




Jun 08, 2022

🌐 NCBI submission protocol for foodborne virus surveillance

Zhihui Yang¹, Ruth Timme², Maria Balkey², Zhihui Yang³¹FDA/CFSAN/OARSA/DMB; ²CFSAN/ORS/DM/MMSB; ³CFSAN/OARSA/DMB

1

dx.doi.org/10.17504/protocols.io.j8nlkkdbxl5r/v1 Zhihui Yang
FDA

INTRODUCTION:

This protocol outlines the steps which ViroTrakr contributors need to follow in order to submit their data to NCBI. It includes how to:

- establish your new BioProject at NCBI;
- link it to ViroTrakr;
- create BioSample numbers for your submission and submit raw sequence data to the SRA database;
- submit assembled data to GenBank and link them to ViroTrakr (optional).

ViroTrakr, a genomic database initiated by CFSAN and housed in NCBI, aims to (1) cover sequences of a wide range of foodborne viruses (e.g., norovirus, hepatitis A virus, sapovirus, etc.) from clinical, food and/or environmental specimens and (2) provide real-time reference sequences for phylogenetic analysis and epidemiologic studies linked to foodborne illnesses. Pipelines are under development with which ViroTrakr will eventually integrate to an international genomic reference database, GenomeTrakr, which has been successfully employed for bacterial pathogen surveillance in multiple countries.

ViroTrakr: [foodborne viruses \(ID 396739\) - BioProject - NCBI \(nih.gov\)](#)

GenomeTrakr: [Multispecies \(ID 593772\) - BioProject - NCBI \(nih.gov\)](#)

Reference: NCBI submission protocol for microbial pathogen surveillance V.5:
[NCBI submission protocol for microbial pathogen surveillance \(protocols.io\)](#)

DOI

dx.doi.org/10.17504/protocols.io.j8nlkkdbxl5r/v1[http://ViroTrakr: foodborne viruses \(ID 396739\) - BioProject - NCBI \(nih.gov\)](http://ViroTrakr: foodborne viruses (ID 396739) - BioProject - NCBI (nih.gov))

Zhihui Yang, Ruth Timme, Maria Balkey, Zhihui Yang 2022. NCBI submission protocol for foodborne virus surveillance. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.j8nlkkdbxl5r/v1>



ViroTrakr, Foodborne virus surveillance, NCBI submission protocol

protocol ,

Apr 08, 2022

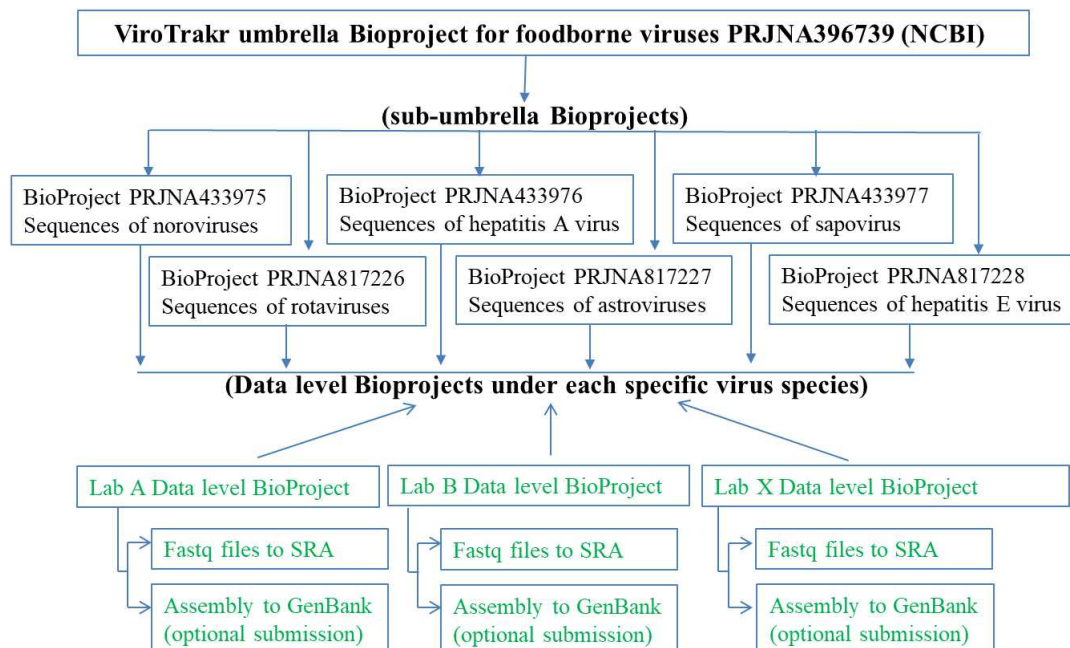
Jun 08, 2022

60501

ViroTrakr data structure

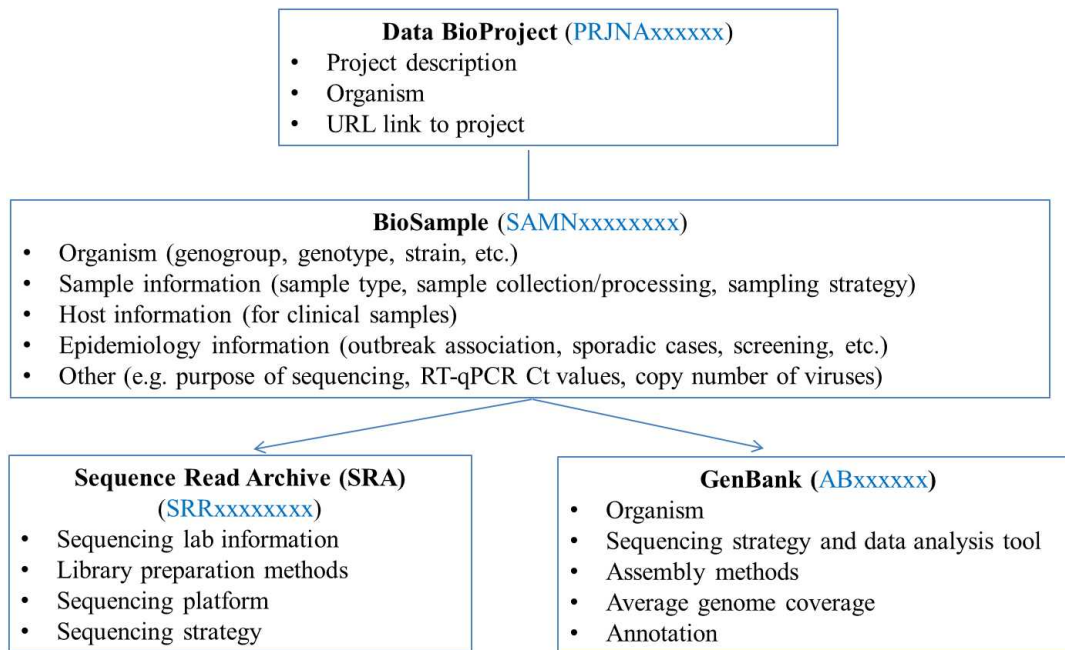
1

ViroTrakr database structure: the ViroTrakr database was established as an umbrella BioProject at NCBI with the structure shown below:



Note: The steps involved in your ViroTrakr submission are highlighted in green;
One data level BioProject per lab or per collaboration project.

Database structure: (cont.) for each data level BioProject:



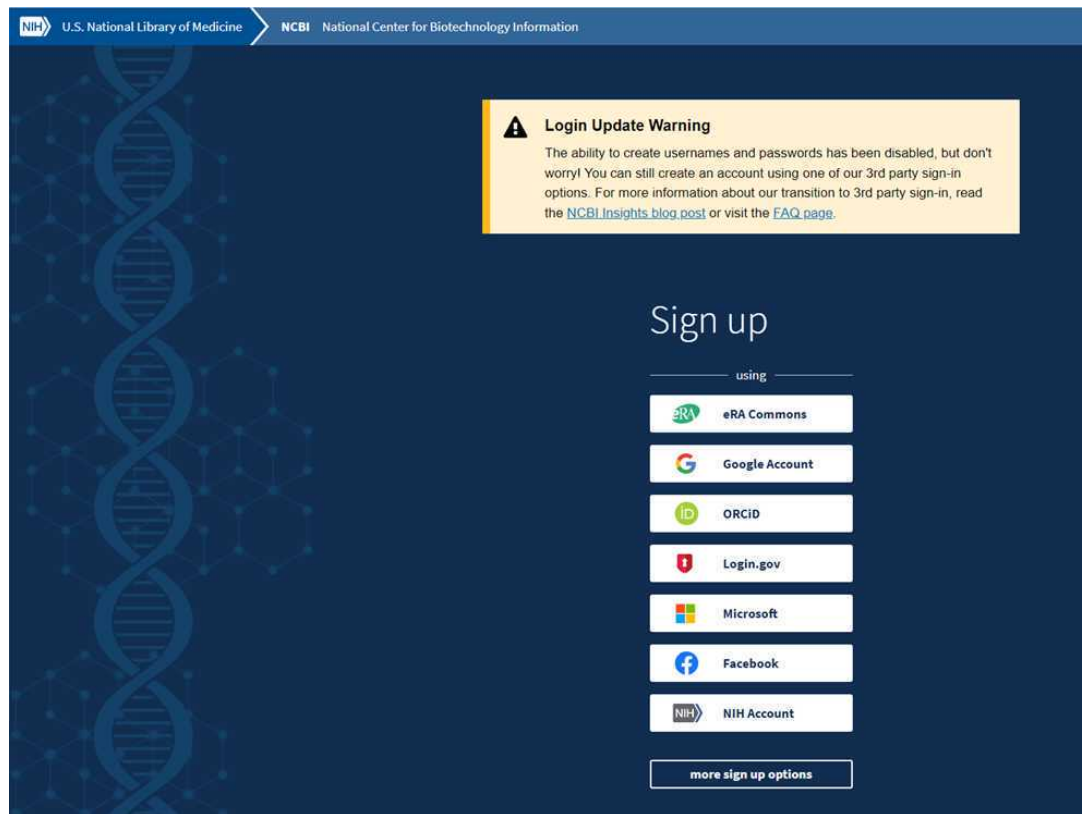
NCBI sign in

2

Getting started

Please refer to “NCBI submission protocol for microbial pathogen surveillance” Section 1 [NCBI submission protocol for microbial pathogen surveillance \(protocols.io\)](https://www.ncbi.nlm.nih.gov/submit/submit_protocol_microbial_pathogen_surveillance.html) for details.

- 2.1 For new users, directly create an account using one of the 3rd party sign-in options:
[Sign up / NCBI \(nih.gov\)](https://www.ncbi.nlm.nih.gov/submit/submit_protocol_microbial_pathogen_surveillance.html)



2.2 For existing users, sign in to our NCBI account: [NCBI Sign In Page \(nih.gov\)](https://www.ncbi.nlm.nih.gov/signin)



Note for existing users: NCBI-managed credentials are the username and password you set at NCBI – these will be retired in June 2022, and access to any My NCBI accounts without a linked 3rd-party login will require going through an access recovery process. Federated account credentials are those set through eRA Commons, Google, or a university or institutional point of access. Your NCBI Accounts and the contents of your current account will not change. You simply need to login a different way: through a third party option. See the FAQ for more information: (<https://ncbiinsights.ncbi.nlm.nih.gov/ncbi-login-retirement-faqs/>).

For existing users, use the steps below to link 3rd party logins to your account:

1. [Sign in directly to NCBI](#) with your username and password.
2. Click your username, which is located on the top right of the browser page.
3. Click “Change” in the “Linked Account” portal.
4. Locate the 3rd party account of your choice using the search bar.
5. You will be transferred to the 3rd party’s sign in page. Enter your credentials there for the 3rd party account.

2.3 You may group, organize and manage your NCBI submission environment for

your lab: please refer to “NCBI submission protocol for microbial pathogen surveillance” Section 1
[NCBI submission protocol for microbial pathogen surveillance \(protocols.io\)](https://protocols.io/j8nlkkdbxl5r/v1) for details.

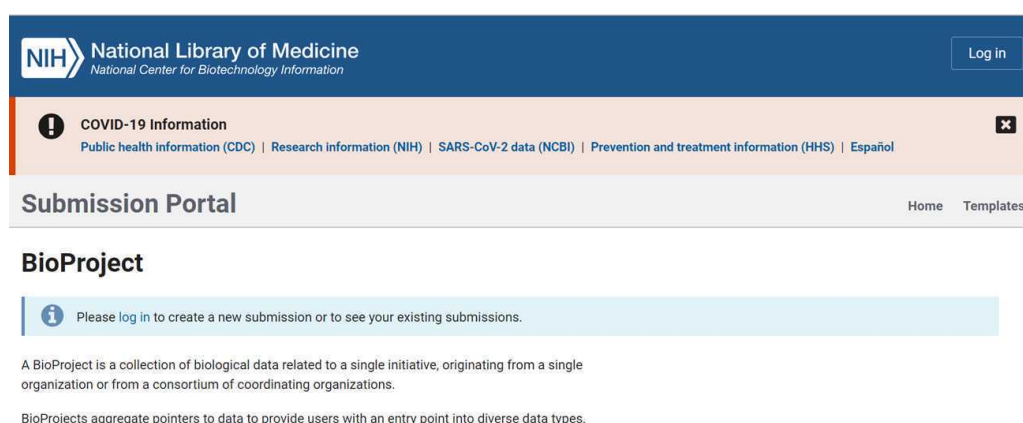
2.4 Log into your NCBI account and you are ready for your NCBI submission.

Creating BioProjects at NCBI

3 Establish your new data level BioProject under the umbrella BioProject ViroTrakr:

Please refer to “NCBI submission protocol for microbial pathogen surveillance” [NCBI submission protocol for microbial pathogen surveillance \(protocols.io\)](https://protocols.io/j8nlkkdbxl5r/v1) step 3 for details.

3.1 [Log into your NCBI account at Submissions | BioProject | Submission Portal \(nih.gov\)](https://ncbi.nlm.nih.gov/submissions/bioproject/):



3.2 Establish a new BioProject by clicking “New submission”:

NIH National Library of Medicine
National Center for Biotechnology Information

Submission Portal

Home My submissions Manage data Groups Templates My profile

BioProject [New submission](#)

ATTN: to update an existing record or recent submission, you can use "Manage data" to make some changes yourself. If you cannot make the desired change there, then please email your request with your BioProject accession or Submission ID included. **Do not** create new submission to update an existing submission!

Short description and brief instructions

9 submissions

Submission	Title	Group	Status	Updated
SUB2910971	Genome sequencing of foodborne viruses by FDA		✓ BioProject: Processed PRJNA396739 : ViroTrakr	Mar 05

[Manage data](#)

3.3

There are seven tabs under each BioProject submission.

Populate "Submitter" tab: (a submission group is highly recommended* for your laboratory)

Submission Portal

Home My submissions Manage data Groups Templates My profile

BioProject submission: SUB11436575 [Delete submission](#)

New

1 SUBMITTER 2 PROJECT TYPE 3 TARGET 4 GENERAL INFO 5 BIOEXAMPLE 6 PUBLICATIONS 7 REVIEW & SUBMIT

Submitter

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

* Email (primary) Email (secondary)
zhihui.yang@fda.hhs.gov zhihui.yang@fda.hhs.gov At least one email should be from the organization's domain.

Group for this submission
☐ No group (affiliation from my personal profile)
☐ 1 member Zhihui Yang's shared submissions
☒ 2 members CFSAN OARSA foodborne pathogen submission group

* Submitting organization Submitting organization URL * Department
FDA, CFSAN, OARSA DMB

Phone Fax

* Street * City * State/Province * Postal code * Country
8301 Muirkirk Rd Laurel MD 20708 USA

[Continue](#) ☒ Update my contact information in profile

*Note: to establish and use a user group for all your submission related to microbial genome surveillance is highly recommended. The reasons are, as mentioned in [NCBI submission protocol for microbial pathogen surveillance \(protocols.io\)](https://dx.doi.org/10.17504/protocols.io.j8nlkdbxl5r/v1):

"- it will link your laboratory's NCBI data ownership to the user group and not to individuals, allowing anyone in the current group to perform updates or retractions and answer inquiries from the NCBI staff, even if there's been a complete turnover of staff since the original data submission.

- it also ensures consistent data ownership across BioProjects, BioSamples, and sequence data. If your laboratory has non-overlapping research groups submitting and managing data at NCBI, multiple user groups can be established to track these efforts separately.”

You may use a submission group which has been already established by your laboratory. You may check the “Group” tab in the submission portal, <https://submit.ncbi.nlm.nih.gov/groups/> for this information. Ask your colleagues to do the same thing, to ensure your laboratory doesn't already have one in place.

If your laboratory doesn't have one proper submission group ready, please refer to [NCBI submission protocol for microbial pathogen surveillance \(protocols.io\)](https://protocols.io) section 1.2 and 1.3 for the details on:

- how to request and create a new user group by emailing to NCBI help staff at submit-help@ncbi.nih.gov
- how to manage your NCBI submission user group by clicking in the “Group” tab of the submission portal <https://submit.ncbi.nlm.nih.gov/groups/>

You may contact NCBI by emailing to submit-help@ncbi.nih.gov if you have any further question regarding submission group and need additional help.

3.4

Populate “Project type” tab (e.g. Raw sequence reads):

* Required fields are marked with * asterisk.

BioProject submission: SUB11436575

New

1 SUBMITTER 2 PROJECT TYPE 3 TARGET 4 GENERAL INFO 5 BIOSAMPLE 6 PUBLICATIONS 7 REVIEW & SUBMIT

Project Type

A general label indicating the primary study goal.

* Project data type

- ☐ Genome sequencing and assembly
- ☒ Raw sequence reads
- ☐ Genome sequencing
- ☐ Assembly
- ☐ Clone ends
- ☐ Epigenomics
- ☐ Exome
- ☐ Map
- ☐ Metagenome
- ☐ Metagenomic assembly
- ☐ Phenotype or Genotype
- ☐ Proteome
- ☐ Random survey
- ☐ Targeted loci cultured
- ☐ Targeted loci environmental
- ☐ Targeted Locus (Loci)
- ☐ Transcriptome or Gene expression
- ☐ Variation
- ☐ Other

3.5

Populate “Target” tab: move cursor to the question marks for description of each item.

Required fields are marked with * asterisk; fields without * asterisk could be left blank.

The screenshot shows the 'Target' tab of a BioProject submission form. The header includes the NIH logo and 'National Library of Medicine National Center for Biotechnology Information'. The user 'zhihui_yang' is logged in. The submission ID is SUB11058927. The progress bar shows '3 TARGET' as the active tab. The 'Target' section has a tooltip: 'The most descriptive organism name for the study (to the species, if relevant)'. A red arrow points to the 'Organism name' field, which is marked with an asterisk. Below it are fields for 'Strain', 'Breed', 'Cultivar', 'Isolate name', and 'Label', each with a question mark icon. A 'Description' field is also present. A 'Continue' button is at the bottom.

Note: choose the most descriptive and valid organism name for your study. For example, “Norwalk virus” instead of “norovirus”, “Homo sapiens” instead of “human” should be used. See [Organism information - BioSample - NCBI \(nih.gov\)](#) and [Home - Taxonomy - NCBI \(nih.gov\)](#) for more information about providing a valid organism name.

3.6

Populate “General Info” tab:

- Choose “release immediately following processing” or a specified date to release your submission;
- Provide a description (e.g., Norwalk virus sequencing) of the study goals and relevance (e.g., NGS of clinical samples as part of norovirus surveillance) under “Public description”;
- Choose a “Relevance” from the provided options;
- Click “Yes” to question “Is your project part of a larger initiative which is already registered with NCBI?”:
 - enter BioProject accession number PRJNA433975 if norovirus sequence data;
 - enter BioProject accession number PRJNA433976 if hepatitis A virus sequence data;
 - enter BioProject accession number PRJNA433977 if sapovirus sequence data;
 - enter BioProject accession number PRJNA817226 if rotavirus sequence data;
 - enter BioProject accession number PRJNA817227 if astrovirus sequence data;
 - enter BioProject accession number PRJNA817228 if hepatitis E virus sequence data.
- You may leave other fields blank.

General Info

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

* Project title

Norovirus virus sequencing

* Public description

Next-generation sequencing of clinical samples as part of norovirus surveillance

Relevance

Medical

* Is your project part of a larger initiative which is already registered with NCBI?

- ☐ No ☒ Yes (not very common)

* Initiative description

ViroTrakr umbrella project for norovirus surveillance (FDA)

* BioProject accession

PRJNA433975

Note: If you are registering a project that is part of an initiative which is already registered in the BioProject database, then please tell us the existing BioProject accession and provide a general description of the larger initiative. For example, the ENCODE project and Human Microbiome project have several subprojects. This information is needed for project linking.

3.7 Leave "BioSample" tab bland, it will be created from a different submission portal.

3.8 Leave "Publications" tab blank or add relevant publications from your group.

3.9 Check your input in "Review and Submit" tab, you may edit if needed or click "submit" to complete your submission.

3.10 The BioProject accession number "PRJNAxxxxxx" will be available within a few minutes on "my submission" page. Meanwhile, you will receive an NCBI email containing these accession numbers, usually within 12 hours.

Creating BioSamples at NCBI

- Submission of reads to SRA for sequence data and associated sequence metadata to BioSample for sample metadata in a single step.

Please refer to “NCBI submission protocol for microbial pathogen surveillance” [NCBI submission protocol for microbial pathogen surveillance \(protocols.io\)](https://protocols.io/ncbi-submission-protocol-for-microbial-pathogen-surveillance) step 2 for details.

4.1 [Log into your NCBI account at Submissions | Sequence Read Archive \(SRA\) | Submission Portal \(nih.gov\)](#);

Establish a SRA submission by clicking “New Submission”:



The screenshot shows the NCBI Submission Portal interface. At the top, there is a navigation bar with links: Home, My submissions, Manage data, Groups, Templates, and My profile. Below this, the main heading is 'Sequence Read Archive (SRA)' with a 'New submission' button highlighted by a red arrow. Underneath, there are several expandable sections for options to preload data: 'Short description and brief instructions', 'Aspera browser plugin upload', 'Aspera command line and FTP upload options', 'Amazon S3 instructions', and 'Google Cloud Storage instructions'. At the bottom, there is a table of submissions. The first submission is titled 'Sequence reads of various foodborne viruses from food, Dec 08 '21' with a status of 'SRA: Processed' and a 'Manage data' button.

Submission	Title	Group	Status	Updated
SUB10780695	Sequence reads of various foodborne viruses from food, Dec 08 '21		✓ SRA: Processed SRR17168754 Download metadata file with SRA accessions View and manage my SRA submission data	Dec 08

There are initially five tabs under each SRA submission.

4.2 Populate “Submitter” tab: (a submission group is highly recommended for your laboratory, see details in step 2.3)

Sequence Read Archive (SRA) submission: SUB11436659

New

1 SUBMITTER 2 GENERAL INFO 3 SRA METADATA 4 FILES 5 REVIEW & SUBMIT

Submitter

★ First (given) name	Middle name	★ Last (family) name
<input type="text" value="Zhihui"/>	<input type="text"/>	<input type="text" value="Yang"/>
★ Email (primary)	Email (secondary)	
<input type="text" value="zhihui.yang@fda.hhs.gov"/>	<input type="text" value="zhihui.yang@fda.hhs.gov"/>	

At least one email should be from the organization's domain.

Group for this submission

☐ No group (affiliation from my personal profile)

☐ 1 member Zhihui Yang's shared submissions

☒ 2 members CFSAN OARSA foodborne pathogen submission group

★ Submitting organization	Submitting organization URL	★ Department		
<input type="text" value="FDA, CFSAN, OARSA"/>	<input type="text"/>	<input type="text" value="DMB"/>		
Phone	Fax			
<input type="text"/>	<input type="text"/>			
★ Street	★ City	★ State/Province	★ Postal code	★ Country
<input type="text" value="8301 Muirkirk Rd"/>	<input type="text" value="Laurel"/>	<input type="text" value="MD"/>	<input type="text" value="20708"/>	<input type="text" value="USA"/>

4.3 Populate "General info" tab:

1 SUBMITTER 2 GENERAL INFO 3 SRA METADATA 4 FILES 5 REVIEW & SUBMIT

General Information

BioProject

BioProject describes the goal of your research effort.

★ Did you already register a BioProject for this research, e.g. for the submission of the reads to SRA and/or of the genome to GenBank?

☒ Yes ☐ No

★ Existing Project

BioSample

The BioSample records the detailed biological and physical properties of the sample that was sequenced. A BioSample can be used in more than one BioProject since it should be used for all the data that were obtained from that sample. Usually SRA data sets are generated from more than one sample.

★ Did you already register a BioSample for this sample, e.g. for the submission of the reads to SRA and/or of the genome to GenBank?

☐ Yes ☒ No

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

★ Projected release date

- Click “Yes” Under BioProject, put in the BioProject accession number established in step 2.

- Click “No” Under BioSample, to indicate you do not have an existing BioSample to associate with this sequence data and will create the BioSample on one of the next steps.

- * Click “Release immediately following processing” or specify a date to release if preferred.

Note: this is important for your first submission especially for data from clinical samples. To protect subject privacy, removal of any human genomics reads from your raw sequencing data could be done with the automated human-read scrubbing tool available in NCBI <https://github.com/ncbi/sra-human-scrubber>. To do so, along with the first data submission, a flag can be set for that BioProject indicating this first data submission and subsequent data submissions for that BioProject will get automatically scrubbed. Specifically, choose “Release on specified date” on your first submission, you may enter a date one week in the future (or longer and you are able to change the date later), meanwhile send the following email to sra@ncbi.nlm.nih.gov as soon as possible:

“Hi SRA help desk,
Please add the human read scrubbing analysis flag to my
BioProject PRJNAXXXXXX, then release my HUPed (delayed release) SRA
submissions.
Thanks,
Your name”

Once the flag is set for that BioProject, you may click “Release immediately following processing” for subsequent data submissions.

- Click “Continue” to next page.

4.4 Populate “BioSampleType” tab:

Preview BioSampleTypes and Attributes on the template page, and select the package that best describes your samples (e.g., select “Pathogen” and “Pathogen: clinical or host-associated” type):

1 SUBMITTER 2 GENERAL INFO 3 BIOSAMPLE TYPE 4 BIOSAMPLE ATTRIBUTES 5 SRA METADATA 6 FILES 7 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.
- ☐ **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.
 - ☒ **Pathogen: clinical or host-associated**
 - ☐ **Pathogen: environmental/food/other**
 - ☐ **Combined pathogen submission**

GSC *MiX*S 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](#).
- ☐ **MIGS Eukaryotic**
Use for eukaryotic genomic sequences. Organism must have lineage [Eukaryota](#).
- ☐ **MIGS Viral**
Use for virus genomic sequences. Organism must have lineage [Viruses](#).
- ☐ **MIMAG Metagenome-assembled Genome**
Use for metagenome-assembled genome sequences produced using computational binning tools that group sequences into individual organism genome assemblies starting from metagenomic data sets. Organism cannot contain the term 'metagenome'. Use the [MIUVIG](#) package for virus genomes. Before creating BioSamples for prokaryotic and eukaryotic MAGs, please read and follow the MAG submission instructions at <https://www.ncbi.nlm.nih.gov/genbank/wgsfaq/#metagen>.

* Note: two additional tabs "BioSample type" and "BioSample attributes" were added to collect information for BioSample creation.

4.5 Click "Continue" at the bottom of "Sample type" tab:

☐ **Microbe**
Use for bacteria or other unicellular microbes when it is not appropriate or advantageous to use [MixS](#), Pathogen or Virus packages.
 ☐ **Model organism or animal**
Use for multicellular samples or cell lines derived from common laboratory model organisms, e.g., mouse, rat, Drosophila, worm, fish, frog, or large mammals including zoo and farm animals.
 ☐ **Metagenome or environmental**
Use for metagenomic and environmental samples when it is not appropriate or advantageous to use [MixS](#) packages.
 ☐ **Invertebrate**
Use for any invertebrate sample.
 ☐ **Human**
WARNING: Only use for human samples or cell lines that have no privacy concerns. For all studies involving human subjects, it is the submitter's responsibility to ensure that the information supplied protects participant privacy in accordance with all applicable laws, regulations and institutional policies. Make sure to remove any direct personal identifiers from your submission. If there are patient privacy concerns regarding making data fully public, please submit samples and data to NCBI's [dbGaP](#) database. [dbGaP](#) has controlled access mechanisms and is an appropriate resource for hosting sensitive patient data. For samples isolated from humans use the Pathogen, Microbe or appropriate [MixS](#) package.
 ☐ **Plant**
Use for any plant sample or cell line.
 ☐ **Viral**
Use for all virus samples not directly associated with disease. Viral pathogens should be submitted using the Pathogen: Clinical or host-associated pathogen package.
 ☐ **Beta-lactamase**
Use for beta-lactamase gene transformants that have sequence and antibiotic resistance data. Please use the 'Supplementary Files' wizard to submit corresponding Sequin and Antibigram files.

specimens. Organism cannot contain the term 'metagenome'.
 ☐ **MIMARKS Survey related**
Use for any type of marker gene sequences, eg, 16S, 18S, 23S, 28S rRNA or COI obtained directly from the environment, without culturing or identification of the organisms. Organism must be a metagenome, where lineage starts with [unclassified sequences](#) and scientific name ends with 'metagenome'.
 ☐ **MIMS Environmental/Metagenome**
Use for environmental and metagenome sequences. Organism must be a metagenome, where lineage starts with [unclassified sequences](#) and scientific name ends with 'metagenome'.
 ☐ **MISAG Single Amplified Genome**
Use for single amplified genome sequences produced by isolating individual cells, amplifying the genome of each cell using whole genome amplification, and then sequencing the amplified DNA. Organism cannot contain the term 'metagenome'.
 ☐ **MIUVIG Uncultivated Virus Genome**
Use for uncultivated virus genome identified in metagenome and metatranscriptome datasets. Organism must have lineage [Viruses](#).

Continue

4.6 Populate "BioSample attributes" tab:

- Click "Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples.
- Click "Download Excel" button under "Attributes file".
- Fill out the downloaded BioSample attributes sheet in excel file, save it in a local folder.
- Click "Choose file" button under "Attributes file", then upload the populated and saved attributes sheet.
- At bottom of the "Attributes" tab, click "Continue":

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

Attributes Required fields are marked with * asterisk

Package Pathogen: clinical or host-associated; version 1.0

★ 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

★ Attributes file

or drag and drop it here

Download Excel or download TSV
For column definitions and examples, please see the sample attributes page.
For more information, please see creating sample attribute file.

Antibiogram – pathogen MIC (optional)

or drag and drop it here

Download Antibigram Excel template
Once the spreadsheet is completed, save as a Text (Tab-delimited) file, and upload it.
For more information, please see providing pathogen MIC antibiogram data.

Antibiogram – mycobacterial non-MIC (optional)

or drag and drop it here

Download Antibigram Excel template
Once the spreadsheet is completed, save as a Text (Tab-delimited) file, and upload it.
For more information, please see providing mycobacterial non-MIC antibiogram data.

4.7 Populate “SRA metadata” tab:

- Click “Upload a file using Excel or text format (tab-delimited)”.
- Click “Download Excel spreadsheet” under “Metadata file”.
- Read instructions under first tab (Contact Info and Instructions) on how to fill up the spreadsheet (next page), fill out the second tab (SRA_data) and save it as a TSV (tab-delimited) file in a local folder.
- Click “Choose file” button under “Metadata file”, then upload the populated spreadsheet that saved in your local folder.
- At the bottom of the “SRA metadata” tab, click “Continue”:

1 SUBMITTER 2 GENERAL INFO 3 BIOSAMPLE TYPE 4 BIOSAMPLE ATTRIBUTES 5 SRA METADATA 6 FILES 7 REVIEW & SUBMIT

SRA Metadata

For more detailed help with SRA submission please read the [SRA Submission Wizard Help](#).

★ How do you want to provide your metadata?

☐ Use built-in table editor

☒ Upload a file using Excel or text format (tab-delimited)

★ Metadata file

or drag and drop it here

Download Excel spreadsheet (designed to make it easier to select the correct metadata values), edit, save and then upload the modified Excel file.

4.7.1: The downloaded metadata excel spreadsheet:

Instructions:

Each column that has a red triangle in the upper-right corner has a comment that can be displayed if you hover over the header.

Some column headers have hyperlinks to NCBI webpages.

The YELLOW columns have drop-down menus that allow you to select from a controlled vocabulary. Once specified for one row, these values can be copied-and-pasted down.

You must save spreadsheet under second tab (SRA_data) as a TSV (tab-delimited file) to upload the TSV file for the SRA metadata tab.

If you created samples previously, provide accessions in the form of **SAMN#** in the column *sample_accession*.

Each row in the template represents a sequencing library with a unique combination of *sample* + *library* + *sequencing strategy* + *layout* + *instrument model*. Each row should have a unique *library_id* that is short and meaningful (like an ID you might use in lab).

When libraries are indeed identical (same combination of *sample* + *library* + *strategy* + *layout* + *instrument model*), all files should be placed in the same row. To do this simply enter the file names consecutively in the same row by adding more columns with headers *filename2*, *filename3*, etc.... PAIRED files must always be listed in the same row.

Provide **exact file names** (including extensions) in the *filename* columns.

File names must be **unique**.

red triangles indicate pop-up comments for that field required for ALL data types required for aligned data paired-end data only

Many of the columns also have data checks - if you received a warning, please verify that you have entered a value from the drop-down menu. Example Below.

NOTE: There are data checks and auto-complete features in this spreadsheet that are not compatible with Libre- and Open-Office. If you use one of these suites, please manually consult the platform and instrument information on the last page.

platform (click for details) instrument_model design_description

LS454

454 GS

454 GS 20

454 GS FLX

454 GS FLX+

Contact Info and Instructions SRA_data Library and Platform Terms

Note: You must save the spreadsheet under the second tab (SRA_data) as a TSV (tab-delimited file) to upload the TSV file for the SRA metadata tab.

4.7.2: Example: fill out the metadata excel spreadsheet:

sample_accession	library_id	title	library_strategy	library_source	library_selection	library_layout	platform	instrument_model	design_description	filetype	filename
1											
2	SAMN13991854	E7951	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
3	SAMN13991855	E7952	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
4	SAMN13991856	E7953	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
5	SAMN13991857	E7954	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
6	SAMN13991858	E7955	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
7	SAMN13991859	E8051	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
8	SAMN13991860	E8052	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
9	SAMN13991861	E8053	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
10	SAMN13991862	E7751	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
11	SAMN13991863	E8054	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
12	SAMN13991864	E7752	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
13	SAMN13991865	E7956	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
14	SAMN13991866	E7957	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
15	SAMN13991867	E8351	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
16	SAMN13991868	E8352	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
17	SAMN13991869	E8353	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
18	SAMN13991870	E8354	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
19	SAMN13991871	E8355	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
20	SAMN13991872	E8356	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
21	SAMN13991873	E8055	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
22	SAMN13991874	E8056	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
23	SAMN13991875	E8057	norovirus outbreak in California	WGS	METAGENOMIC	RANDOM	paired	ILLUMINA	Illumina MiSeq	TruSeq	fastq
24											
25											
26											
27											
28											
29											
30											
31											
32											
33											
34											
35											
36											
37											
38											
39											
40											

4.7.2: Example: fill out the metadata excel spreadsheet (Cont.):

Get the file name of each fastq file and fill out the column:

Name	Date modified	Type	Size
E7751_S1_L001_R1_001.fastq.gz	10/4/2019 5:07 AM	WinZip File	162,929 KB
E7751_S1_L001_R2_001.fastq.gz	10/4/2019 5:07 AM	WinZip File	168,056 KB
E7752_S2_L001_R1_001.fastq.gz	10/4/2019 5:07 AM	WinZip File	162,894 KB
E7752_S2_L001_R2_001.fastq.gz	10/4/2019 5:07 AM	WinZip File	169,028 KB
E7951_S1_L001_R1_001.fastq.gz	10/29/2019 4:29 AM	WinZip File	380,133 KB
E7951_S1_L001_R2_001.fastq.gz	10/29/2019 4:29 AM	WinZip File	394,378 KB
E7952_S2_L001_R1_001.fastq.gz	10/29/2019 4:29 AM	WinZip File	171,481 KB
E7952_S2_L001_R2_001.fastq.gz	10/29/2019 4:29 AM	WinZip File	177,415 KB
E7953_S3_L001_R1_001.fastq.gz	10/29/2019 4:29 AM	WinZip File	217,973 KB
E7953_S3_L001_R2_001.fastq.gz	10/29/2019 4:29 AM	WinZip File	227,443 KB

L39																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

4.8 Populate "Files" tab:

- Click "FTP or Aspera Command Line file preload".
- Click "FTP upload Instructions".

1 SUBMITTER
2 GENERAL INFO
3 BIOSAMPLE TYPE
4 BIOSAMPLE ATTRIBUTES
5 SRA METADATA
6 FILES
7 REVIEW & SUBMIT

Files

- Each file must be listed in the SRA metadata table you uploaded. If you are uploading a tar archive, list each file name, not the archive name.
- Unique file names that do not contain any sensitive information should be used for all files. File names as submitted appear publicly when data is retrieved from the cloud.
- Files can be compressed using **gzip** or **bzip2**, and may be submitted in a **tar** archive, but archiving or compressing your files is not required. **Do not use zip!**

★ How do you want to provide files for this submission?

☐ Web browser upload via HTTP or Aspera Connect plugin
Do not use web browser HTTP upload if you are uploading files over 10 GB or more than 300 files.

☒ **FTP or Aspera Command Line file preload**
Files for a submission must be uploaded into a single folder.

☐ AWS or GCP bucket

Select preload folder Preload folder not selected

Aspera command line upload instructions +

FTP upload instructions +

☐ Autofinish submission

Continue

4.8.1: Read and follow "FTP upload Instructions". Select a proper FTP tool (e.g., FileZilla) to upload your data:

Select preload folder Preload folder not selected

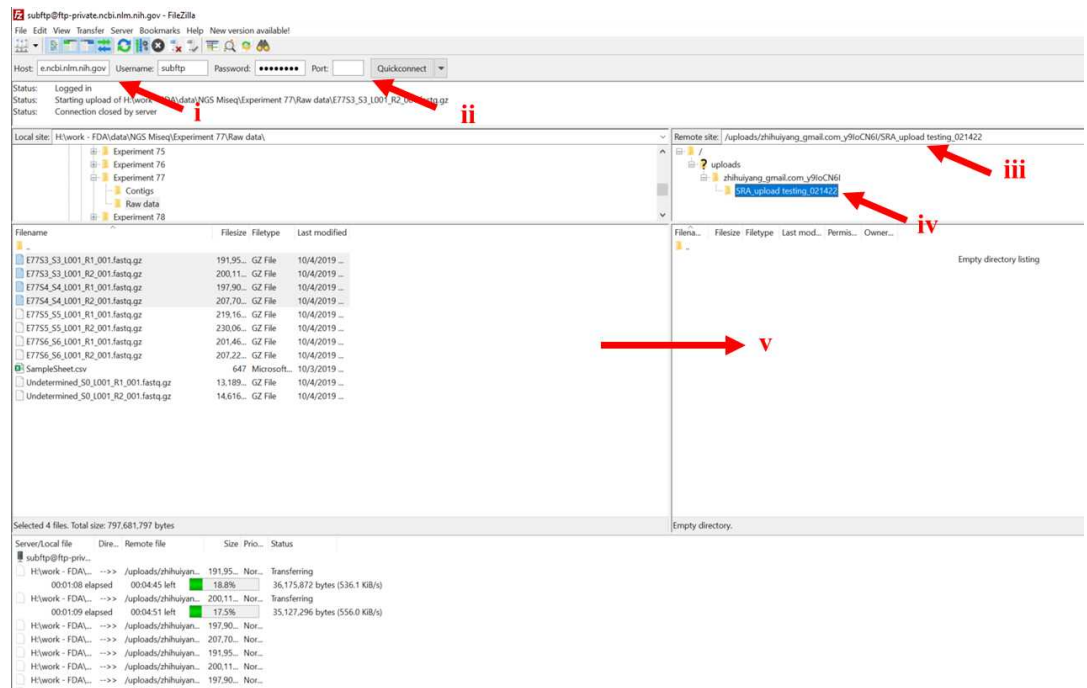
Aspera command line upload instructions +

FTP upload instructions -

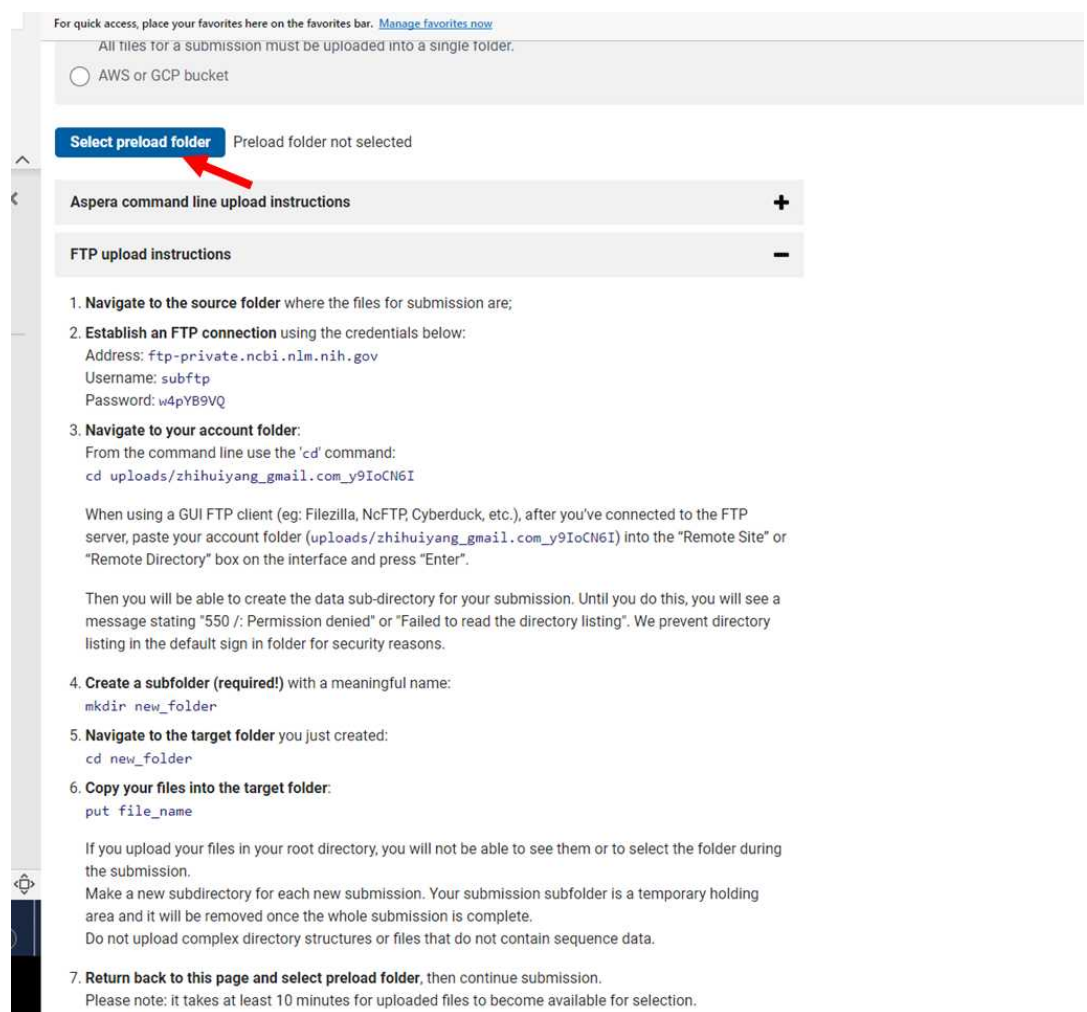
1. **Navigate to the source folder** where the files for submission are;
2. **Establish an FTP connection** using the credentials below:
 Address: ftp-private.ncbi.nlm.nih.gov
 Username: subftp
 Password: w4pYB9VQ
3. **Navigate to your account folder**.
 From the command line use the 'cd' command:
 cd uploads/zhihuiyang_gmail.com_y9IoCN6I
 When using a GUI FTP client (eg: FileZilla, NcFTP, Cyberduck, etc.), after you've connected to the FTP server, paste your account folder (uploads/zhihuiyang_gmail.com_y9IoCN6I) to the "Remote Site" or "Remote Directory" box on the interface and press "Enter".
 Then you will be able to create the data sub-directory for your submission. Until you do this, you will see a message stating "550 /: Permission denied" or "Failed to read the directory listing". We prevent directory listing in the default sign in folder for security reasons.
4. **Create a subfolder (required!)** with a meaningful name:
 mkdir new_folder
5. **Navigate to the target folder** you just created:
 cd new_folder
6. **Copy your files into the target folder**:
 put file_name
 If you upload your files in your root directory, you will not be able to see them or to select the folder during the submission.
 Make a new subdirectory for each new submission. Your submission subfolder is a temporary holding area and it will be removed once the whole submission is complete.
 Do not upload complex directory structures or files that do not contain sequence data.
7. **Return back to this page and select preload folder**, then continue submission.
 Please note: it takes at least 10 minutes for uploaded files to become available for selection.
8. Please complete your submission within 30 days of creating a preload folder. If you upload files and do not submit them, they will be **automatically deleted 30 days after folder creation**.

4.8.2: Open FileZilla

- Copy and paste "Host, Username, Password" to establish FTP connection.
- Port: default for FTP is 21; default for SFTP is 22. Click "Quick connect".
- Copy and paste your directory name "uploads/. ...".
- Create a subfolder (required!) with a meaningful name.
- Start upload your sequence data from your local folder to the created subfolder.



4.8.3: when upload is completed, return to SRA submission page and click “Select preload folder”;



4.8.4: (note: it takes at least 10 minutes for uploaded files to become available)
click “continue” to upload:

1 SUBMITTER2 GENERAL INFO3 BIOSAMPLE TYPE4 BIOSAMPLE ATTRIBUTES5 SRA METADATA6 FILES7 REVIEW & SUBMIT

Files

- Each file must be listed in the SRA metadata table you uploaded. If you are uploading a tar archive, list each file name, not the archive name.
- Unique file names that do not contain any sensitive information should be used for all files. File names as submitted appear publicly when data is retrieved from the cloud.
- Files can be compressed using gzip or bzip2, and may be submitted in a tar archive, but archiving or compressing your files is not required. Do not use zip!

★ How do you want to provide files for this submission?

☐ Web browser upload via HTTP or Aspera Connect plugin
Do not use web browser HTTP upload if you are uploading files over 10 GB or more than 300 files.

☒ FTP or Aspera Command Line file preload
All files for a submission must be uploaded into a single folder.

☐ AWS or GCP bucket

Select preload folder

Preload folder not selected

Aspera command line upload instructions

+

FTP upload instructions

+

☐ Autofinish submission

Continue

4.9 Please review your submission, make necessary changes on any tab, then click the “Submit” button:

Review & Submit

This Sequence Read Archive (SRA) submission will be released **immediately following processing**.

Metadata file [SRA_metadata_testing_02142022.txt](#) (5.5 KB)

BioSample accession	Files
SAMN25859890	<ul style="list-style-type: none"> E77S3_S3_L001_R1_001.fastq.gz (fastq) E77S3_S3_L001_R2_001.fastq.gz (fastq)
SAMN25859891	<ul style="list-style-type: none"> E77S4_S4_L001_R1_001.fastq.gz (fastq) E77S4_S4_L001_R2_001.fastq.gz (fastq)

Submitter [Zhihui Yang](#)
zhihui.yang@fda.hhs.gov

Submitting organization [OARSA/CFSAN/FDA](#)

Submit

- 4.10 The SRA accession number “SRRxxxxxxx” will be available within a few minutes on “my submission” page. You may download the “metadata file with SRA accessions” for your record. Meanwhile, you will receive an NCBI email containing these accession numbers, usually within 12 hours.

Sequence Read Archive (SRA) [New submission](#)

Short description and brief instructions

The SRA accepts genetic data and the associated quality scores produced by next generation sequencing technologies. Please refer to the [File Format Guide](#).

- Files can be compressed using **gzip** or **bzip2**, and may be submitted in a tar archive but archiving and/or compressing your files is not required. **Do not use zip!**
- All file names must be **unique** and **not contain any sensitive information**. File names as submitted appear publicly in the Google and AWS clouds.
- Each file must be listed** in the SRA metadata table. If you are uploading a tar archive, list each file name, not the archive name.
- Use the **preload** option if you are uploading files over **10 GB** or more than **300** files. All files for a submission must be uploaded into a **single folder**. Please complete your submission within 30 days of creating a preload folder. If you upload files and do not submit them, they will be **automatically deleted 30 days after folder creation**.

Before submitting, read the [SRA Submission Wizard Help](#).

Contact sra@ncbi.nlm.nih.gov with any question or concern about your data or submission.

Options to preload data:

- [Aspera browser plugin upload](#)
- [Aspera Command-Line upload](#)
- [FTP upload](#)
- [Amazon S3 Instructions](#)
- [Google Cloud Storage Instructions](#)

13 submissions

Submission	Title	Group	Status	Updated
SUB11073971	Norwalk virus sequencing, Feb 11 '22		<div> <div>SRA: Processed</div> <div>(2 objects)</div> <div>Download metadata file with SRA accessions</div> <div>View and manage my SRA submission data</div> </div>	12:29

[Manage data](#)

5

Submission of the assembled data to GenBank.

Raw sequencing data is required as input for ViroTrakr database deposit and subsequent data analysis. Assemblies or consensus sequences are part of the data analysis in our workflow, its submission to GenBank is optional but encouraged. The GenBank submission of your assembled sequences will:

- “Make your sequence data available in the International Nucleotide Sequence Database Collaboration (INSDC) for global use;
- Ensure your data contribution is included in NCBI Virus, BLAST, RefSeq and other resources;
- Follow FAIR data-sharing principles.”

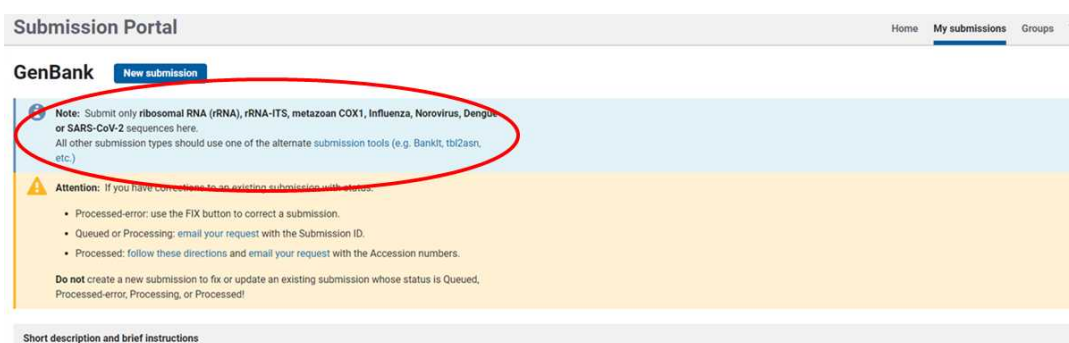
Reference:

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

Note: this GenBank submission requires and assumes that you already have established a BioProject and BioSample(s) from step 1 and step 2 of this protocol.

5.1 Log into your NCBI account at Submissions [GenBank | Submission Portal \(nih.gov\)](#);

Establish a GenBank submission by clicking “New Submission”:



Note: Submission of norovirus assemblies can be directly made at the submission portal, all other submission types could use one of the alternate submission tools (such as [BankIt](#), [tbl2asn](#)) with similar submission steps.

5.2 There are nine tabs under each GenBank submission.

Populate “Submitter” tab:

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

★ 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

★ Which genogroup do these Norovirus sequences represent?

(You will provide the complete genotype later in the submission process.)

☐ GI

☒ GII

☐ GIII

☐ GIV

☐ GV

☐ GVI

☐ other

Review requirements for norovirus submissions

If none of the options above describe your sequences, use Bankit to submit.

Submission title (Optional, not displayed in final records)

Continue

5.3 Populate “Submitter” tab: (a submission group is highly recommended for your laboratory, see details in step 2.3)

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

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

☐ 1 member Zhihui Yang's shared submissions (edit group)

☒ 2 members CFSAN OARSA foodborne pathogen submission group (edit group)
Charlie Seto, you

★ Submitting organization

★ Department

★ Street

★ City

★ State/Province

★ Postal code

★ Country

Contact information

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

★ Email (primary)

★ Email (secondary)

Please provide an alternate email address to ensure that messages are received

★ First (given) name

Middle name

★ Last (family) name

Phone

Fax

Continue ☒ Update my contact information in profile

5.4 Populate “Sequencing technology” tab:

- Choose the method used to obtain these sequences;
- Click “Assembled sequences”;
- Fill in the Assembly information (Assembly program and version/date);
- Click “Continue” to next page.

The screenshot shows the 'Sequencing Technology' tab, which is the third step in a nine-step process. The steps are: 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 'Sequencing Technology' section contains three main parts: 'Method', 'Assembly state', and 'Assembly info'. The 'Method' section asks 'What methods were used to obtain these sequences?' and lists options: Sanger dideoxy sequencing, 454, Helicos, Illumina (selected), IonTorrent, Pacific Biosciences, SOLiD, and Other. The 'Assembly state' section asks 'These sequences are:' and lists options: Unassembled sequence reads and Assembled sequences (each sequence was assembled from two or more overlapping sequence reads) (selected). The 'Assembly info' section has a header 'Name of algorithm, e.g. Newbler or SOAPdenovo' and a table with columns 'Assembly program', 'Version or date', and 'Delete'. There is one row with empty input fields for the first two columns and a red delete button. Below the table is a link 'Add another assembly program'. At the bottom of the form is a blue 'Continue' button.

5.5 Populate “Sequences” tab:

- Click “Release immediately following processing” or specify a data to release if preferred.
- Upload a prepared nucleotide FASTA file by clicking “Choose file”.
- Click “Continue” to next page.

1 SUBMISSION TYPE 2 SUBMITTER 3 SEQUENCING TECHNOLOGY 4 SEQUENCES 5 SEQUENCE PROCESSING 6 SOURCE INFO 7 SOURCE MODIFIERS 8 REFERENCES 9 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](#) 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
aacgatatagagatgtagcggatagagagagga
>Seq2
gtacgataaagagatgtagcggatagagagagga
```

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](#)

Notes:

- Organize your sequence files by type or locus and make one submission for each type.
- Plain text (.txt) nucleotide FASTA files are accepted.
- Use a text editor (for example: Notepad or WordPad) to prepare a file containing the set of nucleotide sequences in FASTA format and save the file as plain text or text.
- You may use the strain, isolate, specimen-voucher, or clone IDs as the sequence_IDs in your FASTA file. If you do this, do not include extra information in the sequence_IDs such as the organism name, etc.

For more information on how to format and organize FASTA files, please see [FASTA file help](#).

- FASTA file example:

```
>SeqID1
CCTTTATCTAATCTTTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCGTGCAGAACTTGGACAAC
TGCAGAACTTGGACAACCTTTATCTAATCTTTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCG
TTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCGTGCAGAACTTGGACAACCTTTATCTAATCT
ATAGTTCAGCCTCCTCATCCGTGCAGAACTTGGACAACCTTTATCTAATCTTTGGAGCATGAGCTGGC
Hard return
>SeqID2
CCTTTATCTAATCTTTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCGTGCAGAACTTGGACAAC
TGCAGAACTTGGACAACCTTTATCTAATCTTTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCG
TTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCGTGCAGAACTTGGACAACCTTTATCTAATCT
ATAGTTGGAACCGCCCTCAGCCTCCTCATCCGTGCAGAACTTGGACAACCTTTATCTAATCTTTGGAGCATGAGCTGGC
```

5.6 Populate “Source Info” tab:

- You may find more information on the question “Do your sequence IDs represent one of these?” by clicking “description of these fields”.
- Click “None of these” if your sequence IDs don’t contain information as described.
- Click “Continue” to next page.

The screenshot shows the 'Source Information' tab, which is the fifth step in the submission process. The progress bar at the top indicates the following steps: 1 SUBMISSION TYPE, 2 SUBMITTER, 3 SEQUENCING TECHNOLOGY, 4 SEQUENCES, 5 SOURCE INFO (current), 6 SOURCE MODIFIERS, 7 REFERENCES, and 8 REVIEW & SUBMIT. The main heading is 'Source Information'. Below it, a message states: 'The first few sequence IDs that we found are: E8351_trimmed_contig_34, E8051_mapping_consensus_7501_bp'. A question follows: '★ Do your sequence IDs represent one of these?'. There are two radio button options: 'Isolate' and 'NONE of these'. The 'NONE of these' option is selected. To the right of the radio buttons, a help icon and text explain: '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.' At the bottom, there is a blue 'Continue' button.

5.7 Populate “Source modifiers” tab:

- Click the “Upload a tab-delimited table (template file provided)” button.

The screenshot shows the 'Source Modifiers' tab, which is the sixth step in the submission process. The progress bar at the top indicates the following steps: 1 SUBMISSION TYPE, 2 SUBMITTER, 3 SEQUENCING TECHNOLOGY, 4 SEQUENCES, 5 SOURCE INFO, 6 SOURCE MODIFIERS (current), 7 REFERENCES, and 8 REVIEW & SUBMIT. The main heading is 'Source Modifiers'. Below it, a message states: 'For each sequence, GenBank requires the following source information: collection-date, country, genotype, host, isolate, and isolation-source.' A section titled 'Current source modifiers - what you have provided so far' is empty. Below this, a help icon and text explain: 'More help: what is a source modifier, description of each modifier, how to provide source modifiers.' A message states: 'If you have already provided all the required information, you can press Continue to proceed.' A question follows: '★ How do you want to apply source modifiers?'. There are two radio button options: 'Use an editable table' and 'Upload a tab-delimited table (template file provided)'. The 'Upload a tab-delimited table (template file provided)' option is selected, and a red arrow points to it. At the bottom, there is a blue 'Continue' button.

4.7:Populate “Source modifiers” tab: (cont.)

Source Modifiers

For each sequence, GenBank requires the following source information:

- collection-date,
- country,
- genotype,
- host,
- isolate, and
- isolation-source.

Current source modifiers - what you have provided so far

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

[See an example Source Modifiers table](#)

 Choose file or drag and drop it here

5. Click Continue to validate the information and follow the instructions.

Continue

* Notes:

• GenBank source modifier template:

Below is a custom version containing direct linkage to the respective BioSample and BioProject records. Populate the template as guided and save it in a txt format.

 [GenBank submission modifiers_ViroTrakr.xlsx](#)

• Click “Choose file” to upload your saved source modifier file.

• Click “Continue” to next page.

Technical Assistance

6 Technical Assistance:

If you are having trouble finalizing your submission, contact the relevant NCBI database for assistance and include your submission ID in the email subject (SUB#####):

BioProject (for any BioProject issues): bioprojecthelp@ncbi.nlm.nih.gov

BioSample (for issues on source metadata): biosamplehelp@ncbi.nlm.nih.gov

SRA (for issues on raw sequencing data): sra@ncbi.nlm.nih.gov

GenBank (for issues on assembled sequences): gb-admin@ncbi.nlm.nih.gov

GenomeTrakr: genomeTrakr@fda.hhs.gov

ViroTrakr: ViroTrakr@fda.hhs.gov

NCBI help desk and account issues: info@ncbi.nlm.nih.gov