



# SARS-CoV-2 GENOMICS: BIOINFORMATICS WORKFLOWS (PANGOLin and COVDB online tools)

21 May 2021

**Daniel Gyamfi Amoako, PhD**

Centre for Respiratory Diseases and Meningitis,  
National Institute for Communicable Diseases,  
A division of the National Health Laboratory Service

 **NATIONAL INSTITUTE FOR  
COMMUNICABLE DISEASES**

Division of the National Health Laboratory Service

# Overview

- PANGOLin (<https://pangolin.cog-uk.io/>)
- Covdb tool (<https://covdb.stanford.edu/sierra/sars2/by-sequences/>)
- GISAID (<https://www.gisaid.org/>)

# PANGOLin (PANGO lineages)

- PHYLOGENETIC ASSIGNMENT OF NAMED GLOBAL OUTBREAK LINEAGES
- **Lineages** are sequences of biological entities connected by ancestry-descent relationships.
- Pangolin was developed to implement the dynamic nomenclature of SARS-CoV-2 lineages, known as the **Pango nomenclature**.
- It allows a user to assign a SARS-CoV-2 genome sequence the most likely lineage (Pango lineage) to SARS-CoV-2 query sequences.
- **Web-based** and command line tool.

# Overview

## PANGO lineages

HOME

PANGO LINEAGES

MAJOR LINEAGES

SUGGEST A NEW LINEAGE

PANGOLIN

PANGOLIN DOCS

INTERNATIONAL LINEAGE REPORT

LLAMA

FREQUENTLY ASKED QUESTIONS

CONTRIBUTORS

HOW TO CITE

<https://cov-lineages.org/pangolin.html>

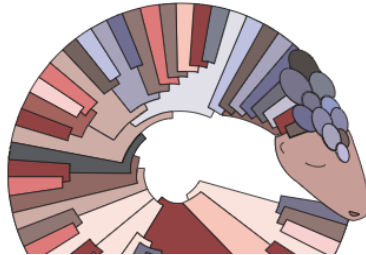


## PANGO lineages

# pangolin

## PHYLOGENETIC ASSIGNMENT OF NAMED GLOBAL OUTBREAK LINEAGES

pangolin was developed to implement the dynamic nomenclature of SARS-CoV-2 lineages, known as the Pango nomenclature. It allows a user to assign a SARS-CoV-2 genome sequence the most likely lineage (Pango lineage) to SARS-CoV-2 query sequences.



# Overview

INTERNATIONAL LINEAGE REPORT ▼

---

LLAMA

---

FREQUENTLY ASKED QUESTIONS


---

CONTRIBUTORS

---

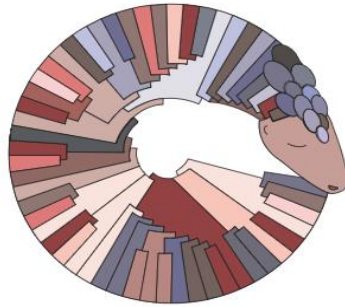
HOW TO CITE

---



---

© SARS-CoV-2 lineages  
Design: [HTML5 UP](#)  
[Andrew Banchich](#)  
[Áine O'Toole](#)



## Command-line tool

GNU General Public License v3.0

## QUICK LINKS

- [How it works](#)
- [Dependencies](#)
- [Usage](#)
- [Model training](#)
- [Installation](#)
- [Output](#)
- [pangolin tutorial](#)
- [Updating](#)
- [Citing pangolin](#)



## Web application

Developed by the [Centre for Genomic Pathogen Surveillance](#).

# Web-Interface (<https://pangolin.cog-uk.io/>)

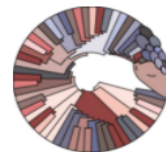
Drag and drop fasta file

Select fasta file to upload

## Pangolin COVID-19

### Lineage Assigner

Phylogenetic Assignment of Named  
Global Outbreak LINEages



You can upload one or more sequences by dragging and dropping a (multi)fasta file or clicking "Select fasta file to upload" and selecting a (multi)fasta file.


This Web Application assigns lineages to COVID-19 sequences based on the methodology described in this [article](#)








The software to assign lineages based on the algorithm that was developed by [Áine O'Toole](#), [Verity Hill](#), [JT McCrone](#), [Emily Scher](#) and [Andrew Rambaut](#).

The source code can be found [here](#)

Recommended browsers   or 

# Uploading files and start analysis

 [Start analysis](#) [Reset entries](#) [Upload another file](#) [Help](#)

File name	Sequence name	Lineage	Assignment Conflict
— READY FOR ANALYSIS 7 sequences			
 Multi-Fasta.fasta	EPI_ISL_2003169		
 Multi-Fasta.fasta	hCoV-19/SouthAfrica/819503/2021		
 Multi-Fasta.fasta	hCoV-19/SouthAfrica/819931/2021		
 Multi-Fasta.fasta	hCoV-19/SouthAfrica/400919678/2021		
 Multi-Fasta.fasta	2019-nCoV_MN908947 S03911-20		
 Multi-Fasta.fasta	N2618		
 Multi-Fasta.fasta	N2625		

# Analysis and Interpretation


























Retry Failed Sequences

Reset entries

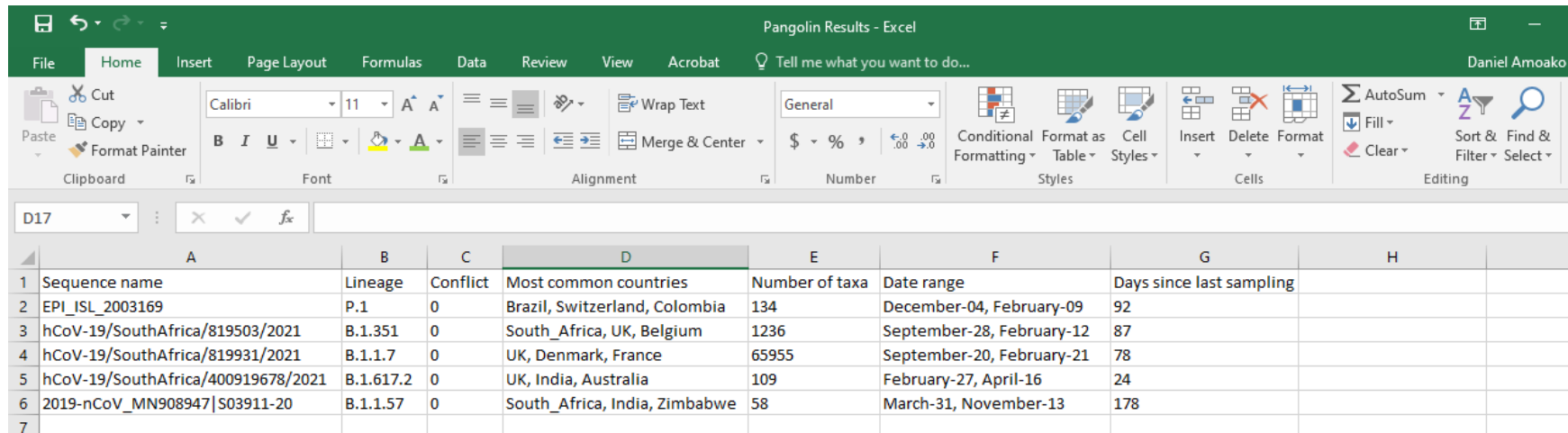
Upload another file

Help

File name	Sequence name	Lineage	Assignment Conflict
— FAILED (Click warning icon for more info) 2 sequences			
 Multi-Fasta.fasta	N2618		
 Multi-Fasta.fasta	N2625		
— ANALYSED (Click tick icon for more info) 5 sequences 			
 Multi-Fasta.fasta	EPI_ISL_2003169	P.1   	0
 Multi-Fasta.fasta	hCoV-19/SouthAfrica/819503/2021	B.1.351   	0
 Multi-Fasta.fasta	hCoV-19/SouthAfrica/819931/2021	B.1.1.7   	0
 Multi-Fasta.fasta	hCoV-19/SouthAfrica/400919678/2021	B.1.617.2   	0.0
 Multi-Fasta.fasta	2019-nCoV_MN908947 S03911-20	B.1.1.57   	0.0



# Downloading Results and Quality Calling



Pangolin Results - Excel

File Home Insert Page Layout Formulas Data Review View Acrobat Tell me what you want to do... Daniel Amoako

Clipboard Font Alignment Number Styles Cells Editing

	A	B	C	D	E	F	G	H
1	Sequence name	Lineage	Conflict	Most common countries	Number of taxa	Date range	Days since last sampling	
2	EPI_ISL_2003169	P.1	0	Brazil, Switzerland, Colombia	134	December-04, February-09	92	
3	hCoV-19/SouthAfrica/819503/2021	B.1.351	0	South_Africa, UK, Belgium	1236	September-28, February-12	87	
4	hCoV-19/SouthAfrica/819931/2021	B.1.1.7	0	UK, Denmark, France	65955	September-20, February-21	78	
5	hCoV-19/SouthAfrica/400919678/2021	B.1.617.2	0	UK, India, Australia	109	February-27, April-16	24	
6	2019-nCoV_MN908947 S03911-20	B.1.1.57	0	South_Africa, India, Zimbabwe	58	March-31, November-13	178	
7								

## Quality Calling:

- Passed\_qc
- Failed\_qc

## Coverage vs Quality Calling

- N\_content
- Sequence length

# Downloading Results and Quality Calling

	A	B	C	D	E	F	G	H	I
65	N5193_S77_L001	B.1.351	0	2.4.2	5/12/2021	1.2.2	passed_qc	9/9 B.1.351 SNPs (0 ref and 0 other)	
66	N5206_S90_L001	B.1.351	0	2.4.2	5/12/2021	1.2.2	passed_qc	6/9 B.1.351 SNPs (0 ref and 3 other)	
67	N5208_S92_L001	B.1.351	0	2.4.2	5/12/2021	1.2.2	passed_qc	9/9 B.1.351 SNPs (0 ref and 0 other)	
68	Negative-2_S192_L001	B.1.351	0	2.4.2	5/12/2021	1.2.2	passed_qc	7/9 B.1.351 SNPs (0 ref and 2 other)	
69	Positive-1_S95_L001	B.1	0	2.4.2	5/12/2021	1.2.2	passed_qc		
70	Positive-3_S287_L001	B.1	0	2.4.2	5/12/2021	1.2.2	passed_qc		
71	N2618_S295_L001	None	0	2.4.2	5/12/2021	1.2.2	fail	seq_len:330	
72	N2625_S302_L001	None	0	2.4.2	5/12/2021	1.2.2	fail	N_content:0.67	
73	N2626_S303_L001	None	0	2.4.2	5/12/2021	1.2.2	fail	N_content:0.55	
74	N2629_S306_L001	None	0	2.4.2	5/12/2021	1.2.2	fail	N_content:0.72	
75	N2630_S307_L001	None	0	2.4.2	5/12/2021	1.2.2	fail	N_content:0.77	
76	N2631_S308_L001	None	0	2.4.2	5/12/2021	1.2.2	fail	N_content:0.67	
77	N2632_S309_L001	None	0	2.4.2	5/12/2021	1.2.2	fail	N_content:0.89	
78	N2634_S311_L001	None	0	2.4.2	5/12/2021	1.2.2	fail	seq_len:6299	

## Quality Calling:

- Passed\_qc
- Failed\_qc

## Coverage (threshold) vs Quality Calling

- N\_content
- Sequence length

Stanford tool (<https://covdb.stanford.edu/sierra/sars2/by-sequences/>)

## ➤ Different options for inputs

The screenshot shows the web interface for the Stanford University CORONAVIRUS ANTIVIRAL & RESISTANCE DATABASE. The header includes the university logo, the database name, and navigation links: HOME, SEARCH, ANALYSIS PROGRAM, DRUGS, TRIALS, CITATION, and a highlighted SUPPORT COVDB button. Below the header, the page title is "SARS-CoV-2 Sequence Analysis". A descriptive paragraph explains that sequences can be entered as plain text or uploaded as a file, with a limit of 100 sequences per file. The main form has three tabs: "Input mutations", "Input sequences" (which is selected), and "Input sequence reads". Under the "Input sequences" tab, there is a "Header:" label followed by a text input field and the word "(optional)". Below this is an "Upload text file:" section with a "Choose File" button, the text "No file chosen", and a "Load Examples" link. A large, empty text area is provided for pasting sequences. At the bottom of the form, there is an "Output options" section with two radio buttons: "HTML" (which is selected) and "Printable HTML". At the very bottom right of the form are "Reset" and "Analyze" buttons.

Stanford University  
**CORONAVIRUS ANTIVIRAL & RESISTANCE DATABASE**  
A Stanford HIVDB team website. Last updated on 5/19/2021, 7:25:11 AM.

HOME SEARCH ANALYSIS PROGRAM DRUGS TRIALS CITATION **SUPPORT COVDB**

### SARS-CoV-2 Sequence Analysis

SARS-CoV-2 Sequences can be entered as plain text if just one sequence is entered. Sequences must be entered using the FASTA format if multiple sequences are entered. Sequences can be pasted in the text box or uploaded using the File Upload option. The upper limit is currently 100 sequences containing ~30000 nucleotides per sequence.

Input mutations **Input sequences** Input sequence reads

Header:  (optional)

Upload text file: **Choose File** No file chosen [Load Examples](#)

Output options

HTML  Printable HTML

Reset Analyze

Stanford tool (<https://covdb.stanford.edu/sierra/sars2/by-sequences/>)

## Outputs

- Sequence summary (Information on the specific genes coverage)
- Pangolin lineage assignment
- Outbreak information
- Sequence quality assessment (graphical visualization)
- Mutation list
- Mutation comments
- MAb susceptibility summary
- Convalescent plasma susceptibility summary
- Plasma from vaccinated persons susceptibility summary

Stanford tool (<https://covdb.stanford.edu/sierra/sars2/by-sequences/>)

## Outputs

- Sequence summary (Information on the specific genes coverage)
- Pangolin lineage assignment
- Outbreak information
- Sequence quality assessment (graphical visualization)
- Mutation list
- Mutation comments
- MAb susceptibility summary
- Convalescent plasma susceptibility summary
- Plasma from vaccinated persons susceptibility summary

# INPUT SEQUENCES AND ANALYSE

## SARS-CoV-2 Sequence Analysis

SARS-CoV-2 Sequences can be entered as plain text if just one sequence is entered. Sequences must be entered using the FASTA format if multiple sequences are entered. Sequences can be pasted in the text box or uploaded using the File Upload option. The upper limit is currently 100 sequences containing ~30000 nucleotides per sequence.

Input mutations   Input sequences   Input sequence reads

Header:  (optional)

Upload text file: [Choose File](#) [Multi-Fasta.fasta](#) [Load Examples](#)

```
>EPI_ISL_2003169
CCTTCCAGGTAACAACCAACCAACTTTTCGATCTCTGTAGATCTGTTCTCTAAACGAACTTAAAACTGTGTGGCTG
TCACTCGGCTGCATGCTTAGTGCACTACGCAGTATAATTAATACTAATTAAGTTCGTTGACAGGACACGAGTAACTCG
TCTATCTTTCGCAGGCTGCTTACGGTTTCGTCGGTGTTCAGCCGATCATCAGCACATCTAGGTTTTGTCCGGGTGTGAC
CGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTCAACGAGAAAAACACACGTCCAACCTCAGTTTGCCTGTTTTACAGGTT
CGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGGTCCTATCAGAGGCACGTCAACATCTTAAAGATGGCAC
TTGTGGCTTAGTAGAAGTTGAAAAAGGGCTTTTGCTCAACTTGAAACAGCCCTATGTTTCATCAAACGTTCCGGATGCTC
```

Output options

HTML    Printable HTML

Reset

Analyze

# OUTPUTS



## 1. EPI\_ISL\_2003169

### Sequence summary

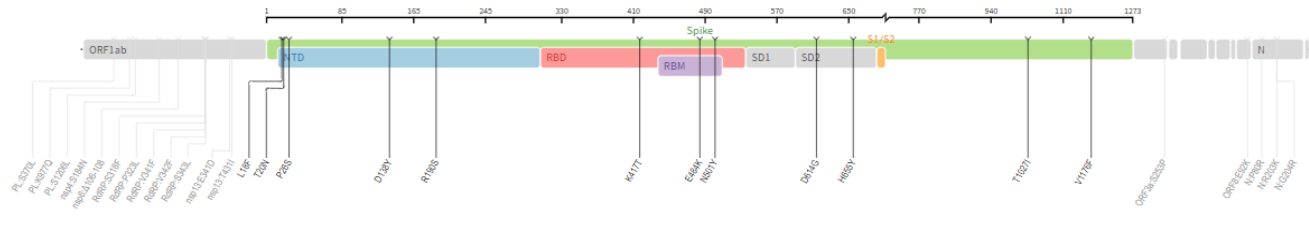
Sequence includes following genes: nsp1 • nsp2 • PLpro • nsp4 • 3CLpro • nsp6 • nsp7 • nsp8 • nsp9 • nsp10 • RdRP • nsp13 • nsp14 • nsp15 • nsp16 • Spike • ORF3a • E • M  
• ORF6 • ORF7a • ORF7b • ORF8 • N • ORF10

PANGO lineage: P.1 (Prob=1.0; pangolin: 2.4.2; pangoLEARN: 2021-05-12)

Outbreak.info: P.1 (n=16,994)

[Pretty pairwise](#)

### Sequence quality assessment



1. EPI\_ISL\_2003169

1

2

3

4

5

6

7

This submission contains 7 sequences.

#### Mutation list

**PLpro:** S370L • K977Q • S1206L

**nsp4:** S184N

**nsp6:** Δ106-108

**RdRP:** S318F • P323L • V341F • V342F • S343L

**nsp13:** E341D • T431I

**Spike:** L18F • T20N • P26S • D138Y • R190S • K417T • E484K • N501Y • D614G • H655Y • T1027I  
• V1176F

**ORF3a:** S253P

**ORF8:** E92K

**N:** P80R • R203K • G204R

#### Mutation comments

Last updated on May 1, 2021

- **nsp6:Δ106**  
Each of the WHO-defined variants of concern, one of the two CDC-defined variants of concern, and a recently described novel E484K-containing U.S. lineage contains the same deletion of amino acids 104 to 106 in the non-structural protein 6 (nsp6) a component of the SARS-CoV-2 membrane-tethered replication complex that is also involved in antagonizes host-responsiveness to interferons<sup>[1][2][3][4]</sup>. Although several additional non-spike mutations have been shared by more than one variant of concern, none have occurred as commonly as this mutation<sup>[5]</sup>.
- **L18F**  
L18F is an NTD mutation present in B.1.351, P.1, and a sub-lineage of B.1.1.7<sup>[6]</sup>. It is associated with reduced susceptibility to several NTD-binding mAbs but by itself does not appear to reduce susceptibility to plasma from convalescent or vaccinated persons<sup>[7][8]</sup>.
- **K417T**  
K417N/T are ACE2-binding site RBM mutations present in the WHO-defined variants of concern B.1.351 (K417N) and P.1 (K417T). Both mutations reduce ACE2 binding<sup>[9][10]</sup> and rarely occur in other virus variants. K417N confers >100-fold reduced susceptibility to ETE<sup>[11]</sup> and ~30-fold reduced susceptibility to CAS<sup>[7]</sup> but appears to retain susceptibility to the remaining mAbs in advanced clinical development<sup>[7]</sup>. K417N/T retain full susceptibility to plasma samples from convalescent persons.
- **E484K**



# SARS-CoV-2-related Database- GISAID (<https://www.gisaid.org/>)

**GISAID** © 2008 - 2021 | Terms of Use | Privacy Notice | Contact

You are logged in as **Daniel Gyamfi Amoako** - [logout](#)

Registered Users | EpiFlu™ | **EpiCoV™** | My profile

EpiCoV™ Search Downloads Upload My Unreleased

## Pandemic coronavirus causing COVID-19

A previously unknown human coronavirus (hCoV-19) was first detected in late 2019 in patients in the City of Wuhan, who suffered from respiratory illnesses including atypical pneumonia, an illness that has become known as coronavirus disease (COVID-19). The coronavirus originated from an animal host and is closely related to the virus responsible for the Severe Acute Respiratory Syndrome coronavirus (SARS).

On 10. January 2020, the first virus genomes and associated data were publicly shared via GISAID. The World Health Organization announced on 11. March 2020 the first coronavirus pandemic. As the pandemic progresses, scientists from around the globe are tracking the virus and its genome sequences to ensure optimal virus diagnostic tests, to track and trace the ongoing outbreak and to identify potential intervention options. Several analyses to assist with these efforts are offered here, including sequence alignments, diagnostic primer and probe coordinates, 3D protein models, drug targets, phylogenetic trees and many more.

Search

by AT&T Singapore

Audacity

BLAST

CoVizu<sup>e</sup>

Emerging Variants

Official GISAID reference sequence

PrimerChecker

Important note: In the [GISAID EpiFlu™ Database Access Agreement](#), you have accepted certain terms and conditions for viewing and using data regarding influenza viruses. To the extent the Database contains data relating to non-influenza viruses, the viewing and use of these data is subject to the same terms and conditions, and by viewing or using such data you agree to be bound by the terms of the [GISAID EpiFlu™ Database Access Agreement](#) in respect of such data in the same manner as if they were data relating to influenza viruses.



# GISAID (General Overview)

- EpiCoV
- Search
- Downloads
- Upload
- BLAST tool
- PrimerChecker

# Using BLAST tool for Lineage confirmation

```
>hCoV-19/SouthAfrica/400919678/2021  
NNNNNNNNNNNNNNNNNNNNNNNAACAACCAACCAACTTTCGATCTCTTGTAGATCT  
GTTCTCTAAACGAACTTTAAATCTGTGTGGCTGTCACCTGGCTGCATGCTTAGTGACT  
CACGCAGTATAATTAATAACTAATTACTGTCGTTGACAGGACACGAGTAACCTGCTATC  
TTCGACGGCTGCTTACGGTTTCGTCGGTTTTGCAGCCGATCATCAGCACATCTAGGTTT  
TGTCGGGTGTGACCGAAAGGTAAGATGGAGACCTTGCCCTGGTTTCAACGAGAAAAC  
ACACGTCCAACCTCAGTTTGCCGTGTTTTACAGGTTCCGACGTGCTCGTACGTGGCTTTGG  
..
```

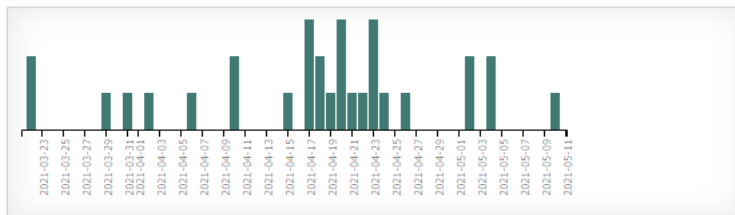
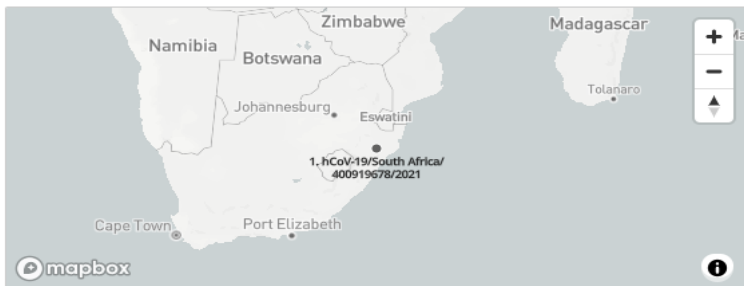
Submit hCoV-19 BLAST

# Identifying similar sequences

Nucleotide BLAST 2.6.0+ | 1,587,108 sequences | Updated 15 hour(s) and 29 minute(s) ago

## 1. Query: hCoV-19/SouthAfrica/400919678/2021

Subject collection date prior to query collection date HSPs ▼



1. Click on the green bars in above histogram for hit locations at specific dates
2. Click on the locations to center on map.

## 1. Subject: hCoV-19/South Africa/400919678/2021 [Download HSP\(s\)](#)

EPI\_ISL\_2013037 Location **Africa / South Africa / KwaZulu-Natal** Collection Date **2021-05-02** Length **29,822bp**

HSP 1 | Identity **100%**, Score **29750**, Bit-Score **54938.9**, eValue **0**, Length **29822**, Gaps **0**, From (Query/Subject) **26/1**, To (Query/Subject) **29847/29822**

```
AAACAAACCAACCAACTTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACCTTAAAAATCTGTGGCTGTCACCTCGGCTGCATGCTTAGTGCACTCACGCAGTATAAATTAATAACTAATTACTGTCGTTGACAGGCACACGAGTAACTCGTCTATCTTCTGCAAGGCTGCT
AAACAAACCAACCAACTTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACCTTAAAAATCTGTGTGGCTGTCACCTCGGCTGCATGCTTAGTGCACTCACGCAGTATAAATTAATAACTAATTACTGTCGTTGACAGGCACACGAGTAACTCGTCTATCTTCTGCAAGGCTGCT
```

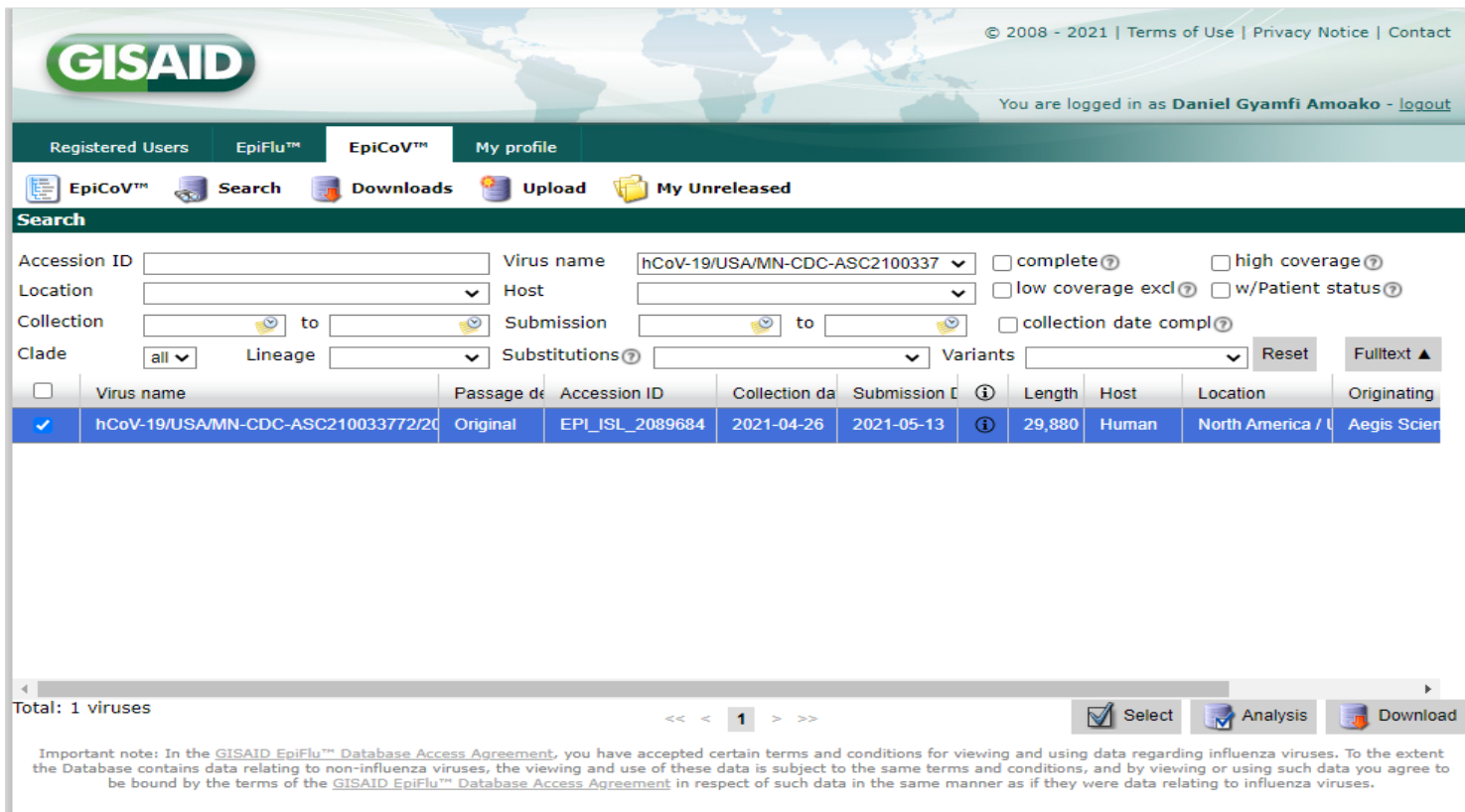
## 2. Subject: hCoV-19/USA/MN-CDC-ASC210033772/2021 [Download HSP\(s\)](#)

EPI\_ISL\_2089684 Location **North America / USA / Minnesota** Collection Date **2021-04-26** Length **29,880bp**

HSP 1 | Identity **99.85%**, Score **29722**, Bit-Score **54887.2**, eValue **0**, Length **29822**, Gaps **1**, From (Query/Subject) **26/16**, To (Query/Subject) **29847/29836**

```
AAACAAACCAACCAACTTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACCTTAAAAATCTGTGTGGCTGTCACCTCGGCTGCATGCTTAGTGCACTCACGCAGTATAAATTAATAACTAATTACTGTCGTTGACAGGCACACGAGTAACTCGTCTATCTTCTGCAAGGCTGCT
AAACAAACCAACCAACTTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACCTTAAAAATCTGTGTGGCTGTCACCTCGGCTGCATGCTTAGTGCACTCACGCAGTATAAATTAATAACTAATTACTGTCGTTGACAGGCACACGAGTAACTCGTCTATCTTCTGCAAGGCTGCT
```

# Searching for the similar sequences



**GISAID** © 2008 - 2021 | [Terms of Use](#) | [Privacy Notice](#) | [Contact](#)

You are logged in as **Daniel Gyamfi Amoako** - [logout](#)

Registered Users | [EpiFlu™](#) | **EpiCoV™** | [My profile](#)

[EpiCoV™](#) | [Search](#) | [Downloads](#) | [Upload](#) | [My Unreleased](#)

### Search

Accession ID:  Virus name:   complete  high coverage

Location:  Host:   low coverage excl  w/Patient status

Collection:  to  Submission:  to   collection date compl

Clade:  Lineage:  Substitutions:  Variants:  [Reset](#) [Fulltext ▲](#)

<input type="checkbox"/>	Virus name	Passage de	Accession ID	Collection da	Submission C	Length	Host	Location	Originating
<input checked="" type="checkbox"/>	hCoV-19/USA/MN-CDC-ASC210033772/20	Original	EPI_ISL_2089684	2021-04-26	2021-05-13	29,880	Human	North America / U	Aegis Scier

Total: 1 viruses

<< < 1 > >> [Select](#) [Analysis](#) [Download](#)


Important note: In the [GISAID EpiFlu™ Database Access Agreement](#), you have accepted certain terms and conditions for viewing and using data regarding influenza viruses. To the extent the Database contains data relating to non-influenza viruses, the viewing and use of these data is subject to the same terms and conditions, and by viewing or using such data you agree to be bound by the terms of the [GISAID EpiFlu™ Database Access Agreement](#) in respect of such data in the same manner as if they were data relating to influenza viruses.

# Confirming the set of mutations in the specific lineage

## Virus Detail

<b>Virus name:</b>	hCoV-19/USA/MN-CDC-ASC210033772/2021
<b>Accession ID:</b>	EPI_ISL_2089684
<b>Type:</b>	betacoronavirus
<b>Clade</b>	G
<b>Pango Lineage</b>	B.1.617.2 (version: 2021-05-12)
<b>AA Substitutions</b>	Spike D614G, Spike D950N, Spike E156G, Spike F157del, Spike L452R, Spike P681R, Spike R158del, Spike T19R, Spike T478K, M I82T, N D63G, N D377Y, N R203M, NS3 S26L, NS7a L116F, NS7a T120I, NS7a V82A, NSP2 P129L, NSP3 P822L, NSP4 A446V, NSP4 P29S, NSP6 V149A, NSP12 G671S, NSP12 P323L, NSP13 P77L, NSP14 P46L
<b>Variant</b>	G/452R.V3 (B.1.617+)
<b>Passage details/history:</b>	Original

## Sample information

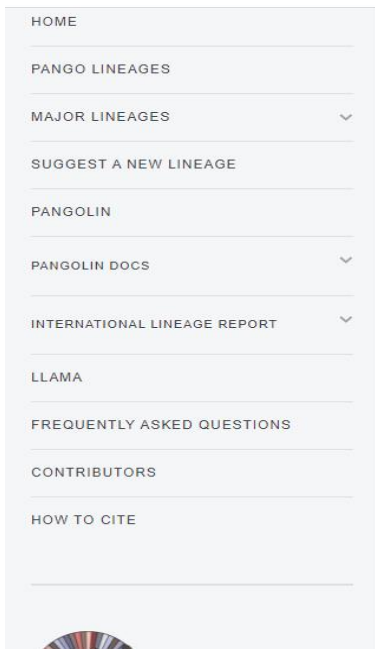
<b>Collection date:</b>	2021-04-26
<b>Location:</b>	North America / USA / Minnesota
<b>Host:</b>	Human
<b>Additional location information:</b>	
<b>Gender:</b>	Female
<b>Patient age:</b>	33
<b>Patient status:</b>	unknown
<b>Specimen source:</b>	Nasal - Anterior Nares
<b>Additional host information:</b>	
<b>Sampling strategy:</b>	
<b>Outbreak:</b>	
<b>Last vaccinated:</b>	
<b>Treatment:</b>	
<b>Sequencing technology:</b>	Illumina NovaSeq 6000
<b>Assembly method:</b>	Dragen COVID Lineage v3.5.1
<b>Coverage:</b>	2,059.93x
<b>Comment:</b>	 Gap of 13 nucleotides when compared to the reference sequence.

## Institute information

<b>Originating lab:</b>	Aegis Sciences Corporation
<b>Address:</b>	501 Great Circle Road, Nashville, TN 37228

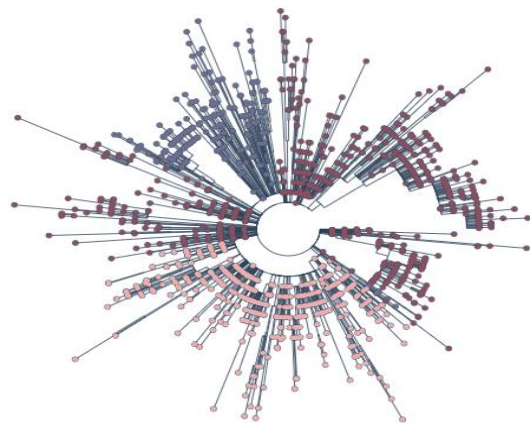
# Identified a new lineage

## What next?



## How to: lineage designation

So you think you've identified a new lineage? The following is a step-by-step guide of how to add your new lineage to the growing list of lineages we maintain and can then be assigned using pangolin.



LLAMA

FREQUENTLY ASKED QUESTIONS

CONTRIBUTORS

HOW TO CITE



© SARS-CoV-2 lineages

Design: HTML5 UP

Andrew Banchich

Aine O'Toole

## What to do

### 1. Does your cluster fulfil the definition of a lineage?

Refer to the [Pango lineage guide](#) to check if your cluster fits the new lineage guidelines.

### 2. Navigate to the new pango-designation repository

Go to the [github.com/cov-lineages/pango-designation](https://github.com/cov-lineages/pango-designation) repository, shown below. Notice the lineage\_notes.txt and lineages.csv files. These include the latest set of manually curated lineage designations.

Search or jump to... Pull requests Issues Marketplace Explore

cov-lineages / pango-designation

Unwatch 4 Star 0 Fork 2

Code Issues 1 Pull requests Actions Projects Wiki Security Insights Settings

master 1 branch 3 tags Go to file Add file Code

aineniamh Adding pango lineage B.1.526 to designation list 124 new sequences, m... 4b79a17 1 hour ago 17 commits

README.md	updating readme	2 hours ago
lineage_notes.txt	Adding pango lineage B.1.526 to designation list 124 new sequences, ...	1 hour ago
lineages.csv	Adding pango lineage B.1.526 to designation list 124 new sequences, ...	1 hour ago

README.md

### Panao desianation

About  
Repository for suggesting new lineages that should be added to the current scheme  
Readme

Releases 3  
pango designations v1.0.2 (Latest)  
40 minutes ago  
+ 2 releases





THANK YOU