

Test_Website_Feburary_2025

 Uploading sediment 16s rRNA data to SRA

Uploading sediment 16s rRNA data to SRA

Uploading sequence data to Genbank or SRA is necessary for publication and important for long-term archival of scientific projects. Use this guiding document alongside instructions on the NCBI website (which will always be more up-to-date than this) to upload your sequencing data.

Use the [NCBI Submission Portal](#) to get started. This portal will direct you to the most appropriate submission tool for your data.

Depending on what type of samples you have, the details of each step may differ. However, the broad strokes of this protocol will remain the same.

Sequence Read Archive (SRA)

For uploading minimally-processed read data from high throughput platforms. This includes unassembled 16s rRNA the Bowen lab receives from the Tufts sequencing facility and Azenta (both use the Illumina MiSeq platform).

Detailed instructions can be found at <https://www.ncbi.nlm.nih.gov/sra/docs/submitportal/#6-sra-metadata>.

Information to collect before starting this process:

1. Project Information

- ☐ Project Title
- ☐ Project Description
- ☐ Grant Information
- ☐ External Information (If Applicable)

2. Sample Metadata (aka Biosample Type, select the [MIMARKS survey sediment package](#) when requested by SRA upload portal)

- ☐ Sample Names
- ☐ Organism of origin
- ☐ Collection Date
- ☐ Depth (below surface)

- ☐ Elevation (above sea level)
- ☐ Broad scale environment description
- ☐ Local scale environment description
- ☐ environmental medium
- ☐ geographical origin of sample
- ☐ latitude and longitude
- ☐ Additional information as collected (not required for upload)
- ☐ IMPORTANT: SRA requires each sample row to have at least one unique identifier - i.e., no two sample rows can share the exact same metadata. They do not count the sample_name, sample_title, or bioproject_accession columns as unique identifiers.
 - ☐ It's often helpful to create a column with a unique identifier for each sample (which can be a modification of the sample name).

3. SRA Metadata

- ☐ Library Strategy
 - ☐ AMPLICON
- ☐ Library Source
 - ☐ METAGENOMIC
- ☐ Library Selection
 - ☐ PCR
- ☐ Library Layout (single or paired)
 - ☐ Paired
- ☐ Platform
 - ☐ Illumina
- ☐ Instrument Model
 - ☐ Illumina Miseq

Notes:

SRA accepts sequencing files in the gzip or bzip2 format, but does not accept zip files. Do not upload zip files to SRA! It will not accept them!

For paired sequences, each sample will have two associated sequence files (one forward, on reverse). The SRA metadata sheet has multiple columns for associated filenames where you can associate both sequence filenames to your samples.

Metadata can be edited after submission here <https://dataview.ncbi.nlm.nih.gov/>). Instructions on how to edit your metadata can be found here <https://www.ncbi.nlm.nih.gov/sra/docs/submitupdate/#how-i-can-review-and-update-my-s>)