## Sequence Read Archive (SRA) Submission

The NCBI <u>Sequence Read Archive (SRA)</u> 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 <u>submission portal guide</u>, <u>quick start guide</u>, and <u>file upload options</u>.

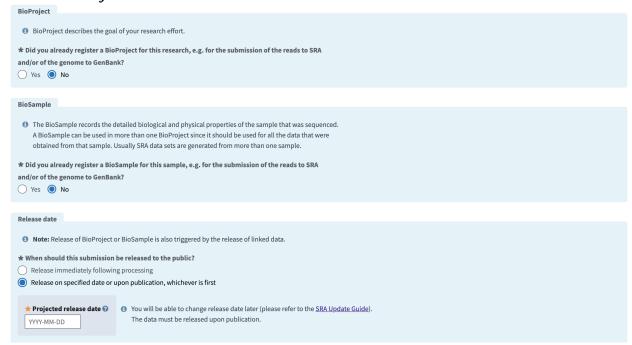
Before starting the submission, you can prepare your personal information, general information about the project and fundings, and the metadata and sample attribute tables. SRA autosaves all the information you provide so you may leave your submission at any step and come back later to complete it.

The submission steps are as follows:

Page 1 - Submitter

* First (given) name Middle name  * Email (primary)	* Last (family) name  Email (secondary)	4 At least one email should be from the organization's domain.
		Actes one characteristic organization 3 domain.
Group for this submission  No group (affiliation from my personal pro  Create group    Allow selected collar	file) borators to read, modify, submit and delete you	rsubmissions
* Submitting organization	Submitting organization URL	* Department
Phone @ Fax @		
* Street	* City State/Province	* Postal code * Country
Continue Update my contact i	nformation in profile	

## Page 2 - General info



# Page 3 - Project info

* Project title ②				٦		
* Public description ?						
Relevance ②						
★ Is your project part of a  No Yes (not very	larger initiative which is alre	eady registered with NCBI?				
External Links						
Description ② URL ②	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
						•
						•
						•
• Add another grant						
Consortium name 🕢		Consortium URL @				
Data provider 🕝		Data provider URL 🕝		Delete		
				•		
◆ Add another data provider						

Continue

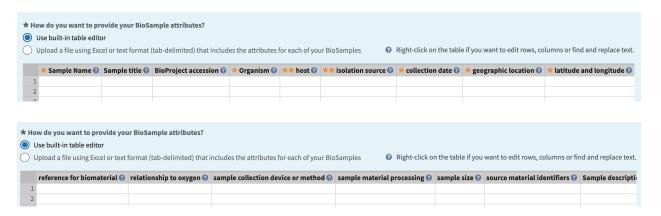
# Page 4 - Biosample type

<b>*</b> S	elect 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.
$\bigcirc$	Microbe
	Use for bacteria or other unicellular microbes when it is not appropriate or advantageous to use MIxS, Pathogen or Virus packages.
$\bigcirc$	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 MIxS packages.
$\bigcirc$	Invertebrate
	Use for any invertebrate sample.
$\bigcirc$	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 MIxS package.
$\bigcirc$	Plant sample
	Use for any plant sample or cell line.
$\bigcirc$	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.
$\bigcirc$	Genome, metagenome or marker sequences (MIxS 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 MIXS.
$\bigcirc$	Beta-lactamase
	Use for beta-lactamase gene transformants that have antibiotic resistance data.

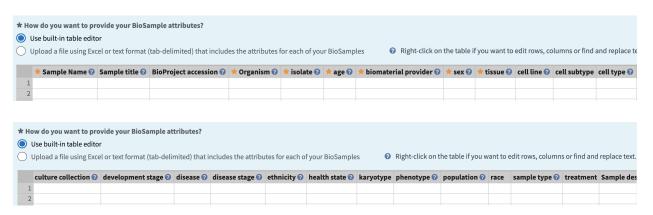
### Page 5 - Biosample attributes

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:



### 2) Human sample:



#### 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 <u>dbGaP submission Guide</u>.

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 <a href="NCBI">NCBI</a> 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).



#### 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.



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