

NCBI SARS-CoV-2 Data Hub

Find sequenced genomes of viruses from patient samples

We want to look at samples from humans, but it's not limited to humans

Download files to your computer from NCBI

Get the sequence data files to logrus

Look at data



National Library of Medicine
National Center for Biotechnology Information

Log in



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](#)



Results Table



Dashboard Visualizations



Variants Overview

New!

Selected Results: 0

Align

Build Phylogenetic Tree

New! Randomized subsets in Downloads

You now have the option of downloading a smaller, randomized subset of the data shown in the Results table. Begin by using filters to refine your dataset, select the Nucleotide, Protein, or RefSeq Genome tab above the table for the datatype you would like to download, then follow the prompts in the Download menu. Our [Help documentation](#) has more information.

Refine Results

Reset

Nucleotide (7,115,779)

Protein (41,160,434)

RefSeq Genome (1)

Select Columns

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

Pango lineage

Random Sampling

Isolate

Proteins

Provirus

Geographic Region

Host

Submitters

Isolation Source

Nucleotide Completeness

- ☐ complete (1,956,955)
- ☐ partial (5,158,824)

Geographic Region

Search All Geo Locations

How to filter by the U.S. states?

- ☐ Africa (24,157)
- ☐ Asia (49,998)
- ☐ Europe (3,632,052)
- ☐ North America (3,358,778)
- ☐ Oceania (30,026)
- ☐ South America (20,316)

Host

Start typing, select an option

- ☐ Human
- ☐ Non-human

Collection Date

From

mm/dd/yyyy

To

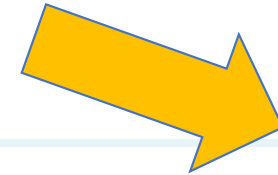
6/19/2023

Submit

Downloads

Get a smaller, randomized subset of the data shown in the Results table. Begin by using filters to refine your dataset, select the Nucleotide, Protein, or RefSeq Genome tab above the table for the download. Then follow the prompts in the Download menu. Our [Help documentation](#) has more information.

Nucleotide (7,115,779)		Protein (41,160,434)		RefSeq Genome (1)		Select Columns		
<input type="checkbox"/>	Accession	Organism Name	Submitters	Organization	Release Date	Pangolin	Isolate	Species
<input type="checkbox"/>	NC_012224	Severe acute respiratory ...	Wang, et al.	National Center for Prote...	2020-01-10	XB	Woman-Hu-1	Severe acute respiratory ...
<input type="checkbox"/>	OR131333	Severe acute respiratory ...	Howard, D., et ...	Centers for Disease Contr...	2023-06-14	XB	PA-CDC-LC1043852	Severe acute respiratory ...
<input type="checkbox"/>	OR131334	Severe acute respiratory ...	Howard, D., et ...	Centers for Disease Contr...	2023-06-14	XB	PA-CDC-LC1043852	Severe acute respiratory ...



Select Columns

Add or Remove Columns From the Results Table

Columns to add:

- [+ SRA Accession](#)
- [+ Genus](#)
- [+ Family](#)
- [+ Sequence Type](#)
- [+ Nuc Completeness](#)
- [+ Genotype](#)
- [+ Segment](#)
- [+ Publications](#)
- [+ Country](#)
- [+ BioSample](#)
- [+ GenBank Title](#)

Displayed columns

(click to remove from the view):

- [✕ Accession](#)
- [✕ Organism Name](#) New!
- [✕ Submitters](#)
- [✕ Organization](#)
- [✕ Release Date](#)
- [✕ Pangolin](#)
- [✕ Isolate](#)
- [✕ Species](#)
- [✕ Molecule type](#)
- [✕ Length](#)
- [✕ Geo Location](#)
- [✕ USA](#)
- [✕ Host](#)
- [✕ Isolation Source](#)
- [✕ Collection Date](#)

Feedback

Nucleotide Completeness **+**

complete **x**

Pango lineage **+**

Random Sampling **+**

Isolate **+**

Proteins **+**

Provirus **+**

Geographic Region **+**

Denmark **x** Norway **x**

Sweden **x**

Host **+**

Homo sapiens (human),
taxid:9606 **x**



Nucleotide (635)						
Protein (6,706)						
RefSeq Genome (0)						
<input checked="" type="checkbox"/>	Accession ⌵	Organism Name New! ⌵	Submitters ⌵	Organization ⌵	Release Date	
<input checked="" type="checkbox"/>	OR079912	Severe acute respiratory ...	Smura,T.P., ...	NTNU, IKOM, Norway	2023-06-	
<input checked="" type="checkbox"/>	OQ843561	Severe acute respiratory ...	Binderup,A...	University of Copenhage...	2023-04-	
<input checked="" type="checkbox"/>	OQ816151	Severe acute respiratory ...	Albert,J., et al.	Karolinska University Ho...	2023-04-	
<input checked="" type="checkbox"/>	OQ816152	Severe acute respiratory ...	Albert,J., et al.	Karolinska University Ho...	2023-04-	
<input checked="" type="checkbox"/>	OQ816154	Severe acute respiratory ...	Albert,J., et al.	Karolinska University Ho...	2023-04-	
<input checked="" type="checkbox"/>	OQ816156	Severe acute respiratory ...	Albert,J., et al.	Karolinska University Ho...	2023-04-	
<input checked="" type="checkbox"/>	OQ816157	Severe acute respiratory ...	Albert,J., et al.	Karolinska University Ho...	2023-04-	

New! Randomized subsets in Downloads

You now have the option of downloading a smaller, randomized subset of the data shown in the Results table. Begin by using filters to refine your dataset, select the Nucleotide, Protein, or RefSeq Genome tab above the table for the datatype you would like to download, then follow the prompts in the Download menu. Our [Help documentation](#) has more information.

Refine Results

Reset

Virus

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

Accession

Sequence Length

Nucleotide (635)

Protein (6,706)

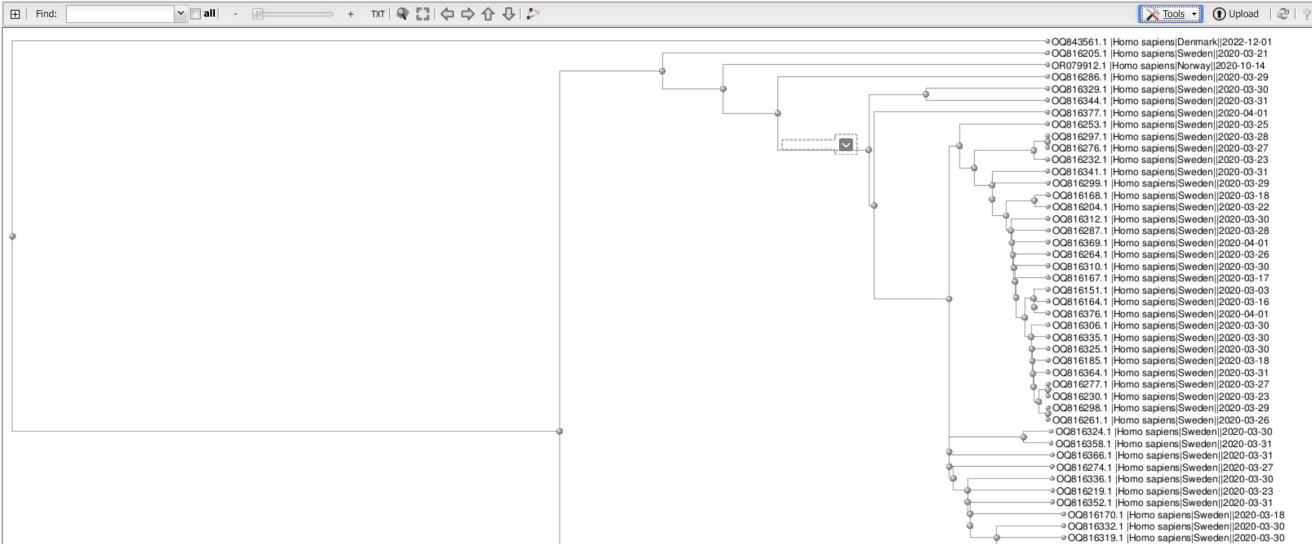
RefSeq Genome (0)

Select Columns

Expand Table

<input checked="" type="checkbox"/>	Accession	Organism Name	Submitters	Organization	Release Date	Pangolin	Isolate	Species	Molecule type	Length	Geo Location	USA	Host	Isolation Source	Collection Date
<input checked="" type="checkbox"/>	OR079912	Severe acute respiratory...	Smura,T.P., ...	NTNU, IKOM, Norway	2023-06-02	B.1.258	P9	Severe acute respirator...	ssRNA(+)	29825	Norway		Homo sapiens	oronasopharynx	2020-03-12
<input checked="" type="checkbox"/>	OQ843561	Severe acute respiratory...	Binderup,A...	University of Copenhag...	2023-04-19	BQ.1.1.20	DK-AHH8	Severe acute respirator...	ssRNA(+)	29752	Denmark		Homo sapiens	oronasopharynx	2020-03-12
<input checked="" type="checkbox"/>	OQ816151	Severe acute respiratory...	Albert,J., et al.	Karolinska University H...	2023-04-14	B.1	01_SE100_21CS503718	Severe acute respirator...	ssRNA(+)	29736	Sweden		Homo sapiens	oronasopharynx	2020-03-12

Phylogenetic Tree



Multiple Alignment



Nucleotide (635)		Protein (6,706)	RefSeq Genome (0)		Select Columns				
<input type="checkbox"/>	Accession	Organism New!	Submitters	Release Date	Pangolin	Isolate	Species	Length	
<input type="checkbox"/>	WIF20394	Severe acute respiratory ...	Smura,T.P., ...	2023-06-02	B.1.258	P9	Severe acute respiratory ...	7096	
<input type="checkbox"/>	WIF20395	Severe acute respiratory ...	Smura,T.P., ...	2023-06-02	B.1.258	P9	Severe acute respiratory ...	4405	
<input type="checkbox"/>	WIF20396	Severe acute respiratory ...	Smura,T.P., ...	2023-06-02	B.1.258	P9	Severe acute respiratory ...	1271	
<input type="checkbox"/>	WIF20397	Severe acute respiratory ...	Smura,T.P., ...	2023-06-02	B.1.258	P9	Severe acute respiratory ...	275	
<input type="checkbox"/>	WIF20398	Severe acute respiratory ...	Smura,T.P., ...	2023-06-02	B.1.258	P9	Severe acute respiratory ...	75	
<input type="checkbox"/>	WIF20399	Severe acute respiratory ...	Smura,T.P., ...	2023-06-02	B.1.258	P9	Severe acute respiratory ...	222	

Results Table

Dashboard Visualizations

Variants Overview New!

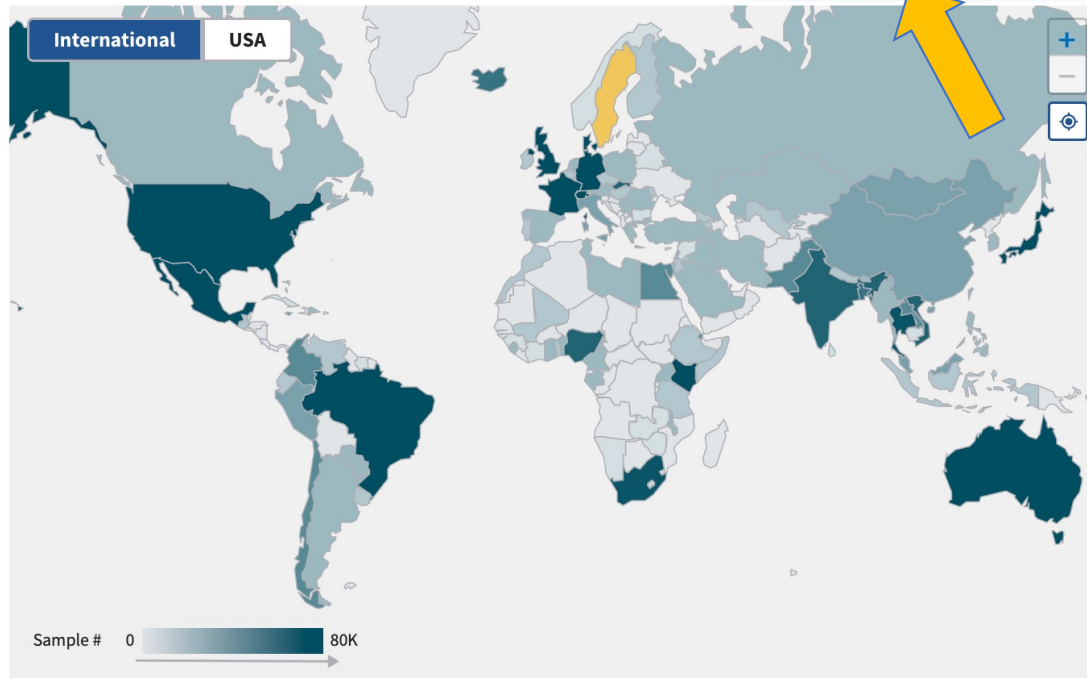
Sequence Type

0 RefSeq Nucleotides **8,433** All Proteins **810** All Nucleotides 0 RefSeq Proteins **616** Complete Nucleotides

Geographic and Time Distribution

Choose locations to select SARS-CoV-2 sequence records by their 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 i



Collection Time i

Weekly ▼

2021 - 2023 ▼

1/1/2021 - 1/7/2021

6/16/2023 - 6/22/20

Release Time i

Weekly ▼

2021 - 2023 ▼

1/1/2021 - 1/7/2021

6/16/2023 - 6/22/20

SARS-CoV-2 Variants Overview

[Data Overview](#)[Lineages](#)[Mutations](#)[New!](#)

OR

[Top 5 growing lineages](#)

Data updated: Jun 19, 2023

[Download Data](#)[Learn more about download format in Help Center](#)Try one of these: **Omicron, BA.2, VBM**

Variants

Variant Card for

XBB, VOC

Defining Mutations:

[Explore sequence data](#)[Additional Information](#)**XBB.1, VOC**

Defining Mutations:

[Explore sequence data](#)[Additional Information](#)**XBB.1.1, VOC**

Pango lineage

XBB

CDC status

VOC

WHO name

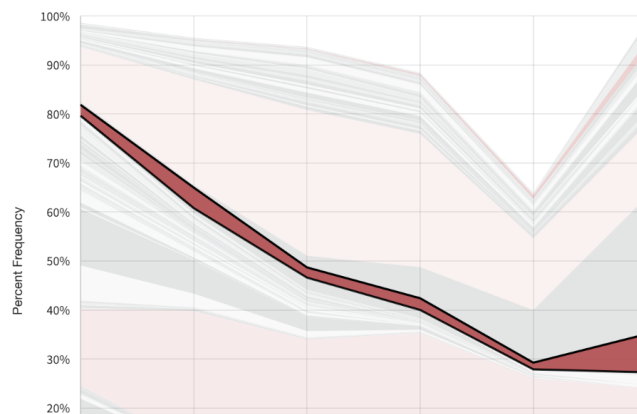
N/A[View all lineages](#)[Explore sequence data](#)

Lineage Frequency

[XBB](#) [Area](#) [Line](#)

Doubling Time

Faster Growth Rate



Geographic Chart

[Map](#) [Table](#)[USA](#)[International](#)

Lineage ↑	All time ↑	Jan. 2023 ↓	Feb. 2023 ↓	Mar. 2023 ↓	Apr. 2023 ↓	May 2023 ↓	Jun. 2023 ↓
Processed*	3187	38339	25460	19196	10402	6200	154
BJ.1	0	6,076	7,007	4,646	2,945	1,229	25
XBB.1.5	0	4,345	5,495	5,721	2,836	914	23
XBB.1	0	219	136	453	658	663	43
B.1.1.529	0	1,555	1,426	1,263	532	333	12
XBB	3,187	859	1,071	402	250	85	12
XBB.1.5.49	0	25	128	229	117	51	2
XBB.2.3	0	4	4	18	27	45	3
XBB.1.5.15	0	109	155	156	79	39	2
XBB.1.5.1	0	55	60	131	65	33	2

SARS-CoV-2 Data Hub

[Download ▾](#)
[Quick Links](#)
[Betacoronavirus BLAST](#)
[CDC Outbreak Information](#)
[Results Table](#)
[Dashboard Visualizations](#)
[Variants Overview](#) **New!**

📢 New! Randomized subsets in Downloads

You now have the option of downloading a smaller, randomized subset of the data shown in the Results table. Begin by using filters to refine your dataset, select the Nucleotide, Protein or RefSeq Genome, and click the [Help documentation](#) has more information.

Refine Results

[Reset](#)

Virus

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

Accession

Sequence Length

Ambiguous Characters

Nucleotide (810)

Protein (8,433)

RefSeq Genome (0)

Expand Table

<input type="checkbox"/>	Accession ▾	Organism Name New! ▾	Submitters ▾	Organization ▾	Release Date ▾	Pangolin ▾	Isolation Source ▾
<input type="checkbox"/>	OQ816151	Severe acute respiratory...	Albert,J., et al.	Karolinska University H...	2023-04-14	B.1	01_S1
<input type="checkbox"/>	OQ816152	Severe acute respiratory...	Albert,J., et al.	Karolinska University H...	2023-04-14	B.1	01_S1
<input type="checkbox"/>	OQ816153	Severe acute respiratory...	Albert,J., et al.	Karolinska University H...	2023-04-14	B.1	01_S1
<input type="checkbox"/>	OQ816154	Severe acute respiratory...	Albert,J., et al.	Karolinska University H...	2023-04-14	B.1	01_S1

Download Results [x]

Step 1 of 3: Select Data

Sequence Data (FASTA format)

☒ Nucleotide

☐ Coding Region

☐ Protein

Accession List

☐ Nucleotide

☐ Protein

☐ Assembly

Results Table

☐ CSV format

☐ XML format

Next

Accession	Sequence Data	Accession List	Results Table
Q0816154	Severe acute respiratory...	Albert,J., et al. Karolinska University H...	2023-04-14 B.1 01_SE100_21CS503085 Severe acute respirator... ssRNA(+) 29736 Sweden
Q0816156	Severe acute respiratory...	Albert,J., et al. Karolinska University H...	2023-04-14 B.1 01_SE100_21CS502014 Severe acute respirator... ssRNA(+) 29736 Sweden

Download Results [x]

Step 3 of 3: Select FASTA definition line

☐ Use default: Accession GenBank Title

☒ Build custom: Accession GenBank Title

Assembly

Organism Name

SRA Accession

Submitters

Accession

GenBank Title

Add Remove

Back Download

Download Results [x]

Step 2 of 3: Select Records

☐ Download Selected Records (200)

☒ Download All Records (635)

☐ Download a randomized subset of all records (up to 635) **New!**

Back Next

Download Results [x]

Step 3 of 3: Select FASTA definition line

☐ Use default: Accession GenBank Title

☒ Build custom: Accession GenBank Title Collection Date Country Geo Location Length Pangolin Release Date SRA Accession Sequence Type

Family

Molecule type

Nuc Completeness

Genotype

Accession

GenBank Title

Collection Date

Country

Back Download

🔔 New! Randomized subsets in Downloads

You now have the option of downloading a smaller, randomized subset of the data shown in the Results table. Begin by using filters to refine your dataset, select the Nucleotide, [documentation](#) has more information.

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

complete

Pango lineage

Random Sampling

Isolate

Proteins

membrane glycoprotein

Provirus

Geographic Region

Denmark

Norway

Sweden

Host

Homo sapiens (human), taxid:9606

Nucleotide (559)		Protein (559)	RefSeq Genome (0)					
<input type="checkbox"/>	Accession	Organism Name New!	Submitters	Release Date	Pangolin	Isolate	Species	
<input type="checkbox"/>	WIF20399	Severe acute respiratory...	Smura,T.P., ...	2023-06-02	B.1.258	P9	Seven	
<input type="checkbox"/>	WGG89488	Severe acute respiratory...	Binderup,A...	2023-04-19	BQ.1.1.20	DK-AHH8	Seven	
<input type="checkbox"/>	WGC55534	Severe acute respiratory...	Albert,J., et al.	2023-04-14	B.1	01_SE100_21CS503718	Seven	
<input type="checkbox"/>	WGC55546	Severe acute respiratory...	Albert,J., et al.	2023-04-14	B.1	01_SE100_21CS504474	Seven	
<input type="checkbox"/>	WGC55570	Severe acute respiratory...	Albert,J., et al.	2023-04-14	B.1	01_SE100_21CS503085	Seven	
<input type="checkbox"/>	WGC55594	Severe acute respiratory...	Albert,J., et al.	2023-04-14	B.1	01_SE100_21CS502014	Seven	
<input type="checkbox"/>	WGC55606	Severe acute respiratory...	Albert,J., et al.	2023-04-14	B.1	01_SE100_21CS502013	Seven	
<input type="checkbox"/>	WGC55630	Severe acute respiratory...	Albert,J., et al.	2023-04-14	B.1	01_SE100_21CS502478	Seven	
<input type="checkbox"/>	WGC55642	Severe acute respiratory...	Albert,J., et al.	2023-04-14	B.1	01_SE100_21CS502479	Seven	
<input type="checkbox"/>	WGC55666	Severe acute respiratory...	Albert,J., et al.	2023-04-14	B.1	01_SE100_21CS502003	Seven	
<input type="checkbox"/>	WGC55678	Severe acute respiratory...	Albert,J., et al.	2023-04-14	B.1	01_SE100_21CS502008	Seven	
<input type="checkbox"/>	WGC55690	Severe acute respiratory...	Albert,J., et al.	2023-04-14	B.1	01_SE100_21CS501379	Seven	
<input type="checkbox"/>	WGC55700	Severe acute respiratory...	Albert,J., et al.	2023-04-14	B.1	01_SE100_21CS502006	Seven	

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

Results Table

Dashboard Visualizations

Variants Overview

New!

New! Randomized subsets in Downloads

You now have the option of downloading a smaller, randomized documentation has more information.

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

complete

Pango lineage

Random Sampling

Isolate

Proteins

membrane glycoprotein

Nucleotide (559)

Accession

WIF20399

WGG89488

WGC55534

WGC55546

WGC55570

WGC55594

WGC55606

WGC55630

WGC55642

WGC55666

WGC55678

Severe acute respiratory...

Albert, J., et al.

2023-04-14

B.1

01_SE100_21CS502478

Severe acute respirator...

222

complete

membrane glycoprotein

Sweden

Severe acute respiratory...

Albert, J., et al.

2023-04-14

B.1

01_SE100_21CS502479

Severe acute respirator...

222

complete

membrane glycoprotein

Sweden

Severe acute respiratory...

Albert, J., et al.

2023-04-14

B.1

01_SE100_21CS502003

Severe acute respirator...

222

complete

membrane glycoprotein

Sweden

Severe acute respiratory...

Albert, J., et al.

2023-04-14

B.1

01_SE100_21CS502008

Severe acute respirator...

222

complete

membrane glycoprotein

Sweden

Download Results

Step 1 of 3: Select Data Type

Sequence Data (FASTA format)

Accession List

Results Table

Nucleotide

Coding Region

Protein

Nucleotide

Protein

Assembly

CSV format

XML format

Next

Now repeat the steps with your 3 countries

Copy your file to logrus

- ssh into logrus
- Make a directory called NCBIdata
- cd to NCBIdata
- Secure copy fasta file from your computer to logrus

Getting familiar with data in files

- How many accession numbers are there for one of your countries?