



Jan 07, 2022

# SARS-CoV-2 NCBI consensus submission protocol: GenBank V.3

Ruth Timme<sup>1</sup>, Emma Griffiths<sup>2</sup>, Lee Katz<sup>3</sup>, Duncan MacCannell<sup>4</sup>, Michael Weigand<sup>4</sup><sup>1</sup>US Food and Drug Administration; <sup>2</sup>University of British Columbia; <sup>3</sup>CDC; <sup>4</sup>Centers for Disease Control and Prevention

1

[dx.doi.org/10.17504/protocols.io.b3bnqime](https://dx.doi.org/10.17504/protocols.io.b3bnqime)[GenomeTrakr](#) [Coronavirus Method Development Community](#) 2Technical Outreach and Assistance for States Team  
Centers for Disease Control and Prevention

Please note that this protocol is public domain, which supersedes the CC-BY license default used by protocols.io.

**This is a SARS-CoV-2 specific protocol that covers the steps needed to submit SARS-CoV-2 consensus sequence to GenBank.**

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](mailto:gb-admin@ncbi.nlm.nih.gov) to set up an account for submitting through the API.

This protocol assumes (and requires) that the user has a BioProject and BioSamples(s) already registered.

**Complete in order::**

**1. [Populate your templates first.](#)**

**2. [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. Users can modify this protocol to just create a BioSample with no linked raw data.

**3. [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.

**Version history:**

V3: Direct links provided to download metadata templates (instead of hosting duplicate files). minor edits throughout the protocol.

DOI

[dx.doi.org/10.17504/protocols.io.b3bnqime](https://dx.doi.org/10.17504/protocols.io.b3bnqime)

Ruth Timme, Emma Griffiths, Lee Katz, Duncan MacCannell, Michael Weigand 2022. SARS-CoV-2 NCBI consensus submission protocol: GenBank. **protocols.io**  
<https://dx.doi.org/10.17504/protocols.io.b3bnqime>  
Ruth Timme



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>

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

protocol ,

Dec 28, 2021

Jan 13, 2022

Dec 28, 2021  Ruth Timme US Food and Drug Administration

Jan 13, 2022  Technical Outreach and Assistance for States Team Centers for Disease Control and Prevention

56398

:

Please note that this protocol is public domain, which supersedes the CC-BY license default used by protocols.io.

**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.

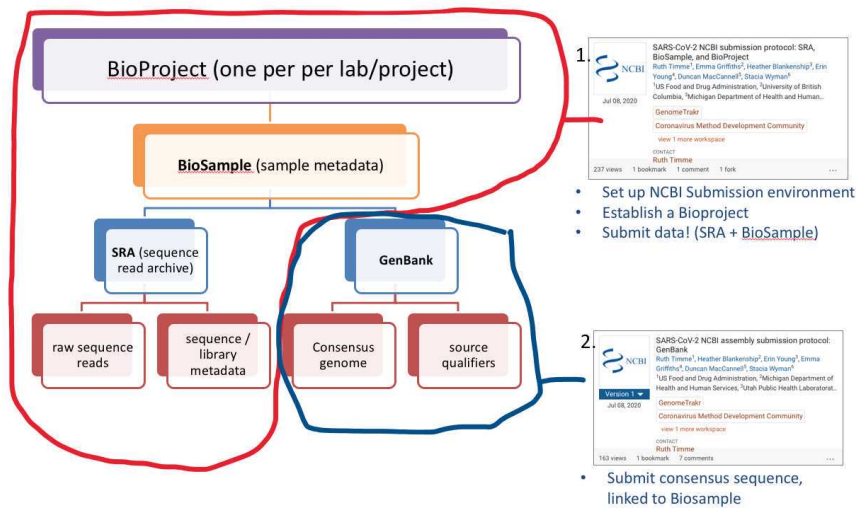
Link to [PHA4GE contextual data specification](#)

**"Ingredients"** to have in place before starting your submissions

- 1.1: Ensure you have a working NCBI user account
- 1.2: Identify your NCBI submission user group or establish a new one if necessary.
- 1.3: Bookmark the link to your submission portal
- 1.4. **BioSample + BioProject accessions in-hand**

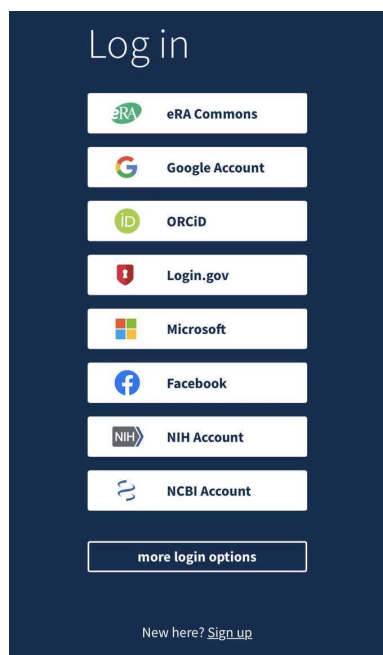
After these steps are complete you can proceed with GenBank data submission in **Step 2**.

# NCBI database structure



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/>



## 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

Home

My submissions

Manage data

Groups

Templates

My 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

Sequence Read Archive

Genome

TSA

BioProject

BioSample

Supplementary Files

API

Filter / Search

From date

To date

Status

Sort by

YYYY-MM-DD

YYYY-MM-DD

Not deleted

Apps

Data archives

Query

Search

Clear

106,699 submissions

Submission

Title

App

Group

Status

Updated

SUB905908

SARS-CoV-2

GenBank

fda

Unfinished at the References step

17:45

SUB5032842

GenomeTrakr Project: New York State Department of Health, Wadsworth Center

BioProject

nys\_g

✓

BioProject: Processed

PRJNA514286: GenomeTrakr Project: Listeria monocytogenes New York State Department of Health, Wadsworth Center (TaxID: 1639)

Locus Tag Prefixes:

E2J45 (SAMN11148736)

E2B07 (SAMN11102767)

locustagprefix.txt

Manage data

08:00

SUB480036

Salmonella enterica Genome sequencing

BioProject

fda

✓

BioProject: Processed

PRJNA424847: GenomeTrakr Project: USDA - Food Safety and Inspection Service (TaxID: 238511)

Locus Tag Prefixes:

A0J62 (SAMN0432083)

A0J63 (SAMN0432086)

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 Populate the two GenBank templates.

Download the current templates in [Step 4 Overview of NCBI's SARS-CoV-2 submission process and the metadata required](#).

1. GenBank structured comment info (mapping/assembly methods)
2. GenBank Source Qualifiers

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).

Save the source qualifier 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.

**\*\*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 (and you will get a validation error in the submission portal).**

## 2.2 Prepare your sequence files.

**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 only **ONE** BioProject at a time. If you are submitting to more than one BioProject, you'll need to prepare one submission for each BioProject.

## 2.3 Click the "New submission" box.

The screenshot shows the GenBank Submission Portal interface. At the top, there's a navigation bar with 'Home', 'My submissions', 'Groups', 'Templates', and 'My profile'. Below this, the 'GenBank' logo is on the left, and a 'New submission' button is highlighted with a red rectangular box. Below the button, there's a light blue box with a note: '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.)'. Below that is an orange box with an 'Attention' icon and text: 'Attention: If you have corrections to an existing submission with status:'. It lists three bullet points: 'Processed-error: use the FIX button to correct a submission.', 'Queued or Processing: email your request with the Submission ID.', and 'Processed: follow these directions and email your request with the Accession numbers.' At the bottom of the orange box, it says: '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**.

The screenshot shows the 'Submission Type' tab in the GenBank Submission Portal. The top navigation bar is the same as in the previous screenshot. Below it, the 'GenBank submission: SUB9059508' is displayed, with a 'Delete submission' button to its right. Below this is a progress bar with steps: 1 SUBMISSION TYPE, 2 SUBMITTER, 3 SEQUENCING TECHNOLOGY, 4 SEQUENCES, 5 SEQUENCE PROCESSING, 6 SOURCE INFO, 7 SOURCE MODIFIERS, 8 REFERENCES, and 9 REVIEW & SUBMIT. The 'SUBMISSION TYPE' step is currently selected. Below the progress bar, the 'Submission Type' section is shown. It has a sub-header 'Submission type' and a note: 'Required fields are marked with \* asterisk'. Under this, there's a section 'What do your sequences contain?' with several radio button options: 'Prokaryotic rRNA/IGS', 'Eukaryotic Nuclear rRNA/ITS', 'Eukaryotic Organelle rRNA', 'Metazoan (multicellular animal) Mitochondrial COX1', 'Influenza virus', 'Norovirus', 'Dengue virus', and 'Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) NEW'. The 'SARS-CoV-2' option is selected. Below this is a note: 'If none of the options above describe your sequences, use BankIt to submit.' At the bottom, there's a text input field for 'Submission title (Optional, not displayed in final records)' and a 'Continue' button.

## 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.

The screenshot shows the 'Submitter' tab of the GenBank submission portal for submission SUB9059508 (SARS-CoV-2). The progress bar indicates the current step is '2 SUBMITTER'. The 'Affiliation' section contains a note about the information being displayed in the final sequence records and a field for the primary address. Below this, the 'Group for this submission' section shows a group with 13 members, including the FDA Center for Food Safety and Applied Nutrition and several individuals. A 'Delete submission' button is visible in the top right corner.

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 'Sequencing Technology' tab of the GenBank submission portal for submission SUB9059508 (SARS-CoV-2). The progress bar indicates the current step is '3 SEQUENCING TECHNOLOGY'. The 'Method' section asks for the methods used to obtain the sequences, with checkboxes for Sanger dideoxy sequencing, 454, Helicos, Illumina, IonTorrent, Pacific Biosciences, SOLID, and Other (selected). 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 includes a table with columns for 'Assembly program', 'Version or date', and 'Delete'. The first row shows 'ARTIC-CoV-bioinform' with version '1.1.0'. A 'Continue' button is 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
aacgatagatagtagtgatccgatagagagga
>Seq2
gtacgataagatagtagtgatccgatagagagga
```

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**HomeMy submissionsGroups

**GenBank submission: SUB9059508**  
SARS-CoV-2

1 SUBMISSION TYPE2 SUBMITTER3 SEQUENCING TECHNOLOGY4 SEQUENCES5 SOURCE INFO6 SOURCE MODIFIERS7 REFERENCES8 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.  
SUB9059508\_GenbankExample\_fasta\_filtered.fsa59.0 kB2021-02-10 15:26Delete

If you have multiple sequences, all of your sequences need to be in one file. Help on FASTA file.

Example FASTA nucleotide format:  
>Seq1  
aacgatatagatagatggtccgatatagagagga  
>Seq2  
gtacgataaagatagatggtccgatatagagagga

If you have multiple sequences, all of your sequences need to be in one file. Help on FASTA file.

Example FASTA nucleotide format:  
>Seq1  
aacgatatagatagatggtccgatatagagagga  
>Seq2  
gtacgataaagatagatggtccgatatagagagga

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"

 protocols.io

9

Citation: Ruth Timme, Emma Griffiths, Lee Katz, Duncan MacCannell, Michael Weigand SARS-CoV-2 NCBI consensus submission protocol: GenBank <https://dx.doi.org/10.17504/protocols.io.b3bnqime>

This is an open access protocol distributed under the terms of the **Creative Commons Attribution License** (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium,

**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

### 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"

**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

### Source Modifiers

Required fields are marked with \* asterisk.  
At least one of the fields marked with ★, †† or ‡ is required.

For each sequence, GenBank requires the following source information:

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

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

[More help: what is a source modifier, description of each modifier, how to provide source modifiers.](#)

★ ★ How do you want to apply source modifiers?

☐ Use an editable table  
☐ Upload a tab-delimited table (template file provided)

[Continue](#)

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

HomeMy submissionsGroupsTemplatesMy profile

GenBank submission: SUB9059508

SARS-CoV-2

Delete submission

1 SUBMISSION TYPE2 SUBMITTER3 SEQUENCING TECHNOLOGY4 SEQUENCES5 SOURCE INFO6 SOURCE MODIFIERS7 REFERENCES8 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

UnpublishedIn-pressPublished

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

HomeMy submissionsGroupsTemplatesMy profile

GenBank submission: SUB9059508

SARS-CoV-2

Delete submission

1 SUBMISSION TYPE2 SUBMITTER3 SEQUENCING TECHNOLOGY4 SEQUENCES5 SOURCE INFO6 SOURCE MODIFIERS7 REFERENCES8 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

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

24040225127

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

SARS-CoV-2/human/USA/CA-IGI-3157/2021.

DR LINK

BioProject: PRJNA386835

BioSample: SAMN06175389

KEYWORDS

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

ORGANISM

Severe acute respiratory syndrome coronavirus 2

REFERENCE

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 Method :: ARTIC-nCoV-bioinformaticsSOP v. 1.1.0

Sequencing Technology :: MinION

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), anatomical part, body product, environmental material, environmental site, collection device, collection method)"

/host="homo sapiens"

/db\_xref="taxon:2697049"

/country="USA:California"

/collection\_date="2021-02-08"

/note="GISAID accession: EPI\_ISL\_123457"

BASE COUNT

8817 a 5426 c 5786 g 9491 t 262 others

ORIGIN

1 agatctgttc tctaaagaa ctttaaatc ttgtgtgctg tctactgctt gcactgcttg

81 tgcactcag cagtaattt aataactat tactgtgctt gacaggac gagtaactg

121 tctactcttc gcagctgctt tagctgcttg tctactgctg agcagctat cagcactct

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: <b>Processed</b> MT683386-MT683418 3 files: <ul style="list-style-type: none"><li>• AccessionReport.tsv</li><li>• flatfile.zip</li><li>• email.txt</li></ul>	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>