Uploading sequences to NCBI’s Sequence Read Archive (SRA)

Preparing sequence data and metadata

1. Add sequence data to BoxSync so it is accessible from computer without downloading files
2. Download IBM’s Aspera Connect

<https://www.ibm.com/aspera/connect/>

Make sure the plug-in/extension has access to read website information

Starting a new SRA project

1. In order to associate the project with an existing contact/group, you must be invited and begin the project within this framework
2. Fill out information, either by downloading an excel template, completing the guided sections, and reuploading, or using the built in table editor

SRA Metadata note:

Insert the exact name for each sequence you will be uploading under ‘filename’ (ex. Sample1\_072023\_001.fastq

For paired sequences or other multiple-part sequences, add each consecutive filename in the ‘filename 2’, ‘filename 3’, etc. columns, so each row contains all of the files pertinent for each sample (ex. Sample1\_\_A | Sample1\_B) (screenshot)

Sequence file upload

1. In the Files tab, select the Aspera Connect plug in option
2. Choose files > select a test sequence > upload

The Aspera Connect extension will open a new window to process the file upload

If a new window doesn’t open, the sequence won’t upload. Troubleshoot the Aspera application and make sure both Connect and the plug-in are installed and have access

Once Aspera is working properly, select all of your sequence files and upload them.

Microbiome metagenomic sequence data

In the Biosample section, select “Metagenome or environmental” under NCBO packages

Seq Center microbiome shotgun metagenomics

--Library strategy: WGS

--Library source: metagenomic

--Library\_selection: random

--Library layout: paired

Platform: illumine

Instrument model: nextseq2000