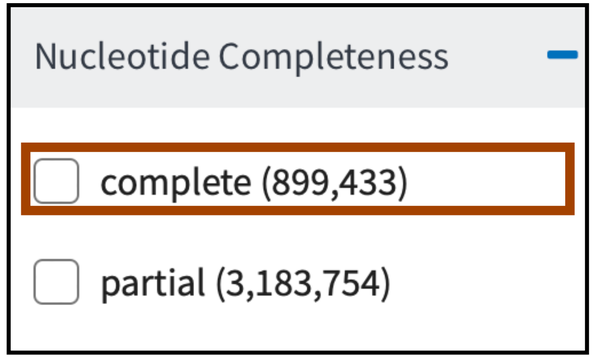
Expand Table

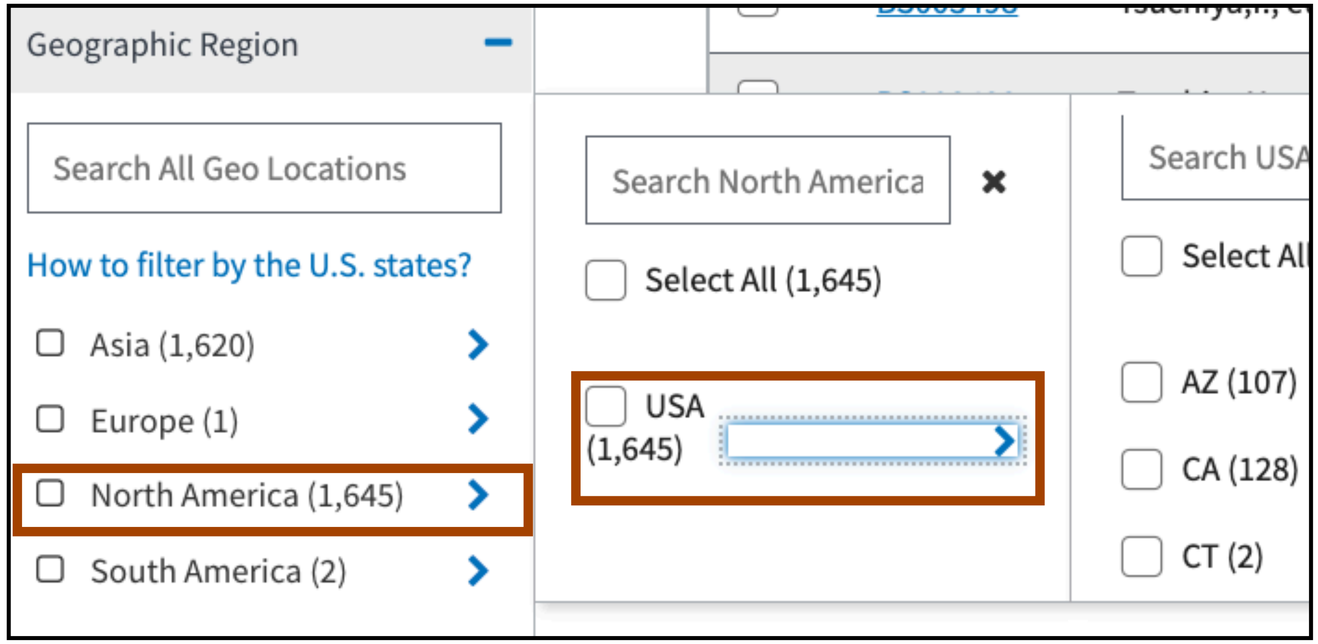
Accession	Submitters	Release Date	Pangolin	Isolate	Species
NC_045512 RefSeq	Wu,F., et al.	2020-01-13	B	Wuhan-Hu-1	Severe acute respirat
OM840138	Andrews,K....	2022-02-27	B.1.2	ID-UI-IIDS-U0847	Severe acute respirat
OM840139	Andrews,K....	2022-02-27	B.1.2	ID-UI-IIDS-U0850	Severe acute respirat
OM840140	Andrews,K....	2022-02-27	B.1	ID-UI-IIDS-U0852	Severe acute respirat

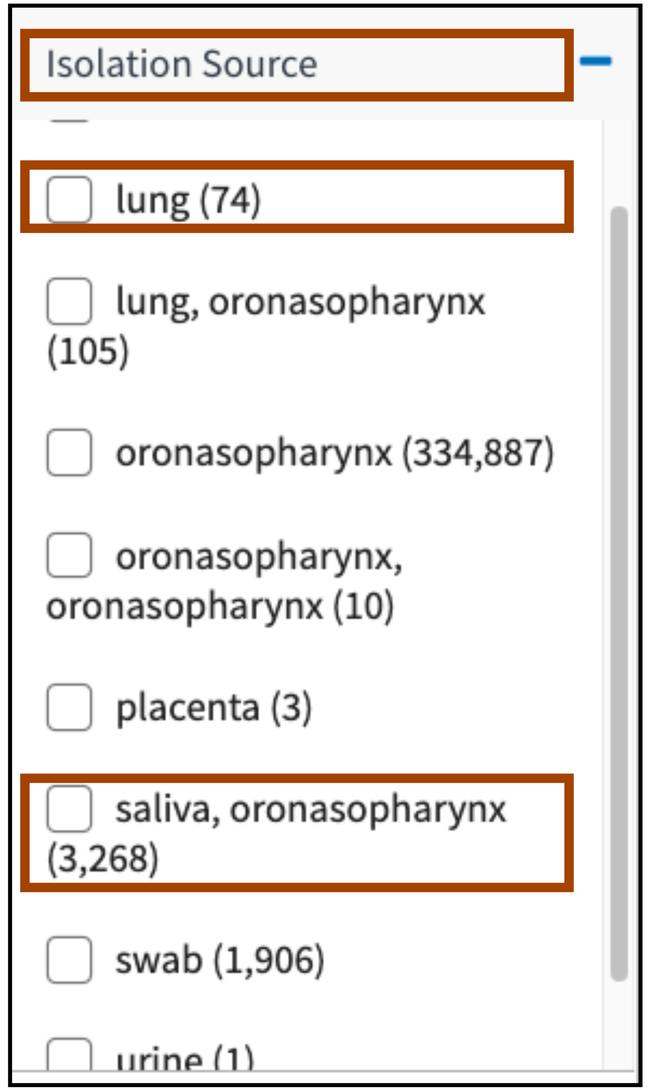
Filtering data

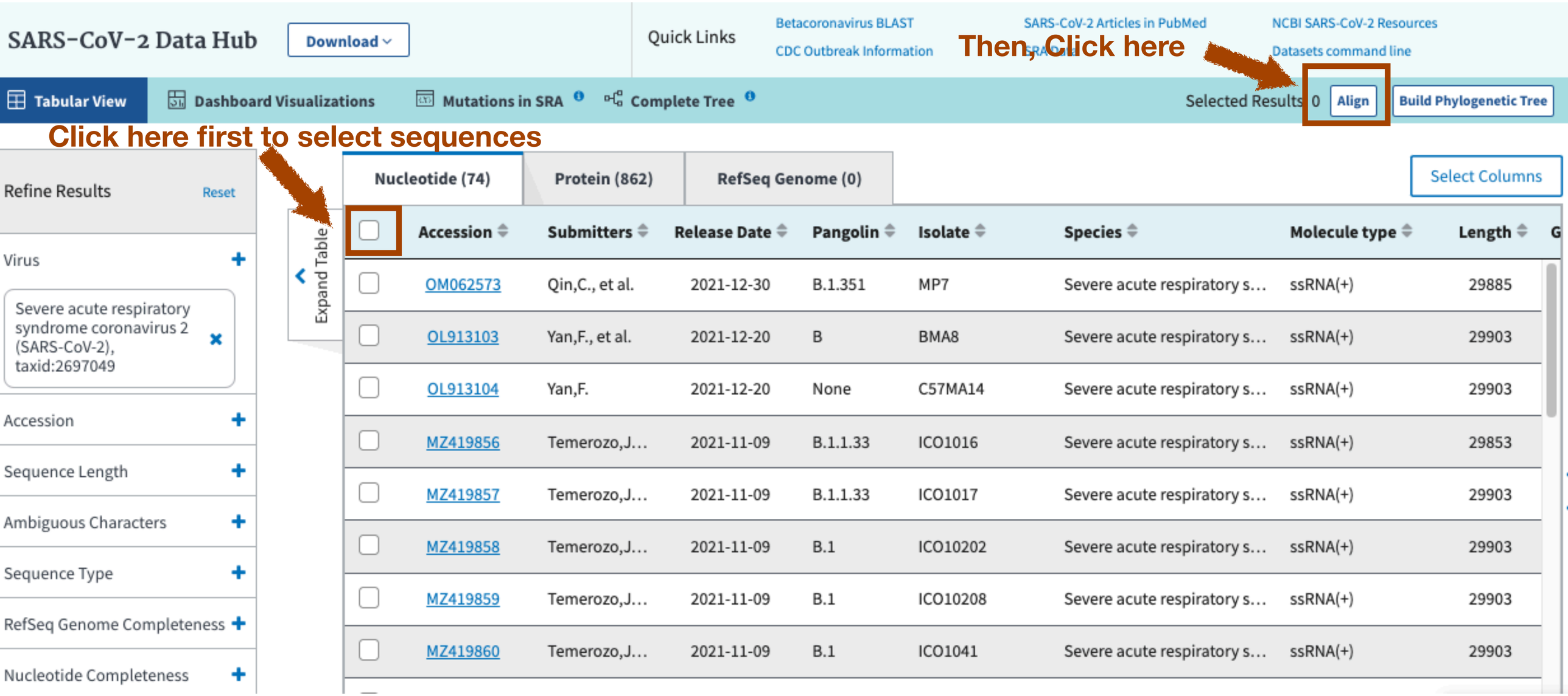
How to filter dataset?





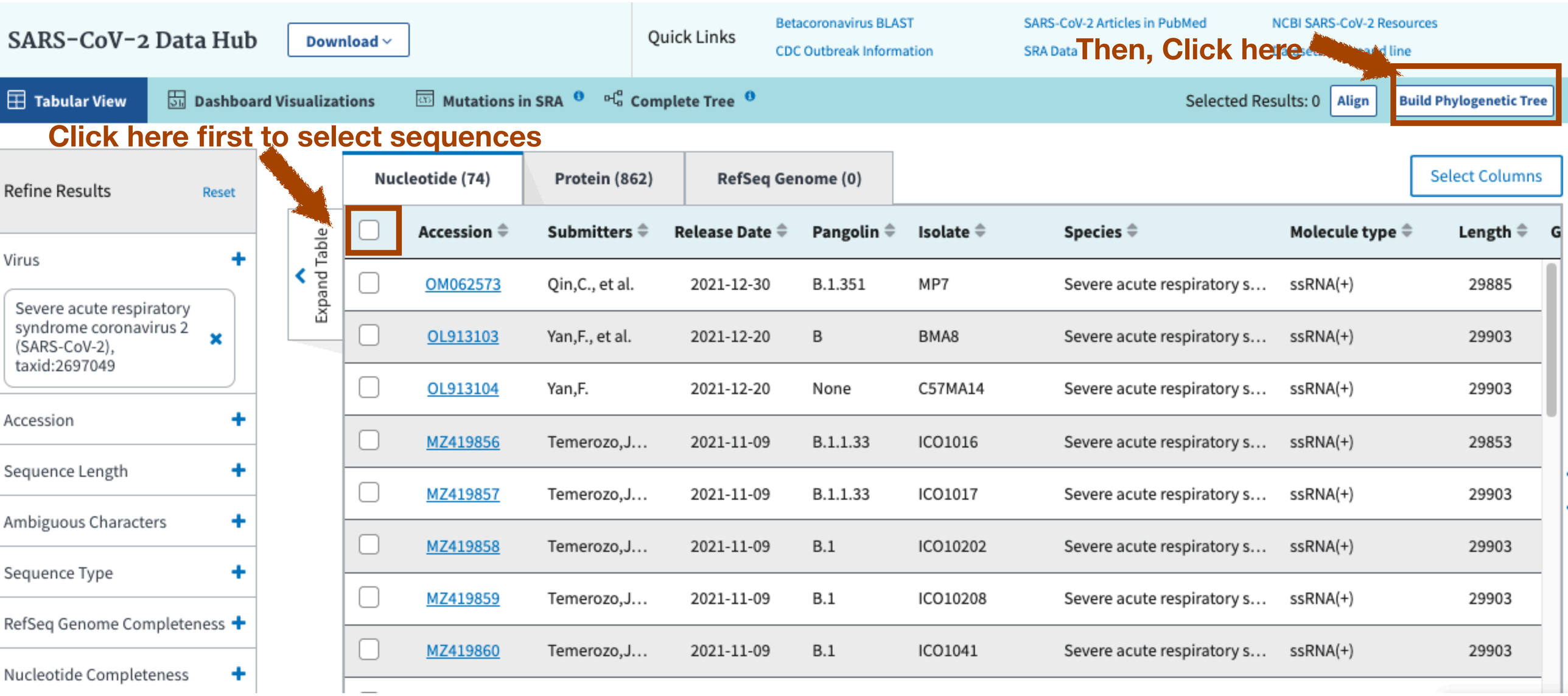





SARS-CoV-2 Data Hub
Download
Quick Links
Betacoronavirus BLAST
CDC Outbreak Information
SARS-CoV-2 Articles in PubMed
NCBI SARS-CoV-2 Resources
Datasets command line
Then, Click here
Tabular View
Dashboard Visualizations
Mutations in SRA
Complete Tree
Selected Results 0
Align
Build Phylogenetic Tree
Click here first to select sequences
Refine Results
Reset
Virus
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049
Accession
Sequence Length
Ambiguous Characters
Sequence Type
RefSeq Genome Completeness
Nucleotide Completeness
Expand Table
Nucleotide (74)
Protein (862)
RefSeq Genome (0)
Select Columns
Accession
Submitters
Release Date
Pangolin
Isolate
Species
Molecule type
Length
OM062573 Qin,C., et al. 2021-12-30 B.1.351 MP7 Severe acute respiratory s… ssRNA(+) 29885
OL913103 Yan,F., et al. 2021-12-20 B BMA8 Severe acute respiratory s… ssRNA(+) 29903
OL913104 Yan,F. 2021-12-20 None C57MA14 Severe acute respiratory s… ssRNA(+) 29903
MZ419856 Temerozo,J… 2021-11-09 B.1.1.33 ICO1016 Severe acute respiratory s… ssRNA(+) 29853
MZ419857 Temerozo,J… 2021-11-09 B.1.1.33 ICO1017 Severe acute respiratory s… ssRNA(+) 29903
MZ419858 Temerozo,J… 2021-11-09 B.1 ICO10202 Severe acute respiratory s… ssRNA(+) 29903
MZ419859 Temerozo,J… 2021-11-09 B.1 ICO10208 Severe acute respiratory s… ssRNA(+) 29903
MZ419860 Temerozo,J… 2021-11-09 B.1 ICO1041 Severe acute respiratory s… ssRNA(+) 29903

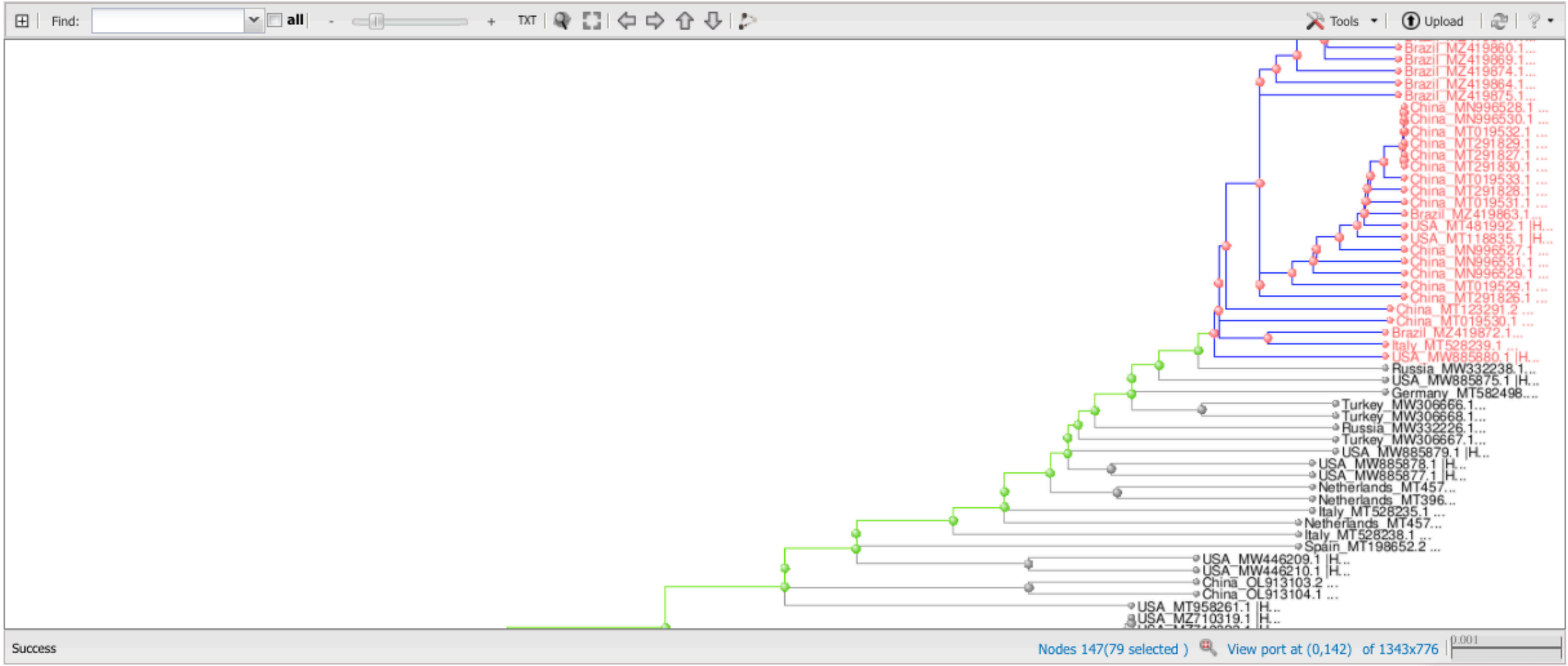
Multiple Alignment

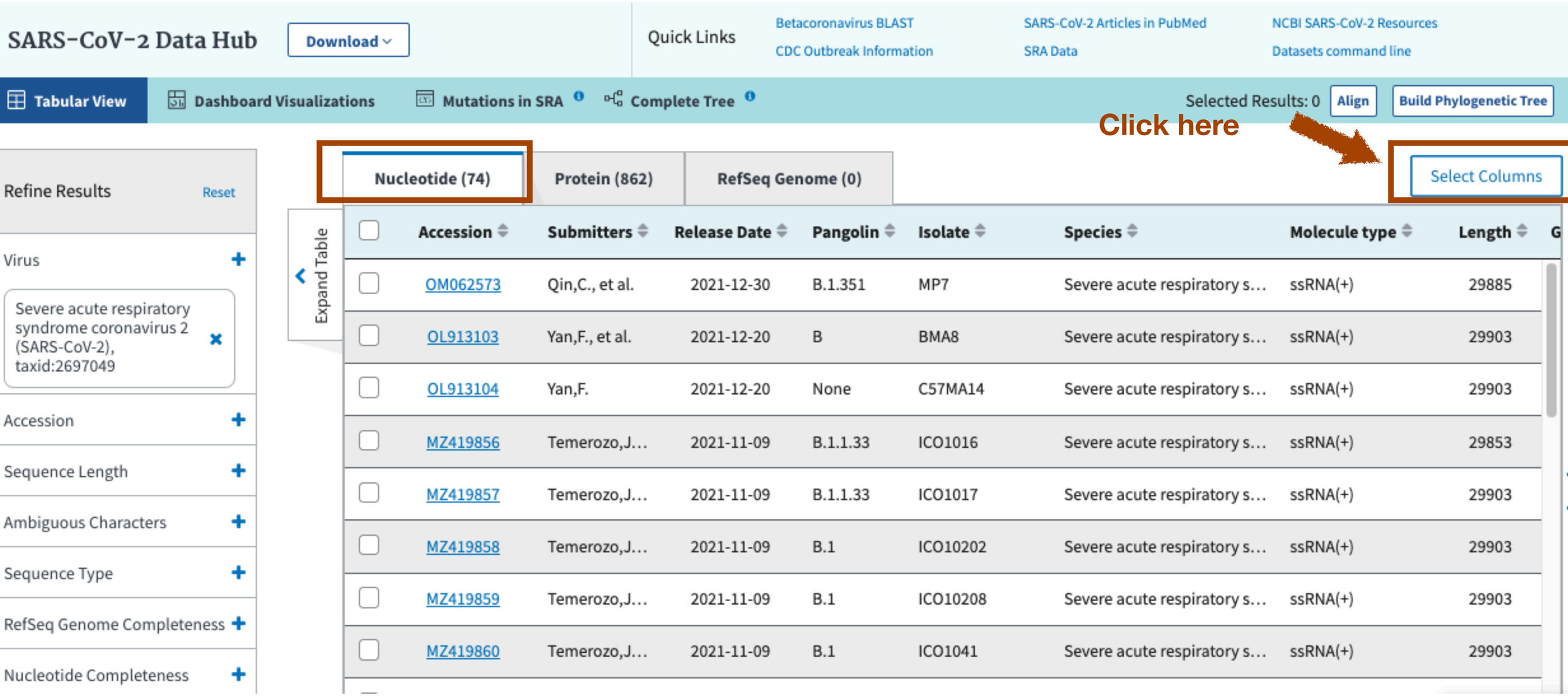
74 rows = number of genomes
29903 columns = number of bases (A , T , C , G)

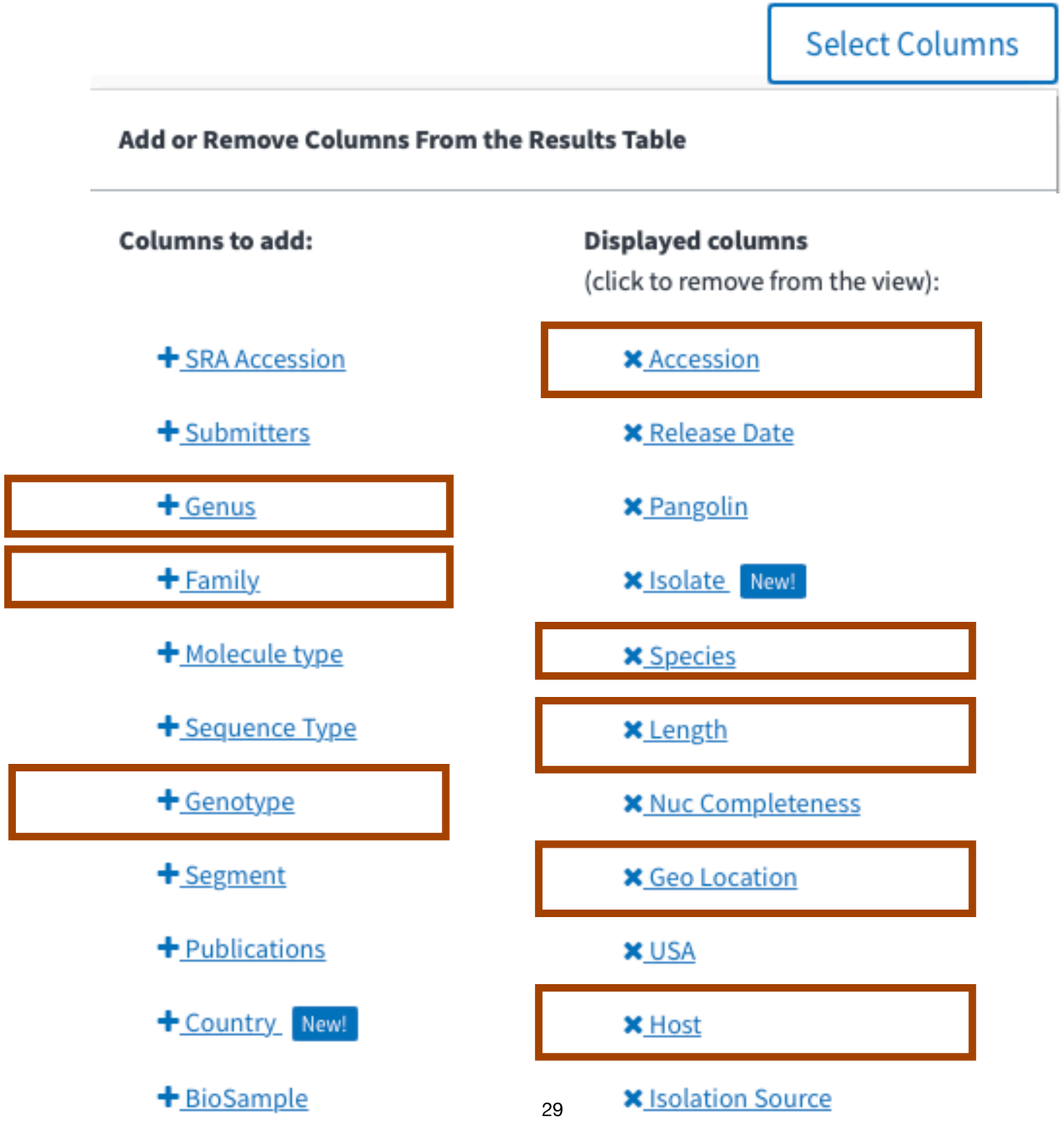

SARS-CoV-2 Data Hub
Download
Quick Links
Betacoronavirus BLAST
CDC Outbreak Information
SARS-CoV-2 Articles in PubMed
SRA Data
NCBI SARS-CoV-2 Resources
Then, Click here
Tabular View
Dashboard Visualizations
Mutations in SRA
Complete Tree
Selected Results: 0
Align
Build Phylogenetic Tree
Click here first to select sequences
Refine Results
Reset
Virus
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049
Accession
Sequence Length
Ambiguous Characters
Sequence Type
RefSeq Genome Completeness
Nucleotide Completeness
Expand Table
Nucleotide (74)
Protein (862)
RefSeq Genome (0)
Select Columns
Accession
Submitters
Release Date
Pangolin
Isolate
Species
Molecule type
Length
OM062573 Qin,C., et al. 2021-12-30 B.1.351 MP7 Severe acute respiratory s… ssRNA(+) 29885
OL913103 Yan,F., et al. 2021-12-20 B BMA8 Severe acute respiratory s… ssRNA(+) 29903
OL913104 Yan,F. 2021-12-20 None C57MA14 Severe acute respiratory s… ssRNA(+) 29903
MZ419856 Temerozo,J… 2021-11-09 B.1.1.33 ICO1016 Severe acute respiratory s… ssRNA(+) 29853
MZ419857 Temerozo,J… 2021-11-09 B.1.1.33 ICO1017 Severe acute respiratory s… ssRNA(+) 29903
MZ419858 Temerozo,J… 2021-11-09 B.1 ICO10202 Severe acute respiratory s… ssRNA(+) 29903
MZ419859 Temerozo,J… 2021-11-09 B.1 ICO10208 Severe acute respiratory s… ssRNA(+) 29903
MZ419860 Temerozo,J… 2021-11-09 B.1 ICO1041 Severe acute respiratory s… ssRNA(+) 29903

We are going to learn other approaches to build this tree that might be better for your data in Workshop 2

Phylogenetic Tree




SARS-CoV-2 Data Hub
Download
Quick Links
Betacoronavirus BLAST
SARS-CoV-2 Articles in PubMed
NCBI SARS-CoV-2 Resources
CDC Outbreak Information
SRA Data
Datasets command line
Tabular View
Dashboard Visualizations
Mutations in SRA
Complete Tree
Selected Results: 0
Align
Build Phylogenetic Tree
Click here
Refine Results
Reset
Virus
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049
Accession
Sequence Length
Ambiguous Characters
Sequence Type
RefSeq Genome Completeness
Nucleotide Completeness
Nucleotide (74)
Protein (862)
RefSeq Genome (0)
Select Columns
Expand Table
Accession
Submitters
Release Date
Pangolin
Isolate
Species
Molecule type
Length
OM062573 Qin,C., et al. 2021-12-30 B.1.351 MP7 Severe acute respiratory s… ssRNA(+) 29885
OL913103 Yan,F., et al. 2021-12-20 B BMA8 Severe acute respiratory s… ssRNA(+) 29903
OL913104 Yan,F. 2021-12-20 None C57MA14 Severe acute respiratory s… ssRNA(+) 29903
MZ419856 Temerozo,J… 2021-11-09 B.1.1.33 ICO1016 Severe acute respiratory s… ssRNA(+) 29853
MZ419857 Temerozo,J… 2021-11-09 B.1.1.33 ICO1017 Severe acute respiratory s… ssRNA(+) 29903
MZ419858 Temerozo,J… 2021-11-09 B.1 ICO10202 Severe acute respiratory s… ssRNA(+) 29903
MZ419859 Temerozo,J… 2021-11-09 B.1 ICO10208 Severe acute respiratory s… ssRNA(+) 29903
MZ419860 Temerozo,J… 2021-11-09 B.1 ICO1041 Severe acute respiratory s… ssRNA(+) 29903


Select Columns
Add or Remove Columns From the Results Table
Columns to add:
Displayed columns
(click to remove from the view):
+ SRA Accession
+ Submitters
+ Genus
+ Family
+ Molecule type
+ Sequence Type
+ Genotype
+ Segment
+ Publications
+ Country New!
+ BioSample
✖ Accession
✖ Release Date
✖ Pangolin
✖ Isolate New!
✖ Species
✖ Length
✖ Nuc Completeness
✖ Geo Location
✖ USA
✖ Host
✖ Isolation Source

Then click here

SARS-CoV-2 Data Hub

Download

Quick Links

Betacoronavirus BLAST

CDC Outbreak Information

SARS-CoV-2 Articles in PubMed

SRA Data

NCBI SARS-CoV-2 Resources

Datasets command line

Tabular View | Dashboard Visualizations | Mutations in SRA | Complete Tree

Selected Results: 74

Align | Build Phylogenetic Tree

Click here first

Refine Results Reset

Virus +

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049

Accession +

Sequence Length +

Ambiguous Characters +

Expand Table

Nucleotide (74) | Protein (862) | RefSeq Genome (0)

Select Columns

Accession	Release Date	Pangolin	Isolate	Species
OL913104	2021-12-20	None	C57MA14	Severe acute respiratory s…
MT582498	2020-06-09	B.3	NRW-02.1	Severe acute respiratory s…
MW306667	2020-11-30	B.1.9	ETLKVET2	Severe acute respiratory s…
MT457400	2020-05-12	B.1.8	NB03_index	Severe acute respiratory s…
MT457401	2020-05-12	B.1.8	NB04_index	Severe acute re…

Feedback

Fasta format

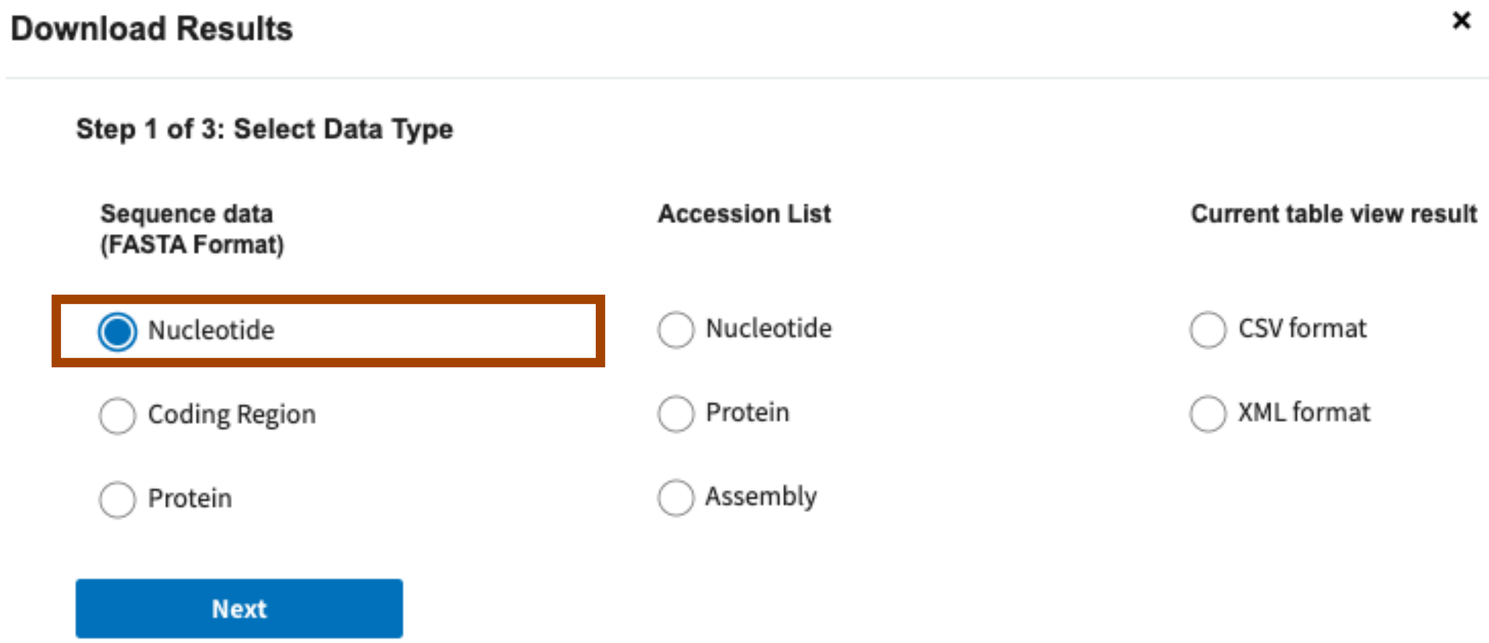
The first line in a FASTA file starts with a “>" (greater-than)

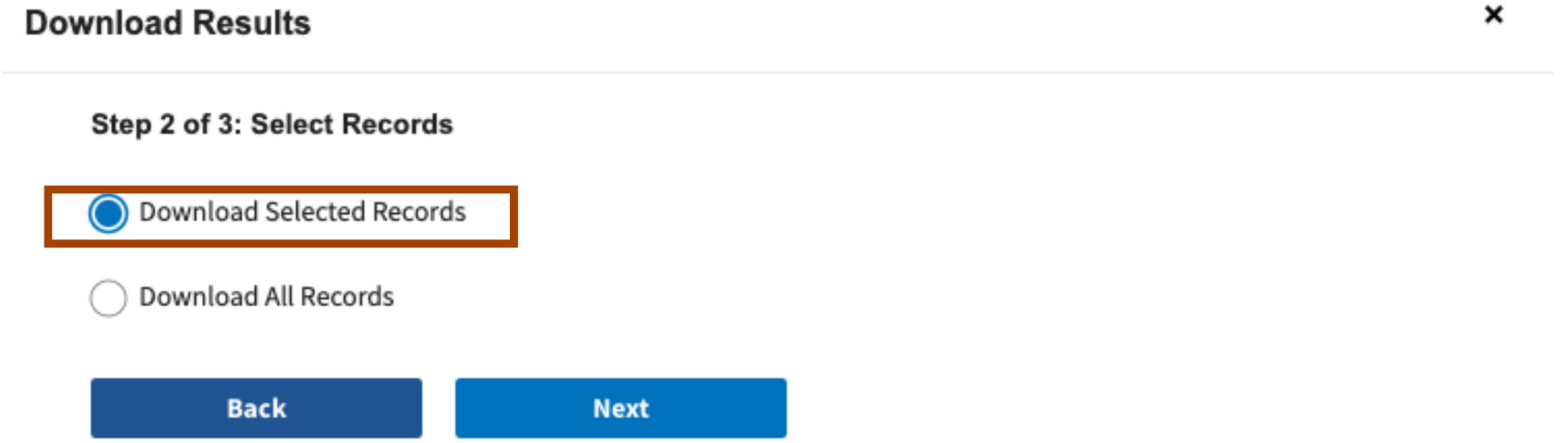
">" (greater-than) is followed by definition (detailed description)

>Header_Line
ATCGACTACCATACCATCGACTA

The next line is the sequence (nucleotide/ protein)

```
>SEQUENCE_1
A
MGQFYVMDDKKTVEQVIAEKEKEFGGKIKIVEFICFEVGEGLEKKTEDFAAEVAAQL
>SEQUENCE_2
SATVSEINSETDFVAKNDQFIALTKDTTAHIQSNSLQSVEELHSSTINGVKFEEYLKSQI
ATIGENLVVRRFATLKAGANGVVNGYIHTNGRVGVVIAAACDSAEVASKSRDLLRQICMH
>SEQUENCE_3
SATVSEINSETDFVAKNDQFIALTKDTTAHIQSNSLQSVEELHSSTINGVKFEEYLKSQI
ATIGENLVVRRFATLKAGANGVVNGYIHTNGRVGVVIAAACDSAEVASKSRDLLRQICMH
```


Download Results
Step 1 of 3: Select Data Type
Sequence data (FASTA Format)
Nucleotide
Coding Region
Protein
Accession List
Nucleotide
Protein
Assembly
Current table view result
CSV format
XML format
Next


Download Results
Step 2 of 3: Select Records
Download Selected Records
Download All Records
Back
Next

Download Results

Step 3 of 3: Select FASTA definition line

Use default : Accession GenBank Title

Build custom : Accession GenBank Title

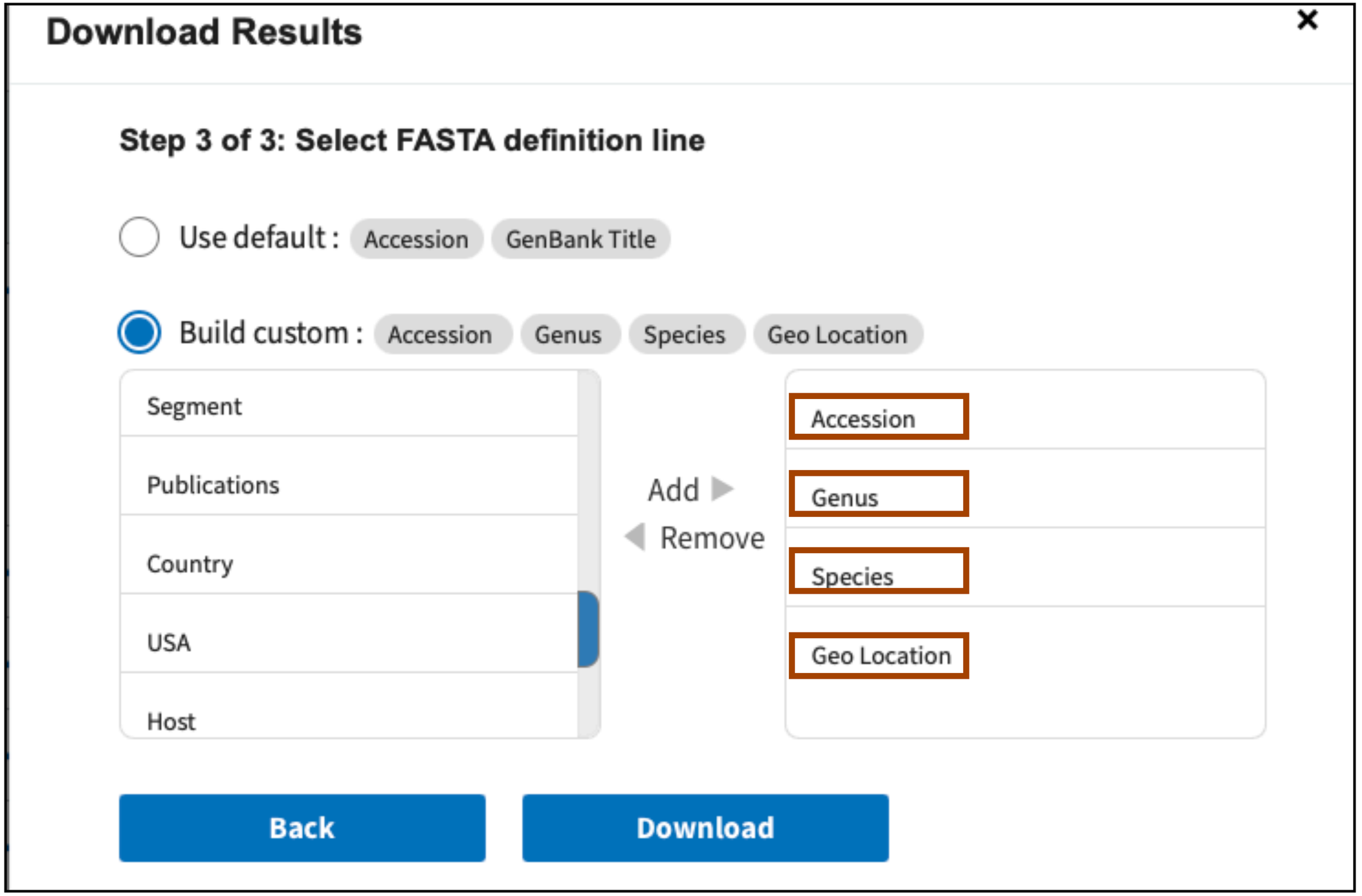
Assembly
SRA Accession
Submitters
Release Date
Pangolin

Add ▶
◀ Remove

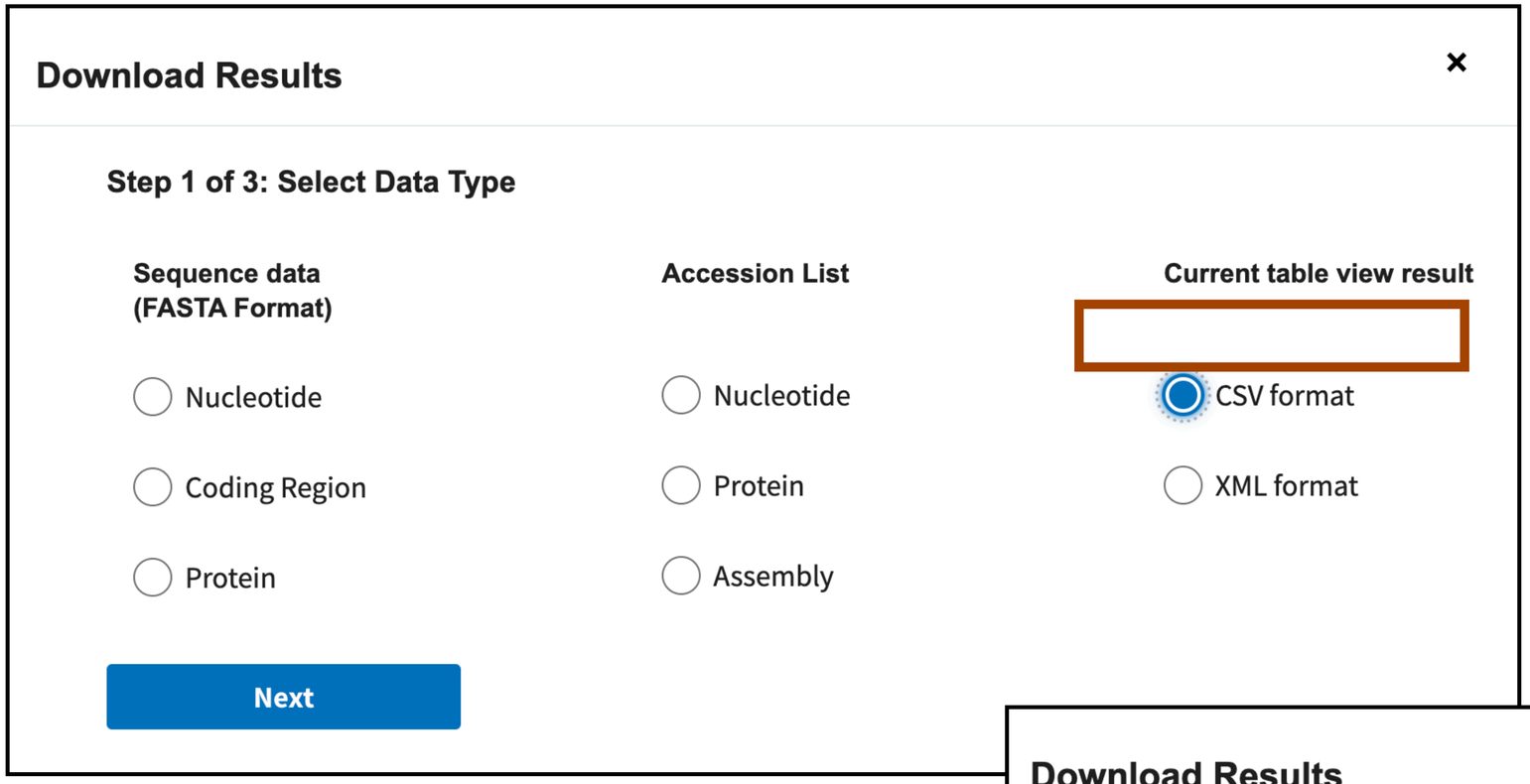
Accession
GenBank Title

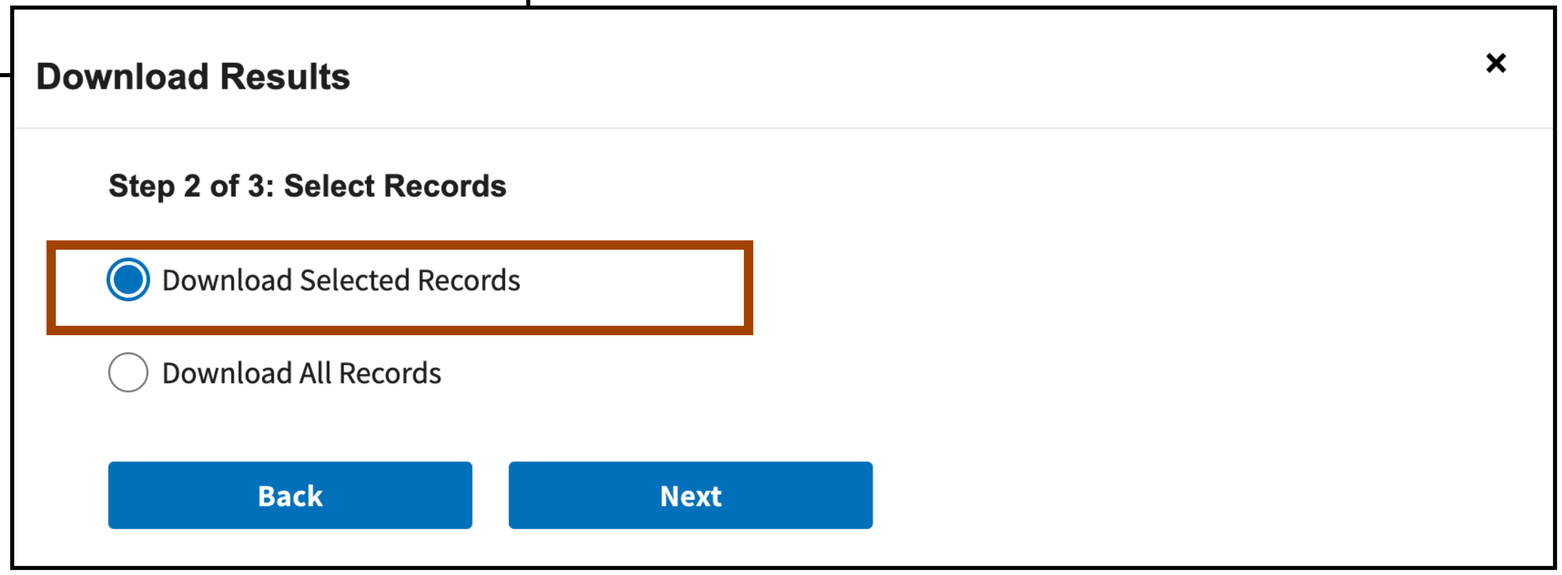
Back

Download


Download Results
Step 3 of 3: Select FASTA definition line
Use default : Accession GenBank Title
Build custom : Accession Genus Species Geo Location
Segment
Publications
Country
USA
Host
Add
Remove
Accession
Genus
Species
Geo Location
Back
Download

Downloaded fasta file


Download Results
Step 1 of 3: Select Data Type
Sequence data (FASTA Format)
Nucleotide
Coding Region
Protein
Accession List
Nucleotide
Protein
Assembly
Current table view result
CSV format
XML format
Next


Download Results
Step 2 of 3: Select Records
Download Selected Records
Download All Records
Back
Next

Download Results

Step 3 of 3: Select columns to include in results set

Note: Columns currently displayed in results table are already selected. You can modify the selection to include the columns you need below.

- [x] Accession
- [ ] SRA Accession
- [x] Submitters
- [x] Release Date
- [x] Pangolin
- [x] Random Sampling
- [x] Isolate
- [x] Species
- [ ] Genus
- [ ] Family
- [x] Molecule type
- [x] Length
- [ ] Sequence Type
- [ ] Nuc Completeness
- [ ] Genotype
- [ ] Segment
- [ ] Publications
- [x] Geo Location
- [ ] Country
- [x] USA
- [x] Host
- [x] Isolation Source
- [x] Collection Date
- [ ] BioSample
- [ ] GenBank Title
- [ ] **Select All**

Accession: (•) without *version* (NC_045512) ( ) with *version* (NC_045512.2)

Back **Download**

Downloaded Table (in csv format)

Genus	Family	Molecule_type	Length	Sequence_Type	Nuc_Completeness	Geo_Location	Country	USA	Host	Isolation_Source	Collec
Betacoronavirus	Coronaviridae	ssRNA(+)	29885	GenBank	complete	China: Beijing	China		Mus musculus	lung	2021-(
Betacoronavirus	Coronaviridae	ssRNA(+)	29903	GenBank	complete	China	China		Rodentia	lung	2020-(
Betacoronavirus	Coronaviridae	ssRNA(+)	29903	GenBank	complete	China	China		Rodentia	lung	2020-(
Betacoronavirus	Coronaviridae	ssRNA(+)	29853	GenBank	complete	Brazil	Brazil		Homo sapiens	lung	2020-(
Betacoronavirus	Coronaviridae	ssRNA(+)	29903	GenBank	complete	Brazil	Brazil		Homo sapiens	lung	2020-(
Betacoronavirus	Coronaviridae	ssRNA(+)	29903	GenBank	complete	Brazil	Brazil		Homo sapiens	lung	2020-(
Betacoronavirus	Coronaviridae	ssRNA(+)	29903	GenBank	complete	Brazil	Brazil		Homo sapiens	lung	2020-(
Betacoronavirus	Coronaviridae	ssRNA(+)	29903	GenBank	complete	Brazil	Brazil		Homo sapiens	lung	2020-(
Betacoronavirus	Coronaviridae	ssRNA(+)	29884	GenBank	complete	Brazil	Brazil		Homo sapiens	lung	2020-(
Betacoronavirus	Coronaviridae	ssRNA(+)	29854	GenBank	complete	Brazil	Brazil		Homo sapiens	lung	2020-(
Betacoronavirus	Coronaviridae	ssRNA(+)	29903	GenBank	complete	Brazil	Brazil		Homo sapiens	lung	2020-(
Betacoronavirus	Coronaviridae	ssRNA(+)	29903	GenBank	complete	Brazil	Brazil		Homo sapiens	lung	2020-(

Workshop 2

1. Transferring files from your local computer to HiPerGator: Tutorial can be found here : File Transfer
2. Making an alignment in Phylip format: Tutorial can be found here : Clustal Alignment

You can then use this alignment file to build a phylogenetic tree.

3. Making a phylogenetic tree using RAxML Tutorial can be found here : RAxML tutorial

Workshop 3

1. Using RShiny and FlexDashboard
2. Loading and formatting data in RMarkdown
3. Creating an interactive Visualization Dashboard

GitHub

SCAN ME

Questions?

http://NatyaHans.github.io/Workshops