



Advanced Molecular Detection Southeast Region Bioinformatics



SC2 Data Submissions, Part 3: GISAID & NCBI Submissions

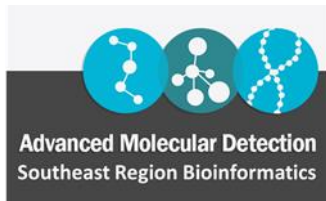
September 17, 2021

BPHL-SEbioinformatics@flhealth.gov

SARS-CoV-2 Data Submission Training Series

- Part 1: General Overview
- Part 2: Sample Review, Batch, and Multi-Fasta File Prep
- **Part 3: Submissions to GISAID and NCBI**
- Part 4: FASTQ de-host and SRA Submissions
- Part 5: Flagged Sample Review, Variant Confirmation, and Assembly Correction

Advanced Molecular Detection
Southeast Region Bioinformatics



Outline



Submit to GISAID



Submit to NCBI BioSample

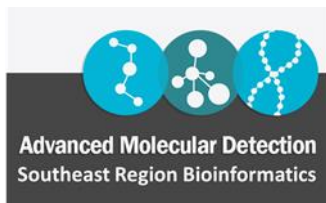
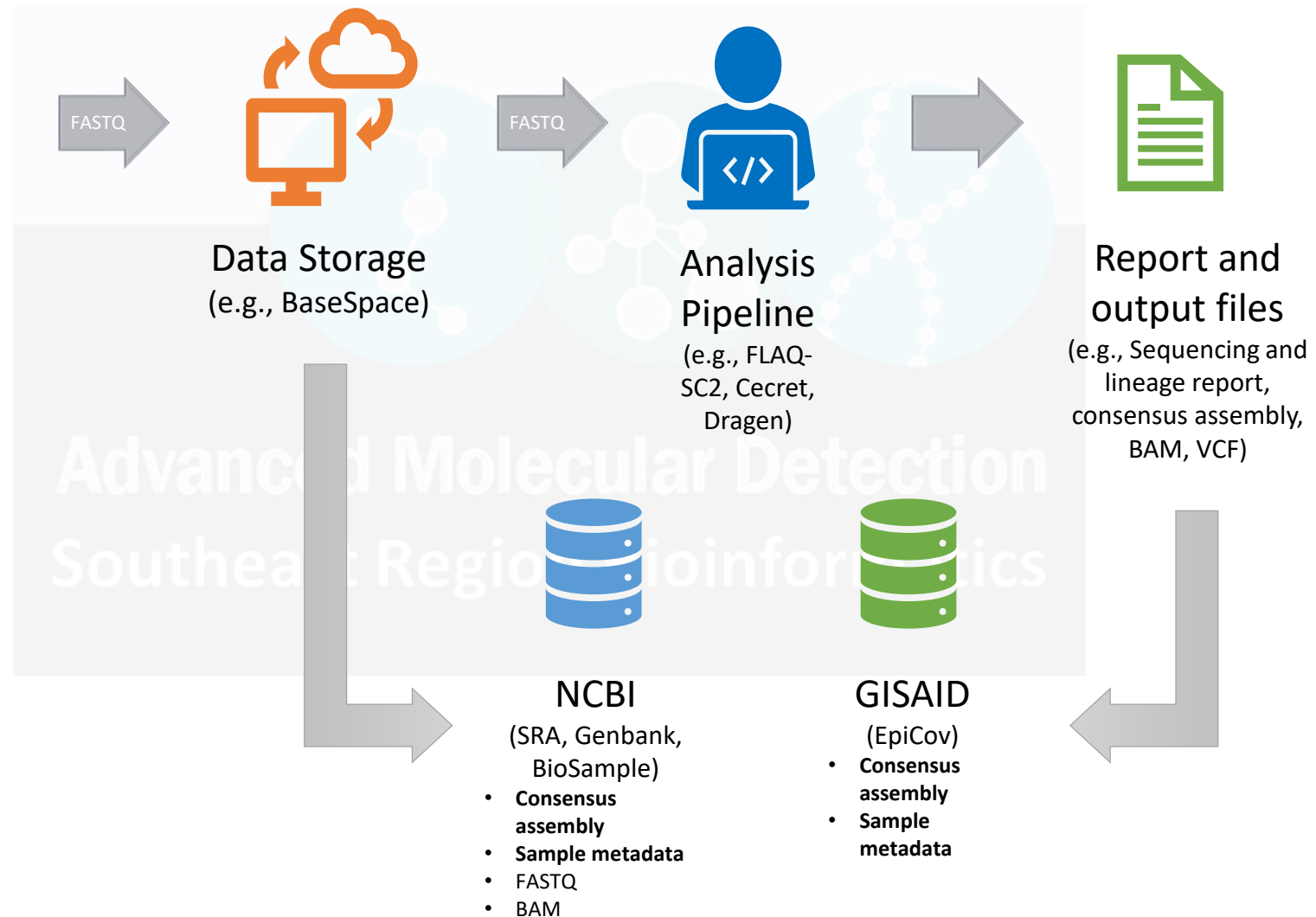


Submit to NCBI Genbank

SARS-CoV-2 Sequencing Workflow



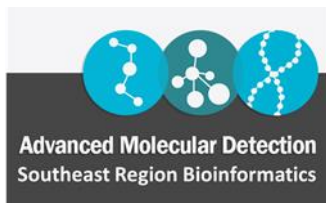
Tiled-Amplicon or
Enrichment-based
sequencing



SC2 Consensus Assembly Submissions to GISAID and NCBI

- Submission Process

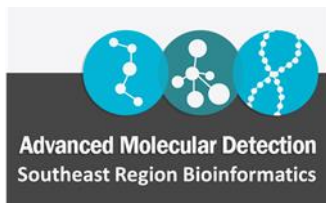
- ✓ Screen passing QC samples for submission (VADR – HiPerGator)
- ✓ Select samples for submission
- ✓ Collect relevant sample metadata needed for submission
- ✓ Assign public repository sample names
- ✓ Prepare formatted multi-fasta files for GISAID and Genbank (HiPerGator)
- ✓ Submit to GISAID – submit metadata template and multi-fasta file
- ✓ Retrieve GISAID accessions
- ✓ Submit to NCBI Biosample - submit metadata template (with linked GISAID accessions)
- ✓ Save NCBI Biosample accessions
- ✓ Submit to NCBI Genbank – submit metadata template (with linked GISAID and Biosample accessions) and multi-fasta file
- ✓ Save NCBI Genbank accessions
- ✓ **SUBMISSION COMPLETE and all data is now linked!!!**



Submit to GISAID (Demo)



- Files required for submission:
 - 1) Metadata template (Excel file)
 - 2) Formatted multi-fasta file
- Submission Steps
 - 1) Log in
 - 2) Select "Upload" Tab
 - 3) Select "Single upload" if only submitting a single sample. Select "Batch upload" if submitting more than one sample.
 - 4) Upload metadata template
 - 5) Upload multi-fasta file
 - 6) Select a "Confirmation Option"
 - 7) Select "Verify and Submit"
 - 8) Retrieve accessions after acceptance



[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).

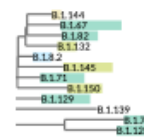
On 10. January 2020, the first virus genomes and associated data were publicly shared via GISAID. The World Health Organization announced on 11. March 2020 the first coronavirus pandemic. As the pandemic progresses, scientists from around the globe are tracking the virus and its genome sequences to ensure 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.

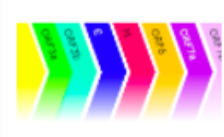


by A*STAR Singapore


[Audacity](#)

[AudacityInstant](#)

[BLAST](#)

[CoVizu^e](#)

[Emerging Variants](#)

[Official GISAID reference sequence](#)

[PrimerChecker](#)

[Spike glycoprotein mutation surveillance](#)

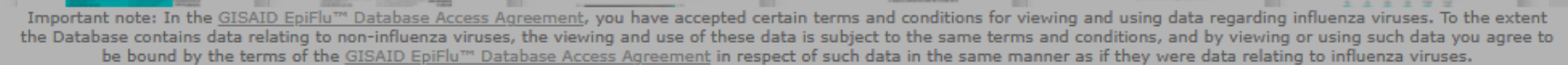
Analysis Update (2021-09-14)

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.



Advanced Molecular Detection
Southeast Region Bioinformatics

by A*STAR Singapore



[Registered Users](#)

[EpiFlu™](#)

[EpiCoV™](#)

[My profile](#)



EpiCoV™



Search



Downloads



Upload

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

Sequences as FASTA*

max size: 32M

Confirmation options

(Default) Notify me about ALL DETECTED FRAMESHIFTS in this submission for reconfirmation of affected sequences

Report

Upload XLS/CSV and FASTA.

Automated Uploading via CLI

Learn more about the Command Line Interface
a streamlined process that permits
uploading of higher volumes of data.

[Download to the CLI Guide](#)



OK



Download Instructions and Template



Contact Curator



Verify and Submit

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.



Advanced Molecular Detection
Southeast Region Bioinformatics

BPHL-SEbioinformatics@flhealth.gov

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


EpiCoV™



Search



Downloads



Upload

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

No file chosen

Sequences as FASTA*

max size: 32M

No file chosen

Confirmation options

(Default) Notify me about ALL DETECTED FRAMESHIFTS in this submission for reconfirmation of affected sequences



Report

Upload XLS/CSV and FASTA.

Check here for updated
metadata templates from
time to time



Download Instructions and Template



Contact Curator



Verify and Submit

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.



Advanced Molecular Detection
Southeast Region Bioinformatics

BPHL-SEbioinformatics@flhealth.gov



EpiCoV™



Search



Downloads



Upload

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*



[20210915_EpiCoV_BulkUpload_FL_BPHL_metadata.xls \(180 kbytes\)](#)

max size: 5M

[Choose File](#)

No file chosen

Sequences as FASTA*



[20210915_FL_BPHL_all_sequences_gisaid.fasta \(3.94 MB\)](#)

max size: 32M

[Choose File](#)

No file chosen

Confirmation options

I confirm ANY FRAMESHIFTS in this submission and request their release without reconfirmation by a Curator



Report

Upload XLS/CSV and FASTA.

Select option →



[Download Instructions and Template](#)



[Contact Curator](#)



[Verify and Submit](#)

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.



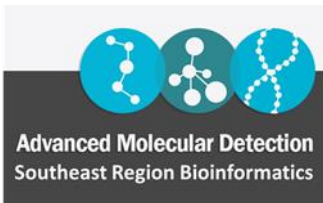
Advanced Molecular Detection
Southeast Region Bioinformatics

BPHL-SEbioinformatics@flhealth.gov

GISAID Confirmation Options - NEW

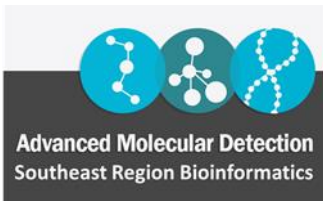
- Select options provides an improved, streamline process to confirming frameshift mutations from “flagged samples”.
 - Faster submission
 - Significantly fewer flagged samples
 - Reduces underrepresentation of “novel” variants when they first emerge
 - Decreases the number of emails back and forth with GISAID curators

Advanced Molecular Detection
Southeast Region Bioinformatics



GISAID Confirmation Options - NEW

- Confirmation Options
 - **(Default) Notify me about ALL DETECTED FRAMESHIFTS in this submission for reconfirmation of affected sequences**
 - Will reject all “flagged” samples
 - Requires resubmission of sequences with a confirmation email
 - **Notify me only about NOT PREVIOUSLY REPORTED FRAMESHIFTS in this submission for reconfirmation of affected sequences**
 - *This is FL-BPHL’s new default
 - Will accept all samples with frameshifts previously reported/confirmed by you or another submitter
 - Will only reject samples with not previously reported frameshift mutations
 - **I confirm ANY FRAMESHIFTS in this submission and request their release without reconfirmation by a Curator**
 - ONLY use this option if you personally have manually reviewed every sample in the submission and confirm every frameshift mutation present.
 - For previously “flagged samples” that are ready for submission



[Registered Users](#)

[EpiFlu™](#)

[EpiCoV™](#)

[My profile](#)



EpiCoV™



Search



Downloads



Upload

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*



[20210915 EpiCoV_BulkUpload_FL_BPHL_metadata.xls \(180 kbytes\)](#)

Sequences as FASTA*

Completed

The data has been successfully submitted. Once curation has successfully been completed, the data will be released and an email confirmation will be sent. Click "My Unreleased" for a list of all submitted data.

OK

max size: 32M

[Choose File](#)

No file chosen

Confirmation options

[Notify me only about NOT PREVIOUSLY REPORTED FRAMESHIFTS in this submission for reconfirmation of affected sequences](#) ▼

Report

[Upload XLS/CSV and FASTA.](#)



[Download Instructions and Template](#)



[Contact Curator](#)



[Verify and Submit](#)

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.

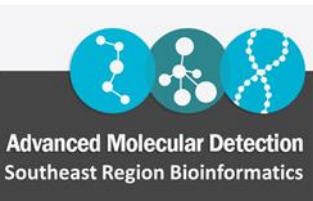


Advanced Molecular Detection
Southeast Region Bioinformatics

BPHL-SEbioinformatics@flhealth.gov

Retrieve GISAID Accessions (Demo)

- GISAID will send an email with the accepted accessions; however, they do not link them back to the sample name.
 - You have to manually download the accessions from <https://gisaid.org/>
- Select the “Search” tab
- Filter by state
- Filter by submission date
- Check all
- Download – “Patient Status Metadata”
- Open in Excel and sort by “Virus Name”



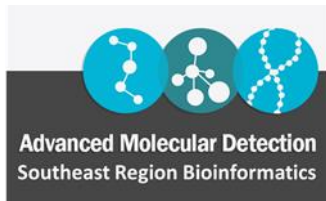
Rejected GISAID Samples

- If any of your samples get rejected/flagged by GISAID, either:
 - Make a record of the rejected samples and re-run `sc2_fasta_for_sub.py` on HiPerGator to re-generate your multi-fasta file for Genbank so your submissions match (i.e., edit your input name file to remove the samples you will no longer be submitting to Genbank at this time until further review) – Submissions need to match!
 - **OR**
 - Review your flagged samples and re-submit to GISAID prior to moving forward with NCBI BioSample/Genbank submissions
 - Make sure to re-run `sc2_fasta_for_sub.py` on HiPerGator if you had to correct any assemblies so your Genbank submissions also reflect these changes, if made.

Submit to NCBI BioSample



- Files required for submission:
 - 1) Metadata template (Excel file) – (Demo)
- Submission steps
 - 1) Go to <https://www.ncbi.nlm.nih.gov/biosample/>
 - 2) Select “Submit”
 - 3) Select “New Submission”
 - 4) Walk-through the submission wizard
 - 5) Attach metadata template
 - 6) Select “Submit”
 - 7) Accessions will be emailed to you in a .txt file **(NOTE: Make sure to sort by sample name in Excel)**



NCBI
Resources
How To
jaxmolecular1
My NCBI
Sign Out

BioSample
BioSample
Advanced
Search
Help

COVID-19 Information

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

BioSample

The BioSample database contains descriptions of biological source materials used in experimental assays.

Using BioSample

- [BioSample Overview](#)
- [BioSample Documentation](#)
- [Submission FAQ](#)
- [Search Help](#)
- [Submit](#)

Sources

- [GenBank](#)
- [SRA](#)
- [Coriell](#)
- [ATCC](#)
- [ICLAC](#)

Authenticated Cell Line

- [Background Search and Submit](#)
- [Browse Human Cell Line STR Profiles](#)
- [Browse Known Misidentified Cell Lines](#)

Example Searches

bacteria of genus <i>Shigella</i> for which SRA data is available	<code>shigella[organism] AND biosample sra[filter]</code>
MIGS/MIMS/MIMARKS.water-complaint samples released in first quarter of 2013	<code>package migs/mims/mimarks water[Properties] AND 2013/1:2013/3[Publication date]</code>
mouse samples for which strain and age information is available	<code>(strain[Attribute Name] AND age[Attribute Name]) AND Mus musculus[organism]</code>
fibroblast cell samples	<code>cell type fibroblast[Attribute]</code>

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

Submission Portal

Home **My submissions** Manage data Templates My profile

BioSample **New submission** Download batch submission template

Note: to update an existing record or recent submission, please email your request.

Short description and brief instructions



Filter / Search

From date YYYY-MM-DD To date YYYY-MM-DD Status Not deleted Sort by ☐ desc

Data archives

Query ?

Search **Clear**

← previous 1 2 3 4 5 6 7 8 9 ... 159 160 161 162 next →

4,038 submissions				
Submission	Title	Group	Status	Updated
SUB10353696	SARS-CoV-2: clinical or host-associated sample		✓ BioSample: Processed Successfully loaded 342 objects. Download attributes file with BioSample accessions	Sep 09
SUB10307142	SARS-CoV-2: clinical or host-associated sample		✓ BioSample: Processed Successfully loaded	Sep 02

gisaid_hcov-19_202....tsv

Show all



Advanced Molecular Detection
Southeast Region Bioinformatics



COVID-19 Information

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)



Submission Portal

[Home](#) [My submissions](#) [Manage data](#) [Templates](#) [My profile](#)

BioSample submission: SUB10388605

[Delete submission](#)

New

1 SUBMITTER

2 GENERAL INFO

3 SAMPLE TYPE

4 ATTRIBUTES

5 DESCRIPTION

6 REVIEW & SUBMIT

Submitter

[i](#) Required fields are marked with * asterisk

* First (given) name Middle name * Last (family) name

FLDOH

FLDOH

* Email (primary)

DLBPHL16WGSJax@flhealth.gov

Email (secondary)

bphl16BioInformatics@flhealth.gov

[i](#) At least one email should be from the organization's domain.

Group for this submission

☒ No group (affiliation from my personal profile)

☐ 109 members PulseNet Next Generation Subtyping Methods Unit



Advanced Molecular Detection
Southeast Region Bioinformatics

BPHL-SEbioinformatics@flhealth.gov

- Fill in Submitter Info (should be auto-filled after the first time you login)
- Click “Continue”

BioSample submission: SUB10388605

[Delete submission](#)

New

1 SUBMITTER 2 GENERAL INFO 3 SAMPLE TYPE 4 ATTRIBUTES 5 DESCRIPTION 6 REVIEW & SUBMIT

General Information

Required fields are marked with * asterisk

Release date

Note: Release of BioProject or BioSample is also triggered by the release of linked data.

* When should this submission be released to the public?

- ☒ Release immediately following processing
- ☐ Release on specified date or upon publication, whichever is first

* Specify if you are submitting a single sample or a file containing multiple samples

- ☒ Batch/Multiple BioSamples
You will be asked to upload a tab-delimited text file that describes each of your samples and their attributes. Submission template files can be downloaded from the Attributes tab or the [templates page](#).
- ☐ Single BioSample
You will be asked to manually complete a web form to describe one sample and its attributes.

Continue



Advanced Molecular Detection
Southeast Region Bioinformatics

BioSample submission: SUB10388605

New

1 SUBMITTER

2 GENERAL INFO

3 SAMPLE TYPE

4 ATTRIBUTES

5 REVIEW & SUBMIT

Sample Type

★ Select the package that best describes your samples.

All packages Packages for MAG submitters Packages for metagenome submitters

(Optional) Filter packages by organism name

Enter the full scientific name of your samples, e.g., Escherichia coli

Reset and show all packages

To filter for relevant BioSample packages, enter the **full scientific name** of the organism of your samples.

- If your BioSamples are derived from a species **not represented in NCBI's Taxonomy database**, enter the genus-level name, e.g., *Escherichia*
- If your BioSamples are derived from **more than one organism**, enter the common species, genus, or family, e.g., *Enterobacteriaceae*
- If your BioSamples are **metagenomic/environmental**, or **metagenome-assembled genomes (MAG)**, select the appropriate tab above
- For more information about organism names, see [Organism information](#).

NCBI packages [More...](#)

☐ SARS-CoV-2: clinical or host-associated

Use for SARS-CoV-2 samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of SARS-CoV-2 cases.

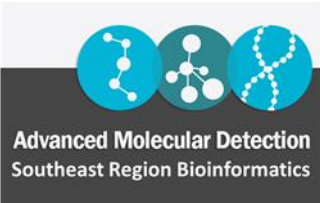
GSC [MIS](#) packages for genomes, metagenomes, and marker sequences [More...](#)

☐ MIGS Cultured Bacterial/Archaeal

Use for cultured bacterial or archaeal genomic sequences. Organism must have lineage [Bacteria](#) or [Archaea](#).

NCBI packages [More...](#)

- ☐ SARS-CoV-2: clinical or host-associated
- Use for SARS-CoV-2 samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of SARS-CoV-2 cases.
- ☐ SARS-CoV-2: wastewater surveillance
- Use for SARS-CoV-2 wastewater surveillance samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of SARS-CoV-2 cases.
- ☐ Pathogen
- Use for pathogen samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of pathogens.
- ☐ Microbe
- Use for bacteria or other unicellular microbes when it is not appropriate or advantageous to use [MIS](#), Pathogen or Virus packages.



BioSample submission: SUB10388605

New

1 SUBMITTER 2 GENERAL INFO 3 SAMPLE TYPE 4 ATTRIBUTES 5 REVIEW & SUBMIT

Attributes

* How do you want to provide your BioSample attributes?

- ☐ Use built-in table editor
- ☐ Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples

Continue



Advanced Molecular Detection
Southeast Region Bioinformatics

Submission Portal

BioSample submission: SUB10388605

New

1 SUBMITTER

2 GENERAL INFO

3 SAMPLE TYPE

4 ATTRIBUTES

5 REVIEW & SUBMIT

Attributes


★ How do you want to provide your BioSample attributes?

☐ Use built-in table editor

☒ Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples

 Choose file

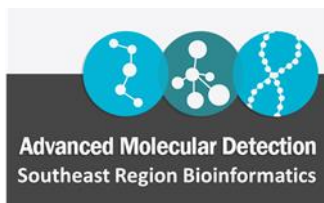
or drag and drop it here

 Template for BioSample package **SARS-CoV-2: clinical or host-associated; version 1.0**

[Download Excel](#) [Download TSV](#)

For column explanations and examples, please see the [sample attributes page](#).

For more information, please see [creating sample attribute file](#).



Submission Portal

BioSample submission: SUB10388605

New

1 SUBMITTER 2 GENERAL INFO 3 SAMPLE TYPE 4 ATTRIBUTES 5 REVIEW & SUBMIT

Attributes

★ How do you want to provide your BioSample attributes?

☐ Use built-in table editor

☒ Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples

20210916_NCBI_SC2_BioSample_Submission_FL_BPHL_137batch.xlsx 41.1 kB 2021-09-16 08:18

Delete

• Template for BioSample package SARS-CoV-2: clinical or host-associated, version 1.0

[Download Excel](#) [Download TSV](#)

For column explanations and examples, please see the [sample attributes page](#).

For more information, please see [creating sample attribute file](#).

Continue



Advanced Molecular Detection
Southeast Region Bioinformatics

Submission Portal

BioSample submission: SUB10388605

SARS-CoV-2: clinical or host-associated sample

1 SUBMITTER 2 GENERAL INFO 3 SAMPLE TYPE 4 ATTRIBUTES 5 REVIEW & SUBMIT

Review & Submit

This BioSample submission will be released **immediately following processing.**

Submitter Information

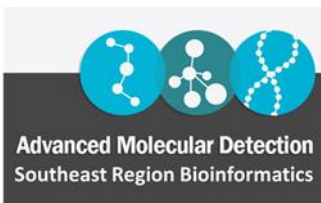
Submitter FLDOH FLDOH
DLBPHL16WGSJax@flhealth.gov

General Information

Package SARS-CoV-2: clinical or host-associated; version 1.0

Attribute file 20210916_NCBI_SC2_BioSample_Submission_FL_BPHL_137batch_vley (11.1 KB)

Submit



Submission Portal

BioSample [New submission](#) [Download batch submission template](#)

Note: to update an existing record or recent submission, please email your request.

Short description and brief instructions +

Filter / Search

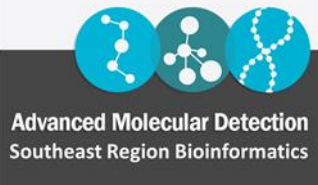
From date YYYY-MM-DD To date YYYY-MM-DD Status Not deleted Sort by desc

Data archives +

Query Search Clear

4,039 submissions

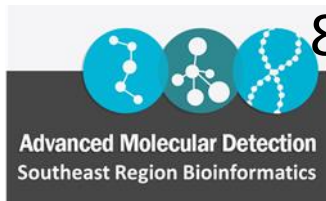
Submission	Title	Group	Status	Updated
SUB10388605	SARS-CoV-2: clinical or host-associated sample		Submitted Awaiting processing.	08:19
SUB10353696	SARS-CoV-2: clinical or host-associated sample		BioSample: Processed Successfully loaded 342 objects.	Sep 09



Submit to NCBI GenBank

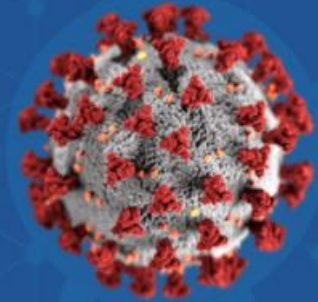


- Files required for submission:
 - 1) Metadata template (Excel file) – (Demo)
 - 2) Formatted multi-fasta file
- Submission steps
 - 1) Go to <https://submit.ncbi.nlm.nih.gov/sarscov2/>
 - 2) Select GenBank “Submit”
 - 3) Select “New Submission”
 - 4) Walk-through the submission wizard
 - 5) Attach multi-fasta file
 - 6) Attach metadata template
 - 7) Select “Submit” at end
 - 8) Accessions will be emailed to you in a .txt file **(NOTE: Make sure to sort by sample name in Excel – many times they are not in order)**



Submit SARS-CoV-2 sequences

Add your SARS-CoV-2 sequence data to the growing public archive



Easily submit assembled & raw read SARS-CoV-2 data on the web or via XML upload for COVID-19 response.
NCBI is here to help.

GenBank

i Submitted 2021-09-09

Submit assembled reads of SARS-CoV-2 with FASTA files and source metadata. Annotation for SARS-CoV-2 is not required.

Accessions in 2 hours (avg)

[Learn more](#)

[Submit](#)

Sequence Read Archive (SRA)

i Submitted 2020-03-04

Submit unassembled reads of SARS-CoV-2 with BioProject, BioSample, metadata and NGS files.

Accessions in 2 hours (avg)

[Learn more](#)

[Submit](#)



Advanced Molecular Detection
Southeast Region Bioinformatics

GenBank

[New submission](#)

Note: Submit only **ribosomal RNA (rRNA)**, **rRNA-ITS**, **metazoan COX1**, **Influenza**, **Norovirus**, **Dengue** or **SARS-CoV-2** sequences here.
All other submission types should use one of the alternate [submission tools](#) (e.g. [BankIt](#), [tbl2asn](#), etc.)

Attention: If you have corrections to an existing submission with status:

- Processed-error: use the FIX button to correct a submission.
- Queued or Processing: [email your request](#) with the Submission ID.
- Processed: [follow these directions](#) and [email your request](#) with the Accession numbers.

Do not create a new submission to fix or update an existing submission whose status is Queued, Processed-error, Processing, or Processed!

Filter / Search

From date To date Status Sort by

YYYY-MM-DD YYYY-MM-DD Not deleted ☐ desc

Data archives +

Query ?

Search

Clear

Short description and brief instructions

+

56 submissions

← previous 1 2 3 next →

Submission	Title	Group	Status	Updated
SUB10354468	SARS-CoV-2		✓ GenBank: Processed OK077177-OK077518 3 files: <ul style="list-style-type: none"> • AccessionReport.tsv • flatfile.txt • email.txt 	Sep 09
SUB10308020	SARS-CoV-2		✓ GenBank: Processed OK018350-OK019012	Sep 02

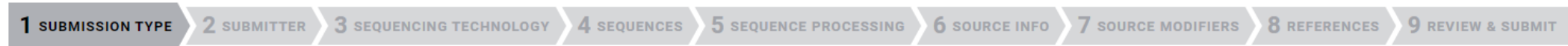


Advanced Molecular Detection
Southeast Region Bioinformatics

GenBank submission: SUB10388662

[Delete submission](#)

New



Submission Type

i Required fields are marked with ***** asterisk

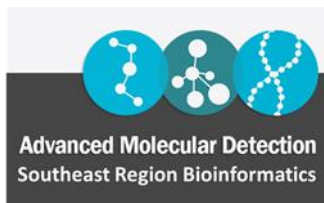
***** What do your sequences contain?

- ☐ rRNA or rRNA-ITS **i**
- ☐ COX1 from metazoan mitochondria **i**
- ☐ SARS-CoV-2, Influenza, Norovirus, or Dengue virus

i If none of the options above describe your sequences, use [Bankit](#) to submit.

Submission title (Optional, not displayed in final records) **i**

Continue



GenBank submission: SUB10388662

Delete submission

New

1 SUBMISSION TYPE 2 SUBMITTER 3 SEQUENCING TECHNOLOGY 4 SEQUENCES 5 SEQUENCE PROCESSING 6 SOURCE INFO 7 SOURCE MODIFIERS 8 REFERENCES 9 REVIEW & SUBMIT

Submission Type

Required fields are marked with * asterisk

* What do your sequences contain?

- ☐ rRNA or rRNA-ITS ?
- ☐ COX1 from metazoan mitochondria ?
- ☒ SARS-CoV-2, Influenza, Norovirus, or Dengue virus

* Which virus?

- ☐ SARS-CoV-2
- ☐ Influenza virus
- ☐ Norovirus
- ☐ Dengue virus

If none of the options above describe your sequences, use [Bankit](#) to submit.

Submission title (Optional, not displayed in final records) ?

Continue

Advanced Molecular Detection
Southeast Region Bioinformatics

GenBank submission: SUB10388662

SARS-CoV-2

[Delete submission](#)**1 SUBMISSION TYPE** **2 SUBMITTER** 3 SEQUENCING TECHNOLOGY 4 SEQUENCES 5 SEQUENCE PROCESSING 6 SOURCE INFO 7 SOURCE MODIFIERS 8 REFERENCES 9 REVIEW & SUBMIT**Submitter**

i Required fields are marked with * asterisk

Affiliation

- i The information you give here will be displayed in the final sequence records.
For address details, provide the primary address where work was done to generate the data in this submission.

Group for this submission

- ☒ 0 members No group
☐ 109 members PulseNet Next Generation Subtyping Methods Unit

[Create group](#)

i Allow selected collaborators to read, modify, submit and delete your submissions

*** Submitting organization**

FLDOH

*** Department**

BPHL

*** Street**

1217 N Pearl St

*** City**

Jacksonville

*** State/Province**

FL

*** Postal code**

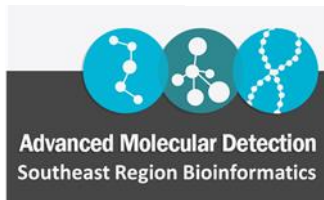
32202

*** Country**

USA

Contact information

- i GenBank may use this information to contact you about your submission, it will not be displayed in the final sequence records.



- Fill in Submitter Info (should be auto-filled after the first time you login)
- Click “Continue”

GenBank submission: SUB10388662

SARS-CoV-2

1 SUBMISSION TYPE

2 SUBMITTER

3 SEQUENCING TECHNOLOGY

4 SEQUENCES

5 SEQUENCE PROCESSING

6 SOURCE INFO

7 SOURCE MODIFIERS

8 REFERENCES

9 REVIEW & SUBMIT

Sequencing Technology

[i](#) Required

Method

★ What methods were used to obtain these sequences? [?](#)

- ☐ Sanger dideoxy sequencing
- ☐ 454
- ☐ Helicos
- ☐ Illumina
- ☐ IonTorrent
- ☐ Pacific Biosciences
- ☐ SOLiD
- ☐ Other

Assembly state

These sequences are:

- ☐ Unassembled sequence reads
- ☐ Assembled sequences (each sequence was assembled from two or more overlapping sequence reads)

Assembly state

These sequences are:

- ☐ Unassembled sequence reads
- ☒ Assembled sequences (each sequence was assembled from two or more overlapping sequence reads)

Assembly Information

★ Assembly program ?

★ Version or date ?

Delete

bwa

0.7.17



ivar

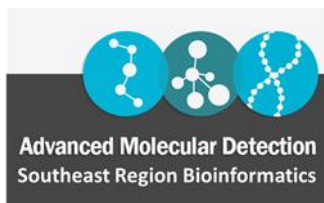
1.3.1



+ Add another assembly program

**Put the name of the pipeline
you used or the actual
assembly program**

Continue



Sequences

Release date

Note: Release of BioProject or BioSample is also triggered by the release of linked data.

* When should this submission be released to the public?

- ☒ Release immediately following processing
- ☐ Release on specified date or upon publication, whichever is first

Sequences

* Upload a nucleotide **FASTA** formatted file.

[Choose file](#) or drag and drop it here

? If you have multiple sequences, all of your sequences need to be in one file. [Help on FASTA file.](#)

Example FASTA nucleotide format:

```
>Seq1
aaccgatatagagatagtgatccgatatagagagagga
>Seq2
gtacgataaagagatagtgatccgatatagagagagga
```

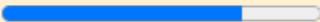
i Use the latest version of the [Aspera Connect plugin](#) for faster file uploads. If a pop-up box about '**fasp protocol**' is displayed, click 'Allow' or 'Open' to let [Aspera Connect](#) handle file uploads more efficiently.

Continue

GenBank submission: SUB10388662

SARS-CoV-2



Please wait! Processing the data.  (PROGRESS: 6 out of 8) Checking for runs of ambiguous bases

1 SUBMISSION TYPE > 2 SUBMITTER > 3 SEQUENCING TECHNOLOGY > **4 SEQUENCES** ⌚ > 5 SEQUENCE PROCESSING > 6 SOURCE INFO > 7 SOURCE MATERIAL

Sequences

Release date

Note: Release of BioProject or BioSample is also triggered by the release of linked data.

★ When should this submission be released to the public?

- ☒ Release immediately following processing
- ☐ Release on specified date or upon publication, whichever is first



Advanced Molecular Detection
Southeast Region Bioinformatics

Sequences

Required fields are



Warning: The following sequence(s) were trimmed of ambiguous bases:

Sequence_ID

SARS-CoV-2/human/USA/FL-BPHL-6688/2021
SARS-CoV-2/human/USA/FL-BPHL-6625/2021
SARS-CoV-2/human/USA/FL-BPHL-6627/2021
SARS-CoV-2/human/USA/FL-BPHL-6628/2021
SARS-CoV-2/human/USA/FL-BPHL-6631/2021

... and more (Complete table can be found [here](#)).

**This warning means Ns
were trimmed off the
end of the genome**



Warning: Found one or more string of NNN's (length > 10):

Sequence-IDs

SARS-CoV-2/human/USA/FL-BPHL-6624/2021
SARS-CoV-2/human/USA/FL-BPHL-6625/2021
SARS-CoV-2/human/USA/FL-BPHL-6626/2021
SARS-CoV-2/human/USA/FL-BPHL-6627/2021
SARS-CoV-2/human/USA/FL-BPHL-6629/2021

... and more (Complete table can be found [here](#)).

**This warning means Ns
were detected in the
genome.**

**No action is required.
Click Continue.**



Advanced Molecular Detection
Southeast Region Bioinformatics

Only select the second option if you know for a fact that your assembly program uses a single N to represent a string of Ns of any length. Most programs do not do this.

What do the internal NNN's represent?

- i** The nucleotide sequence(s) in your file contain strings of internal NNN's (length > 10). Please answer the question below and click Continue at the bottom of the page.

★ Please explain what the strings of internal NNNs represent

- ☒ A region of estimated length between the sequenced regions based on an alignment to similar sequences or genome
- ☐ A region of unknown length between the sequenced regions

Release date

- i** **Note:** Release of BioProject or BioSample is also triggered by the release of linked data.

★ When should this submission be released to the public?

- ☒ Release immediately following processing
- ☐ Release on specified date or upon publication, whichever is first

Sequences

★ Upload a nucleotide **FASTA** formatted file. To upload a new file, you must delete your previous file.

SUB10388662_20210915_FL_BPHL_all_sequences_genbank_fasta_filtered.fsa 3.9 MB 2021-09-16 08:30

Delete

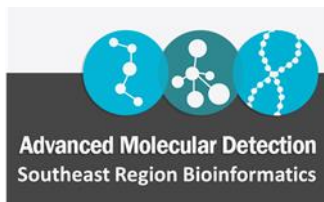
- ?** If you have multiple sequences, all of your sequences need to be in one file. [Help on FASTA file.](#)

Example FASTA nucleotide format:

```
>Seq1
aaccgatatagagatagtgatccgatatagagagagga
>Seq2
gtacgataaagagatagtgatccgatatagagagagga
```

- i** Use the latest version of the [Aspera Connect plugin](#) for faster file uploads.

Click Continue again.



GenBank submission: SUB10388662[Delete submission](#)

SARS-CoV-2

1 SUBMISSION TYPE **2 SUBMITTER** **3 SEQUENCING TECHNOLOGY** **4 SEQUENCES** **5 SEQUENCE PROCESSING** 6 SOURCE INFO 7 SOURCE MODIFIERS 8 REFERENCES 9 REVIEW & SUBMIT**Sequence Processing**

i Required fields are marked with * asterisk

Option to automatically remove failed sequences

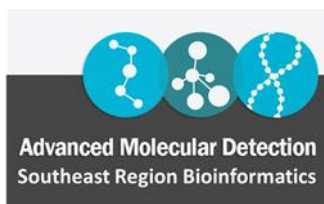
i If errors are found on sequences during **processing**, they will be removed from this submission and the successful sequences accessioned. You will receive a detailed report on these errors.

* During processing, should NCBI remove sequences with errors and process the rest?

☐ Yes☐ No[Continue](#)

- Choose “Yes” if you want NCBI to accept all passing samples and reject “flagged samples” for manual review. This is the RECOMMENDED option if you are not yet trained to manually review samples or do not have the time to review in real-time. Flagged samples should be ZERO or very few if you pre-screened with VADR.

- Choose “NO” if you want NCBI to hold your entire submission until you address flagged samples. Choose this option if you have already manually reviewed flagged samples and are re-submitting (or plan to do the review in real-time). This requires a confirmation email to Genbank curators to confirm each frameshift/stop codon mutation.

**BPHL-SEbioinformatics@flhealth.gov**

GenBank submission: SUB10388662

SARS-CoV-2

Delete submission



Source Information

Required fields are marked with * asterisk

The first few sequence IDs that we found are:

SARS-CoV-2/human/USA/FL-BPHL-6688/2021
SARS-CoV-2/human/USA/FL-BPHL-6624/2021
SARS-CoV-2/human/USA/FL-BPHL-6625/2021
SARS-CoV-2/human/USA/FL-BPHL-6626/2021
SARS-CoV-2/human/USA/FL-BPHL-6627/2021
...

★ Do your sequence IDs represent one of these?

Isolate

NONE of these

Values for these are typically alpha-numeric sample codes used in your laboratory to track individual samples. Select 'NONE of these' if it does not describe your sequence IDs or the sequence IDs contain more information than the descriptions of these fields.

The sequence IDs will be used as the in the final sequence records.

Continue

Advanced Molecular Detection
Southeast Region Bioinformatics

BPHL-SEbioinformatics@flhealth.gov

Select “Isolate” if your sample name is the same as your Isolate name

41

GenBank submission: SUB10388662

SARS-CoV-2

- 1 SUBMISSION TYPE
- 2 SUBMITTER
- 3 SEQUENCING TECHNOLOGY
- 4 SEQUENCES
- 5 SEQUENCE PROCESSING
- 6 SOURCE MODIFIERS

Source Modifiers

For each sequence, GenBank requires the following source information:

- collection-date,
- country,
- host, and
- isolate.

Current source modifiers - what you have provided so far (only first 50 rows are displayed)

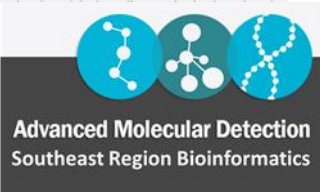
More help: what is a source modifier, description of each modifier, how to provide source modifiers.

If you have already provided all the required information, you can press Continue to proceed.

How do you want to apply source modifiers?

- ☐ Use a form to apply the same value for all sequences
- ☐ Use an editable table
- ☒ Upload a tab-delimited table (template file provided)

Continue



Current source modifiers - what you have provided so far (only first 50 rows are displayed)

More help: what is a source modifier, description of each modifier, how to provide source modifiers.

If you have already provided all the required information, you can press Continue to proceed.

How do you want to apply source modifiers?

Apply source modifiers by uploading a tab-delimited table

1. Download source modifier template table.
2. Edit the downloaded table in Microsoft Excel or another editor.

See an example Source Modifiers table

3. Save the table as a tab-delimited text file.
4. Upload your saved table file.
- Choose file

 or drag and drop it here
5. Click Continue to validate the information and follow the instructions.

Continue

GenBank submission: SUB10388662

Delete submission

SARS-CoV-2

1 SUBMISSION TYPE 2 SUBMITTER 3 SEQUENCING TECHNOLOGY 4

References

Sequence authors

Who should be publicly credited as the submitter of this sequence data? Enter

Sequence authors from your recent submissions (Optional)

Apply sequ

* First (given) name MI ? * Last (family) name Delete



+ Add another sequence author

Sequence authors from your recent submissions (Optional)

Schmedes,S., Li,J., Tarigopula,N. and Blanton,J.

Apply sequence authors

* First (given) name MI ? * Last (family) name Delete



Sarah

Schmedes



Jiaqi

Li



Namratha

Tarigopula



Jason

Blanton



+ Add another sequence author

Names will appear in your records as:

Schmedes,S., Li,J., Tarigopula,N. and Blanton,J.

Reference

* Publication status

☒ Unpublished ☐ In-progress ☐ Published

Continue



Advanced Molecular Detection
Southeast Region Bioinformatics

BPHL-SEbioinformatics@flhealth.gov

Reference

* Publication status

☒ Unpublished ☐ In-progress ☐ Published

Reference title

Reference authors

☐ Same as sequence authors ☐ Specify authors

Continue

GenBank submission: SUB10388662

SARS-CoV-2

[Delete submission](#)

1 SUBMISSION TYPE 2 SUBMITTER 3 SEQUENCING TECHNOLOGY 4 SEQUENCES 5 SEQUENCE PROCESSING 6 SOURCE INFO 7 SOURCE MODIFIERS 8 REFERENCES 9 REVIEW & SUBMIT

Review & Submit

To proceed please review your submission, make any necessary changes using the tabs/steps above, then click on the Submit button below.

You have requested that your sequence data be released **immediately following processing**.

Submitter

Submitter FLDOH FLDOH
DLBPHL16WGSJax@flhealth.gov
bphl16Bioinformatics@flhealth.gov

Institution FLDOH

[Submit](#)

GenBank Record Preview

Why is some information missing/different in this GenBank record preview? +

LOCUS SARS-CoV-2/human/USA/FL-BPHL-6688/2021 29819 bp DNA linear
VRL 16-SEP-2021
DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate
SARS-CoV-2/Human/USA/FL-BPHL-6688/2021.



Advanced Molecular Detection
Southeast Region Bioinformatics

BPHL-SEbioinformatics@flhealth.gov

Note: Submit only ribosomal RNA (rRNA), rRNA-ITS, metazoan COX1, Influenza, Norovirus, Dengue or SARS-CoV-2 sequences here.
All other submission types should use one of the alternate submission tools (e.g. BankIt, tbl2asn, etc.)

Attention: If you have corrections to an existing submission with status:

- Processed-error: use the FIX button to correct a submission.
- Queued or Processing: email your request with the Submission ID.
- Processed: follow these directions and email your request with the Accession numbers.

Do not create a new submission to fix or update an existing submission whose status is Queued, Processed-error, Processing, or Processed!

Filter / Search

From date YYYY-MM-DD To date YYYY-MM-DD Status Not deleted Sort by desc

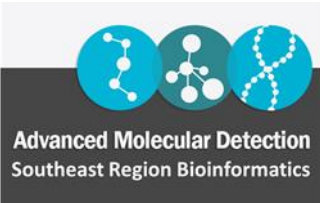
Data archives +

Query Search Clear

Short description and brief instructions +

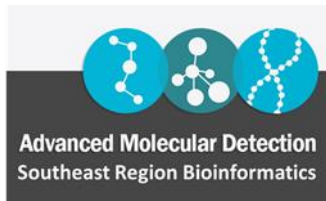
57 submissions

Submission	Title	Group	Status	Updated
SUB10388662	SARS-CoV-2		<div>Submitted</div> <div>Awaiting processing.</div>	08:39
SUB10354468	SARS-CoV-2		<div>GenBank: Processed</div> <div>OK077177-OK077518</div> <div>3 files:</div>	Sep 09



SC2 Consensus Assembly Recap

- You can prepare and submit up to 1,000 genomes for submission to Genbank at once.
 - GISAID requirement is based on file size, but it is ~1,000 genomes as well.
- Batch by sequencing technology (e.g., Illumina, ONT) and assembly type (i.e., assembly program/pipeline) (for Genbank)
- You can leave your scripts in a central “sequence_submissions” directory on HiPerGator
 - No need to have to re-enter your email each time in your sbatch script – just need to edit options that change (e.g., input name file).
- BRR will update scripts when there is an update to VADR.

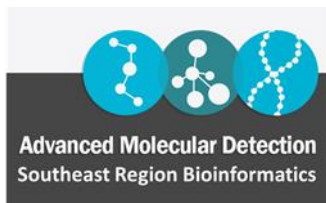


SC2 Consensus Assembly Recap

- When there is an update to VADR, go back and re-run your previously “flagged”/REVIEW samples
 - The last update significantly decreased the number of samples needing review and their new status is now “PASS”.
- If you need to correct or update a submission/sample(s) (e.g., update FASTA, metadata), send an email and the affected accession(s) to:
 - GISAID - hcov-19@gisaid.org
 - NCBI BioSample - biosamplehelp@ncbi.nlm.nih.gov
 - NCBI GenBank - gb-admin@ncbi.nlm.nih.gov
 - Please update any past accessions with “purpose of sequencing” tagging using this method.

Next Trainings

- Follow-up calls with each jurisdiction for hands-on submission walk-throughs, if requested
- **TBD** – SARS-CoV-2 Data Submissions, Part 4: FASTQ de-host and SRA Submissions
- **TBD** – SARS-CoV-2 Data Submissions, Part 5: Flagged Sample Review, Variant Confirmation, and Assembly Correction
- The recording from each training, slides, and associated training materials will be available at <https://github.com/StaPH-B/southeast-region>.
- If you are a new HiPerGator user, please contact sarah.schmedes@flhealth.gov to schedule your one-on-one trainings (e.g., Linux 101, compute environment setup, BS CLI)!!!





Advanced Molecular Detection Southeast Region Bioinformatics

Questions???

BPHL-SEbioinformatics@flhealth.gov

Sarah Schmedes, PhD

Lead Bioinformatician

BRR/WFD Lead, Southeast Region

Jiaqi Li, PhD

Bioinformatician

Jason Blanton, PhD

Molecular Administrator

State Sequencing Coordinator