

# DNAnexus viral-ngs pipeline tutorial for SARS-CoV-2 genome assembly

Version 1.0 30-06-2021

Idowu Olawoye; [olawoyei0303@run.edu.ng](mailto:olawoyei0303@run.edu.ng)

Paul E. Oluniyi; [oluniyip@run.edu.ng](mailto:oluniyip@run.edu.ng)

Kemi Adedotun-Sulaiman; [kemi.sulaiman555@gmail.com](mailto:kemi.sulaiman555@gmail.com)

ACEGID, Redeemer's University, Ede, Osun State, Nigeria

# 1. Creating Project

[+ New Project](#)

To create a new project, click on New Project on the right top corner of the screen and fill in your project name followed by setting up your billing account.

## 1.1. Importing viral-ngs pipeline


To import viral-ngs pipeline from the global DNAnexus platform into your new project. Search for “viral-ngs” in the search bar at the top of the webpage. Then select “**Broad viral-ngs dxWDL CI - Public**” project to open it.

Click on “**build**”

Click on “**quay.io**”




Click on “**broadinstitute**”

Click on “**viral-pipelines**”

Select the appropriate pipeline folder by clicking on the box next to the version number such as 2.1.19.0 ☐  2.1.19.0-rc111 Then click on the Copy button on the top right corner of the

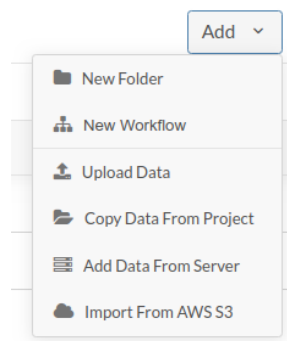
webpage. This will display your available Projects where you can select the Project destination for the viral-ngs pipeline.

In your Project, create a New Folder by clicking on New Folder on the bottom of the dialog

screen    and name it “pipelines”. Then you can copy the viral-ngs pipeline into the “pipelines” folder in your project.

## 2. FASTQ to Unaligned BAM conversion

To convert your raw FASTQ files from your Sequencing machines to unaligned BAM files. Upload the FASTQ files (paired-end or single read files) to a designated folder within your



project by clicking on the Add button and selecting Upload Data on the Top right corner of the screen.

This will take you to a screen where you can select files to upload from your computer.

## 2.1. Running FASTQtoUBAM

Open the pipelines folder and select the appropriate viral-ngs pipeline version. **FastqToUBAM** can be found in the **assemble\_refbased** folder under **pipelines**. Under the ANALYSIS SETTINGS tab, select the output folder for this workflow. Click on ANALYSIS INPUTS and select the paired-end FASTQ reads (fastq\_1 and fastq\_2). Type the **library and sample name** (which should be the same). Then click **Start Analysis**.

The screenshot shows the 'FastqToUBAM' analysis settings interface. At the top, there are three tabs: 'ANALYSIS SETTINGS', 'ANALYSIS INPUTS' (which is active), and 'STAGE SETTINGS'. Below the tabs, there is a header bar with the app name 'FastqToUBAM' and a toggle for 'Enable Batch' set to 'OFF'. The main content area is divided into two sections. The top section, 'ANALYSIS INPUTS', contains two input fields: 'fastq\_1' with a file path 'CV1372\_S6\_L001\_R1\_001.fastq.gz' and 'fastq\_2' with a file path 'CV1372\_S6\_L001\_R2\_001.fastq.gz'. The bottom section, 'COMMON', contains several input fields: 'docker' (quay.io/broadinstitute/viral-core:2.1.19), 'library\_name' (CV1372), 'platform\_name', 'platform\_unit', 'readgroup\_name', 'run\_date', 'sample\_name' (CV1372), and 'sequencing\_center'.

Field	Value
fastq_1	CV1372_S6_L001_R1_001.fastq.gz
fastq_2	CV1372_S6_L001_R2_001.fastq.gz
docker	quay.io/broadinstitute/viral-core:2.1.19
library_name	CV1372
platform_name	
platform_unit	
readgroup_name	
run_date	
sample_name	CV1372
sequencing_center	

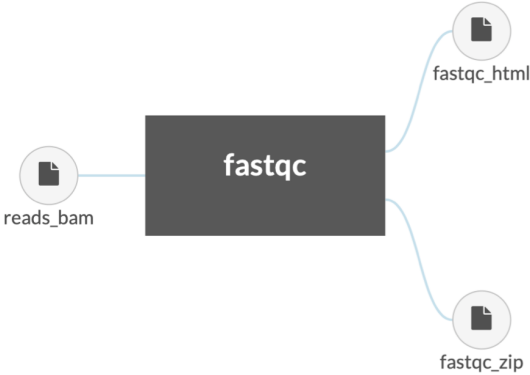
## 3. FastQC

To run a quality check of your unaligned BAM files you can run FastQC. FastQC can be found in the **assemble\_refbased** folder under the **pipelines** folder.

DNAnexus PROJECTS TOOLS HELP

Run Analysis fastqc Required field not configured Start Analysis

Reset zoom



**ANALYSIS SETTINGS** ANALYSIS INPUTS 1 STAGE SETTINGS

**Execution Name**  
fastqc

**Execute in Project**  
NCDC\_Unknown\_and\_COVID19 (project-Fk81K200K1XkGp1Z0Fv1 ...)

**Execution Output Folder**  
/ ...

**Priority** ⓘ  
Normal

**Spending Limit** ⓘ  
No Limit

After clicking on FastQC you should be directed to a window that looks like this.

Under the **ANALYSIS SETTINGS** tab select your desired folder as the **Execution Output Folder**. Keep other settings as default.

Under the **ANALYSIS INPUTS** tab

ANALYSIS SETTINGS **ANALYSIS INPUTS 1** STAGE SETTINGS

FASTQC Enable Batch OFF

**fastqc** ⓘ About this app

\* reads\_bam Select File

**COMMON**

\* docker quay.io/broadinstitute/viral-core:2.1.19

Here you select your input, which must be a **.BAM** file. Once this has been selected you can now begin your analysis by clicking on the green **Start Analysis** button on the top right corner of the screen.

## 4. Assemble rebased

To assemble your genome against a reference genome you can use the workflow **assemble\_rebased** under the **assemble\_rebased** folder nested in the **pipelines** folder.

The screenshot shows the DNAnexus 'Run Analysis' interface for the 'assemble\_rebased' workflow. The top navigation bar includes the DNAnexus logo, 'PROJECTS', 'TOOLS', and 'HELP' menus, along with search, notification, and user profile icons. Below the navigation bar, the 'Run Analysis' tab is active, showing the workflow name 'assemble\_rebased'. A yellow warning icon indicates 'Required fields not configured'. A green 'Start Analysis' button is in the top right corner. The main area is divided into two panels. The left panel displays a complex workflow diagram with numerous nodes and connecting lines. The right panel, titled 'ANALYSIS SETTINGS', contains the following fields: 'Execution Name' (set to 'assemble\_rebased'), 'Execute in Project' (set to 'NCDC\_Unknown\_and\_COVID19 (project-Fk81K200K1XkGp1Z0Fv19kZ)'), 'Execution Output Folder' (set to '/'), 'Priority' (set to 'Normal'), and 'Spending Limit' (set to 'No Limit'). Below these fields is an 'ADVANCED' section with a note about SSH access for developer debugging features, with a link to 'More about SSH Access'.

After clicking on **assemble\_rebased** you should be directed to a window that looks like this.

Under the **ANALYSIS SETTINGS** tab select your desired folder as the **Execution Output Folder**. Keep other settings as default.



ASSEMBLE\_REFBASED ▾

Enable Batch ☐ OFF

common

[? About this app](#)**novocraft\_license**

\*.lic

Select File

**reads\_unmapped\_bams**

\*.bam

Select File (Array)

**reference\_fasta**

\*.fasta

Select File

**trim\_coords\_bed**

\*.bed

Select File

Under the ANALYSIS INPUTS tab

Leave **novocraft\_license** and **trim\_coords\_bed** blank. Input your desired

## 5. Batch Input runs

This can be used to input multiple BAM files for assemble\_refbased pipeline. To do this, click on Enable Batch option in ANALYSIS INPUT. Once you do this, the batch option will be enabled

Enable Batch ☒ ON

like this

and at the same time, a box will appear next to

**reads\_unmapped\_bams**. Click on the box and the **Select File** box will change to **Configure Batch**

**reads\_unmapped\_bams**

\*.bam

Configure Batch...



Click on **Configure Batch** and it should take you to a page where you can select the BAM files needed for the run. On the left side of the page, click on Add files to Batch



Add Files to Batch

and then you can select the BAM files, then check the **Input Fields**

box. Uncheck **Create subfolder for each run** if you want all your files to be saved in a designated folder already created by you but leave it checked if you want the batch run to create a subfolder for each of the runs in your designated folder. After this, click on **Complete changes** which will take you back to ANALYSIS INPUT. Then you can **Start Analysis**.

## 6. Concatenate FASTA files

In order to merge multiple FASTA file outputs from your assemble\_refbased pipeline into a single multi-FASTA file, you can use the concatenate pipeline for this. This can be found in **pipelines > 2.1.19.0.rc-111 > sarscov2\_illumina\_full**. Then click on the Concatenate applet to open it. From here you can configure your output folder and select your array of input files. NOTE: DO NOT ENABLE BATCH FOR THIS ANALYSIS.

Input your output file name e.g. merged.fasta

Run Analysis.