



UNIVERSITY OF TM
KWAZULU-NATAL
INYUVESI
YAKWAZULU-NATALI



INTRODUCTION TO GISAID

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



EDGEWOOD CAMPUS



HOWARD COLLEGE CAMPUS



NELSON R MANDELA SCHOOL OF MEDICINE



PIETERMARITZBURG CAMPUS



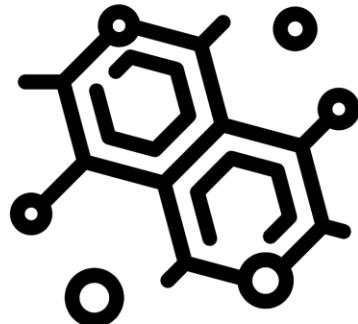
WESTVILLE CAMPUS

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

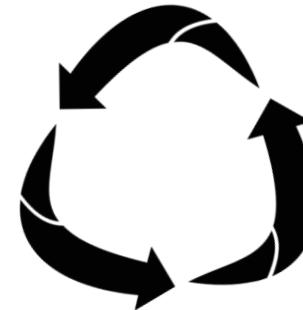
Why is it important to publish our data?



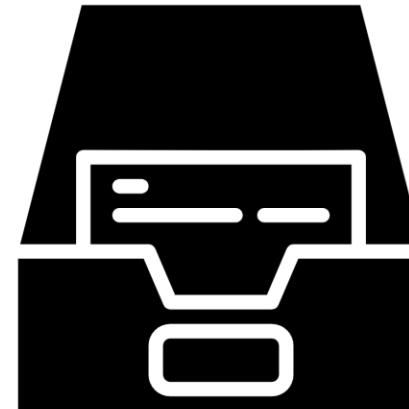
- Fast track science



- Metadata QC and Standardization



- Reproducible research

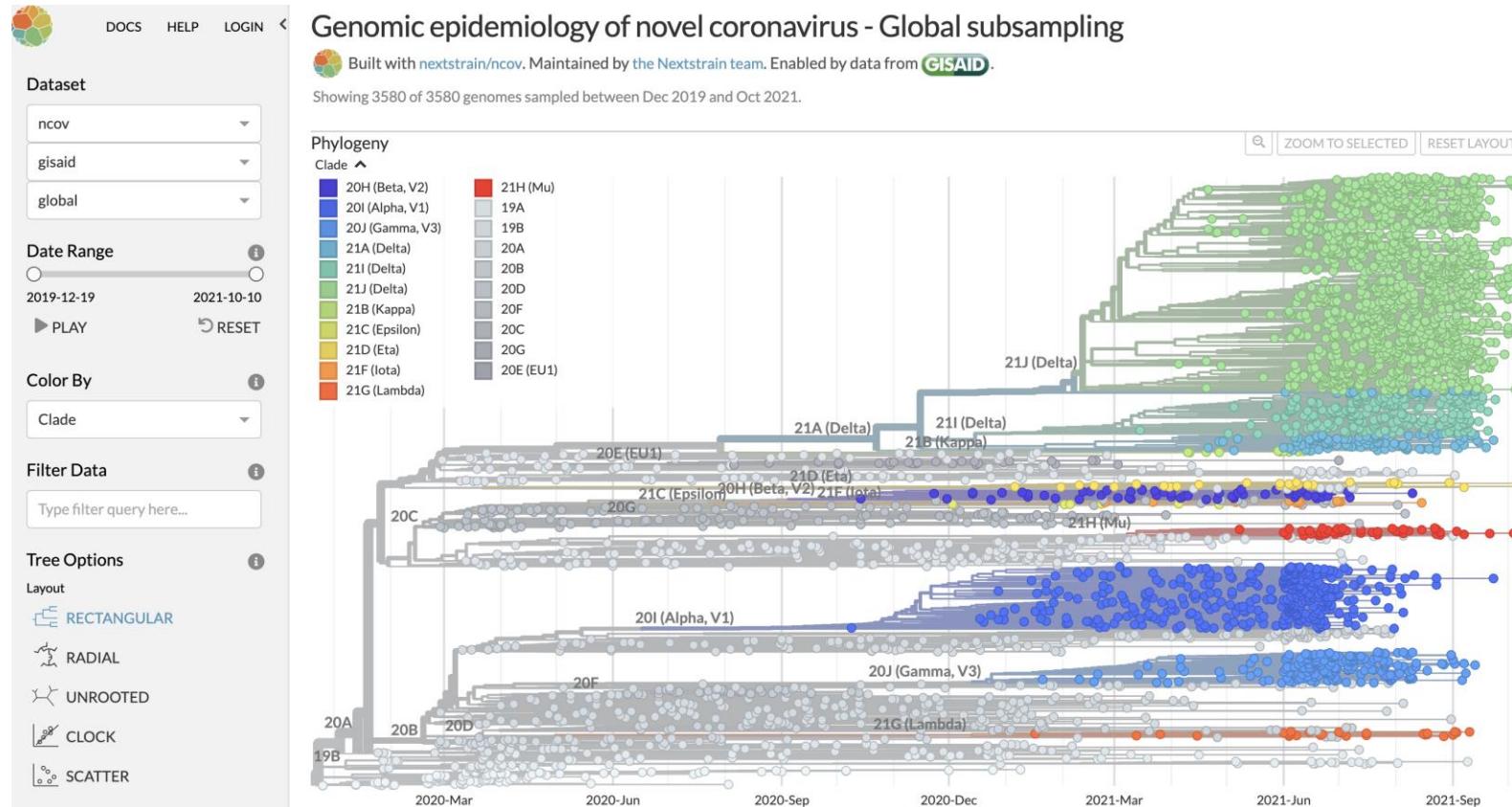


- Permanent storage

<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?

[Login](#)[About us](#)[Database Features](#)[Events](#)[Collaborations](#)[References](#)[Registration](#)[Help](#)

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.



hCoV-19 Submission Tracking



hCoV-19 Tracking of Variants



Genomic epidemiology of hCoV-19



hCoV-19 data sharing via GISAID

~~3,398,338~~ submissions **4,206,853**

Enabled by data shared via GISAID

CoVGlobe

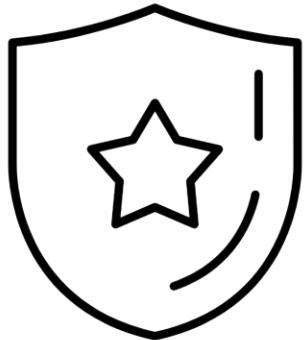
The Francis Crick Institute (United Kingdom)
Aggregate three-week rolling averages of lineages in different locations

GISAID Resources

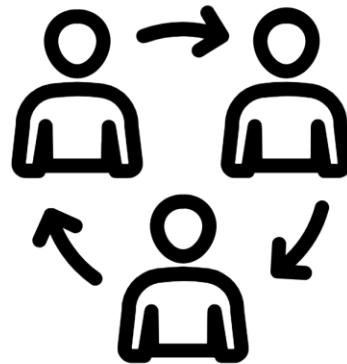
[Free Access Credentials](#)

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

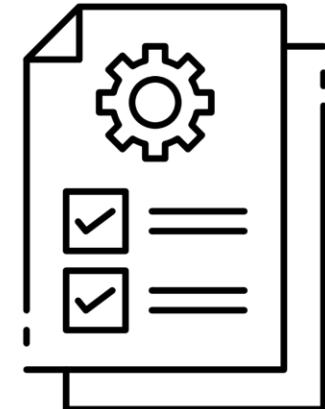
Why GISAID?



❖ Recognition



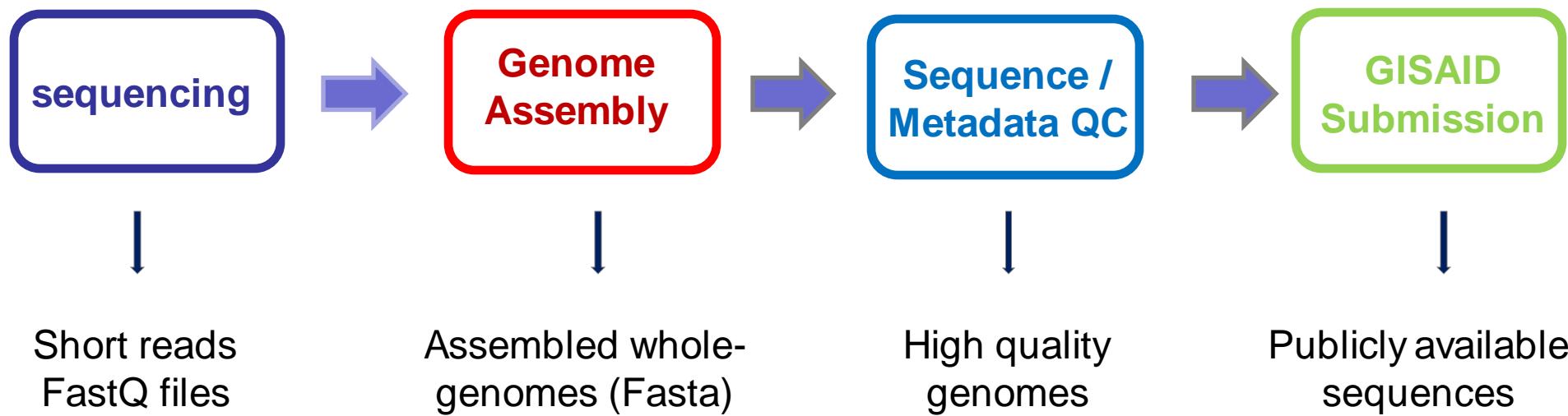
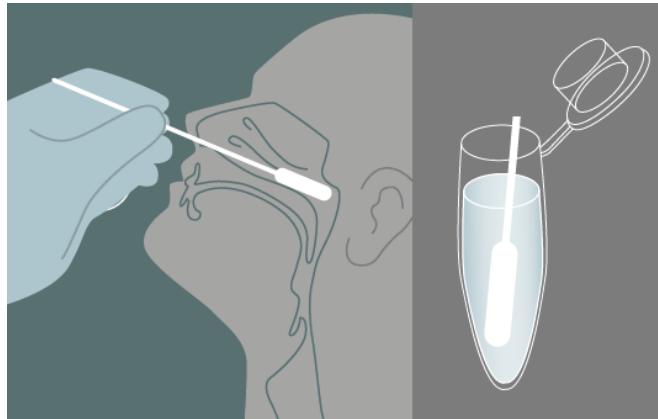
❖ Collaboration



❖ Quality Control

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

Data Production Workflow



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**



SeqKit - Ultrafast FASTA/Q kit

Nextclade beta



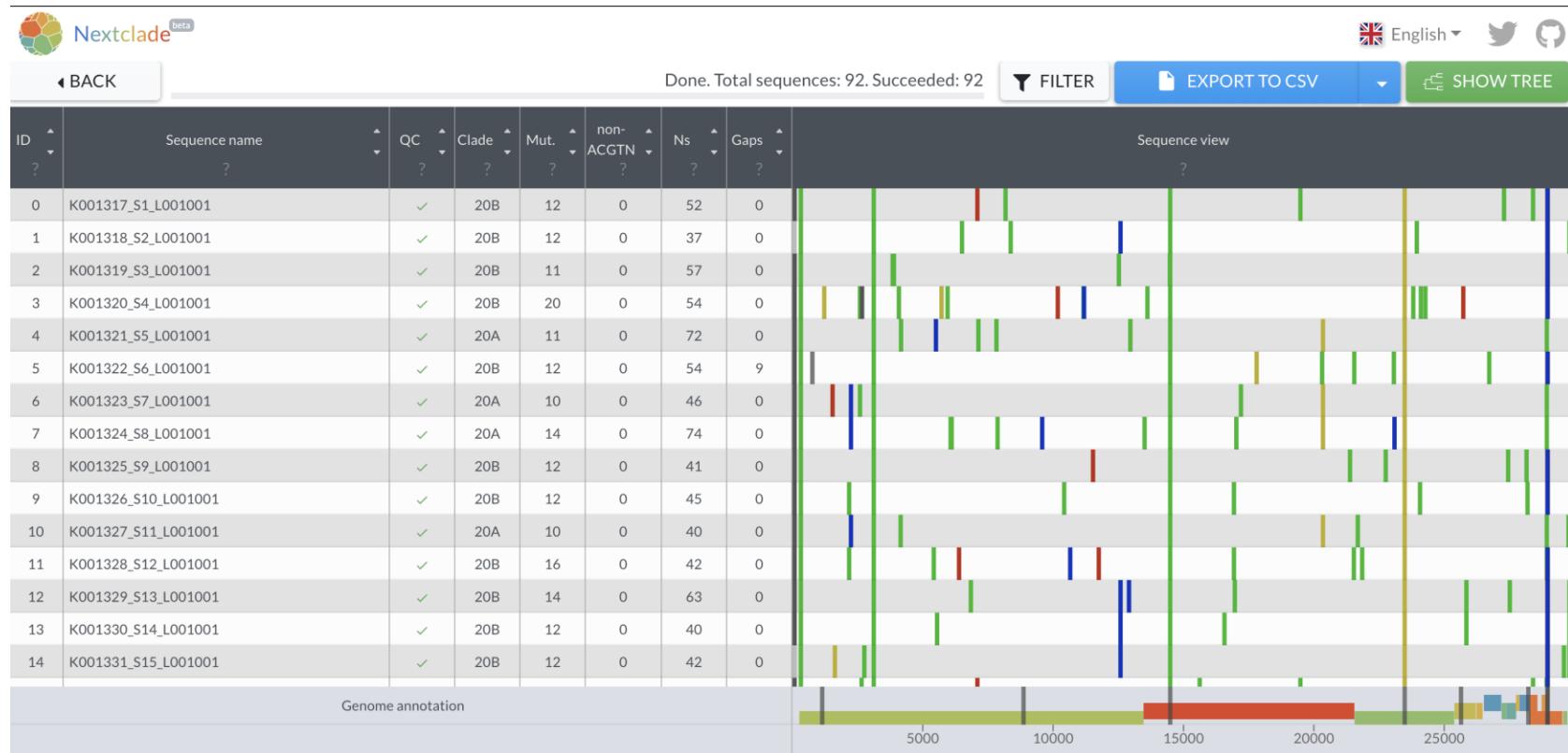
Nextstrain

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 case 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.overallScc	qc.overallSta	totalSubstitu	totalDeletio	totalInsertio	totalAminoac	totalAminoac	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,T21458		7
K019116	21A (Delta)	10126.32235	bad	4	36	0	7	18	27470	0	0	T6607A,C17 ² 26785-26820		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,A1 18914,2281		11
K019114	20G	9303.978056	bad	23	9	6	16	1	26238	0	1	T928C,G9321936,944-951		12
K019112	19A	9027.146145	bad	5	5	0	3	0	25873	0	0	T18672C,C18 7433-7437		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,T16116384,23137		17
K019108	20I (Alpha, V 446.494475	bad		27	20	0	14	6	5604	0	2	C241T,C913T 11288-11296		81
K019163	20B	546.302500	bad	39	10	9	26	0	4836	0	4	C1059T,C303 24457-24465		84
K019100	20I (Alpha, V 290.680123	bad		28	20	0	14	6	4434	0	4	C241T,C913T 11288-11296		85
K019131	20I (Alpha, V 331.106547	bad		43	36	0	19	6	3994	0	4	C829T,C913T 11288-11296		87
K019167	20I (Alpha, V 140.791238	bad		26	21	0	15	3	2780	0	4	C913T,C3037 7118-7119,1		91
K019410	20B	91.878398	mediocre	28	0	0	15	0	2573	0	4	C222T,C241T,C1190T,G18		91
K019099	20I (Alpha, V 67.534623	mediocre		31	20	0	18	6	1207	0	4	C241T,C913T 11288-11296		96
K019106	20I (Alpha, V 64.069273	mediocre		30	20	0	16	6	1055	0	4	C241T,C913T 11288-11296		96
K019156	20D	56.820864	mediocre	36	12	0	20	4	504	0	4	C241T,C843T 21983-21994		98
K019107	20I (Alpha, V 56.250000	mediocre		29	20	0	16	6	156	0	4	C241T,C913T 11288-11296		99
K019110	20I (Alpha, V 56.250000	mediocre		29	20	0	16	6	157	0	4	C241T,C913T 11288-11296		99
K019115	20I (Alpha, V 4.340278	good		38	19	0	23	6	281	0	4	G210T,C241T 11288-11296		99
K019117	20I (Alpha, V 56.250000	mediocre		29	20	0	16	6	157	0	4	C241T,C913T 11288-11296		99
K019128	20I (Alpha, V 21.006944	good		36	19	0	22	6	293	0	1	G210T,C241T 11288-11296		99
K019158	20I (Alpha, V 56.944444	mediocre		29	20	0	16	6	210	0	4	C241T,C913T 11288-11296		99
K019165	20I (Alpha, V 56.466629	mediocre		32	20	0	17	6	356	0	4	C241T,C913T 11288-11296		99

Love Coverage

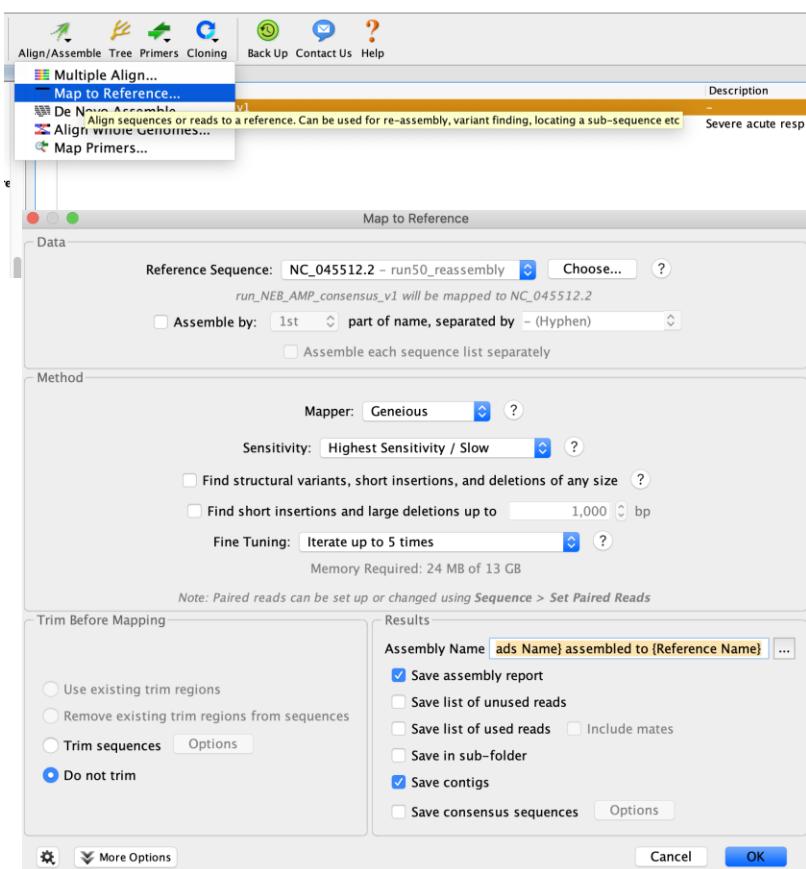
High Coverage

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

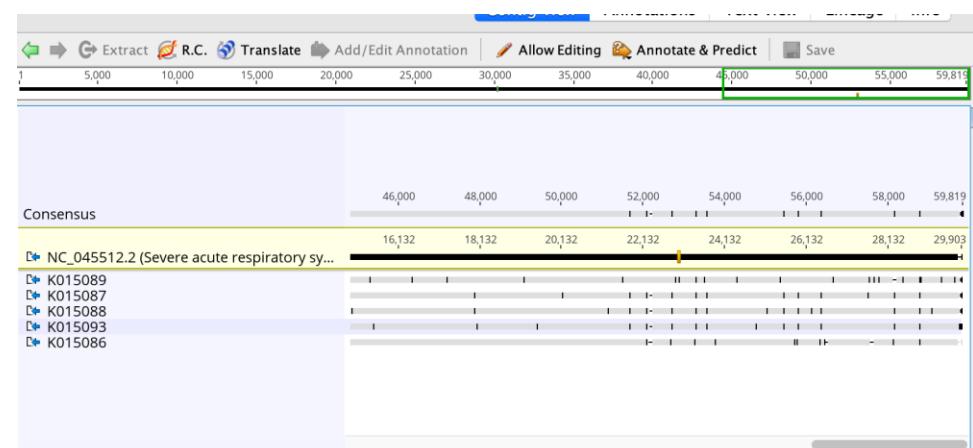
GENEIOUS – cleaning sequences

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 unknown
Patient age	mandatory	e.g. 65 or 7 months, or unknown
Patient status	mandatory	e.g. Hospitalized, Released, Live, Deceased, or unknown
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

Screenshot of Microsoft Excel showing two sheets of data for preparing GISAID files.

Sheet 1: CAF_run312_consensus_gisaid_n6

This sheet contains 8 rows of data. The columns represent various metadata fields:

	A	B	C	D	E	F	G	H	I	J	K	L
1	submitter	fn	covv_virus_name	covv_type	covv_paa	covv_collection_date	covv_location	covv_add_location	covv_host	covv_add_host_info	covv_sampling_strategy	covv_gen
2	tuliodna	fm	covv_virus_name	covv_type	covv_paa	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

Sheet 2: CAF_run312_consensus_gisaid_n6

This sheet contains 14 rows of data. The columns represent sequencing and assembly details:

	N	O	P	Q	R	S	T	U	V	W	X	Y	
1	ent_age	covv_patient_status	covv_specimen	covv_outbreak	covv_last_vaccin	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	ent_age	covv_patient_status	covv_specimen	covv_outbreak	covv_last_vaccin	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					KRISP, KZN Research Innov
4	27		SWAB_NASO_OROP				Ion Torrent	Genome Detective 1.133					KRISP, KZN Research Innov
5	33		SWAB_NASO_OROP				Ion Torrent	Genome Detective 1.133					KRISP, KZN Research Innov
6	54		SWAB_NASO_OROP				Ion Torrent	Genome Detective 1.133					KRISP, KZN Research Innov
7	57		SWAB_NASO_OROP				Ion Torrent	Genome Detective 1.133					KRISP, KZN Research Innov
8	36		SWAB_NASO_OROP				Ion Torrent	Genome Detective 1.133					KRISP, KZN Research Innov
9													
10													
11													
12													
13													
14													

Cont...

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
```

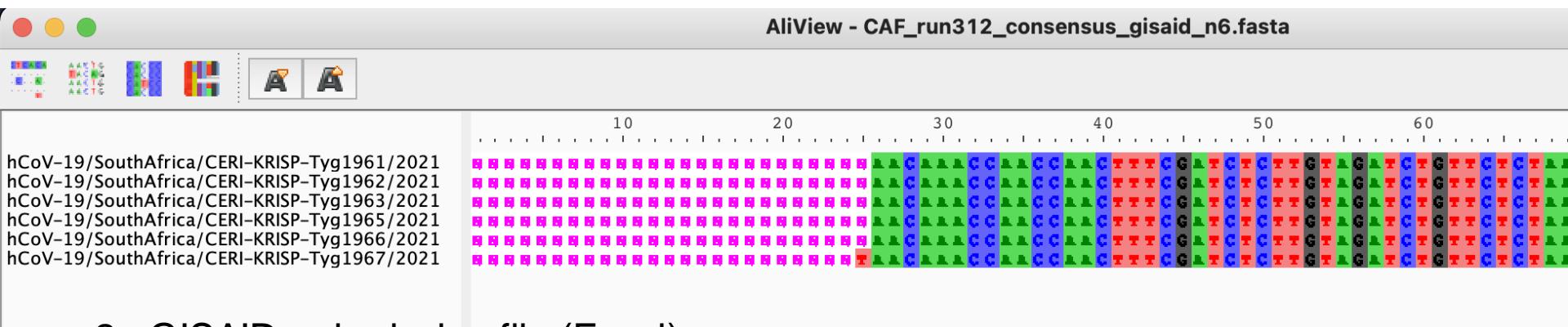
One liner to batch update

```
$ 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_pاس;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_pاس;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
9												

GISAID Upload

GISAID

© 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

The logo for the Max Planck Institute (MPI) Informatik, featuring a circular emblem with a profile of a head and the letters "MPI" in a stylized font next to it.

max planck institut
informatik

Ensure that you have a valid account

GISAID Upload

epicov.org/epi3/frontend#958d

GISAID

© 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 bat coronavirus (SAR-CoV).

On 10 January 2020, The World Health Organization announced the first sequences to enable identification of potential targets, including sequencing of the virus genome, phylogenetic analysis, and other analyses.

 Single upload

 Batch upload

 by A*STAR Singapore

 Audacity

 AudacityInstant

 BLAST

 CoVizue

 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 that 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.

The screenshot shows the GISAID hCoV-19 Batch Upload interface. At the top, there's a navigation bar with links for Registered Users, EpiFlu™, EpiCoV™ (which is selected), My profile, and Administration. Below the navigation bar are several icons: EpiCoV™, Browse, Downloads, Upload, Batch Upload, My Unreleased, and Curation. A dark green header bar displays the title "GISAID hCoV-19 Batch Upload". Below it, a red text box contains instructions: "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 main form area has three sections: "Metadata as Excel or CSV*" with a file input field showing "No file selected.", "Sequences as FASTA*" with a file input field showing "No file selected.", and "Report" with a text input field showing "Upload XLS/CSV and FASTA.". At the bottom left is a button "Download Instructions and Template" with a checkmark icon. At the bottom right are two buttons: "Contact Curator" with an envelope icon and "Check and Submit" with a checkmark icon. A dashed arrow from the left points to the "Download Instructions and Template" button, and another dashed arrow from the right points to the "Check and Submit" button.

2. Download instructions

3. Upload your data

Download Instructions and Template

Contact Curator

Check and Submit



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 shows the GISAID hCoV-19 Batch Upload interface. At the top, there are tabs for Registered Users, EpiFlu™, EpiCoV™ (selected), My profile, and Administration. Below the tabs are links for EpiCoV™, Browse, Downloads, Upload, Batch Upload, My Unreleased, and Curation. The main area is titled "GISAID hCoV-19 Batch Upload" and contains instructions: "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." There are two upload sections: "Metadata as Excel or CSV*" and "Sequences as FASTA*". Both sections have a file input field showing "Test.csv (35.1 kbytes)" and "Test.fasta (1.20 MB)" respectively, with a note "max size: 5M" and "max size: 32M". A "Browse..." button is present in each. A modal dialog box titled "Message" appears with the text "Errors in Batch-Upload-Data! Please read report." and an "OK" button. In the bottom left, a "Report" link leads to a box containing the error message: "Excel-Sheet contains 44 submissions Error: Sequence Virus name not found in FASTA Error: Sequence Virus name not found in FASTA". At the bottom right, there are buttons for "Download Instructions and Template", "Contact Curator", and "Check and Submit". The "Check and Submit" button is circled in black. Dashed arrows point from callout boxes labeled "5. Error Report" and "4. Check and Submit" to their respective locations on the page.



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

GISAID

© 2008 - 2021 | Terms of Use | Privacy Notice | Contact
You are logged in as Houriiyah Tegally - logout

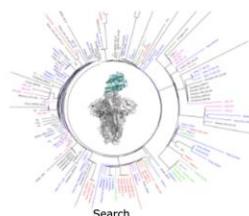
Registered Users EpiFlu™ EpiCoV™ My profile

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.

A circular phylogenetic tree showing the relationships between various SARS-CoV-2 variants. The tree is rooted at the center and branches outwards into numerous clades, each represented by a different color. Labels on the branches indicate specific variants or lineages.

Analysis Tools:

- Audacity
- AudacityInstant
- BLAST
- Covizut[®]
- Emerging Variants
- Official GISAID reference sequence
- PrimerChecker
- Spike glycoprotein mutation surveillance

GISAID

© 2008 - 2021 | Terms of Use | Privacy Notice | Contact
You are logged in as Houriiyah 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

The screenshot shows the GISAID search interface. A large black arrow points from the top-left towards the search bar. Several search fields are highlighted with red circles: 'Virus name', 'Location', 'Collection', 'Submission', 'Variants', and a blue circle highlights 'Total: 4,248,757 viruses' at the bottom left. The search results table includes columns for Virus name, Passage date, Accession ID, Collection date, Submission date, Length, Host, Location, and Originating. The first few rows of the table are:

	Virus name	Passage date	Accession ID	Collection da	Submission d	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
<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
<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
<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
<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
<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
<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
<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
<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
<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
<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
<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
Total: 4,248,757 viruses									

At the bottom right, there are buttons for 'Select', 'Analysis', and 'Download'.

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 search interface with the following parameters highlighted by red circles:

- Location: South Africa
- Collection date: 2021-08-01 to 2021-08-31
- Variants: VOC Delta GK/478K.V1

The results table displays 2,026 viruses, with the last entry shown as:

	Virus name	Passage de	Accession ID	Collection da	Submission D	Length	Host	Location	Originating
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955117	2021-08-30	2021-10-07	29,769	Human	Africa / South Afr	NHLS Unive
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955110	2021-08-31	2021-10-07	29,166	Human	Africa / South Afr	NHLS Unive
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955098	2021-08-31	2021-10-07	28,858	Human	Africa / South Afr	NHLS Unive
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955080	2021-08-30	2021-10-07	29,700	Human	Africa / South Afr	NHLS Unive
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955074	2021-08-30	2021-10-07	29,488	Human	Africa / South Afr	NHLS Unive
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955066	2021-08-30	2021-10-07	28,865	Human	Africa / South Afr	NHLS Unive
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955060	2021-08-30	2021-10-07	29,481	Human	Africa / South Afr	NHLS Unive
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955056	2021-08-31	2021-10-07	29,769	Human	Africa / South Afr	NHLS Unive
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955041	2021-08-30	2021-10-07	29,769	Human	Africa / South Afr	NHLS Unive
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955036	2021-08-30	2021-10-07	29,481	Human	Africa / South Afr	NHLS Unive
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955031	2021-08-30	2021-10-07	29,481	Human	Africa / South Afr	NHLS Unive
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955013	2021-08-31	2021-10-07	29,768	Human	Africa / South Afr	NHLS Unive
	hCoV-19/South Africa/UFS-VIRO-NGS-MS0	Original	EPI_ISL_4955003	2021-08-30	2021-10-07	29,487	Human	Africa / South Afr	NHLS Unive

Total: 2,026 viruses

Navigation: << < 1 2 3 4 5 > >>

Buttons: Select, Analysis, Download

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 Houriiyah Tegally - [logout](#)

EpiCoV™ Search Downloads Upload

Search

Accession ID Virus name **KRISP** complete high coverage
 low coverage excl w/Patient status
Location South Africa Host
Collection 2021-08-01 to 2021-08-31 Submission to
 collection date compl
Clade all Lineage Substitutions Variants VOC Delta GK/478K.V1

<input type="checkbox"/>	Virus name	Passage de	Accession ID	Collection da	Submission d		Length	Host	Location	Originating
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022470/202	Original	EPI_ISL_4575203	2021-08-14	2021-09-29		29,853	Human	Africa / South Afr	ZARV/NHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022485/202	Original	EPI_ISL_4575200	2021-08-12	2021-09-29		29,853	Human	Africa / South Afr	ZARV/NHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022436/202	Original	EPI_ISL_4575191	2021-08-07	2021-09-29		29,852	Human	Africa / South Afr	ZARV/NHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022462/202	Original	EPI_ISL_4575190	2021-08-02	2021-09-29		29,852	Human	Africa / South Afr	ZARV/NHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K023564/202	Original	EPI_ISL_4575189	2021-08-21	2021-09-29		29,852	Human	Africa / South Afr	ZARV/NHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K023567/202	Original	EPI_ISL_4575185	2021-08-20	2021-09-29		29,792	Human	Africa / South Afr	ZARV/NHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022453/202	Original	EPI_ISL_4575184	2021-08-04	2021-09-29		29,853	Human	Africa / South Afr	ZARV/NHL
<input type="checkbox"/>	hCoV-19/South Africa/KRISP-K022432/202	Original	EPI_ISL_4575183	2021-08-10	2021-09-29		29,853	Human	Africa / South Afr	ZARV/NHL
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
Total: 537 viruses										

<< < 1 2 3 4 5 > >>

Select Analysis Download

Interacting with the GISAID database

Downloading selections of the data

The screenshot shows the EpiCoV™ interface with a search results table and a 'Download' dropdown menu.

Search Results:

Accession ID	Virus name	Host	Collection Date	Lineage	Clade	Location	Originating Lab	
KRISP								
hCoV-19/South Africa/KRISP-K022						Africa / South Afr	ZARV/NHL	
hCoV-19/South Africa/KRISP-K022						Africa / South Afr	ZARV/NHL	
hCoV-19/South Africa/KRISP-K022						Africa / South Afr	ZARV/NHL	
hCoV-19/South Africa/KRISP-K022						Africa / South Afr	ZARV/NHL	
hCoV-19/South Africa/KRISP-K023						Africa / South Afr	ZARV/NHL	
hCoV-19/South Africa/KRISP-K023						Africa / South Afr	ZARV/NHL	
hCoV-19/South Africa/KRISP-K022						Africa / South Afr	ZARV/NHL	
hCoV-19/South Africa/KRISP-K022						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

Total: 537 viruses

Download Options:

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

Download

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_ISL_genbank_acc	date	region	country	division	location	region_expo	country_expo	division_expi	segment	length	host	age	sex	Nextstrain_c	pangolin_lineage	GISAID_ID	
2	hCoV-19/Sol betacoronavirus EPI_ISL_372?			03/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29853	Human	37	Female	?	AY.4	GK		
3	hCoV-19/Sol betacoronavirus EPI_ISL_372?			03/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29859	Human	41	Male	?	AY.33	GK		
4	hCoV-19/Sol betacoronavirus EPI_ISL_372?			03/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29859	Human	22	Female	?	AY.19	GK		
5	hCoV-19/Sol betacoronavirus EPI_ISL_372?			03/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29848	Human	34	Male	?	B.1.617.2	GK		
6	hCoV-19/Sol betacoronavirus EPI_ISL_372?			03/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29853	Human	49	Female	?	AY.4	GK		
7	hCoV-19/Sol betacoronavirus EPI_ISL_372?			03/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29853	Human	58	Male	?	AY.4	GK		
8	hCoV-19/Sol betacoronavirus 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 betacoronavirus EPI_ISL_379?			02/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29859	Human	97	Female	?	AY.4	GK		
10	hCoV-19/Sol betacoronavirus EPI_ISL_379?			03/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29859	Human	26	Male	?	AY.4	GK		
11	hCoV-19/Sol betacoronavirus 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 betacoronavirus EPI_ISL_379?			02/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29859	Human	29	Male	?	AY.4	GK		
13	hCoV-19/Sol betacoronavirus EPI_ISL_379?			02/08/2021	Africa	South Africa	Gauteng	City of Johan Africa	South Africa	Gauteng	genome	29859	Human	unknown	Female	?	AY.4	GK		
14	hCoV-19/Sol betacoronavirus 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 betacoronavirus 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 betacoronavirus EPI_ISL_379?			02/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29853	Human	46	Female	?	AY.4	GK		
17	hCoV-19/Sol betacoronavirus EPI_ISL_379?			02/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29874	Human	48	Male	?	AY.4	GK		
18	hCoV-19/Sol betacoronavirus 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 betacoronavirus EPI_ISL_379?			08/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29850	Human	34	Female	?	AY.4	GK		
20	hCoV-19/Sol betacoronavirus EPI_ISL_379?			02/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29874	Human	44	Female	?	AY.4	GK		
21	hCoV-19/Sol betacoronavirus EPI_ISL_379?			06/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29853	Human	62	Male	?	AY.4	GK		
22	hCoV-19/Sol betacoronavirus EPI_ISL_379?			04/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29844	Human	51	Female	?	AY.4	GK		
23	hCoV-19/Sol betacoronavirus EPI_ISL_379?			07/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29853	Human	26	Male	?	B.1.617.2	GK		
24	hCoV-19/Sol betacoronavirus EPI_ISL_379?			06/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29818	Human	30	unknown	?	AY.4	GK		
25	hCoV-19/Sol betacoronavirus EPI_ISL_379?			02/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29859	Human	97	Female	?	AY.4	GK		
26	hCoV-19/Sol betacoronavirus EPI_ISL_379?			07/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29852	Human	50	Female	?	AY.4	GK		
27	hCoV-19/Sol betacoronavirus EPI_ISL_379?			02/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29758	Human	15	Female	?	AY.38	GK		
28	hCoV-19/Sol betacoronavirus 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 betacoronavirus EPI_ISL_379?			02/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29853	Human	27	Female	?	AY.4	GK		
30	hCoV-19/Sol betacoronavirus EPI_ISL_379?			01/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal genome		29796	Human	44	Female	?	AY.38	GK		
31	hCoV-19/Sol betacoronavirus EPI_ISL_379?			10/08/2021	Africa	South Africa	KwaZulu-Natal eThekweni	Africa	South Africa	KwaZulu-Natal 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 shows the GISAID database interface. At the top, there are tabs for Registered Users, EpiFlu™, EpiCoV™ (which is selected), and My profile. Below the tabs are buttons for EpiCoV™, Search, Downloads, and Upload. The main area is titled 'Search' and shows a search form with fields for Accession ID, Virus name (KRISP), Location (South africa), Collection date (2021-08-01 to 2021-08-15), Host, and various filters like complete, high coverage, low coverage excl, and w/Patient status. A 'Download' button is highlighted, showing a dropdown menu with options: Sequences (FASTA), Patient status metadata (selected), Sequencing technology metadata, Dates and Location, Input for the Augur pipeline, and Acknowledgement (Supplemental table). The results table lists 537 viruses, mostly from South Africa. The first few rows are: hCoV-19/South Africa/KRISP-K022, hCoV-19/South Africa/KRISP-K022, hCoV-19/South Africa/KRISP-K022, hCoV-19/South Africa/KRISP-K022, hCoV-19/South Africa/KRISP-K023, hCoV-19/South Africa/KRISP-K023, hCoV-19/South Africa/KRISP-K022, hCoV-19/South Africa/KRISP-K022, hCoV-19/South Africa/KRISP-K023561/202, hCoV-19/South Africa/KRISP-K022434/202, hCoV-19/South Africa/KRISP-K022442/202, hCoV-19/South Africa/KRISP-K022435/202, and hCoV-19/South Africa/KRISP-K022473/202. The table includes columns for Virus name, Host, Location, Originating, and a 'Select' checkbox. At the bottom, there are navigation links (Back, Download, Analysis, and another Download button), a page number (1 of 5), and a 'Select' checkbox.

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/Sol EPI_ISL_3722	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/Sol 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,Spik						
4	HCoV-19/Sol EPI_ISL_3722	03/08/2021	Africa / Sout Human					Female	22	Mild - not requiring suppl	Original	Nasopharyngeal and oropl	AY.19	GK	(N57a_L116F_N_D63G_N_R203M,NSP12_G6715,NSP2_P129L,NS3_S26L,NS						
5	HCoV-19/Sol 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_NS						
6	HCoV-19/Sol EPI_ISL_3722	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,Spik						
7	HCoV-19/Sol EPI_ISL_3722	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,Spik						
8	HCoV-19/Sol EPI_ISL_3795	06/08/2021	Africa / Sout Human					Sentinal Scre Male	42	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NSP3_E195D,NS3_S26L,Spik						
9	HCoV-19/Sol EPI_ISL_3795	02/08/2021	Africa / Sout Human					Sentinal Scre Female	97	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NS3_S26L,Spik						
10	HCoV-19/Sol EPI_ISL_3795	03/08/2021	Africa / Sout Human					Sentinal Scre Male	26	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NS3_S26L,Spik						
11	HCoV-19/Sol EPI_ISL_3795	02/08/2021	Africa / Sout Human					Sentinal Scre Female	71	unknown	Original	Nasopharyngeal and oropl	AY.5	GK	(N_D63G,NSP12_G6715,NS3_S26L,Spik						
12	HCoV-19/Sol EPI_ISL_3795	02/08/2021	Africa / Sout Human					Sentinal Scre Male	29	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G,NSP4_R122L_N_R203M,NSP12_G6715,NS3_S26L,NS3						
13	HCoV-19/Sol EPI_ISL_3795	02/08/2021	Africa / Sout Human					Staff	Female	unknown	unknown	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NS3_S26L,Spik						
14	HCoV-19/Sol 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,Spik						
15	HCoV-19/Sol EPI_ISL_3795	06/08/2021	Africa / Sout Human					Sentinal Scre Female	45	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C,NSP4_S34F_N_D63G_N_R203M,NSP12_G6715,NS3_S26L,NSP12						
16	HCoV-19/Sol EPI_ISL_3795	02/08/2021	Africa / Sout Human					Sentinal Scre Female	46	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NS3_S26L,Spik						
17	HCoV-19/Sol EPI_ISL_3795	02/08/2021	Africa / Sout Human					Sentinal Scre Male	48	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NS3_S26L,Spik						
18	HCoV-19/Sol EPI_ISL_3795	05/08/2021	Africa / Sout Human					Sentinal Scre Female	32	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NS3_S26L,Spik						
19	HCoV-19/Sol EPI_ISL_3795	08/08/2021	Africa / Sout Human					Sentinal Scre Female	34	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NSP2_T223,NS3_S26L,Spik						
20	HCoV-19/Sol EPI_ISL_3795	02/08/2021	Africa / Sout Human					Sentinal Scre Female	44	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NSP8_T137,NSP5_G2835,NS						
21	HCoV-19/Sol EPI_ISL_3795	06/08/2021	Africa / Sout Human					Sentinal Scre Male	62	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G,NSP3_S284G_N_R203M,NSP12_G6715,NS3_S26L,Spik						
22	HCoV-19/Sol EPI_ISL_3795	04/08/2021	Africa / Sout Human					Sentinal Scre Female	51	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NS3_S26L,Spik						
23	HCoV-19/Sol EPI_ISL_3795	07/08/2021	Africa / Sout Human					Sentinal Scre Male	26	unknown	Original	Nasopharyngeal and oropl	B.1.617.2	GK	(NSP3_L1802F_N_G215C,E_P71LN_D63G_N_R203M,NSP12_G6715,NS3_A3						
24	HCoV-19/Sol EPI_ISL_3795	06/08/2021	Africa / Sout Human					Sentinal Scre unknown	30	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NSP2_A318V,Spik						
25	HCoV-19/Sol EPI_ISL_3795	02/08/2021	Africa / Sout Human					Sentinal Scre Female	97	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NS3_S26L,Spik						
26	HCoV-19/Sol EPI_ISL_3795	07/08/2021	Africa / Sout Human					Sentinal Scre Female	50	unknown	Original	Nasopharyngeal and oropl	AY.4	GK	(N_G215C_N_D63G_N_R203M,NSP12_G6715,NSP13_K271T,NS3_S26L,Spik						
27	HCoV-19/Sol EPI_ISL_3795	02/08/2021	Africa / Sout Human					Sentinal Scre Female	15	unknown	Original	Nasopharyngeal and oropl	AY.48	GK	(N_D63G_N_R203M,NSP12_G6715,NSP6_T181U,NS3_S26L,NSP12_G778C,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

Interacting with the GISAID database

Downloading selections of the data

Clade	all	Lineage	Substitutions	Variants	Reset	Fulltext ▲			
	Virus name	Passage date	Accession ID	Collection date	Submission	Length	Host	Location	Originating lab
	hCoV-19/USA/CO-CDPHE-2101733519	Original	EPI_ISL_5055855	2021-09-09	2021-10-12	29,007	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101729463	Original	EPI_ISL_5055854	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101731818	Original	EPI_ISL_5055853	2021-09-10	2021-10-12	29,481	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101730983	Original	EPI_ISL_5055852	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101730809	Original	EPI_ISL_5055851	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101734190	Original	EPI_ISL_5055850	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101730330	Original	EPI_ISL_5055849	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101736002	Original	EPI_ISL_5055848	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101736420	Original	EPI_ISL_5055847	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101729602	Original	EPI_ISL_5055846	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101731853	Original	EPI_ISL_5055845	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101729442	Original	EPI_ISL_5055844	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101729461	Original	EPI_ISL_5055843	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101729447	Original	EPI_ISL_5055842	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101729437	Original	EPI_ISL_5055841	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101729435	Original	EPI_ISL_5055840	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101729411	Original	EPI_ISL_5055839	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101729402	Original	EPI_ISL_5055838	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101729381	Original	EPI_ISL_5055837	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101729358	Original	EPI_ISL_5055836	2021-09-10	2021-10-12	29,769	Human	North America /	Colorado Depar
	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
	hCoV-19/USA/CO-CDPHE-2101729098	Original	EPI_ISL_5055833	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar
	hCoV-19/USA/CO-CDPHE-2101729084	Original	EPI_ISL_5055832	2021-09-09	2021-10-12	29,769	Human	North America /	Colorado Depar

Interacting with the GISAID database

Downloading selections of the data

Interacting with the GISAID database

Downloading selections of the data

Clade	Virus name	Passage date	Accession ID	Collection date	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

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 links for 'Registered Users', 'EpiFlu™', 'EpiCoV™' (which is highlighted in blue), and 'My profile'. Below the navigation bar is a search bar with fields for 'Accession ID', 'Virus name', 'Location', 'Collection', '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 search bar is a table displaying virus data. The table has columns for 'Virus name', 'Passage date', 'Accession ID', 'Collection da', 'Submission D', 'Length', 'Host', 'Location', and 'Originating'. Each row represents a specific virus entry. At the bottom of the table, it says 'Total: 4,248,757 viruses'. Below the table are navigation buttons for page numbers (1, 2, 3, 4, 5) and links for 'Select', 'Analysis', and 'Download'. A red circle highlights the 'Downloads' button in the top navigation bar.

	Virus name	Passage date	Accession ID	Collection da	Submission D	i	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 Hygie

Interacting with the GISAID database

Downloading global aggregated of the data

The screenshot shows the GISAID database interface for downloading global aggregated data. It includes sections for Alignment and proteins, Submission and Variant statistics, Download packages, and Genomic epidemiology.

- Alignment and proteins:** Includes MSA (full1008, unmasked1006, masked1006) and Archive files.
- Submission and Variant statistics:** Includes Global by month (xlsx), USA by month (xlsx), Variants VoC/Vol (xlsx), and Clade/Lineage, Variants (json).
- Download packages:** Includes FASTA, metadata, Variant surveillance, per clade, per lineage, and Dates + Locations.
- Genomic epidemiology:** Includes Audacity, Custom Selection, FASTA, metadata, and nextregions. The "Custom Selection" option is highlighted with a red oval.
- Region-specific Auspice source files:** Includes Global, Africa, Asia, Europe, North America, South America, and Oceania.

A large blue arrow points from the "Region-specific Auspice source files" section to the text below, indicating a relationship between the two.

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

