**Uploading Hells Canyon 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 “Raw sequence reads”, “Metagenomes”, and “Metagenomic assembly”.
   2. Sample scope – “Environment”
5. Target
   1. Environmental sample name: “Brownlee Reservoir water column microbial community”
6. General Info
   1. Set release date to 2023-01-01
   2. Project title: “Mercury-methylating microbial communities under diverse redox conditions in stratified freshwater reservoir”
   3. Public description: “The metagenomic data in this project are from samples collected as part of a large project to understand methylmercury production in the Hells Canyon Complex on the Snake River. Extensive geochemical characterization of the system, including paired with these metagenomes, can be found in USGS data release.
   4. Relevance: “Environmental”
   5. Project not part of larger initiative.
   6. External links: Geochemical data release: <https://www.sciencebase.gov/catalog/item/5ee3a43b82ce3bd58d7e1c48>.
7. Skip the BioSample submission for now.
8. No publication yet.

BioProject submission: [SUB12048451](https://submit.ncbi.nlm.nih.gov/subs/bioproject/SUB12048451)

BioProject ID = PRJNA878929

**Step 2:** Make BioSamples. There is a single BioSample for each sampling event (which corresponds to a specific date, river mile, and depth). 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 2023-01-01
   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\_upload/NCBI\_info\_HCC\_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\_upload/NCBI\_info\_HCC\_MIMS.xlsx.
   5. Upload the xlsx file to NCBI.
7. Download the attributes file that includes the BioSample IDs. Saved here: dataEdited/metagenomes/NCBI\_upload/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 “FTP upload”. Instruction below are from the website:
   * 1. **Navigate to the source folder** where the files for submission are on GLBRC**: `cd ~/HellsCanyon/dataRaw/metagenomes`**
   1. **Set up screen: `screen -S HCC\_MG\_upload`**
   2. **Establish an FTP connection** using the credentials below:
      * + Address: ftp ftp-private.ncbi.nlm.nih.gov
        + Username: subftp
        + Password: w4pYB9VQ
   3. **Navigate to your account folder**: cd uploads/petersonben50\_gmail.com\_2kD7rlpc
   4. **Create a subfolder (required!)** with a meaningful name: mkdir HCC\_MG
   5. **Navigate to the target folder** you just created: cd HCC\_MG
   6. **Copy your files into the target folder**:
      * + prompt
        + mput KMBP00\*
3. After the FTP has finished, go back to <https://submit.ncbi.nlm.nih.gov/subs/sra/> and click on New Submission
4. Submitting this with my UW-Madison info.
5. General Info
   1. Provide the BioProject ID
   2. Did already register BioSample IDs
   3. Set release date to 2023-01-01
6. SRA Metadata:
   1. Batch upload. Excel sheet is here: dataEdited/metagenomes/NCBI\_upload/SRA\_metadata\_HCC\_acc.xlsx. Back to code/metagenomes/metagenome\_metadata.R to prep the data. Save out to “dataEdited/metagenomes/NCBI\_upload/temp\_NCBI\_info\_SRA.csv”, then add data to xlsx file.
7. Files: “FTP or Aspera Command Line file preload”
   1. Select the HCC\_data\_MG folder.
8. Review and confirm.