



Version 2

Apr 09, 2021

SARS-CoV-2 NCBI consensus submission protocol: GenBank V.2

In 1 collection

Ruth E Timme¹, Heather M Blankenship², Pavitra Roychoudhury³, Erin L Young⁴, Emma Griffiths⁵, Duncan MacCannell⁶, Stacia Wyman⁷, Lee Katz⁸

¹US Food and Drug Administration; ²Michigan Department of Health and Human Services; ³University of Washington;

⁴Utah Public Health Laboratory; ⁵University of British Columbia; ⁶Centers for Disease Control and Prevention;

⁷Innovative Genomics Institute, UC Berkeley; ⁸CDC

1 Works for me dx.doi.org/10.17504/protocols.io.bid7ka9n

GenomeTrakr StaPH-B 1 more workspace

Ruth Timme
US Food and Drug Administration

ABSTRACT

PURPOSE:

This is a SARS-CoV-2 specific protocol that covers the steps needed to establish a new NCBI submission environment for your laboratory, including the creation of new BioProject(s) and submission groups. Once these are set up, the protocol then walks through the process for submitting raw reads to SRA and sample metadata to BioSample through the Submission portal.

For new submitters, there's quite a bit of groundwork that needs to be established before a laboratory can start its first data submission. We recommend that one person in the laboratory take a few days to get everything set up in advance of when you expect to do your first data submission.

If you need a pipeline for frequent or large volume submissions, follow Step 1 in the [SARS-CoV-2 NCBI submission protocol: SRA, BioSample, and BioProject](#) to get your NCBI submission environment established, then contact gb-admin@ncbi.nlm.nih.gov to set up an account for submitting through the API.

These protocols cover submission using NCBI's Submission Portal web-interface.

Complete in order (1 then 2):

1. [SARS-CoV-2 NCBI submission protocol: SRA, BioSample, and BioProject](#)

- Step-by-step instructions for establishing a new NCBI laboratory submission account and for creating and linking a new BioProject to an existing umbrella effort.
- SARS-CoV-2 raw data submission to SRA (Sequence Read Archive) and metadata to BioSample.

2. [SARS-CoV-2 NCBI consensus submission protocol: GenBank \(included protocol\)](#)

Required: established BioProject and BioSamples

- Submit SARS-CoV-2 assemblies to NCBI GenBank, linking to existing BioProject, BioSamples, and raw data.

Metadata templates, explained! [SOP for populating NCBI submission templates](#)

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Griffiths, E. J. et al. The PHA4GE SARS-CoV-2 Contextual Data Specification for Open Genomic Epidemiology. (2020) doi:10.20944/preprints202008.0220.v1.
<https://www.preprints.org/manuscript/202008.0220/v1>

DOI

dx.doi.org/10.17504/protocols.io.bid7ka9n

PROTOCOL CITATION

Ruth E Timme, Heather M Blankenship, Pavitra Roychoudhury, Erin L Young, Emma Griffiths, Duncan MacCannell, Stacia Wyman, Lee Katz 2021. SARS-CoV-2 NCBI consensus submission protocol: GenBank. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.bid7ka9n>

MANUSCRIPT CITATION please remember to cite the following publication along with this protocol

Griffiths, E. J. et al. The PHA4GE SARS-CoV-2 Contextual Data Specification for Open Genomic Epidemiology. (2020) doi:10.20944/preprints202008.0220.v1.
<https://www.preprints.org/manuscript/202008.0220/v1>

COLLECTIONS ⓘ

 **SARS-CoV-2 NCBI submission workflow + guidance for structuring and releasing metadata**

KEYWORDS

NCBI submission, pathogen surveillance, SARS-CoV-2, covid-19, genomic epidemiology, GenBank

LICENSE

————— This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

CREATED

Jul 09, 2020

LAST MODIFIED

Apr 09, 2021

PROTOCOL INTEGER ID

39071

PARENT PROTOCOLS

Part of collection

[SARS-CoV-2 NCBI submission workflow + guidance for structuring and releasing metadata](#)

BEFORE STARTING

This protocol has two sections:

Section 1: ensuring your NCBI submission environment is established

Section 2: SARS-CoV-2 submission of assemblies or consensus sequences to GenBank.

Associated protocols:

- [SOP for populating the NCBI submission templates](#) (e.g. source modifiers for GenBank)
- [NCBI submission to BioProject, SRA, and BioSample](#). Also includes NCBI account set-up for new users (Step 1)
- [NCBI Data Curation protocol](#) for making updates, corrections, or retractions to your data.

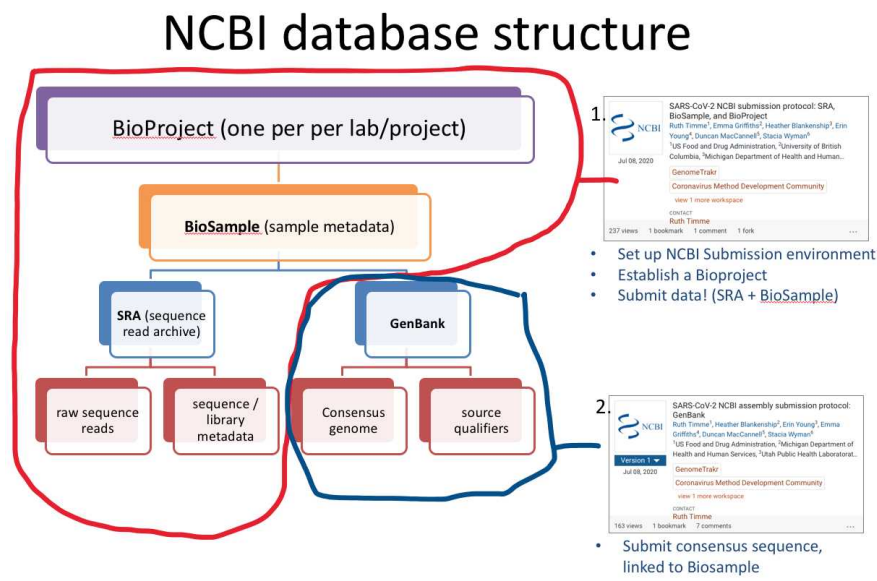
Link to [PHA4GE contextual data specification](#)

"Ingredients" to have in place before starting your submissions

1. **1.1:** Ensure you have a working NCBI user account
1. **1.2:** Identify your NCBI submission user group or establish a new one if necessary.
1. **1.3:** Bookmark the link to your submission portal

1.4. BioSample + BioProject assessments in-hand

After these steps are complete you can proceed with GenBank data submission in **Step 2**.



The NCBI database structure and how this protocol (#2) fits into the submission process.

1.1 Sign in to your **NCBI user account**: <https://www.ncbi.nlm.nih.gov/account/>

NCBI Resources How To

COVID-19 is an emerging, rapidly changing threat. Get the latest public health information from CDC. Get the latest research from NIH. Find NCBI SARS-CoV-2 literature, sequence, and clinical data.

Sign in to NCBI

Sign in with

Google NIH Login Commons

See more 3rd party sign in options

OR

Sign in directly to NCBI

NCBI Username

Password

☐ Keep me signed in

Sign In

[Forgot NCBI username or password?](#)

[Register for an NCBI account](#)

1.2 Ensure you have an NCBI user group established and correct permissions are assigned for you to submit.

List of submission groups: <https://submit.ncbi.nlm.nih.gov/groups/>

Submission Portal

HomeMy submissionsManage dataGroupsTemplatesMy profile

Groups

Search

Group Id	Full name	Aliases	Department	Institution	Contact email	Members
fda	FDA Center for Food Safety and Applied Nutrition	FDA/CFSAN	CFSAN-ORS-DM-MMSB	US Food and Drug Administration	GenomeTrakr@fda.hhs.gov	13
fda_ny	FDA/CFSAN/NY_State		CFSAN-ORS-DM-MMSB	US Food and Drug Administration	Ruth.Timme@fda.hhs.gov	6
fda_mdh	FDA/CFSAN/MDH		CFSAN-ORS-DM-MMSB	US Food and Drug Administration	Ruth.Timme@fda.hhs.gov	6

If you don't have a submission group established, please follow this protocol to create one for your laboratory group:

<https://www.protocols.io/edit/sars-cov-2-ncbi-submission-protocol-sra-biosample-bf7bjrin>

1.3 Bookmark "my submissions" at NCBI: <https://submit.ncbi.nlm.nih.gov/subs/>. This is your homepage for tracking your NCBI submissions.

If you see a blank page with a yellow box in the upper right corner saying "please login", click this link and login using the credentials created in **Step 1.1**.

Submission Portal

Home

My submissions

Manage data

Groups

Templates

My profile

Your submissions

Start a new submission

GenBank

BioProject

Sequence Read Archive

BioSample

Genome

Supplementary Files

TSA

API

Filter / Search

From date

To date

Status

Sort by

YYYY-MM-DD

YYYY-MM-DD

Not deleted

desc

Apps

+

Data archives

+

Query

Search

Clear

106,699 submissions

— previous

1

2

3

4

5

6

7

8

9

...

4,265

4,266

4,267

4,268

next →

Submission	Title	App	Group	Status	Updated
SUB0909508	SARS-CoV-2	GenBank	fda	Unfinished at the References step	17:46
SUB0502842	GenomeTrakr Project: New York State Department of Health, Wadsworth Center	BioProject	ny_ny	BioProject: Processed PRJNA542481 - GenomeTrakr Project: Listeria monocytogenes New York State Department of Health, Wadsworth Center (TXMID: 1639) Locus Tag Prefixes: <ul style="list-style-type: none">E2141 (SAMN11148776)E2807 (SAMN11162767) locustagprefix.txt	Manage data 08:00
SUB480036	Salmonella enterica Genome sequencing	BioProject	fda	BioProject: Processed PRJNA242847 - GenomeTrakr Project: USDA - Food Safety and Inspection Service (TXid: 289021) Locus Tag Prefixes: <ul style="list-style-type: none">AQJ62 (SAMN04532063)AQJ63 (SAMN04532066) locustagprefix.txt	Manage data 07:30

1.4 Identify your lab's BioProject accession. Does your laboratory have an established BioProject for this effort?

If not please follow instructions in **Step 3** of <https://www.protocols.io/edit/sars-cov-2-ncbi-submission-protocol-sra-biosample-br8ym9xw> for creating a new one.

Data submission (assemblies to GenBank)

2 GenBank consensus (or assembly) submission of SARS-CoV-2:

SARS-CoV-2 landing page: <https://submit.ncbi.nlm.nih.gov/sarscov2/>

Click "submit" under GenBank.

2.1 For all sequences you intend to submit at this time (one BioProject per submission):

1. Gather associated BioSample accessions and metadata previously registered in <https://www.protocols.io/edit/sars-cov-2-ncbi-submission-protocol-sra-biosample-br8ym9xw> along with three pieces of information describing the sequencing method and assembly:

- Sequencing method.** Populate with the PHA4GE field "sequencing instrument"
- Assembly program/pipeline.** Populate with the name from the PHA4GE field "assembly method"
- Version** of the assembly program. Populate with the version from the PHA4GE field "assembly method"

A	B	C	D	E
BioSample Accession	sample_name	seq. method	assembly program	assembly version or date
SAMN15460792	CA-IGI-0042	MinION	ARTIC-nCoV-bioinformaticsSOP	1.1.0
SAMN15460793	CA-IGI-0031	MinION	ARTIC-nCoV-bioinformaticsSOP	1.1.0

Example of two BioSamples and associated sample_name IDs

2. Concatenate all SARS-CoV-2 consensus sequences into a single fasta file, where the fasta headers contain the "sample_name" submitted to the BioSample.

Example FASTA file for two sequence submissions:

```
>CA-IGI-0042
ATCGATCGGTACCTAAGGATCGATCGGTACCTAAGGATCGATCGGTACCTAAGG....
>CA-IGI-0031
ATCGATCGGTACCTAAGGATCGATCGGTACCTAAGGATCGATCGGTACCTAAGG....
```

NCBI allows batch submissions to one BioProject at a time. If you are submitting to more than one BioProject, you'll need to prepare multiple submissions.

2.2 Download and populate the PHA4GE GenBank source modifiers template:

 [PHA4GE_GenBank-source_modifiers_Feb2021.v2.xlsx](#)

Guidance:

- Follow **Step 4** in the [SOP for populating the NCBI submission templates](#) for populating the source modifiers.
- Consult the current [PHA4GE contextual data specification](#) where relevant.

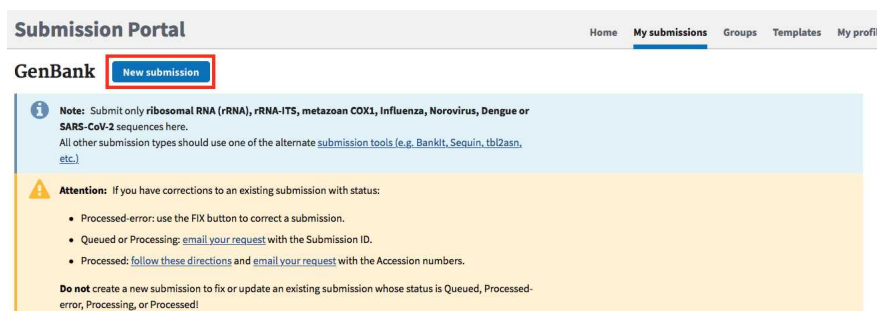
Populate the metadata spreadsheets for each isolate you intend to submit (you can submit metadata for a single isolate, entire MiSeq run, or for a large collection of isolates you intend to submit in batch).

****Ensure that the BioProject and BioSample(s) were registered using the same NCBI user group. If you are not listed as an owner on the BioProject/BioSample(s) you will not be able to properly link the new assembly data to existing records.**

the

Save the excel spreadsheet as a tab-delimited text file (.tsv) and **ensure that the date field is formatted correctly** (e.g. 2020-04-20) in the text file.

2.3 Click the "New submission" box.



Submission Portal Home **My submissions** Groups Templates My profile

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, Sequin, 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!

2.4 SUBMISSION TYPE tab:

Select the "**SARS-CoV-2**" option and click **Continue**.

Submission Portal Home My submissions Groups Templates My profile

GenBank submission: SUB9059508 [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 *

Submission type

* What do your sequences contain?

- ☐ Prokaryotic rRNA/IGS
- ☐ Eukaryotic Nuclear rRNA/ITS
- ☐ Eukaryotic Organelle rRNA
- ☐ Metazoan (multicellular animal) Mitochondrial COXI
- ☐ Influenza virus
- ☐ Norovirus
- ☐ Dengue virus
- ☐ Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) **NEW**

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

Submission title (Optional, not displayed in final records)

[Continue](#)

2.5 SUBMITTER tab:

Populate with submitter info. The “submitter” is the name of the person, or user group, who is physically doing the submissions, not a supervisor or PI.

**Must be the same person or group that submitted the associated BioSamples and BioProject.

Select the appropriate submission group name for your laboratory and check the contact information below.

If you do not have a submission group available to click, see **Steps 1.2-1.3 in the [SRA submission protocol](#) to establish a new one for your laboratory, or to add your name to a group already established for your lab.

Submission Portal Home My submissions Groups Templates My profile

GenBank submission: SUB9059508 [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

Submitter Required fields are marked with *

Affiliation

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
- ☒ 13 members [FDA Center for Food Safety and Applied Nutrition \(edit group\)](#)
Anjanette Johnston, Yan Luo, Errol Strain, Justin Payne, Narjol Gonzalez-Escalona, Hugh Rand, Maria Bailey, fda service, Maria Hoffmann, Jayanthi Gangiredla, Julie Haendiges, Yu Wang, you

Click "Continue" to proceed.

2.6 SEQUENCING TECHNOLOGY tab:

This information will get populated as a structured comment on the GenBank record.

Pull the sequencing method and assembly information gathered in **Step 2.1**.

Method: sequencing technology or platform.

Assembly State: Click "Assembled sequences".

Assembly information: Specify program/pipeline AND version.

The screenshot shows the 'Submission Portal' for GenBank. The user is logged in as 'SUB9059508' and is on the 'Sequencing Technology' tab. The progress bar shows steps 1 through 9, with 'Sequencing Technology' being the current step. The 'Method' section asks 'What methods were used to obtain these sequences?' and lists options: Sanger dideoxy sequencing, 454, Helicos, Illumina, IonTorrent, Pacific Biosciences, SOLID, and Other (selected). The 'Method' field is filled with 'MinION'. The 'Assembly state' section has two options: 'Unassembled sequence reads' and 'Assembled sequences (each sequence was assembled from two or more overlapping sequence reads)' (selected). The 'Assembly Information' section has a table with columns for 'Assembly program', 'Version or date', and 'Delete'. The first row shows 'ARTIC-nCoV-bioinform' and '1.1.0'. There is a 'Continue' button at the bottom.

Click **Continue** to proceed.

2.7 SEQUENCES tab:

Release date: Click "Release immediately following processing" for all routine surveillance isolates.

Sequences:

Sequences can be uploaded one at a time (one per submission), or as a batch upload in a single concatenated FASTA file (<https://submit.ncbi.nlm.nih.gov/genbank/help/#fasta>) when you are submitting multiple isolates in one submission. See **Step 2.1** for guidance on formatting your FASTA file.

Fasta headers must include a unique ID that links the sequence to the source modifiers

For example:

FASTA header:

>CA-IGI-0042

Source modifier template

ID from Sequence_ID column in metadata workbook: CA-IGI-0042

Click "Choose File" to browse and upload your .fasta file:

Submission Portal

GenBank submission: SUB7534548
SARS-CoV-2

1 SUBMISSION TYPE

2 SUBMITTER

3 SEQUENCING TECHNOLOGY

4 SEQUENCES

5 SOURCE MODIFIERS

6 REFERENCES

7 REVIEW & SUBMIT

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 no file selected

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

Example FASTA nucleotide format:

```
>Seq1
aacgatatagagatagtgatccgatagagagagga
>Seq2
gtacgataaagagatagtgatccgatagagagagga
```

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

Continue

Click "Continue" and respond to any validation issues.

Common validation issues:

Ambiguous bases were trimmed warning. Ambiguous bases are non- A, T, G, or Cs. NCBI trims terminal Ns first at the 5' end, then looks to see if 50% of the first 10 bases are ambiguous and trim to last ambiguous. If more than 30% of the first 50 are ambiguous, we trim to the last ambiguous and then recheck the 5' end. If that is fine, we follow the same steps on the 3' end. This procedure is run again if we trimmed vector from an end. NCBI removes sequences that are greater than 50% ambiguous after the trimming. They also remove sequences with internal vector.

String of NNNs: If your assembly contains strings of internal NNNs (from mapping to a reference genome), you will get a warning asking for you for more information:

- Click "A region of estimated length between the sequenced regions based on an alignment to similar sequences or genome" if the NNNs were caused by the reference-based assembly.
- Click "Release immediately following processing".

Submission Portal Home **My submissions** Groups

GenBank submission: SUB9059508
SARS-CoV-2

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

Sequences ⓘ Required fields

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

Sequence-IDs
CA-IGI-3157
CA-IGI-3197

What do the internal NNN's represent?

❗ 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

ⓘ 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.

File Name	Size	Created	Actions
SUB9059508_GenbankExample_fasta_filtered.fsa	59.0 kB	2021-02-10 15:26	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
aacgatatagagatagtcgatatagagagagga
>Seq2
gtacgataaagagatagtcgatatagagagagga
```

ⓘ 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](#)

Click "Continue" again.

2.8 SOURCE INFO tab:

GenBank will attempt to pull out IDs from the fasta headers. For our case these should be the 'sequence_ID' in the source modifier table (not the isolate).

Click "none of the above"

Submission Portal Home **My submissions** Groups Templates My profile

GenBank submission: SUB9059508 [Delete submission](#)

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

Source Information ⓘ Required fields are marked with * asterisk

ⓘ The first few sequence IDs that we found are:

CA-IGI-3157
CA-IGI-3197

★ Do your sequence IDs represent one 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.

☐ Isolate

☒ NONE of these

[Continue](#)

2.9 SOURCE MODIFIERS tab:

Guidance for populating this metadata outlined in **Step 2.2**.

- Click "Upload a tab-delimited table"

The screenshot shows the 'Submission Portal' for GenBank submission SUB9059508 (SARS-CoV-2). The progress bar indicates the current step is '6 SOURCE MODIFIERS'. The 'Source Modifiers' section explains that for each sequence, GenBank requires source information: collection-date, country, host, and isolate. It notes that required fields are marked with an asterisk and that at least one field marked with a star, a dagger, or a double dagger is required. A message states: 'If you have already provided all the required information, you can press Continue to proceed.' A link for more help is provided. A section titled 'How do you want to apply source modifiers?' offers two options: 'Use an editable table' (selected) and 'Upload a tab-delimited table (template file provided)'. A 'Continue' button is at the bottom.

Upload the csv file created from populating the **PHA4GE GenBank source modifiers template** in **Step 2.2**. Upload this file by clicking on the "upload a tab-delimited text file" link. Ensure that the first column in this spreadsheet, "Sequence_ID" contains an ID that matches *exactly* the ID used in your FASTA file headers.

ERRORS: If you are not listed as an owner on the BioProject/BioSample you will see an error here stating that these are "Unknown", e.g. Error: Unknown BioProject accession(s).

Solve this by updating the ownership on previous submissions OR, if you haven't created a submission group, see **Steps 1.2-1.3** in the [SRA submission protocol](#) to establish a new one for your laboratory, or to add your name to a group already established for your lab.

Click Continue.

2.10 REFERENCES tab:

Sequence Authors: Enter names here from your NCBI submission user group (can be a sub-set of the names or the full submission group list).

Publication status: For routine surveillance submissions choose "Unpublished".

Submission Portal Home My submissions Groups Templates My profile

GenBank submission: SUB9059508 SARS-CoV-2 [Delete submission](#)

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

References

Sequence authors

Who should be publicly credited as the submitter of this sequence data? Enter authors below. Drag and drop to reorder authors.

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

[Add another sequence author](#)

Reference

* Publication status

☐ Unpublished ☐ In press ☐ Published

[Continue](#)

Click Continue.

2.11 REVIEW & SUBMIT tab:

Review the submission and Genbank record preview, paying close attention to correct linkage of BioProject and BioSample, plus any other metadata submitted as source qualifiers (in the FEATURES->source section).

Submission Portal Home My submissions Groups Templates My profile

GenBank submission: SUB9059508 SARS-CoV-2 [Delete submission](#)

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

Review & Submit

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

Submitter

Submitter: Ruth Timme
ruth.timme@fda.hhs.gov
retimme@gmail.com

Institution: US Food and Drug Administration

Department: CFSAN-ORS-DM-MMSB

Street: 5100 Paint Branch Parkway

City: College Park

State: MD

Postal code: 20740

Country: USA

Phone: 2404025127

Sequence authors

Ruth E. Timme

References

Publication status: unpublished

Authors: same as sequence authors

Sequencing Technology

Methods: Other: MiniON

Assembly state: assembled

Assembly Programs: ARTIC-nCoV-bioinformaticsSOP (1.1.0)

Uploaded files

GenBank Record Preview

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

LOCUS CA-IGI-3157 29782 bp DNA linear VRL 11-FEB-2021

DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate

ACCESSION SRR161111.1

VERSION SRR161111.1

DBLINK BioProject: PRJNA186835

KEYWORDS Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

SOURCE Severe acute respiratory syndrome coronavirus 2

ORGANISM Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Coronavirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.

REFERENCE 1 (bases 1 to 29782)

AUTHORS Time, R.E.

TITLE Direct Submission

JOURNAL Submitted (11-FEB-2021) CFSAN-ORS-DM-MMSB, FDA Center for Food Safety and Applied Nutrition, 5100 Paint Branch Parkway, College Park, MD 20740, USA

COMMENT #Assembly-Data-START#

Assembly Method : ARTIC-nCoV-bioinformaticsSOP v. 1.1.0

Sequencing Technology : MiniON

#Assembly-Data-END#

FEATURES

source

1..29782

/organism="Severe acute respiratory syndrome coronavirus 2"

/mol_type="genomic DNA"

/isolate="SARS-CoV-2/human/USA/CA-IGI-3157/2021"

/isolation_source="clinical; See additional sample source fields for further information (i.e. anatomical material, environmental site, body product, environmental material, anatomical part, collection device, collection method)"

/host="Homo sapiens"

/db_xref="taxon:2697049"

/country="USA:California"

/collection_date="2021-02-08"

/note="GISAID accession: EPI_ISL_1234567"

BASE COUNT 8817 a 5426 c 5786 g 5491 t 262 others

ORIGIN

1 agatctgttc tctaaagaa ctttaaatc tgtgtgttc tctctgctt gcatctgtt

61 tgccttcag cagttatatt actactact tctgtgttc gacagacac ggtgacttc

121 tatattctt gcagctgct taccgtttc tctgtgttc agcgcacat cagcacatc

Click Submit when ready.

2.12 GenBank accessions:

The status of your submission will be updated once it is processed and can be tracked within the "My Submissions" tab: <https://submit.ncbi.nlm.nih.gov/subs/>.

GenBank accessions will be listed here, under "GenBank: Processed" and available for download in the AccessionReport.tsv file.

Submission ↕	Title ↕	App ↕	Status ↕	Updated ↕
SUB7560154	SARS-CoV-2	GenBank	✓ GenBank: Processed MT683386-MT683418 3 files: <ul style="list-style-type: none"> • AccessionReport.tsv • flatfile.zip • email.txt 	Jul 01

Sequences with no annotation issues will be listed as **Processed**. Submissions with annotation discrepancies will be marked as **Error** and a Fix button will appear. A report is emailed to you and listed on the submissions page with the detailed issues. If the data is incorrect, click the Fix button and you will return to the sequences page of your submission to upload a corrected file.

If you have evidence that the discrepancy is due to a naturally occurring mutation, send an email to **gb-admin@ncbi.nlm.nih.gov** with the SUB number and evidence.

2.13 Important data stewardship and curation notes:

- Develop an internal method for storing and tracking your GenBank accessions! They are required for making future updates to your records.

For updates to your GenBank records follow the NCBI Curation Protocol hosted by GenomeTrakr: <https://www.protocols.io/view/ncbi-data-curation-protocol-bacaiase>