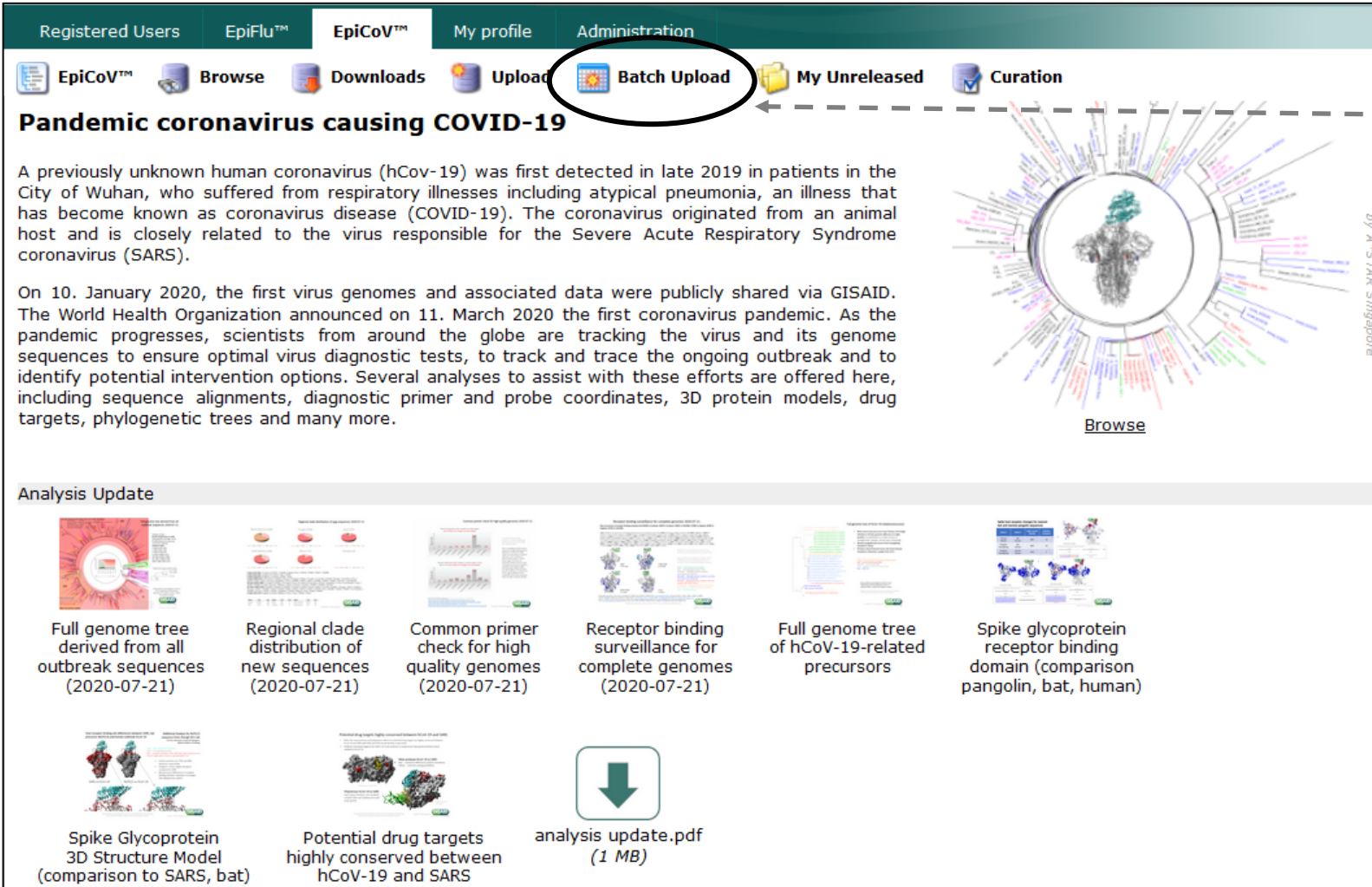


GISAID EpiCoV™ Data Batch Upload Instructions

1. Under the EpiCoV™ tab select Batch Upload



The screenshot shows the GISAID EpiCoV™ web interface. The top navigation bar includes tabs for 'Registered Users', 'EpiFlu™', 'EpiCoV™', 'My profile', and 'Administration'. Below this, a secondary bar contains icons for 'EpiCoV™', 'Browse', 'Downloads', 'Upload', 'Batch Upload' (circled in red), 'My Unreleased', and 'Curation'. A dashed line points from the 'Batch Upload' button to a callout box on the right. The main content area features a section titled 'Pandemic coronavirus causing COVID-19' with a paragraph of text and a circular phylogenetic tree. Below this is an 'Analysis Update' section with six thumbnails: 'Full genome tree derived from all outbreak sequences (2020-07-21)', 'Regional clade distribution of new sequences (2020-07-21)', 'Common primer check for high quality genomes (2020-07-21)', 'Receptor binding surveillance for complete genomes (2020-07-21)', 'Full genome tree of hCoV-19-related precursors', and 'Spike glycoprotein receptor binding domain (comparison pangolin, bat, human)'. At the bottom, there are three more items: 'Spike Glycoprotein 3D Structure Model (comparison to SARS, bat)', 'Potential drug targets highly conserved between hCoV-19 and SARS', and a download icon for 'analysis update.pdf (1 MB)'.

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.

[Browse](#)

Analysis Update

- Full genome tree derived from all outbreak sequences (2020-07-21)
- Regional clade distribution of new sequences (2020-07-21)
- Common primer check for high quality genomes (2020-07-21)
- Receptor binding surveillance for complete genomes (2020-07-21)
- Full genome tree of hCoV-19-related precursors
- Spike glycoprotein receptor binding domain (comparison pangolin, bat, human)

- Spike Glycoprotein 3D Structure Model (comparison to SARS, bat)
- Potential drug targets highly conserved between hCoV-19 and SARS
- analysis update.pdf (1 MB)

1. Batch Upload

In case you are unable to see the 'Batch Upload' button, please re-login or refresh

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



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”).

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*

Test.csv (35.1 kbytes)

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

Sequences as FASTA*

Test.fasta (1.20 MB)

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

Report

Excel-Sheet contains 44 submissions
Error: Sequence Virus name not found in FASTA
Error: Sequence Virus name not found in FASTA

Download Instructions and Template

Contact Curator Check and Submit

5. Error Report

4. Check and Submit



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

More details on GISAID EpiCoV CSV-Batch-Import:

- We also provide the Comma-Separated-Values (CSV) batch import option for our batch submitters. You can import a programmatically generated CSV-file or a text file saved as CSV-file, into the GISAID curation-space using the web-interface.
- A separate FASTA-File is needed for the sequences.
- While uploading the CSV-File it gets automatically converted into the Excel-Template-File. Use this Excel-File when downloading, editing and re-uploading the metadata again in case of any changes.

CSV-Format specification:

- File-Extension: .csv
- File-Encoding: UTF-8
- Line-Endings: CR/LF or LF
- Field-Separator: , (comma)
- String-Delimiter: " (double quotes)
- The first line of file is a header
- For comma separated list of GISAID-internal field-names (see example on next page)
- Same rules for mandatory fields as in the XLS template
- Maximum file size: 1000 submissions

Example of csv file:

```
submitter,fn,covv_virus_name,covv_type,covv_passage,covv_collection_date,covv_location,covv_add_location,covv_host,covv_add_host_info,covv_gender,covv_patient_age,covv_patient_status,covv_specimen,covv_outbreak,covv_last_vaccinated,covv_treatment,covv_seq_technology,covv_assembly_method,covv_coverage,covv_orig_lab,covv_orig_lab_addr,covv_provider_sample_id,covv_subm_lab,covv_subm_lab_addr,covv_subm_sample_id,covv_authors
lanman,20200629.gisaid.fa,hCoV-19/England/07/2020,betacoronavirus,Original,2020-03-24,United Kingdom / England,,Human,,unknown,unknown,unknown,unknown,unknown,unknown,unknown,OXFORD_NANOPORE,unknown,unknown,University, "Institute of Microbiology and Infection, University-Street, UK",BIRM-5E407,COVID-19 Genomics UK (COG-UK) Consortium,United Kingdom,LON-07, "Jane Doe, John Doe"
...
```

To use commas as part of your entries, please use quotes as shown above for “covv_orig_lab” and "covv_authors" columns (header is mandatory and should be in the same format).