

“SAR-COV-2 WGS analysis module”

User guide to execute SAR-COV-2 whole genome variant analysis

Introduction

IBDC offers end-to-end analysis services for analysing covid-19 sequence samples. It provides an easy to run analysis pipeline for the users having no prior experience of programming/command-line data analysis tools. Users can submit their sequence datasets and IBDC servers will execute the analysis 'on-the-fly' and provide results.

Steps for initiating WGS SAR-COV-2 analysis job at IBDC

- 1) Visit IBDC website and move to 'Covid Portal'
- 2) Sign up in the 'Covid Portal'. *(If you are already a registered user, directly go to step 4)*
- 3) The user will receive a 'Welcome email' from IBDC along with *two* signin credentials, each for login to the '**Analysis portal**' and the '**ftp server**', respectively.
- 4) The users first needs to upload their data sets at the IBDC ftp server (ip address: 117.211.200.18) using the ftp login credentials (provided in the email). The data should be uploaded in the folder entitled '**sample_data**'. The ftp access credentials can also be viewed in '**User Detail**' section on the user's dashboard. User can use any third party FTP client to upload the data.
- 5) Login to the IBDC 'Covid Portal' web portal (using 'Analysis portal' login credentials) and visit the '**Analyse Data**' section on the user 'Dashboard'.
- 6) The user can view two directories. '**sample_data/**' and '**result/**'.
- 7) The uploaded data files (step 4) can be viewed inside '**sample_data**' directory .
- 8) Select the sample type from the dropdown menu. Following type of datasets are valid.

IBDC SAR-COV-2 WGS analysis module

1. Illumina PE reads (fastq.gz)
 2. Illumina SE reads (fastq.gz)
 3. Assembled SAR COV-2 sequences (fasta).
- 9) Verify the data files and then click on '**Click To Analyse**' option. This will initiate the analysis at the IBDC server. It may take sometime depending on the file size and workload. You will receive an email when the analysis is done. **Note:** *A user can only trigger one analysis job at a time with as many number of samples at one go. The '**Run WGS covid analysis**' tab will be disabled once one job is running for each user. Once the previously running job is completed, the **Run WGS covid analysis** will be enabled for submission of another new set of analysis job.*

Accessing the results

- 1) The analysis results will be made available as follows:
 - a) results will be stored in the '**result**' directory available at ftp location.
 - b) The analysis results will also be shared via registered email id.
 - c) so be accessed from the '**Analysis Reports**' section on the on the user dashboard.
- 2) All the results will be available for download for a peroid of 15 days and will be deleted thereafter.
- 3) The result file naming format will be as : *UserID_JobID_Results.tar.gz*

Initiate another set of analysis job

- 1) To initiate another set of analysis job, the user should upload new fastq data files [follow [filenaming convention section](#)] to his/her '**sample_data**' folder available at his/her ftp location.
- 2) The user should remove earlier files (already analyzed) from the ftp folder before running new analysis.

File naming convention

The user must follow the file naming convention to proceed with the smooth execution of the analysis.

1. Illumina PE reads (fastq format)

Sample name should be named as follows:

Sample1_R1.fastq.gz

Sample1_R2.fastq.gz

Sample2_R1.fastq.gz

Sample2_R2.fastq.gz

2. Illumina SE reads (fastq format)

Sample name should be named as follows:

Sample1.fastq.gz

Sample2.fastq.gz

3. Assembled SAR COV-2 sequences (fasta format).

Kindly upload one fasta sequence for one sample. Kindly do not uploaded a multifasta sequences from several samples as one file. The sample name should be named as follows:

Sample1.fasta

Sample2.fasta

IBDC SAR-COV-2 WGS analysis module

Further queries:

The user can communicate to us at support@ibdc.rcb.res.in. Kindly include the following 'subject' line in the email: 'Covid analysis:JobID'. The 'JobID' is generated for each analysis job.
