

Nov 14, 2025

Version 2

Uploading Metadata and Creating SRA Submission Using mixS Package for Metagenomic samples Within NCBI Submission Portal 2.0 V.2



Version 1 is forked from [NCBI submission protocol for microbial pathogen surveillance](#)



DOI

<https://dx.doi.org/10.17504/protocols.io.dm6gpm5o8gzp/v2>

Christopher Duda¹, Amanda Windsor¹, Brandon Kocurek¹, Andrea Ottesen¹, Karen Jarvis¹, Taylor Richter¹, Padmini Ramachandran¹, Christopher Grim¹

¹FDA

GenomeTrakr

Tech. support email: genomeTrakr@fda.hhs.gov



Christopher Duda

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.dm6gpm5o8gzp/v2>



Protocol Citation: Christopher Duda, Amanda Windsor, Brandon Kocurek, Andrea Ottesen, Karen Jarvis, Taylor Richter, Padmini Ramachandran, Christopher Grim 2025. Uploading Metadata and Creating SRA Submission Using mixS Package for Metagenomic samples Within NCBI Submission Portal 2.0. **protocols.io**

<https://dx.doi.org/10.17504/protocols.io.dm6gpm5o8gzp/v2> Version created by **Christopher Duda**

Manuscript citation:

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>.

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: In development

We are still developing and optimizing this protocol

Created: November 14, 2025

Last Modified: November 17, 2025

Protocol Integer ID: 232452

Keywords: NCBI submission, GenomeTrakr, submission protocol for microbial pathog..., ncbi pathogen detection, new ncbi submission environment, ncbi submission environment, step instructions for data submission, ncbi submission, ncbi for analysis, first data submission, data submission, general submission to sra, ncbi, submission protocol, guidance for new submitter, metadata package guidance, large volume submission, laboratory group, describing general submission, new bioproject, anticipated first submission, submission, bioproject, insdc standard data structure, creating new bioproject, essential step, based submission, laboratory, other protocol, us lab, metagenomics, submitting metagenomic data, sra submission using mixs package for metagenomic sample, mixs metadata package to ncbi, metagenomic sample, metagenomic data from any food, step instructions for data submission, metagenomic sequence, ncbi submission environment, new ncbi submission environment, mixs metadata package, using mixs metadata package, creating sra submission using mixs pack

Disclaimer

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



Abstract

PURPOSE: This document provides detailed instructions on how to submit metagenomic sequences (targeted and shotgun) data and associated contextual data to NCBI, while adhering to the INSDC standard data structure, "**Pathogen DOM**", <https://genomicsstandardsconsortium.github.io/mixs/>. The protocol includes essential steps to create a new NCBI submission environment for your laboratory group, using **MixS** metadata package which is crucial to have in place before data are submitted. After these initial setups, the the remaining protocol focuses on step-by-step instructions for data submission.

SCOPE: This protocol is intended for any laboratory submitting metagenomic data from any food, farm, environmental samples using MixS metadata package to NCBI.

GUIDANCE FOR NEW SUBMITTERS: Before initiating your first data submission, there is significant preparatory work required. We advise designating a team member to spend several days setting up the necessary systems well before your anticipated first submission.

Watch NCBI's [10min video tutorial](#) describing general submission of whole genome sequences (WGS) to SRA.

ADVICE FOR FREQUENT/LARGE VOLUME SUBMISSIONS: Start by following Step 1 to establish your NCBI submission environment. For ongoing or large-scale submissions, email gb-admin@ncbi.nlm.nih.gov to arrange an account for API-based submissions.

Before start

This protocol has three sections:

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

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

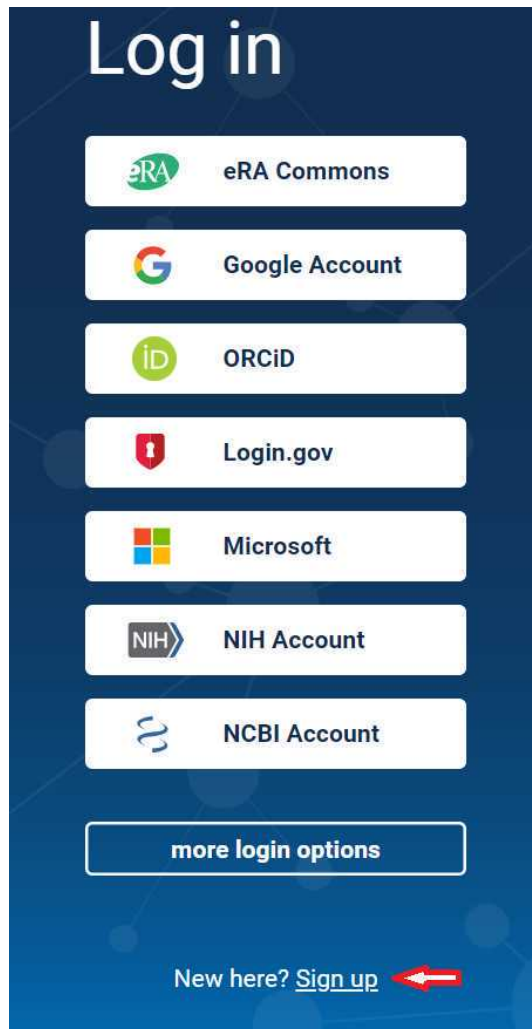
1.4: Bookmark the link to your submission portal

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.



The signup link is at the bottom of the page.

Choose a signup option that works for your institution.

1.2 Establish an NCBI submission user group for your laboratory.

We recommend using this user group for *all* NCBI submissions related to your lab's pathogen genome surveillance submissions.

This approach will link data submitted by your lab to the **user group** and not to individuals doing the submissions, allowing anyone in the current submission 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 were submitted.

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, if needed, to manage these efforts separately.

Your laboratory might already have a submission group established! Sign into your personal NCBI account, then check the "Group" tab in the Submission Portal ([Submission Portal](#) | [NCBI](#) | [NLM](#) | [NIH](#)). Ask your colleagues to do the same to ensure your laboratory does not already have one in place.

Submission Portal

Home My submissions Manage data Groups Templates [My profile](#)

Personal Information

* First (given) name: Christopher * Middle name: Robert * Last (family) name: Duda

* Email (primary): Christopher.Duda@fda.hhs.gov [Change primary email](#) ☒ Send email notifications on submission status change

Email (secondary):

* Institution: HHS * Institution URL: * Department: FDA

Phone (O): Fax (O):

* Street: 2001 Campus Drive * City: College Park * State/Province: MD * Postal code: 20741 * Country: USA

[Save](#) [Back](#)

Submission Portal

Groups

Group Id	Full name	Aliases	Department	Institution
561400-shared-submissions	Christopher Duda's shared submissions		HFP-OLQAS-OAMT-DFSG-MDEB	FDA

View of the "Groups" tab, when selected from the NCBI Submission Portal

Click on this link to verify your membership in NCBI user groups:
<https://submit.ncbi.nlm.nih.gov/groups/>

Creating a new submission group:

1. On your NCBI profile page (<https://submit.ncbi.nlm.nih.gov/accounts/profile/>), scroll to the bottom of the page and click on the "**Create group for shared submissions**" button.

Phone Fax

★ Street ★ City ★ State/Province ★ Postal code ★ Country

Campus Drive College Park MD 20740 USA

Save **Back**

Groups

FDA Center for Food Safety and Applied Nutrition (admin, read, modify, submit, delete)

14 members: Anjanette Johnston, Yan Luo, Errol Strain, Ruth Timme, Justin Payne, Narjol Gonzalez-Escalona, Hugh Rand, Maria Balkey, Maria Hoffmann, Jayanthi Gangiredla, Julie Haendiges, Yu Wang, Elizabeth Hunter, you

Create group for shared submissions



Note

The "Create group for shared submissions" button will not exist, if the user has not filled in all of the required profile information, marked with an asterisk (*) on the profile page.

2. On the resulting page, fill in the required information to create, at minimum, a short name, full name, and contact information for this submission group.

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

Groups > Christopher Duda's shared submissions

[Profile](#) [Members](#) [Invites](#) [Submissions](#)

Open invites tab to invite collaborators to join your group.

★ Short name ★ Full name

Christopher Duda Christopher Duda's shared submissions

Aliases Description

URL

★ Email Alternative email

Christopher.Duda@fda.hhs.gov

★ Institution ★ Department

FDA HPP-CDAS-GAMTDFIS-MDEB

Phone ID Fax ID

★ Street ★ City ★ State/Province ★ Postal code ★ Country

5001 Campus Drive College Park MD 20741 USA

Save

3. To invite members, use either the **"Invite members"** button at the top of the next, "Members" tab or from the "Invites" tab itself to navigate to the invite tab and add the invitees' emails to the text box. Click the **"Invite Members"** button when finished.

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.

First name	Last name	Email	Receives email notifications	Read	Modify	Submit	Delete	Admin	Remove member
			☐	☒	☐	☐	☐	☒	<button>Remove</button>
			☐	☒	☒	☒	☒	☒	<button>Remove</button>

This user list should be kept current as members/staff enter and leave the laboratory.

Permissions levels:



- **READ:** primarily for collaborators who would like to view 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.

The "Submissions" tab will show a breakdown of how many submissions have been made by this group:

Submission Portal	
Groups > fda	
Profile	Members
Invites	Submissions
All submissions	109148
api	108144
bioproject	150
biosample	573
genbank	6
sra	96
wgs	105
wgs1	8
wgs_batch	66

- 1.4 **Bookmark "My submissions"** at NCBI: <https://submit.ncbi.nlm.nih.gov/subs/>. This is the page where you view and track current and past submissions.

Start a new submission

- GenBank
- Sequence Read Archive
- Genome
- TSA
- BioProject
- BioSample
- API
- dbGaP
- GTR
- ClinVar

Filter / Search

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

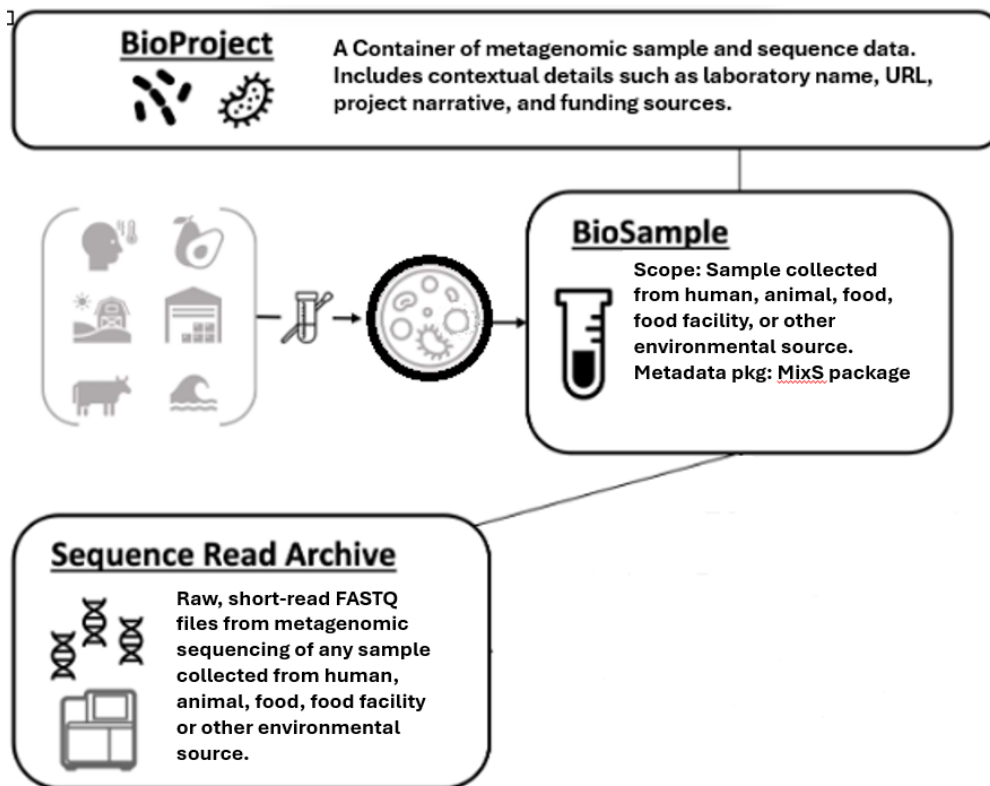
Apps: + Data archives: +

Query: Search Clear

Submission	Title	App	Status	Updated
SUB13033276	Detection of multiple Salmonella serotypes from enrichment cultures using shotgun metagenomic profiling, Apr 05 23	Sequence Read Archive (SRA)	✓ SRA: Processed (50 objects) Download metadata file with SRA accessions View and manage my SRA submission data	Apr 05 2023
SUB13030092	MIMS Environmental/Metagenome sample	BioSample	✓ BioSample: Processed (Details) Download attributes file with BioSample accessions	Apr 05 2023
SUB13028722	Detection of multiple Salmonella serotypes from enrichment cultures using shotgun metagenomic profiling	BioProject	✓ BioProject: Processed PRJNA952520: Detection of multiple Salmonella serotypes from enrichment cultures using shotgun metagenomic profiling	Apr 05 2023
SUB2401223	Enrichment dynamics of <i>Listeria monocytogenes</i> and the associated microbiome from naturally contaminated ice cream linked to a listeriosis outbreak, Feb 13 17	Sequence Read Archive (SRA)	✓ SRA: Processed (42 objects) Download metadata file with SRA accessions View and manage my SRA submission data	Jan 28 2021
SUB8676904	Investigating the distribution of strains of <i>E. amylovora</i> and streptomycin resistance in apple orchards in New York using CRISPR profiles: a six-year follow-up, Dec 02 20	Sequence Read	✓ SRA: Processed (15 objects)	Dec 02 2020

Data submission (BioSample and SRA)

2 Data submission (Sample metadata, SRA metadata, and raw sequence data), compliant with the Pathogen DOM data structure.



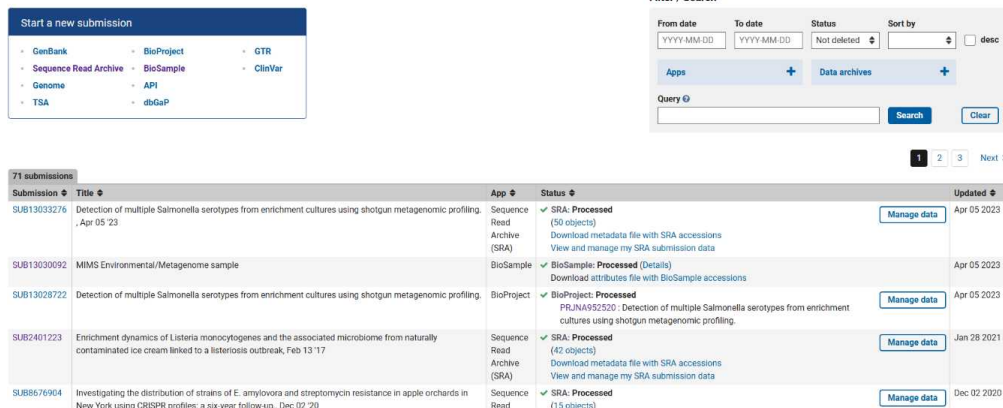
Overview detailing the scope of BioProjects, BioSamples, and Sequence Read Archive (SRA) submissions along with associated metadata standards for enteric pathogen surveillance. Abbreviations: Pkg – Package.

Note

Arrange your submissions according to their corresponding BioProjects, ensuring that each submission workflow is dedicated to a single BioProject. In cases where your data encompass multiple BioProjects, initiate a distinct submission for each BioProject separately.

2.1 Navigate to the **My Submissions** page in the NCBI Submission Portal:
<https://submit.ncbi.nlm.nih.gov/subs/>

Click "Sequence Read Archive" to start a submission.



Start a new submission

- GenBank
- BioProject
- GTR
- Sequence Read Archive
- BioSample
- ClinVar
- Genome
- API
- TSA
- dbGaP

Filter / Search

From date: YYYY-MM-DD To date: YYYY-MM-DD Status: Not deleted Sort by: [dropdown] [checkbox] desc

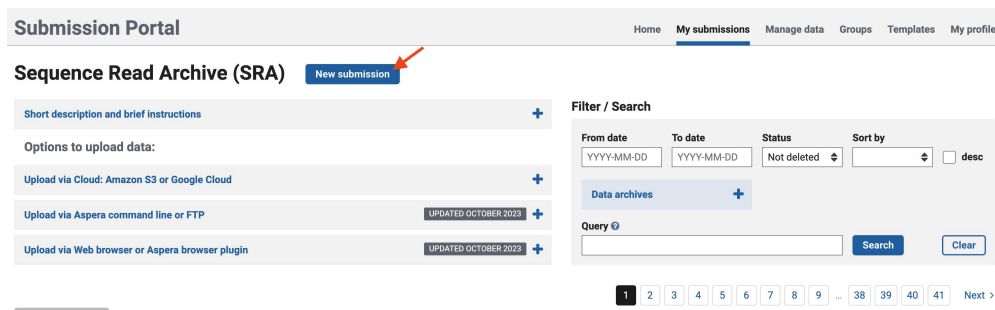
Apps: [plus] Data archives: [plus]

Query [input] [Search] [Clear]

71 submissions

Submission	Title	App	Status	Updated
SUB13033276	Detection of multiple Salmonella serotypes from enrichment cultures using shotgun metagenomic profiling, Apr 05 '23	Sequence Read Archive (SRA)	✓ SRA: Processed (50 objects) Download metadata file with SRA accessions View and manage my SRA submission data	Apr 05 2023
SUB13030092	MIMS Environmental/Metagenome sample	BioSample	✓ BioSample: Processed (Details) Download attributes file with BioSample accessions	Apr 05 2023
SUB13028722	Detection of multiple Salmonella serotypes from enrichment cultures using shotgun metagenomic profiling	BioProject	✓ BioProject: Processed PRJNA52520: Detection of multiple Salmonella serotypes from enrichment cultures using shotgun metagenomic profiling	Apr 05 2023
SUB2401223	Enrichment dynamics of Listeria monocytogenes and the associated microbiome from naturally contaminated ice cream linked to a listeriosis outbreak, Feb 13 '17	Sequence Read Archive (SRA)	✓ SRA: Processed (42 objects) Download metadata file with SRA accessions View and manage my SRA submission data	Jan 28 2021
SUB8676904	Investigating the distribution of strains of E. amylovora and streptomycin resistance in apple orchards in New York using CRISPR profiles: a six-year follow-up, Dec 02 '20	Sequence Read	✓ SRA: Processed (15 objects)	Dec 02 2020

2.2 Click the "New submission" button.



Submission Portal

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

Sequence Read Archive (SRA) **New submission**

Short description and brief instructions [plus]

Options to upload data:

- Upload via Cloud: Amazon S3 or Google Cloud [plus]
- Upload via Aspera command line or FTP [plus] UPDATED OCTOBER 2023
- Upload via Web browser or Aspera browser plugin [plus] UPDATED OCTOBER 2023

Filter / Search

From date: YYYY-MM-DD To date: YYYY-MM-DD Status: Not deleted Sort by: [dropdown] [checkbox] desc

Data archives: [plus]

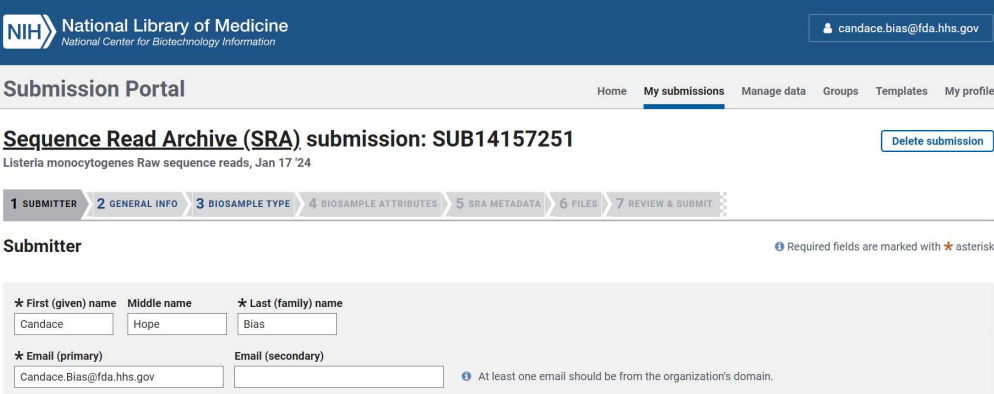
Query [input] [Search] [Clear]

1 2 3 4 5 6 7 8 9 ... 38 39 40 41 Next >

2.3 SUBMITTER tab:

Populate with submitter info. The "submitter" is the name of the person AND 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.



The screenshot shows the NIH Submission Portal interface. At the top, the NIH logo and 'National Library of Medicine' are visible, along with a user email 'candace.bias@fda.hhs.gov'. The 'Submission Portal' header includes navigation links: Home, My submissions (active), Manage data, Groups, Templates, and My profile. Below this, the submission title 'Sequence Read Archive (SRA) submission: SUB14157251' is displayed, with a 'Delete submission' button. A progress bar indicates the current step is '1 SUBMITTER', followed by '2 GENERAL INFO', '3 BIOSAMPLE TYPE', '4 BIOSAMPLE ATTRIBUTES', '5 SRA METADATA', '6 FILES', and '7 REVIEW & SUBMIT'. The 'Submitter' section contains the following fields:

- First (given) name: Candace
- Middle name: Hope
- Last (family) name: Bias
- Email (primary): Candace.Bias@fda.hhs.gov
- Email (secondary): [Empty field]

A note at the bottom right of the form states: 'Required fields are marked with * asterisk' and 'At least one email should be from the organization's domain.'

2.4 GENERAL INFO tab:

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

2. Release date: Choose "Release immediately following processing".

3. Click Continue.



📘 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

PRJNA429788 Food microbiome Organization: FDA [Clear field](#)

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

[Continue](#)

Example of filled in "General Info" tab. Please use the BioProject accession necessary for your organism and project.

2.5 BIOSAMPLE TYPE tab:

Here is where you can select the package type for your submission. Click on the "packages for metagenomic submitters" tab.



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 facilities;
4. environmental samples (farm, water, and the environment).
US public health agencies have created customized versions of this package that include more specific guidance, controlled vocabulary picklists, and sub-packages for each of the 4 sample types.
+ [@Phylorespository](#)

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 [MIUVIS](#) package for virus genomes. Before creating BioSamples for prokaryotic and eukaryotic MAGs, please read and follow the [MAG submission instructions](#).
- ☐ **MIMARKS Specimens**
Use for any type of marker gene sequences, eg. 16S, 18S, 23S, 28S rRNA or COI obtained from cultured or voucher-

Then click on one of the 4 packages in the highlighted box. Each use case will be explained below.

Sample Type

* Select the package that best describes your samples.

All packages

Packages for MAG submitters

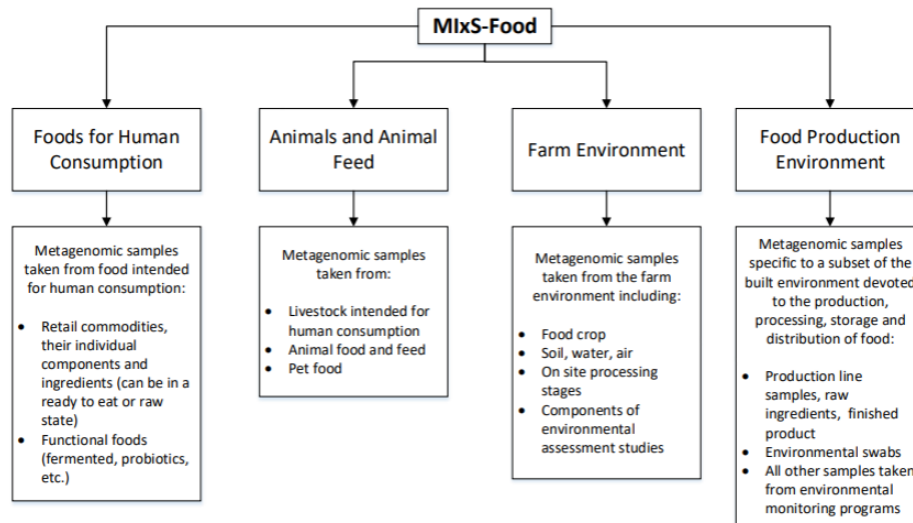
Packages for metagenome submittersNCBI packages [More...](#)

- ☐ **Metagenome or environmental**
Use for metagenomic and environmental samples when it is not appropriate or advantageous to use [MiXS](#) packages.

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

- ☐ **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'.
 - ☐ agriculture
 - ☐ air
 - ☐ built
 - ☐ food-animal and animal feed
 - ☐ food-farm environment
 - ☐ food-food production facility
 - ☐ food-human foods
 - ☐ host-associated
 - ☐ human-associated
 - ☐ human-gut
 - ☐ human-oral
 - ☐ human-skin
 - ☐ human-vaginal
 - ☐ hydrocarbon resources-cores
 - ☐ hydrocarbon resources-fluids/swabs
 - ☐ microbial mat/biofilm

What package do I select?



Animal and animal feed: A collection of terms appropriate when collecting samples and performing sequencing of samples obtained from farm animals and their feed

Farm environment: A collection of terms appropriate when collecting samples and performing sequencing of samples obtained from the farm environment, including soil, manure, and food harvesting equipment.

Food production facility: A collection of terms appropriate when collecting samples and performing sequencing of samples obtained from food production facilities.

Human Foods: A collection of terms appropriate when collecting samples and performing sequencing of samples obtained from human food products.

Once your package is selected, click the continue button.

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**".

For an example of each term: [Full term table - mixs](#)

Each BioSample package has a unique excel document that will be included in each section below.

There are a number of columns on this document that require specific information. What information to put in each column is explained on this webpage [BioSample Attributes](#) or in the following list.

***** Means Mandatory Value***** But please fill as many fields as possible!

Name:

- ***** Sample_name *****: Sample Name is a name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.

- **Sample_title**: Title of the sample

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

- **bioproject_accession**: The accession number of the BioProject(s) to which the BioSample belongs. If the BioSample belongs to more than one BioProject, enter multiple bioproject_accession columns. A valid BioProject accession has prefix PRJN, PRJE or PRJD, e.g., PRJNA12345.
- ***** organism *****: The most descriptive organism name for this sample (to the species, if possible). It is OK to submit an organism name that is not in our database. In the case of a new species, provide the desired organism name, and our taxonomists may assign a provisional taxID. In the case of unidentified species, choose the appropriate Genus and include 'sp.', e.g., "Escherichia sp.". When sequencing a genome from a non-metagenomic source, include a strain or isolate name too, e.g., "Pseudomonas sp. UK4". For more information about providing a valid organism, including new species, metagenomes (microbiomes) and metagenome-assembled genomes, see <https://www.ncbi.nlm.nih.gov/biosample/docs/organism/>.

Environment:

- ***** collection_date *****: the date on which the sample was collected; date/time ranges are supported by providing two dates from among the supported value formats, delimited by a forward-slash character; collection times are supported by adding "T", then the hour and minute after the date, and must be in Coordinated Universal Time (UTC), otherwise known as "Zulu Time" (Z); supported formats include "DD-Mmm-YYYY", "Mmm-YYYY", "YYYY" or ISO 8601 standard "YYYY-mm-dd", "YYYY-mm", "YYYY-mm-ddThh:mm:ss"; e.g., 30-Oct-1990, Oct-1990, 1990, 1990-10-30, 1990-10, 21-Oct-1952/15-Feb-1953, 2015-10-11T17:53:03Z; valid non-ISO dates will be automatically transformed to ISO format
- ***** env_broad_scale *****: Add terms that identify the major environment type(s) where your sample was collected. Recommend subclasses of biome

[ENVO:00000428]. Multiple terms can be separated by one or more pipes e.g.: mangrove biome [ENVO:01000181]|estuarine biome [ENVO:01000020]

- ***** env_local_scale *****: Add terms that identify environmental entities having causal influences upon the entity at time of sampling, multiple terms can be separated by pipes, e.g.: shoreline [ENVO:00000486]|intertidal zone [ENVO:00000316]
- ***** env_medium *****: Add terms that identify the material displaced by the entity at time of sampling. Recommend subclasses of environmental material [ENVO:00010483]. Multiple terms can be separated by pipes e.g.: estuarine water [ENVO:01000301]|estuarine mud [ENVO:00002160]
- ***** geo_loc_name *****: Geographical origin of the sample; use the appropriate name from this list https://www.insdc.org/subsubmitting-standards/geo_loc_name-qualifier-vocabulary/. Use a colon to separate the country or ocean from more detailed information about the location, eg "Canada: Vancouver" or "Germany: halfway down Zugspitze, Alps"
- ***** lat_lon *****: The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[d.ddd] N|S d[dd.ddd] W|E", eg, 38.98 N 77.11 W
- ***** omics_observ_id *****: A unique identifier of the omics-enabled observatory (or comparable time series) your data derives from. This identifier should be provided by the OMICON ontology; if you require a new identifier for your time series, contact the ontology's developers. Information is available here: <https://github.com/GLOMICON/omicon>. This field is only applicable to records which derive from an omics time-series or observatory.

Package specific attributes:

- Each package has specific attributes that you will need to include. Please reference the package step.

Nucleic Acid Sequence Source:

- **collection_method**: Process used to collect the sample, e.g., bronchoalveolar lavage (BAL)
- **ref_biomaterial**: Primary publication or genome report
- **rel_to_oxygen**: Is this organism an aerobe, anaerobe? Please note that aerobic and anaerobic are valid descriptors for microbial environments, eg, aerobe, anaerobe, facultative, microaerophilic, microanaerobe, obligate aerobe, obligate anaerobe, missing, not applicable, not collected, not provided, restricted access

- **samp_collect_device:** Method or device employed for collecting sample
- **samp_mat_process:** Processing applied to the sample during or after isolation
- **samp_size:** Amount or size of sample (volume, mass or area) that was collected
- **samp_vol_we_dna_ext:** volume (mL) or weight (g) of sample processed for DNA extraction
- **size_frac:** Filtering pore size used in sample preparation, e.g., 0-0.22 micrometer
- **source_material_id:** unique identifier assigned to a material sample used for extracting nucleic acids, and subsequent sequencing. The identifier can refer either to the original material collected or to any derived sub-samples.
- **isolation_source:** Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived.
- **neg_cont_type:** The substance or equipment used as a negative control in an investigation, e.g., distilled water, phosphate buffer, empty collection device, empty collection tube, DNA-free PCR mix, sterile swab, sterile syringe
- **pos_cont_type:** The substance, mixture, product, or apparatus used to verify that a process which is part of an investigation delivers a true positive
- **description:** Description of the sample.

3 Food-animal and animal feed:

Web Link for specific attributes: [BioSample Attributes](#) | [Submission Portal](#)



Food animal Animal feed.xlsx

*** **Means Mandatory Value***** But please fill as many fields as possible!

- ***** coll_site_geo_feat ***:** Text or terms that describe the geographic feature where the food sample was obtained by the researcher. This field encourages selected terms listed under the following ontologies: anthropogenic geographic feature (http://purl.obolibrary.org/obo/ENVO_00000002), for example agricultural fairground [ENVO:01000986]; garden [ENVO:00000011] or any of its subclasses; market [ENVO:01000987]; water well [ENVO:01000002]; or human construction (http://purl.obolibrary.org/obo/ENVO_00000070), e.g., grocery store [GENEPIO:0001020]
- ***** food_origin***:** A reference to a place on the Earth, by its name or by its geographical location that describes the origin of the food commodity, either in terms

of its cultivation or production. This field encourages terms listed under geographic location (http://purl.obolibrary.org/obo/GAZ_00000448), e.g., Thailand

- ***** food_prod_char*****: Descriptors of the food production system such as wild caught, free-range, organic, free-range, industrial, dairy, beef
- ***** food_product_type *****: A food product type is a class of food products that is differentiated by its food composition (e.g., single- or multi-ingredient), processing and/or consumption characteristics. This does not include brand name products but it may include generic food dish categories. This field encourages terms under food product type (<http://purl.obolibrary.org/obo/FOODON:03400361>). For terms related to food product for an animal, consult food product for animal (http://purl.obolibrary.org/obo/FOODON_03309997). If the proper descriptor is not listed please use text to describe the food type. Multiple terms can be separated by one or more pipes, e.g., multi-component food product [FOODON:00002501]
- ***** ifsac_category *****: The IFSAC food categorization scheme has five distinct levels to which foods can be assigned, depending upon the type of food. First, foods are assigned to one of four food groups (aquatic animals, land animals, plants, and other). Food groups include increasingly specific food categories; dairy, eggs, meat and poultry, and game are in the land animal food group, and the category meat and poultry is further subdivided into more specific categories of meat (beef, pork, other meat) and poultry (chicken, turkey, other poultry). Finally, foods are differentiated by differences in food processing (such as pasteurized fluid dairy products, unpasteurized fluid dairy products, pasteurized solid and semi-solid dairy products, and unpasteurized solid and semi-solid dairy products. An IFSAC food category chart is available from <https://www.cdc.gov/foodsafety/ifsac/projects/food-categorization-scheme.html> PMID: 28926300, e.g., Plants:Produce:Vegetables:Herbs:Dried Herbs
- ***** intended_consumer *****: Food consumer type, human or animal, for which the food product is produced and marketed, e.g., human
- ***** purpose_of_sampling *****: the reason that the sample was collected, e.g., active surveillance in response to an outbreak, active surveillance not initiated by an outbreak, clinical trial, cluster investigation, environmental assessment, farm sample, field trial, for cause, industry internal investigation, market sample, passive surveillance, population based studies, research, research and development
- **animal_am_dur**: The duration of time (days) that the antimicrobial was administered to the food animal
- **animal_am_freq**: The frequency per day that the antimicrobial was administered to the food animal

- **animal_am_route:** The route by which the antimicrobial is administered into the body of the food animal
- **animal_am_use:** The prescribed intended use of or the condition treated by the antimicrobial given to the food animal by any route of administration
- **animal_body_cond:** Body condition scoring is a production management tool used to evaluate overall health and nutritional needs of a food animal, e.g., normal, over conditioned, under conditioned
- **animal_diet:** If the isolate is from a food animal, the type of diet eaten by the food animal. Please list the main food staple and the setting, if appropriate. For a list of acceptable animal feed terms or categories, please see <http://www.feedipedia.org>. Multiple terms may apply and can be separated by pipes. Food product for animal covers foods intended for consumption by domesticated animals. Consult http://purl.obolibrary.org/obo/FOODON_03309997. If the proper descriptor is not listed please use text to describe the food type. Multiple terms can be separated by one or more pipes. If the proper descriptor is not listed please use text to describe the food product type
- **animal_feed equip:** Description of the feeding equipment used for livestock. This field accepts terms listed under feed delivery (http://opendata.inra.fr/EOL/EOL_0001757). Multiple terms can be separated by one or more pipes
- **animal_sex:** The sex and reproductive status of the food animal, e.g., castrated female, castrated male, intact female, intact male
- **bacterial_density:** Number of bacteria in sample, as defined by bacteria density (http://purl.obolibrary.org/obo/GENEPIO_0000043)
- **cons_food_stor_dur:** The storage duration of the food commodity by the consumer, prior to onset of illness or sample collection. Indicate the timepoint written in ISO 8601 format
- **cons_food_stor_temp:** Temperature at which food commodity was stored by the consumer, prior to onset of illness or sample collection
- **cons_purch_date:** The date a food product was purchased by consumer
- **cons_qty_purchased:** The quantity of food purchased by consumer

- **cult_isol_date:** A culture isolation date is a date-time entity marking the end of a process in which a sample yields a positive result for the target microbial analyte(s) in the form of an isolated colony or colonies, e.g., 5/24/2020
- **cult_result:** Any result of a bacterial culture experiment reported as a binary assessment, e.g., absent, active, inactive, negative, no, present, positive, yes
- **cult_result_org:** Taxonomic information about the cultured organism(s)
- **cult_target:** The target microbial analyte in terms of investigation scope. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>
- **enrichment_protocol:** The microbiological workflow or protocol followed to test for the presence or enumeration of the target microbial analyte(s). Please provide a PubMed or DOI reference for published protocols
- **experimental_factor:** Variable aspect of experimental design
- **food_additive:** A substance or substances added to food to maintain or improve safety and freshness, to improve or maintain nutritional value, or improve taste, texture and appearance. This field encourages terms listed under food additive (http://purl.obolibrary.org/obo/FOODON_03412972). Multiple terms can be separated by one or more pipes, but please consider limiting this list to the top 5 ingredients listed in order as on the food label. See also, <https://www.fda.gov/food/food-ingredients-packaging/overview-food-ingredients-additives-colors>, e.g., xanthan gum [FOODON:03413321]
- **food_contact_surf:** The specific container or coating materials in direct contact with the food. Multiple values can be assigned. This field encourages terms listed under food contact surface (http://purl.obolibrary.org/obo/FOODON_03500010), e.g., aluminum surface [FOODON:03500042]
- **food_contain_wrap:** Type of container or wrapping defined by the main container material, the container form, and the material of the liner lids or ends. Also type of container or wrapping by form; prefer description by material first, then by form. This field encourages terms listed under food container or wrapping (http://purl.obolibrary.org/obo/FOODON_03490100), e.g., bottle or jar [FOODON:03490203]
- **food_cooking_proc:** The transformation of raw food by the application of heat. This field encourages terms listed under food cooking

(http://purl.obolibrary.org/obo/FOODON_03450002), e.g., food blanching [FOODON:03470175]

- **food_dis_point:** A reference to a place on the Earth, by its name or by its geographical location that refers to a distribution point along the food chain. This field accepts terms listed under geographic location (http://purl.obolibrary.org/obo/GAZ_00000448). Reference: Adam Diamond, James Barham. Moving Food Along the Value Chain: Innovations in Regional Food Distribution. U.S. Dept. of Agriculture, Agricultural Marketing Service. Washington, DC. March 2012. <http://dx.doi.org/10.9752/MS045.03-2012>
- **food_dis_point_city:** A reference to a place on the Earth, by its name or by its geographical location that refers to a distribution point along the food chain. This field accepts terms listed under geographic location (http://purl.obolibrary.org/obo/GAZ_00000448). Reference: Adam Diamond, James Barham. Moving Food Along the Value Chain: Innovations in Regional Food Distribution. U.S. Dept. of Agriculture, Agricultural Marketing Service. Washington, DC. March 2012. <http://dx.doi.org/10.9752/MS045.03-2012>
- **food_ingredient:** In this field, please list individual ingredients for multi-component food [FOODON:00002501] and simple foods that is not captured in food_type. Please use terms that are present in FoodOn. Multiple terms can be separated by one or more pipes, but please consider limiting this list to the top 5 ingredients listed in order as on the food label. See also, <https://www.fda.gov/food/food-ingredients-packaging/overview-food-ingredients-additives-colors>
- **food_pack_capacity:** The maximum number of product units within a package
- **food_pack_integrity:** A term label and term id to describe the state of the packing material and text to explain the exact condition. This field encourages terms listed under food packing medium integrity (http://purl.obolibrary.org/obo/FOODON_03530218), e.g., food packing medium compromised [FOODON:00002517]
- **food_pack_medium:** The medium in which the food is packed for preservation and handling or the medium surrounding homemade foods, e.g., peaches cooked in sugar syrup. The packing medium may provide a controlled environment for the food. It may also serve to improve palatability and consumer appeal. This includes edible packing media (e.g. fruit juice), gas other than air (e.g. carbon dioxide), vacuum packed, or packed with aerosol propellant. This field encourages terms under food packing medium (http://purl.obolibrary.org/obo/FOODON_03480020). Multiple terms may apply and can be separated by pipes, e.g., vacuum-packed [FOODON:03480027]

- **food_preserv_proc:** The methods contributing to the prevention or retardation of microbial, enzymatic or oxidative spoilage and thus to the extension of shelf life. This field encourages terms listed under food preservation process (http://purl.obolibrary.org/obo/FOODON_03470107), e.g., food fermentation [FOODON:00001304]
- **food_prior_contact:** The material the food contacted (e.g., was processed in) prior to packaging. This field accepts terms listed under material of contact prior to food packaging (http://purl.obolibrary.org/obo/FOODON_03530077). If the proper descriptor is not listed please use text to describe the material of contact prior to food packaging
- **food_prod_synonym:** Other names by which the food product is known by (e.g., regional or non-English names), e.g., pinot gris
- **food_product_qual:** Descriptors for describing food visually or via other senses, which is useful for tasks like food inspection where little prior knowledge of how the food came to be is available. Some terms like "food (frozen)" are both a quality descriptor and the output of a process. This field accepts terms listed under food product by quality (http://purl.obolibrary.org/obo/FOODON_00002454)
- **food_quality_date:** The date recommended for the use of the product while at peak quality, this date is not a reflection of safety unless used on infant formula this date is not a reflection of safety and is typically labeled on a food product as ""best if used by,"" best by,"" ""use by,"" or ""freeze by"", e.g., 5/24/2020
- **food_source:** Individual organism or category of organisms from which the food product or its major ingredient is derived, e.g., giant tiger prawn [FOODON:03412612]
- **food_source_age:** The age of the food source host organism. Depending on the type of host organism, age may be more appropriate to report in days, weeks, or years
- **food_trace_list:** The FDA is proposing to establish additional traceability recordkeeping requirements (beyond what is already required in existing regulations) for persons who manufacture, process, pack, or hold foods the Agency has designated for inclusion on the Food Traceability List. The Food Traceability List (FTL) identifies the foods for which the additional traceability records described in the proposed rule would be required. The term ""Food Traceability List"" (FTL) refers not only to the foods specifically listed (<https://www.fda.gov/media/142303/download>), but also to any foods that contain listed foods as ingredients, e.g., tropical tree fruits
- **food_trav_mode:** A descriptor for the method of movement of food commodity along the food distribution system. This field accepts terms listed under travel mode

(http://purl.obolibrary.org/obo/GENEPIO_0001064). If the proper descriptor is not listed please use text to describe the mode of travel. Multiple terms can be separated by one or more pipes

- **food_trav_vehic:** A descriptor for the mobile machine which is used to transport food commodities along the food distribution system. This field accepts terms listed under vehicle (http://purl.obolibrary.org/obo/ENVO_01000604). If the proper descriptor is not listed please use text to describe the mode of travel. Multiple terms can be separated by one or more pipes
- **food_treat_proc:** Used to specifically characterize a food product based on the treatment or processes applied to the product or any indexed ingredient. The processes include adding, substituting or removing components or modifying the food or component, e.g., through fermentation. Multiple values can be assigned. This field accepts terms listed under food treatment process (http://purl.obolibrary.org/obo/FOODON_03460111)
- **haccp_term:** Hazard Analysis Critical Control Points (HACCP) food safety terms; This field accepts terms listed under HACCP guide food safety term (http://purl.obolibrary.org/obo/FOODON_03530221)
- **host_am:** The class(es) or name(s) (generic or brand) of the antimicrobial(s) given to the food animal within the last 30 days, e.g., tetracycline [CHEBI:27902]
- **host_group_size:** The number of food animals of the same species that are maintained together as a unit, i.e. a herd or flock, e.g., 80
- **host_housing:** Description of the housing system of the livestock. This field encourages terms listed under terrestrial management housing system (http://opendata.inra.fr/EOL/EOL_0001605), e.g., pen [EOL:0001902]
- **lot_number:** A distinctive alpha-numeric identification code assigned by the manufacturer or distributor to a specific quantity of manufactured material or product within a batch. The submitter should provide lot number of the item followed by the item name for which the lot number was provided
- **microb_cult_med:** A culture medium used to select for, grow, and maintain prokaryotic microorganisms. Can be in either liquid (broth) or solidified (e.g. with agar) forms. This field accepts terms listed under microbiological culture medium (http://purl.obolibrary.org/obo/MICRO_0000067). If the proper descriptor is not listed please use text to describe the culture medium

- **misc_param:** any other measurement performed or parameter collected, that is not listed here
- **organism_count:** total count of any organism per gram or volume of sample, should include name of organism followed by count; can include multiple organism counts
- **part_plant_animal:** The anatomical part of the organism being involved in food production or consumption; e.g., a carrot is the root of the plant (root vegetable). This field accepts terms listed under part of plant or animal (http://purl.obolibrary.org/obo/FOODON_03420116)
- **perturbation:** type of perturbation, e.g. chemical administration, physical disturbance, etc., coupled with time that perturbation occurred; can include multiple perturbation types
- **pool_dna_extracts:** were multiple DNA extractions mixed? how many?
- **repository:** the name of the institution where the sample or DNA extract is held or "sample not available" if the sample was used in its entirety for analysis or otherwise not retained
- **samp_collect_method:** The method employed for collecting the sample
- **samp_pooling:** Physical combination of several instances of like material, e.g., RNA extracted from samples or dishes of cell cultures into one big aliquot of cells. Please provide a short description of the samples that were pooled
- **samp_rep_biol:** Measurements of biologically distinct samples that show biological variation
- **samp_rep_tech:** Repeated measurements of the same sample that show independent measures of the noise associated with the equipment and the protocols
- **samp_source_mat_cat:** This is the scientific role or category that the subject organism or material has with respect to an investigation. This field accepts terms listed under specimen source material category (http://purl.obolibrary.org/obo/GENEPIO_0001237 or http://purl.obolibrary.org/obo/OBI_0100051)
- **samp_stor_device:** The container used to store the sample. This field accepts terms listed under container (http://purl.obolibrary.org/obo/NCIT_C43186). If the proper descriptor is not listed please use text to describe the storage device

- **samp_stor_media:** The liquid that is added to the sample collection device prior to sampling. If the sample is pre-hydrated, indicate the liquid media the sample is pre-hydrated with for storage purposes. This field accepts terms listed under microbiological culture medium (http://purl.obolibrary.org/obo/MICRO_0000067). If the proper descriptor is not listed please use text to describe the sample storage media
- **samp_store_dur:** Sample storage duration
- **samp_store_loc:** Sample storage location
- **samp_store_temp:** Sample storage temperature
- **samp_transport_cont:** Container in which the sample was stored during transport. Indicate the location name, e.g., bottle, cooler, glass vial, plastic vial, vendor supplied container
- **samp_transport_dur:** The duration of time from when the sample was collected until processed. Indicate the duration for which the sample was stored written in ISO 8601 format
- **samp_transport_temp:** Temperature at which sample was transported, e.g., -20 or 4 degree Celsius
- **serovar_or_serotype:** A characterization of a cell or microorganism based on the antigenic properties of the molecules on its surface. Indicate the name of a serovar or serotype of interest. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>
- **spikein_amr:** Qualitative description of a microbial response to antimicrobial agents. Bacteria may be susceptible or resistant to a broad range of antibiotic drugs or drug classes, with several intermediate states or phases. This field accepts terms under antimicrobial phenotype (http://purl.obolibrary.org/obo/ARO_3004299)
- **spikein_antibiotic:** Antimicrobials used in research study to assess effects of exposure on microbiome of a specific site. Please list antimicrobial, common name and/or class and concentration used for spike-in
- **spikein_count:** Total cell count of any organism (or group of organisms) per gram, volume or area of sample, should include name of organism followed by count. The method that was used for the enumeration (e.g., qPCR, atp, mpn, etc.) should also be provided, e.g., total prokaryotes; 3.5e7 cells per ml; qPCR

- **spikein_growth_med:** A liquid or gel containing nutrients, salts, and other factors formulated to support the growth of microorganisms, cells, or plants (National Cancer Institute Thesaurus). A growth medium is a culture medium which has the disposition to encourage growth of particular bacteria to the exclusion of others in the same growth environment. In this case, list the culture medium used to propagate the spike-in bacteria during preparation of spike-in inoculum. This field accepts terms listed under microbiological culture medium (http://purl.obolibrary.org/obo/MICRO_0000067). If the proper descriptor is not listed please use text to describe the spike in growth media
- **spikein_metal:** Heavy metals used in research study to assess effects of exposure on microbiome of a specific site. Please list heavy metals and concentration used for spike-in
- **spikein_org:** Taxonomic information about the spike-in organism(s). This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>. Multiple terms can be separated by pipes
- **spikein_serovar:** Taxonomic information about the spike-in organism(s) at the serovar or serotype level. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>. Multiple terms can be separated by pipes
- **spikein_strain:** Taxonomic information about the spike-in organism(s) at the strain level. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>. Multiple terms can be separated by pipes
- **study_design:** Epidemiological or omics research design context that this biosample was used in.
- **study_inc_dur:** Sample incubation duration if unpublished or unvalidated method is used. Indicate the timepoint written in ISO 8601 format
- **study_inc_temp:** Sample incubation temperature if unpublished or unvalidated method is used
- **study_timecourse:** For time-course research studies involving samples of the food commodity, indicate the total duration of the time-course study

- **study_tmnt:** A process in which the act is intended to modify or alter some other material entity. From the study design, each treatment is comprised of one level of one or multiple factors. This field accepts terms listed under treatment (http://purl.obolibrary.org/obo/MCO_0000866). If the proper descriptor is not listed please use text to describe the study treatment. Multiple terms can be separated by one or more pipes
- **temp:** temperature of the sample at time of sampling
- **timepoint:** Time point at which a sample or observation is made or taken from a biomaterial as measured from some reference point. Indicate the ti

4 Food-human foods:

Web Link for specific attributes: [BioSample Attributes](#) | [Submission Portal](#)



*** **Means Mandatory Value***** But please fill as many fields as possible!

- ***** coll_site_geo_feat ***:** Text or terms that describe the geographic feature where the food sample was obtained by the researcher. This field encourages selected terms listed under the following ontologies: anthropogenic geographic feature (http://purl.obolibrary.org/obo/ENVO_00000002), for example agricultural fairground [ENVO:01000986]; garden [ENVO:00000011] or any of its subclasses; market [ENVO:01000987]; water well [ENVO:01000002]; or human construction (http://purl.obolibrary.org/obo/ENVO_00000070), e.g., grocery store [GENEPIO:0001020]
- ***** food_product_type ***:** A food product type is a class of food products that is differentiated by its food composition (e.g., single- or multi-ingredient), processing and/or consumption characteristics. This does not include brand name products but it may include generic food dish categories. This field encourages terms under food product type (<http://purl.obolibrary.org/obo/FOODON:03400361>). For terms related to food product for an animal, consult food product for animal (http://purl.obolibrary.org/obo/FOODON_03309997). If the proper descriptor is not listed please use text to describe the food type. Multiple terms can be separated by one or more pipes, e.g., multi-component food product [FOODON:00002501]
- ***** ifsac_category ***:** The IFSAC food categorization scheme has five distinct levels to which foods can be assigned, depending upon the type of food. First, foods are assigned to one of four food groups (aquatic animals, land animals, plants, and other). Food groups include increasingly specific food categories; dairy, eggs, meat

and poultry, and game are in the land animal food group, and the category meat and poultry is further subdivided into more specific categories of meat (beef, pork, other meat) and poultry (chicken, turkey, other poultry). Finally, foods are differentiated by differences in food processing (such as pasteurized fluid dairy products, unpasteurized fluid dairy products, pasteurized solid and semi-solid dairy products, and unpasteurized solid and semi-solid dairy products). An IFSAC food category chart is available from <https://www.cdc.gov/foodsafety/ifsac/projects/food-categorization-scheme.html> PMID: 28926300, e.g., Plants:Produce:Vegetables:Herbs:Dried Herbs

- **bacterial_density:** Number of bacteria in sample, as defined by bacteria density (http://purl.obolibrary.org/obo/GENEPIO_0000043)
- **cons_food_stor_dur:** The storage duration of the food commodity by the consumer, prior to onset of illness or sample collection. Indicate the timepoint written in ISO 8601 format
- **cons_food_stor_temp:** Temperature at which food commodity was stored by the consumer, prior to onset of illness or sample collection
- **cons_purch_date:** The date a food product was purchased by consumer
- **cons_qty_purchased:** The quantity of food purchased by consumer
- **cult_isol_date:** A culture isolation date is a date-time entity marking the end of a process in which a sample yields a positive result for the target microbial analyte(s) in the form of an isolated colony or colonies, e.g., 5/24/2020
- **cult_result:** Any result of a bacterial culture experiment reported as a binary assessment, e.g., absent, active, inactive, negative, no, present, positive, yes
- **cult_result_org:** Taxonomic information about the cultured organism(s)
- **cult_target:** The target microbial analyte in terms of investigation scope. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>
- **dietary_claim_use:** The microbiological workflow or protocol followed to test for the presence or enumeration of the target microbial analyte(s). Please provide a PubMed or DOI reference for published protocols
- **enrichment_protocol:** The microbiological workflow or protocol followed to test for the presence or enumeration of the target microbial analyte(s). Please provide a

PubMed or DOI reference for published protocols

- **experimental_factor:** Variable aspect of experimental design
- **ferm_chem_add:** Any chemicals that are added to the fermentation process to achieve the desired final product
- **ferm_chem_add_perc:** The amount of chemical added to the fermentation process
- **ferm_headspace_oxy:** The amount of headspace oxygen in a fermentation vessel
- **ferm_medium:** The growth medium used for the fermented food fermentation process, which supplies the required nutrients. Usually this includes a carbon and nitrogen source, water, micronutrients and chemical additives
- **ferm_ph:** The pH of the fermented food fermentation process
- **ferm_rel_humidity:** The relative humidity of the fermented food fermentation process
- **ferm_temp:** The temperature of the fermented food fermentation process
- **ferm_time:** The time duration of the fermented food fermentation process
- **ferm_vessel:** The type of vessel used for containment of the fermentation
- **food_additive:** A substance or substances added to food to maintain or improve safety and freshness, to improve or maintain nutritional value, or improve taste, texture and appearance. This field encourages terms listed under food additive (http://purl.obolibrary.org/obo/FOODON_03412972). Multiple terms can be separated by one or more pipes, but please consider limiting this list to the top 5 ingredients listed in order as on the food label. See also, <https://www.fda.gov/food/food-ingredients-packaging/overview-food-ingredients-additives-colors>, e.g., xanthan gum [FOODON:03413321]
- **food_allergen_label:** A label indication that the product contains a recognized allergen. This field accepts terms listed under dietary claim or use (http://purl.obolibrary.org/obo/FOODON_03510213)
- **food_contact_surf:** The specific container or coating materials in direct contact with the food. Multiple values can be assigned. This field encourages terms listed under food contact surface (http://purl.obolibrary.org/obo/FOODON_03500010), e.g., aluminum surface [FOODON:03500042]

- **food_contain_wrap:** Type of container or wrapping defined by the main container material, the container form, and the material of the liner lids or ends. Also type of container or wrapping by form; prefer description by material first, then by form. This field encourages terms listed under food container or wrapping (http://purl.obolibrary.org/obo/FOODON_03490100), e.g., bottle or jar [FOODON:03490203]
- **food_cooking_proc:** The transformation of raw food by the application of heat. This field encourages terms listed under food cooking (http://purl.obolibrary.org/obo/FOODON_03450002), e.g., food blanching [FOODON:03470175]
- **food_dis_point:** A reference to a place on the Earth, by its name or by its geographical location that refers to a distribution point along the food chain. This field accepts terms listed under geographic location (http://purl.obolibrary.org/obo/GAZ_00000448). Reference: Adam Diamond, James Barham. Moving Food Along the Value Chain: Innovations in Regional Food Distribution. U.S. Dept. of Agriculture, Agricultural Marketing Service. Washington, DC. March 2012. <http://dx.doi.org/10.9752/MS045.03-2012>
- **food_ingredient:** In this field, please list individual ingredients for multi-component food [FOODON:00002501] and simple foods that is not captured in food_type. Please use terms that are present in FoodOn. Multiple terms can be separated by one or more pipes, but please consider limiting this list to the top 5 ingredients listed in order as on the food label. See also, <https://www.fda.gov/food/food-ingredients-packaging/overview-food-ingredients-additives-colors>
- **food_name_status:** A datum indicating that use of a food product name is regulated in some legal jurisdiction. This field accepts terms listed under food product name legal status (http://purl.obolibrary.org/obo/FOODON_03530087)
- **food_origin:** A reference to a place on the Earth, by its name or by its geographical location that describes the origin of the food commodity, either in terms of its cultivation or production. This field encourages terms listed under geographic location (http://purl.obolibrary.org/obo/GAZ_00000448), e.g., Thailand
- **food_pack_capacity:** The maximum number of product units within a package
- **food_pack_integrity:** A term label and term id to describe the state of the packing material and text to explain the exact condition. This field encourages terms listed under food packing medium integrity (http://purl.obolibrary.org/obo/FOODON_03530218), e.g., food packing medium compromised [FOODON:00002517]

- **food_pack_medium:** The medium in which the food is packed for preservation and handling or the medium surrounding homemade foods, e.g., peaches cooked in sugar syrup. The packing medium may provide a controlled environment for the food. It may also serve to improve palatability and consumer appeal. This includes edible packing media (e.g. fruit juice), gas other than air (e.g. carbon dioxide), vacuum packed, or packed with aerosol propellant. This field encourages terms under food packing medium (http://purl.obolibrary.org/obo/FOODON_03480020). Multiple terms may apply and can be separated by pipes, e.g., vacuum-packed [FOODON:03480027]
- **food_preserv_proc:** The methods contributing to the prevention or retardation of microbial, enzymatic or oxidative spoilage and thus to the extension of shelf life. This field encourages terms listed under food preservation process (http://purl.obolibrary.org/obo/FOODON_03470107), e.g., food fermentation [FOODON:00001304]
- **food_prior_contact:** The material the food contacted (e.g., was processed in) prior to packaging. This field accepts terms listed under material of contact prior to food packaging (http://purl.obolibrary.org/obo/FOODON_03530077). If the proper descriptor is not listed please use text to describe the material of contact prior to food packaging
- **food_prod_char:** Descriptors of the food production system such as wild caught, free-range, organic, free-range, industrial, dairy, beef
- **food_prod_synonym:** Other names by which the food product is known by (e.g., regional or non-English names), e.g., pinot gris
- **food_product_qual:** Descriptors for describing food visually or via other senses, which is useful for tasks like food inspection where little prior knowledge of how the food came to be is available. Some terms like "food (frozen)" are both a quality descriptor and the output of a process. This field accepts terms listed under food product by quality (http://purl.obolibrary.org/obo/FOODON_00002454)
- **food_quality_date:** The date recommended for the use of the product while at peak quality, this date is not a reflection of safety unless used on infant formula this date is not a reflection of safety and is typically labeled on a food product as ""best if used by,"" best by,"" ""use by,"" or ""freeze by"", e.g., 5/24/2020
- **food_source:** Individual organism or category of organisms from which the food product or its major ingredient is derived, e.g., giant tiger prawn [FOODON:03412612]

- **food_trace_list:** The FDA is proposing to establish additional traceability recordkeeping requirements (beyond what is already required in existing regulations) for persons who manufacture, process, pack, or hold foods the Agency has designated for inclusion on the Food Traceability List. The Food Traceability List (FTL) identifies the foods for which the additional traceability records described in the proposed rule would be required. The term "Food Traceability List" (FTL) refers not only to the foods specifically listed (<https://www.fda.gov/media/142303/download>), but also to any foods that contain listed foods as ingredients, e.g., tropical tree fruits
- **food_trav_mode:** A descriptor for the method of movement of food commodity along the food distribution system. This field accepts terms listed under travel mode (http://purl.obolibrary.org/obo/GENEPIO_0001064). If the proper descriptor is not listed please use text to describe the mode of travel. Multiple terms can be separated by one or more pipes
- **food_trav_vehic:** A descriptor for the mobile machine which is used to transport food commodities along the food distribution system. This field accepts terms listed under vehicle (http://purl.obolibrary.org/obo/ENVO_01000604). If the proper descriptor is not listed please use text to describe the mode of travel. Multiple terms can be separated by one or more pipes
- **food_treat_proc:** Used to specifically characterize a food product based on the treatment or processes applied to the product or any indexed ingredient. The processes include adding, substituting or removing components or modifying the food or component, e.g., through fermentation. Multiple values can be assigned. This field accepts terms listed under food treatment process (http://purl.obolibrary.org/obo/FOODON_03460111)
- **genetic_mod:** Genetic modifications of the genome of an organism, which may occur naturally by spontaneous mutation, or be introduced by some experimental means, e.g. specification of a transgene or the gene knocked-out or details of transient transfection
- **haccp_term:** Hazard Analysis Critical Control Points (HACCP) food safety terms; This field accepts terms listed under HACCP guide food safety term (http://purl.obolibrary.org/obo/FOODON_03530221)
- **intended_consume:** Food consumer type, human or animal, for which the food product is produced and marketed, e.g., human
- **lot_number:** A distinctive alpha-numeric identification code assigned by the manufacturer or distributor to a specific quantity of manufactured material or product

within a batch. The submitter should provide lot number of the item followed by the item name for which the lot number was provided

- **microb_cult_med:** A culture medium used to select for, grow, and maintain prokaryotic microorganisms. Can be in either liquid (broth) or solidified (e.g. with agar) forms. This field accepts terms listed under microbiological culture medium (http://purl.obolibrary.org/obo/MICRO_0000067). If the proper descriptor is not listed please use text to describe the culture medium
- **microb_start:** Any type of microorganisms used in food production. This field accepts terms listed under live organisms for food production (http://purl.obolibrary.org/obo/FOODON_0344453)
- **microb_start_count:** Total cell count of starter culture per gram, volume or area of sample and the method that was used for the enumeration (e.g. qPCR, atp, mpn, etc.) should also be provided, e.g., total prokaryotes; 3.5e7 cells per ml; qPCR
- **microb_start_inoc:** The amount of starter culture used to inoculate a new batch
- **microb_start_prep:** Information about the protocol or method used to prepare the starter inoculum
- **microb_start_source:** The source from which the microbial starter culture was sourced. If commercially supplied, list supplier
- **microb_start_taxid:** Please include Genus species and strain ID, if known of microorganisms used in food production. For complex communities, pipes can be used to separate two or more microbes
- **misc_param:** any other measurement performed or parameter collected, that is not listed here
- **num_samp_collect:** The number of samples collected during the current sampling event
- **organism_count:** total count of any organism per gram or volume of sample, should include name of organism followed by count; can include multiple organism counts
- **part_plant_animal:** The anatomical part of the organism being involved in food production or consumption; e.g., a carrot is the root of the plant (root vegetable). This field accepts terms listed under part of plant or animal (http://purl.obolibrary.org/obo/FOODON_03420116)

- **perturbation:** type of perturbation, e.g. chemical administration, physical disturbance, etc., coupled with time that perturbation occurred; can include multiple perturbation types
- **pool_dna_extracts:** were multiple DNA extractions mixed? how many?
- **purpose_of_sampling:** the reason that the sample was collected, e.g., active surveillance in response to an outbreak, active surveillance not initiated by an outbreak, clinical trial, cluster investigation, environmental assessment, farm sample, field trial, for cause, industry internal investigation, market sample, passive surveillance, population based studies, research, research and development
- **repository:** the name of the institution where the sample or DNA extract is held or "sample not available" if the sample was used in its entirety for analysis or otherwise not retained
- **samp_pooling:** Physical combination of several instances of like material, e.g., RNA extracted from samples or dishes of cell cultures into one big aliquot of cells. Please provide a short description of the samples that were pooled
- **samp_rep_biol:** Measurements of biologically distinct samples that show biological variation
- **samp_rep_tech:** Repeated measurements of the same sample that show independent measures of the noise associated with the equipment and the protocols
- **samp_source_mat_cat:** This is the scientific role or category that the subject organism or material has with respect to an investigation. This field accepts terms listed under specimen source material category (http://purl.obolibrary.org/obo/GENEPIO_0001237 or http://purl.obolibrary.org/obo/OBI_0100051)
- **samp_stor_device:** The container used to store the sample. This field accepts terms listed under container (http://purl.obolibrary.org/obo/NCIT_C43186). If the proper descriptor is not listed please use text to describe the storage device
- **samp_stor_media:** The liquid that is added to the sample collection device prior to sampling. If the sample is pre-hydrated, indicate the liquid media the sample is pre-hydrated with for storage purposes. This field accepts terms listed under microbiological culture medium (http://purl.obolibrary.org/obo/MICRO_0000067). If the proper descriptor is not listed please use text to describe the sample storage media

- **samp_store_dur:** Sample storage duration
- **samp_store_loc:** Sample storage location
- **samp_store_temp:** Sample storage temperature
- **samp_transport_cont:** Container in which the sample was stored during transport. Indicate the location name, e.g., bottle, cooler, glass vial, plastic vial, vendor supplied container
- **samp_transport_dur:** The duration of time from when the sample was collected until processed. Indicate the duration for which the sample was stored written in ISO 8601 format
- **samp_transport_temp:** Temperature at which sample was transported, e.g., -20 or 4 degree Celsius
- **serovar_or_serotype:** A characterization of a cell or microorganism based on the antigenic properties of the molecules on its surface. Indicate the name of a serovar or serotype of interest. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>
- **spikein_amr:** Qualitative description of a microbial response to antimicrobial agents. Bacteria may be susceptible or resistant to a broad range of antibiotic drugs or drug classes, with several intermediate states or phases. This field accepts terms under antimicrobial phenotype (http://purl.obolibrary.org/obo/ARO_3004299)
- **spikein_antibiotic:** Antimicrobials used in research study to assess effects of exposure on microbiome of a specific site. Please list antimicrobial, common name and/or class and concentration used for spike-in
- **spikein_count:** Total cell count of any organism (or group of organisms) per gram, volume or area of sample, should include name of organism followed by count. The method that was used for the enumeration (e.g., qPCR, atp, mpn, etc.) should also be provided, e.g., total prokaryotes; 3.5e7 cells per ml; qPCR
- **spikein_growth_med:** A liquid or gel containing nutrients, salts, and other factors formulated to support the growth of microorganisms, cells, or plants (National Cancer Institute Thesaurus). A growth medium is a culture medium which has the disposition to encourage growth of particular bacteria to the exclusion of others in the same growth environment. In this case, list the culture medium used to propagate the spike-in bacteria during preparation of spike-in inoculum. This field accepts terms listed

under microbiological culture medium

(http://purl.obolibrary.org/obo/MICRO_0000067). If the proper descriptor is not listed please use text to describe the spike in growth media

- **spikein_metal:** Heavy metals used in research study to assess effects of exposure on microbiome of a specific site. Please list heavy metals and concentration used for spike-in
- **spikein_org:** Taxonomic information about the spike-in organism(s). This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>. Multiple terms can be separated by pipes
- **spikein_serovar:** Taxonomic information about the spike-in organism(s) at the serovar or serotype level. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>. Multiple terms can be separated by pipes
- **spikein_strain:** Taxonomic information about the spike-in organism(s) at the strain level. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>. Multiple terms can be separated by pipes
- **study_design:** Epidemiological or omics research design context that this biosample was used in.
- **study_inc_dur:** Sample incubation duration if unpublished or unvalidated method is used. Indicate the timepoint written in ISO 8601 format
- **study_inc_temp:** Sample incubation temperature if unpublished or unvalidated method is used
- **study_timecourse:** For time-course research studies involving samples of the food commodity, indicate the total duration of the time-course study
- **study_tmnt:** A process in which the act is intended to modify or alter some other material entity. From the study design, each treatment is comprised of one level of one or multiple factors. This field accepts terms listed under treatment (http://purl.obolibrary.org/obo/MCO_0000866). If the proper descriptor is not listed please use text to describe the study treatment. Multiple terms can be separated by one or more pipes

- **temp:** temperature of the sample at time of sampling
- **timepoint:** Time point at which a sample or observation is made or taken from a biomaterial as measured from some reference point. Indicate the timepoint written in ISO 8601 format

5 Food-food production facility:

Web Link for specific attributes: [BioSample Attributes](#) | [Submission Portal](#)



*** **Means Mandatory Value***** But please fill as many fields as possible!

- ***** coll_site_geo_feat ***:** Text or terms that describe the geographic feature where the food sample was obtained by the researcher. This field encourages selected terms listed under the following ontologies: anthropogenic geographic feature (http://purl.obolibrary.org/obo/ENVO_00000002), for example agricultural fairground [ENVO:01000986]; garden [ENVO:00000011] or any of its subclasses; market [ENVO:01000987]; water well [ENVO:01000002]; or human construction (http://purl.obolibrary.org/obo/ENVO_00000070), e.g., grocery store [GENEPIO:0001020]
- ***** food_contact_surf ***:** The specific container or coating materials in direct contact with the food. Multiple values can be assigned. This field encourages terms listed under food contact surface (http://purl.obolibrary.org/obo/FOODON_03500010), e.g., aluminum surface [FOODON:03500042]
- ***** food_product_qual ***:** Descriptors for describing food visually or via other senses, which is useful for tasks like food inspection where little prior knowledge of how the food came to be is available. Some terms like "food (frozen)" are both a quality descriptor and the output of a process. This field accepts terms listed under food product by quality (http://purl.obolibrary.org/obo/FOODON_00002454)
- ***** food_product_type ***:** A food product type is a class of food products that is differentiated by its food composition (e.g., single- or multi-ingredient), processing and/or consumption characteristics. This does not include brand name products but it may include generic food dish categories. This field encourages terms under food product type (<http://purl.obolibrary.org/obo/FOODON:03400361>). For terms related to food product for an animal, consult food product for animal (http://purl.obolibrary.org/obo/FOODON_03309997). If the proper descriptor is not

listed please use text to describe the food type. Multiple terms can be separated by one or more pipes, e.g., multi-component food product [FOODON:00002501]

- ***** ifsac_category *****: The IFSAC food categorization scheme has five distinct levels to which foods can be assigned, depending upon the type of food. First, foods are assigned to one of four food groups (aquatic animals, land animals, plants, and other). Food groups include increasingly specific food categories; dairy, eggs, meat and poultry, and game are in the land animal food group, and the category meat and poultry is further subdivided into more specific categories of meat (beef, pork, other meat) and poultry (chicken, turkey, other poultry). Finally, foods are differentiated by differences in food processing (such as pasteurized fluid dairy products, unpasteurized fluid dairy products, pasteurized solid and semi-solid dairy products, and unpasteurized solid and semi-solid dairy products. An IFSAC food category chart is available from <https://www.cdc.gov/foodsafety/ifsac/projects/food-categorization-scheme.html> PMID: 28926300, e.g., Plants:Produce:Vegetables:Herbs:Dried Herbs
- ***** samp_mat_type *****: The type of material from which the sample was obtained. For the Hydrocarbon package, samples include types like core, rock trimmings, drill cuttings, piping section, coupon, pigging debris, solid deposit, produced fluid, produced water, injected water, swabs, etc. For the Food Package, samples are usually categorized as food, body products or tissues, or environmental material. This field accepts terms listed under environmental specimen (http://purl.obolibrary.org/obo/GENEPIO_0001246)
- ***** samp_source_mat_cat *****: This is the scientific role or category that the subject organism or material has with respect to an investigation. This field accepts terms listed under specimen source material category (http://purl.obolibrary.org/obo/GENEPIO_0001237 or http://purl.obolibrary.org/obo/OBI_0100051)
- ***** samp_stor_device *****: The container used to store the sample. This field accepts terms listed under container (http://purl.obolibrary.org/obo/NCIT_C43186). If the proper descriptor is not listed please use text to describe the storage device
- ***** samp_stor_media *****: The liquid that is added to the sample collection device prior to sampling. If the sample is pre-hydrated, indicate the liquid media the sample is pre-hydrated with for storage purposes. This field accepts terms listed under microbiological culture medium (http://purl.obolibrary.org/obo/MICRO_0000067). If the proper descriptor is not listed please use text to describe the sample storage media
- **air_temp**: temperature of the air at the time of sampling

- **area_samp_size:** The total amount or size (volume (ml), mass (g) or area (m2)) of sample collected
- **avg_occup:** Daily average occupancy of room. Indicate the number of person(s) daily occupying the sampling room
- **bacterial_density:** Number of bacteria in sample, as defined by bacteria density (http://purl.obolibrary.org/obo/GENEPIO_0000043)
- **biocide_used:** Substance intended for preventing, neutralizing, destroying, repelling, or mitigating the effects of any pest or microorganism; that inhibits the growth, reproduction, and activity of organisms, including fungal cells; decreases the number of fungi or pests present; deters microbial growth and degradation of other ingredients in the formulation. Indicate the biocide used on the location where the sample was taken. Multiple terms can be separated by pipes, e.g., Quaternary ammonium compound|SterBac
- **cult_isol_date:** A culture isolation date is a date-time entity marking the end of a process in which a sample yields a positive result for the target microbial analyte(s) in the form of an isolated colony or colonies, e.g., 5/24/2020
- **cult_result:** Any result of a bacterial culture experiment reported as a binary assessment, e.g., absent, active, inactive, negative, no, present, positive, yes
- **cult_result_org:** Taxonomic information about the cultured organism(s)
- **cult_target:** The target microbial analyte in terms of investigation scope. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>
- **dietary_claim_use:** These descriptors are used either for foods intended for special dietary use as defined in 21 CFR 105 or for foods that have special characteristics indicated in the name or labeling. This field accepts terms listed under dietary claim or use (http://purl.obolibrary.org/obo/FOODON_03510023). Multiple terms can be separated by one or more pipes, but please consider limiting this list to the most prominent dietary claim or use
- **enrichment_protocol:** The microbiological workflow or protocol followed to test for the presence or enumeration of the target microbial analyte(s). Please provide a PubMed or DOI reference for published protocols

- **env_monitoring_zone:** An environmental monitoring zone is a formal designation as part of an environmental monitoring program, in which areas of a food production facility are categorized, commonly as zones 1-4, based on likelihood or risk of foodborne pathogen contamination. This field accepts entries of zones 1-4, e.g., Zone 1
- **experimental_factor:** Variable aspect of experimental design
- **facility_type:** Establishment details about the type of facility where the sample was taken. This is independent of the specific product(s) within the facility, e.g., manufacturing-processing
- **food_additive:** A substance or substances added to food to maintain or improve safety and freshness, to improve or maintain nutritional value, or improve taste, texture and appearance. This field encourages terms listed under food additive (http://purl.obolibrary.org/obo/FOODON_03412972). Multiple terms can be separated by one or more pipes, but please consider limiting this list to the top 5 ingredients listed in order as on the food label. See also, <https://www.fda.gov/food/food-ingredients-packaging/overview-food-ingredients-additives-colors>, e.g., xanthan gum [FOODON:03413321]
- **food_allergen_label:** A label indication that the product contains a recognized allergen. This field accepts terms listed under dietary claim or use (http://purl.obolibrary.org/obo/FOODON_03510213)
- **food_contain_wrap:** Type of container or wrapping defined by the main container material, the container form, and the material of the liner lids or ends. Also type of container or wrapping by form; prefer description by material first, then by form. This field encourages terms listed under food container or wrapping (http://purl.obolibrary.org/obo/FOODON_03490100), e.g., bottle or jar [FOODON:03490203]
- **food_cooking_proc:** The transformation of raw food by the application of heat. This field encourages terms listed under food cooking (http://purl.obolibrary.org/obo/FOODON_03450002), e.g., food blanching [FOODON:03470175]
- **food_dis_point:** A reference to a place on the Earth, by its name or by its geographical location that refers to a distribution point along the food chain. This field accepts terms listed under geographic location (http://purl.obolibrary.org/obo/GAZ_00000448). Reference: Adam Diamond, James Barham. Moving Food Along the Value Chain: Innovations in Regional Food

Distribution. U.S. Dept. of Agriculture, Agricultural Marketing Service. Washington, DC. March 2012. <http://dx.doi.org/10.9752/MS045.03-2012>

- **food_dis_point_city:** A reference to a place on the Earth, by its name or by its geographical location that refers to a distribution point along the food chain. This field accepts terms listed under geographic location (http://purl.obolibrary.org/obo/GAZ_00000448). Reference: Adam Diamond, James Barham. Moving Food Along the Value Chain: Innovations in Regional Food Distribution. U.S. Dept. of Agriculture, Agricultural Marketing Service. Washington, DC. March 2012. <http://dx.doi.org/10.9752/MS045.03-2012>
- **food_ingredient:** In this field, please list individual ingredients for multi-component food [FOODON:00002501] and simple foods that is not captured in food_type. Please use terms that are present in FoodOn. Multiple terms can be separated by one or more pipes, but please consider limiting this list to the top 5 ingredients listed in order as on the food label. See also, <https://www.fda.gov/food/food-ingredients-packaging/overview-food-ingredients-additives-colors>
- **food_name_status:** A datum indicating that use of a food product name is regulated in some legal jurisdiction. This field accepts terms listed under food product name legal status (http://purl.obolibrary.org/obo/FOODON_03530087)
- **food_origin:** A reference to a place on the Earth, by its name or by its geographical location that describes the origin of the food commodity, either in terms of its cultivation or production. This field encourages terms listed under geographic location (http://purl.obolibrary.org/obo/GAZ_00000448), e.g., Thailand
- **food_pack_capacity:** The maximum number of product units within a package
- **food_pack_integrity:** A term label and term id to describe the state of the packing material and text to explain the exact condition. This field encourages terms listed under food packing medium integrity (http://purl.obolibrary.org/obo/FOODON_03530218), e.g., food packing medium compromised [FOODON:00002517]
- **food_pack_medium:** The medium in which the food is packed for preservation and handling or the medium surrounding homemade foods, e.g., peaches cooked in sugar syrup. The packing medium may provide a controlled environment for the food. It may also serve to improve palatability and consumer appeal. This includes edible packing media (e.g. fruit juice), gas other than air (e.g. carbon dioxide), vacuum packed, or packed with aerosol propellant. This field encourages terms under food packing medium (http://purl.obolibrary.org/obo/FOODON_03480020). Multiple terms may apply and can be separated by pipes, e.g., vacuum-packed [FOODON:03480027]

- **food_preserv_proc:** The methods contributing to the prevention or retardation of microbial, enzymatic or oxidative spoilage and thus to the extension of shelf life. This field encourages terms listed under food preservation process (http://purl.obolibrary.org/obo/FOODON_03470107), e.g., food fermentation [FOODON:00001304]
- **food_prior_contact:** The material the food contacted (e.g., was processed in) prior to packaging. This field accepts terms listed under material of contact prior to food packaging (http://purl.obolibrary.org/obo/FOODON_03530077). If the proper descriptor is not listed please use text to describe the material of contact prior to food packaging
- **food_prod_char:** Descriptors of the food production system such as wild caught, free-range, organic, free-range, industrial, dairy, beef
- **food_prod_synonym:** Other names by which the food product is known by (e.g., regional or non-English names), e.g., pinot gris
- **food_quality_date:** The date recommended for the use of the product while at peak quality, this date is not a reflection of safety unless used on infant formula this date is not a reflection of safety and is typically labeled on a food product as ""best if used by,"" best by,"" ""use by,"" or ""freeze by,"" e.g., 5/24/2020
- **food_source:** Individual organism or category of organisms from which the food product or its major ingredient is derived, e.g., giant tiger prawn [FOODON:03412612]
- **food_trace_list:** The FDA is proposing to establish additional traceability recordkeeping requirements (beyond what is already required in existing regulations) for persons who manufacture, process, pack, or hold foods the Agency has designated for inclusion on the Food Traceability List. The Food Traceability List (FTL) identifies the foods for which the additional traceability records described in the proposed rule would be required. The term ""Food Traceability List"" (FTL) refers not only to the foods specifically listed (<https://www.fda.gov/media/142303/download>), but also to any foods that contain listed foods as ingredients, e.g., tropical tree fruits
- **food_trav_mode:** A descriptor for the method of movement of food commodity along the food distribution system. This field accepts terms listed under travel mode (http://purl.obolibrary.org/obo/GENEPIO_0001064). If the proper descriptor is not listed please use text to describe the mode of travel. Multiple terms can be separated by one or more pipes

- **food_trav_vehic:** A descriptor for the mobile machine which is used to transport food commodities along the food distribution system. This field accepts terms listed under vehicle (http://purl.obolibrary.org/obo/ENVO_01000604). If the proper descriptor is not listed please use text to describe the mode of travel. Multiple terms can be separated by one or more pipes
- **food_treat_proc:** Used to specifically characterize a food product based on the treatment or processes applied to the product or any indexed ingredient. The processes include adding, substituting or removing components or modifying the food or component, e.g., through fermentation. Multiple values can be assigned. This field accepts terms listed under food treatment process (http://purl.obolibrary.org/obo/FOODON_03460111)
- **freq_clean:** The number of times the sample location is cleaned, e.g., daily, weekly, monthly, quarterly, annually
- **genetic_mod:** Genetic modifications of the genome of an organism, which may occur naturally by spontaneous mutation, or be introduced by some experimental means, e.g. specification of a transgene or the gene knocked-out or details of transient transfection
- **haccp_term:** Hazard Analysis Critical Control Points (HACCP) food safety terms; This field accepts terms listed under HACCP guide food safety term (http://purl.obolibrary.org/obo/FOODON_03530221)
- **hygienic_area:** The subdivision of areas within a food production facility according to hygienic requirements. This field accepts terms listed under hygienic food production area (<http://purl.obolibrary.org/obo/ENVO>). Please add a term that most accurately indicates the hygienic area your sample was taken from according to the definitions provided
- **indoor_surf:** type of indoor surface
- **intended_consumer:** Food consumer type, human or animal, for which the food product is produced and marketed, e.g., human
- **lot_number:** A distinctive alpha-numeric identification code assigned by the manufacturer or distributor to a specific quantity of manufactured material or product within a batch. The submitter should provide lot number of the item followed by the item name for which the lot number was provided
- **microb_cult_med:** A culture medium used to select for, grow, and maintain prokaryotic microorganisms. Can be in either liquid (broth) or solidified (e.g. with

agar) forms. This field accepts terms listed under microbiological culture medium (http://purl.obolibrary.org/obo/MICRO_0000067). If the proper descriptor is not listed please use text to describe the culture medium

- **misc_param:** any other measurement performed or parameter collected, that is not listed here
- **num_samp_collect:** The number of samples collected during the current sampling event
- **organism_count:** total count of any organism per gram or volume of sample, should include name of organism followed by count; can include multiple organism counts
- **part_plant_animal:** The anatomical part of the organism being involved in food production or consumption; e.g., a carrot is the root of the plant (root vegetable). This field accepts terms listed under part of plant or animal (http://purl.obolibrary.org/obo/FOODON_03420116)
- **pool_dna_extracts:** were multiple DNA extractions mixed? how many?
- **prod_label_claims:** Labeling claims containing descriptors such as wild caught, free-range, organic, free-range, industrial, hormone-free, antibiotic free, cage free
- **purpose_of_sampling:** the reason that the sample was collected, e.g., active surveillance in response to an outbreak, active surveillance not initiated by an outbreak, clinical trial, cluster investigation, environmental assessment, farm sample, field trial, for cause, industry internal investigation, market sample, passive surveillance, population based studies, research, research and development
- **repository:** the name of the institution where the sample or DNA extract is held or "sample not available" if the sample was used in its entirety for analysis or otherwise not retained
- **room_dim:** The length, width and height of sampling room
- **samp_collect_method:** The method employed for collecting the sample
- **samp_floor:** The floor of the building, where the sampling room is located, e.g., 1st floor, 2nd floor, basement, lobby
- **samp_loc_condition:** The condition of the sample location at the time of sampling, e.g., damaged, new, rupture, visible signs of mold-mildew, visible weariness repair

- **samp_pooling:** Physical combination of several instances of like material, e.g., RNA extracted from samples or dishes of cell cultures into one big aliquot of cells. Please provide a short description of the samples that were pooled
- **samp_rep_biol:** Measurements of biologically distinct samples that show biological variation
- **samp_rep_tech:** Repeated measurements of the same sample that show independent measures of the noise associated with the equipment and the protocols
- **samp_room_id:** Sampling room number. This ID should be consistent with the designations on the building floor plans
- **samp_store_dur:** Sample storage duration
- **samp_store_loc:** Sample storage location
- **samp_store_temp:** Sample storage temperature
- **samp_surf_moisture:** Degree of water held on a sampled surface, e.g., intermittent moisture, not present, submerged
- **samp_transport_cont:** Container in which the sample was stored during transport. Indicate the location name, e.g., bottle, cooler, glass vial, plastic vial, vendor supplied container
- **samp_transport_dur:** The duration of time from when the sample was collected until processed. Indicate the duration for which the sample was stored written in ISO 8601 format
- **samp_transport_temp:** Temperature at which sample was transported, e.g., -20 or 4 degree Celsius
- **ster_meth_samp_room:** The method used to sterilize the sampling room. This field accepts terms listed under electromagnetic radiation (http://purl.obolibrary.org/obo/ENVO_01001026). If the proper descriptor is not listed, please use text to describe the sampling room sterilization method. Multiple terms can be separated by pipes
- **study_design:** Epidemiological or omics research design context that this biosample was used in.

- **study_inc_dur:** Sample incubation duration if unpublished or unvalidated method is used. Indicate the timepoint written in ISO 8601 format
- **study_inc_temp:** Sample incubation temperature if unpublished or unvalidated method is used
- **study_timecourse:** For time-course research studies involving samples of the food commodity, indicate the total duration of the time-course study
- **study_tmnt:** A process in which the act is intended to modify or alter some other material entity. From the study design, each treatment is comprised of one level of one or multiple factors. This field accepts terms listed under treatment (http://purl.obolibrary.org/obo/MCO_0000866). If the proper descriptor is not listed please use text to describe the study treatment. Multiple terms can be separated by one or more pipes
- **surf_material:** surface materials at the point of sampling
- **timepoint:** Time point at which a sample or observation is made or taken from a biomaterial as measured from some reference point. Indicate the timepoint written in ISO 8601 format

6 Food-Farm environment:

Web Link for specific attributes: [BioSample Attributes](#) | [Submission Portal](#)



Food Farm Enviro.xlsx

*** **Means Mandatory Value***** But please fill as many fields as possible!

- ***** biotic_regm ***:** Information about treatment(s) involving use of biotic factors, such as bacteria, viruses or fungi
- ***** chem_administration ***:** list of chemical compounds administered to the host or site where sampling occurred, and when (e.g. antibiotics, N fertilizer, air filter); can include multiple compounds. For Chemical Entities of Biological Interest ontology (CHEBI) (v1.72), please see <http://bioportal.bioontology.org/visualize/44603>
- ***** depth ***:** Depth is defined as the vertical distance below surface, e.g. for sediment or soil samples depth is measured from sediment or soil surface, respectively. Depth can be reported as an interval for subsurface samples.
- ***** food_product_type ***:** A food product type is a class of food products that is differentiated by its food composition (e.g., single- or multi-ingredient), processing and/or consumption characteristics. This does not include brand name products but it

may include generic food dish categories. This field encourages terms under food product type (<http://purl.obolibrary.org/obo/FOODON:03400361>). For terms related to food product for an animal, consult food product for animal (http://purl.obolibrary.org/obo/FOODON_03309997). If the proper descriptor is not listed please use text to describe the food type. Multiple terms can be separated by one or more pipes, e.g., multi-component food product [FOODON:00002501

- ***** ifsac_category *****: The IFSAC food categorization scheme has five distinct levels to which foods can be assigned, depending upon the type of food. First, foods are assigned to one of four food groups (aquatic animals, land animals, plants, and other). Food groups include increasingly specific food categories; dairy, eggs, meat and poultry, and game are in the land animal food group, and the category meat and poultry is further subdivided into more specific categories of meat (beef, pork, other meat) and poultry (chicken, turkey, other poultry). Finally, foods are differentiated by differences in food processing (such as pasteurized fluid dairy products, unpasteurized fluid dairy products, pasteurized solid and semi-solid dairy products, and unpasteurized solid and semi-solid dairy products). An IFSAC food category chart is available from <https://www.cdc.gov/foodsafety/ifsac/projects/food-categorization-scheme.html> PMID: 28926300, e.g., Plants:Produce:Vegetables:Herbs:Dried Herbs
- ***** samp_mat_type *****: The type of material from which the sample was obtained. For the Hydrocarbon package, samples include types like core, rock trimmings, drill cuttings, piping section, coupon, pigging debris, solid deposit, produced fluid, produced water, injected water, swabs, etc. For the Food Package, samples are usually categorized as food, body products or tissues, or environmental material. This field accepts terms listed under environmental specimen (http://purl.obolibrary.org/obo/GENEPIO_0001246)
- **adjacent_environment**: Description of the environmental system or features that are adjacent to the sampling site. This field accepts terms under ecosystem (http://purl.obolibrary.org/obo/ENVO_01001110) and human construction (http://purl.obolibrary.org/obo/ENVO_00000070). Multiple terms can be separated by pipes
- **air_flow_impede**: Presence of objects in the area that would influence or impede air flow through the air filter, e.g., obstructed, unobstructed
- **air_pm_concen**: concentration of substances that remain suspended in the air, and comprise mixtures of organic and inorganic substances (PM10 and PM2.5); can report multiple PM's by entering numeric values preceded by name of PM
- **ances_data**: Information about either pedigree or other ancestral information description, e.g., parental variety in case of mutant or selection, A/3*B (meaning [(A x

B) x B] x B)

- **anim_water_method:** Description of the equipment or method used to distribute water to livestock. This field accepts terms listed under water delivery equipment (http://opendata.inra.fr/EOL/EOL_0001653). Multiple terms can be separated by pipes
- **animal_diet:** If the isolate is from a food animal, the type of diet eaten by the food animal. Please list the main food staple and the setting, if appropriate. For a list of acceptable animal feed terms or categories, please see <http://www.feedipedia.org>. Multiple terms may apply and can be separated by pipes. Food product for animal covers foods intended for consumption by domesticated animals. Consult http://purl.obolibrary.org/obo/FOODON_03309997. If the proper descriptor is not listed please use text to describe the food type. Multiple terms can be separated by one or more pipes. If the proper descriptor is not listed please use text to describe the food product type
- **animal_feed equip:** Description of the feeding equipment used for livestock. This field accepts terms listed under feed delivery (http://opendata.inra.fr/EOL/EOL_0001757). Multiple terms can be separated by one or more pipes
- **animal_intrusion:** Identification of animals intruding on the sample or sample site including invertebrates (such as pests or pollinators) and vertebrates (such as wildlife or domesticated animals). This field encourages terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also encourages identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>. Multiple terms can be separated by pipes, e.g., large flies
- **conduc:** electrical conductivity of water
- **crop_rotation:** whether or not crop is rotated, and if yes, rotation schedule
- **crop_yield:** Amount of crop produced per unit or area of land
- **cult_isol_date:** A culture isolation date is a date-time entity marking the end of a process in which a sample yields a positive result for the target microbial analyte(s) in the form of an isolated colony or colonies, e.g., 5/24/2020
- **cult_result:** Any result of a bacterial culture experiment reported as a binary assessment, e.g., absent, active, inactive, negative, no, present, positive, yes
- **cult_result_org:** Taxonomic information about the cultured organism(s)

- **cult_target:** The target microbial analyte in terms of investigation scope. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>
- **date_extr_weath:** Date of unusual weather events that may have affected microbial populations. Multiple terms can be separated by pipes, listed in reverse chronological order
- **enrichment_protocol:** The microbiological workflow or protocol followed to test for the presence or enumeration of the target microbial analyte(s). Please provide a PubMed or DOI reference for published protocols
- **extr_weather_event:** Unusual weather events that may have affected microbial populations. Multiple terms can be separated by pipes, listed in reverse chronological order, e.g., hail
- **farm equip:** List of relevant equipment used for planting, fertilization, harvesting, irrigation, land levelling, residue management, weeding or transplanting during the growing season. This field encourages terms listed under agricultural implement (http://purl.obolibrary.org/obo/AGRO_00000416). Multiple terms can be separated by pipes, e.g., combine harvester [AGRO:00000473]
- **farm equip_san:** Method used to sanitize growing and harvesting equipment including type and concentration of sanitizing solution and frequency of sanitization
- **farm equip_san_freq:** The number of times farm equipment is cleaned. Frequency of cleaning might be on a daily basis, weekly, monthly, quarterly or annually
- **farm equip_shared:** List of planting, growing or harvesting equipment shared with other farms
- **farm_water_source:** Source of water used on the farm for irrigation of crops or watering of livestock, e.g., water well
- **fertilizer_admin:** Type of fertilizer or amendment added to the soil or water for the purpose of improving substrate health and quality for plant growth. This field encourages terms listed under agronomic fertilizer (http://purl.obolibrary.org/obo/AGRO_00002062). Multiple terms may apply and can be separated by pipes, listing in reverse chronological order, e.g., fish emulsion [AGRO:00000082]

- **fertilizer_date:** Date of administration of soil amendment or fertilizer. Multiple terms may apply and can be separated by pipes, listing in reverse chronological order
- **food_clean_proc:** The process of cleaning food to separate other environmental materials from the food source. Multiple terms can be separated by pipes, e.g., rinsed with water|scrubbed with brush
- **food_contact_surf:** The specific container or coating materials in direct contact with the food. Multiple values can be assigned. This field encourages terms listed under food contact surface (http://purl.obolibrary.org/obo/FOODON_03500010), e.g., aluminum surface [FOODON:03500042]
- **food_contain_wrap:** Type of container or wrapping defined by the main container material, the container form, and the material of the liner lids or ends. Also type of container or wrapping by form; prefer description by material first, then by form. This field encourages terms listed under food container or wrapping (http://purl.obolibrary.org/obo/FOODON_03490100), e.g., bottle or jar [FOODON:03490203]
- **food_harvest_proc:** A harvesting process is a process which takes in some food material from an individual or community of plant or animal organisms in a given context and time, and outputs a precursor or consumable food product. This may include a part of an organism or the whole, and may involve killing the organism
- **food_pack_medium:** The medium in which the food is packed for preservation and handling or the medium surrounding homemade foods, e.g., peaches cooked in sugar syrup. The packing medium may provide a controlled environment for the food. It may also serve to improve palatability and consumer appeal. This includes edible packing media (e.g. fruit juice), gas other than air (e.g. carbon dioxide), vacuum packed, or packed with aerosol propellant. This field encourages terms under food packing medium (http://purl.obolibrary.org/obo/FOODON_03480020). Multiple terms may apply and can be separated by pipes, e.g., vacuum-packed [FOODON:03480027]
- **food_preserv_proc:** The methods contributing to the prevention or retardation of microbial, enzymatic or oxidative spoilage and thus to the extension of shelf life. This field encourages terms listed under food preservation process (http://purl.obolibrary.org/obo/FOODON_03470107), e.g., food fermentation [FOODON:00001304]
- **food_prod_char:** Descriptors of the food production system such as wild caught, free-range, organic, free-range, industrial, dairy, beef

- **food_quality_date:** The date recommended for the use of the product while at peak quality, this date is not a reflection of safety unless used on infant formula this date is not a reflection of safety and is typically labeled on a food product as ""best if used by,"" best by,"" ""use by,"" or ""freeze by"", e.g., 5/24/2020
- **food_source:** Individual organism or category of organisms from which the food product or its major ingredient is derived, e.g., giant tiger prawn [FOODON:03412612]
- **food_trav_mode:** A descriptor for the method of movement of food commodity along the food distribution system. This field accepts terms listed under travel mode (http://purl.obolibrary.org/obo/GENEPIO_0001064). If the proper descriptor is not listed please use text to describe the mode of travel. Multiple terms can be separated by one or more pipes
- **food_trav_vehic:** A descriptor for the mobile machine which is used to transport food commodities along the food distribution system. This field accepts terms listed under vehicle (http://purl.obolibrary.org/obo/ENVO_01000604). If the proper descriptor is not listed please use text to describe the mode of travel. Multiple terms can be separated by one or more pipes
- **food_treat_proc:** Used to specifically characterize a food product based on the treatment or processes applied to the product or any indexed ingredient. The processes include adding, substituting or removing components or modifying the food or component, e.g., through fermentation. Multiple values can be assigned. This fields accepts terms listed under food treatment process (http://purl.obolibrary.org/obo/FOODON_03460111)
- **genetic_mod:** Genetic modifications of the genome of an organism, which may occur naturally by spontaneous mutation, or be introduced by some experimental means, e.g. specification of a transgene or the gene knocked-out or details of transient transfection
- **growth_habit:** Characteristic shape, appearance or growth form of a plant species, e.g., erect, semi-erect, spreading, prostrate
- **growth_medium:** A liquid or gel containing nutrients, salts, and other factors formulated to support the growth of microorganisms, cells, or plants (National Cancer Institute Thesaurus). The name of the medium used to grow the microorganism
- **host_age:** Age of host at the time of sampling
- **host_dry_mass:** measurement of dry mass

- **host_genotype:**
- **host_group_size:** The number of food animals of the same species that are maintained together as a unit, i.e. a herd or flock, e.g., 80
- **host_height:** the height of subject
- **host_housing:** Description of the housing system of the livestock. This field encourages terms listed under terrestrial management housing system (http://opendata.inra.fr/EOL/EOL_0001605), e.g., pen [EOL:0001902]
- **host_length:** the length of subject
- **host_phenotype:** Host phenotype
- **host_subspecf_genlin:** Information about the genetic distinctness of the host organism below the subspecies level e.g., serovar, serotype, biotype, ecotype, variety, cultivar, or any relevant genetic typing schemes like Group I plasmid. Subspecies should not be recorded in this term, but in the NCBI taxonomy. Supply both the lineage name and the lineage rank separated by a colon, e.g., biovar:abc123
- **host_tot_mass:** total mass of the host at collection, the unit depends on host
- **humidity:** amount of water vapour in the air, at the time of sampling
- **intended_consumer:** Food consumer type, human or animal, for which the food product is produced and marketed, e.g., human
- **lot_number:** A distinctive alpha-numeric identification code assigned by the manufacturer or distributor to a specific quantity of manufactured material or product within a batch. The submitter should provide lot number of the item followed by the item name for which the lot number was provided
- **mechanical_damage:** information about any mechanical damage exerted on the plant; can include multiple damages and sites
- **misc_param:** any other measurement performed or parameter collected, that is not listed here
- **organism_count:** total count of any organism per gram or volume of sample, should include name of organism followed by count; can include multiple organism counts

- **part_plant_animal:** The anatomical part of the organism being involved in food production or consumption; e.g., a carrot is the root of the plant (root vegetable). This field accepts terms listed under part of plant or animal (http://purl.obolibrary.org/obo/FOODON_03420116)
- **perturbation:** type of perturbation, e.g. chemical administration, physical disturbance, etc., coupled with time that perturbation occurred; can include multiple perturbation types
- **ph:** type of perturbation, e.g. chemical administration, physical disturbance, etc., coupled with time that perturbation occurred; can include multiple perturbation types
- **ph_meth:** reference or method used in determining pH
- **plant_growth_med:** Type of the media used for growing sampled plants, e.g., soil [ENVO:00001998]
- **plant_part_maturity:** A description of the stage of development of a plant or plant part based on maturity or ripeness. This field accepts terms listed under degree of plant maturity (http://purl.obolibrary.org/obo/FOODON_03530050)
- **plant_reprod_crop:** Plant reproductive part used in the field during planting to start the crop, e.g., plant cutting, pregerminated seed, ratoon, seed, seedling, whole mature plant
- **plant_water_method:** Description of the equipment or method used to distribute water to crops. This field encourages terms listed under irrigation process (http://purl.obolibrary.org/obo/AGRO_00000006). Multiple terms can be separated by pipes, e.g., drip irrigation process [AGRO:00000056]
- **previous_land_use:** previous land use and dates
- **prod_label_claims:** Labeling claims containing descriptors such as wild caught, free-range, organic, free-range, industrial, hormone-free, antibiotic free, cage free
- **purpose_of_sampling:** the reason that the sample was collected, e.g., active surveillance in response to an outbreak, active surveillance not initiated by an outbreak, clinical trial, cluster investigation, environmental assessment, farm sample, field trial, for cause, industry internal investigation, market sample, passive surveillance, population based studies, research, research and development
- **rel_location:** Location of sample to other parts of the farm, e.g. under crop plant, near irrigation ditch, from the dirt road, from air above crops, nearby river, e.g., furrow

- **repository:** the name of the institution where the sample or DNA extract is held or "sample not available" if the sample was used in its entirety for analysis or otherwise not retained
- **root_cond:** Relevant rooting conditions such as field plot size, sowing density, container dimensions, number of plants per container
- **root_med_carbon:** Source of organic carbon in the culture rooting medium, e.g., sucrose
- **root_med_macronutr:** Measurement of the culture rooting medium macronutrients (N,P, K, Ca, Mg, S), e.g., KH₂PO₄ (170mg/L)
- **root_med_micronutr:** Measurement of the culture rooting medium micronutrients (Fe, Mn, Zn, B, Cu, Mo), e.g., H₃BO₃ (6.2mg/L)
- **root_med_ph:** Measurement of the culture rooting medium micronutrients (Fe, Mn, Zn, B, Cu, Mo), e.g., H₃BO₃ (6.2mg/L)
- **salinity:** salinity measurement
- **salinity_meth:** reference or method used in determining salinity
- **samp_pooling:** Physical combination of several instances of like material, e.g., RNA extracted from samples or dishes of cell cultures into one big aliquot of cells. Please provide a short description of the samples that were pooled
- **samp_source_mat_cat:** This is the scientific role or category that the subject organism or material has with respect to an investigation. This field accepts terms listed under specimen source material category (http://purl.obolibrary.org/obo/GENEPIO_0001237 or http://purl.obolibrary.org/obo/OBI_0100051)
- **samp_store_dur:** Sample storage duration
- **samp_store_temp:** Sample storage temperature
- **season:** The season when sampling occurred. Any of the four periods into which the year is divided by the equinoxes and solstices. This field accepts terms listed under season (http://purl.obolibrary.org/obo/NCIT_C94729)
- **season_humidity:** Average humidity of the region throughout the growing season

- **season_precpt:** The average of all seasonal precipitation values known, or an estimated equivalent value derived by such methods as regional indexes or Isohyetal maps
- **season_temp:** Mean seasonal temperature
- **serovar_or_serotype:** A characterization of a cell or microorganism based on the antigenic properties of the molecules on its surface. Indicate the name of a serovar or serotype of interest. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>
- **size_frac_low:** Refers to the mesh/pore size used to pre-filter/pre-sort the sample. Materials larger than the size threshold are excluded from the sample
- **size_frac_up:** Refers to the mesh/pore size used to retain the sample. Materials smaller than the size threshold are excluded from the sample
- **soil_conductivity:** Conductivity of soil at time of sampling
- **soil_cover:** Description of the material covering the sampled soil. This field accepts terms under ENVO:00010483, environmental material
- **soil_ph:** The pH of soil at time of sampling
- **soil_porosity:** Porosity of soil or deposited sediment is volume of voids divided by the total volume of sample
- **soil_temp:** Temperature of soil at the time of sampling
- **soil_texture_class:** One of the 12 soil texture classes use to describe soil texture based on the relative proportion of different grain sizes of mineral particles [sand (50 um to 2 mm), silt (2 um to 50 um), and clay (<2 um)] in a soil, e.g., clay, clay loam, loam, loamy sand, sand, sandy clay, sandy clay loam, sandy loam, silt, silty clay, silty clay loam, silt loam
- **soil_texture_meth:** reference or method used in determining soil texture
- **soil_type:** soil series name or other lower-level classification
- **soil_type_meth:** reference or method used in determining soil series name or other lower-level classification

- **solar_irradiance:** the amount of solar energy that arrives at a specific area of a surface during a specific time interval
- **spikein_antibiotic:** Antimicrobials used in research study to assess effects of exposure on microbiome of a specific site. Please list antimicrobial, common name and/or class and concentration used for spike-in
- **spikein_count:** Total cell count of any organism (or group of organisms) per gram, volume or area of sample, should include name of organism followed by count. The method that was used for the enumeration (e.g., qPCR, atp, mpn, etc.) should also be provided, e.g., total prokaryotes; 3.5e7 cells per ml; qPCR
- **spikein_growth_med:** A liquid or gel containing nutrients, salts, and other factors formulated to support the growth of microorganisms, cells, or plants (National Cancer Institute Thesaurus). A growth medium is a culture medium which has the disposition to encourage growth of particular bacteria to the exclusion of others in the same growth environment. In this case, list the culture medium used to propagate the spike-in bacteria during preparation of spike-in inoculum. This field accepts terms listed under microbiological culture medium (http://purl.obolibrary.org/obo/MICRO_0000067). If the proper descriptor is not listed please use text to describe the spike in growth media
- **spikein_metal:** Heavy metals used in research study to assess effects of exposure on microbiome of a specific site. Please list heavy metals and concentration used for spike-in
- **spikein_org:** Taxonomic information about the spike-in organism(s). This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>. Multiple terms can be separated by pipes
- **spikein_serovar:** Taxonomic information about the spike-in organism(s) at the serovar or serotype level. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>. Multiple terms can be separated by pipes
- **spikein_strain:** Taxonomic information about the spike-in organism(s) at the strain level. This field accepts terms under organism (http://purl.obolibrary.org/obo/NCIT_C14250). This field also accepts identification numbers from NCBI under <https://www.ncbi.nlm.nih.gov/taxonomy>. Multiple terms can be separated by pipes

- **temp:** temperature of the sample at time of sampling
- **tillage:** note method(s) used for tilling
- **timepoint:** Time point at which a sample or observation is made or taken from a biomaterial as measured from some reference point. Indicate the timepoint written in ISO 8601 format
- **tot_nitro:** total nitrogen content of the sample
- **tot_nitro_cont_meth:** Reference or method used in determining the total nitrogen
- **tot_org_c_meth:** reference or method used in determining total organic C
- **tot_org_carb:** Definition for soil: total organic C content of the soil units of g C/kg soil. Definition otherwise: total organic carbon content
- **turbidity:** turbidity measurement
- **ventilation_rate:** ventilation rate of the system in the sampled premises
- **ventilation_type:** ventilation system used in the sampled premises
- **water_frequency:** Number of water delivery events within a given period of time
- **water_ph:** The pH measurement of the sample, or liquid portion of sample, or aqueous phase of the fluid
- **water_source_adjac:** Description of the environmental features that are adjacent to the farm water source. This field accepts terms under ecosystem (http://purl.obolibrary.org/obo/ENVO_01001110) and human construction (http://purl.obolibrary.org/obo/ENVO_00000070). Multiple terms can be separated by pipes
- **water_source_shared:** Other users sharing access to the same water source, e.g., multiple users, agricultural, multiple users, no sharing
- **wind_direction:** wind direction is the direction from which a wind originates
- **wind_speed:** speed of wind measured at the time of sampling

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 your populated **SRA metadata template** you downloaded from the packages section.

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**".

7.1 Files tab:

Each laboratory will establish its own path for transferring files. Select the radio button corresponding to the means you will use.

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. Firewalls may prevent use of Aspera or AWS routes of submission.

Note

It is generally not recommended to check the "Autofinish submission" box as this would not allow you to make corrections, if needed.

Sequence Read Archive (SRA) submission: SUB14157251

Listeria monocytogenes Raw sequence reads, Jan 17 '24



Files

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

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

- ☐ Web browser upload via HTTP or Aspera Connect plugin
Do not use web browser HTTP upload if you are uploading files over 10 GB or more than 300 files.
- ☐ FTP or Aspera Command Line file preload
All files for a submission must be uploaded into a single folder.
- ☐ AWS or GCP bucket

☐ Autofinish submission

- Try to complete the submission automatically. Basic checks will be done on the files in the background. If there are no errors, this submission will be submitted for you. If the checks fail, you will get email notification to come back to fix the problems.

Continue

7.2 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.

Note

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

7.3 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.

7.4 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.

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

Submission Portal

Home My submissions Manage data Groups Templates My profile

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

7.5 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. Some edits can be made within the submission portal and others need to be done via email.

Safety information

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 a new run was submitted. It's important to keep track of both IDs, even if one was retracted.

BioProject Creation

8 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, and 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 portion of the 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.

8.1

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

The screenshot shows the NCBI Submission Portal interface. At the top, the NIH logo and 'U.S. National Library of Medicine' are visible. The 'Submission Portal' header includes navigation links: Home, My submissions (active), Manage data, Groups, Templates, and My profile. A 'New' button is also present. Below the header, the 'Your submissions' section contains a 'Start a new submission' box on the left. This box lists various submission types: GenBank, Sequence Read Archive, Genome, TSA, BioProject (highlighted with a red rectangle), BioSample, Supplementary Files, and API. To the right of this box is a search and filter area with fields for 'From date', 'To date', 'Status' (set to 'Not deleted'), and 'Sort by'. Below these are buttons for 'Apps' and 'Data archives'. A 'Query' search bar with 'Search' and 'Clear' buttons is also present. At the bottom, a table displays submission details. The table has columns for Submission, Title, App, Group, Status, and Updated. The first row shows a submission with ID 'SUB6510311', title 'UI-less submission 2019-11-04', App 'API', Group 'fda', Status 'BioSample: Processed Successfully loaded SAMN13192483 (TaxID: 1639)', and Updated '07:15'.

Submission	Title	App	Group	Status	Updated
SUB6510311	UI-less submission 2019-11-04	API	fda	BioSample: Processed Successfully loaded SAMN13192483 (TaxID: 1639)	07:15

8.2 Click the “New submission” box:

U.S. National Library of Medicine
National Center for Biotechnology Information

Submission Portal

Home My submissions Manage data Groups Templates My profile

BioProject New submission

ATTN: to update an existing record or recent submission, please email your request with your BioProject ID or Submission ID included. Do not create new submission to update an existing submission!

Short description and brief instructions +

Filter / Search

From date To date Status Not deleted Sort by desc

Data archives +

Query Search Clear

187 submissions

Submission	Title	Group	Status	Updated
SUB8320199	Listeria monocytogenes Database: isolates from FDA-CFSAN	fda	BioProject: Processed PRJNA215355: Listeria monocytogenes Database: isolates from FDA-CFSAN (TaxID: 1639) locustagrefix.txt	Nov 03

8.3 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.

8.4 Project type tab:

Project data type: Genome sequencing and assembly.

Sample scope:

For a **Data BioProject**: Select multi-species. This will allow you to submit multiple different species to the BioProject.

8.5 Target tab:

For a **Data BioProject**: Populate ONLY the Organism name here:

For targeted-pathogen BioProjects:

Organism name = Include a Genus name, e.g., *Salmonella* sp.

For non-targeted pathogens

Organism name = "bacteria"



Create a description of the scope of the project (e.g. "enteric bacteria").

8.6 **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: environmental.

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.

Note

Note: We advise against linking data BioProjects to multiple umbrella BioProjects.

8.7 **BioSample tab:**

Leave blank!! You will create biosamples separately.

8.8 **Publications tab:**

If relevant, include publications from your laboratory.

8.9 **Review and Submit tab:**

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



Submission Portal

BioProject submission: SUB12915263

FDA BioProject: bacterial pathogens



This submission was copied from SUB12914688.

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

Review & Submit

This BioProject submission will be released **immediately following processing**.

Submitter

Submitter: Ruth Timme
ruth.timme@fda.hhs.gov
Submitting organization: US Food and Drug Administration

Project type

Sample scope: Multispecies
Target description: Bacterial pathogens not included in the major taxa scoped for GenomeTrakr.

Target

Organism name (taxid): bacterial pathogens

General information

Project details

Parent project ID: PRJNA706684
Parent project description: GenomeTrakr umbrella for diverse pathogen species
Project type: genome-sequencing
Title: FDA BioProject: bacterial pathogens
Description: Whole genome sequencing of pure-cultured bacterial pathogens for FDA surveillance and outbreak investigations.
Relevance: Environmental

Example for a new non-targeted BioProject

- 8.10 The **BioProject accession** will be available within a few minutes on the “My Submissions” page of the Submission portal in the format “PRJNAxxxxxx.” You will also receive an email containing the new accession.

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

Submission Portal

Home My submissions Manage data Groups Templates My profile

Your submissions

Start a new submission

- GenBank
- Sequence Read Archive
- Genome
- TSA
- BioProject
- BioSample
- Supplementary Files
- ASR

Filter / Search

From date To date Status Not deleted Sort by desc

Apps + Data archives +

Query Q Search Clear

92,767 submissions

Submission	Title	App	Group	Status	Updated
SUB12915263	GenomeTrakr Project: Minnesota Department of Health	BioProject	10a	BioProject: Processed	10/23