



UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY

100
1918 · 2018



forward together · saam vorentoe · masiye phambili

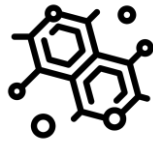
PUBLIC DATABASES FOR GENOMIC SURVEILLANCE (NCBI-SRA)

SAN EMMANUEL JAMES, DEREK TSHIABUILA, TULIO DE OLIVEIRA
ET. AL

- **Why publish data?**
- **What about raw reads?**
- **Interaction with NCBI-SRA (Upload/Interaction/Downloads and uploads)**
 - Intro
 - Resources (Literature, analysis tools)
 - KRISP on SRA
 - SRA downloads
 - SRA uploads
- **Summary**
- **Acknowledgements**



Why Publish Data?



- ☐ Fast track science



- ☐ Reproducible research

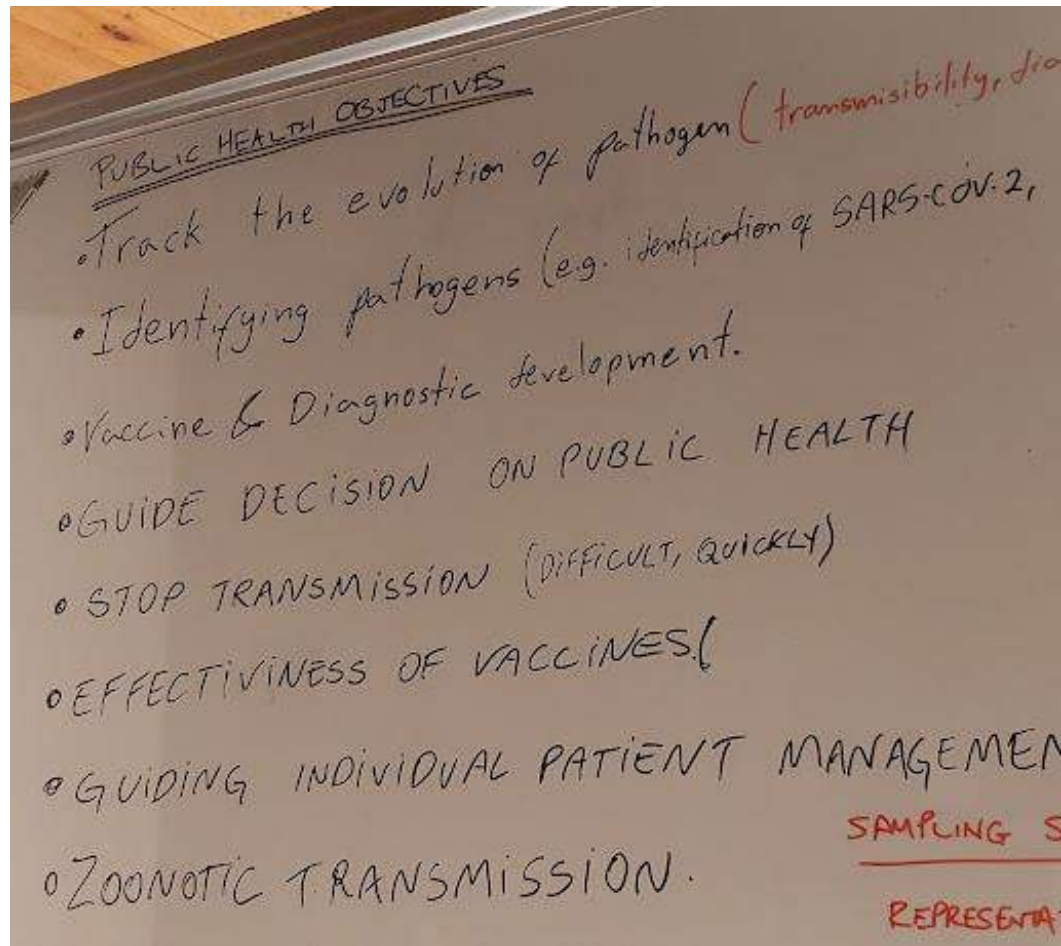


- ☐ Metadata QC and Standardization



- ☐ Permanent storage

- Evolutionary, transmission and spread analyses particularly require context
- Intra patient evolutionary analyses and co-infections



Rapid replacement of the Beta variant by the Delta variant in South Africa

Houriyyah Tegally, Euan Wilkinson, Christian L. Althaus, Marta Giovanetti, James Emmanuel San, Jennifer Giandhari, Sureshnee Pillay, Yeshnee Naidoo,

Download packages



[Back](#)

Doi: <https://doi.org/10.1101/2021.09.23.21264018>

Cell

Volume 184, Issue 20, 30 September 2021, Pages 5189-5200.e7



Article

The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages

Darren P. Martin¹, Steven Weaver², Houriiyah Tegally³, James Emmanuel San³, Stephen D. Shank², Euan

Method details

Global SARS-CoV-2 dataset preparation

- 1 We downloaded and curated GISAID sequence data, removed sequences that contained too many ambiguous or unresolved nucleotides, and identified all <https://doi.org/10.1016/j.cell.2021.09.003>



National Center for Biotechnology (NCBI)



ncbi.nlm.nih.gov/sars-cov-2/

NIH National Library of Medicine
National Center for Biotechnology Information

Search NCBI Search

COVID-19 Information
Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

NCBI SARS-CoV-2 Resources

Quick Navigation Guide

- Sequence Submission
- Literature
- Sequence-Related Resources
- Clinical Resources
- Other Websites

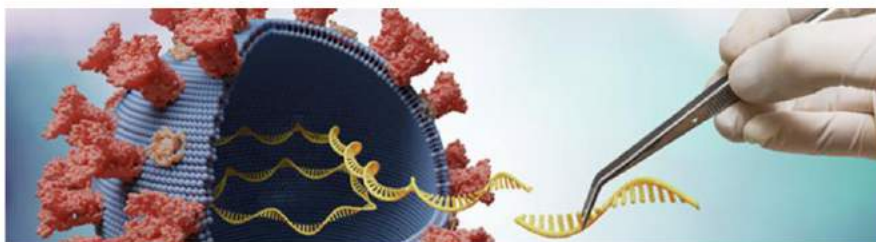
SARS-CoV-2 Data

1,365,799 <u>SRA runs</u>	2,111,990 <u>Nucleotide records</u>	3,215 <u>ClinicalTrials.gov</u>
192,749 <u>PubMed</u>	227,476 <u>PMC</u>	

EN



Submit SARS-CoV-2 Sequences



Add assembled & raw read data to the growing public archive

[Submit Now](#)

SARS-CoV-2 Literature

Articles referencing SARS-CoV-2 and COVID-19 in **PubMed**

[View in PubMed](#)

Free full-text content in **PubMed Central (PMC)**, including preprints, from the Public Health Emergency COVID-19 Initiative, suitable for text mining and secondary analysis

[Access PMC](#)

LitCovid: Comprehensive curated literature collection regarding the 2019 novel Coronavirus

[Access LitCovid](#)

Full-text COVID-19 guidelines and reviews in **Bookshelf**

[View in Bookshelf](#)

EN

[Feedback](#)



NCBI SARS-CoV-2 Sequence Analysis Tools



Click to go back, hold to see history

SARS-CoV-2 Sequence Resources

Genome Reference Sequence (NC_045512)

NCBI RefSeq SARS-CoV-2 genome annotation

Download Annotation

NCBI RefSeq SARS-CoV-2 genome sequence record

View Record

NCBI RefSeq SARS-CoV-2 genome graphical display

View Display

NCBI Gene SARS-CoV-2 curated gene records

View Records

Explore the Data

Search a **BLAST** database of Betacoronavirus nucleotide sequences

Run BLAST

Search, retrieve, and analyze sequences and other content in the **NCBI Virus SARS-CoV-2 Data Hub** Interactive Dashboard

Explore in NCBI Virus

↑
Back
To
Top

Feedback



[ncbi.nlm.nih.gov/sars-cov-2/](#)

Download viral genome and protein sequences, annotation, and a data report from **NCBI Datasets**

Explore in NCBI Datasets

Get the latest list of SARS-CoV-2 nucleotide sequences. You can query these IDs in **GenBank**

Download Accession List

SARS-CoV-2 **SRA** dataset on the Registry of Open Data on AWS (Amazon Web Services)

Explore in AWS

SARS-CoV-2 next-generation sequencing runs in **SRA**

View in SRA

Coronaviridae family-containing **SRA** runs

Download from SRA Run Selector

SARS-CoV-2 protein structures, domains, and sequences available through **NCBI Structure**

View Structure Data

SARS-CoV-2 related compounds, substances, pathways, bioassays, and more in **PubChem**

View in PubChem

Genome expression studies related to SARS-CoV-2 in **GEO**

View in GEO

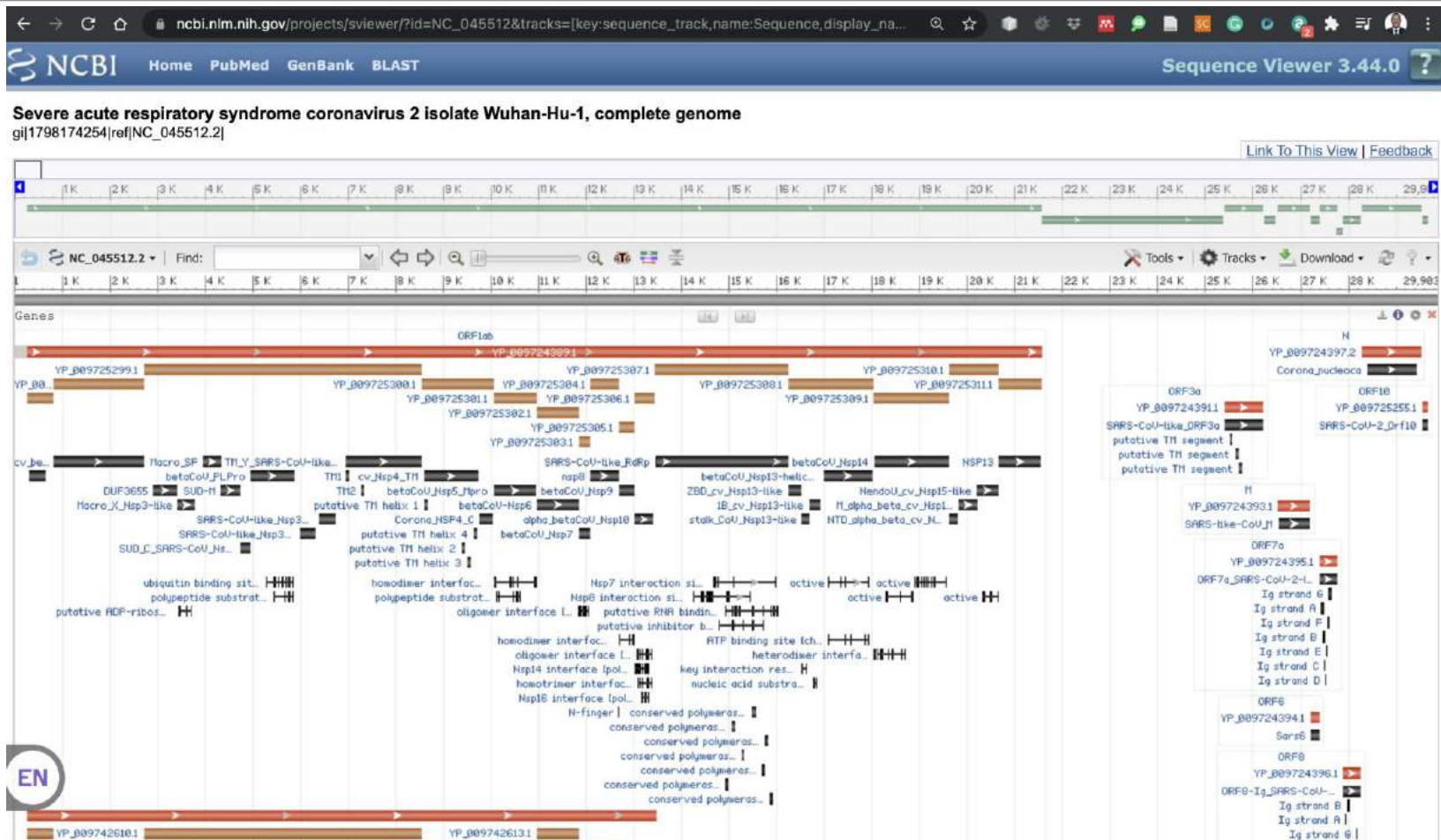
EN

Back To Top

Feedback



NCBI SARS-CoV-2 Sequence Viewer



NCBI RefSeq SARS-CoV-2 genome graphical display

View Display





NCBI Sequence Read Archive (SRA)



ncbi.nlm.nih.gov/sra/docs/srdownload/

NCBI Resources How To

houriyahategally My NCBI Sign Out

SRA SRA Advanced Search

Getting Started Submission Quick Start Search and Download SRA in the Cloud

COVID-19 Information

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

Download SRA sequences from Entrez search results

- [Obtain search results](#)
- [Obtain run accessions](#)
- [Download sequence data files using SRA Toolkit](#)
- [Download metadata associated with SRA data](#)
- [Download sequence data from the Run Browser](#)
- [Download SRA sequence data from the Cloud](#)
- [Contact SRA](#)

Obtain search results

Task: find RNA-Seq records for lymph node tissue in BALB/c mice in [SRA Entrez](#)

To learn how to use Advanced Search Builder please refer to [Search in SRA](#)

- In the Entrez search bar enter the query: `((("mus musculus"[Organism]) AND BALB/c*) AND "lymph") AND "rna seq"[Strategy]`.
- To limit your search to only aligned data add to the above query **AND aligned data**[Properties].
- Click the checkboxes next to records (experiments) to select data of interest. Leave all checkboxes unchecked to select all records (experiments) from your search.

Obtain run accessions

Run accessions are used to download SRA data. To download a list of Run accessions selected from your Entrez search ([Example](#)):

- Click **Send to** on the top of the page, check the radiobutton **File**, select **Accession List**.
Save this file to the location from which you are running the SRA Toolkit.

Getting Started

- [Getting Started](#)

Downloading SRA Data

- [Download Guide](#)
- [dbGaP Download Guide](#)
- [Data Delivery to Cloud](#)

Search SRA

[Go to:](#)

- [Search in SRA Entrez](#)
- [Cloud-based Search](#)
- [Search in BigQuery](#)
- [Search in Athena](#)

Browsing

[Go to:](#)

- [SRA Run Selector](#)
- [Samples](#)
- [Analyses](#)
- [Data for a single RUN](#)

SRA-BLAST

- [SRA BLAST How To](#)

<https://www.ncbi.nlm.nih.gov/sra/docs/srdownload/>





KRISP Data on NCBI SRA



ncbi.nlm.nih.gov/bioproject/PRJNA636748/

Display Settings: ▾ Send to: ▾

KRISP Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus sequencing in South-Africa Accession: PRJNA636748 ID: 636748

Sequencing of SARS-CoV-2 viral samples in South-Africa from COVID-19 positive patients during the 2020 pandemic by the KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP)

Accession	PRJNA636748
Data Type	Raw sequence reads
Scope	Multispecies
Submission	Registration date: 2-Jun-2020 KwaZulu-Natal Research Innovation and Sequencing Platform (UKZN)
Relevance	Medical

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
SRA Experiments	6196
OTHER DATASETS	
BioSample	6434

SRA Data Details

Parameter	Value
Data volume, Gbases	395
Data volume, Tbytes	0.23

NAVIGATE UP
This project is a component of the INSDC SARS-CoV-2 Viral Sequencing Data

NAVIGATE ACROSS
563 additional projects are components of the INSDC SARS-CoV-2 Viral Sequencing Data.

ncbi.nlm.nih.gov/sra/?term=PRJNA636748

NCBI Resources ▾ How To ▾

SRA SRA PRJNA636748 Create alert Advanced

COVID-19 Information
[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#)

Access Public (6,196) Summary ▾ 20 per page ▾ Send to: ▾

Source RNA (6,196)

Library Layout paired (5,367) single (629)

Platform Illumina (5,368) Oxford Nanopore (828)

Strategy other (6,196)

Data in Cloud GS (0,196) S3 (6,100)

File Type fastq (6,196)

[Clear all](#) [Show additional filters](#)

COVID-19 Information
[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#)

KRISP Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus sequencing in South-Africa - BioProject
Sequencing of SARS-CoV-2 viral samples in South-Africa from COVID-19 positive patients during the 2020 pandemic by the KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP)
Raw sequence reads project
Accession: PRJNA636748

View results as an expanded interactive table using the RunSelector. [Send results to Run selector](#)

Search results
Items: 1 to 20 of 6196

1. [=B9&" - SARS-CoV-2 sequencing in South Africa"](#)
1 ILLUMINA (Illumina MiSeq) run: 146,473 spots, 39.5M bases, 22.3Mb downloads
Accession: SRX12577546

2. [=B71&" - SARS-CoV-2 sequencing in South Africa"](#)
1 ILLUMINA (Illumina MiSeq) run: 272,183 spots, 85.7M bases, 51.2Mb downloads
Accession: SRX12577547

❑ Accession: **PRJNA636748** - KRISP Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus sequencing in South-Africa

❑ Reads for all sequencing runs deposited, whether they lead to a whole genome or not



Filters List

- 1 ☐ AvgSpotLen
- 2 ☐ Bases
- 3 ☐ BioSampleModel
- 4 ☐ Bytes
- 5 ☐ Center Name
- 6 ☐ geo_loc_name_country
- 7 ☐ geo_loc_name_country_continent
- 8 ☐ Instrument
- 9 ☐ isolation_source
- 10 ☐ LibraryLayout
- 11 ☐ Platform

Common Fields

BioProject	PRJNA636748
Consent	PUBLIC
Assay Type	AMPLICON
DATASTORE filetype	FASTQ, SRA
DATASTORE provider	GS, NCBI, S3
DATASTORE region	gs.US, ncbl.public, s3.us-east-1
Host_disease	COVID-19
Host	Homo sapiens
lat_lon	missing

Select

	Runs	Bytes	Bases	Download	Cloud Data Delivery	Computing
Total	6196	215.38 Gb	395.33 G	Metadata or Accession List		
Selected	0	0	0	Metadata or Accession List or JWT Cart	Deliver Data	Galaxy

Found 6,196 Items

Search within results

Q

Clear

<

1

1

124

>

<input checked="" type="checkbox"/>	<input type="checkbox"/>	Run	BioSample	AvgSpotLen	Bases	BioSampleModel	Bytes	Center Name
<input type="checkbox"/>	1	SRR11907531	SAMN15082663	418	315.51 M	Pathogen.cl	160.79 Mb	KWAZULU-NATAL RESEARCH INNOVATION AND SEQUENCING PLATFORM (UKZN)
<input type="checkbox"/>	2	SRR11907532	SAMN15082662	366	42.19 M	Pathogen.cl	24.06 Mb	KWAZULU-NATAL RESEARCH INNOVATION AND SEQUENCING PLATFORM (UKZN)
<input type="checkbox"/>	3	SRR11907533	SAMN15082661	335	488.50 M	Pathogen.cl	257.05 Mb	KWAZULU-NATAL RESEARCH INNOVATION AND SEQUENCING PLATFORM (UKZN)

ments/SraRunTable.txt ↕ master

Run	Assay Type	AvgSpotLen	Bases	BioProject	BioSample	BioSampleModel	Bytes	Center
SRR12338189	AMPLICON	343	1873803	PRJNA636748	SAMN15659322	Pathogen.cl	1050992	KW
SRR12338190	AMPLICON	325	71008503	PRJNA636748	SAMN15659321	Pathogen.cl	36557811	I
SRR12338191	AMPLICON	359	73174619	PRJNA636748	SAMN15659320	Pathogen.cl	37395039	I
SRR12338192	AMPLICON	400	44132730	PRJNA636748	SAMN15659319	Pathogen.cl	23842991	I
SRR12338193	AMPLICON	396	59871923	PRJNA636748	SAMN15659318	Pathogen.cl	31417528	I
SRR12338194	AMPLICON	302	67879749	PRJNA636748	SAMN15659317	Pathogen.cl	34469758	I
SRR12338195	AMPLICON	371	149794023	PRJNA636748	SAMN15659173	Pathogen.cl	83624938	I
SRR12338196	AMPLICON	358	66922830	PRJNA636748	SAMN15659316	Pathogen.cl	37600414	I
SRR12338197	AMPLICON	265	2997979	PRJNA636748	SAMN15659315	Pathogen.cl	1688852	KW
SRR12338198	AMPLICON	289	1331802	PRJNA636748	SAMN15659314	Pathogen.cl	780482	KWA
SRR12338199	AMPLICON	282	66913648	PRJNA636748	SAMN15659313	Pathogen.cl	36421098	I
SRR12338200	AMPLICON	296	5360287	PRJNA636748	SAMN15659312	Pathogen.cl	2920938	KW
SRR12338201	AMPLICON	359	73043188	PRJNA636748	SAMN15659311	Pathogen.cl	38810146	I
SRR12338202	AMPLICON	305	64118571	PRJNA636748	SAMN15659310	Pathogen.cl	33090971	I
SRR12338203	AMPLICON	338	58099965	PRJNA636748	SAMN15659309	Pathogen.cl	30350106	I
SRR12338204	AMPLICON	377	83476098	PRJNA636748	SAMN15659308	Pathogen.cl	42391344	I
SRR12338205	AMPLICON	221	66075846	PRJNA636748	SAMN15659307	Pathogen.cl	35549225	I
SRR12338206	AMPLICON	361	172113752	PRJNA636748	SAMN15659172	Pathogen.cl	97098129	I
SRR12338207	AMPLICON	326	2759556	PRJNA636748	SAMN15659306	Pathogen.cl	1595485	KW

SRR12338189
SRR12338190
SRR12338191
SRR12338192
SRR12338193
SRR12338194
SRR12338195
SRR12338196
SRR12338197
SRR12338198
SRR12338199
SRR12338200
SRR12338201
SRR12338202
SRR12338203
SRR12338204
SRR12338205
SRR12338206
SRR12338207



- ☐ Download and configure sratoolkit

```
$ vdb-config --interactive
```

- ☐ Get help

```
$ prefetch -h
```

- ☐ Download single file

```
$ prefetch -T fastq SRR15003883
```

- ☐ Batch download

```
$ prefetch -T fastq --option-file C.1.2_SRA_IDs.txt
```

C.1.2_SRA_IDS.txt

```
SRR15003883
SRR15003795
SRR15061003
SRR15060993
SRR15256763
SRR15256721
SRR15256662
SRR15256631
SRR15683627
SRR15683215
SRR15683166
SRR15683536
SRR15683419
SRR15683710
```

<https://github.com/ncbi/sra-tools/wiki/01.-Downloading-SRA-Toolkit>

<https://github.com/ncbi/sra-tools/wiki/03.-Quick-Toolkit-Configuration>

Submission of SARS-CoV-2 Raw Reads (FASTQ) to the Sequence Reads Archive (SRA)

SRA – Data Submission

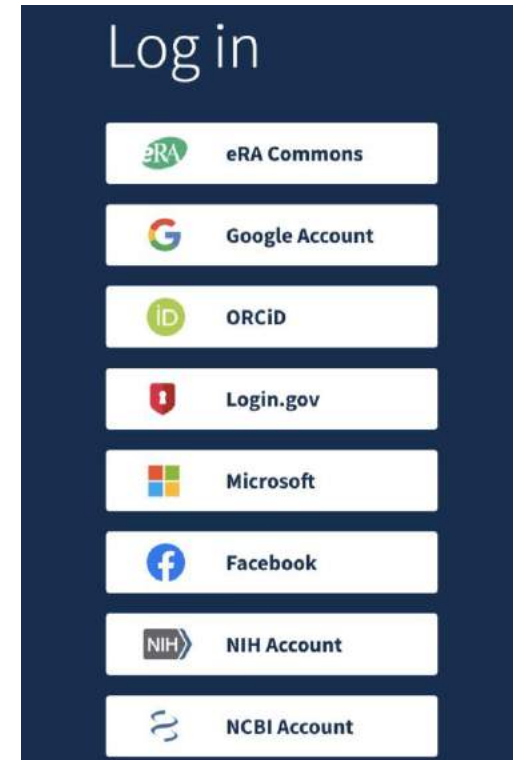
- There are 4 major steps to SRA submissions:
 1. Logging in to NCBI
 2. Creating a BioProject for the submissions
 3. Creating BioSamples for each of the samples being submitted
 4. Uploading sequencing reads with associated metadata
- Files can be compressed using gzip or bzip2, and may be submitted in a tar
- All file names must be unique and not contain any sensitive information
- Each file must be listed in the SRA metadata table

NCBI Submission Portal



Step I: Logging in to NCBI

- ❖ A login is required to submit to the SRA
- ❖ You can login by opening the website:
https://account.ncbi.nlm.nih.gov/?back_url=https%3A//submit.ncbi.nlm.nih.gov/subs/sra/
- ❖ Any of the methods alongside can be used to login





Step 2: Creating a BioSample

Submission Portal

1 **My submissions** Manage data Templates My profile

Your submissions

Filter / Search

Start a new submission

- GenBank 2
- **Sequence Read Archive**
- Genome
- TSA
- BioProject
- BioSample
- Supplementary Files
- API

From date YYYY-MM-DD To date YYYY-MM-DD Status Not deleted Sort by ☐ desc

Apps + Data archives +

Query Search Clear



Submission Portal

3 Home **My submissions** Manage data Templates My profile

Sequence Read Archive (SRA) **New submission**



Sequence Read Archive (SRA) submission: SUB10598803

New

1 SUBMITTER 2 GENERAL INFO 3 SRA METADATA 4 FILES 5 REVIEW & SUBMIT

Submitter

★ First (given) name	Middle name	★ Last (family) name
<input type="text" value="KRISP"/>	<input type="text"/>	<input type="text" value="UKZN"/>
★ Email (primary)	Email (secondary)	
<input type="text" value="krispilluminabasespace@gmail.com"/>	<input type="text"/>	

At least one email should be from the organization's domain.

Group for this submission

☐ No group (affiliation from my personal profile)

☒ 4 members: Houriyah Tegally's shared submissions
Houriyah Tegally, sanemmanueljames@gmail.com, Derek Tshiabulila, you

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

★ Submitting organization	Submitting organization URL	★ Department
<input type="text" value="KwaZulu-Natal Research Innovation and"/>	<input type="text"/>	<input type="text" value="College of Health Sciences, UKZN"/>
Phone ⓘ	Fax ⓘ	
<input type="text"/>	<input type="text"/>	
★ Street	★ City	State/Province
<input type="text" value="K-RITH Tower Building, Nelson R Mandela"/>	<input type="text" value="Durban"/>	<input type="text" value="KZN"/>
	★ Postal code	★ Country
	<input type="text" value="0000"/>	<input type="text" value="South Africa"/>

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



Sequence Read Archive (SRA) submission: SUB10598803

New

1 SUBMITTER 2 GENERAL INFO 3 SRA METADATA 4 FILES 5 REVIEW & SUBMIT

General Information


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

Project Info

★ Project title ?

KRISP SARS-CoV-2 Sequencing in South Africa

★ Public description ?

Kwazulu-Natal Research Innovation and Sequencing Platform sequencing
Severe Acute Respiratory-2 C

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](#)

Step 3: Registering a BioSample

1 SUBMITTER 2 GENERAL INFO 3 PROJECT INFO 4 BIOSAMPLE TYPE 5 BIOSAMPLE ATTRIBUTES 6 SRA METADATA 7 FILES 8 REVIEW & SUBMIT

Sample Type

★ Select the package that best describes your samples.

All packages Packages for MAG submitters Packages for metagenome submitters

(Optional) Filter packages by organism name

Enter the full scientific name of your samples, e.g., *Escherichia coli*

Reset and show all packages

- To filter for relevant BioSample packages, enter the **full scientific name** of the organism of your samples.
 - If your BioSamples are derived from a species **not represented in NCBI's Taxonomy database**, enter the genus-level name, e.g., *Escherichia*
 - If your BioSamples are derived from **more than one organism**, enter the common species, genus, or family, e.g., *Enterobacteriaceae*
 - If your BioSamples are **metagenomic/environmental**, or **metagenome-assembled genomes (MAG)**, select the appropriate tab above
 - For more information about organism names, see [Organism information](#).

NCBI packages [More...](#)

- ☒ **SARS-CoV-2: clinical or host-associated**
Use for SARS-CoV-2 samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of SARS-CoV-2 cases.
- ☐ **SARS-CoV-2: wastewater surveillance**
Use for SARS-CoV-2 wastewater surveillance samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of SARS-CoV-2 cases.
- ☐ **Pathogen**
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.

GSC [MixS](#) packages for genomes, metagenomes, and marker sequences [More...](#)

- ☐ **MIGS Cultured Bacterial/Archaeal**
Use for cultured bacterial or archaeal genomic sequences. Organism must have lineage [Bacteria](#) or [Archaea](#).
- ☐ **MIGS Eukaryotic**
Use for eukaryotic genomic sequences. Organism must have lineage [Eukaryota](#).
- ☐ **MIGS Viral**
Use for virus genomic sequences. Organism must have lineage [Viruses](#).
- ☐ **MIMAG Metagenome-assembled Genome**
Use for metagenome-assembled genome sequences produced using computational binning tools that group sequences into individual organism genome assemblies starting from



1 SUBMITTER 2 GENERAL INFO 3 PROJECT INFO 4 BIOSAMPLE TYPE 5 BIOSAMPLE ATTRIBUTES

Attributes

Package SARS-CoV-2: clinical or host-associated; version 1.0

★ 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

	★ Sample Name ?	Sample title ?	BioProject accession ?	★ Organism ?
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

★ 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

★ Attributes file

[Choose file](#) or drag and drop it here

- [Template for BioSample package SARS-CoV-2: clinical or host-associated; version 1.0](#)
[Download Excel](#) [Download TSV](#)
 For column explanations and examples, please see the [sample attributes page](#).
 For more information, please see [creating sample attribute file](#).



BioSample Attributes


[illegible]



Step 4: Data Upload












































1 SUBMITTER 2 GENERAL INFO 3 PROJECT INFO 4 BIOSAMPLE TYPE 5 BIOSAMPLE ATTRIBUTES 6 SRA METADATA

SRA Metadata

 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)


	★ Sample name	★ Library ID 	★ Title 	★ Library strategy	★ Library source
1	<input type="text"/>				
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					

★ How do you want to provide your metadata?

- ☐ Use built-in table editor
☒ Upload a file using Excel or text format (tab-delimited)

★ Metadata file

 [Choose file](#) or drag and drop it here

 [Download Excel spreadsheet](#) (designed to make it easier to select the correct metadata values), edit, save and then upload the modified Excel file.

	A	B	C	D	E	F	G	H	I	J	K	L	M
	sample_name	library_ID	library_strategy	library_source	library_selection	library_layout	platform	instrument_model	design_description	filetype	filename	filename2	
2	K010943	K010943	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010943_S1_L001_R1_001.fastq.gz	K010943_S1_L001_R2_001.fastq.gz	
3	K010944	K010944	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010944_S2_L001_R1_001.fastq.gz	K010944_S2_L001_R2_001.fastq.gz	
4	K010945	K010945	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010945_S3_L001_R1_001.fastq.gz	K010945_S3_L001_R2_001.fastq.gz	
5	K010946	K010946	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010946_S4_L001_R1_001.fastq.gz	K010946_S4_L001_R2_001.fastq.gz	
6	K010947	K010947	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010947_S5_L001_R1_001.fastq.gz	K010947_S5_L001_R2_001.fastq.gz	
7	K010948	K010948	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010948_S6_L001_R1_001.fastq.gz	K010948_S6_L001_R2_001.fastq.gz	
8	K010949	K010949	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010949_S7_L001_R1_001.fastq.gz	K010949_S7_L001_R2_001.fastq.gz	
9	K010950	K010950	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010950_S8_L001_R1_001.fastq.gz	K010950_S8_L001_R2_001.fastq.gz	
10	K010951	K010951	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010951_S9_L001_R1_001.fastq.gz	K010951_S9_L001_R2_001.fastq.gz	
11	K010952	K010952	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010952_S10_L001_R1_001.fastq.gz	K010952_S10_L001_R2_001.fastq.gz	
12	K010954	K010954	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010954_S11_L001_R1_001.fastq.gz	K010954_S11_L001_R2_001.fastq.gz	
13	K010955	K010955	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010955_S12_L001_R1_001.fastq.gz	K010955_S12_L001_R2_001.fastq.gz	
14	K010956	K010956	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010956_S13_L001_R1_001.fastq.gz	K010956_S13_L001_R2_001.fastq.gz	
15	K010957	K010957	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010957_S14_L001_R1_001.fastq.gz	K010957_S14_L001_R2_001.fastq.gz	
16	K010958	K010958	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010958_S15_L001_R1_001.fastq.gz	K010958_S15_L001_R2_001.fastq.gz	
17	K010959	K010959	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010959_S16_L001_R1_001.fastq.gz	K010959_S16_L001_R2_001.fastq.gz	
18	K010960	K010960	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010960_S17_L001_R1_001.fastq.gz	K010960_S17_L001_R2_001.fastq.gz	
19	K010961	K010961	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010961_S18_L001_R1_001.fastq.gz	K010961_S18_L001_R2_001.fastq.gz	
20	K010962	K010962	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010962_S19_L001_R1_001.fastq.gz	K010962_S19_L001_R2_001.fastq.gz	
21	K010963	K010963	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010963_S20_L001_R1_001.fastq.gz	K010963_S20_L001_R2_001.fastq.gz	
22	K010964	K010964	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010964_S21_L001_R1_001.fastq.gz	K010964_S21_L001_R2_001.fastq.gz	
23	K010965	K010965	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010965_S22_L001_R1_001.fastq.gz	K010965_S22_L001_R2_001.fastq.gz	
24	K010966	K010966	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010966_S23_L001_R1_001.fastq.gz	K010966_S23_L001_R2_001.fastq.gz	
25	K010967	K010967	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010967_S24_L001_R1_001.fastq.gz	K010967_S24_L001_R2_001.fastq.gz	
26	K010968	K010968	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010968_S25_L001_R1_001.fastq.gz	K010968_S25_L001_R2_001.fastq.gz	
27	K010969	K010969	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010969_S26_L001_R1_001.fastq.gz	K010969_S26_L001_R2_001.fastq.gz	
28	K010970	K010970	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010970_S27_L001_R1_001.fastq.gz	K010970_S27_L001_R2_001.fastq.gz	
29	K010971	K010971	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010971_S28_L001_R1_001.fastq.gz	K010971_S28_L001_R2_001.fastq.gz	
30	K010972	K010972	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010972_S29_L001_R1_001.fastq.gz	K010972_S29_L001_R2_001.fastq.gz	
31	K010973	K010973	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010973_S30_L001_R1_001.fastq.gz	K010973_S30_L001_R2_001.fastq.gz	
32	K010974	K010974	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010974_S31_L001_R1_001.fastq.gz	K010974_S31_L001_R2_001.fastq.gz	
33	K010975	K010975	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010975_S32_L001_R1_001.fastq.gz	K010975_S32_L001_R2_001.fastq.gz	
34	K010976	K010976	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010976_S33_L001_R1_001.fastq.gz	K010976_S33_L001_R2_001.fastq.gz	
35	K010977	K010977	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010977_S34_L001_R1_001.fastq.gz	K010977_S34_L001_R2_001.fastq.gz	
36	K010978	K010978	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010978_S35_L001_R1_001.fastq.gz	K010978_S35_L001_R2_001.fastq.gz	
37	K010979	K010979	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010979_S36_L001_R1_001.fastq.gz	K010979_S36_L001_R2_001.fastq.gz	
38	K010980	K010980	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010980_S37_L001_R1_001.fastq.gz	K010980_S37_L001_R2_001.fastq.gz	
39	K010981	K010981	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010981_S38_L001_R1_001.fastq.gz	K010981_S38_L001_R2_001.fastq.gz	
40	K010982	K010982	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010982_S39_L001_R1_001.fastq.gz	K010982_S39_L001_R2_001.fastq.gz	
41	K010983	K010983	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010983_S40_L001_R1_001.fastq.gz	K010983_S40_L001_R2_001.fastq.gz	
42	K010984	K010984	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010984_S41_L001_R1_001.fastq.gz	K010984_S41_L001_R2_001.fastq.gz	
43	K010985	K010985	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010985_S42_L001_R1_001.fastq.gz	K010985_S42_L001_R2_001.fastq.gz	
44	K010986	K010986	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010986_S43_L001_R1_001.fastq.gz	K010986_S43_L001_R2_001.fastq.gz	
45	K010987	K010987	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010987_S44_L001_R1_001.fastq.gz	K010987_S44_L001_R2_001.fastq.gz	
46	K010988	K010988	Amplicon	Viral RNA	RT-PCR	paired	ILLUMINA	ILLUMINA MiSeq	Amplicons were generated using	fastq	K010988_S45_L001_R1_001.fastq.gz	K010988_S45_L001_R2_001.fastq.gz	

Data Upload Options

1 SUBMITTER 2 GENERAL INFO 3 PROJECT INFO 4 BIOSAMPLE TYPE 5 BIOSAMPLE ATTRIBUTES 6 SRA METADATA 7 FILES

Files

- 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.
- ☐ AWS or GCP bucket

☐ Autofinish submission ?

Continue





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

☐ AWS or GCP bucket

Select preload folder Preload folder not selected

Aspera command line upload instructions

You may use the following command to upload files via **Aspera Command-Line**:

```
ascp -i <path/to/key_file> -QT -l100m -k1 -d <path/to/folder/containing_files>
subasp@upload.ncbi.nlm.nih.gov:uploads/krispilluminabasespace_gmail.com_zYSMLYcy
```

Where:

<path/to/key_file> must be an absolute path, e.g.: /home/keys/aspera.openssh

<path/to/folder/containing_files> needs to specify the local folder that contains all of the files to upload.

Get the [key file](#).

If you upload your files in your root directory, you will not be able to see them or to select the folder during the submission.

Make a new subdirectory for each new submission. Your submission subfolder is a temporary holding area and it will be removed once the whole submission is complete.

Do not upload complex directory structures or files that do not contain sequence data.

Return back to this page and select preload folder, then continue submission.

Please note: it takes at least 10 minutes for uploaded files to become available for selection.

```
/usr/local/biotools/SRA/aspera_connect/.aspera/connect/bin/ascp -i /usr/local/biotools/SRA/aspera_connect/aspera.openssh -QT -l
100m -k1 -v -d "/analyses2/houriiyah/nCoV_genomes/sra_2"
subasp@upload.ncbi.nlm.nih.gov:uploads/krispilluminabasespace_gmail.com_zYSMLYcy
```


Select preload folder Preload folder not selected

Aspera command line upload instructions +

FTP upload instructions -

1. **Navigate to the source folder** where the files for submission are;

2. **Establish an FTP connection** using the credentials below:

Address: `ftp-private.ncbi.nlm.nih.gov`

Username: `subftp`

Password: `w4pYB9vQ`

3. **Navigate to your account folder:**

From the command line use the 'cd' command:

`cd uploads/krispilluminabasespace_gmail.com_3X21iMw0`

When using a GUI FTP client (eg: Filezilla, NcFTP, Cyberduck, etc.), after you've connected to the FTP server, paste your account folder (`uploads/krispilluminabasespace_gmail.com_3X21iMw0`) into the "Remote Site" or "Remote Directory" box on the interface and press "Enter".

Then you will be able to create the data sub-directory for your submission. Until you do this, you will see a message stating "550 /: Permission denied" or "Failed to read the directory listing". We prevent directory listing in the default sign in folder for security reasons.

4. **Create a subfolder (required!)** with a meaningful name:

`mkdir new_folder`

5. **Navigate to the target folder** you just created:

`cd new_folder`

6. **Copy your files into the target folder:**

`put file_name`

If you upload your files in your root directory, you will not be able to see them or to select the folder during the submission.

Make a new subdirectory for each new submission. Your submission subfolder is a temporary holding area and it will be removed once the whole submission is complete.

Do not upload complex directory structures or files that do not contain sequence data.

7. **Return back to this page and select preload folder**, then continue submission.

Please note: it takes at least 10 minutes for uploaded files to become available for selection.

104 submissions					
Submission ↕	Title ↕	App ↕	Group	Status ↕	Updated ↕
SUB10550543	KRISP Sequencing of SARS-CoV-2 in Cameroon, Oct 22 '21	Sequence Read Archive (SRA)	Houriiyah Tegally's shared submissions	<div> <div>✓</div> <div> BioProject: Processed PRJNA773611 : KRISP Sequencing of SARS-CoV-2 in Cameroon </div> </div> <div> <div>Manage data</div> </div> <div> <div>✓</div> <div> BioSample: Processed Successfully loaded (92 objects) Download attributes file with BioSample accessions </div> </div> <div> <div>✓</div> <div> SRA: Processed (92 objects) Download metadata file with SRA accessions View and manage my SRA submission data </div> </div>	Oct 23
SUB10550577	KRISP Sequencing of SARS-CoV-2 in Mozambique, Oct 22 '21	Sequence Read Archive (SRA)	Houriiyah Tegally's shared submissions	<div> <div>✓</div> <div> BioSample: Processed Successfully loaded SAMN22502415 : K025848 (TaxID: 2697049) Download attributes file with BioSample accessions </div> </div> <div> <div>Manage data</div> </div> <div> <div>✓</div> <div> SRA: Processed SRR16548011 Download metadata file with SRA accessions View and manage my SRA submission data </div> </div>	Oct 23
SUB10539215	KRISP Sequencing of SARS-CoV-2 in Mozambique, Oct 20 '21	Sequence Read Archive (SRA)	Houriiyah Tegally's shared submissions	<div> <div>✓</div> <div> BioSample: Processed Successfully loaded (96 objects) Download attributes file with BioSample accessions </div> </div> <div> <div>Manage data</div> </div> <div> <div>✓</div> <div> SRA: Processed (96 objects) Download metadata file with SRA accessions View and manage my SRA submission data </div> </div>	Oct 20

“As a scientific community, we have responsibility to produce and share high quality data. NCBI and the sister databases provide excellent option for permanent storage and several tools to submit, retrieve, analyse, visualize and download datasets and analysis results”.



Acknowledgements

