

# Navigating NCBI Using the NCBI Accession Number

## Locating the Pangolin Lineage

- 1- You do not need a user account to follow these instructions. Navigate to the NCBI database at: <https://www.ncbi.nlm.nih.gov/search/>. Under COVID-19 resources click on **NCBI Virus**.

The screenshot shows the NCBI homepage. At the top is the NIH logo and 'National Library of Medicine' text. Below is a search bar. A prominent orange banner for 'COVID-19 Information' is visible, with links to CDC, NIH, SARS-CoV-2 data, and HHS. Under the 'News' section, there are recent blog posts and research news. At the bottom, the 'COVID-19 resources' section features three tiles: 'NCBI Virus' (highlighted with an orange arrow), 'LitCovid', and 'BLAST'. The 'NCBI Virus' tile describes it as the most up-to-date set of SARS-CoV-2 nucleotide and protein sequences.

- 2- You will see the SARS-CoV-2 Data Hub. On the left-hand menu, under **Refine Results**, click **Accession**.

The screenshot shows the SARS-CoV-2 Data Hub interface. It has a top navigation bar with 'Download' and 'Quick Links'. Below is a tabular view of the data. On the left, a 'Refine Results' sidebar is visible, with 'Accession' highlighted by an orange arrow. The main table displays columns for 'Nucleotide (1,221,198)', 'Protein (6,032,887)', and 'RefSeq Genome (1)'. The 'Accession' column is expanded, showing a list of sequences with their respective submitters, release dates, and isolate names.

Accession	Submitters	Release Date	Pangolin	Isolate
<input type="checkbox"/> <a href="#">NC_045512</a>	Wu,F., et al.	2020-01-13	B	Wuhan-Hu-1
<input type="checkbox"/> <a href="#">BS001137</a>	Mitsunaga,...	2021-08-24		KH879
<input type="checkbox"/> <a href="#">BS001138</a>	Mitsunaga,...	2021-08-24		KH886
<input type="checkbox"/> <a href="#">BS001139</a>	Mitsunaga,...	2021-08-24		KH887
<input type="checkbox"/> <a href="#">BS001140</a>	Mitsunaga,...	2021-08-24		KH892
<input type="checkbox"/> <a href="#">BS001141</a>	Mitsunaga,...	2021-08-24		KH893

- 3- This will expand the pane > Enter in the **ACCESSION NUMBER** provided on the result report > then click **Submit**.

Refine Results [Reset](#)

Virus +

Wuhan seafood market pneumonia virus, taxid:2697049 x

Accession -

MZ851949

Submit

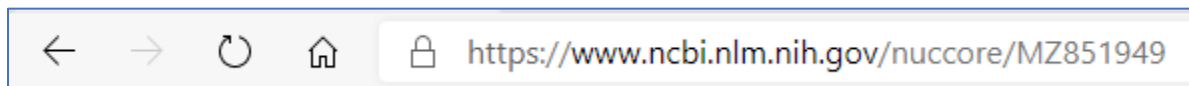
- 4- Result for the Accession number will show up on the table to the right. If you are only looking for the lineage, look under the Pangolin heading. Your use of NCBI is complete.

Nucleotide (1)		Protein (11)		RefSeq Genome (0)					Select Columns
<input type="checkbox"/>	Accession	Submitters	Release Date	Pangolin	Isolate	Species	Molecule type	Length	
<input type="checkbox"/>	<a href="#">MZ851949</a>	Razzaque,R.	2021-08-20	B.1.617.2	OSPHL01837	Severe acute respiratory s...	ssRNA(+)	2971	

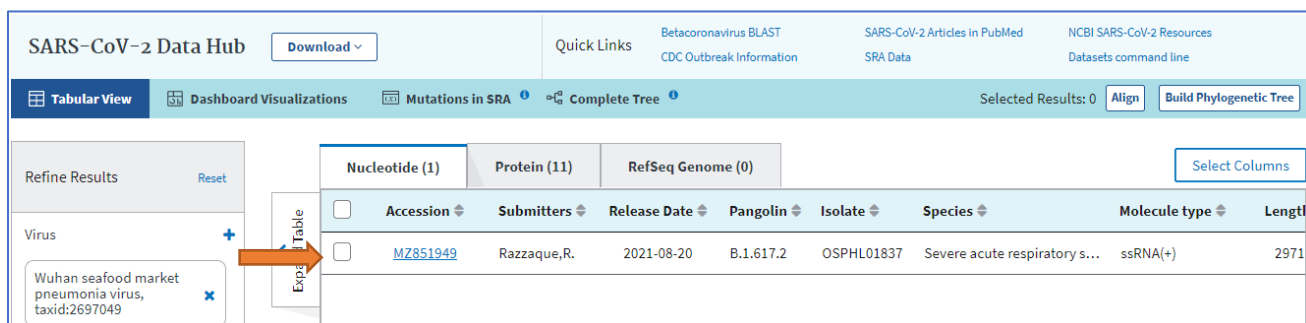
## Locating the FASTA Sequence

**Option #1:** If only FASTA sequence is desired, you can obtain this quickly by entering the following into your browser: [https://www.ncbi.nlm.nih.gov/nuccore/\[AccessionNumber\]](https://www.ncbi.nlm.nih.gov/nuccore/[AccessionNumber])

Here is an example: <https://www.ncbi.nlm.nih.gov/nuccore/MZ851949>



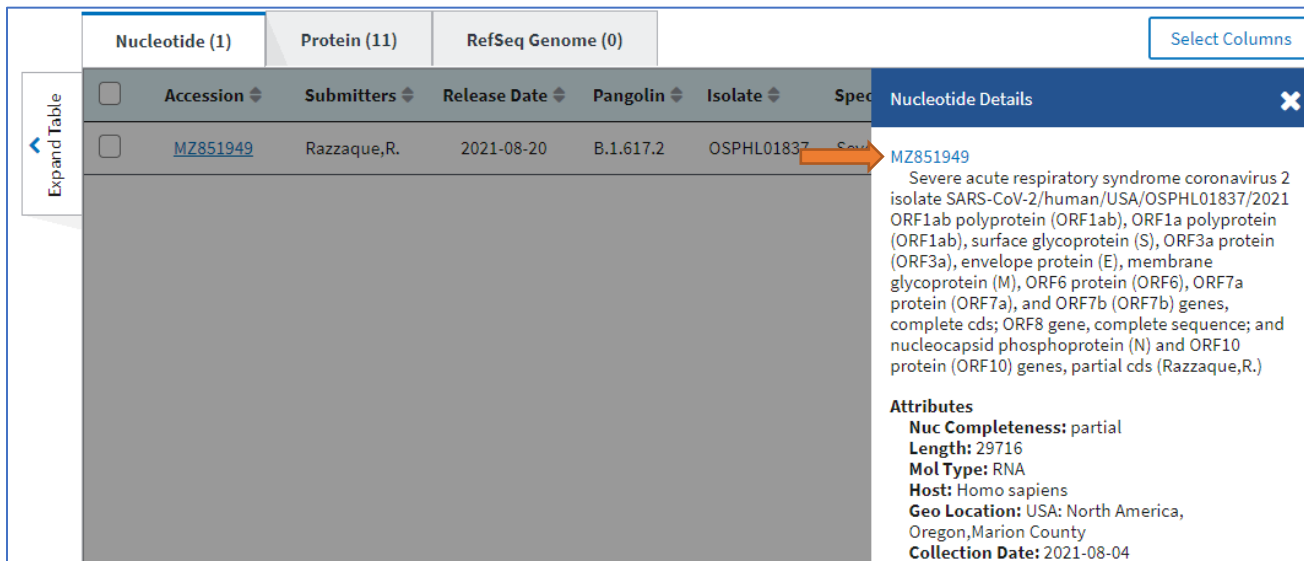
**Option #2:** If you have already navigated to the screen shown below, click on the **Accession number**.



The screenshot shows the SARS-CoV-2 Data Hub interface. At the top, there's a 'Download' button and 'Quick Links' for Betacoronavirus BLAST, CDC Outbreak Information, SARS-CoV-2 Articles in PubMed, SRA Data, and NCBI SARS-CoV-2 Resources. Below this is a navigation bar with 'Tabular View' selected, along with 'Dashboard Visualizations', 'Mutations in SRA', and 'Complete Tree'. On the right, it says 'Selected Results: 0' with 'Align' and 'Build Phylogenetic Tree' buttons. A 'Refine Results' sidebar on the left shows 'Virus' with a search box containing 'Wuhan seafood market pneumonia virus, taxid:2697049'. The main table has tabs for 'Nucleotide (1)', 'Protein (11)', and 'RefSeq Genome (0)'. The table columns are 'Accession', 'Submitters', 'Release Date', 'Pangolin', 'Isolate', 'Species', 'Molecule type', and 'Length'. One row is visible with the accession number 'MZ851949'. An orange arrow points from the 'Accession' column to the 'MZ851949' link.

Accession	Submitters	Release Date	Pangolin	Isolate	Species	Molecule type	Length
<a href="#">MZ851949</a>	Razzaque,R.	2021-08-20	B.1.617.2	OSPHL01837	Severe acute respiratory s...	ssRNA(+)	2971

5- A window will open to show more information about the sequence. Click on the Accession Number.



The screenshot shows the 'Nucleotide Details' window for accession MZ851949. The window is titled 'Nucleotide Details' and has a close button. It displays the following information:

- Accession:** [MZ851949](#)
- Submitters:** Razzaque,R.
- Release Date:** 2021-08-20
- Pangolin:** B.1.617.2
- Isolate:** OSPHL01837
- Species:** Severe acute respiratory syndrome coronavirus 2

**Attributes:**

- Nuc Completeness:** partial
- Length:** 29716
- Mol Type:** RNA
- Host:** Homo sapiens
- Geo Location:** USA: North America, Oregon, Marion County
- Collection Date:** 2021-08-04

An orange arrow points from the 'MZ851949' link in the table to the 'Nucleotide Details' window.

- 6- A new browser tab will open showing all the details and sequence information.

NCBI Resources How To

Nucleotide Nucleotide Search Help

Advanced

COVID-19 Information  
Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

GenBank MZ851941.1

FASTA Graphics

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/OSPHL01731/2021 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein...

LOCUS MZ851941 29787 bp RNA linear VRL 20-AUG-2021

DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/OSPHL01731/2021 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), and ORF7b (ORF7b) genes, complete cds; ORF8 gene, complete sequence; and nucleocapsid phosphoprotein (N) and ORF10 protein (ORF10) genes, complete cds.

ACCESSION MZ851941

VERSION MZ851941.1

KEYWORDS .

SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

ORGANISM Severe acute respiratory syndrome coronavirus 2

REFERENCE 1 (bases 1 to 29787)

AUTHORS Razzaque, R.

TITLE Direct Submission

JOURNAL Submitted (20-AUG-2021) Virology and Immunology, Oregon State Public Health Laboratory, 7202 NE Evergreen Parkway, Suite 100, Hillsboro, OR 97124, USA

COMMENT ##Assembly-Data-START##  
Assembly Method : generateConsensus v. v5.16..5  
Sequencing Technology : IonTorrent  
##Assembly-Data-END##

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

NCBI Virus

Retrieve, view, and download SARS-CoV-2 coronavirus genomic and protein sequences.

Recent activity

Turn Off Clear

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2

MZ733622 : MZ733659[accn] (0)

MZ766085 : MZ766117[accn] (0)

MZ668347 : MZ668382[accn] (36)

MZ682035 : MZ682069[accn] (0)

See more...

- 7- Click on **FASTA** to get the FASTA sequence to be used for further analysis through other platforms.

GenBank Send to:

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/OSPHL01837/2021 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein...

GenBank: MZ851949.1

FASTA Graphics

Go to:

LOCUS MZ851949 29716 bp RNA linear VRL 20-AUG-2021

DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/OSPHL01837/2021 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), and ORF7b (ORF7b) genes, complete cds; ORF8 gene, complete sequence; and nucleocapsid phosphoprotein (N) and ORF10 protein (ORF10) genes, partial cds.

ACCESSION MZ851949

VERSION MZ851949.1

KEYWORDS .

SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

ORGANISM Severe acute respiratory syndrome coronavirus 2

Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Coronavirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus

8- FASTA sequence will open in this format:

**Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/OSPHL01837/2021 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein...**

GenBank: MZ851949.1

[GenBank](#) [Graphics](#)

>MZ851949.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/OSPHL01837/2021 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), and ORF7b (ORF7b) genes, complete cds; ORF8 gene, complete sequence; and nucleocapsid phosphoprotein (N) and ORF10 protein (ORF10) genes, partial cds

```
GTAGATCTGTTCTCTAAACGAACTTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCA
CGCAGTATAATTAAATAACTAATTACTGTCTTGACAGGACACGAGTAACCTGCTCTATCTTCTGCAGGCTG
CTTACGGTTTCTGTCGGTTTTCGACGCCGATCATCAGCACATCTAGGTTTGTCCGGGTGTGACCGAAAGGT
AAGATGGAGAGCCTTGTCCCTGGTTTCAACGAGAAACACACGTCCTCACTGCTTGTCCCTGTTTACAGG
TTCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGACGCTCAACATCT
TAAAGATGGCACTTGTGGCTTAGTAGAAGTTGAAAAAGGCGTTTGGCTCAACTTGAACAGCCCTATGTG
TTCATCAAAACGTTTCGGATGCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCAGAACTCG
AAGGCATTACGTACGGTCTGAGTGGTGAGACACTTGGTGTCTTGTCCCTCATGTGGCGAAATACCACT
GGCTTACCGCAAGGTTCTTCTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGAT
CTAAAGTCATTTGACTTAGGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTCAAGAAAACTGGAACA
CTAAACATAGCAGTGGTGTACCCGTGAACCTCATGCTGAGCTTAACGGAGGGGCATACACTCGCTATGT
CGATAACAACTTCTGTGGCCTGATGGCTACCCCTTGTAGTGCACTTAAAGACCTTCTAGCACGTGCTGGT
AAAGCTTCATGCACTTTGTCCGAACAACCTGGACTTTATTGACACTAAGAGGGGTGTACTGCTGCCGTG
AACATGAGCATGAATTTGCTTGGTACACGGAACGTTCTGAAAGAGCTATGAATTCAGACACCTTTTGA
AATTAATTTGGCAAGAAATTTGACACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
TCTATCCAGTTGCGTCCACCAATGAATGCAACCAATGTGCTTTCAACTCTCATGAAGTGTGATCATTG
TGGTGAACTTCATGGCAGAGGGCGATTTTGTAAAGCACTTGCGAATTTTGTGGCACTGAGAATTTG
ACTAAAGAGGTGCCACTACTTGTGTTACTTACCCCAAAATGCTGTTGTTAAATTTATTGTCCAGCAT
GTCACAATTGAGAAGTAGGACCTGAGCATAGTCTTGCCGAATACCATATGAATCTGGCTTGAACCACT
TCTTCTGAAGGGTGGTGCATTTATGCTTGGAGGCTGTGTTCTTATGTTGGTTGCCATAACAAG
TGTGCCATTGGGTTCCAGTGCTAGCGCTAACATAGGTTGTAACCATACAGGTGTTGTTGGAGAAGGTT
CCGAAGGCTTAAATGACAACCTTCTTGAATACTCCAAAAAGAGAAAGTCAACATCAATATTGTTGGTGA
CTTTAAACTTAATGAAGAGATCGCCATTATTTTGGCATCTTTTCTGCTTCCACAAGTGCTTTTGTGGAA
ACTGTGAAGGTTTGGATTATAAGCATTCAACAAATTTGTTGAATCCTGTGGTAATTTTAAAGTTACAA
AAGGAAAAGCTAAAAAAGGTGCTTGAATATTGGTGAACAGAAATCAATACTGAGTCTCTTTATGCATT
TGCATCAGAGGCTGCTCGTGTGTACGATCAATTTTCTCCCGCACTCTTGAACCTGCTCAAAATCTGTG
CGTGTTTTACAGAAGGCCGCTATAACAATACTAGATGGAATTTACAGTATTCAGTGAAGTCTTATGATG
CTATGATGTTACATCTGATTTGGCTACTAACAATCTAGTTGTAATGGCTTACATTACAGGTGGTGTGT
TCAGTTGACTTCGCACTGGCTAACTAACATCTTTGGCACTGTTTATGAAAACTCAACCCGCTCTTGAT
TGGCTTGAAGAGAGGTTTAAAGGAAGGTGTAGAGTTTCTTAGAGACGGTTGGGAAATTTGTTAAATTTATCT
CAACCTGTGCTTGTGAAATTTGCGGTGGACAAATTTGACCTGTGCAAGGAAATTAAGGAGAGTGTCA
GACATCTTTAAGCTTGTAAATAAATTTTGGCTTTGTGTGCTGACTTATCATTATTGNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
CCAGAGAAGAACTGGCTACTCATGCCTCTAAAAGCCCCAAAAGAAATATCTTCTTAGAGGGGAGAAAC
ACTTCCACAGAAAGTGTAAACAGAGGAAGTTGTCTTGAAGAACTGGTGATTACAAACATTAGAACAACT
ACTAGTGAAGCTGTTGAAGCTCCATTGGTTGGTACACCAAGTTTGTATTAAACGGGCTTATGTTGCTCGAAA
TCAAGACACAGAAAGTACTGTGCCCTTGACCTAATATGATGGTAACAAACAATACCTTACACTCAA
AGGCGGTGCACCAACAAGGTTACTTTGGTGATGACACTGTGATAGAAGTGCAAGGTTACAAGAGTGTG
AATATCACTTTTGAACCTTGATGAAAGGATTGATAAGTACTTAATGAGAAGTGCTCGCTATACAGTTG
AACTCGGTACAGAAAGTAAAGTGGCTGTGTTGTGGCAGATGCTGTCATAAAAACTTTGCAACCACT
```