**Uploading BLiMMP metagenomes to NCBI**

All raw metagenomic data from this study will be made available through NCBI. This document outlines how I uploaded the data to NCBI.

Log in to NCBI using ORCID

**Step 1:** Make BioProject on NCBI

1. Go to the BioProjection section of Your Submissions: <https://submit.ncbi.nlm.nih.gov/subs/bioproject/>
2. Click new submission.
3. Submitting this with my UW-Madison info.
4. Project Type
   1. Project data type listed as “Metagenomes” and “Metagenomic assembly”.
   2. Sample scope – “Environment”
5. Target
   1. Environmental sample name: “Lake Mendota water column microbial community”
6. General Info
   1. Set release date to 2023-10-01
   2. Project title: “Microbial community mediating mercury methylation in Lake Mendota water column”
   3. Public description: “The metagenomic and metatranscriptomic data stored here was paired with extensive geochemical measurements and enriched stable isotope mercury tracer incubations to identify biogeochemical drivers of mercury methylation in the water column of Lake Mendota.”
   4. Relevance: “Environmental”
   5. Project not part of larger initiative.
7. Skip the BioSample submission for now.
8. No publication yet.

BioProject submission: SUB12012965

BioProject ID = PRJNA876614

**Step 2:** Make BioSamples. There is a single BioSample for each sampling event (which corresponds to a specific date and depth). The 2021 sampling with thus include duplicate metagenomes and triplicate metatranscriptomes. Metadata aggregation is here: code/metagenomes/metagenome\_metadata.R

1. Go to BioSample section of Your Submissions: <https://submit.ncbi.nlm.nih.gov/subs/biosample/>.
2. Click new submission.
3. Submitting this with my UW-Madison info.
4. General Info
   1. Set release date to 2023-10-01
   2. Upload as “Batch/Multiple BioSamples”
5. Sample Type
   1. Navigate to [Packages for metagenome submitters](https://submit.ncbi.nlm.nih.gov/biosample/template/). Select MIMS Environmental/Metagenome – Water.
6. Attributes
   1. Set to upload a file using Excel
   2. Excel sheet saved here: “dataEdited/metagenomes/NCBI\_info\_MIMS.xlsx”
   3. Script to general entries here: code/metagenomes/metagenome\_metadata.R. Details of setting up the columns is provided there as well.
   4. Save out data from R to temporary csv file, then copy the data over to dataEdited/metagenomes/NCBI\_info\_MIMS.xlsx. Manually replace the dates to use the appropriate date style.
   5. Upload the xlsx file to NCBI.
7. Download the attributes file that includes the BioSample IDs. Saved here: dataEdited/metagenomes/NCBI\_info\_MIMS\_with\_BioSample.txt

**Step 3:** Upload raw (untrimmed, unassembled) metagenomes to the Sequence Read Archive (SRA).

1. <https://submit.ncbi.nlm.nih.gov/subs/sra/>
2. We’ll upload using the “Aspera command line and FTP upload options”. To do this, we’ll first request a preload folder. It doesn’t show much when you click on this.
3. **Navigate to the source folder** where the files for submission are
   1. **Navigate to folder: `cd ~/BLiMMP/dataRaw/metagenomes`**
4. **Set up screen: `screen -S BLiMMP\_MG\_upload`**
5. **Establish an FTP connection** using the credentials below:
   1. Address: ftp ftp-private.ncbi.nlm.nih.gov
   2. Username: subftp
   3. Password: w4pYB9VQ
6. **Navigate to your account folder**: cd uploads/petersonben50\_gmail.com\_2kD7rlpc
7. **Create a subfolder (required!)** with a meaningful name: mkdir BLI\_data\_MG
8. **Navigate to the target folder** you just created: cd BLI\_data\_MG
9. **Copy your files into the target folder**:
   1. Prompt
   2. mput \*
10. After the FTP has finished, go back to <https://submit.ncbi.nlm.nih.gov/subs/sra/> and click on New Submission
11. Submitting this with my UW-Madison info.
12. General Info
    1. Provide the BioProject ID
    2. Did already register BioSample IDs
    3. Set release date to 2023-10-01
13. SRA Metadata:
    1. Batch upload. Excel sheet is here: dataEdited/metagenomes/NCBI\_SRA\_metadata\_acc.xlsx. Back to code/metagenomes/metagenome\_metadata.R to prep the data. Save out to “dataEdited/metagenomes/temp\_NCBI\_info\_SRA.csv”, then add data to xlsx file.
14. Files: “FTP or Aspera Command Line file preload”
    1. Select the BLI\_data\_MG folder.
15. Review and confirm.