



UNIVERSITY OF
KWAZULU-NATAL™
INYUVESI
YAKWAZULU-NATALI



INTRODUCTION TO GISAID

SAN EMMANUEL JAMES, HOURIYAH TEGALLY, EDUAN WILKINSON, TULIO DE OLIVEIRA ET. AL



EDGEWOOD CAMPUS



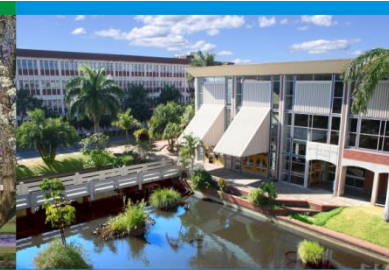
HOWARD COLLEGE CAMPUS



NELSON R MANDELA SCHOOL OF MEDICINE



PIETERMARITZBURG CAMPUS



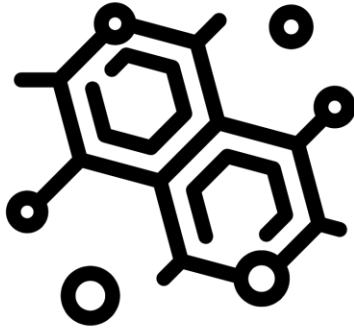
WESTVILLE CAMPUS

UKZN INSPIRING GREATNESS

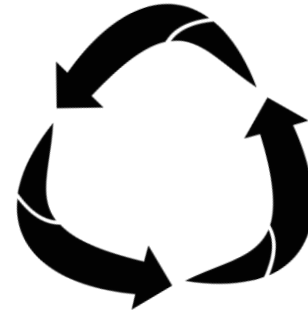
Outline

- **Why publish our data**
- **What is and why GISAID?**
- **Quality Control of sequences prior to submission**
 - Using Nextclade and geneious
 - Consensus v2,..., vFINAL
- **Quality control of metadata prior to submission**
 - Mandatory fields and data formats
- **Submission to the GISAID EpiCoV Database**
- **Interacting with GISAID EpiCoV Database**
 - Navigating the data
 - Downloading selections of the data
 - Downloading global aggregated data

Why is it important to publish our data?



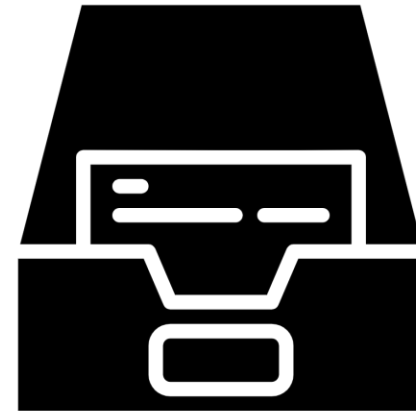
- Fast track science



- Reproducible research



- Metadata QC and Standardization

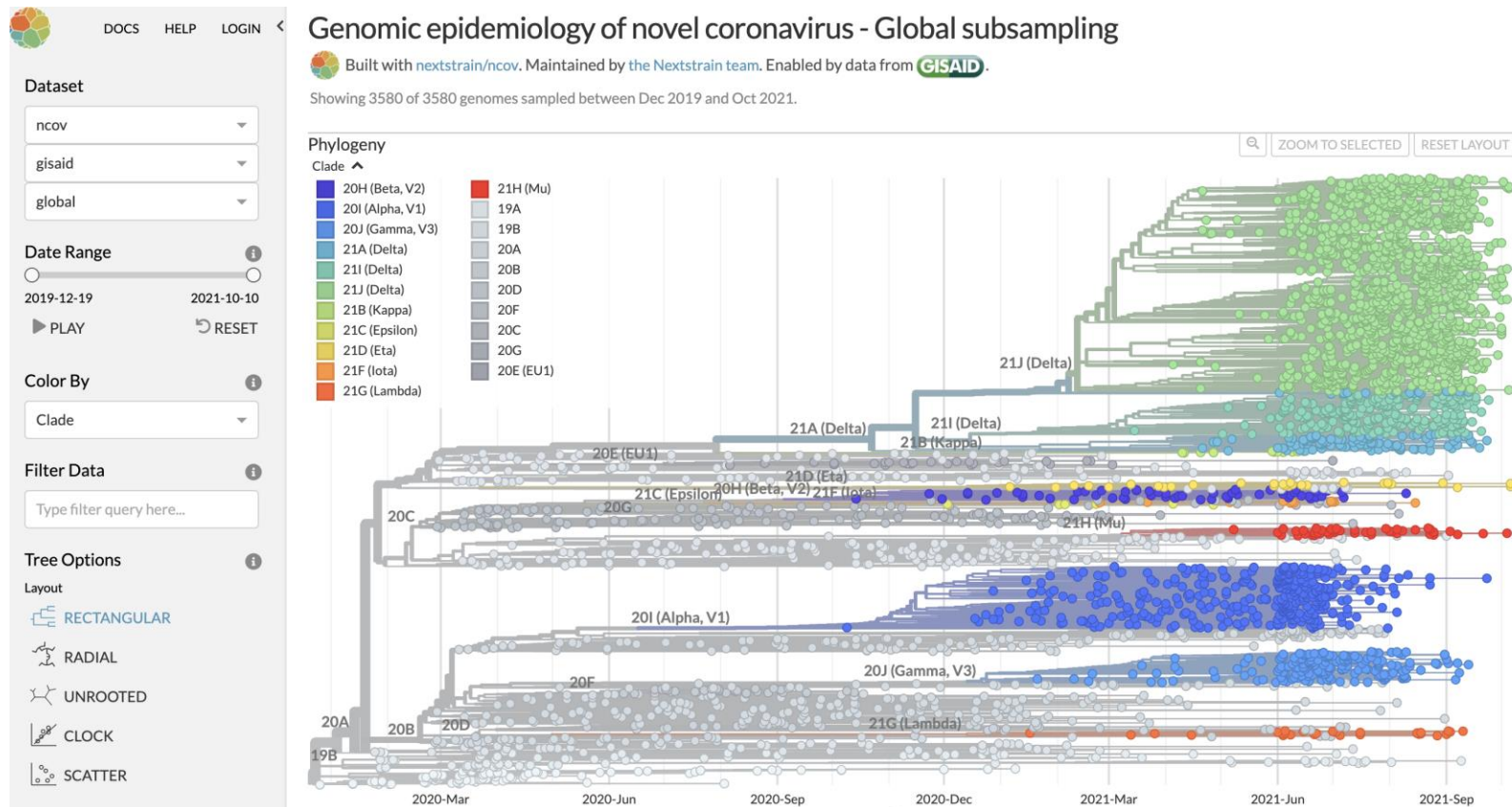


- Permanent storage

<https://www.gisaid.org/>

Why is it important to publish our data?

- ☐ Enables real-time genomic epidemiology to inform public health response



Regional phylogenies maintained by Nextstrain and updated regularly as new data is released
<https://nextstrain.org/ncov/gisaid/global>

What is GISAID?



About us Database Features Events Collaborations References Registration Help

Login

In Focus

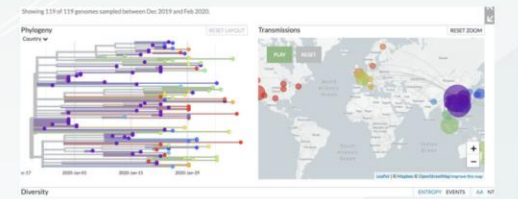
COVID-19 lineages and variants

GISAID's EpiCoV database employs tools to assign phylogenetic clades and lineages to genetic sequences of the pandemic coronavirus. One such tool is the Pango nomenclature by [Rambaut et al \(2020\)](#) which takes a granular approach to classify and describe viral evolution with detailed lineages.

As new lineages become more widespread, additional genetic markers emerge. Lineage definitions may be updated to allow researchers to track these separately and permit a more fine-grained picture of how a variant is circulating. When these updates occur, all genomes in EpiCoV undergo reclassification by Pango which can lead to temporary fluctuations in the tallies of variants. Overinterpretation of these changes in numbers should be avoided.



Genomic epidemiology of hCoV-19



hCoV-19 data sharing via GISAID

~~3,388,336~~
submissions **4,206,853**

Enabled by data shared via GISAID

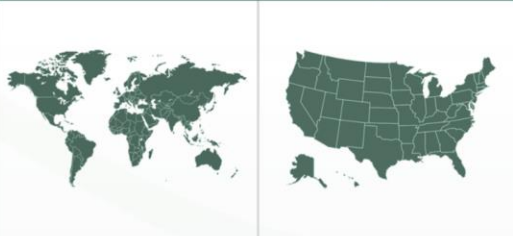
CoVglobe

The Francis Crick Institute (United Kingdom)
Aggregate three-week rolling averages of lineages in different locations

GISAID Resources

Free Access Credentials

hCoV-19 Submission Tracking



hCoV-19 Tracking of Variants

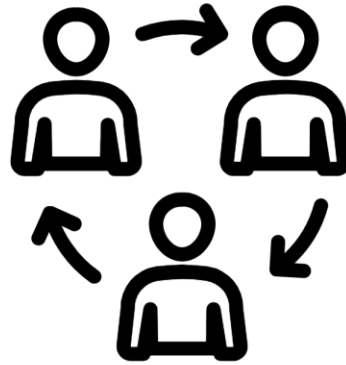


<https://www.gisaid.org/>

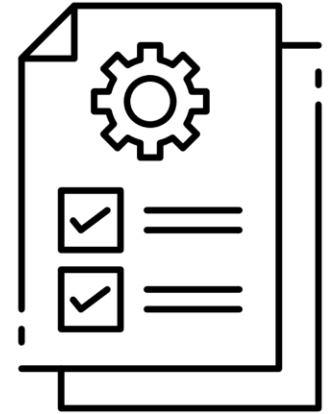
Why GISAID?



❖ Recognition



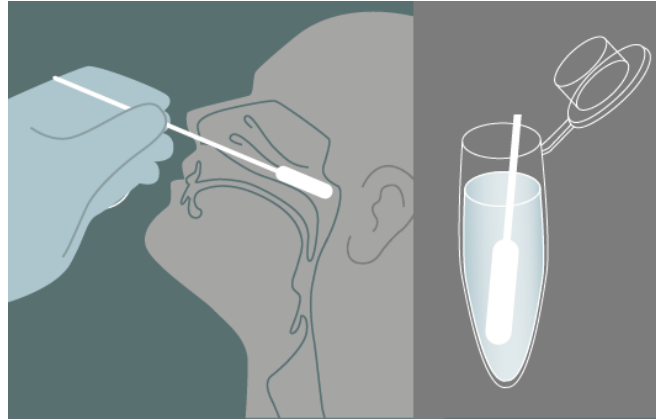
❖ Collaboration



❖ Quality Control

<https://www.gisaid.org/>

Data Production Workflow



sequencing

Genome
Assembly

Sequence /
Metadata QC

GISAID
Submission

Short reads
FastQ files

Assembled whole-
genomes (Fasta)

High quality
genomes

Publicly available
sequences

Quality Control (QC) - Sequences

- Does the sequence have sufficient genome coverage?
- Indels not a multiple of 3 (i.e. frameshifting)
- Mutations not causing premature stop codons
- Very large indels
- Clustered SNPs

Quality Control (QC)

Curation of sequence artifacts

Consensus v1 → Consensus v2/v3/v4... = **FINAL GENOME**

The logo for Geneious, featuring the word "geneious" in a lowercase, sans-serif font. The "gene" part is orange, and the "ious" part is dark grey. A stylized orange icon resembling a pair of pliers or a molecular structure is positioned above the "s".The logo for Nextclade, with the word "Nextclade" in a multi-colored, sans-serif font. A small grey box with the word "beta" is located to the right of the "e".The logo for Nextstrain, with the word "Nextstrain" in a multi-colored, sans-serif font.

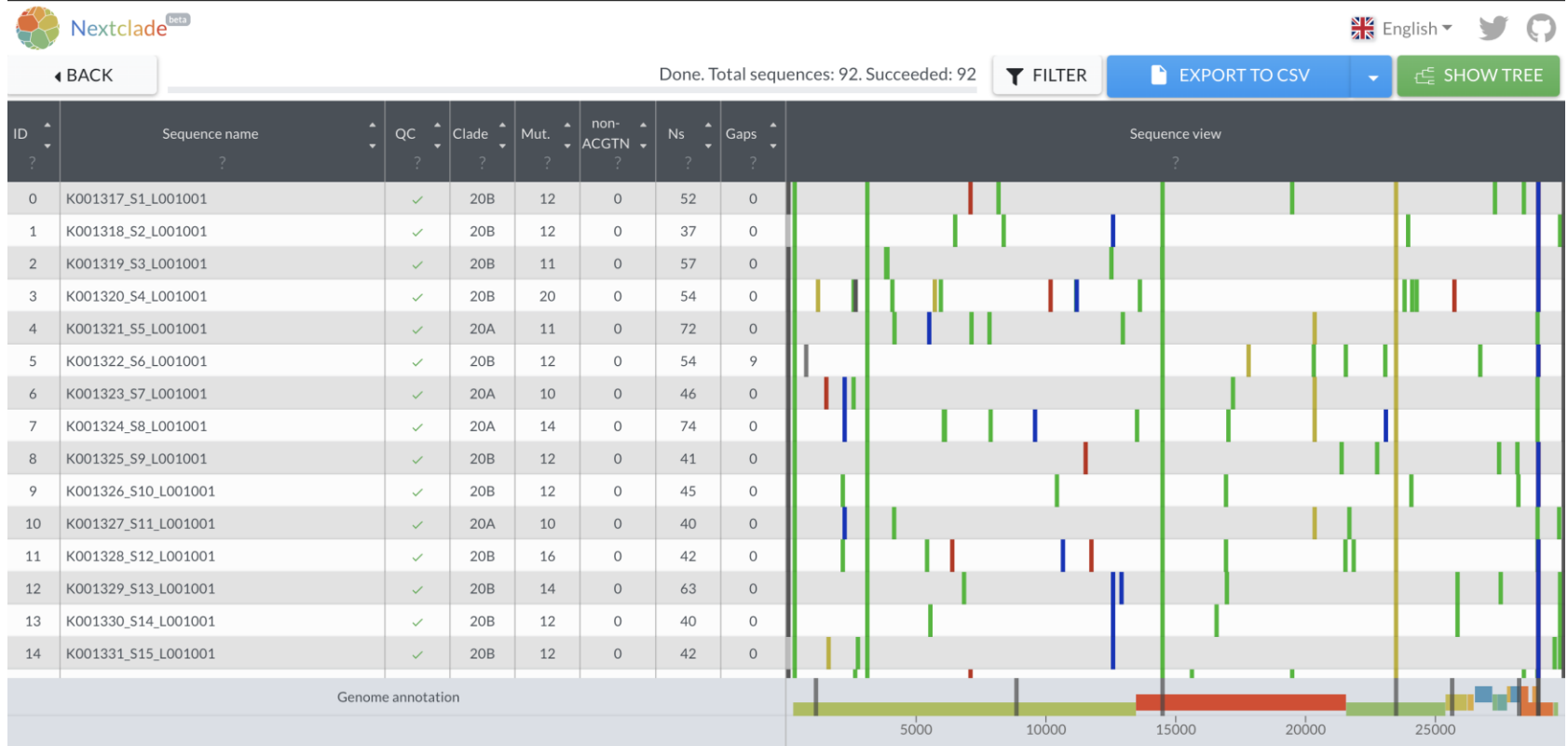
SeqKit - Ultrafast FASTA/Q kit

Identify mutations, open reading frames, indels, etc.

Sequence QC using Nextclade



Clade assignment, mutation calling,
and sequence quality checks



Sequencing QC using Nextclade



Clade assignment, mutation calling,
and sequence quality checks

- Calculating sequence coverage %
 - $=((29903 - \text{totalMissing}) / 29903) * 100$
- Investigating indels to be corrected (we remove all mutations and indels that are not a multiple of three that cause frameshifts and premature stop codons)

N	O
deletions	insertions
11288-11297,22287-22296,26579-26586,26590-26596,27722-27724	26603:G,26606:ATGGAA
11288-11297,22287-22296	
11288-11297,22287-22296	
28248-28254,28274	
14930	14906:G
10229-10232,13836,14693-14700,18704,24669-24671,29775	10233:GCC,14681:C,14703:G,14712:GTGGTG,18709:C,24665:CT,29771:A
11288-11297,22287-22296	
17458-17460,17465-17467,23288-23290,28254	17469:ACTC

Remove low coverage seqs.

seqName	clade	qc.overallSc	qc.overallSta	totalSubstitu	totalDeletio	totalInsertio	totalAminoa	totalAminoa	totalMissing	totalNonACG	totalPcrPrim	substitutions	deletions	coverage
K019120	19B	11656.00137	bad	2	0	0	1	0	29450	0	0	C10138T,A10323G		2
K019137	19A	11373.82750	bad	1	0	0	0	0	29095	0	0	C3037T		3
K019141	21A (Delta)	10965.19244	bad	1	0	0	0	0	28573	0	0	G22813T		4
K019145	21A (Delta)	10726.05444	bad	2	0	0	2	0	28263	0	0	C21618G,G22813T		5
K019119	19A	10312.02639	bad	1	0	0	0	0	27718	0	0	G5008C		7
K019121	19A	10330.08662	bad	14	0	0	1	7	27703	0	0	G10097A,A21456C,T21456C		7
K019116	21A (Delta)	10126.32235	bad	4	36	0	7	18	27470	0	0	T6607A,C17326785-26820C		8
K019122	19A	10127.48753	bad	0	1	0	0	0	27396	0	0	16873		8
K019148	19B	10173.33729	bad	2	0	0	2	0	27533	0	0	G22813T,A23063T		8
K019091	20C	9569.187160	bad	4	0	0	1	0	26712	0	0	C3037T,C14408T,T16176C		11
K019111	21B (Kappa)	9690.127901	bad	3	2	0	1	0	26568	0	0	A11201G,A118914,22813T		11
K019114	20G	9303.978056	bad	23	9	6	16	1	26238	0	1	T928C,G9321936,944-951G		12
K019112	19A	9027.146145	bad	5	5	0	3	0	25873	0	0	T18672C,C187433-7437T		13
K019102	20C	8379.029150	bad	5	0	0	4	0	25015	0	1	C6543T,G17320T,A23403C		16
K019159	20B	8512.865312	bad	19	16	13	15	0	24785	0	0	C5388A,T16116384,23137T		17
K019108	20I (Alpha, V	446.494475	bad	27	20	0	14	6	5604	0	2	C241T,C913T11288-11296G		81
K019163	20B	546.302500	bad	39	10	9	26	0	4836	0	4	C1059T,C30324457-24465G		84
K019100	20I (Alpha, V	290.680123	bad	28	20	0	14	6	4434	0	4	C241T,C913T11288-11296G		85
K019131	20I (Alpha, V	331.106547	bad	43	36	0	19	6	3994	0	4	C829T,C913T11288-11296G		87
K019167	20I (Alpha, V	140.791238	bad	26	21	0	15	3	2780	0	4	C913T,C30371118-7119,11288-11296G		91
K019410	20B	91.878398	mediocre	28	0	0	15	0	2573	0	4	C222T,C241T,C1190T,G1833T		91
K019099	20I (Alpha, V	67.534623	mediocre	31	20	0	18	6	1207	0	4	C241T,C913T11288-11296G		96
K019106	20I (Alpha, V	64.069273	mediocre	30	20	0	16	6	1055	0	4	C241T,C913T11288-11296G		96
K019156	20D	56.820864	mediocre	36	12	0	20	4	504	0	4	C241T,C843T21983-21994G		98
K019107	20I (Alpha, V	56.250000	mediocre	29	20	0	16	6	156	0	4	C241T,C913T11288-11296G		99
K019110	20I (Alpha, V	56.250000	mediocre	29	20	0	16	6	157	0	4	C241T,C913T11288-11296G		99
K019115	20I (Alpha, V	4.340278	good	38	19	0	23	6	281	0	4	G210T,C241T11288-11296G		99
K019117	20I (Alpha, V	56.250000	mediocre	29	20	0	16	6	157	0	4	C241T,C913T11288-11296G		99
K019128	20I (Alpha, V	21.006944	good	36	19	0	22	6	293	0	1	G210T,C241T11288-11296G		99
K019158	20I (Alpha, V	56.944444	mediocre	29	20	0	16	6	210	0	4	C241T,C913T11288-11296G		99
K019165	20I (Alpha, V	56.466629	mediocre	32	20	0	17	6	356	0	4	C241T,C913T11288-11296G		99

Love Coverage

High Coverage

```
$ seqkit grep -r -v -f bad_ids.txt consensus_seqs.fasta > good_seqs.fasta
```

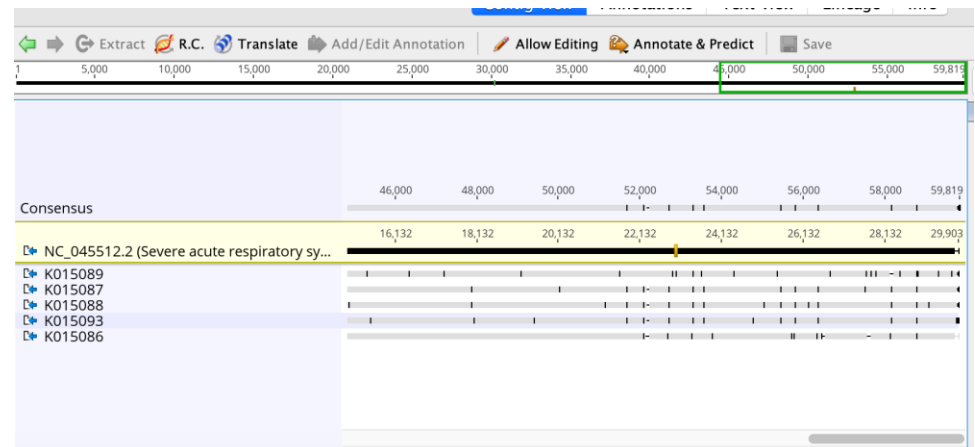
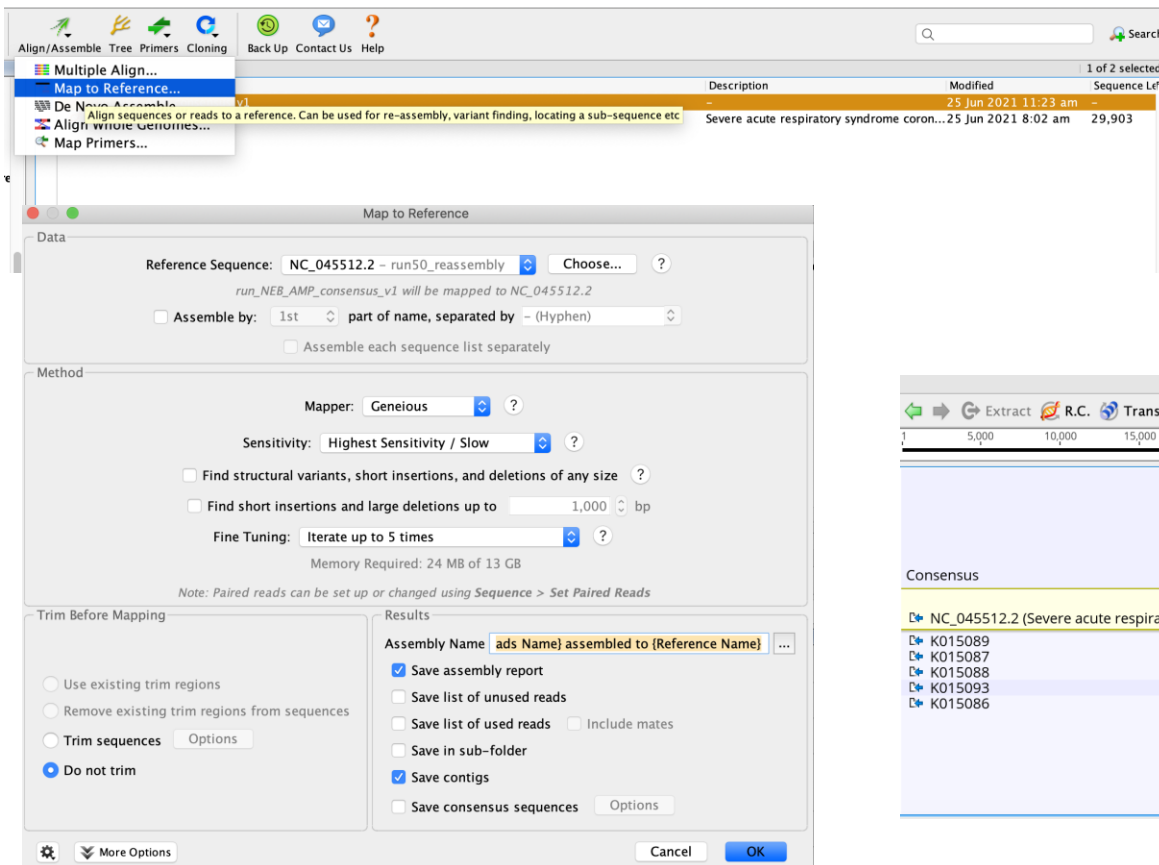
GENEIOUS – cleaning sequences



Geneious Prime - <https://www.geneious.com/download/>

- You need the SARS-CoV-2 reference file
- Map consensus_v1 to ref, manually edit spurious indels

- Save edits
- Export fasta file back and save as v2
- Run consensus v2 through nextclade again to identify any remaining spurious indels



Quality Control (QC) - Metadata



EpiCoV hCoV-19 bulk upload

Version: 2021-02-24

Column information		
Submitter	mandatory	enter your GISAID-Username
FASTA filename	mandatory	the filename that contains the sequence without path (e.g. all_sequences.fasta not c:\users\meier\docs\all_sequences.fasta)
Virus name	mandatory	e.g. hCoV-19/Netherlands/Gelderland-01/2020 (Must be FASTA-Header from the FASTA file all_sequences.fasta)
Type	mandatory	default must remain "betacoronavirus"
Passage details/history	mandatory	e.g. Original, Vero
Collection date	mandatory	Date in the format YYYY or YYYY-MM or YYYY-MM-DD
Location	mandatory	e.g. Europe / Germany / Bavaria / Munich
Additional location information		e.g. Cruise Ship, Convention, Live animal market
Host	mandatory	e.g. Human, Environment, Canine, Manis javanica, Rhinolophus affinis, etc
Additional host information		e.g. Patient infected while traveling in ...
Sampling Strategy		e.g. Sentinel surveillance (ILI), Sentinel surveillance (ARI), Sentinel surveillance (SARI), Non-sentinel-surveillance (hospital), Non-sentinel-surveillance (GP network), Longitudinal sampling on same patient(s), S gene dropout
Gender	mandatory	Male, Female, or <i>unknown</i>
Patient age	mandatory	e.g. 65 or 7 months, or <i>unknown</i>
Patient status	mandatory	e.g. Hospitalized, Released, Live, Deceased, or <i>unknown</i>
Specimen source		e.g. Sputum, Alveolar lavage fluid, Oro-pharyngeal swab, Blood, Tracheal swab, Urine, Stool, Cloakal swab, Organ, Feces, Other
Outbreak		Date, Location e.g. type of gathering, Family cluster, etc.
Last vaccinated		provide details if applicable
Treatment		Include drug name, dosage
Sequencing technology	mandatory	e.g. Illumina Miseq, Sanger, Nanopore MinION, Ion Torrent, etc.
Assembly method		e.g. CLC Genomics Workbench 12, Geneious 10.2.4, SPAdes/MEGAHIT v1.2.9, UGENE v. 33, etc.
Coverage		e.g. 70x, 1,000x, 10,000x (average)
Originating lab	mandatory	Where the clinical specimen or virus isolate was first obtained
Address	mandatory	
Sample ID given by the originating laboratory		
Submitting lab	mandatory	Where sequence data have been generated and submitted to GISAID
Address	mandatory	
Sample ID given by the submitting laboratory		
Authors	mandatory	a comma separated list of Authors with complete First followed by Last Name
Comment	leave empty	do not use this column
Comment Icon	leave empty	do not use this column

★ Critical field for genomic epidemiology

Missing data? ...*unknown*

Preparing GISAID Files

CAF_run312_consensus_gisaid_n6 — Saved to my Mac

Home Insert Draw Page Layout Formulas Data Review View Tell me

Calibri (Body) 11

General

Share Comments

	A	B	C	D	E	F	G	H	I	J	K	L
1	submitter	fn	covv_virus_name	covv_type	covv_pass	covv_collection_date	covv_location	covv_add_location	covv_host	covv_add_host_info	covv_sampling_strategy	covv_gen
2	submitter	fn	covv_virus_name	covv_type	covv_pass	covv_collection_date	covv_location	covv_add_location	covv_host	covv_add_host_info	covv_sampling_strategy	covv_gen
3	tuliodna	CAF_run312_consensus_gisaid_n6.fasta	hCoV-19/SouthAfrica/CERI-KRISP-Tyg1961/2021	betacoronavirus	Original	2021-07-29	Africa / SouthAfrica / Western Cape Province	Tygerberg	Human		Routine	Male
4	tuliodna	CAF_run312_consensus_gisaid_n6.fasta	hCoV-19/SouthAfrica/CERI-KRISP-Tyg1962/2021	betacoronavirus	Original	2021-07-29	Africa / SouthAfrica / Western Cape Province	Tygerberg	Human		Routine	Male
5	tuliodna	CAF_run312_consensus_gisaid_n6.fasta	hCoV-19/SouthAfrica/CERI-KRISP-Tyg1963/2021	betacoronavirus	Original	2021-07-29	Africa / SouthAfrica / Western Cape Province	Tygerberg	Human		Routine	Male
6	tuliodna	CAF_run312_consensus_gisaid_n6.fasta	hCoV-19/SouthAfrica/CERI-KRISP-Tyg1965/2021	betacoronavirus	Original	2021-07-29	Africa / SouthAfrica / Western Cape Province	Eastern	Human		Routine	Male
7	tuliodna	CAF_run312_consensus_gisaid_n6.fasta	hCoV-19/SouthAfrica/CERI-KRISP-Tyg1966/2021	betacoronavirus	Original	2021-07-29	Africa / SouthAfrica / Western Cape Province	Eastern	Human		Routine	Female
8	tuliodna	CAF_run312_consensus_gisaid_n6.fasta	hCoV-19/SouthAfrica/CERI-KRISP-Tyg1967/2021	betacoronavirus	Original	2021-07-29	Africa / SouthAfrica / Western Cape Province	Eastern	Human		Routine	Female

Cont...

	N	O	P	Q	R	S	T	U	V	W	X	Y	
1	ant_age	covv_patient_status	covv_specimen	covv_outbreak	covv_last_vaccinat	covv_treatment	covv_seq_technology	covv_assembly_method	covv_coverage	covv_orig_lab	covv_orig_lab_addr	covv_provider_sample_id	covv_subm_lab
2	ant_age	covv_patient_status	covv_specimen	covv_outbreak	covv_last_vaccinat	covv_treatment	covv_seq_technology	covv_assembly_method	covv_coverage	covv_orig_lab	covv_orig_lab_addr	covv_provider_sample_id	covv_subm_lab
3	46		SWAB_NASO_OROP				Ion Torrent	Genome Detective 1.133		Division of Medical Virology, N Division of Medical Virology, Depart Tyg1961		KRISP, KZn Research Innov	
4	27		SWAB_NASO_OROP				Ion Torrent	Genome Detective 1.133		Division of Medical Virology, N Division of Medical Virology, Depart Tyg1962		KRISP, KZn Research Innov	
5	33		SWAB_NASO_OROP				Ion Torrent	Genome Detective 1.133		Division of Medical Virology, N Division of Medical Virology, Depart Tyg1963		KRISP, KZn Research Innov	
6	54		SWAB_NASO_OROP				Ion Torrent	Genome Detective 1.133		Division of Medical Virology, N Division of Medical Virology, Depart Tyg1965		KRISP, KZn Research Innov	
7	57		SWAB_NASO_OROP				Ion Torrent	Genome Detective 1.133		Division of Medical Virology, N Division of Medical Virology, Depart Tyg1966		KRISP, KZn Research Innov	
8	36		SWAB_NASO_OROP				Ion Torrent	Genome Detective 1.133		Division of Medical Virology, N Division of Medical Virology, Depart Tyg1967		KRISP, KZn Research Innov	

Modify Sequence Names

Create a text file...*gisaid_ids.txt*

 **gisaid_ids.txt**

🔧 ~/temp/Collaborations/covid/GISAID/runs167_168/gisaid_ids.txt ↕

```
10 K020867,hCoV-19/Malawi/KRISP-K020867/2021
11 K020868,hCoV-19/Malawi/KRISP-K020868/2021
12 K020869,hCoV-19/Malawi/KRISP-K020869/2021
13 K020870,hCoV-19/Malawi/KRISP-K020870/2021
14 K020873,hCoV-19/Malawi/KRISP-K020873/2021
15 K020874,hCoV-19/Malawi/KRISP-K020874/2021
16 K020875,hCoV-19/Malawi/KRISP-K020875/2021
17 K020876,hCoV-19/Malawi/KRISP-K020876/2021
18 K020879,hCoV-19/Malawi/KRISP-K020879/2021
19 K020881,hCoV-19/Malawi/KRISP-K020881/2021
20 K020882,hCoV-19/Malawi/KRISP-K020882/2021
21 K020883,hCoV-19/Malawi/KRISP-K020883/2021
22 K020884,hCoV-19/Malawi/KRISP-K020884/2021
23 K020886,hCoV-19/Malawi/KRISP-K020886/2021
24 K020887,hCoV-19/Malawi/KRISP-K020887/2021
25 K020895,hCoV-19/Malawi/KRISP-K020895/2021
26 K020906,hCoV-19/Malawi/KRISP-K020906/2021
27 K020908,hCoV-19/Malawi/KRISP-K020908/2021
28 K020909,hCoV-19/Malawi/KRISP-K020909/2021
29 K020910,hCoV-19/Malawi/KRISP-K020910/2021
30 K020911,hCoV-19/Malawi/KRISP-K020911/2021
31 K020913,hCoV-19/Malawi/KRISP-K020913/2021
```

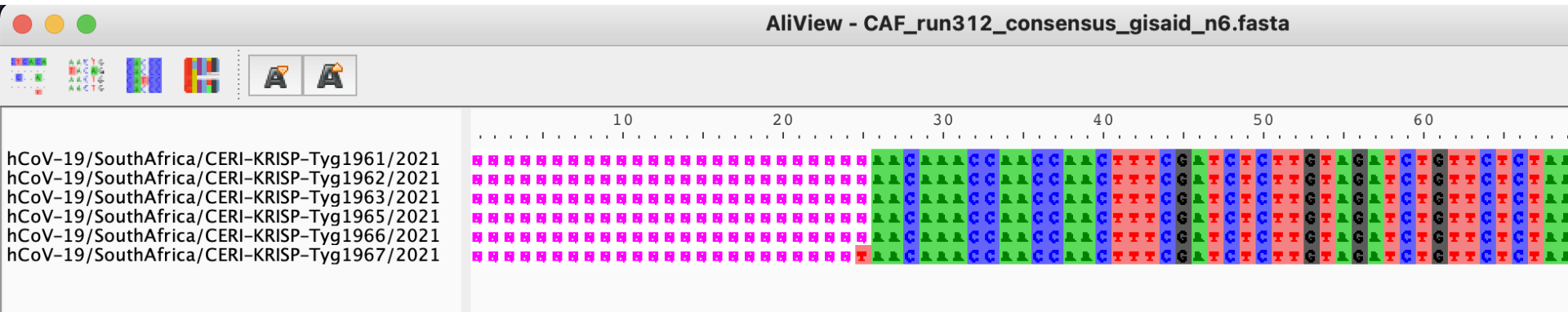
One liner to batch update

```
$ for s in `cat gisaid_ids.txt`; do
on=$(echo $s | cut -d "," -f1); nn=$(echo
    $s | cut -d "," -f2); sed -i "
        "s#${on}#${nn}#g"
    run666_gisaid_consensus_n721.fasta;
done
```

Final GISAID Submission Files

Files required:

1. Final sequence fasta file with sequences appropriately named



2. GISAID submission file (Excel)

	A	B	C	D	E	F	G	H	I	J	K	L
1	submitter	fn	covv_virus_name	covv_type	covv_pass	covv_collection_date	covv_location	covv_add_location	covv_host	covv_add_host_info	covv_sampling_strategy	covv_gender
2	submitter	fn	covv_virus_name	covv_type	covv_pass	covv_collection_date	covv_location	covv_add_location	covv_host	covv_add_host_info	covv_sampling_strategy	covv_gender
3	tuliodna	CAF_run312_consensus_gisaid_n6.fasta	hCoV-19/SouthAfrica/CERI-KRISP-Tyg1961/2021	betacoronavirus	Original	2021-07-29	Africa / SouthAfrica / Western Cape Province	Tygerberg	Human		Routine	Male
4	tuliodna	CAF_run312_consensus_gisaid_n6.fasta	hCoV-19/SouthAfrica/CERI-KRISP-Tyg1962/2021	betacoronavirus	Original	2021-07-29	Africa / SouthAfrica / Western Cape Province	Tygerberg	Human		Routine	Male
5	tuliodna	CAF_run312_consensus_gisaid_n6.fasta	hCoV-19/SouthAfrica/CERI-KRISP-Tyg1963/2021	betacoronavirus	Original	2021-07-29	Africa / SouthAfrica / Western Cape Province	Tygerberg	Human		Routine	Male
6	tuliodna	CAF_run312_consensus_gisaid_n6.fasta	hCoV-19/SouthAfrica/CERI-KRISP-Tyg1965/2021	betacoronavirus	Original	2021-07-29	Africa / SouthAfrica / Western Cape Province	Eastern	Human		Routine	Male
7	tuliodna	CAF_run312_consensus_gisaid_n6.fasta	hCoV-19/SouthAfrica/CERI-KRISP-Tyg1966/2021	betacoronavirus	Original	2021-07-29	Africa / SouthAfrica / Western Cape Province	Eastern	Human		Routine	Female
8	tuliodna	CAF_run312_consensus_gisaid_n6.fasta	hCoV-19/SouthAfrica/CERI-KRISP-Tyg1967/2021	betacoronavirus	Original	2021-07-29	Africa / SouthAfrica / Western Cape Province	Eastern	Human		Routine	Female
9												

GISAID Upload



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

Registration

Platform Login

Username

Password

[Forgot your password or username?](#)



hosted by the
Federal Republic
of Germany

developed by



Ensure that you have a valid account

GISAID Upload

epicov.org/epi3/frontend#958d

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

You are logged in as **James Emmanuel San** - [logout](#)

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

EpiCoV™ | Search | Downloads | Upload

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 SARS coronavirus (SARS-CoV).

On 10. January 2020, the World Health Organization (WHO) declared the outbreak a pandemic. The WHO is currently sequencing the genomes of the virus to identify potential drug targets, phylogenetic relationships, and other characteristics.

by A*STAR Singapore

Single upload

Batch upload

Audacity | AudacityInstant | BLAST | CoVizu^e | Emerging Variants | Official GISAID reference sequence

PrimerChecker | Spike glycoprotein

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 Upload

2. Click on “Download Instructions and Template” for details on the bulk upload submission template and instructions.
3. Complete the Metadata information in the XLS/CSV file and ensure all entries have the associated FASTA sequences in the FASTA file – Upload your “XLS/CSV” file and “FASTA” file using Browse.

Registered Users EpiFlu™ EpiCoV™ My profile Administration

EpiCoV™ Browse Downloads Upload Batch Upload My Unreleased Curation

GISAID hCoV-19 Batch Upload

Upload genetic sequence as single FASTA-File and metadata, available clinical and epidemiological data, geographical as well as species-specific data as XLS or CSV. Data will be reviewed by a curator prior to release. An email confirmation will be issued upon release.

Metadata as Excel or CSV*

max size: 5M Browse... No file selected.

Sequences as FASTA*

max size: 32M Browse... No file selected.

Report

Upload XLS/CSV and FASTA.

Download Instructions and Template Contact Curator Check and Submit

2. Download instructions

3. Upload your data



GISAID Upload

4. Click on “Check and Submit” to verify your submission and submit to the EpiCoV Curation Team. If virus name in both XLS/CSV and FASTA file matches and the required fields are filled out in a correct format, a confirmation message will appear.
5. If errors appear at the *Report* section (as shown below), check next slide for resolving the issues. If errors are persistent, please use the option **Contact curator** option, mention the errors and we will help you resolve them (located beside “check and Submit”).

The screenshot displays the GISAID hCoV-19 Batch Upload interface. At the top, there are navigation tabs for 'Registered Users', 'EpiFlu™', 'EpiCoV™', 'My profile', and 'Administration'. Below these are icons for 'EpiCoV™', 'Browse', 'Downloads', 'Upload', 'Batch Upload', 'My Unreleased', and 'Curation'. The main heading is 'GISAID hCoV-19 Batch Upload'. A red notice states: 'Upload genetic sequence as single FASTA-File and metadata, available clinical and epidemiological data, geographical as well as species-specific data as XLS or CSV. Data will be reviewed by a curator prior to release. An email confirmation will be issued upon release.'

The interface is divided into three main sections:

- Metadata as Excel or CSV*:** Shows a file named 'Test.csv (35.1 kbytes)' with a 'max size: 5M' limit and a 'Browse...' button. A 'No file selected' message is visible.
- Sequences as FASTA*:** Shows a file named 'Test.fasta (1.20 MB)' with a 'max size: 32M' limit and a 'Browse...' button. A 'No file selected' message is visible.
- Report:** Contains the following text: 'Excel-Sheet contains 44 submissions', 'Error: Sequence Virus name not found in FASTA', and 'Error: Sequence Virus name not found in FASTA'.

At the bottom, there are three buttons: 'Download Instructions and Template', 'Contact Curator', and 'Check and Submit'. The 'Check and Submit' button is circled in red. A 'Message' dialog box is overlaid on the interface, stating: 'Errors in Batch-Upload-Data! Please read report.' with an 'OK' button.

5. Error Report

4. Check and Submit



GISAID Upload

Common Error Messages during batch upload

Error Message	Info	Solution
Submission XXX is not unique in FASTA	A duplicate dataset of one and the same virus has been detected, i.e. a copy of the same sequence.	Remove one of the sequences from the FASTA file.
Sequence XXX not found in FASTA	The Sequence XXX is missing in the submitted FASTA-File, e.g. a typo the FASTA-header.	Add the missing Sequence with correct FASTA-Header to the FASTA-File
Please use correct Excel-Template	You did not submit an Excel-Sheet that is based on the GISAID hCoV Batch Upload Template.	Use the GISAID hCoV Batch Upload Template.
Column XXX is missing	You removed a column from the template or modified the first line of the Excel-Sheet.	Use the GISAID hCoV Batch Upload Template without modification.
Mandatory value YYY for XXX is missing	You forgot to fill in values in this mandatory column.	Fill in all mandatory columns marked in red. Use “unkown” if no value is available.
Submission XXX already exists	The virus-name you are trying to submit already exists in the system – released or in curation. Maybe you try to resubmit an already submitted Batch-Upload.	Use a new virus-name.
Unknown Submitter YYY for XXX	The username in the Excel-Sheet is not known to the system.	Ensure the Submitter column matches the GISAID Username
Submitter YYY not allowed for XXX	As a submitter you can only submit viruses for yourself.	Use your username as submitter.
Virus name XXX format	The Virus-Name in the Excel-Sheet has a wrong format.	Please respect the Format of the Virus-Name in the Excel-Sheet.
Virus name XXX contains invalid letters	The Virus-Name must not contain Non-ASCII Characters like Accents and Umlauts.	Replace all Accents, Umlauts and other special Characters with Standart letters.
Comment / Comment type not allowed	You must leave the Column Comment and Comment Type empty. They are for curators only.	Leave the Column Comment and Comment Type empty.
Collection date XXX format	The Date-Format of the Submission is wrong.	Accepted Date Formats are: YYYY-MM-DD or YYYY-MM or YYYY. Format the Collection date as “Text” not as “Date”.
Only type ‘betacoronavirus’ allowed	The value of the Type-Column must be “betacoronavirus”	Fill in “betacoronavirus” as Type

Interacting with the GISAID database

GISAID offers a number of services in addition to the database

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

You are logged in as Hourliyah Tegally - [logout](#)

Registered Users EpiFlu™ EpiCoV™ My profile

EpiCoV™ Search Downloads Upload

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](#)

Audacity AudacityInstant BLAST CoVizu® Emerging Variants Official GISAID reference sequence PrimerChecker Spike glycoprotein mutation surveillance

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

You are logged in as Hourliyah Tegally - [logout](#)

Registered Users EpiFlu™ EpiCoV™ My profile

EpiCoV™ Search Downloads Upload

PrimerChecker Spike glycoprotein mutation surveillance

Analysis Update (2021-10-08)

- Full genome tree derived from all outbreak sequences
- Timecourse of clade distribution in collected sequences
- Timecourse of variant distribution in all submitted sequences
- Timecourse of Delta variant sublineage distribution
- Percentage of Delta variant sublineages in collected sequences
- Regional clade distribution of new sequences
- Regional distribution of variants in new sequences
- Regional distribution of variants in collected sequences
- Distribution of collection dates of new sequences
- Breakdown of new sequences by clade, then by territory (page 1)
- Breakdown of new sequences by clade, then by territory (page 2)
- Receptor binding surveillance for complete genomes (page 1)
- Receptor binding surveillance for complete genomes (page 2)
- Receptor binding surveillance for complete genomes (page 3)
- Receptor binding surveillance for complete genomes (page 4)
- Common primer check for high quality genomes
- Full genome tree of hCoV-19-related precursors
- [analysis update.pdf](#)

Interacting with the GISAID database

Navigating the data

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

You are logged in as **Houriyyah Tegally** - [logout](#)

Register EpiFlu™ EpiCoV™ My profile

EpiCoV™ Search Downloads Upload

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

<input type="checkbox"/>	Virus name	Passage de	Accession ID	Collection da	Submission C	<input type="checkbox"/>	Length	Host	Location	Originating
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812279/2021	Original	EPI_ISL_5054789	2021-08-23	2021-10-12	<input type="checkbox"/>	29,765	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812268/2021	Original	EPI_ISL_5054788	2021-08-24	2021-10-12	<input type="checkbox"/>	29,598	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812004/2021	Original	EPI_ISL_5054787	2021-08-23	2021-10-12	<input type="checkbox"/>	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812281/2021	Original	EPI_ISL_5054786	2021-08-23	2021-10-12	<input type="checkbox"/>	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812277/2021	Original	EPI_ISL_5054785	2021-08-23	2021-10-12	<input type="checkbox"/>	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812301/2021	Original	EPI_ISL_5054784	2021-08-23	2021-10-12	<input type="checkbox"/>	29,598	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1811999/2021	Original	EPI_ISL_5054783	2021-08-23	2021-10-12	<input type="checkbox"/>	29,598	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812027/2021	Original	EPI_ISL_5054782	2021-08-23	2021-10-12	<input type="checkbox"/>	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IL-SHL-1812284/2021	Original	EPI_ISL_5054781	2021-08-20	2021-10-12	<input type="checkbox"/>	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812007/2021	Original	EPI_ISL_5054780	2021-08-23	2021-10-12	<input type="checkbox"/>	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812252/2021	Original	EPI_ISL_5054779	2021-08-23	2021-10-12	<input type="checkbox"/>	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812237/2021	Original	EPI_ISL_5054778	2021-08-24	2021-10-12	<input type="checkbox"/>	29,598	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812218/2021	Original	EPI_ISL_5054777	2021-08-23	2021-10-12	<input type="checkbox"/>	29,769	Human	North America / U	State Hygie

Total: 4,248,757 viruses

<< < 1 2 3 4 5 > >>

Select

Interacting with the GISAID database

Navigating the data

Example Query Results: South Africa sequences 2,026 Delta genomes from patients testing positive in August

The screenshot shows the GISAID database search interface. The search criteria are: Location: South Africa; Collection: 2021-08-01 to 2021-08-31; Variants: VOC Delta GK/478K.V1. The results table shows 13 rows of data, with a total of 2,026 viruses. The table columns include Virus name, Passage date, Accession ID, Collection date, Submission date, Length, Host, Location, and Originating institution.

	Virus name	Passage date	Accession ID	Collection date	Submission date	Length	Host	Location	Originating
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955117	2021-08-30	2021-10-07	29,769	Human	Africa / South Afr	NHLS Unive
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955110	2021-08-31	2021-10-07	29,166	Human	Africa / South Afr	NHLS Unive
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955098	2021-08-31	2021-10-07	28,858	Human	Africa / South Afr	NHLS Unive
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955080	2021-08-30	2021-10-07	29,700	Human	Africa / South Afr	NHLS Unive
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955074	2021-08-30	2021-10-07	29,488	Human	Africa / South Afr	NHLS Unive
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955066	2021-08-30	2021-10-07	28,865	Human	Africa / South Afr	NHLS Unive
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955060	2021-08-30	2021-10-07	29,481	Human	Africa / South Afr	NHLS Unive
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955056	2021-08-31	2021-10-07	29,769	Human	Africa / South Afr	NHLS Unive
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955041	2021-08-30	2021-10-07	29,769	Human	Africa / South Afr	NHLS Unive
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955036	2021-08-30	2021-10-07	29,481	Human	Africa / South Afr	NHLS Unive
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955031	2021-08-30	2021-10-07	29,481	Human	Africa / South Afr	NHLS Unive
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955013	2021-08-31	2021-10-07	29,768	Human	Africa / South Afr	NHLS Unive
<input type="checkbox"/>	hCoV-19/South Africa/UFS-VIRO-NGS-MSK	Original	EPI_ISL_4955003	2021-08-30	2021-10-07	29,487	Human	Africa / South Afr	NHLS Unive

Total: 2,026 viruses

Interacting with the GISAID database

Navigating the data

Example Query Results: South Africa sequences 2,026 Delta genomes from patients testing positive in August

The KRISP node of the NGS-SA sequenced 537 of those

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

You are logged in as **Houriyyah Tegally** - [logout](#)

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

EpiCoV™ Search Downloads Upload

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

<input type="checkbox"/>	Virus name	Passage date	Accession ID	Collection date	Submission date	<input type="checkbox"/>	Length	Host	Location	Originating
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022470/202	Original	EPI_ISL_4575203	2021-08-14	2021-09-29	<input type="checkbox"/>	29,853	Human	Africa / South Afr	ZARV/INHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022485/202	Original	EPI_ISL_4575200	2021-08-12	2021-09-29	<input type="checkbox"/>	29,853	Human	Africa / South Afr	ZARV/INHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022436/202	Original	EPI_ISL_4575191	2021-08-07	2021-09-29	<input type="checkbox"/>	29,852	Human	Africa / South Afr	ZARV/INHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022462/202	Original	EPI_ISL_4575190	2021-08-02	2021-09-29	<input type="checkbox"/>	29,852	Human	Africa / South Afr	ZARV/INHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K023564/202	Original	EPI_ISL_4575189	2021-08-21	2021-09-29	<input type="checkbox"/>	29,852	Human	Africa / South Afr	ZARV/INHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K023567/202	Original	EPI_ISL_4575185	2021-08-20	2021-09-29	<input type="checkbox"/>	29,792	Human	Africa / South Afr	ZARV/INHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022453/202	Original	EPI_ISL_4575184	2021-08-04	2021-09-29	<input type="checkbox"/>	29,853	Human	Africa / South Afr	ZARV/INHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022432/202	Original	EPI_ISL_4575183	2021-08-10	2021-09-29	<input type="checkbox"/>	29,853	Human	Africa / South Afr	ZARV/INHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K023561/202	Original	EPI_ISL_4575171	2021-08-21	2021-09-29	<input type="checkbox"/>	29,801	Human	Africa / South Afr	ZARV/INHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022434/202	Original	EPI_ISL_4575170	2021-08-09	2021-09-29	<input type="checkbox"/>	29,853	Human	Africa / South Afr	ZARV/INHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022442/202	Original	EPI_ISL_4575169	2021-08-06	2021-09-29	<input type="checkbox"/>	29,852	Human	Africa / South Afr	ZARV/INHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022435/202	Original	EPI_ISL_4575167	2021-08-07	2021-09-29	<input type="checkbox"/>	29,853	Human	Africa / South Afr	ZARV/INHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022473/202	Original	EPI_ISL_4575166	2021-08-13	2021-09-29	<input type="checkbox"/>	29,853	Human	Africa / South Afr	ZARV/INHL

Total: 537 viruses

<< < 1 2 3 4 5 > >>

Select Analysis Download

Interacting with the GISAID database

Downloading selections of the data

The screenshot displays the GISAID EpiCoV™ search interface. The search criteria are: Virus name: KRISP, Location: South Africa, Collection: 2021-08-01 to 2021-08-31, Clade: all. A 'Download' modal window is open, showing the following options:

- Sequences (FASTA)
- Patient status metadata
- Sequencing technology metadata
- Dates and Location
- Input for the Augur pipeline
- Acknowledgement (Supplemental table)

The background shows a table of search results with columns for Virus name, Accession ID, Date, and Location. The first few rows are:

Virus name	Accession ID	Date	Location
hCoV-19/South Africa/KRISP-K022	hCoV-19/South Africa/KRISP-K022	2021-08-01	South Africa
hCoV-19/South Africa/KRISP-K022	hCoV-19/South Africa/KRISP-K022	2021-08-01	South Africa
hCoV-19/South Africa/KRISP-K022	hCoV-19/South Africa/KRISP-K022	2021-08-01	South Africa
hCoV-19/South Africa/KRISP-K022	hCoV-19/South Africa/KRISP-K022	2021-08-01	South Africa

Input for Augur pipeline options:

- Curated metadata containing dates and location of collection of specimen, lineage assignments, originating lab, authors, basic host information and GISAID accession IDs
- FASTA file

Interacting with the GISAID database

Downloading selections of the data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T		
1	strain	virus	gisaid_epi_is	genbank_acc	date	region	country	division	location	region_expo	country_e	division_exp	segment	length	host	age	sex	Nextstrain_c	pangolin_lin	GISAID		
2	hCoV-19/Sol. betacoronavi EPI_ISL_372?				03/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29853	Human		37	Female	?	AY.4	GK	
3	hCoV-19/Sol. betacoronavi EPI_ISL_372?				03/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29859	Human		41	Male	?	AY.33	GK	
4	hCoV-19/Sol. betacoronavi EPI_ISL_372?				03/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29859	Human		22	Female	?	AY.19	GK	
5	hCoV-19/Sol. betacoronavi EPI_ISL_372?				03/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29848	Human		34	Male	?	B.1.617.2	GK	
6	hCoV-19/Sol. betacoronavi EPI_ISL_372?				03/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29853	Human		49	Female	?	AY.4	GK	
7	hCoV-19/Sol. betacoronavi EPI_ISL_372?				03/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29853	Human		58	Male	?	AY.4	GK	
8	hCoV-19/Sol. betacoronavi EPI_ISL_379?				06/08/2021	Africa	South Africa	Gauteng	City of Johan Africa	South Africa	Gauteng	genome		29860	Human		42	Male	?	AY.4	GK	
9	hCoV-19/Sol. betacoronavi EPI_ISL_379?				02/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29859	Human		97	Female	?	AY.4	GK	
10	hCoV-19/Sol. betacoronavi EPI_ISL_379?				03/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29859	Human		26	Male	?	AY.4	GK	
11	hCoV-19/Sol. betacoronavi EPI_ISL_379?				02/08/2021	Africa	South Africa	Gauteng	City of Johan Africa	South Africa	Gauteng	genome		29853	Human		71	Female	?	AY.5	GK	
12	hCoV-19/Sol. betacoronavi EPI_ISL_379?				02/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29859	Human		29	Male	?	AY.4	GK	
13	hCoV-19/Sol. betacoronavi EPI_ISL_379?				02/08/2021	Africa	South Africa	Gauteng	City of Johan Africa	South Africa	Gauteng	genome		29859	Human	unknown		71	Female	?	AY.4	GK
14	hCoV-19/Sol. betacoronavi EPI_ISL_379?				07/08/2021	Africa	South Africa	Gauteng	City of Johan Africa	South Africa	Gauteng	genome		29852	Human		9	Male	?	AY.4	GK	
15	hCoV-19/Sol. betacoronavi EPI_ISL_379?				06/08/2021	Africa	South Africa	Gauteng	City of Johan Africa	South Africa	Gauteng	genome		29853	Human		45	Female	?	AY.4	GK	
16	hCoV-19/Sol. betacoronavi EPI_ISL_379?				02/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29853	Human		46	Female	?	AY.4	GK	
17	hCoV-19/Sol. betacoronavi EPI_ISL_379?				02/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29874	Human		48	Male	?	AY.4	GK	
18	hCoV-19/Sol. betacoronavi EPI_ISL_379?				05/08/2021	Africa	South Africa	Gauteng	City of Johan Africa	South Africa	Gauteng	genome		29859	Human		32	Female	?	AY.4	GK	
19	hCoV-19/Sol. betacoronavi EPI_ISL_379?				08/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29850	Human		34	Female	?	AY.4	GK	
20	hCoV-19/Sol. betacoronavi EPI_ISL_379?				02/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29874	Human		44	Female	?	AY.4	GK	
21	hCoV-19/Sol. betacoronavi EPI_ISL_379?				06/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29853	Human		62	Male	?	AY.4	GK	
22	hCoV-19/Sol. betacoronavi EPI_ISL_379?				04/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29844	Human		51	Female	?	AY.4	GK	
23	hCoV-19/Sol. betacoronavi EPI_ISL_379?				07/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29853	Human		26	Male	?	B.1.617.2	GK	
24	hCoV-19/Sol. betacoronavi EPI_ISL_379?				06/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29818	Human		30	unknown	?	AY.4	GK	
25	hCoV-19/Sol. betacoronavi EPI_ISL_379?				02/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29859	Human		97	Female	?	AY.4	GK	
26	hCoV-19/Sol. betacoronavi EPI_ISL_379?				07/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29852	Human		50	Female	?	AY.4	GK	
27	hCoV-19/Sol. betacoronavi EPI_ISL_379?				02/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29758	Human		15	Female	?	AY.38	GK	
28	hCoV-19/Sol. betacoronavi EPI_ISL_379?				06/08/2021	Africa	South Africa	Gauteng	City of Johan Africa	South Africa	Gauteng	genome		29852	Human		29	Female	?	AY.4	GK	
29	hCoV-19/Sol. betacoronavi EPI_ISL_379?				02/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29853	Human		27	Female	?	AY.4	GK	
30	hCoV-19/Sol. betacoronavi EPI_ISL_379?				01/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29796	Human		44	Female	?	AY.38	GK	
31	hCoV-19/Sol. betacoronavi EPI_ISL_379?				10/08/2021	Africa	South Africa	KwaZulu-Nat eThekwi	Africa	South Africa	KwaZulu-Nat	genome		29854	Human		57	Male	?	AY.4	GK	

Input for Augur pipeline options:

- Curated metadata containing dates and location of collection of specimen, lineage assignments, originating lab, authors, basic host information and GISAID accession IDs
- FASTA file

Interacting with the GISAID database

Downloading selections of the data

The screenshot displays the GISAID database search interface. The search criteria are set to Virus name: KRISP, Location: South Africa, and Collection: 2021-08-01 to 2021-08-31. A 'Download' dialog box is open, showing the following options:

- Sequences (FASTA)
- Patient status metadata
- Sequencing technology metadata
- Dates and Location
- Input for the Augur pipeline
- Acknowledgement (Supplemental table)

The background table shows the following data:

Accession ID	Virus name	Host	Location	Originating				
hCoV-19/South Africa/KRISP-K022	hCoV-19/South Africa/KRISP-K022	Human	Africa / South Afr	ZARV/NHL!				
hCoV-19/South Africa/KRISP-K022	hCoV-19/South Africa/KRISP-K022	Human	Africa / South Afr	ZARV/NHL!				
hCoV-19/South Africa/KRISP-K022	hCoV-19/South Africa/KRISP-K022	Human	Africa / South Afr	ZARV/NHL!				
hCoV-19/South Africa/KRISP-K022	hCoV-19/South Africa/KRISP-K022	Human	Africa / South Afr	ZARV/NHL!				
hCoV-19/South Africa/KRISP-K023	hCoV-19/South Africa/KRISP-K023	Human	Africa / South Afr	ZARV/NHL!				
hCoV-19/South Africa/KRISP-K023	hCoV-19/South Africa/KRISP-K023	Human	Africa / South Afr	ZARV/NHL!				
hCoV-19/South Africa/KRISP-K022	hCoV-19/South Africa/KRISP-K022	Human	Africa / South Afr	ZARV/NHL!				
hCoV-19/South Africa/KRISP-K022	hCoV-19/South Africa/KRISP-K022	Human	Africa / South Afr	ZARV/NHL!				
hCoV-19/South Africa/KRISP-K022	hCoV-19/South Africa/KRISP-K022	Human	Africa / South Afr	ZARV/NHL!				
hCoV-19/South Africa/KRISP-K023561/202	Original	EPI_ISL_4575171	2021-08-21	2021-09-29	29,801	Human	Africa / South Afr	ZARV/NHL!
hCoV-19/South Africa/KRISP-K022434/202	Original	EPI_ISL_4575170	2021-08-09	2021-09-29	29,853	Human	Africa / South Afr	ZARV/NHL!
hCoV-19/South Africa/KRISP-K022442/202	Original	EPI_ISL_4575169	2021-08-06	2021-09-29	29,852	Human	Africa / South Afr	ZARV/NHL!
hCoV-19/South Africa/KRISP-K022435/202	Original	EPI_ISL_4575167	2021-08-07	2021-09-29	29,853	Human	Africa / South Afr	ZARV/NHL!
hCoV-19/South Africa/KRISP-K022473/202	Original	EPI_ISL_4575166	2021-08-13	2021-09-29	29,853	Human	Africa / South Afr	ZARV/NHL!

Patient Status Metadata:

- More detailed information on the patient from whom specimen was collected: Sampling strategy, clinical status, vaccination status...

Interacting with the GISAID database

Downloading selections of the data

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Virus name	Accession ID	Collection da	Location	Host	Additional lo	Sampling str	Gender	Patient age	Patient statu	Last vaccinat	Passage	Specimen	Additional hc	Lineage	Clade	AA Substitutions				
2	hCoV-19/Soc	EPI_ISL_3721	03/08/2021	Africa /	Sout Human			Female	37	Mild - not requiring suppl	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,N_L331V,NS3_S26L,NSP3_V1						
3	hCoV-19/Soc	EPI_ISL_3722	03/08/2021	Africa /	Sout Human			Male	41	Mild - not requiring suppl	Original	Nasopharyngeal and oropl	AY.33	GK	(N_G215C,N_D377Y,M_I82T,N_D63G,Spike_D950N,N_R203M,NS7b_T40I,N						
4	hCoV-19/Soc	EPI_ISL_3722	03/08/2021	Africa /	Sout Human			Female	22	Mild - not requiring suppl	Original	Nasopharyngeal and oropl	AY.19	GK	(NS7a_L116F,N_D63G,N_R203M,NSP12_G6715,NSP2_P129L,NS3_S26L,NS						
5	hCoV-19/Soc	EPI_ISL_3722	03/08/2021	Africa /	Sout Human			Male	34	Mild - not requiring suppl	Original	Nasopharyngeal and oropl	B.1.617.2	GK	(N_D377Y,M_I82T,NSP6_V149A,N_D63G,Spike_D950N,N_R203M,NSP3_P8						
6	hCoV-19/Soc	EPI_ISL_3721	03/08/2021	Africa /	Sout Human			Female	49	Mild - not requiring suppl	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,N_A119S,NS3_S26L,Spike_P6						
7	hCoV-19/Soc	EPI_ISL_3721	03/08/2021	Africa /	Sout Human			Male	58	Mild - not requiring suppl	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NSP3_L312F,NS3_S26L,Spike						
8	hCoV-19/Soc	EPI_ISL_3795	06/08/2021	Africa /	Sout Human		Sentinel Scr	Male	42	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NSP3_E195D,NS3_S26L,Spike					
9	hCoV-19/Soc	EPI_ISL_3795	02/08/2021	Africa /	Sout Human		Sentinel Scr	Female	97	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NS3_S26L,Spike_P681R,NSP3					
10	hCoV-19/Soc	EPI_ISL_3795	03/08/2021	Africa /	Sout Human		Sentinel Scr	Male	26	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NS3_S26L,Spike_P681R,NSP3					
11	hCoV-19/Soc	EPI_ISL_3795	02/08/2021	Africa /	Sout Human		Sentinel Scr	Female	71	unknown		Original	Nasopharyngeal and oropl	AY.5	GK	(N_D63G,NSP12_G6715,NS3_S26L,Spike_P681R,NSP3_P1228L,Spike_E156a					
12	hCoV-19/Soc	EPI_ISL_3795	02/08/2021	Africa /	Sout Human		Sentinel Scr	Male	29	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,NSP4_R112C,N_R203M,NSP12_G6715,NS3_S26L,NS3					
13	hCoV-19/Soc	EPI_ISL_3795	02/08/2021	Africa /	Sout Human		Staff	Female	unknown	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NS3_S26L,Spike_P681R,NSP3					
14	hCoV-19/Soc	EPI_ISL_3795	07/08/2021	Africa /	Sout Human			Male	9	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NS3_S26L,Spike_P681R,Spike					
15	hCoV-19/Soc	EPI_ISL_3795	06/08/2021	Africa /	Sout Human		Sentinel Scr	Female	45	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NS3_S26L,Spike_P681R,NSP12					
16	hCoV-19/Soc	EPI_ISL_3795	02/08/2021	Africa /	Sout Human		Sentinel Scr	Female	46	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NS3_S26L,Spike_P681R,Spike					
17	hCoV-19/Soc	EPI_ISL_3795	02/08/2021	Africa /	Sout Human		Sentinel Scr	Male	48	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NS3_S26L,Spike_P681R,Spike					
18	hCoV-19/Soc	EPI_ISL_3795	05/08/2021	Africa /	Sout Human		Sentinel Scr	Female	32	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NS3_S26L,Spike_P681R,NSP3					
19	hCoV-19/Soc	EPI_ISL_3795	08/08/2021	Africa /	Sout Human		Sentinel Scr	Female	34	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NSP2_T231,NS3_S26L,Spike					
20	hCoV-19/Soc	EPI_ISL_3795	02/08/2021	Africa /	Sout Human		Sentinel Scr	Female	44	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NSP8_T1371,NSP5_G283S,NS					
21	hCoV-19/Soc	EPI_ISL_3795	06/08/2021	Africa /	Sout Human		Sentinel Scr	Male	62	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,NSP3_S284G,N_R203M,NSP12_G6715,NS3_S26L,Spike					
22	hCoV-19/Soc	EPI_ISL_3795	02/08/2021	Africa /	Sout Human		Sentinel Scr	Female	51	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NS3_S26L,Spike_P681R,NSP3					
23	hCoV-19/Soc	EPI_ISL_3795	07/08/2021	Africa /	Sout Human		Sentinel Scr	Male	26	unknown		Original	Nasopharyngeal and oropl	B.1.617.2	GK	(NSP3_L1802F,N_G215C,E_P71L,N_D63G,N_R203M,NSP12_G6715,NS3_A3					
24	hCoV-19/Soc	EPI_ISL_3795	06/08/2021	Africa /	Sout Human		Sentinel Scr	unknown	30	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NSP2_A318V,Spike_H1159Y,N					
25	hCoV-19/Soc	EPI_ISL_3795	02/08/2021	Africa /	Sout Human		Sentinel Scr	Female	97	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NS3_S26L,Spike_P681R,NSP3					
26	hCoV-19/Soc	EPI_ISL_3795	07/08/2021	Africa /	Sout Human		Sentinel Scr	Female	50	unknown		Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,N_D63G,N_R203M,NSP12_G6715,NSP13_R271T,NS3_S26L,Spike					
27	hCoV-19/Soc	FPI_ISI_3795	02/08/2021	Africa /	Sout Human		Sentinel Scr	Female	15	unknown		Original	Nasopharyngeal and oropl	AY.38	GK	(N_D63G,N_R203M,NSP12_G6715,NSP6_T1811,NS3_S26L,NSP12_G7742,S					

Patient Status Metadata:

- More detailed information on the patient from whom specimen was collected: Sampling strategy, clinical status, vaccination status...

Interacting with the GISAID database

Downloading selections of the data

The screenshot displays the GISAID EpiCoV™ interface. At the top, the GISAID logo is on the left, and the copyright notice "© 2008 - 2021 | Terms of Use | Privacy Notice | Contact" is on the right. Below the logo, the user is logged in as "Houriiyah Tegally" with a "logout" link. The main navigation bar includes "Registered Users", "EpiFlu™", "EpiCoV™", and "My profile".

The "EpiCoV™" section is active, showing a search bar and a list of search results. A modal window is open, displaying download options for the selected data. The modal is divided into three sections:

- FASTA, metadata, Variant surveillance, per clade, per lineage, Dates + Locations**: Six download options, each with a green arrow icon.
- Submission and Variant statistics**: Four download options: "Global by month (xlsx)", "USA by month (xlsx)", "Variants VoC/Vol (xlsx)", and "Clade/Lineage, Variants (json)".
- Genomic epidemiology**: Five download options: "Audacity", "Custom Selection" (circled in red), "FASTA", "metadata", and "nextregions".

At the bottom of the modal, there is a "Back" button. The background shows a table of search results with columns for Accession ID, Location, Clade, Virus name, and other metadata. The total number of viruses is 4,249,775.

Interacting with the GISAID database

Downloading selections of the data

Clade Lineage Substitutions Variants

<input type="checkbox"/>	Virus name	Passage d	Accession ID	Collection d	Submission	Length	Host	Location	Originating lab
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101733519	Original	EPI_ISL_5055855	2021-09-09	2021-10-12	29,007	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729463	Original	EPI_ISL_5055854	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101731818	Original	EPI_ISL_5055853	2021-09-10	2021-10-12	29,481	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101730983	Original	EPI_ISL_5055852	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101730809	Original	EPI_ISL_5055851	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101734190	Original	EPI_ISL_5055850	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101730330	Original	EPI_ISL_5055849	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101736002	Original	EPI_ISL_5055848	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101736420	Original	EPI_ISL_5055847	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729602	Original	EPI_ISL_5055846	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101731853	Original	EPI_ISL_5055845	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729442	Original	EPI_ISL_5055844	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729461	Original	EPI_ISL_5055843	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729447	Original	EPI_ISL_5055842	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729437	Original	EPI_ISL_5055841	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729435	Original	EPI_ISL_5055840	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729411	Original	EPI_ISL_5055839	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729402	Original	EPI_ISL_5055838	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729381	Original	EPI_ISL_5055837	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729358	Original	EPI_ISL_5055836	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729322	Original	EPI_ISL_5055835	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729115	Original	EPI_ISL_5055834	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729098	Original	EPI_ISL_5055833	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729084	Original	EPI_ISL_5055832	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar

Back Total: 4,249,775 viruses << < 1 2 3 4 5 > >> Select

Interacting with the GISAID database

Downloading selections of the data

The screenshot displays the GISAID database search interface. A modal dialog box titled "Select combination of entries" is open, allowing users to specify a list of accession numbers for download. The dialog includes a text input field containing a list of EPI_ISL numbers (e.g., EPI_ISL_3799010 to EPI_ISL_3799042) and a "Choose file" button. Below the input field, there are radio buttons for "Accession Numbers" (selected) and "Sample ID given by originating lab". At the bottom of the dialog, there are buttons for "Back", "CSV", "EPI-SET", and "OK". The "OK" button is circled in red. The background shows a table of search results with columns for Accession ID, Virus name, Location, Host, and other metadata.

Accession ID	Virus name	Location	Host	Submission Date	Collection Date	Length	Host	Region	Country	Originating lab
hCoV-19/USA/CO-CDPHE-2101729322	Original	EPI_ISL_5055835	2021-09-10	2021-10-12	29,769	Human	North America	Colorado	Depar	
hCoV-19/USA/CO-CDPHE-2101729115	Original	EPI_ISL_5055834	2021-09-09	2021-10-12	29,769	Human	North America	Colorado	Depar	

Interacting with the GISAID database

Downloading selections of the data

Clade Lineage Substitutions Variants

<input type="checkbox"/>	Virus name	Passage d	Accession ID	Collection d	Submission	Length	Host	Location	Originating lab
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101733519	Original	EPI_ISL_5055855	2021-09-09	2021-10-12	29,007	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729463	Original	EPI_ISL_5055854	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101731818	Original	EPI_ISL_5055853	2021-09-10	2021-10-12	29,481	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101730983	Original	EPI_ISL_5055852	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101730809	Original	EPI_ISL_5055851	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101734190	Original	EPI_ISL_5055850	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101730330	Original	EPI_ISL_5055849	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101736002	Original	EPI_ISL_5055848	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101736420	Original	EPI_ISL_5055847	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729602	Original	EPI_ISL_5055846	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101731853	Original	EPI_ISL_5055845	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729442	Original	EPI_ISL_5055844	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729461	Original	EPI_ISL_5055843	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729447	Original	EPI_ISL_5055842	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729437	Original	EPI_ISL_5055841	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729435	Original	EPI_ISL_5055840	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729411	Original	EPI_ISL_5055839	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729402	Original	EPI_ISL_5055838	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729381	Original	EPI_ISL_5055837	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729358	Original	EPI_ISL_5055836	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729322	Original	EPI_ISL_5055835	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729115	Original	EPI_ISL_5055834	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729098	Original	EPI_ISL_5055833	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
<input type="checkbox"/>	hCoV-19/USA/CO-CDPHE-2101729084	Original	EPI_ISL_5055832	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar

Back Total: 4,249,775 viruses << < 1 2 3 4 5 > >>

Interacting with the GISAID database

Downloading global aggregated of the data

The screenshot shows the GISAID database interface. At the top, there is a navigation bar with the GISAID logo and user information: "© 2008 - 2021 | Terms of Use | Privacy Notice | Contact" and "You are logged in as Houriiyah Tegally - [logout](#)". Below this is a menu with "Registered Users", "EpiFlu™", "EpiCoV™", and "My profile". The "EpiCoV™" section is active, showing "Search", "Downloads" (circled in red), and "Upload" buttons.

The "Search" section contains various filters: "Accession ID", "Virus name", "Location", "Host", "Collection", "Submission", "Clade", "Lineage", "Substitutions", and "Variants". There are also checkboxes for "complete", "high coverage", "low coverage excl", "w/Patient status", and "collection date compl".

Below the filters is a table of search results. The table has columns: "Virus name", "Passage de", "Accession ID", "Collection da", "Submission C", "Length", "Host", "Location", and "Originating". The first few rows show hCoV-19 sequences from the USA/IA-SHL region, collected in 2021 and submitted in 2021. The table is paginated, showing results 1 through 5.

At the bottom of the table, there is a "Total: 4,248,757 viruses" and a "Select" button. There are also "Analysis" and "Download" buttons at the bottom right.

<input type="checkbox"/>	Virus name	Passage de	Accession ID	Collection da	Submission C	Length	Host	Location	Originating
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812279/2021	Original	EPI_ISL_5054789	2021-08-23	2021-10-12	29,765	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812268/2021	Original	EPI_ISL_5054788	2021-08-24	2021-10-12	29,598	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812004/2021	Original	EPI_ISL_5054787	2021-08-23	2021-10-12	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812281/2021	Original	EPI_ISL_5054786	2021-08-23	2021-10-12	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812277/2021	Original	EPI_ISL_5054785	2021-08-23	2021-10-12	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812301/2021	Original	EPI_ISL_5054784	2021-08-23	2021-10-12	29,598	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1811999/2021	Original	EPI_ISL_5054783	2021-08-23	2021-10-12	29,598	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812027/2021	Original	EPI_ISL_5054782	2021-08-23	2021-10-12	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IL-SHL-1812284/2021	Original	EPI_ISL_5054781	2021-08-20	2021-10-12	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812007/2021	Original	EPI_ISL_5054780	2021-08-23	2021-10-12	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812252/2021	Original	EPI_ISL_5054779	2021-08-23	2021-10-12	29,769	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812237/2021	Original	EPI_ISL_5054778	2021-08-24	2021-10-12	29,598	Human	North America / U	State Hygie
<input type="checkbox"/>	hCoV-19/USA/IA-SHL-1812218/2021	Original	EPI_ISL_5054777	2021-08-23	2021-10-12	29,769	Human	North America / U	State Hvae

Interacting with the GISAID database

Downloading global aggregated of the data

Alignment and proteins



MSA
full1008
(277 MB)



MSA
unmasked1006
(199 MB)



MSA
masked1006
(191 MB)



Archive



allprot1011
(711 MB)



spikeprot1011
(74 MB)



allnuc1010
(895 MB)



spikenuc1010
(91 MB)



FASTA



metadata



Variant
surveillance



per clade



per lineage



Dates +
Locations

Submission and Variant statistics



Global
by month
(xlsx)



USA
by month
(xlsx)



Variants
VoC/Vol
(xlsx)



Clade/Lineage,
Variants
(json)

Genomic epidemiology



Audacity



Custom
Selection



FASTA



metadata



nextregions



Back

Download packages



FASTA



metadata



Variant
surveillance



per clade



per lineage



Dates +
Locations



Back

Region-specific Auspice source files



Global



Africa



Asia



Europe



North
America



South
America



Oceania



Corresponds to the global or regional phylogenetic builds maintained by Nexstrain:

- <https://nextstrain.org/ncov/gisaid/global>
- <https://nextstrain.org/ncov/gisaid/africa>
- Etc...

Summary

As a scientific community, we have responsibility to produce and share high quality data. The GISAID infrastructure offers a robust framework that can help improve the quality of data we generate.

Acknowledgements



UKZN INSPIRING GREATNESS