

Sequence Read Archive (SRA) Submission

The NCBI [Sequence Read Archive \(SRA\)](#) is the largest publicly available repository of high throughput sequencing data. Here, I present my personal experience of uploading raw sequences on this repository. For more details, I encourage you to consult the [submission portal guide](#), [quick start guide](#), and [file upload options](#).

Before starting the submission, you can prepare your personal information, general information about the project and fundings, and the metadata and sample attribute tables. SRA autosaves all the information you provide so you may leave your submission at any step and come back later to complete it.

The submission steps are as follows:

Page 1 - Submitter

★ First (given) name	Middle name	★ Last (family) name
<input type="text"/>	<input type="text"/>	<input type="text"/>
★ Email (primary)	Email (secondary)	
<input type="text"/>	<input type="text"/>	
<small>At least one email should be from the organization's domain.</small>		

Group for this submission


No group (affiliation from my personal profile)

[Create group](#) Allow selected collaborators to read, modify, submit and delete your submissions

★ Submitting organization	Submitting organization URL	★ Department
<input type="text"/>	<input type="text"/>	<input type="text"/>
Phone <small>?</small>	Fax <small>?</small>	
<input type="text"/>	<input type="text"/>	
★ Street	★ City	State/Province
<input type="text"/>	<input type="text"/>	<input type="text"/>
	★ Postal code	★ Country
	<input type="text"/>	<input type="text"/>

[Continue](#) ☒ Update my contact information in profile


Page 2 - General info**BioProject**

 BioProject describes the goal of your research effort.

★ Did you already register a BioProject for this research, e.g. for the submission of the reads to SRA and/or of the genome to GenBank?

☐ Yes ☒ No


BioSample

 The BioSample records the detailed biological and physical properties of the sample that was sequenced.
A BioSample can be used in more than one BioProject since it should be used for all the data that were obtained from that sample. Usually SRA data sets are generated from more than one sample.

★ Did you already register a BioSample for this sample, e.g. for the submission of the reads to SRA and/or of the genome to GenBank?


☐ Yes ☒ No

Release date


 **Note:** Release of BioProject or BioSample is also triggered by the release of linked data.

★ When should this submission be released to the public?

☐ Release immediately following processing
☒ Release on specified date or upon publication, whichever is first

★ **Projected release date** 

YYYY-MM-DD

 You will be able to change release date later (please refer to the [SRA Update Guide](#)).
The data must be released upon publication.

Page 3 - Project info

★ Project title ?

★ Public description ?

Relevance ?

★ Is your project part of a larger initiative which is already registered with NCBI?

☒ No

☐ Yes (not very common)

External Links

Description ?

URL ?

Delete

+ Add another link

Select your grants

?

Use this tool to look up grants from many subscribed governmental funding agencies (eg NIH, CDC, FDA and VA) and some non-governmental funding sources (eg HHMI and Autism Speaks). You can search by grant number, title or grantee name. If your grant is not included, you can select the "Add grants manually" option within this tool to add your grant.

+ Add grants

Add your grants manually

?

If your grant is funded by an HHS or VA agency, then use the "Select your grants" tool above to import your grant information into this submission. However, if your grant is not in that tool, then please enter the information manually here.

Grant ID ?	Grant title ?	Agency ?	Agency abbr. ?	Delete
<div></div>	<div></div>	<div></div>	<div></div>	<div></div>
<div></div>	<div></div>	<div></div>	<div></div>	<div></div>
<div></div>	<div></div>	<div></div>	<div></div>	<div></div>

+ Add another grant

Consortium name ?

Consortium URL ?

Data provider ?

Data provider URL ?

Delete

+ Add another data provider

Continue

*Page 4 - Biosample type***★ Select the package that best describes your samples:**

- ☐ **Pathogen affecting public health**
Use for pathogen samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of pathogens.
- ☐ **Microbe**
Use for bacteria or other unicellular microbes when it is not appropriate or advantageous to use MlXS, Pathogen or Virus packages.
- ☐ **Model organism or animal sample**
Use for multicellular samples or cell lines derived from common laboratory model organisms, e.g., mouse, rat, Drosophila, worm, fish, frog, or large mammals including zoo and farm animals.
- ☒ **Metagenome or environmental sample**
Use for metagenomic and environmental samples when it is not appropriate or advantageous to use MlXS packages.
- ☐ **Invertebrate**
Use for any invertebrate sample.
- ☐ **Human sample**
WARNING: Only use for human samples or cell lines that have no privacy concerns. For all studies involving human subjects, it is the submitter's responsibility to ensure that the information supplied protects participant privacy in accordance with all applicable laws, regulations and institutional policies. Make sure to remove any direct personal identifiers from your submission. If there are patient privacy concerns regarding making data fully public, please submit samples and data to NCBI's dbGaP database. [dbGaP](#) has controlled access mechanisms and is an appropriate resource for hosting sensitive patient data. For samples isolated from humans use the Pathogen, Microbe or appropriate MlXS package.
- ☐ **Plant sample**
Use for any plant sample or cell line.
- ☐ **Virus sample**
Use for all virus samples not directly associated with disease. Viral pathogens should be submitted using the Pathogen: Clinical or host-associated pathogen package.
- ☐ **Genome, metagenome or marker sequences (MlXS compliant)**
Use for genomes, metagenomes, and marker sequences. These samples include specific attributes that have been defined by the Genome Standards Consortium (GSC) to formally describe and standardize sample metadata for genomes, metagenomes, and marker sequences. The samples are validated for compliance based on the presence of the required core attributes as described in [MlXS](#).
- ☐ **Beta-lactamase**
Use for beta-lactamase gene transformants that have antibiotic resistance data.

Page 5 - Biosample attributes

The column titles in this section change depending on the sample type. The fields marked with * are required (copy-paste or upload the .tsv file). If you have more measured attributes other than what is in the table or columns with different titles, you may add those columns to the end of the table. Here, I present examples of two types of samples, metagenome or environmental sample and human sample.

1) Metagenome or environmental sample:

★ How do you want to provide your BioSample attributes?

☒ Use built-in table editor

☐ Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples

Right-click on the table if you want to edit rows, columns or find and replace text.

	★ Sample Name ?	Sample title ?	BioProject accession ?	★ Organism ?	★ host ?	★ isolation source ?	★ collection date ?	★ geographic location ?	★ latitude and longitude ?
1									
2									

★ How do you want to provide your BioSample attributes?

☒ Use built-in table editor

☐ Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples

Right-click on the table if you want to edit rows, columns or find and replace text.

	reference for biomaterial ?	relationship to oxygen ?	sample collection device or method ?	sample material processing ?	sample size ?	source material identifiers ?	Sample description
1							
2							

2) Human sample:

★ How do you want to provide your BioSample attributes?

☒ Use built-in table editor

☐ Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples

Right-click on the table if you want to edit rows, columns or find and replace text.

	★ Sample Name ?	Sample title ?	BioProject accession ?	★ Organism ?	★ isolate ?	★ age ?	★ biomaterial provider ?	★ sex ?	★ tissue ?	cell line ?	cell subtype	cell type ?
1												
2												

★ How do you want to provide your BioSample attributes?

☒ Use built-in table editor

☐ Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples

Right-click on the table if you want to edit rows, columns or find and replace text.

	culture collection ?	development stage ?	disease ?	disease stage ?	ethnicity ?	health state ?	karyotype	phenotype ?	population ?	race	sample type ?	treatment	Sample description
1													
2													

Note:

Before depositing human data into the public SRA database, make sure that you have consent from the donating individual to make this data available in an unprotected database. Do not transmit unconsented human data intended for dbGaP submissions to the public SRA database. Please refer to [dbGaP submission Guide](#).

Human metagenomic studies may contain human sequences and require that the donor provide consent to archive their data in an unprotected database. If you would like to archive human metagenomic sequences in the public SRA database, you may contact [NCBI](#) prior to submission.

Page 6 - SRA metadata

In this section, you provide the necessary information related to your metadata. The fields marked with * are required (copy-paste or upload the .tsv file).

For more detailed help with SRA submission please read the [SRA Submission Wizard Help](#).

★ How do you want to provide your metadata?

☒ Use built-in table editor

☐ Upload a file using Excel or text format (tab-delimited)

Right-click on the table if you want to edit rows, columns or find and replace text.

name	★ Library ID	★ Title	★ Library strategy	★ Library source	★ Library selection	★ Library layout	★ Platform	★ Instrument model	★ Design description
------	--------------	---------	--------------------	------------------	---------------------	------------------	------------	--------------------	----------------------

Page 7 - Files

There are different ways to upload raw read files. I chose the FTP upload using FileZilla (it's free) since it was the most straightforward. Files should be uploaded in one folder and can be compressed (gzip, bzip2 or in a tar archive), although it is not required.

Connect to NCBI through FileZilla:

Host: ftp-private.ncbi.nlm.nih.gov
 Username: use provided username
 Password: use provided password

In the "Remote site" bar, add this manually: uploads/user@email.com_XXXXX (use provided address)
 Make a new folder in the "uploads" folder, which will be your submission directory where you will transfer all the raw sequence files.

Drag the files from the local site to the new folder on the Remote site (this will take some time, depending on the size of your files and your internet upload speed).

Finally, it will take about 10 minutes for uploaded files to become available on SRA.

Each file must be listed in the [SRA metadata table you uploaded](#). If you are uploading a tar archive, list each file name, not the archive name.

Unique file names that do not contain any sensitive information should be used for all files. File names as submitted appear publicly when data is retrieved from the cloud.

Files can be compressed using gzip or bzip2, and may be submitted in a tar archive, but archiving or compressing your files is not required. **Do not use zip!**

★ How do you want to provide files for this submission?

☐ Web browser upload via HTTP or Aspera Connect plugin
 Do not use web browser HTTP upload if you are uploading files over 10 GB or more than 300 files.

☒ FTP or Aspera Command Line file preload
 All files for a submission must be uploaded into a single folder.

☐ Amazon S3 bucket

Select preload folder Selected folder: **FTP** aca16s_demultiplexedData_mp (628 files)

Aspera command line upload instructions +

FTP upload instructions +

☐ Autofinish submission

Load folder and go to the next (last) page to review your submission or simply auto finish it.