

SARS-CoV-2 Sequence Submission to NCBI

Contents

Unassembled reads	1
Sequence Read Archive (SRA)	1
Assembled reads	1
GenBank via BankIt – Annotation not required	1

Unassembled reads

Sequence Read Archive (SRA)

In order to submit to SRA, a submission requires:

1. A BioProject, which is a general description of data you are submitting. The project allows you to group all of your samples and data under a single accession.
2. BioSample(s) which is just a digital description of your biological samples.
3. SRA metadata which describes how the sequencing was done (i.e. type of instrument, type of library, file names, etc).
4. Next-gen sequencing files which can be either compressed with gz, bz2 or uncompressed. We cannot accept ZIP files.

Submitting via SRA's Submission Portal wizard:

1. [Log in to the SRA Submission Portal Wizard](#).
2. Create new SRA submission (click on the button in the top left **New submission**).
3. Register your project (BioProject) (This only needs to be done a single time and you can re-use the accession PRJNA##### in subsequent submissions)
4. Register your biological samples (BioSamples), you will want to choose **"Pathogen affecting public health"** package and then chose **"Clinical or host-associated pathogen."** We suggest using the built-in table editor to fill out the metadata as its easy to copy and paste in the columns.
5. For organism, please enter "Severe acute respiratory syndrome coronavirus 2" as the name of the organism.
6. Submit **SRA metadata** - information that will link your project, samples and file names.
7. Upload **sequence data files**. Please refer to [SRA File Upload help](#).


Assembled reads

GenBank via BankIt – Annotation not required

In order to submit assembled data, a submission requires:

1. FASTA file of sequences with a unique sequence_identifier for each sequence
2. Tab-delimited file containing source metadata for each sequence, where the first column is the sequence_identifier used in the FASTA file. The metadata should contain:
 - a. isolate (preferably in the format SARS-CoV-2/isolate/host/year/location as outlined by ICTV)
 - b. complete collection date (including month and day if known) in the format dd-Mon-YYYY
 - c. country/geographic location: for example, USA: Maryland
 - d. host
 - e. isolation-source (for example, oral swab)

Submitting assembled sequences via BankIt:

1. [Log in to Bankit](#)
2. Select SARS-CoV-2 submission type
3. Create new submission (click on the button ).
4. Fill in web forms with prompted information. The tab-delimited source file will be uploaded on the Source Modifiers tab.
5. On the Features (overview) form, just click continue. You do not need to add features, we will do this for you using [VADR](#)
6. Bulk submissions of annotated sequences may be done using tbl2asn. If you would like further assistance with [tbl2asn](#), please let us know.