FEB 26, 2024

DOI:
dx.doi.org/10.17504/protocols.io.36wgq3kbklk5/v1

Protocol Citation: Maria Balkey, Ruth Timme, Tina Lusk Pfefer, Candace Hope Bias 2024. Querying for Bacterial Pathogen Genomic Data at NCBI. **protocols.io** https://dx.doi.org/10.17504/protocols.io.36wgq3kbklk5/v1

Protocol status: Working
We use this protocol and it's working

Querying for Bacterial Pathogen Genomic Data at NCBI

Maria Balkey[1], Ruth Timme[1], Tina Lusk Pfefer[1], Candace Hope Bias[1]

[1]US Food and Drug Administration

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

Maria Balkey

ABSTRACT

PURPOSE: This document provides detailed instructions on how to find bacterial pathogen genomic data and associated contextual information at NCBI, specifically at the BioSample, SRA, and Pathogen Detection databases.

SCOPE: This protocol is intended for use by any laboratory submitting WGS data of bacterial pathogens to NCBI for analysis within NCBI Pathogen Detection. This includes US labs connected to GenomeTrakr, NARMS, Vet-LIRN, NAHLN, and other international networks and submitters.

BEFORE START INSTRUCTIONS

This protocol has three sections:

- **Section 1:** NCBI-SRA Run Selector
- **Section 2:** NCBI-BioSample
- **Section 3:** NCBI-SRA
- **Section 4**: NCBI-Pathogen Detection

Created: Jan 16, 2024

Last Modified: Feb 26, 2024

PROTOCOL integer ID: 93618

Keywords: genomic pathogen surveillance, INSDC, NCBI

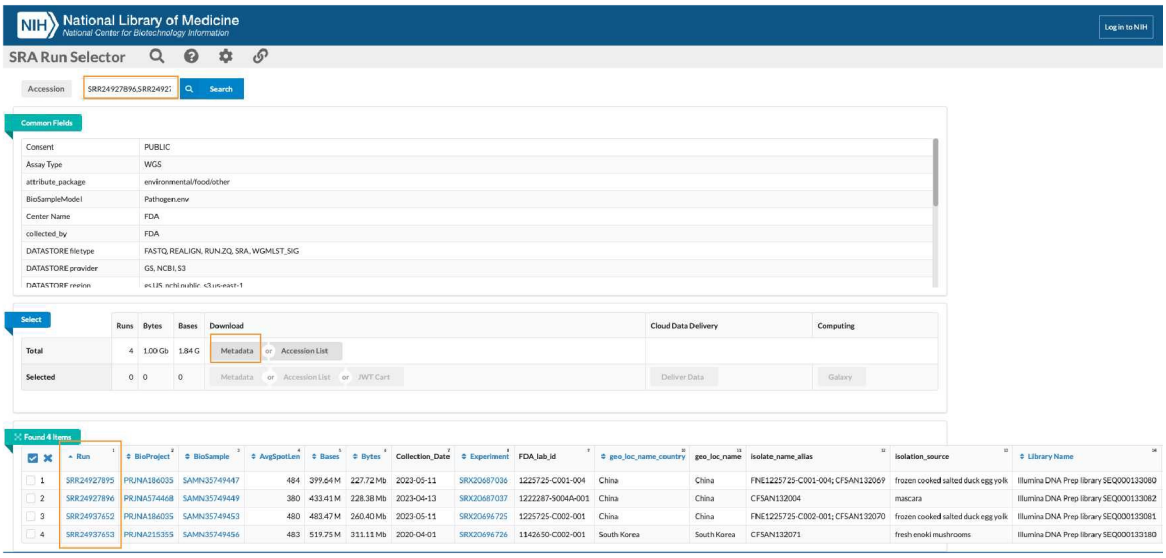
NCBI SRA (Sequence Read Archive) Run Sector

1 WGS submissions of pathogen genomes will become public and searchable almost immediately in SRA Run Selector.

NCBI's **SRA Run Selector** serves as a platform for querying contextual data, or metadta, from *both* BioSample and SRA. This is a good first place to check for a recent submission, obtaining a single table containing a suite of NCBI accessions and metadata for both samples and sequence data. For our application, users could enter a single or comma-separated list of IDs.

Example query with SRA run accessions: SRR24927895,SRR24927896,SRR24937652,SRR24937653

Additionally, a BioProject accession (or multiple BioProject accessions) can be entered to retrieve a complete table of submissions linked to that BioProject(s).



Isolate and experimental metadata displayed in tabular format.

Download table: Click on the -Metadata- button to download the tab-delimited file.

NCBI Sequencing Read Archive (SRA)

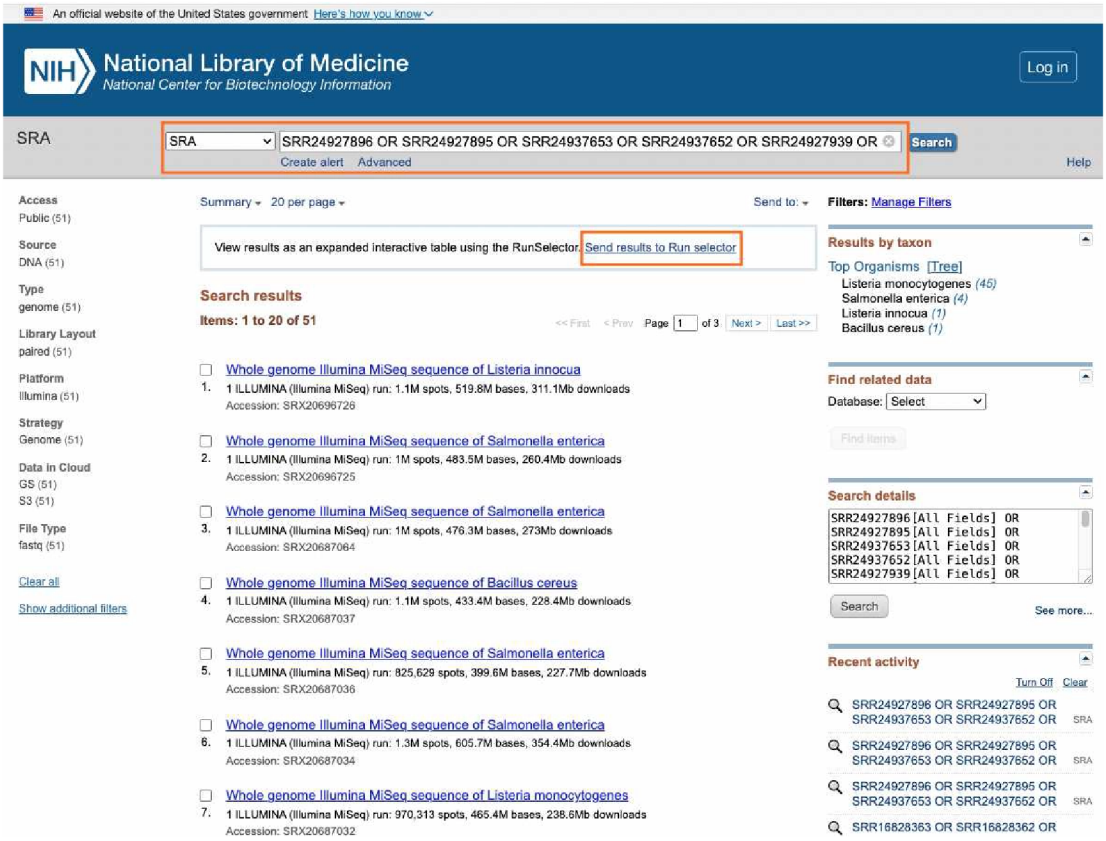
2 Submissions should be available in SRA within 1-3 days.

Strain identifiers and/or NCBI accessions can be used as queries. Multiple isolate identifiers can be included in the search box by using the " OR " separator.

Example with SRA run accessions: SRR24927896 OR SRR24927895 OR SRR24937653 OR SRR24937652.

Users can click on the SRA records themselves, or they can send the results to SRA run selector for a tabular output.

Click **-Send results to Run selector-**

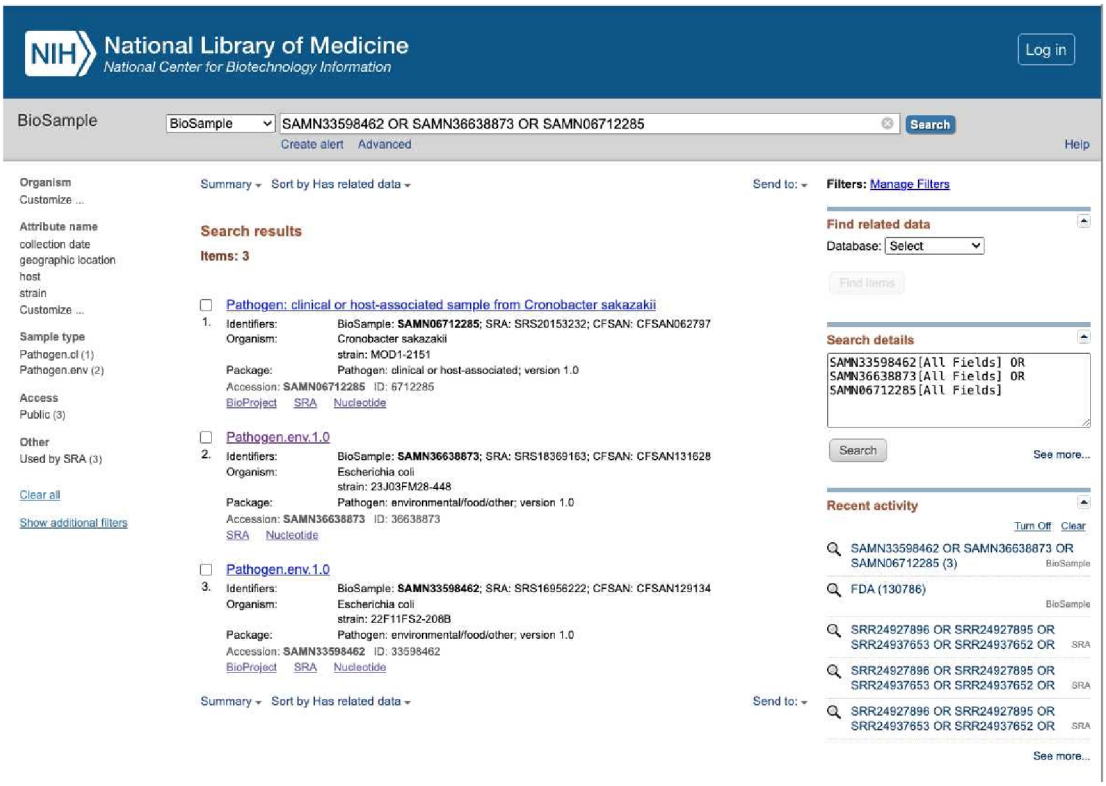


NCBI BioSample

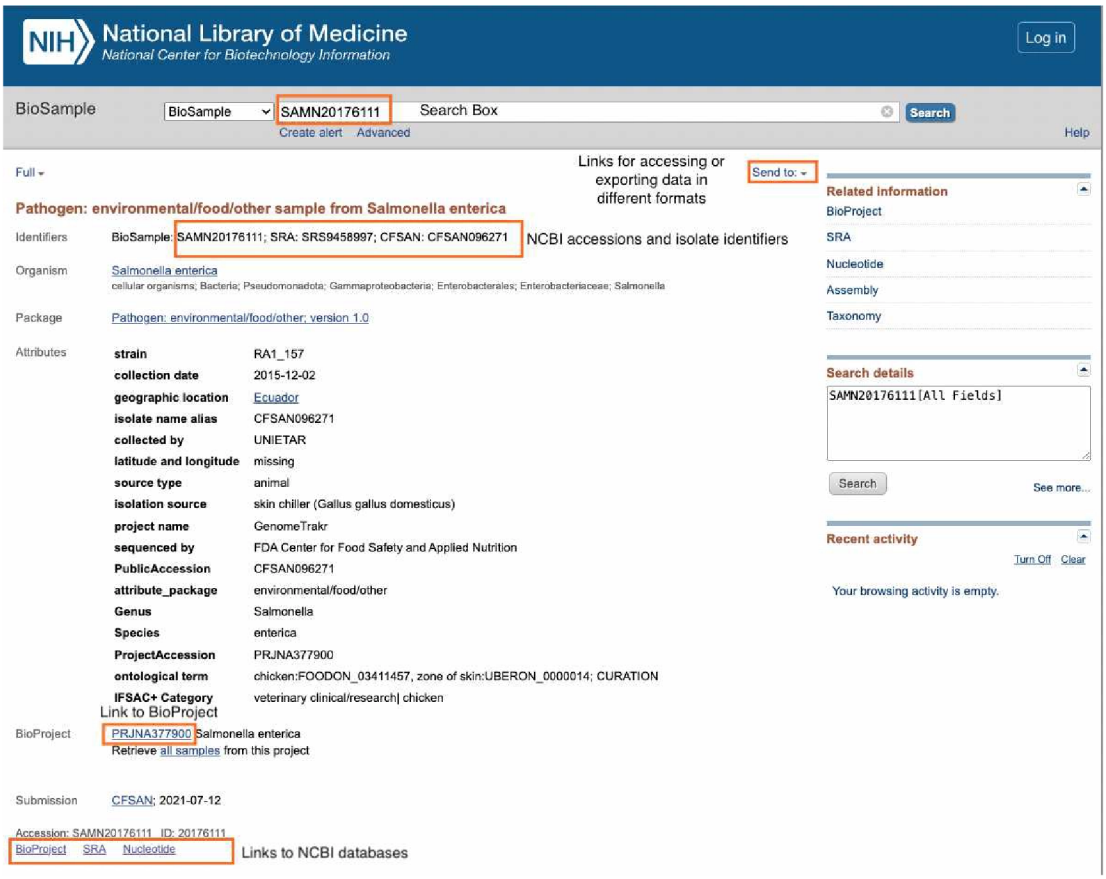
3 Submissions should be available in BioSample within 1-3 days.

Strain identifiers and/or NCBI accessions can be used to query NCBI BioSample. Multiple isolate identifiers can be included in the search box using the "OR" separator.

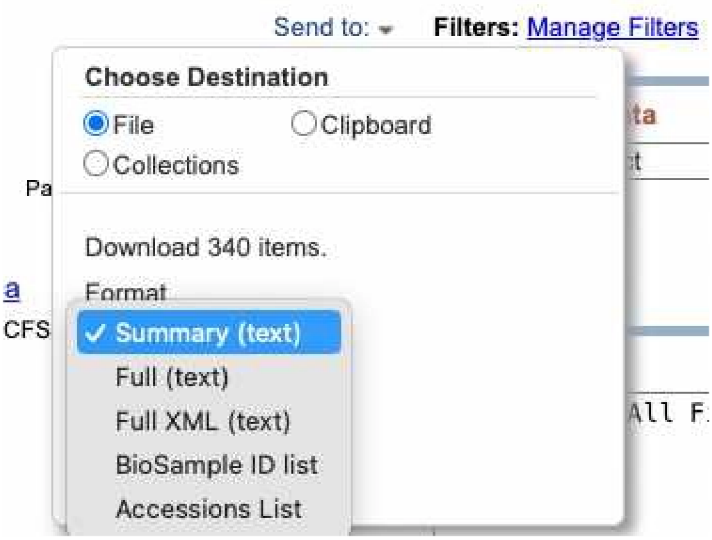
Example with BioSample accessions: SAMN33598462 OR SAMN36638873 OR SAMN06712285



Users can click to view each BioSample query result to review the metadata submitted for that sample (or isolate for our use case). Each BioSample record also includes links to the connected BioProject, and sequenced data generated from that BioSample, for example, raw reads in SRA, and/or genome assemblies in the Nucleotide database, or GenBank.



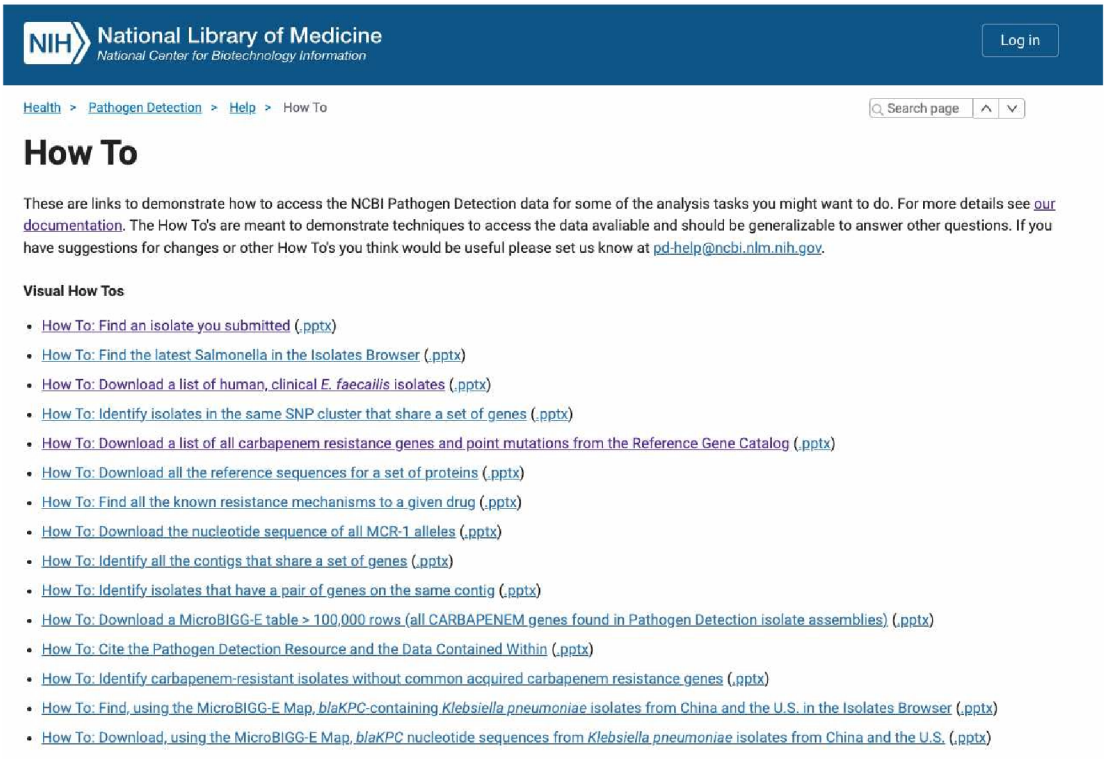
BioSample query results can be exported in different formats by clicking on **-Send To-**.



NCBI Pathogen Detection

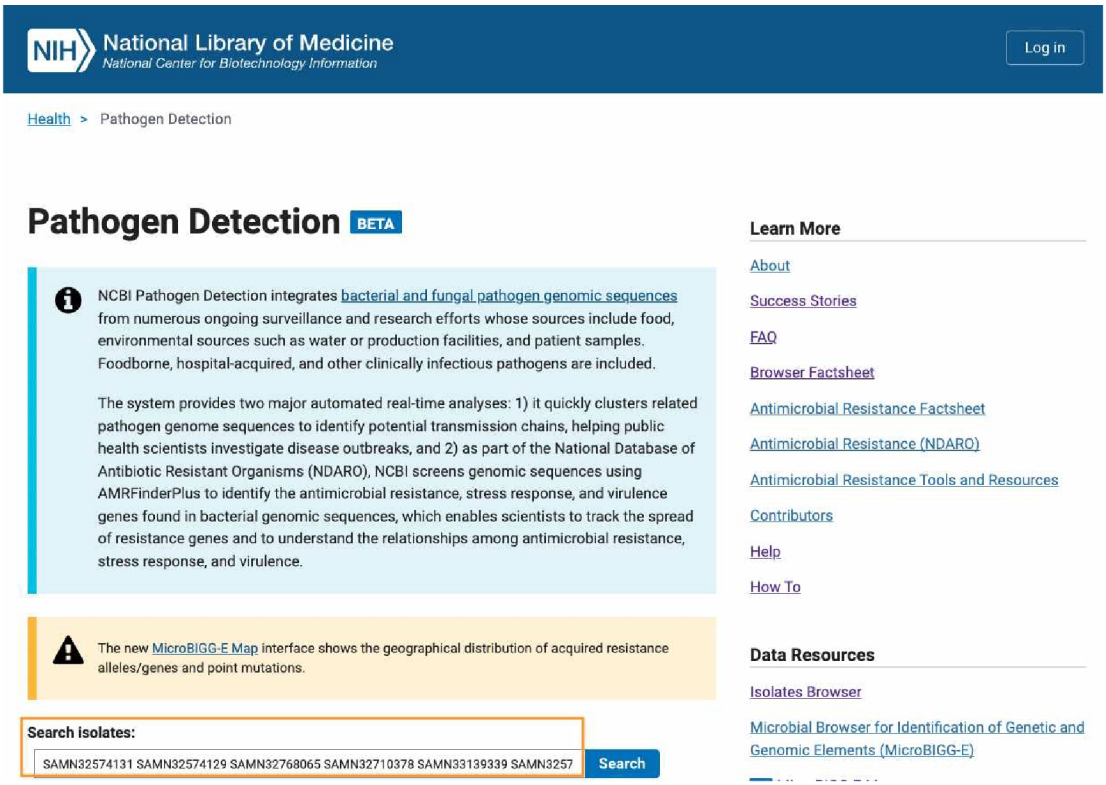
4 NCBI Pathogen Detection performs cluster analysis and genotyping screening for antimicrobial resistance, stress response and virulence genes. Visit NCBI-Pathogen Detection HowTo for extensive documentation on how to access analysis results.

Analysis results expected in <1 day for all organisms except *Salmonella,* which requires about two days for cluster results.



Strain identifiers and/or NCBI accessions can be used to query NCBI Pathogen Detection. Delimiters are **not** needed for querying this database.

Example query with BioSample accessions: SAMN33598462 SAMN36638873 SAMN06712285. Direct cut/paste from an excel table also works here.



For each query, users will see results summarized within two tables:

Table 1. Cluster-level results

Table 2. Isolate-level results