

Oct 04, 2024 Version 2

Populating metadata templates for NCBI submissions using PulseNet 2.0 V.2

DOI

dx.doi.org/10.17504/protocols.io.3byl4qn4ovo5/v2

Ruth Timme¹, Maria Balkey¹, Tina Lusk Pfefer¹, Candace Hope Bias²

¹US Food and Drug Administration; ²US FDA-CFSAN



Maria Balkey

US Food and Drug Administration

OPEN  ACCESS



DOI: **dx.doi.org/10.17504/protocols.io.3byl4qn4ovo5/v2**

Protocol Citation: Ruth Timme, Maria Balkey, Tina Lusk Pfefer, Candace Hope Bias 2024. Populating metadata templates for NCBI submissions using PulseNet 2.0. **protocols.io** **<https://dx.doi.org/10.17504/protocols.io.3byl4qn4ovo5/v2>** Version created by **Maria Balkey**

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

Protocol status: Working

We use this protocol and it's working

Created: February 09, 2024

Last Modified: October 04, 2024

Protocol Integer ID: 107187

Keywords: NCBI submission, biosample, SRA, metadata, bioproject, PulseNet, PulseNet 2.0



Abstract

PURPOSE: to define the standard operating procedure for collecting isolate metadata using PulseNet 2.0 for submission of food/environmental isolates to NCBI.

SCOPE: to provide a standardized procedure to collect isolate metadata using PulseNet 2.0 for submission of food/environmental isolates to NCBI.

RESPONSIBILITIES: SOP Responsible Officials: Ruth Timme, Maria Balkey

The GenomeTrakr Network Management will be responsible to monitor GenomeTrakr submissions processed through PulseNet 2.0 and ensure that all GT labs are familiar with the mandatory metadata fields required for submission of GenomeTrakr sequencing records to NCBI.


V3: Added dropdown menus from controlled vocabulary to sequenced by and project name to metadata template *PulseNet_Bionumerics_Isolate_Metadata*

V4: Changes in metadata template *PulseNet_Bionumerics_Isolate_Metadata*.

- Added dropdown menus from controlled vocabulary to collected_by , SourceCountryState
- Added fields: collected by, isolation source
- Added mapping table of attribute names.
- Remove requirement to send biosample update to NCBI to make changes on sequenced by and project name.

V5: Changes implemented to handle NCBI submissions for GenomeTrakr through the PulseNet 2.0 platform.

Safety warnings

 We will be updating this protocol with each new PulseNet 2.0 upgrade that impacts NCBI submissions



1 Data Ownership

Laboratory scientists that use PN2.0 to submit GenomeTrakr data to NCBI should be included in the PulseNet NCBI submission group.

Please reach out to PulseNet to ensure your name is included in this group.

Membership in the PulseNet submission group will grant you the correct privileges for viewing and managing your GenomeTrakr submissions within the NCBI submission portal. If you login to NCBI and ***don't*** see your GenomeTrakr submissions, you likely not a member.

2 Metadata SampleSheet preparation

Populate the metadata spreadsheet form before uploading your sequencing run or linking NCBI sequencing records within PulseNet 2.0.

Please download the following metadata template, Guidance is included in the first tab.



MetadataTemplatePN20.xlsx 172KB

Once you have filled out the template, import metadata according **PulseNet Instructions**

Note

In order to access **PulseNet Instructions**, make sure you have successfully logged in at the SAMS Partner Portal.

3 Creation of NCBI Submission Templates

Before submitting data to NCBI, set up **sample specific** submission templates in PulseNet 2.0 according to **PulseNet Instructions**, following GenomeTrakr mapping guidelines included in this protocol.

Map each field name to the corresponding field value as described in steps

⇒ [go to step #3.1](#) for Food isolates

⇒ [go to step #3.2](#) for Food Facility inspection isolates

⇒ [go to step #3.3](#) for Animal isolates

⇒ [go to step #3.4](#) for Farm or other environmental isolates.

Make sure to add the the GenomeTrakr-required fields to each template depending on the source type designation, using the **Add Field** functionality in the Pipeline Template section.

The creation of NCBI Submission Templates in PulseNet 2.0 will only be done once, then the saved templates can be used for all future submissions.

3.1 NCBI Template for **Food Isolates**

A	B	C
Field Name	Field Type	Field Value
Strain name	Mapping	WGS_id
Submitter Provided	Mapping	Key
SPUID namespace	Value	GenomeTrakr
BioProject accession	Mapping	BioProject
Author	Mapping	Sequenced_by
Serovar	Mapping	SEROTYPE_WGS
SourceType	Mapping	SourceType
<i>Isolate Name Alias if SourceType equals Human</i>	<i>Value</i>	<i>missing</i>
Isolate Name Alias if SourceType does not equal Human	Mapping	Key
Isolation source if SourceType does not equal Human	Mapping	SourceSite, TypeDetails
<i>Isolation source if SourceType equals Human</i>	<i>Mapping</i>	<i>missing</i>
<i>Geographical origin if SourceType equals Human</i>	<i>Mapping</i>	<i>missing</i>
Geographical origin if SourceType does not equal Human	Mapping	Geo_loc_name
Organism name	Mapping	Organism
<i>Collection / isolation date if SourceType equals Human</i>	<i>Value</i>	<i>missing</i>
Collection / isolation date if SourceType does not equal Human	Mapping	IsolatDate YYYY-mm-DD
<i>Collected by if SourceType equals Human</i>	<i>Mapping</i>	<i>missing</i>
Collected by if SourceType does not equal Human	Mapping	LabID
Instrument model	Mapping	Instrument_Model
Library strategy	Value	WGS
Library source	Value	GENOMIC
Library selection	Value	RANDOM
Library layout	Value	PAIRED



A	B	C
Library name	Mapping	Library
Project name	Mapping	Project
<i>Sequenced by if SourceType equals Human</i>	<i>Value</i>	<i>missing</i>
Sequenced by if SourceType does not equal Human	Mapping	Sequenced_by
ADD FIELD...		
intended consumer	Mapping	Intended_consumer
food product origin geographic location	Mapping	Food_origin
food processing method	Mapping	Food_processing_method
purpose of sampling	Mapping	Purpose_of_sampling

Save submission template according to [PulseNet Instructions](#) as **-GenomeTrakr-FoodGroup-**.

3.2 NCBI Template for **Food Facility Environment Isolates**

A	B	C
Field Name	Field Type	Field Value
Strain name	Mapping	WGS_id
Submitter Provided	Mapping	Key
SPUID namespace	Value	GenomeTrakr
BioProject accession	Mapping	BioProject
Author	Mapping	Sequenced_by
Serovar	Mapping	SEROTYPE_WGS
SourceType	Mapping	SourceType
<i>Isolate Name Alias if SourceType equals Human</i>	<i>Value</i>	<i>missing</i>
Isolate Name Alias if SourceType does not equal Human	Mapping	Key
Isolation source if SourceType does not equal Human	Mapping	SourceSite, TypeDetails
<i>Isolation source if SourceType equals Human</i>	<i>Mapping</i>	<i>missing</i>
<i>Geographical origin if SourceType equals Human</i>	<i>Mapping</i>	<i>missing</i>
Geographical origin if SourceType does not equal Human	Mapping	Geo_loc_name
Organism name	Mapping	Organism



A	B	C
Collection / isolation date if SourceType equals Human	Value	missing
Collection / isolation date if SourceType does not equal Human	Mapping	IsolatDate YYYY-mm-DD
Collected by if SourceType equals Human	Mapping	missing
Collected by if SourceType does not equal Human	Mapping	LabID
Instrument model	Mapping	Instrument_Model
Library strategy	Value	WGS
Library source	Value	GENOMIC
Library selection	Value	RANDOM
Library layout	Value	PAIRED
Library name	Mapping	Library
Project name	Mapping	Project
Sequenced by if SourceType equals Human	Value	missing
Sequenced by if SourceType does not equal Human	Mapping	Sequenced_by
ADD FIELD...		
facility type	Mapping	Facility_type
food type processed in facility	Mapping	Food_type_processed
purpose of sampling	Mapping	Purpose_of_sampling

Save submission template according to **PulseNet Instructions** as **-GenomeTrakr-FoodFacilityEnvironment-**.

3.3 NCBI Template for **Animal Isolates**

A	B	C
Field Name	Field Type	Field Value
Strain name	Mapping	WGS_id
Submitter Provided	Mapping	Key
SPUID namespace	Value	GenomeTrakr
BioProject accession	Mapping	BioProject
Author	Mapping	Sequenced_by
Serovar	Mapping	SEROTYPE_WGS
SourceType	Mapping	SourceType
Isolate Name Alias if SourceType equals Human	Value	missing



A	B	C
Isolate Name Alias if SourceType does not equal Human	Mapping	Key
Isolation source if SourceType does not equal Human	Mapping	SourceSite, TypeDetails
<i>Isolation source if SourceType equals Human</i>	<i>Mapping</i>	<i>missing</i>
<i>Geographical origin if SourceType equals Human</i>	<i>Mapping</i>	<i>missing</i>
Geographical origin if SourceType does not equal Human	Mapping	Geo_loc_name
Organism name	Mapping	Organism
<i>Collection / isolation date if SourceType equals Human</i>	<i>Value</i>	<i>missing</i>
Collection / isolation date if SourceType does not equal Human	Mapping	IsolatDate YYYY-mm-DD
<i>Collected by if SourceType equals Human</i>	<i>Mapping</i>	<i>missing</i>
Collected by if SourceType does not equal Human	Mapping	LabID
Instrument model	Mapping	Instrument_Model
Library strategy	Value	WGS
Library source	Value	GENOMIC
Library selection	Value	RANDOM
Library layout	Value	PAIRED
Library name	Mapping	Library
Project name	Mapping	Project
<i>Sequenced by if SourceType equals Human</i>	<i>Value</i>	<i>missing</i>
Sequenced by if SourceType does not equal Human	Mapping	Sequenced_by
ADD FIELD...		
host	Mapping	Host
animal environment	Mapping	Animal_env
purpose of sampling	Mapping	Purpose_of_sampling

Save submission template according to **PulseNet Instructions** as **-GenomeTrakr-AnimalHost-**.

3.4 NCBI Template for **Farm Environment Isolates**



A	B	C
Field Name	Field Type	Field Value
Strain name	Mapping	WGS_id
Submitter Provided	Mapping	Key
SPUID namespace	Value	GenomeTrakr
BioProject accession	Mapping	BioProject
Author	Mapping	Sequenced_by
Serovar	Mapping	SEROTYPE_WGS
SourceType	Mapping	SourceType
<i>Isolate Name Alias if SourceType equals Human</i>	<i>Value</i>	<i>missing</i>
Isolate Name Alias if SourceType does not equal Human	Mapping	Key
Isolation source if SourceType does not equal Human	Mapping	SourceSite, TypeDetails
<i>Isolation source if SourceType equals Human</i>	<i>Mapping</i>	<i>missing</i>
<i>Geographical origin if SourceType equals Human</i>	<i>Mapping</i>	<i>missing</i>
Geographical origin if SourceType does not equal Human	Mapping	Geo_loc_name
Organism name	Mapping	Organism
<i>Collection / isolation date if SourceType equals Human</i>	<i>Value</i>	<i>missing</i>
Collection / isolation date if SourceType does not equal Human	Mapping	IsolatDate YYYY-mm-DD
<i>Collected by if SourceType equals Human</i>	<i>Mapping</i>	<i>missing</i>
Collected by if SourceType does not equal Human	Mapping	LabID
Instrument model	Mapping	Instrument_Model
Library strategy	Value	WGS
Library source	Value	GENOMIC
Library selection	Value	RANDOM
Library layout	Value	PAIRED
Library name	Mapping	Library
Project name	Mapping	Project
<i>Sequenced by if SourceType equals Human</i>	<i>Value</i>	<i>missing</i>
Sequenced by if SourceType does not equal Human	Mapping	Sequenced_by



A	B	C
ADD FIELD...		
local-scale environmental context	Mapping	Env_local_scale
environmental medium	Mapping	Env_medium
purpose of sampling	Mapping	Purpose_of_sampling

Save submission template according to **PulseNet Instructions** as **-GenomeTrakr-FarmEnvironment-**.

4 **NCBI Submissions**

Submit data to NCBI according to **PulseNet Instructions**, if you have several submissions from different sample types, make sure to group records accordingly and select the appropriate custom **NCBI GenomeTrakr-labeled Template**. If NCBI accessions are not available at PulseNet 2.0 in 1 business day, please contact NCBI (**pd-help@ncbi.nlm.nih.gov**) and PulseNet to troubleshoot issues with submissions.

Contact GenomeTrakr by email **genometrakr@fda.hhs.gov** if issues with submissions are delayed for more than 3 days. GenomeTrakr can support urgent submissions if needed.