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

Version 1.0 30-06-2021

Idowu Olawoye; olawoyei0303@run.edu.ng

Paul E. Oluniyi; oluniyip@run.edu.ng

Kemi Adedotun-Sulaiman; kemi.sulaiman555@gmail.com

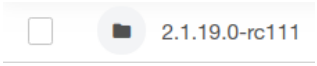
ACEGID, Redeemer's University, Ede, Osun State, Nigeria

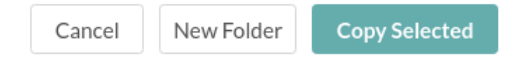
1. Creating Project

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

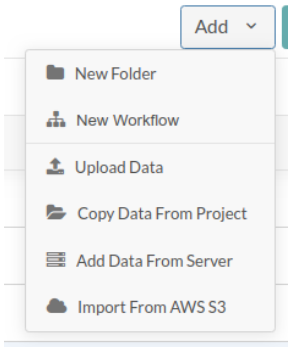
To import viral-ngs pipeline from the global DNAnexus platform into your new project. Search for “viral-ngs” in the search bar at the top of the webpage. Then select “**Broad viral-ngs dxWDL CI - Public**” project to open it.
Click on “**build**”
Click on “**quay.io**”
Click on “**broadinstitute**”
Click on “**viral-pipelines**”
Select the appropriate pipeline folder by clicking on the box next to the version number such as 2.1.19.0 

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

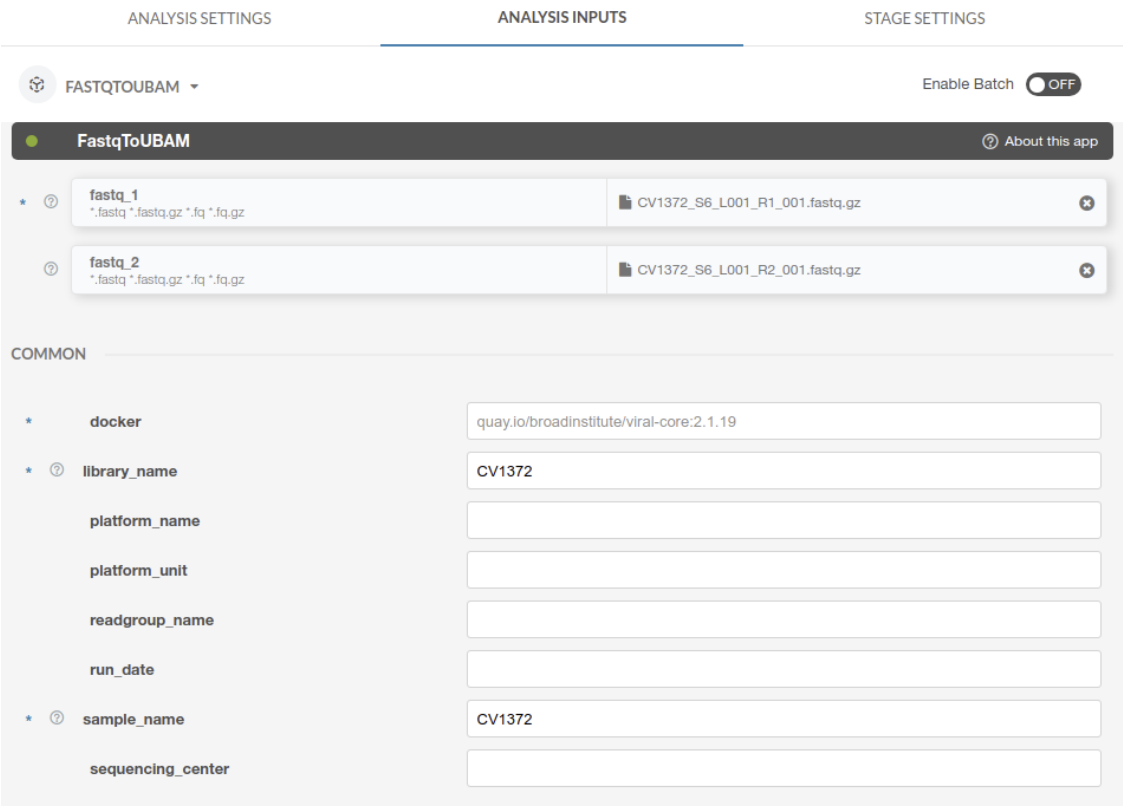
To convert your raw FASTQ files from your Sequencing machines to unaligned BAM files. Upload the FASTQ files (paired-end or single read files) to a designated folder within your



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.

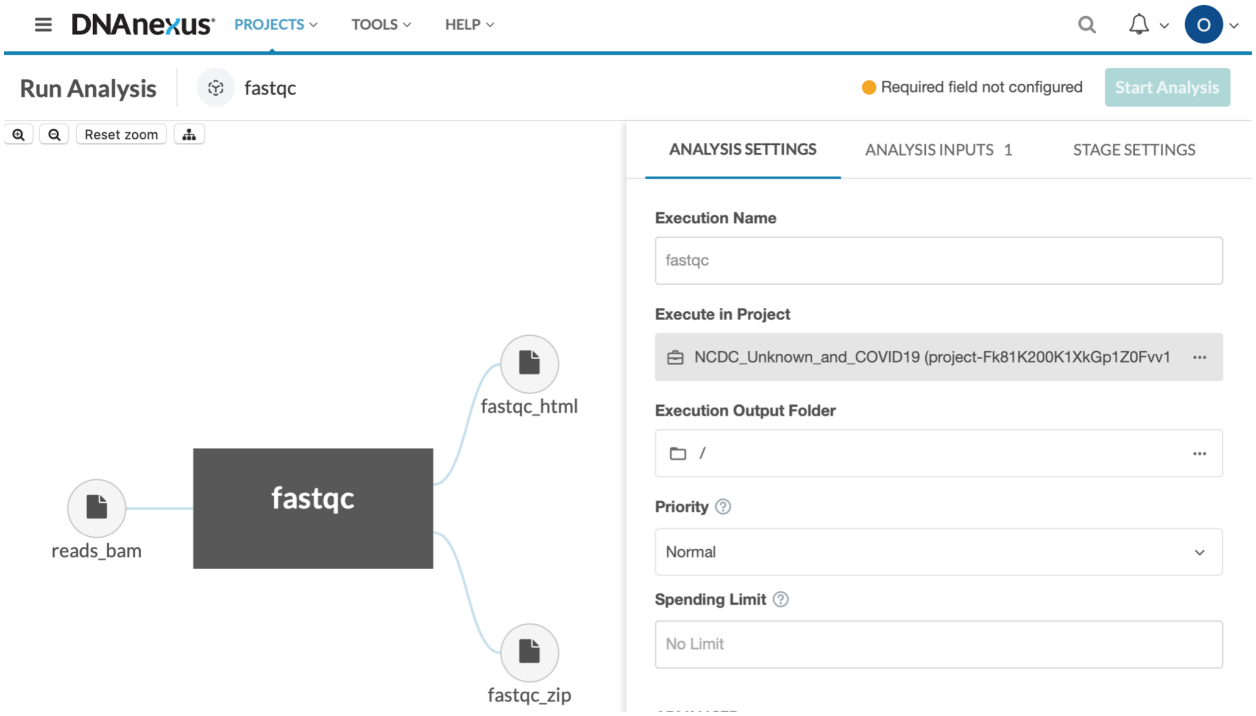
2.1. Running FASTQtoUBAM

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**.



3. FastQC

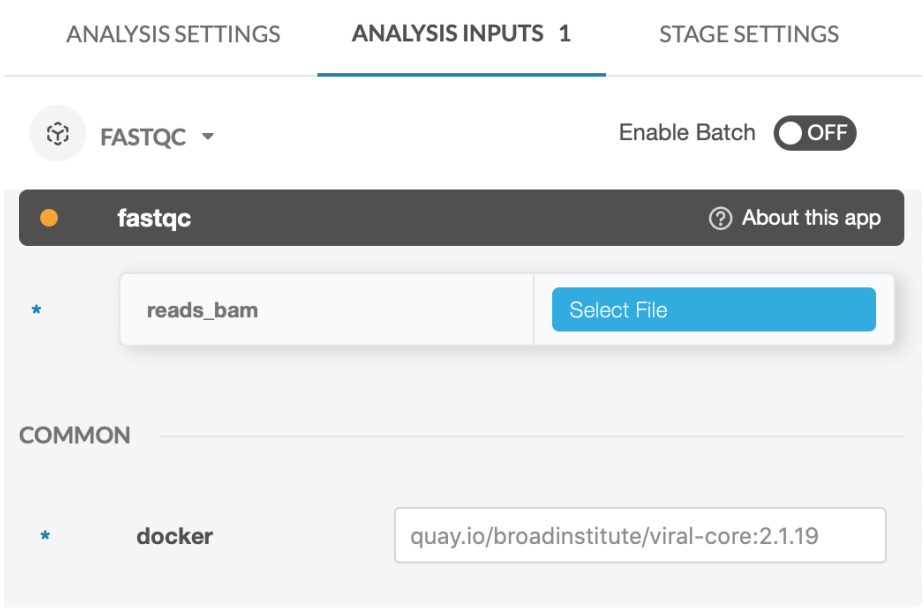
To run a quality check of your unaligned BAM files you can run FastQC. FastQC can be found in the **assemble_refbased** folder under the **pipelines** folder.



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.

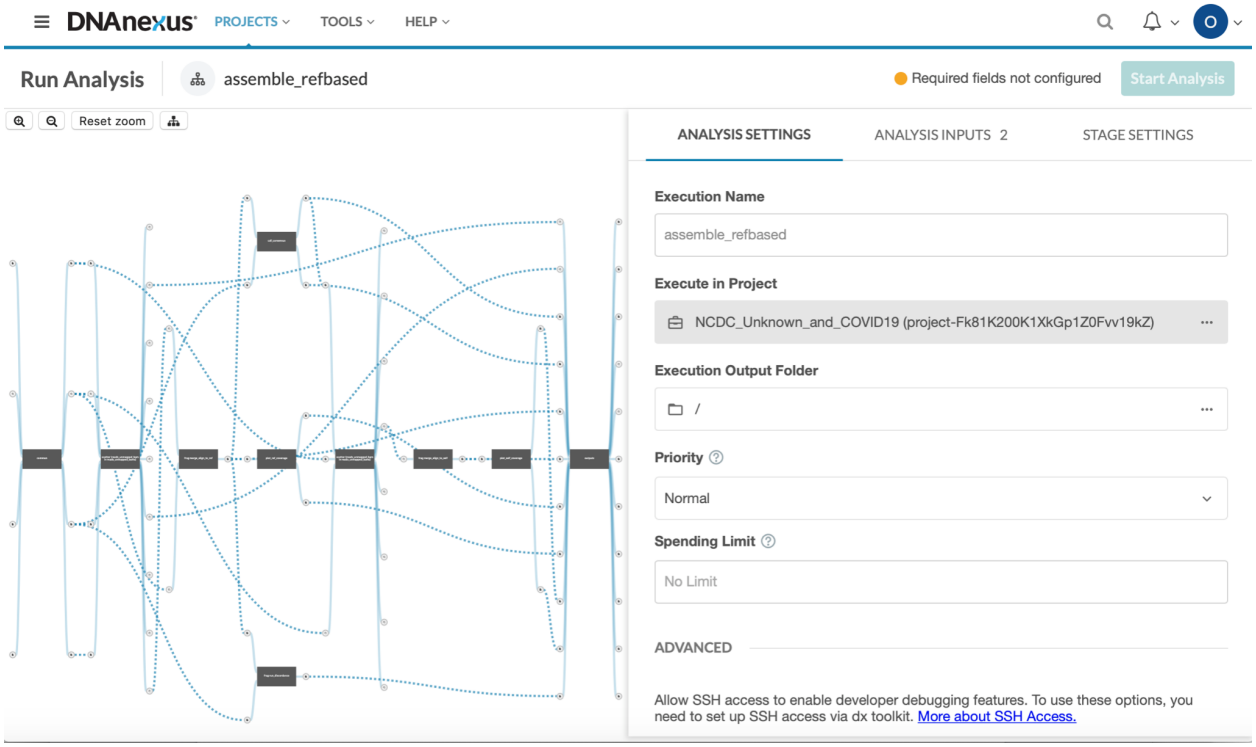
Under the **ANALYSIS INPUTS** tab



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.

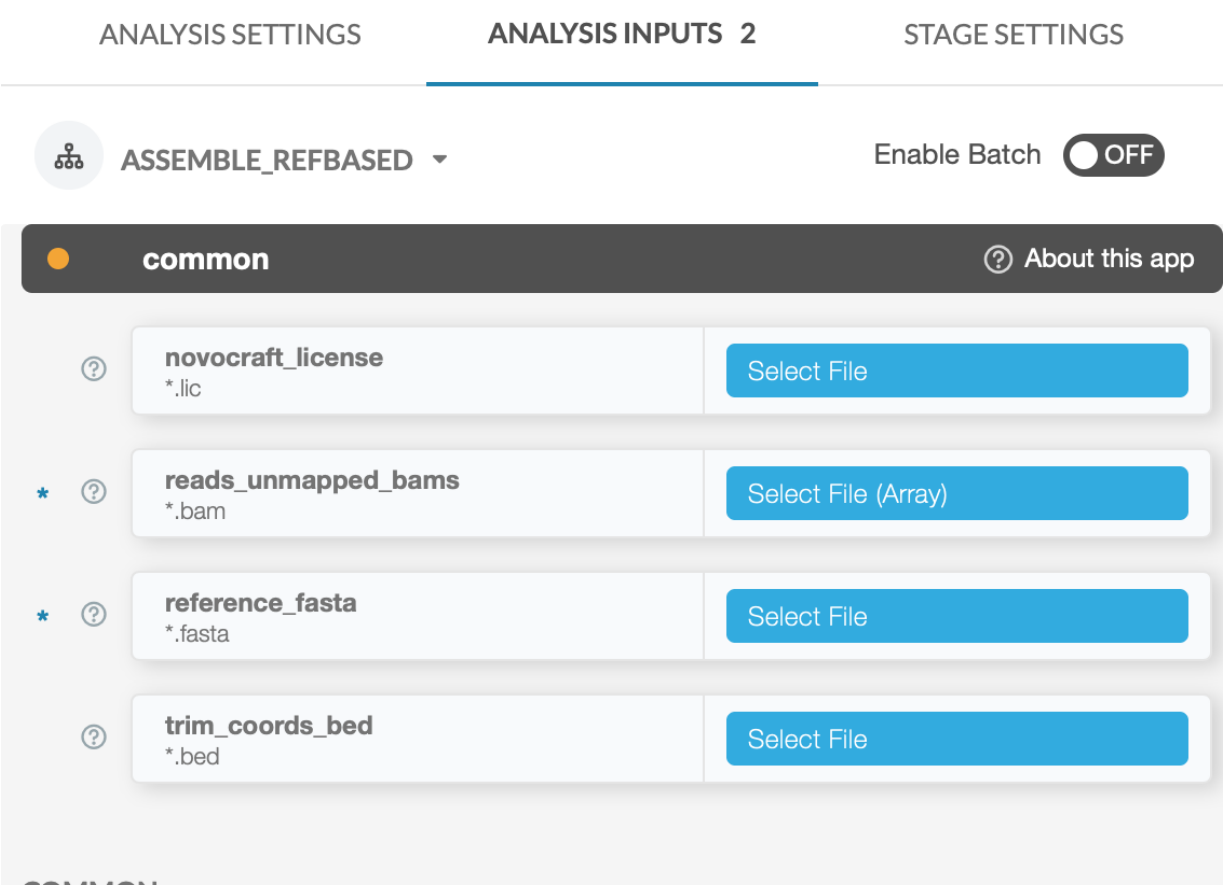
4. Assemble refbased

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



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

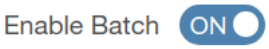
Under the ANALYSIS SETTINGS tab select your desired folder as the Execution Output Folder. Keep other settings as default.



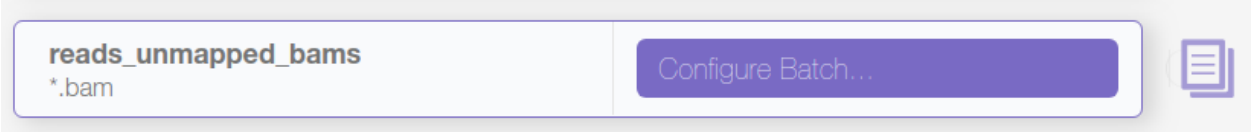
Under the ANALYSIS INPUTS tab
Leave **novocraft_license** and **trim_coords_bed** blank. Input your desired

5. Batch Input runs

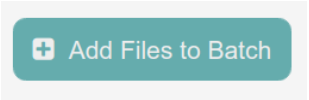
This can be used to input multiple BAM files for assemble_refbased pipeline. To do this, click on Enable Batch option in ANALYSIS INPUT. Once you do this, the batch option will be enabled

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**



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



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.