

Dec 16, 2025

Version 2

NCBI ViroTrakr submission protocol for foodborne virus surveillance_version 2 V.2

DOI

<https://dx.doi.org/10.17504/protocols.io.j8nlkkdbxl5r/v2>

Zhihui Yang¹, Ruth Timme², Maria Balkey²

¹FDA/HFP/OLOAS/OAMT; ²FDA/HFP/OLOAS/OSCCS

GenomeTrakr

ViroTrakr



Zhihui Yang

FDA

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.j8nlkkdbxl5r/v2>

External link: [http://ViroTrakr: foodborne viruses \(ID 396739\) - BioProject - NCBI \(nih.gov\)](http://ViroTrakr: foodborne viruses (ID 396739) - BioProject - NCBI (nih.gov))

Protocol Citation: Zhihui Yang, Ruth Timme, Maria Balkey 2025. NCBI ViroTrakr submission protocol for foodborne virus surveillance_version 2. protocols.io <https://dx.doi.org/10.17504/protocols.io.j8nlkkdbxl5r/v2> Version created by **Zhihui Yang**

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

Protocol status: Working

We use this protocol and it's working

Created: December 15, 2025

Last Modified: December 16, 2025

Protocol Integer ID: 235061

Keywords: Foodborne virus surveillance, NCBI submission protocol, ncbi submission protocol for microbial p..., submission protocol for foodborne virus ..., foodborne virus surveillance introductio..., foodborne virus, bacterial pathogen surveillance in multi..., bacterial pathogen surveillance, microbial pathogen surveillance, virotrakr contributor, international genomic reference database, virotrakr, pathogen, norovirus, new bioproject at ncbi, foodborne illness, data to ncbi, raw sequence data, genometrakr, sapovirus, genomic database, raw sequence data to the sra database, data to genbank, time reference sequences for phylogenet..., ncbi submission protocol, ncbi submission, new bioproject, ncbi, phylogenetic analysis, biosample number, ViroTrakr, Foodborne viruses, NCBI submission protocol, Foodborne virus surveillance, ncbi virotrakr submission protocol for foodborne virus surveillance-version, foodborne viral pathogen surveillance, ncbi submission protocol for microbial pathogen surveillance, foodborne virus surveillance-version, foodborne v



Abstract

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 sequences to GenBank and link them to ViroTrakr.

ViroTrakr, a genomic database initiated by FDA 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. Data analysis pipelines have been established for norovirus and Hepatitis A virus and are under development for other species. Similar to GenomeTrakr for bacterial pathogen surveillance, ViroTrakr is expected to be employed for foodborne viral pathogen surveillance.

ViroTrakr:

foodborne viruses (ID 396739) - BioProject - NCBI (nih.gov)

GenomeTrakr:

Multispecies (ID 593772) - BioProject - NCBI (nih.gov)

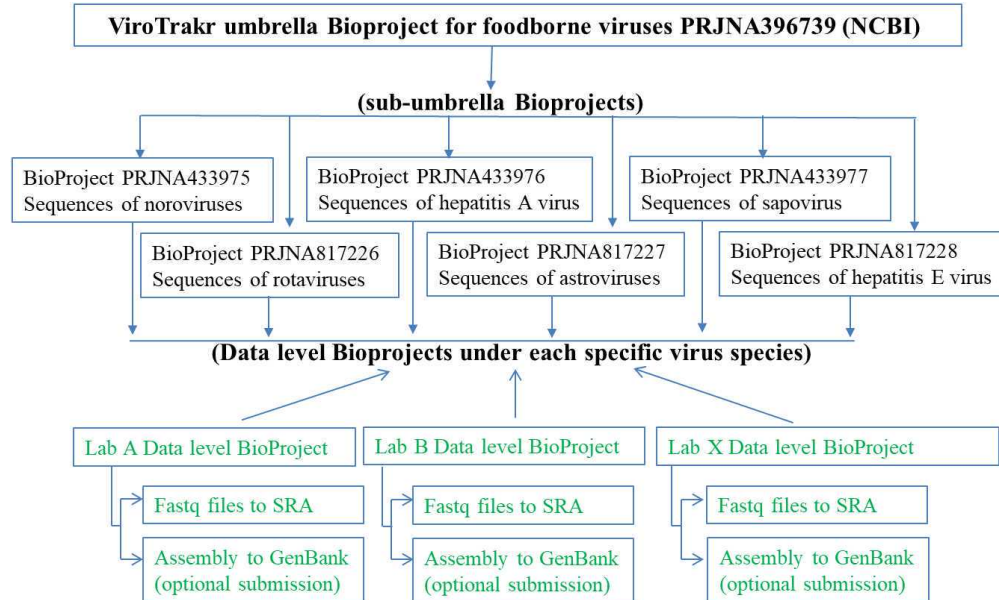
Reference: NCBI submission protocol for microbial pathogen surveillance V.10:

NCBI submission protocol for microbial pathogen surveillance (protocols.io)

ViroTrakr data structure

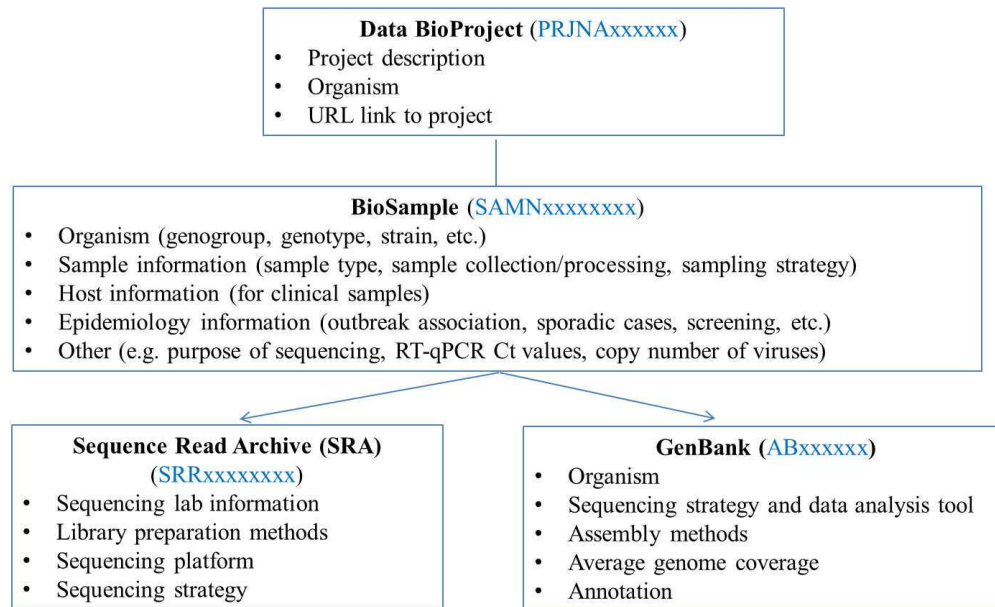
1

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

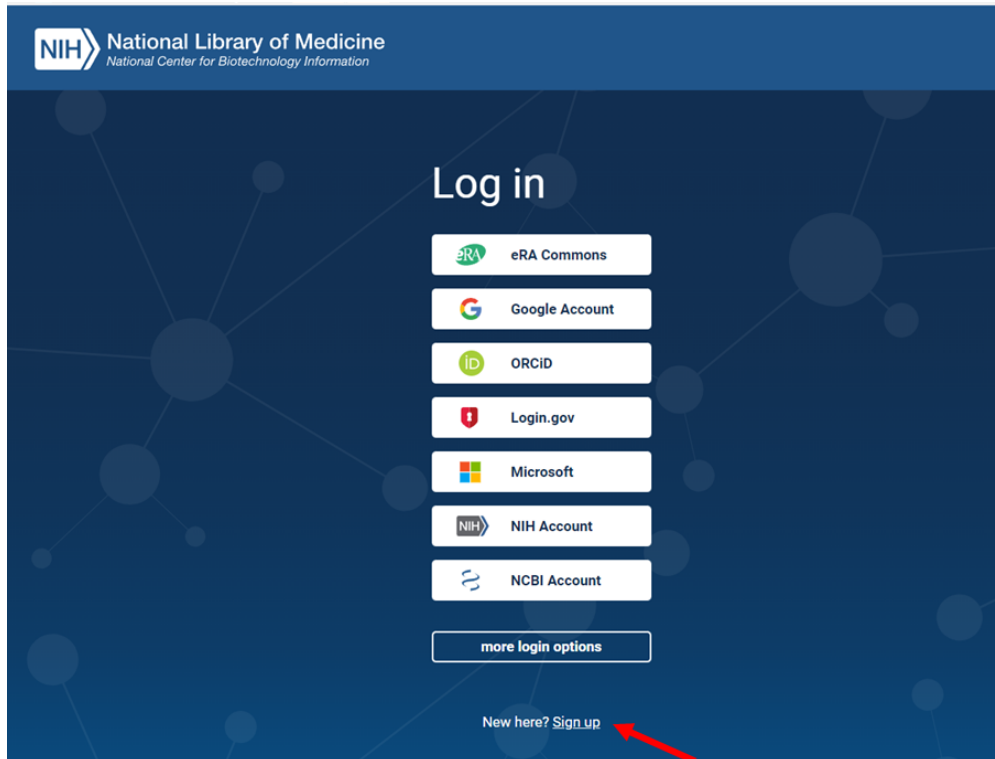
2

Getting started

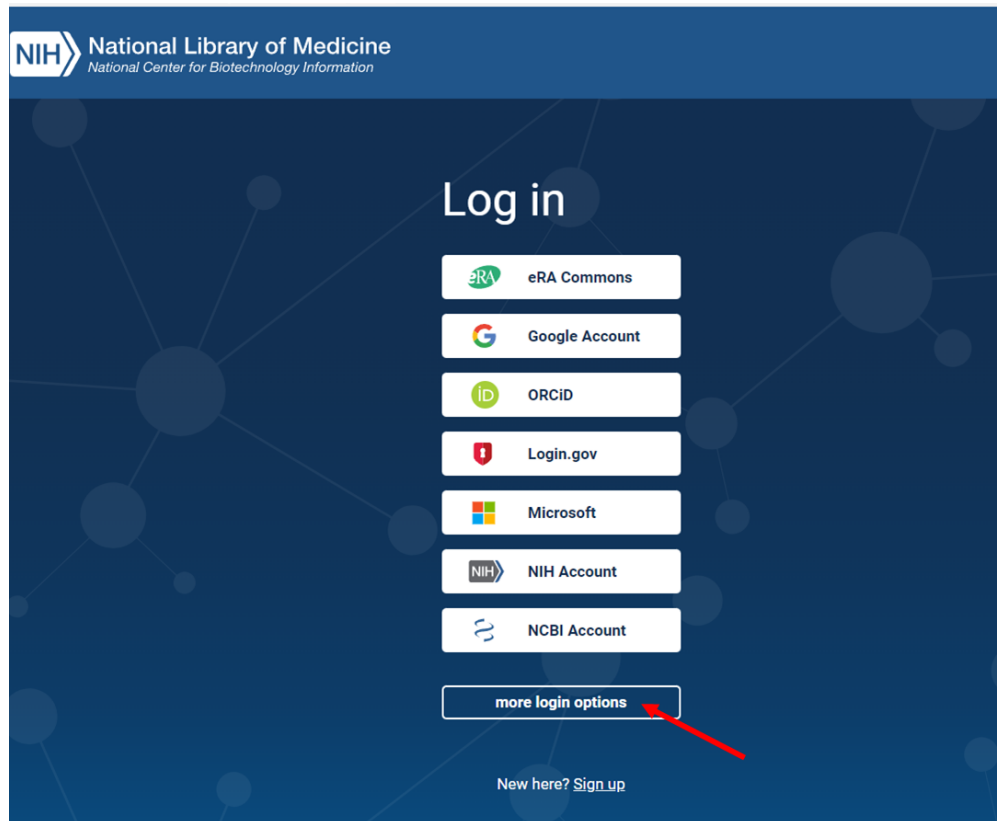
Note: if you are an existing NCBI user, you may skip this step.

Please refer to “NCBI submission protocol for microbial pathogen surveillance” Section 1 **NCBI submission protocol for microbial pathogen surveillance (protocols.io)** 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).



- 2.2 For existing users, sign in to your NCBI account on the 3rd party's sign in page: [Log in / NCBI \(nih.gov\)](#)



For existing users (cont.), click "more login options" to find the 3rd party that you will sign in with:



Find your login provider

start typing ...

Close

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

2

29 Mayıs University, Turkey

A

A*STAR - Agency for Science, Technology and Research, Singapore

A. T. Still University

AAF Virtual Home, Australia

aai.lab.maeen.sa, Saudi Arabia

AAI@EduHr Single Sign-On Service, Croatia

Aalborg University, Denmark

Aalto University, Finland

Aarhus School of Architecture, Denmark

Aarhus School of Marine and Technical Engineering, Denmark

Aarhus University, Denmark

AARNet, Australia

Aba Teachers University, China

ABC - Academia Brasileira de Ciencias, Brazil

Abertay University, United Kingdom

Aberystwyth University, United Kingdom

ABES - French Bibliographic Agency for Higher Education, France

Abingdon and Witney College, United Kingdom

Absalon University College, Denmark

Academia Cotopaxi - cotopaxi, Ecuador

ACADEMIA d.o.o., Slovenia

Academia das Ciências de Lisboa, Portugal

Academia Militar Portugal

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)** for details.

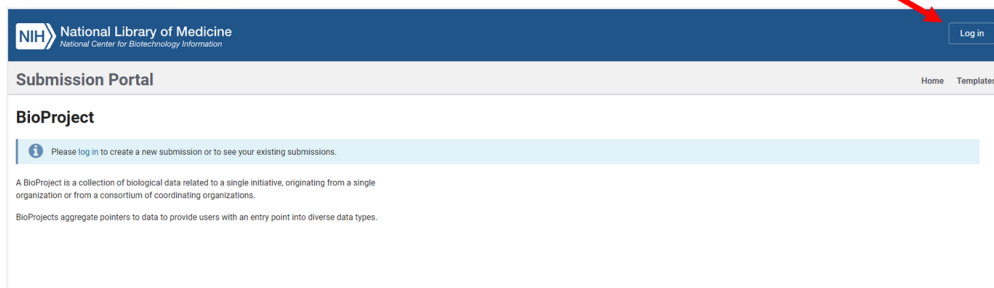
2.4 Log into your NCBI account and you are ready for your NCBI submission.

Creating BioProjects at NCBI

3 **Establish your new data level BioProject under the umbrella BioProject ViroTrakr:**
Note: this step will link your data submission to ViroTrakr.

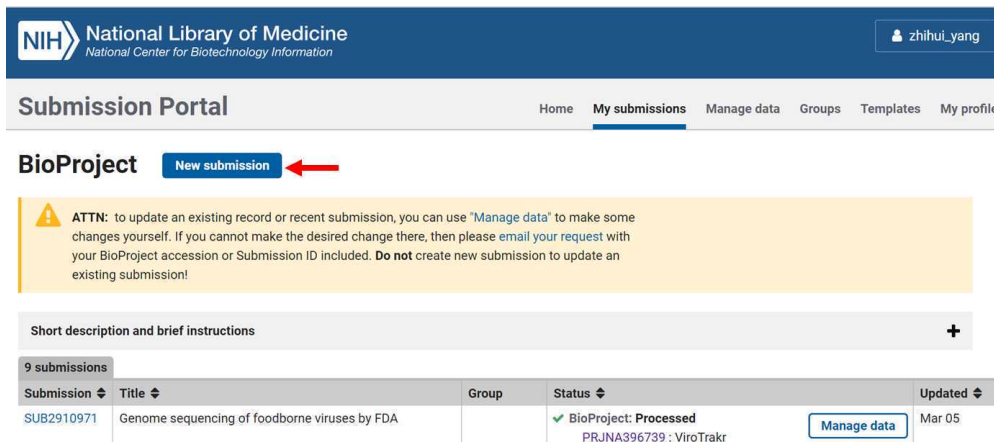
Please refer to “NCBI submission protocol for microbial pathogen surveillance” **NCBI submission protocol for microbial pathogen surveillance (protocols.io)** step 3 for details.

3.1 **Log into your NCBI account at Submissions | BioProject | Submission Portal (nih.gov):**



3.2

Establish a new BioProject by clicking "New submission":



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: SUB14290256 [Delete submission](#)

New

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

Submitter Required fields are marked with *

* 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 (edit group)
☒ 1 member CFSAN OARSA foodborne pathogen submission group (edit group) you

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

Phone Fax
8301 Muirkirk Rd

* 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 \(protocols.io\)](#):

“- 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 \(protocols.io\)](#) 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.

The screenshot shows the NCBI BioProject submission portal for a new submission (SUB14290256). The user is logged in as 'zhuhai_yang'. The 'Project Type' tab is selected, showing a list of project types with checkboxes. The 'Raw sequence reads' option is checked. A note indicates that required fields are marked with an asterisk.

NIH National Library of Medicine
National Center for Biotechnology Information

Submission Portal

Home My submissions Manage data Groups Templates My profile

BioProject submission: SUB14290256

New

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

Project Type

Required fields are marked with * asterisk

* 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 (Loc)
- ☐ Transcriptome or Gene expression
- ☐ Variation
- ☐ Other



NIH National Library of Medicine
National Center for Biotechnology Information

Submission Portal

Home My submissions Manage data Groups Templates My profile

BioProject submission: SUB14290256

New

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

Project Type

Required fields are marked with * asterisk

* 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.



Submission Portal

BioProject submission: SUB14290256

New

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

Target

The most descriptive organism name for the study (to the species, if relevant).

★ Organism name ?

Strain ?

Breed ?

Cultivar ?

Isolate name ?

Label ?

Description ?

Continue

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.



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

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

Norwalk virus sequencing

★ Public description

NGS of clinical samples as part of norovirus surveillance

3943 characters left

Relevance

Medical

★ Is your project part of an umbrella project which is already registered with NCBI?

☐ No ☒ Yes (not very common)

★ Umbrella BioProject description

This project is a component of the initiative umbrella bioproject ViroTrakr; this information is needed for project linking.

★ BioProject accession

PRJNA433975

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

External links

- 3.7 Leave "BioSample" tab blank, it will be created from a different submission portal (SRA submission).
- 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 and SRA submission

- 4 Note: to protect subject privacy for data submission of clinical samples, removal of any human genomics reads from the raw sequencing data is important. Two steps are included in our pipeline to confirm the absence of human reads from the submitted data:

First, the ViroTrakr BioProject has been flagged for automated human-read scrubbing by NCBI. With this setting, any data submission to this BioProject will automatically get scrubbed for human reads with the tool: <https://github.com/ncbi/sra-human-scrubber>.

Second, an additional step is included in our data analysis pipelines to automatically remove any human reads from the samples by mapping the reads to human hg38 chromosome and collect the unmapped (dehosting) reads for downstream analysis.

Please refer to "NCBI submission protocol for microbial pathogen surveillance" [NCBI submission protocol for microbial pathogen surveillance \(protocols.io\)](#) 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. At the top, there's a navigation bar with 'Home', 'My submissions', 'Manage data', 'Groups', 'Templates', and 'My pr'. Below this, the 'Sequence Read Archive (SRA)' section has a 'New submission' button highlighted with a red arrow. Underneath are several expandable sections: 'Short description and brief instructions', 'Options to preload data:', 'Aspera browser plugin upload', 'Aspera command line and FTP upload options', 'Amazon S3 instructions', and 'Google Cloud Storage instructions'. At the bottom, there's a table titled '12 submissions'. The first row is highlighted, showing a submission ID 'SUB10780695', a title 'Sequence reads of various foodborne viruses from food, Dec 08 21', a status 'SRA: Processed', and a 'Manage data' button. The status also includes 'SRR17168754' and links to 'Download metadata file with SRA accessions' and 'View and manage my SRA submission data'.

There are initially five tabs under each BioSample submission.



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

Sequence Read Archive (SRA) submission: SUB14290411

New

1 SUBMITTER 2 GENERAL INFO 3 SRA METADATA 4 FILES 5 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 ([edit group](#))
- ☒ 1 member CFSAN OARSA foodborne pathogen submission group ([edit group](#))
you

★ 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

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.

★ Do you already have a BioProject accession number for this research?

☒ Yes ☐ No (a BioProject will be created within this submission)

★ Existing BioProject

PRJNXXXXXX

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.

★ Do you already have BioSample accession numbers for these samples?

☐ Yes ☒ No (BioSamples will be created within this submission)

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

📘 Please allow 24-48 hours for propagation of the data to the NCBI SRA public site.

Continue

- Click “Yes” Under BioProject and enter the BioProject accession number established in step 3.

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

- * 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 sequencing raw data can 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 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 may change the date later), meanwhile send the following emailing to sra@ncbi.nlm.nih.gov as soon as possible:



Hi, SRA help desk,

Please add the human read scrubbing analysis flag to my BioProjectPRJNXXXXXX, 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 "BioSample Type" tab:

Preview BioSample Types and Attributers on the template page, and select the package that best describes your samples (e.g., select "One Health Enteric" package):



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.
- ☒ **One Health Enteric**
Use for microbial isolates that are collected for genomic surveillance of enteric pathogens. Sample spaces include the following: 1. human/animal hosts; 2. food samples; 3. food

GSC **MiXs** 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](#).

* Note: in addition to the initial 5 tabs, two additional tabs "BioSample type" and "BioSample attributes" were added to collect information for BioSample creation.

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 [MlxS](#), 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 [MlxS](#) 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 [MlxS](#) 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



4.5 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 BIOSAMPLE TYPE

4 BIOSAMPLE ATTRIBUTES

5 SRA METADATA

6 FILES

7 REVIEW & SUBMIT

Attributes

Package One Health Enteric; 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

1

★ Attributes file

or drag and drop it here

3

2

Template for BioSample package One Health Enteric; version 1.0
[Download Excel](#) or [download TSV](#)
For column explanations 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 Antibiogram 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 Antibiogram 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](#).

4.6 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 out 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":



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

[Choose 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.

Continue

4.7.1: The downloaded metadata excel spreadsheet:

	A	B	C	D	E	F	G	H	I
1									
2		Instructions:							
3		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.							
4		Some column headers have hyperlinks to NCBI webpages.							
5		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.							
6									
7		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.							
8		If you created samples previously, provide accessions in the form of SAMN# in the column <i>sample_accession</i> .							
9		Each row in the template represents a sequencing library with a unique combination of <i>sample</i> + <i>library</i> + <i>sequencing strategy</i> + <i>layout</i> + <i>instrument model</i> . Each row should have a unique library_id that is short and meaningful (like an ID you might use in lab).							
10		When libraries are indeed identical (same combination of <i>sample</i> + <i>library</i> + <i>strategy</i> + <i>layout</i> + <i>instrument model</i>), 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 <i>filename2</i> , <i>filename3</i> , etc.... PAIRED files must always be listed in the same row.							
11		Provide exact file names (including extensions) in the <i>filename</i> columns.							
12		File names must be unique .							
13									
14		red triangles indicate pop-up comments for that field							
15									
16		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.							
17		NOTE: There are data checks and autocomplete 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.							
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									

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_name	library_id	title	library_strategy	library_source	library_selection	library_layout	platform	instrument_model	design_description	filetype	filename	filename2
1	AB-NOV-GIB-2	RB282352	Sequencing of norovirus	WGA	METAGENOMIC	RANDOM	paired	Illumina	Illumina MiSeq	non-targeted WGS	fastq	RB282352_S2_L001_R1_001.fastq.gz	RB282352_S2_L001_F
2	AB-NOV-GIB-4	RB282353	Sequencing of norovirus	WGA	METAGENOMIC	RANDOM	paired	Illumina	Illumina MiSeq	non-targeted WGS	fastq	RB282353_S3_L001_R1_001.fastq.gz	RB282353_S3_L001_F
3	YB060523	RB282355	Sequencing of norovirus	WGA	METAGENOMIC	RANDOM	paired	Illumina	Illumina MiSeq	non-targeted WGS	fastq	RB282355_S5_L001_R1_001.fastq.gz	RB282355_S5_L001_F
4													
5													
6													
7													

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

Get the file name of each fastq file and fill out the columns (filename and filename 2):

Name	Status	Date modified	Type	Size
 R82823S2_S2_L001_R1_001.fastq.gz		9/1/2023 9:01 AM	GZ File	442,747 KB
 R82823S2_S2_L001_R2_001.fastq.gz		9/1/2023 9:01 AM	GZ File	467,491 KB
 R82823S3_S3_L001_R1_001.fastq.gz		9/1/2023 9:01 AM	GZ File	420,902 KB
 R82823S3_S3_L001_R2_001.fastq.gz		9/1/2023 9:01 AM	GZ File	452,733 KB
 R82823S5_S5_L001_R1_001.fastq.gz		9/1/2023 9:01 AM	GZ File	287,159 KB
 R82823S5_S5_L001_R2_001.fastq.gz		9/1/2023 9:01 AM	GZ File	320,962 KB

4.7 Populate "Files" tab:

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



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 S3 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: `w4pY89VQ`
3. Navigate to your account folder.
From the command line use the 'cd' command:
`cd uploads/zhuiyang_gmail.com_y9toCH6I`

When using a GUI FTP client (eg: FileZilla, NoFTP Cyberduck etc.), after you've connected to the FTP server, paste your account folder (`uploads/zhuiyang_gmail.com_y9toCH6I`) into 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.

The screenshot displays the SRA FTP interface with the following details:

- Host:** encbi.nlm.nih.gov
- Username:** subftp
- Password:** [Redacted]
- Port:** [Redacted]
- Status:** Disconnected from server
- Local site:** FDA/paper and manuscripts/NCBI submissions/Raw data submission to SRA/SRA submission_VT2_VT4 and YB_Mar2024/Raw data/
 - Raw data of E275R submission to SRA, WI 2015 outbreak
 - Raw data of G251 GridON reads, April 2021
 - Raw data of mono saps and astre samples- Dec 2018
 - Raw data of stool samples from Julia_2023
 - Raw data of stool samples from VT_2023
 - Raw sequencing data for SRA submission_food paper_IJFM 2017
 - Raw sequencing data for SRA submission_HAV SNP
 - Raw sequencing data for SRA submission_moro and HAV- Dec 2018
 - SRA submission_E1653
 - SRA submission_VT2_VT4 and YB_Mar2024
 - Raw data
- Remote site:** /uploads/zhuiyang.gmail.com_y9icCN6i/SRA_upload_030622
 - uploads
 - zhuiyang.gmail.com_y9icCN6i
 - SRA_upload_030622
- File lists:**
 - Local site file list:**

Filename	Filesize	Filetype	Last modified
R8282352_S2_L001_R1_001.fastq.gz	453,372,476	GZ File	9/1/2023 9:01:07 AM
R8282352_S2_L001_R2_001.fastq.gz	478,710,509	GZ File	9/1/2023 9:01:13 AM
R8282353_S3_L001_R1_001.fastq.gz	431,001,514	GZ File	9/1/2023 9:01:07 AM
R8282353_S3_L001_R2_001.fastq.gz	463,597,844	GZ File	9/1/2023 9:01:13 AM
R8282355_S5_L001_R1_001.fastq.gz	294,050,258	GZ File	9/1/2023 9:01:07 AM
R8282355_S5_L001_R2_001.fastq.gz	328,664,812	GZ File	9/1/2023 9:01:13 AM
 - Remote site file list:**

Filename	Filesize	Filetype	Last modified	Permissions	Owner/Group
R8282352_S2_L001_R1...	453,372,476	GZ File	3/6/2024 11:50:...	adfrw (0664)	subftp sub...
R8282352_S2_L001_R2...	478,710,509	GZ File	3/6/2024 11:48:...	adfrw (0664)	subftp sub...
R8282353_S3_L001_R1...	431,001,514	GZ File	3/6/2024 11:51:...	adfrw (0664)	subftp sub...
R8282353_S3_L001_R2...	463,597,844	GZ File	3/6/2024 11:54:...	adfrw (0664)	subftp sub...
R8282355_S5_L001_R1...	294,050,258	GZ File	3/6/2024 11:54:...	adfrw (0664)	subftp sub...
R8282355_S5_L001_R2...	328,664,812	GZ File	3/6/2024 11:55:...	adfrw (0664)	subftp sub...
- Summary:**
 - Selected 6 files. Total size: 2,448,399,413 bytes
 - 6 files. Total size: 2,448,399,413 bytes

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



For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

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 -

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/zhuiyang_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/zhuiyang_gmail.com_y9IoCN6I`) into 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.

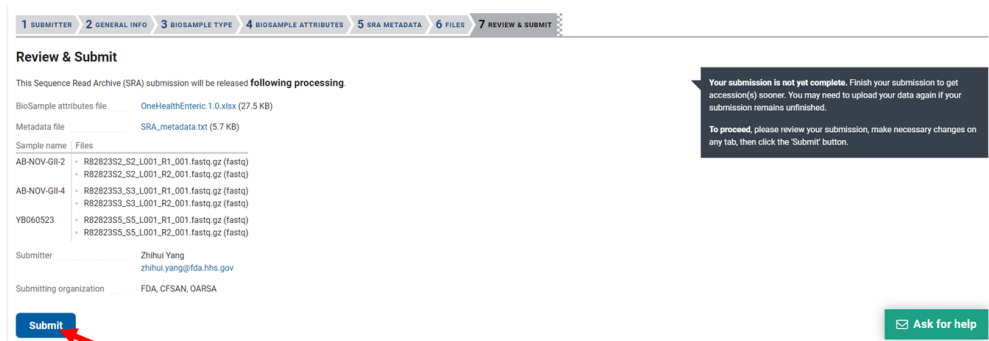
4.8.4: (note: it takes at least 10 minutes for uploaded files to become available) click "continue" to upload:

Do not modify or move the files you are submitting from this preload directory until you see the "SRA: Processed" status is present for the submission.

Aspera command line instructions UPDATED FEBRUARY 2024 +

FTP instructions UPDATED FEBRUARY 2024 -

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



4.9 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

Options to upload data:

- Upload via Cloud: Amazon S3 or Google Cloud
- Upload via Aspera command line or FTP
- Upload via Web browser or Aspera browser plugin

Submission	Title	Group	Status
SUB14290411	Raw sequence reads of multiple foodborne viruses from outbreaks or sporadic cases in US, Mar 05 '24	fda_cfsan_oarsa	✓ BioSample: Processed (Details) - Successfully loaded (3 objects) - SAMN40282907: AB-NOV-GII-2 (TaxID: 11983) - SAMN40282908: AB-NOV-GII-4 (TaxID: 11983) - SAMN40282909: YB060523 (TaxID: 11983) • We will automatically transform the attribute value(s) you provided as follows. (3 objects) - SAMN40282907: AB-NOV-GII-2 collection date: May, 2023 - SAMN40282908: AB-NOV-GII-4 collection date: May, 2023 - SAMN40282909: YB060523 collection date: May, 2023 Download attributes file with BioSample accessions ⊗ SRA: Processing

BioSample number SAMNxxx were created at this step.

Submission of assembled sequences to NCBI GenBank.

5

Submission of the assembled data to GenBank.

Raw sequencing data is required as input for ViroTrakr database deposit and subsequent de novo assembly data analysis. Assemblies or consensus sequences are part of the data analysis in our workflow, its submission to GenBank is highly 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 \(protocols.io\)](#)

Notes:

This GenBank submission assumes that you already have established a BioProject and BioSample(s) from step 3 and step 4 of this protocol. Otherwise, please refer to step 5.7 on how to link your sequence submission to ViroTrakr.

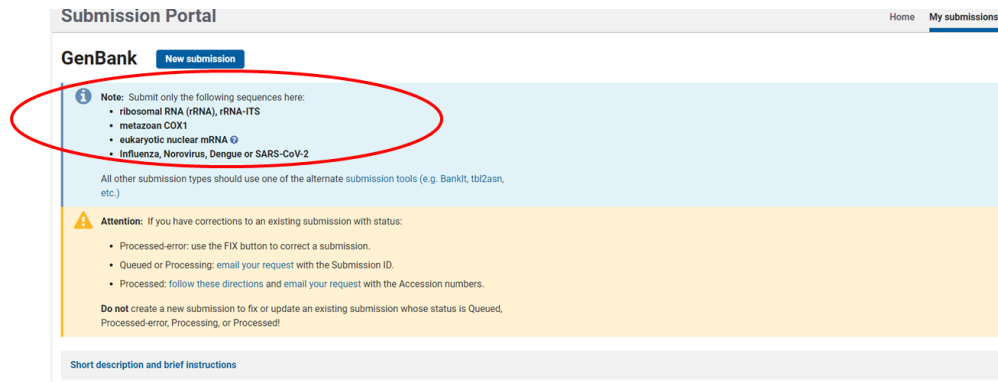
Assembled sequences of norovirus can be directly submitted through GenBank Submission Portal which offers annotation pipelines for certain species including Norovirus. **Submissions | GenBank | Submission Portal (nih.gov).**

Assembled sequences of other foodborne viruses, with annotation provided by the submitter, should use one of the alternate submission tools (e.g., BankIt, tbl2asn, etc.):

<https://www.ncbi.nlm.nih.gov/WebSub/?tool=genbank>

<https://www.ncbi.nlm.nih.gov/genbank/table2asn>

- 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 "Submission type" 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?**

☐ ribosomal RNA (rRNA) or rRNA-ITS ⓘ

☐ COX1 from metazoan mitochondria ⓘ

☐ Eukaryotic nuclear mRNA ⓘ **UPDATED**

☒ 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

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

Review requirements for norovirus submissions

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

Continue

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



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" button at the bottom 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

Sequencing Technology

Method

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

☐ Sanger dideoxy sequencing

☐ 454

☐ Helicos

☒ Illumina

☐ IonTorrent

☐ Pacific Biosciences

☐ SOLID

☐ Other

Assembly state

* These sequences are:

☐ Unassembled sequence reads

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

Assembly info [Name of algorithm, e.g. Newbler or SOAPdenovo](#)

* Assembly program ?	* Version or date ?	Delete
		Delete

[Add another assembly program](#)

[Continue](#)

5.5 Populate "Sequences" tab:

- Click "Release immediately following processing" or specify a date 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.

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
aacgatatagatagatgacccgatagagagaga
>Seq2
gtacgataaagatagatgacccgatagagagaga
```

Use the latest version of the Aspera Connect plugin for faster file uploads. If a pop-up box about 'Asp protocol' is displayed, click 'Allow' or 'Open' to let Aspera Connect handle file uploads more efficiently.

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_ID in your FASTA file. If you do this, do not include extra information in the sequence_ID 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
TGCAGAACTTGGACAACCCCTTTATCTAATCTTTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCG
TTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCGTGCAGAACTTGGACAACCCCTTTATCTAATCT
ATAGTTCAGCCTCCTCATCCGTGCAGAACTTGGACAACCCCTTTATCTAATCTTTGGAGCATGAGCTGGC

>SeqID2
CCTTTATCTAATCTTTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCGTGCAGAACTTGGACAAC
TGCAGAACTTGGACAACCCCTTTATCTAATCTTTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCG
TTGGAGCATGAGCTGGCATAGTTGGAACCGCCCTCAGCCTCCTCATCCGTGCAGAACTTGGACAACCCCTTTATCTAATCT
ATAGTTGGAACCGCCCTCAGCCTCCTCATCCGTGCAGAACTTGGACAACCCCTTTATCTAATCTTTGGAGCATGAGCTGGC
```

5.6 Populate "Source Info" tab:

- You may find more information on the question "Do your sequences 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.

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

Source Information

The first few sequence IDs that we found are:

AB-NOV-GII-2_R8282352_Consensus_sequence
AB-NOV-GII-4_R8282353_Consensus_sequence

★ 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

5.7 Populate "Source modifiers" tab:

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



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

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?

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

Continue

4.7:Populate "Source modifiers" tab: (cont.)

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

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.

[See an example Source Modifiers table](#)

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

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

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

Continue

* Notes: to link your submitted sequences to ViroTrakr, you may include the BioSample numbers and BioProject numbers of your sequences to your submission. To do so, you may choose one of

the options below: 1. download the source modifier template; 2. customize a source modifier template, see details below.

- Customized GenBank source modifier template:

Below is a customized version of source modifier table which contains direct linkage to the respective BioSample and BioProject records. Populate the template as guided and save it in a txt format.

 Customized source modifiers_for V... 104B

An example of the customized source modifiers table:

 GenBank submission modifiers_Viro...

- Click "Choose file" to upload your populated and saved source modifier file.
- Click "Continue" to next page.

Alternatively, you may send the relevant information to GenBank (BioSample number, BioProject number, etc.) at any time during and after the submission process, or as an update after you receive Genbank accession numbers. Please send the email to: **gb-admin@ncbi.nlm.nih.gov** (for replies/updates to records in GenBank)

- 5.8 Populate "References" tab:
Click "Continue" to next page.



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

- 5.9 Populate "Review & Submit" tab:
Review your submission, make any necessary changes using the tabs/steps above, then click on the "Submit" button below.

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 following processing

Submitter
Submitter: Zhihui Yang
zhihui.yang@fda.hhs.gov
zhihui.yang@fda.hhs.gov
Institution: FDA, CFSAN, OARSA
Department: DMB
Street: 8301 Muirkirk Rd
City: Laurel
State: MD
Postal code: 20708
Country: USA
Sequence authors
- Zhihui Yang
- Samantha Wales
References
Publication status: unpublished
Authors: Not provided
Sequencing Technology
Methods: Illumina

GenBank Record Preview
What you need to know about this record preview
LOCUS AB-NOV-GII-2_R0202352_Consensus_seque 7523 bp DNA linear
DEFINITION Norovirus GII isolate AB-NOV-GII-2.
ACCESSION VRL 07-MAR-2024
VERSION
DBLINK BioProject: PRJNA490509
BioSample: SAMN40282507
KEYWORDS
SOURCE Norovirus GII
ORGANISM Norovirus GII
Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
Picornavirales; Caliciviridae; Norovirus; Norwalk virus.
REFERENCE 1 (bases 1 to 7523)
AUTHORS Yang,Z. and Wales,S.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2024) DMB, CFSAN OARSA foodborne pathogen
submission group, 8301 Muirkirk Rd, Laurel, MD 20708, USA
COMMENT ##Assembly-Data-START##
Assembly Method : CLC GMB v. 23
Sequencing Technology : Illumina
##Assembly-Data-END##
FEATURES
source Location/Qualifiers
1..7523
/organism="Norovirus GII"
/mol_type="genomic DNA"
/isolate="AB-NOV-GII-2"
/isolation_source="feces"
/host="Homo sapiens"
/db_xref="taxon:122929"
/country="Canada"
/collection_date="May-2023"
/genotype="GII.6[P1]"
BASE COUNT 2076 a 1890 c 1893 g 1664 t
ORIGIN
1 ataacatcta acacacacac cactacattt acacacacac actctctcaa ccaaatatc

To proceed please review your submission
necessary changes using the tabs/steps
the Submit button below.

Submit

Notes: the DBLINK (BioSample and BioProject) will be included and shown on the GenBank Record Preview, if you chose the option 1 or 2 in step 5.7 above.

Notes:

- The GenBank accession number will be available within a couple of hours on "my submission" page. Meanwhile, you will receive an NCBI email containing these accession numbers, usually within 12 hours;

- If a processed-error shows up, you may correct a submission by using the FIX button associated with the error information.

GenBank [New submission](#)

Note: Submit only the following sequences here:

- ribosomal RNA (rRNA), rRNA-ITS
- metazoan COX1
- eukaryotic nuclear mRNA
- Influenza, Norovirus, Dengue or SARS-CoV-2

All other submission types should use one of the alternate submission tools (e.g. BankIt, tbl2asn, etc.)

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

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

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

Short description and brief instructions

17 submissions

Submission	Title	Group	Status
SUB14293207	Norovirus GII	fda_cfsan_oarsa	Submitted Awaiting processing.
CT1013492506	Norovirus GII	Zhihui Vang's	GenBank: Processed

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