



Version 4

Apr 14, 2021

NCBI submission protocol for microbial pathogen surveillance V.4

In 1 collection

Ruth E Timme¹, Maria Balkey², Robyn Randolph³, Julie Haendiges¹, Sai Laxmi Gubbala Venkata⁴, William Wolfgang⁵, Errol Strain³

¹US Food and Drug Administration;²Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, College Park, Maryland, USA;³Center for Veterinary Medicine, U.S. Food and Drug Administration, College Park, Maryland, USA;⁴Bacteriology Laboratory, Wadsworth Center, New York State Department of Health, Albany, New York, USA;⁵Wadsworth Center NYSDOH*In Development* dx.doi.org/10.17504/protocols.io.bgrajv2e[GenomeTrakr](#) [StaPH-B](#) 1 more workspace

Ruth Timme
US Food and Drug Administration

ABSTRACT

PURPOSE: Step-by-step instructions for submitting pathogen whole genome sequence data to NCBI and to the NCBI Pathogen Detection portal. This protocol covers the steps needed to establish a new NCBI submission environment for your laboratory, including the creation of new BioProject(s) and submission groups. Once these are step up, the protocol then walks through the process for submitting raw reads to SRA and sample metadata to BioSample through the Submission portal.

SCOPE: for use by any laboratory submitting WGS data for species under active surveillance within NCBI's Pathogen Detection. (This includes US laboratories in GenomeTrakr, NARMS, Vet-LIRN, PulseNet, and other non-US networks and submitters).

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

If you need a pipeline for frequent or large volume submissions, follow Step 1 to get your NCBI submission environment established, then contact gb-admin@ncbi.nlm.nih.gov to set up an account for submitting through the API.

This protocol covers submission using **NCBI's Submission Portal web-interface**.

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Timme, Ruth E., William J. Wolfgang, Maria Balkey, Sai Laxmi Gubbala Venkata, Robyn Randolph, Marc Allard, and Errol Strain. "Optimizing Open Data to Support One Health: Best Practices to Ensure Interoperability of Genomic Data from Bacterial Pathogens." *One Health Outlook* 2, no. 1 (October 19, 2020): 20. <https://doi.org/10.1186/s42522-020-00026-3>.

DOI

dx.doi.org/10.17504/protocols.io.bgrajv2e

PROTOCOL CITATION

Ruth E Timme, Maria Balkey, Robyn Randolph, Julie Haendiges, Sai Laxmi Gubbala Venkata, William Wolfgang, Errol Strain 2021. NCBI submission protocol for microbial pathogen surveillance. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.bgrajv2e>

Timme, Ruth E., William J. Wolfgang, Maria Balkey, Sai Laxmi Gubbala Venkata, Robyn Randolph, Marc Allard, and Errol Strain. "Optimizing Open Data to Support One Health: Best Practices to Ensure Interoperability of Genomic Data from Bacterial Pathogens." *One Health Outlook* 2, no. 1 (October 19, 2020): 20. <https://doi.org/10.1186/s42522-020-00026-3>.


COLLECTIONS ⓘ

 [GenomeTrakr data collection and submission workflow](#)

KEYWORDS

NCBI submission, WGS, pathogen surveillance, GenomeTrakr

LICENSE

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

CREATED

May 21, 2020

LAST MODIFIED

Apr 14, 2021

PROTOCOL INTEGER ID

37378

PARENT PROTOCOLS

Part of collection

[GenomeTrakr data collection and submission workflow](#)

BEFORE STARTING

This protocol has three sections:

- **Section 1:** Setting up NCBI accounts (for new users)
- **Section 2:** Single-step data submission to SRA for raw reads and associated sequence metadata and to BioSample for sample metadata
- **Section 3:** Detailed steps for creating a BioProject (usually done once during the account set-up)

Associated protocols:

- [Guidance for populating GenomeTrakr metadata templates \(BioSample and SRA\)](#)
- [NCBI Data Curation protocol](#) for making updates, corrections, or retractions to your data.

Establish submission environment at NCBI

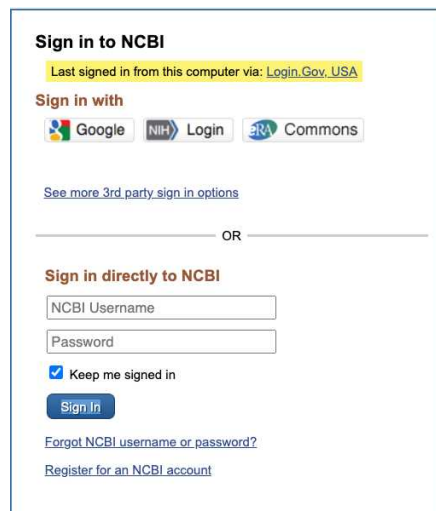
1 Set up a new NCBI submission environment for your lab:

- 1.1:** Create an NCBI user account
- 1.2:** Set up an NCBI submission user group for your lab
- 1.4:** Bookmark the link to your submission portal
- 1.5.** Identify or establish new BioProjects (detailed in **Step 3**)

Ready for data submission:

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

- 1.1 **Create an NCBI user account** at NCBI: <https://www.ncbi.nlm.nih.gov/account>. This will be your own individual user account at NCBI.



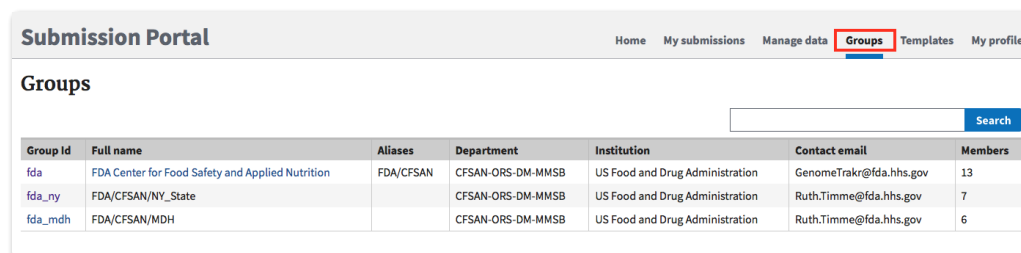
README: [** Important Changes to NCBI Accounts Coming in 2021**](#)

- 1.2 **Establish an NCBI submission user group for your laboratory.**

We recommend using this user group for all NCBI submissions related to microbial genome surveillance. This 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.

User groups also ensure 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.

Your laboratory might already have a submission group established! Check the "Group" tab in the submission portal, <https://submit.ncbi.nlm.nih.gov/groups/>. Ask your colleagues to do the same thing, to ensure your laboratory doesn't already have one in place.



Group id	Full name	Aliases	Department	Institution	Contact email	Members
fda	FDA Center for Food Safety and Applied Nutrition	FDA/CFSAN	CFSAN-ORS-DM-MMSB	US Food and Drug Administration	GenomeTrakr@fda.hhs.gov	13
fda_ny	FDA/CFSAN/NY_State		CFSAN-ORS-DM-MMSB	US Food and Drug Administration	Ruth.Timme@fda.hhs.gov	7
fda_mdh	FDA/CFSAN/MDH		CFSAN-ORS-DM-MMSB	US Food and Drug Administration	Ruth.Timme@fda.hhs.gov	6

Creating a new submission group:

1. Submit an email request to submit-help@ncbi.nlm.nih.gov containing the following information:

Dear NCBI help staff,

Please establish a new user group for my laboratory.
I'm including the following information to help set up the group:

Short name of the group (abbreviation, e.g. "fda_ny")
Full name of the group (e.g. "NY Wadsworth microbial pathogen submission group")
Contact email(s) to start the group
Institution and department or group
Physical address including country
Primary contact person, first and last name plus email.

* if you have existing submissions you want to be converted, please request the ownership change in this email.
i.e., Please assign this new user group to the following BioProjects and linked data.

Thank you

2. Look for an email reply entitled "NCBI Submission Portal Group invitation" and click on the enclosed link to accept the invitation.

1.3 Managing your NCBI submission user group.

After a user group has been established it can be edited for membership and permissions by clicking in the "group" tab of the submission portal (<https://submit.ncbi.nlm.nih.gov/groups/>), then on the Group Id hyperlink, e.g 'fda_ny' in the above example.

Users with admin privileges can update contact information in the "profile" tab and membership in the "Members" tab. New members can be invited by clicking on the "Invite members" link.

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

Groups: fda_ny

Profile **Members** Invites

[Invite members](#)

Username	First name	Last name	Email	Receives email notifications	Read	Modify	Submit	Delete	Admin	Remove member
klimke	Bill	Klimke	[REDACTED]	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="button" value="Remove"/>
retimme	Ruth	Timme	[REDACTED]	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="button" value="Remove"/>
justin.payne	Justin	Payne	[REDACTED]	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="button" value="Remove"/>
wwolfgang	William	Wolfgang	[REDACTED]	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="button" value="Remove"/>
samwirth	PulseNet	PulseNet	[REDACTED]	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="button" value="Remove"/>
slg04	Sai	Gubbala	[REDACTED]	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="button" value="Remove"/>

This user group should be kept up-to-date as members enter and leave the laboratory.

Permissions levels:

- READ: primarily for collaborators who can see the submissions, but not edit them.
- MODIFY, SUBMIT, DELETE: Permissions to submit, modify, or retract data (members usually have all or none of these permissions)
- ADMIN: Can invite or remove members of the submission group. Ensure that at least one (or more) members of your group have ADMIN privileges.

1.4 Bookmark "my submissions" at NCBI: <https://submit.ncbi.nlm.nih.gov/subs/>. This is the page

where you view and track all of your past submissions.

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

The screenshot shows the NIH Submission Portal interface. At the top, there's a header with the NIH logo and navigation links like Home, My submissions, Manage data, Groups, Templates, and My profile. Below the header, there's a section titled "Your submissions" with a "Start a new submission" button and a list of submission types: GenBank, BioProject, Sequence Read Archive, BioSample, Genome, Supplementary Files, TSA, and API. To the right, there's a "Filter / Search" section with fields for "From date", "To date", "Status" (set to "Not deleted"), and "Sort by". Below these are buttons for "Apps" and "Data archives", and a "Query" field with "Search" and "Clear" buttons. The main content area shows a list of submissions, with a table below it displaying details for two submissions: SUB6510311 and SUB6502971. The table columns are Submission, Title, App, Group, Status, and Updated.

Submission	Title	App	Group	Status	Updated
SUB6510311	UI-less submission 2019-11-04	API	fda	✓ BioSample: Processed Successfully loaded SAMN13192483 (TaxID: 1639) ✓ SRA: Processed (2 objects)	07:15
SUB6502971	UI-less submission 2019-11-01	API	fda	✓ BioSample: Processed Successfully loaded SAMN13207727 (TaxID: 670) ✓ SRA: Processed (2 objects)	06:44

1.5 Identify or establish new BioProjects (Umbrella and/or Data BioProjects)

Umbrella BioProjects. If you are already part of a surveillance network, (e.g. GenomeTrakr, NARMS, Vet-LIRN, or PulseNet) you will use one of their established umbrella bioprojects. For reference, here is a list of Umbrella projects for GenomeTrakr and Vet-LIRN.

GenomeTrakr Umbrella projects:

- *Salmonella enterica* PRJNA183844
- *Listeria monocytogenes* PRJNA514048
- *Escherichia coli* PRJNA230919
- *Vibrio parahaemolyticus* PRJNA245885
- *Campylobacter* sp. PRJNA258021
- *Clostridium botulinum* PRJNA290488

Vet-LIRN Umbrella projects:

- *Salmonella enterica* PRJNA314607
- *Escherichia coli* and *Shigella* PRJNA316449
- *Staphylococcus* PRJNA316451

If you need to establish a new umbrella BioProject, follow instructions in **Step 3** with modifications for creating a new Umbrella BioProject (**Step 3.12**).

Data BioProjects. Does your laboratory have an established data BioProject for this effort? If not please follow the instructions in **Step 3** for creating a new one.

More information:

Learn more about **data** vs **umbrella** BioProjects in Step 3

2 Data submission (source metadata and sequence data):

This protocol follows a one-step data submission process where the source metadata is submitted through the same submission workflow as the sequence data.

Before submission, ensure that your sequences meet the quality control (QC) thresholds for your surveillance network. You can follow your own internal QC process or use FDA's free GalaxyTrakr platform:

<https://www.protocols.io/view/assessing-sequence-quality-in-galaxytrakr-bdvfi63n>

Navigate to the Submission Portal management page: <https://submit.ncbi.nlm.nih.gov/subs/>

Submission Portal Home My submissions Manage data Groups Templates My profile

Your submissions

Start a new submission

- GenBank
- BioProject
- Sequence Read Archive**
- BioSample
- Genome
- Supplementary Files
- TSA
- API

Filter / Search

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

Apps: + Data archives: +

Query: Search Clear

104,089 submissions

Submission	Title	App	Group	Status	Updated
SUB8323567	GenomeTrakr Project: Listeria monocytogenes New York State Department of Health, Wadsworth Center, Oct 15 '20	Sequence Read Archive (SRA)	nys_gt	Submitted Awaiting processing.	10:25
SUB8308472	Pathogen: environmental/food/other sample	BioSample	nc_thakur_lab_gt	BioSample: Processed Successfully loaded (24 objects) Download attributes file with BioSample accessions	Oct 14
SUB8322783	UI-less submission 2020-10-14	API	fda	SRA: Processing	Oct 14
SUB8322788	UI-less submission 2020-10-14	API	fda	SRA: Processing	Oct 14

Click "Sequence Read Archive" to start a submission.

2.1 Download and populate the sample (BioSample) and sequence (SRA) metadata templates:

Organize your submissions by BioProject, only submitting to a single BioProject per submission workflow. Populate the metadata spreadsheets for each isolate you intend to submit (you can submit metadata for a single isolate or a collection of isolates under a single BioProject).

1. BioSample: GenomeTrakr pathogen template. (This is an extension of NCBI's combined pathogen package)

[One Health Enteric Package v0.2.xlsx](#)

2. SRA metadata template for raw sequence data (download directly from NCBI, no modification needed for GenomeTrakr compliance).

Download template here:

ftp://ftp-trace.ncbi.nlm.nih.gov/sra/metadata_table/SRA_metadata.xlsx

Follow the GenomeTrakr guidelines for populating these metadata templates: "[GenomeTrakr SOP for populating the NCBI submission templates](#)".

2.2 Click the "New submission" box.

The screenshot shows the NIH Submission Portal interface. At the top, there's a navigation bar with 'Home', 'My submissions', 'Manage data', 'Groups', 'Templates', and 'My profile'. Below this, the 'Sequence Read Archive (SRA)' section is visible. A red arrow points to the 'New submission' button. To the right of this button is a 'Filter / Search' section with fields for 'From date', 'To date', 'Status', and 'Sort by'. Below these are 'Data archives' and a 'Query' search bar. On the left, there are expandable sections for 'Short description and brief instructions', 'Options to preload data:', 'Aspera browser plugin upload', 'Aspera command line and FTP upload options', and 'Amazon S3 instructions'. At the bottom, there's a table of '51 submissions' with columns for 'Submission', 'Title', 'Group', 'Status', and 'Updated'.

2.3 SUBMITTER tab:

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

Select the appropriate submission group name (see **Step 1.2** for creating a new submission group), and describe the submitting organization or laboratory name. This will be auto-populated from the contact info you included in your NCBI user account. Click "Continue" to proceed.

Sequence Read Archive (SRA) submission: SUB9477908

GenomeTrakr Project: US Food and Drug Administration, Apr 14 '21

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

Submitter

★ First (given) name Middle name ★ Last (family) name
Ruth Timme

★ Email (primary) Email (secondary)
ruth.timme@fda.hhs.gov retimme@gmail.com

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

Group for this submission

☐ No group (affiliation from my personal profile)

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

2.4 GENERAL INFO tab:

1. BioProject: Did you already have a data BioProject for this effort? If not please follow instructions in **Step 3** for creating a new data or umbrella BioProject. Return back to this sub-step with the **data BioProject** accession in hand.

Click **"Yes"** and paste in your data BioProject accession, e.g. PRJNA614995.

Sequence Read Archive (SRA) submission: SUB9477908

GenomeTrakr Project: US Food and Drug Administration, Apr 14 '21

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

PRJNA186035
PRJNA186035: GenomeTrakr Project: US Food and Drug Administration

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

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

Continue

2. **BioSample**: Click "NO" here. You will be registering BioSamples within this current submission

3. **Release date**: Choose "Release immediately following processing".

4. Click **Continue**.

2.5 BIOSAMPLE TYPE tab:

You are choosing the appropriate metadata package here for your sample (i.e. what kind of samples are you submitting?).

Select **"Pathogen"**, then **"Combined pathogen submission"** for microbial pathogen submissions.

Submission Portal

Sequence Read Archive (SRA) submission: SUB9477908

GenomeTrakr Project: US Food and Drug Administration, Apr 14 '21

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...](#)

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

Microbe

Use for bacteria or other unicellular microbes when it is not appropriate or advantageous to use [MIS](#), 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 [MIS](#) 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

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

MIS Cultured Bacterial/Archaeal

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

MIS Eukaryotic

Use for eukaryotic genomic sequences. Organism must have lineage [Eukaryota](#).

MIS 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 [MIS](#) package for virus genomes.

MIMARKS Specimen

Use for any type of marker gene sequences, eg, 16S, 18S, 23S, 28S rRNA or COI obtained from cultured or voucher-identifiable 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

2.6 BIOSAMPLE ATTRIBUTES tab:

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

Then click "Choose File" and browse to your populated metadata template

If you have not yet populated your GenomeTrakr BioSample metadata template, download the current One Health Enteric Package in **Step 2.1**.

Sequence Read Archive (SRA) submission: SUB9477908

GenomeTrakr Project: US Food and Drug Administration, Apr 14 '21

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

Attributes

★ How do you want to provide your BioSample attributes?

☐ Use built-in table editor

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

One Health Enteric Package v0.2.xlsx 32.5 kB 2021-04-14 16:57 [Delete](#)

📎 Template for BioSample package **Pathogen: combined; version 1.0**

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

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

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

Antibiogram — pathogen MIC (optional)

[Choose file](#) 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)

[Choose file](#) 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](#).

[Continue](#)

Antibiogram data: please provide if you have it!

Click "**Continue**".

NCBI will do a validation check on your metadata. Resolve any red "errors" reported back by editing the spreadsheet and replacing the uploaded file. Review any yellow "Warnings" and proceed if everything looks ok.

Click "**Continue**".

2.7 SRA METADATA tab:

Choose: "Upload a file using Excel or text format (tab-delimited)".

Sequence Read Archive (SRA) submission: SUB9477908

GenomeTrakr Project: US Food and Drug Administration, Apr 14 '21

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

SRA_metadata-GTexample.xlsx 53.7 kB 2021-04-14 17:11 [Delete](#)

- Use one of these ways to fill out SRA metadata:
- [Download Excel spreadsheet](#) (designed to make it easier to select the correct metadata values), edit, save and then upload the modified Excel file.

[Continue](#)

Upload the populated **SRA metadata template** downloaded in **Step 2.1**.

Click "**Continue**".

NCBI will do a validation check on your sequence metadata. Resolve any red "errors" reported back by editing the spreadsheet and replacing the uploaded file. Review any yellow "Warnings" and proceed if everything looks ok.

Click "**Continue**".

2.8 Files tab:

Each laboratory will establish its own path for transferring files.

In general, selecting the web browser option should work for uploading ~48 sequences at a time. For a more stable internet connection, your laboratory can use FTP or Aspera. Directions for doing so pop up after clicking the FTP radio button.

- 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.
- ☐ Amazon S3 bucket

It is generally not recommended to check the Auto-submission box as this would not allow you to edit corrections if needed.

2.9 REVIEW & SUBMIT tab:

Check over your entire submission, then click submit.

If corrections are needed, you can go back and select individual tabs to edit your submission.

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#####):

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

SRA (for raw sequence or sequence metadata issues): sra@ncbi.nlm.nih.gov

2.10 BioSample accessions:

BioSample accessions will be automatically created upon submission and will be available on the “my submissions” page of the Submission portal by clicking on “## objects” within the submission record. You can also download by clicking the “Download attributes file with BioSample accessions”. Accessions will start with SAMNxxxxxxx. You will also receive an email within 12 hours, but typically much faster, containing these same accessions.

The screenshot shows the BioSample submission portal interface. At the top, there are links for "New submissions" and "Download batch submission template". Below this is a note: "Note: to update an existing record or recent submission, please email your request". A section titled "Short description and brief instructions" is visible. On the right, there is a "Filter / Search" section with fields for "From date", "To date", "Status" (set to "Not started"), and "Sort by". There is also a "Data archive" section with a "+" icon and a "Query ID" field with a "Search" button. The main content area shows a table of submissions. The first submission is titled "Pathogen: environmental/food/water sample" with submission ID "SUB412776". The "Status" column shows "BioSample: Processed (2 responses)". A list of 24 objects is displayed, each with a BioSample accession number (e.g., SAMN088348, SAMN088349, etc.) and a corresponding CFSA accession number. The table is paginated, showing 24 objects per page. At the bottom, there are links for "Download attributes file with the sample accessions" and "Download attributes file with the sample accessions".

2.11 SRA Accessions:

SRA run accessions will be available on the “my submissions” page of the Submission portal by clicking on “## objects” within the submission record. You can also download by clicking the “Download metadata file with SRA accession”. Accessions will start with SRRxxxxxxx.” You will also receive an

email with these same accessions within 24 hours, but typically much faster, containing these same accessions.

Submission Portal

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

Short description and brief instructions +

Options to preload data:

Aspera browser plugin upload +

Aspera command line and FTP upload options +

Amazon S3 instructions +

Filter / Search

From date To date Status Sort by

Not deleted + desc

Data archives +

Query Search Clear

30 submissions

Submission	Title	Group	Status	Updated
SUB5616822	GenomeTrakr Project: US Food and Drug Administration, May 14 '19	fda	✓ SRA: Processed (2 objects) • SRR9052981 • SRR9052982 Download metadata file with SRA accessions View and manage my SRA submission data	May 14
SUB5110190	GenomeTrakr Project: Bacillus cereus, New York State Department of Health, Wadsworth Center, Jan 31 '19	fda_ny	✓ SRA: Processed (6 objects) Download metadata file with SRA accessions View and manage my SRA submission data	Jan 31
SUB5106005	GenomeTrakr Project: NY State Dept. of Health, Wadsworth Center, Jan 30 '19	fda_ny	✓ SRA: Processed (20 objects) Download metadata file with SRA accessions View and manage my SRA submission data	Jan 30

2.12 Important data stewardship and curation notes:

- Develop an internal method for storing and tracking your BioSample and SRR accessions! They are required for making future updates to your records.
- For updates, corrections, or retractions to your BioSample and SRA records, follow the guidance provided in the NCBI Curation Protocol: <https://www.protocols.io/view/ncbi-data-curation-protocol-bacajase>. Some edits can be made within the submission portal and others need to be done via email.



Caution: It is possible for a single BioSample to have more than one SRR ID. Two scenarios include:

1. Two runs were submitted for the same isolate/BioSample, which is not generally recommended for surveillance. Follow **Step 3** in the NCBI curation protocol to retract one of them).
2. if the initial submission was retracted and new a new run was submitted. It's important to keep track of both IDs, even if one was retracted.

BioProject Creation

3 Create a new BioProject

BioProjects are an organizing tool at NCBI that pulls together different kinds of data submitted across multiple NCBI databases. Each BioProject has a unique URL, providing a home page with a title, description, links to lab websites, publications, funding resources associated with a particular project, along with links to the deposited data. A basic **data BioProject** holds actual sequence data, assemblies, and their associated metadata. An **umbrella BioProject** is a way to group two or more data BioProjects together, which is useful for coordinating disease surveillance and for looking across the grouped BioProjects in a single view.

This protocol describes the steps for creating a new data BioProject linked to an existing umbrella BioProject (usually established by a coordinating group, e.g. GenomeTrakr, NARMS, Vet-LIRN).

*If you need to create a new Umbrella BioProject, modifications are summarized in Step 3.12.

3.1

Navigate to the “My Submissions” page, <https://submit.ncbi.nlm.nih.gov/subs/>, and click “BioProject” in the “Start a new submission” box.

3.2 Click the “New submission” box:

3.3 Submitter tab:

Populate with submitter info. An NCBI "submitter" is the name of the person or submission group who is managing the submissions, not a supervisor or PI.

Select the appropriate submission group name (see **Step 1.2** for creating a new submission group), and describe the submitting organization or laboratory name. This will be auto-populated from the contact info you included in your NCBI user account.

3.4 **Project type tab:**

Project data type: Genome sequencing and assembly.

Sample scope:

*For a **Data BioProject**:* select multi-isolate. This will allow you to submit multiple isolates or strains of the same species.

3.5 **Target tab:**

*For a **Data BioProject**:* Populate ONLY the Organism name here, usually Genus species, or just Genus if your laboratory does not determine species, e.g., *Salmonella enterica*. Leave the strain info and Description fields blank.

Leave the strain info and Description fields blank.

3.6 **General info tab:**

Click "Release immediately following processing".

Include a brief title describing the effort.

- **Data BioProject Title:** e.g., "GenomeTrakr Project: NY State Dept. of Health, Wadsworth Center".

Public Description: e.g., "Whole-genome sequencing of pure-cultured microbial pathogens as part of XXXX surveillance effort."

Relevance: medical.

Is your project part of a larger initiative that is already registered at NCBI?

- **Data BioProjects.** Click "Yes" and include a brief description and umbrella BioProject accession number (see **Step 1.5**). This will properly link your data project to the umbrella.

General Info

Required fields are marked with *

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

GenomeTrakr Project: Kentucky Department for Public Health

Public description

Whole genome sequencing of cultured Salmonella enterica as part of the US Food and Drug Administration's WGS surveillance effort for the rapid traceback of foodborne pathogens.

Relevance

Medical

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

- ☐ No ☒ Yes (not very common)

Initiative description

GenomeTrakr network

BioProject accession

PRJNA183844

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.

External Links

Description

Kentucky Department for Public Health - Division of Lat

URL

<https://chfs.ky.gov/agencies/dph/dls/Pages/default.asp>

Delete

Add another link

3.7 BioSample tab:

Leave blank!! You will create biosamples separately.

3.8 Publications tab:

If relevant, include publications from your laboratory.

3.9 Review and Submit tab:

Check if everything looks correct and edit if necessary, then click "submit."

3.10 The BioProject accession will be available within a few minutes on the "my submissions" page of the Submission portal, "PRJNAxxxxx." You will also receive an email containing the new accession.

- 3.11 If you are part of a coordinated surveillance effort, like GenomeTrakr, please alert the coordinating body that a new BioProject was created under an existing umbrella.

For GenomeTrakr, contact genomeTrakr@fda.hhs.gov

3.12 Creating a new Umbrella BioProject:

Proceed as outlined in the above steps with the following modifications:

PROJECT TYPE tab:

For an ***Umbrella BioProject***: select multi-species. This will allow you to link multiple data BioProjects representing different species under a single umbrella.

TARGET tab:

For an ***Umbrella BioProject***: Leave the Organism name field blank. Include a list or description of species you intend to include in this effort. E.g. "bacterial foodborne pathogens", or "SARS-Cov-2"

GENERAL INFO tab:

Umbrella BioProject Title: e.g. "Microbial pathogen surveillance at NY State Dept. of Health, Wadsworth Center."

Is your project part of a larger initiative that is already registered at NCBI?

- For an Umbrella BioProject: click "NO"

The last step is to email bioprojecthelp@ncbi.nlm.nih.gov:

Example email:

"Dear BioProject and PD help teams,

Please convert the PRJNA##### to an Umbrella BioProject. Our laboratory will be submitting data under the XXX effort (SARS-CoV-2, GenomeTrakr, Vet-LIRN, NARMS, HAI, or more general pathogen surveillance).

I'd be happy to provide any additional details you might need.

Thank you, "

After the conversion is complete you can use the new Umbrella accession to properly link any new data BioProjects being created.

3.13 Important data stewardship and curation notes:

- Develop an internal method for storing and tracking your BioProject accessions! They are required for every BioSample and sequence data submission to ensure proper linkage.
- Bookmark URLs for each of your data BioProjects to monitor the public-facing view of your submissions.
e.g. Virginia DCLS's GenomeTrakr *Salmonella* BP:
<https://www.ncbi.nlm.nih.gov/bioproject/219491>
- For updates to your BioProjects, follow the guidance provided in the NCBI Curation Protocol: <https://www.protocols.io/view/ncbi-data-curation-protocol-bacaiase>. Some edits can be made within the submission portal and others need to be done via email.