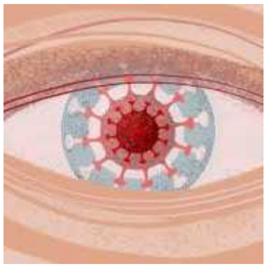
May 01, 2023

SARS-CoV-2 consensus genome reconstruction, quality control, and lineage analysis



DOI
dx.doi.org/10.17504/protocols.io.14egn2kp6g5d/v1

Benjamin Schwessinger[1]

[1]Australian National University

Benjamin Schwessinger
Australian National University

DOI: **https://dx.doi.org/10.17504/protocols.io.14egn2kp6g5d/v1**

Protocol Citation: Benjamin Schwessinger 2023. SARS-CoV-2 consensus genome reconstruction, quality control, and lineage analysis. **protocols.io** **https://dx.doi.org/10.17504/protocols.io.14egn2kp6g5d/v1**

Protocol status: Working
We use this protocol and it's working

Created: April 30, 2023

Last Modified: February 06, 2024

Protocol Integer ID: 81199

Keywords: sequencing sar, cov2 genome sequence, genomic surveillance of sar, cov2 genome, cov2 from real world rna sample, genomic surveillance, cov2 incursion scenario, genomic, protein sequence, cov2 with case interview data, real world rna sample, consensus genome coverage, rna samples per research group, genome, spread into of sar, rna sample, base pair differences relative to the original sar, lineage assignment, sequencing protocol, official sar, available sar, rna, sar, phylogenetics, cov2 in the community, cov2, nextclade lineage, original sar, anu biosecurity course, consensus reconstruction from raw read, lineage naming, mutation, pangolin lineage, mutations in the spike protein, outbreak, act pathology, genetic information

Abstract

This protocols is part of the ANU Biosecurity mini-research project #2 "An SARS-COV2 incursion scenario: Genomics, phylogenetics, and incursions." This mini-research project is modeled on the yearly Quality Assurance Program of The Royal College of Pathologists of Australia (RCPAQAP), we take part in together with ACT Pathology.

This research project is split into two major parts, identical to how the official RCPAQAP is run every year.

Part #1 is focusing on the 'wet- lab' by sequencing SARS-COV2 from real world RNA samples provided by ACT Pathology especially for our ANU biosecurity course (Thank YOU!). Here you will amplify and sequence five (5) RNA samples per research group. You will assess the SARS-COV2 genome sequences for their lineage assignments using online programs, put sequences into a global context, estimate the collection date based on genetic information, and describe mutations in the spike protein.

Part #2 is focusing on the 'dry-lab' by investigating a hypothetical incursion scenario in the so-called city Fantastica. You will combine genomic surveillance of SARS-COV2 with case interview data to trace the spread into of SARS-COV2 in the community and into high risk settings. We will provide you with real publicly available SARS-COV2 genome and fantasized case interviews. You will put these two together to trace the spread and suggest potential improvements in containment strategies with a focus on high risk settings.

This protocol describes the analysis component of Part #1. The metrics you are suppose to report for each of your samples are mostly borrowed from the official SARS-CoV-2 QAP. Don't worry if not all of these mean something to you at the moment as we will explain them again during the prac. In case all/most of your samples have < 50% genome coverage please also include the analysis of MakeUp for points 1 to 7 and TimeMakeUp for point 8. You can access the MakeUp data **here** (ANU only). Make sure to read the README file so you understand what each item relates to.

The metrics you have to report for each of our samples (or MakeUps) include the following.

1. Consensus genome coverage.
2. Average read depth; You might want to include detailed read depth plots here as well.
3. Pangolin Lineage.
4. NextClade Lineage.
5. Base pair differences relative to the original SARS-CoV-2 genome.
6. Amino acid replacements and deletions in the S (spike) protein sequence.
7. Evaluation if your and/or the MakeUp samples would make the QC cut-off 90% genome coverage and other metrics that you deem important for QC. Would you 'flag' any of your samples as standing out e.g. being negative control?

8. Approximate sampling date of your sequences for month and year.

You must report the versions of all tools used in your report and the day the analysis was performed. This is extremely important for reporting as lineage naming and such change VERY frequently during the pandemic and any outbreak.

This protocol is applicable for week 9.

The following links might be useful for your report:

The original publication that describes the sequencing protocol is here: https://academic.oup.com/biomethods/article/5/1/bpaa014/5873518?login=true

Original sources describing the consensus reconstruction from raw reads are here: https://artic.network/ncov-2019, https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html, **https://labs.epi2me.io/**, https://www.nature.com/articles/s41467-020-20075-6

Other websites and resources used in the protocol: **https://igv.org/**, **https://clades.nextstrain.org/**, **https://pangolin.cog-uk.io/**, https://genome.ucsc.edu/cgi-bin/hgPhyloPlace

Guidelines

You must have read, understood, and follow the health and safety instructions provided in the "Overview Mini-Research Project #2 BIOL3106/6106" provided on Wattle (ANU learning portal).

You must have signed and returned one copy of the "Student Safety Declaration Form For Practical Class Work" before starting any laboratory work.

You must have read and understood the Hazard Sheets (Risk assessment) of all chemicals listed bellow in the "Safety Warnings" section. These Hazard Sheets are provided on Wattle as part of the "Overview Mini-Research Project #2 BIOL3106/6106" document.

Troubleshooting

Safety warnings

This protocol does not require any hazardous substances or infectious agents. However, maintain a proper posture while working on your computer.

Before start

You must study the protocol carefully before you start. If anything is unclear post questions directly here on protocols.io.

Section I: Consensus genome reconstruction using Epi2Me and the 'wf-artic'.

1 This section aims to reconstruct each samples consensus genome based on the publication here https://academic.oup.com/biomethods/article/5/1/bpaa014/5873518?login=true and an adaptation of the artic Sars-CoV-2 analytics workflow https://artic.network/ncov-2019

The principle steps are:

1. Read filtering on quality and length.
2. Aligning reads to the reference.
3. Normalizing to 200x per base read depth on average.
4. Trimming of primer sequences.
5. Variant calling for both pools independently.
6. Combining variant calls of both pools.
7. Generation of sample's consensus genome reconstruction.
8. Generation of QC statistics like genome coverage, average read coverage etc.
9. Additional analysis like lineage calling, variant quantification, variant impact, etc.

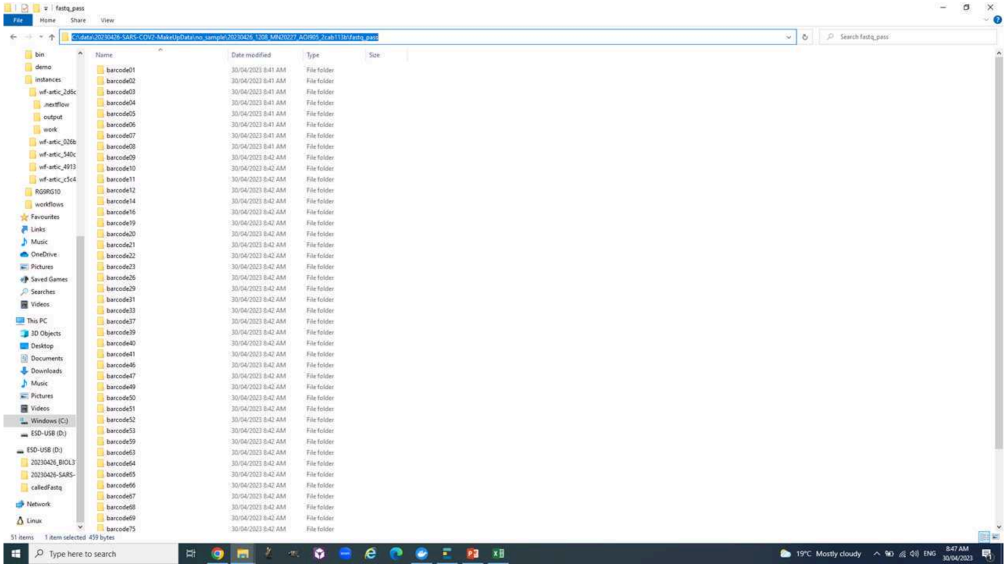
We will run all of this within Epi2Me till step 8. The input for the Epi2Me analysis will be the "fastq_pass" folder of the MinKNOW basecalling output. This folder contains one "barcodeXX" folder for each barcode identified by the basecalling software.

Overall, the Epi2Me "wf-artic" workflow takes two inputs.

1. The basecalled fastq output of MinKNOW (or equivalent e.g. guppy).
2. A sample sheet in csv (comma separated values) format with the columns "barcode", "alias", "type".

You might also want to execute this section on the MakeUp dataset if your data was of poor quality.

2 Your basecalled and demutiplexed, by barcode, fastq data should be in the 'fastq_pass' folder be in a similar to this C:\data\20230426-SARS-COV2-MakeUpData\no_sample\20230426_1208_MN20227_AOI905_2cab113b\fastq_pass.



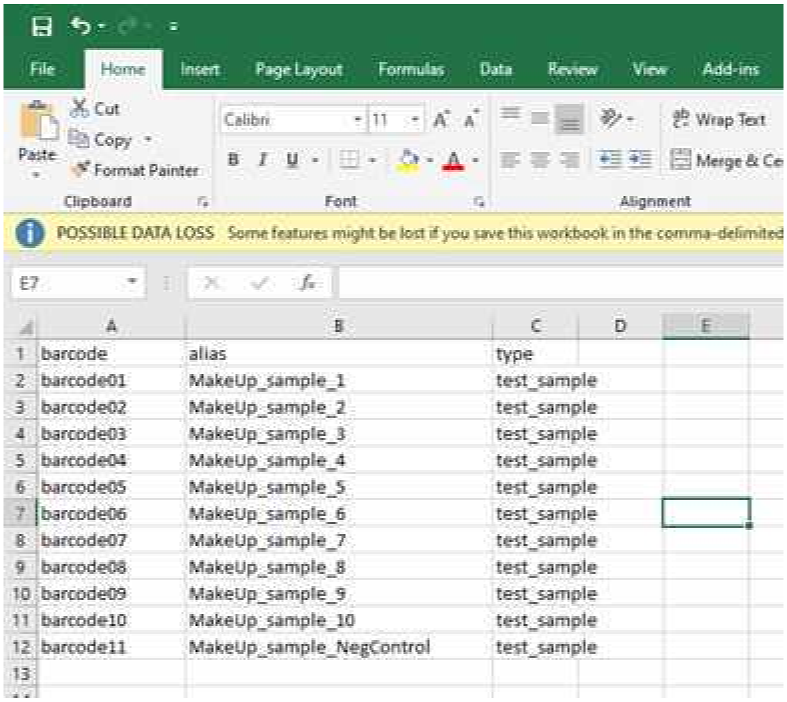
C:\data\20230426-SARS-COV2-MakeUpData\no_sample\20230426_1208_MN20227_AOI905_2cab113b\fastq_pass

Name	Date modified	Type	Size
barcode01	30/04/2023 8:41 AM	File folder	
barcode02	30/04/2023 8:41 AM	File folder	
barcode03	30/04/2023 8:41 AM	File folder	
barcode04	30/04/2023 8:41 AM	File folder	
barcode05	30/04/2023 8:41 AM	File folder	
barcode06	30/04/2023 8:41 AM	File folder	
barcode07	30/04/2023 8:41 AM	File folder	
barcode08	30/04/2023 8:41 AM	File folder	
barcode09	30/04/2023 8:42 AM	File folder	
barcode10	30/04/2023 8:42 AM	File folder	
barcode11	30/04/2023 8:42 AM	File folder	
barcode12	30/04/2023 8:42 AM	File folder	
barcode14	30/04/2023 8:42 AM	File folder	
barcode16	30/04/2023 8:42 AM	File folder	
barcode19	30/04/2023 8:42 AM	File folder	
barcode20	30/04/2023 8:42 AM	File folder	
barcode21	30/04/2023 8:42 AM	File folder	
barcode22	30/04/2023 8:42 AM	File folder	
barcode23	30/04/2023 8:42 AM	File folder	
barcode26	30/04/2023 8:42 AM	File folder	
barcode29	30/04/2023 8:42 AM	File folder	
barcode31	30/04/2023 8:42 AM	File folder	
barcode33	30/04/2023 8:42 AM	File folder	
barcode37	30/04/2023 8:42 AM	File folder	
barcode39	30/04/2023 8:42 AM	File folder	
barcode40	30/04/2023 8:42 AM	File folder	
barcode41	30/04/2023 8:42 AM	File folder	
barcode46	30/04/2023 8:42 AM	File folder	
barcode47	30/04/2023 8:42 AM	File folder	
barcode49	30/04/2023 8:42 AM	File folder	
barcode50	30/04/2023 8:42 AM	File folder	
barcode51	30/04/2023 8:42 AM	File folder	
barcode52	30/04/2023 8:42 AM	File folder	
barcode53	30/04/2023 8:42 AM	File folder	
barcode59	30/04/2023 8:42 AM	File folder	
barcode63	30/04/2023 8:42 AM	File folder	
barcode64	30/04/2023 8:42 AM	File folder	
barcode65	30/04/2023 8:42 AM	File folder	
barcode66	30/04/2023 8:42 AM	File folder	
barcode67	30/04/2023 8:42 AM	File folder	
barcode68	30/04/2023 8:42 AM	File folder	
barcode69	30/04/2023 8:42 AM	File folder	
barcode75	30/04/2023 8:42 AM	File folder	

51 items 1 item selected 459 bytes

fastq_pass folder with all the demultiplexed barcodeXX folders containing fastq files for each barcode

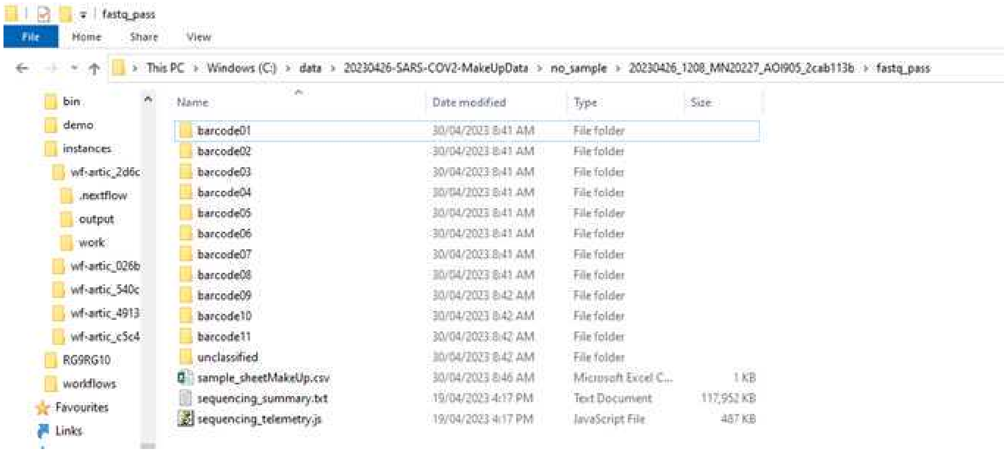
3 We need to update the sample sheet to reflect your specific run and samples. This sheet should be located in the same folder your 'fastq_pass' barcode folders. It should look similar to the one below. (Done for you.)



POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited

	A	B	C	D	E
1	barcode	alias	type		
2	barcode01	MakeUp_sample_1	test_sample		
3	barcode02	MakeUp_sample_2	test_sample		
4	barcode03	MakeUp_sample_3	test_sample		
5	barcode04	MakeUp_sample_4	test_sample		
6	barcode05	MakeUp_sample_5	test_sample		
7	barcode06	MakeUp_sample_6	test_sample		
8	barcode07	MakeUp_sample_7	test_sample		
9	barcode08	MakeUp_sample_8	test_sample		
10	barcode09	MakeUp_sample_9	test_sample		
11	barcode10	MakeUp_sample_10	test_sample		
12	barcode11	MakeUp_sample_NegControl	test_sample		
13					

4 Before we start with the workflow we need to make a backup (e.g. copy of the 'fastq_pass' folder on the desktop for now) and then delete all the barcode folders we have not specified in the sample sheet. We also need to delete all folders of barcodes we have used but have not received any data.

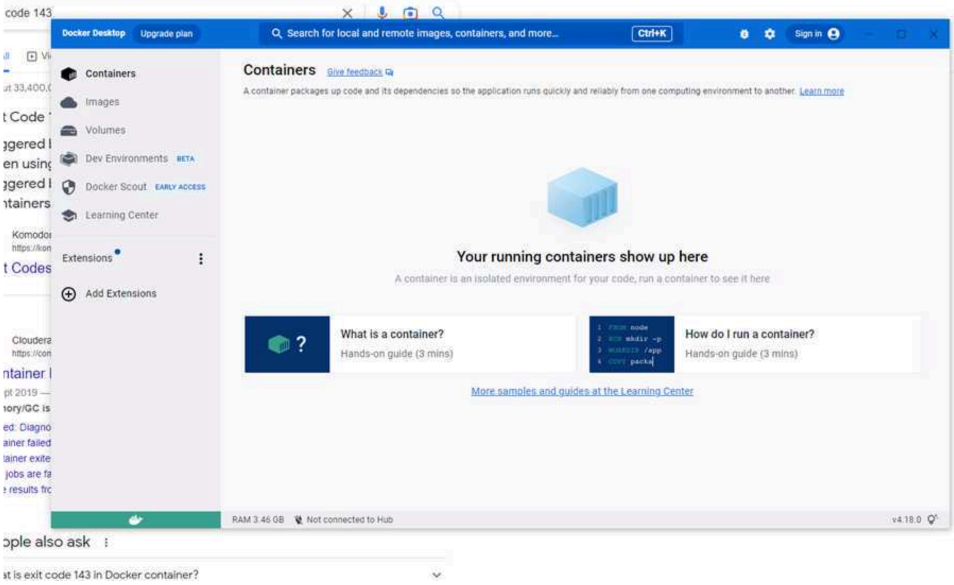


Name	Date modified	Type	Size
barcode01	30/04/2023 8:41 AM	File folder	
barcode02	30/04/2023 8:41 AM	File folder	
barcode03	30/04/2023 8:41 AM	File folder	
barcode04	30/04/2023 8:41 AM	File folder	
barcode05	30/04/2023 8:41 AM	File folder	
barcode06	30/04/2023 8:41 AM	File folder	
barcode07	30/04/2023 8:41 AM	File folder	
barcode08	30/04/2023 8:41 AM	File folder	
barcode09	30/04/2023 8:42 AM	File folder	
barcode10	30/04/2023 8:42 AM	File folder	
barcode11	30/04/2023 8:42 AM	File folder	
unclassified	30/04/2023 8:42 AM	File folder	
sample_sheetMakeUp.csv	30/04/2023 8:46 AM	Microsoft Excel C...	1 KB
sequencing_summary.txt	19/04/2023 4:17 PM	Text Document	117,952 KB
sequencing_telemetry.js	19/04/2023 4:17 PM	JavaScript File	487 KB

Fastq_pass folder after clean-up and with sample sheet.

5 Now we are ready to set-up and run the Epi2Me workflow for artic analysis.

6 Start the "Docker App" and make sure it is running.



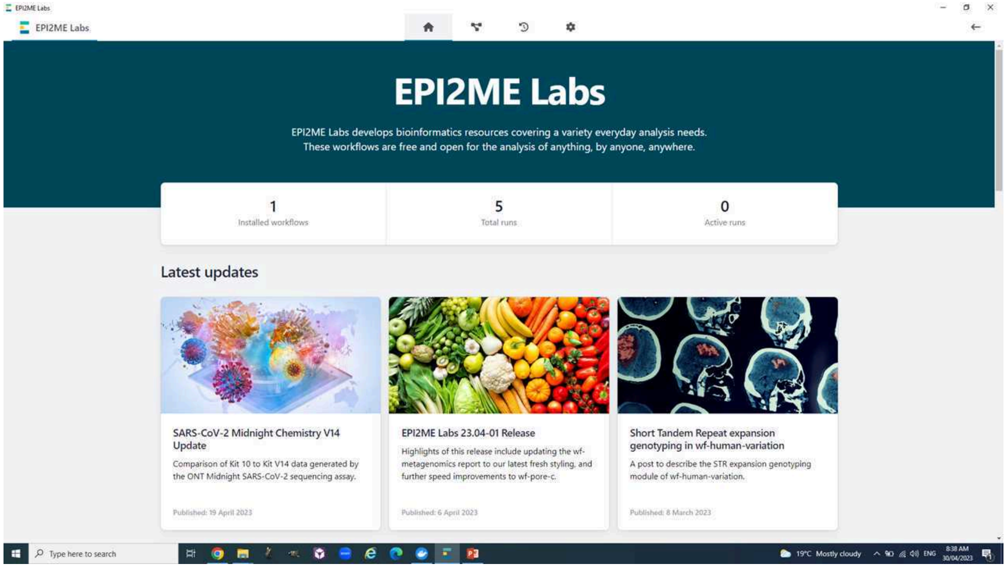
Containers

A container packages up code and its dependencies so the application runs quickly and reliably from one computing environment to another.

Your running containers show up here

A container is an isolated environment for your code, run a container to see it here

7 Start Epi2Me.



EPI2ME Labs

EPI2ME Labs develops bioinformatics resources covering a variety everyday analysis needs.
These workflows are free and open for the analysis of anything, by anyone, anywhere.

1 Installed workflows | 5 Total runs | 0 Active runs

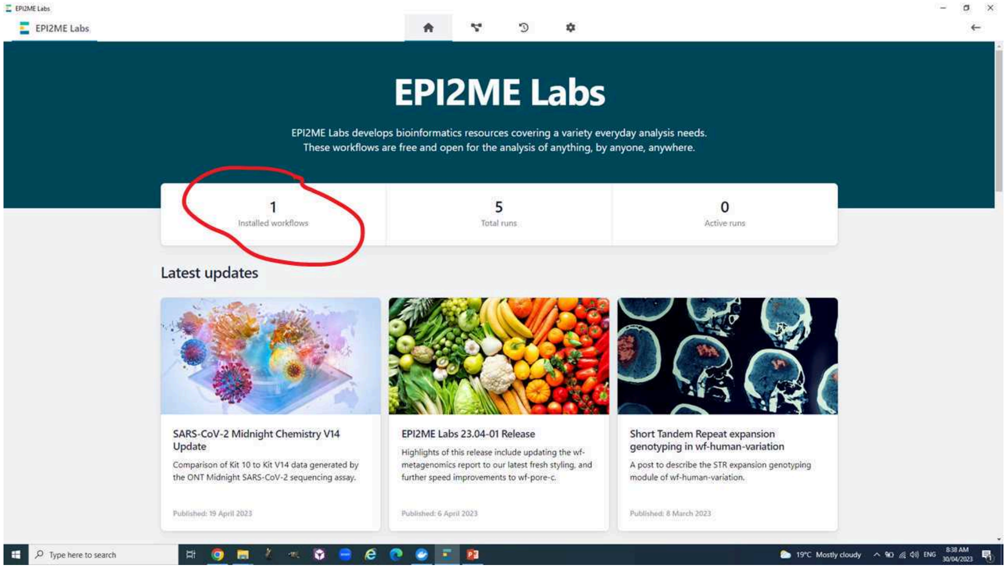
Latest updates

SARS-CoV-2 Midnight Chemistry V14 Update
Comparison of Kit 10 to Kit V14 data generated by the ONT Midnight SARS-CoV-2 sequencing assay.
Published: 19 April 2023

EPI2ME Labs 23.04-01 Release
Highlights of this release include updating the wf-metagenomics report to our latest fresh styling, and further speed improvements to wf-pore-c.
Published: 6 April 2023

Short Tandem Repeat expansion genotyping in wf-human-variation
A post to describe the STR expansion genotyping module of wf-human-variation.
Published: 8 March 2023

8 Click the "Installed workflows" bottom.



EPI2ME Labs

EPI2ME Labs develops bioinformatics resources covering a variety everyday analysis needs.
These workflows are free and open for the analysis of anything, by anyone, anywhere.

1 Installed workflows | 5 Total runs | 0 Active runs

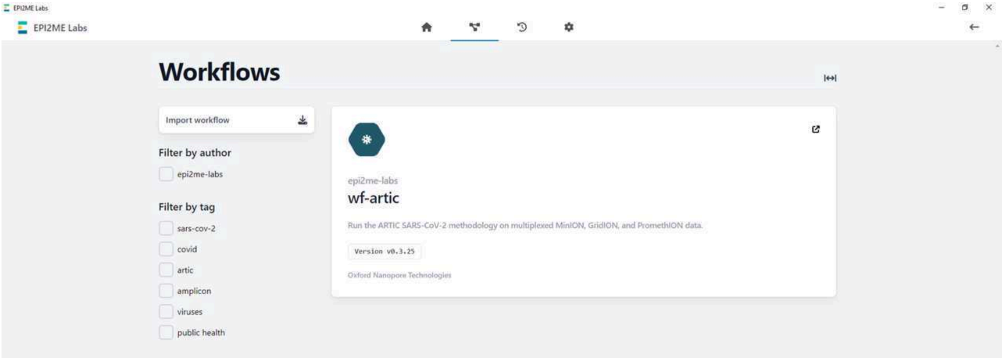
Latest updates

SARS-CoV-2 Midnight Chemistry V14 Update
Comparison of Kit 10 to Kit V14 data generated by the ONT Midnight SARS-CoV-2 sequencing assay.
Published: 19 April 2023

EPI2ME Labs 23.04-01 Release
Highlights of this release include updating the wf-metagenomics report to our latest fresh styling, and further speed improvements to wf-pore-c.
Published: 6 April 2023

Short Tandem Repeat expansion genotyping in wf-human-variation
A post to describe the STR expansion genotyping module of wf-human-variation.
Published: 8 March 2023

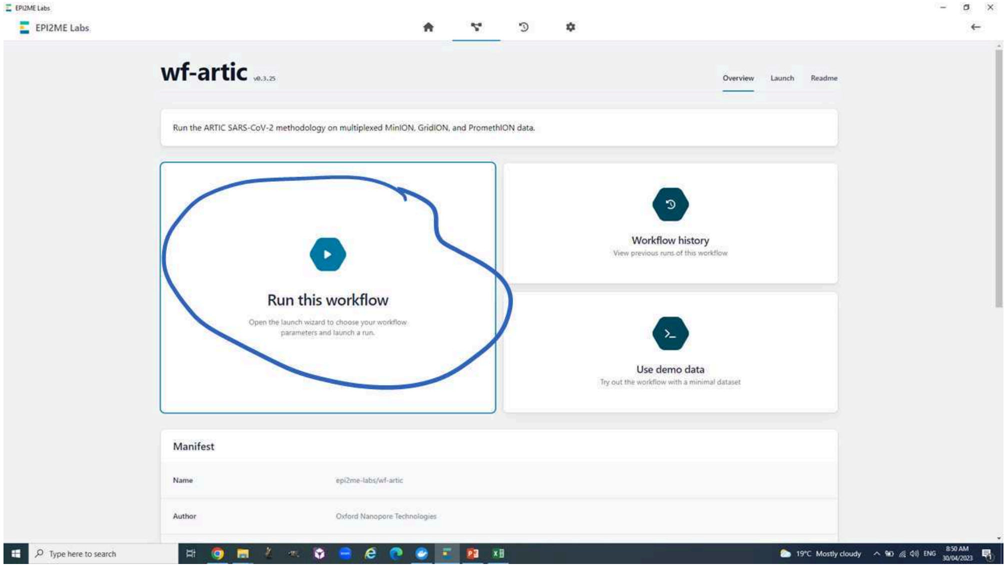
9 This should show "wf-artic" to be available.



Workflows

epi2me-labs
wf-artic
Run the ARTIC SARS-CoV-2 methodology on multiplexed MinION, GridION, and PromethION data.
Version v0.3.25
Oxford Nanopore Technologies

10 Now select the "wf-artic".



wf-artic v0.3.25

Run the ARTIC SARS-CoV-2 methodology on multiplexed MinION, GridION, and PromethION data.

Open the launch wizard to choose your workflow parameters and launch a run.

View previous runs of this workflow

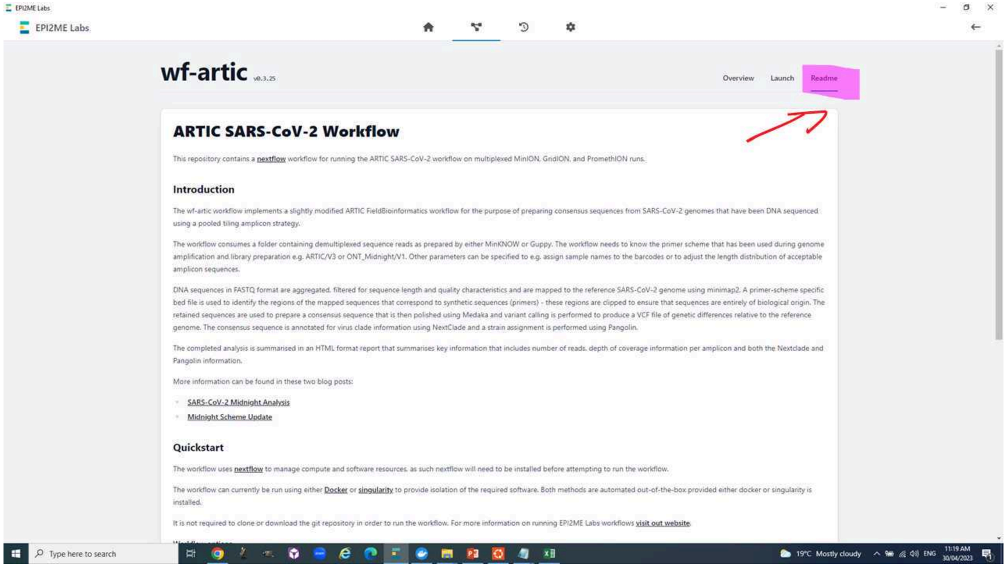
Try out the workflow with a minimal dataset

Manifest

Name epi2me-labs/wf-artic

Author Oxford Nanopore Technologies

10.1 Site note. Also consider reading the "README" section to get more background for your report.



wf-artic v0.3.25

ARTIC SARS-CoV-2 Workflow

This repository contains a nextflow workflow for running the ARTIC SARS-CoV-2 workflow on multiplexed MinION, GridION, and PromethION runs.

Introduction

The wf-artic workflow implements a slightly modified ARTIC FieldBioinformatics workflow for the purpose of preparing consensus sequences from SARS-CoV-2 genomes that have been DNA sequenced using a pooled tiling amplicon strategy.

The workflow consumes a folder containing demultiplexed sequence reads as prepared by either MinKNOW or Guppy. The workflow needs to know the primer scheme that has been used during genome amplification and library preparation e.g. ARTIC/V3 or ONT_Midnight/V1. Other parameters can be specified to e.g. assign sample names to the barcodes or to adjust the length distribution of acceptable amplicon sequences.

DNA sequences in FASTQ format are aggregated, filtered for sequence length and quality characteristics and are mapped to the reference SARS-CoV-2 genome using minimap2. A primer-scheme specific bed file is used to identify the regions of the mapped sequences that correspond to synthetic sequences (primers) - these regions are clipped to ensure that sequences are entirely of biological origin. The retained sequences are used to prepare a consensus sequence that is then polished using Medaka and variant calling is performed to produce a VCF file of genetic differences relative to the reference genome. The consensus sequence is annotated for virus clade information using NextClade and a strain assignment is performed using Pangolin.

The completed analysis is summarised in an HTML format report that summarises key information that includes number of reads, depth of coverage information per amplicon and both the Nextclade and Pangolin information.

More information can be found in these two blog posts:

- SARS-CoV-2 Midnight Analysis
- Midnight Scheme Update

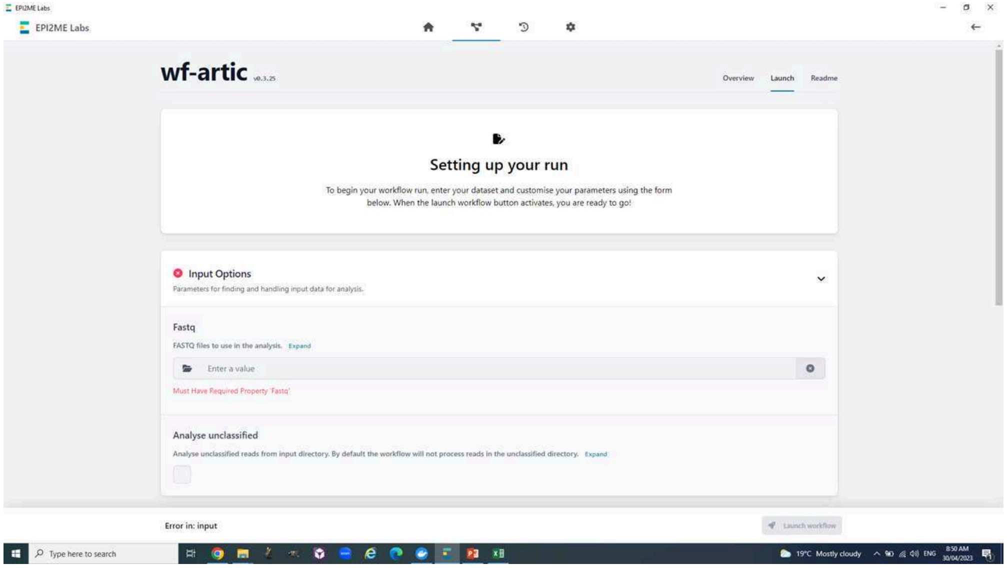
Quickstart

The workflow uses nextflow to manage compute and software resources, as such nextflow will need to be installed before attempting to run the workflow.

The workflow can currently be run using either Docker or singularity to provide isolation of the required software. Both methods are automated out-of-the-box provided either docker or singularity is installed.

It is not required to clone or download the git repository in order to run the workflow. For more information on running EPI2ME Labs workflows visit out website.

11 Now we are ready to set up our "wf-artic" specific to our samples and wet-lab protocols.



wf-artic v0.3.25

Setting up your run

To begin your workflow run, enter your dataset and customise your parameters using the form below. When the launch workflow button activates, you are ready to go!

Input Options

Parameters for finding and handling input data for analysis.

Fastq

FASTQ files to use in the analysis. Expand

Enter a value

Must Have Required Property 'Fastq'

Analyse unclassified

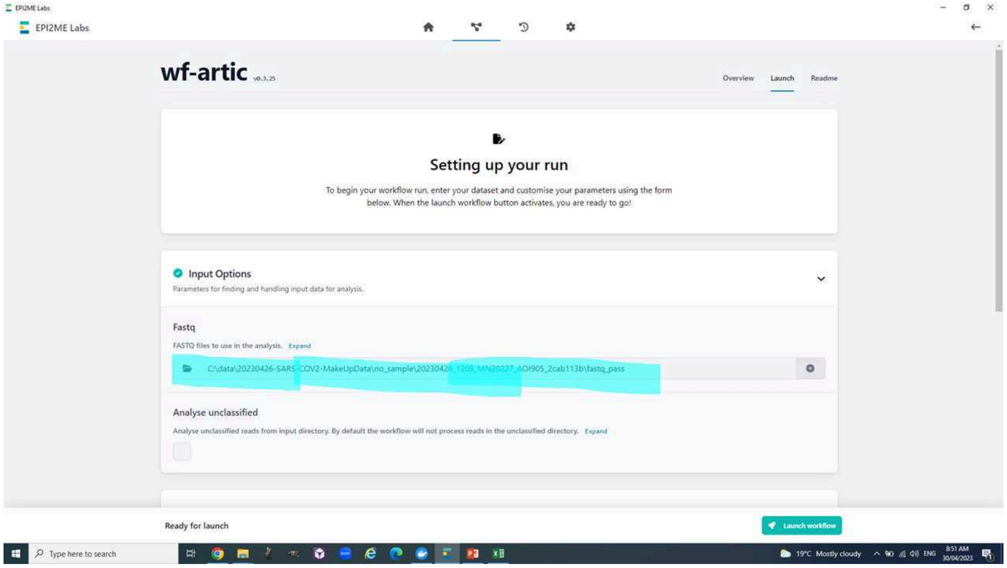
Analyse unclassified reads from input directory. By default the workflow will not process reads in the unclassified directory. Expand

□

Error in: input

12 Tell the program where to find your 'fastq_pass' folder.

Similar to "C:\data\20230426-SARS-COV2-MakeUpData\no_sample\20230426_1208_MN20227_AOI905_2cab113b\fastq_pass".



wf-artic v0.3.25

Setting up your run

To begin your workflow run, enter your dataset and customise your parameters using the form below. When the launch workflow button activates, you are ready to go!

Input Options

Parameters for finding and handling input data for analysis.

Fastq

FASTQ files to use in the analysis.

C:\data\20230426-SARS-COV2-MakeUpData\no_sample\20230426_1208_MN20227_AOI905_2cab113b\fastq_pass

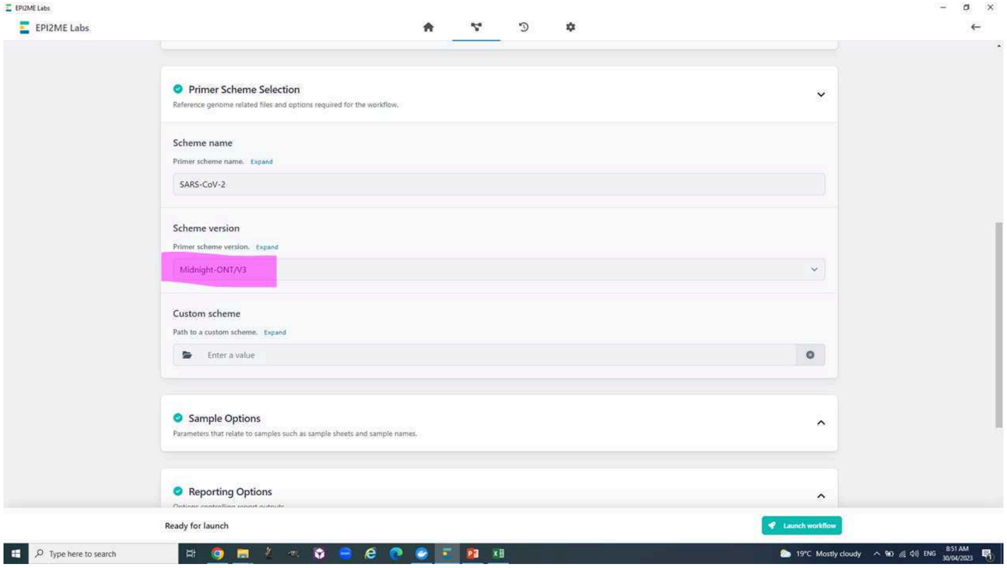
Analyse unclassified

Analyse unclassified reads from input directory. By default the workflow will not process reads in the unclassified directory.

□

Ready for launch

13 Now set up all the other fields.



Primer Scheme Selection

Reference genome related files and options required for the workflow.

Scheme name

Primer scheme name.

SARS-CoV-2

Scheme version

Primer scheme version.

Midnight-ONT/V3

Custom scheme

Path to a custom scheme.

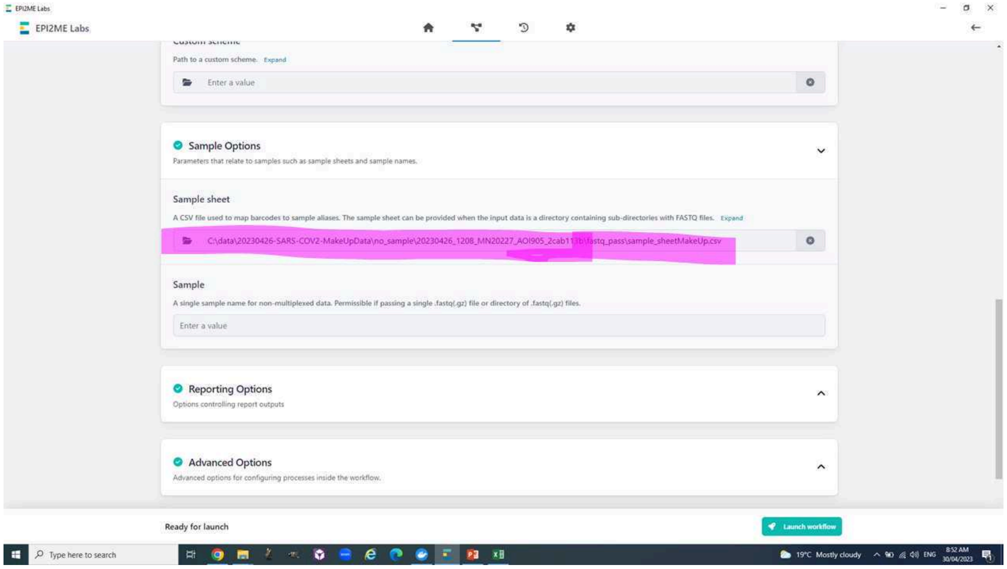
Enter a value

Sample Options

Parameters that relate to samples such as sample sheets and sample names.

Reporting Options

Ready for launch



Custom scheme

Path to a custom scheme. Expand

Enter a value

Sample Options

Parameters that relate to samples such as sample sheets and sample names.

Sample sheet

A CSV file used to map barcodes to sample aliases. The sample sheet can be provided when the input data is a directory containing sub-directories with FASTQ files. Expand

C:\data\20230426-SARS-COV2-MakeUpData\no_sample\20230426_1208_MN20227_AOI905_2cab111b\fastq_pass\sample_sheetMakeUp.csv

Sample

A single sample name for non-multiplexed data. Permissible if passing a single .fastq(.gz) file or directory of .fastq(.gz) files.

Enter a value

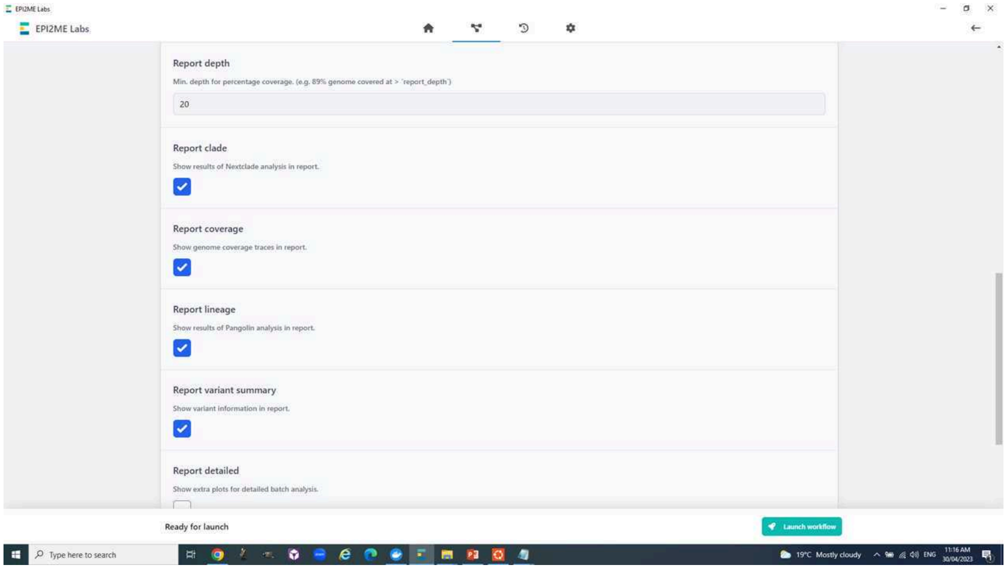
Reporting Options

Options controlling report outputs

Advanced Options

Advanced options for configuring processes inside the workflow.

Ready for launch



Report depth

Min. depth for percentage coverage. (e.g. 89% genome covered at > `report_depth`)

20

Report clade

Show results of Nextclade analysis in report.

☑

Report coverage

Show genome coverage traces in report.

☑

Report lineage

Show results of Pangolin analysis in report.

☑

Report variant summary

Show variant information in report.

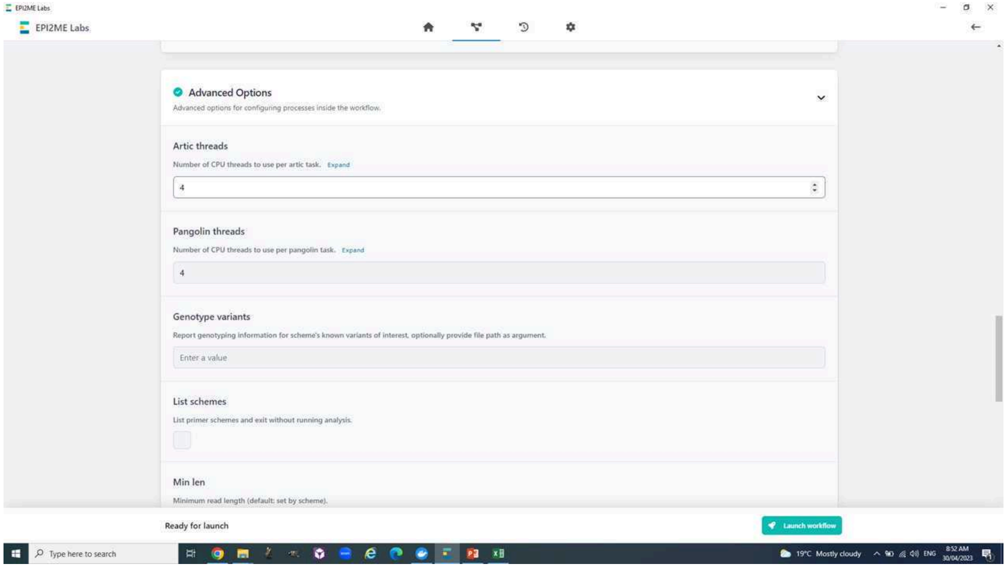
☑

Report detailed

Show extra plots for detailed batch analysis.

□

Ready for launch



Advanced Options

Advanced options for configuring processes inside the workflow.

Artic threads

Number of CPU threads to use per artic task.

4

Pangolin threads

Number of CPU threads to use per pangolin task.

4

Genotype variants

Report genotyping information for scheme's known variants of interest, optionally provide file path as argument.

Enter a value

List schemes

List primer schemes and exit without running analysis.

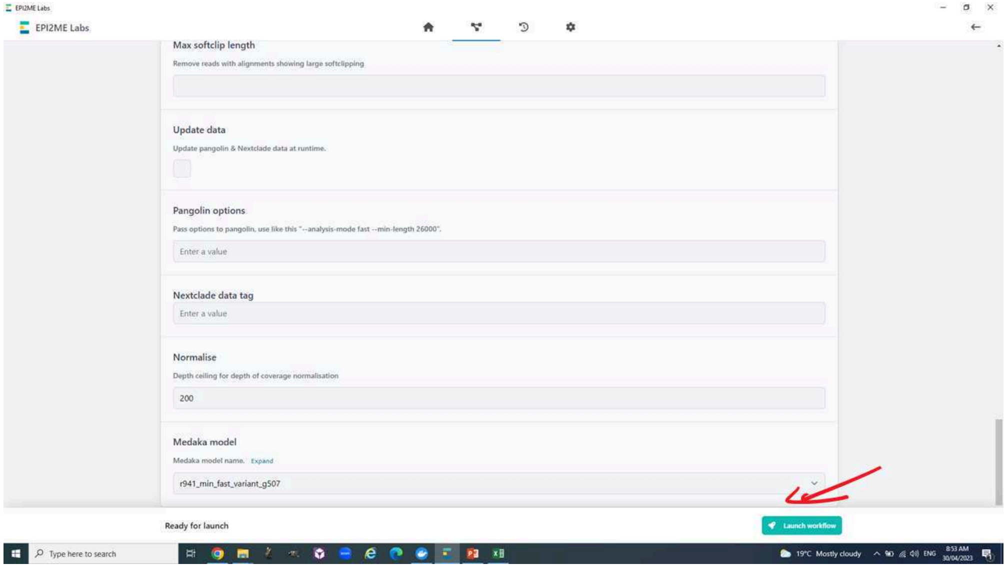
□

Min len

Minimum read length (default: set by scheme).

Ready for launch

14 Hit the "Launch Workflow" bottom.



Max softclip length

Remove reads with alignments showing large softclipping

Update data

Update pangolin & Nextclade data at runtime.

□

Pangolin options

Pass options to pangolin, use like this "--analysis-mode fast --min-length 26000".

Enter a value

Nextclade data tag

Enter a value

Normalise

Depth ceiling for depth of coverage normalisation

200

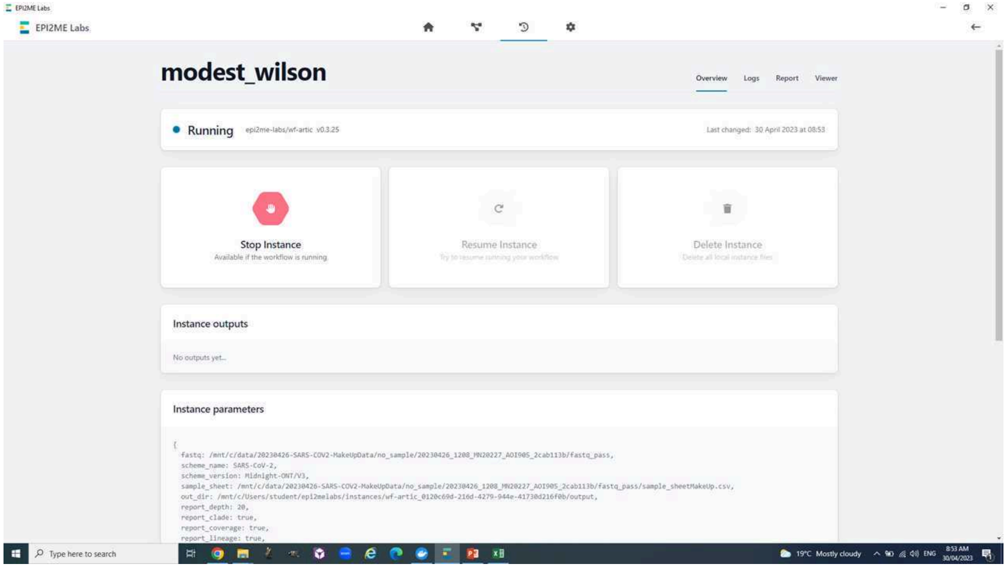
Medaka model

Medaka model name.

r941_min_fast_variant_g507

Ready for launch

15 Now all should be running smoothly.


EPI2ME Labs
modest_wilson
Overview
Logs
Report
Viewer
Running
epi2me-labs/wf-artic v0.3.25
Last changed: 30 April 2023 at 08:53
Stop Instance
Available if the workflow is running
Resume Instance
Try to resume running your workflow
Delete Instance
Delete all local instance files
Type here to search
19°C Mostly cloudy
8:53 AM
30/04/2023

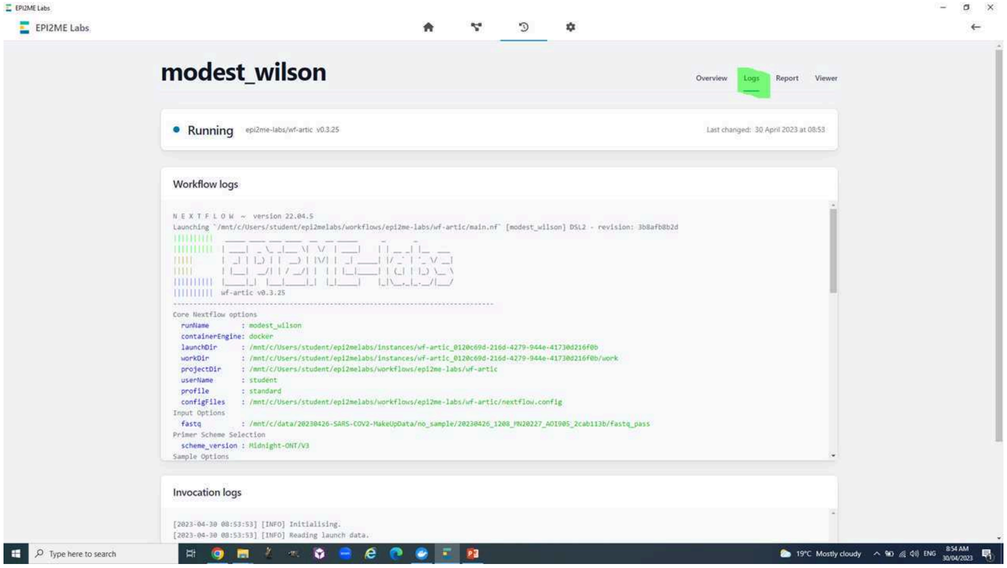
Instance outputs

No outputs yet...

Instance parameters

```
{
  fastq: /mnt/c/data/20230426-SARS-COV2-MakeUpData/no_sample/20230426_1208_MN20227_AOI905_2cab113b/fastq_pass,
  scheme_name: SARS-CoV-2,
  scheme_version: Midnight-ONT/V3,
  sample_sheet: /mnt/c/data/20230426-SARS-COV2-MakeUpData/no_sample/20230426_1208_MN20227_AOI905_2cab113b/fastq_pass/sample_sheetMakeUp.csv,
  out_dir: /mnt/c/Users/student/epi2melabs/instances/wf-artic_0120c69d-216d-4279-944e-41730d216f0b/output,
  report_depth: 20,
  report_clade: true,
  report_coverage: true,
  report_lineage: true,
```

... and this after 2-5 minutes.



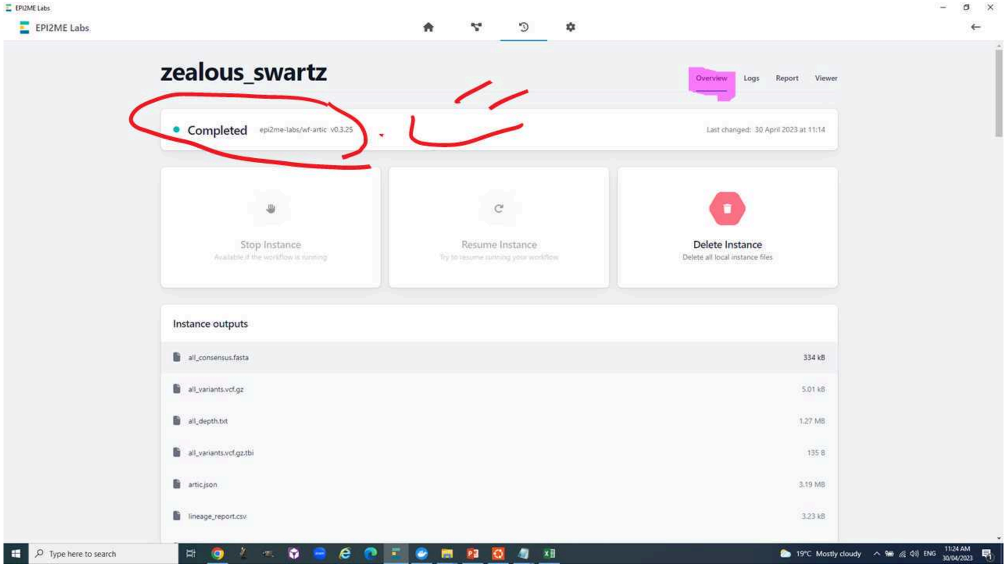
Workflow logs

```
N E X T F L O W  ~  version 22.04.5
Launching `/mnt/c/Users/student/epi2melabs/workflows/epi2me-labs/wf-artic/main.nf` [modest_wilson] DSL2 - revision: 3b8afb8b2d
||||||||||  wf-artic v0.3.25
--------------------------------------------------------------------------------
Core Nextflow options
  runName        : modest_wilson
  containerEngine: docker
  launchDir      : /mnt/c/Users/student/epi2melabs/instances/wf-artic_0120c69d-216d-4279-944e-41730d216f0b
  workDir        : /mnt/c/Users/student/epi2melabs/instances/wf-artic_0120c69d-216d-4279-944e-41730d216f0b/work
  projectDir     : /mnt/c/Users/student/epi2melabs/workflows/epi2me-labs/wf-artic
  userName       : student
  profile        : standard
  configFiles    : /mnt/c/Users/student/epi2melabs/workflows/epi2me-labs/wf-artic/nextflow.config
Input Options
  fastq          : /mnt/c/data/20230426-SARS-COV2-MakeUpData/no_sample/20230426_1208_MN20227_AOI905_2cab113b/fastq_pass
Primer Scheme Selection
  scheme_version : Midnight-ONT/V3
Sample Options
```

Invocation logs

```
[2023-04-30 08:53:53] [INFO] Initialising.
[2023-04-30 08:53:53] [INFO] Reading launch data.
```

16 If all goes well... you will see the following complete screen in about 45 minutes.



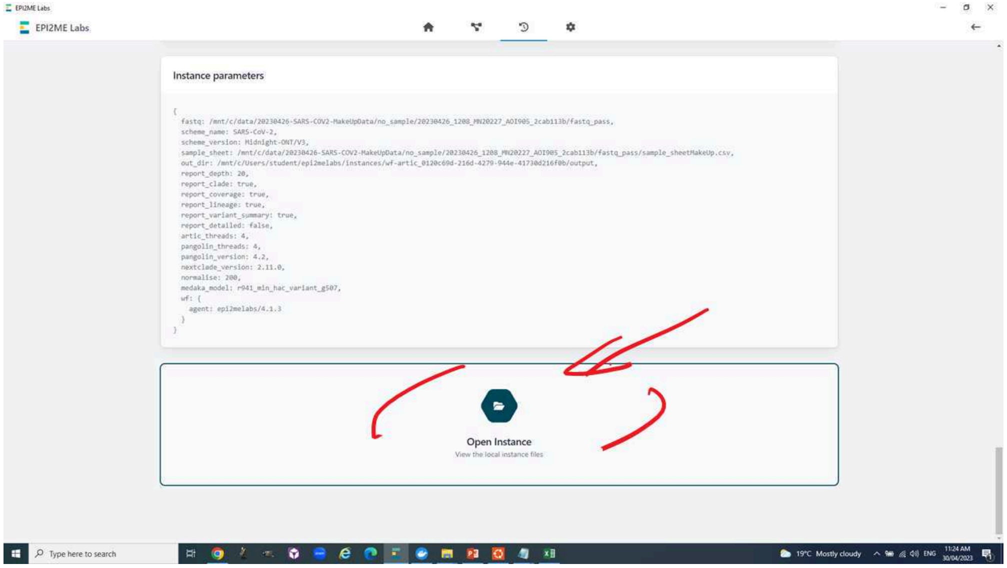
zealous_swartz

Completed epi2me-labs/wf-artic v0.3.25 Last changed: 30 April 2023 at 11:14

Instance outputs

all_consensus.fasta	334 kB
all_variants.vcf.gz	5.01 kB
all_depth.txt	1.27 MB
all_variants.vcf.gz.tbi	135 B
artic.json	3.19 MB
lineage_report.csv	3.23 kB

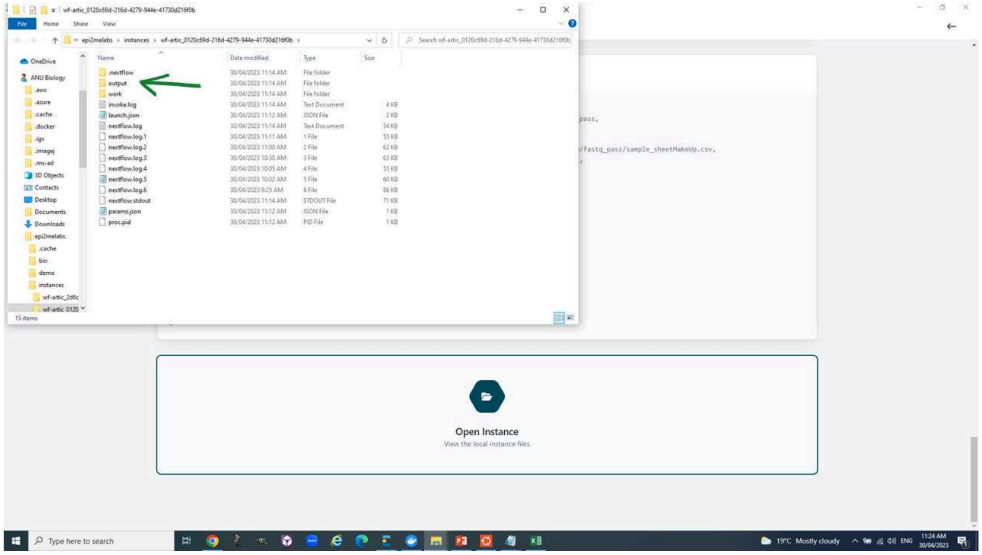
17 Now you can open the instance output by hitting the following link at the bottom of the page.



Instance parameters

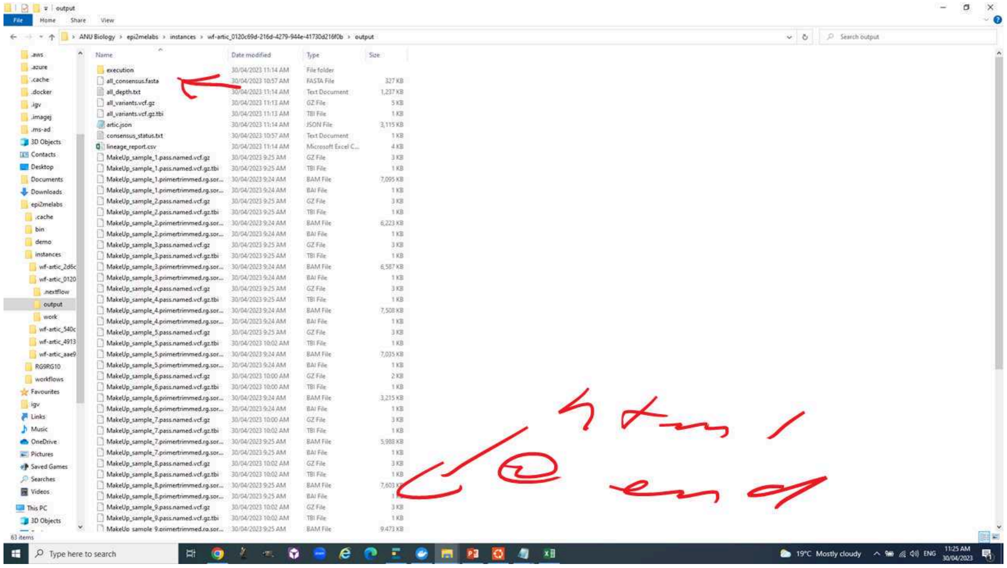
```
{
  fastq: /mnt/c/data/20230426-SARS-COV2-MakeUpData/no_sample/20230426_1208_MN20227_AOI905_2cab113b/fastq_pass,
  scheme_name: SARS-CoV-2,
  scheme_version: Midnight-ONT/V3,
  sample_sheet: /mnt/c/data/20230426-SARS-COV2-MakeUpData/no_sample/20230426_1208_MN20227_AOI905_2cab113b/fastq_pass/sample_sheetMakeUp.csv,
  out_dir: /mnt/c/Users/student/epi2melabs/instances/wf-artic_0120c69d-216d-4279-944e-41730d216f0b/output,
  report_depth: 20,
  report_clade: true,
  report_coverage: true,
  report_lineage: true,
  report_variant_summary: true,
  report_detailed: false,
  artic_threads: 4,
  pangolin_threads: 4,
  pangolin_version: 4.2,
  nextclade_version: 2.11.0,
  normalise: 200,
  medaka_model: r941_min_hac_variant_g507,
  wf: {
    agent: epi2melabs/4.1.3
  }
}
```

18 This will open the folder of your specific workflow instance that also contains your "output" folder.



Name	Date modified	Type	Size
.nextflow	30/04/2023 11:14 AM	File folder	
output	30/04/2023 11:14 AM	File folder	
work	30/04/2023 11:14 AM	File folder	
invoke.log	30/04/2023 11:14 AM	Text Document	4 KB
launch.json	30/04/2023 11:12 AM	JSON File	2 KB
nextflow.log	30/04/2023 11:14 AM	Text Document	54 KB
nextflow.log.1	30/04/2023 11:11 AM	1 File	55 KB
nextflow.log.2	30/04/2023 11:00 AM	2 File	62 KB
nextflow.log.3	30/04/2023 10:30 AM	3 File	63 KB
nextflow.log.4	30/04/2023 10:05 AM	4 File	55 KB
nextflow.log.5	30/04/2023 10:02 AM	5 File	60 KB
nextflow.log.6	30/04/2023 9:25 AM	6 File	86 KB
nextflow.stdout	30/04/2023 11:14 AM	STDOUT File	71 KB
params.json	30/04/2023 11:12 AM	JSON File	1 KB
proc.pid	30/04/2023 11:12 AM	PID File	1 KB

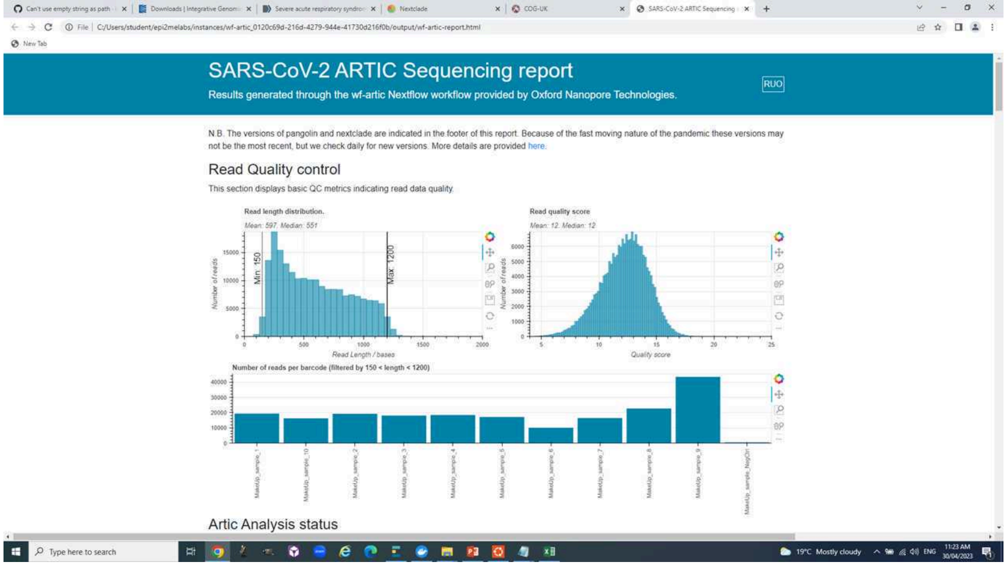
19 This "output" folder contains LOADS of data. The most important is the "all_consensus.fasta" and the "wf-artic-report.html". We will also make use of some of the other files.



Name	Date modified	Type	Size
execution	30/04/2023 11:14 AM	File folder	
all_consensus.fasta	30/04/2023 10:57 AM	FASTA File	327 KB
all_depth.txt	30/04/2023 11:14 AM	Text Document	1,237 KB
all_variants.vcf.gz	30/04/2023 11:13 AM	GZ File	5 KB
all_variants.vcf.gz.tbi	30/04/2023 11:13 AM	TBI File	1 KB
artic.json	30/04/2023 11:14 AM	JSON File	3,115 KB
consensus_status.txt	30/04/2023 10:57 AM	Text Document	1 KB
lineage_report.csv	30/04/2023 11:14 AM	Microsoft Excel C...	4 KB
MakeUp_sample_1.pass.named.vcf.gz	30/04/2023 9:25 AM	GZ File	3 KB
MakeUp_sample_1.pass.named.vcf.gz.tbi	30/04/2023 9:25 AM	TBI File	1 KB
MakeUp_sample_1.primertrimmed.rg.sor...	30/04/2023 9:24 AM	BAM File	7,095 KB
MakeUp_sample_1.primertrimmed.rg.sor...	30/04/2023 9:24 AM	BAI File	1 KB
MakeUp_sample_2.pass.named.vcf.gz	30/04/2023 9:25 AM	GZ File	3 KB
MakeUp_sample_2.pass.named.vcf.gz.tbi	30/04/2023 9:25 AM	TBI File	1 KB
MakeUp_sample_2.primertrimmed.rg.sor...	30/04/2023 9:24 AM	BAM File	6,223 KB
MakeUp_sample_2.primertrimmed.rg.sor...	30/04/2023 9:24 AM	BAI File	1 KB
MakeUp_sample_3.pass.named.vcf.gz	30/04/2023 9:25 AM	GZ File	3 KB
MakeUp_sample_3.pass.named.vcf.gz.tbi	30/04/2023 9:25 AM	TBI File	1 KB
MakeUp_sample_3.primertrimmed.rg.sor...	30/04/2023 9:24 AM	BAM File	6,587 KB
MakeUp_sample_3.primertrimmed.rg.sor...	30/04/2023 9:24 AM	BAI File	1 KB
MakeUp_sample_4.pass.named.vcf.gz	30/04/2023 9:25 AM	GZ File	3 KB
MakeUp_sample_4.pass.named.vcf.gz.tbi	30/04/2023 9:25 AM	TBI File	1 KB
MakeUp_sample_4.primertrimmed.rg.sor...	30/04/2023 9:24 AM	BAM File	7,508 KB
MakeUp_sample_4.primertrimmed.rg.sor...	30/04/2023 9:24 AM	BAI File	1 KB
MakeUp_sample_5.pass.named.vcf.gz	30/04/2023 9:25 AM	GZ File	3 KB
MakeUp_sample_5.pass.named.vcf.gz.tbi	30/04/2023 10:02 AM	TBI File	1 KB
MakeUp_sample_5.primertrimmed.rg.sor...	30/04/2023 9:24 AM	BAM File	7,035 KB
MakeUp_sample_5.primertrimmed.rg.sor...	30/04/2023 9:24 AM	BAI File	1 KB
MakeUp_sample_6.pass.named.vcf.gz	30/04/2023 10:00 AM	GZ File	2 KB
MakeUp_sample_6.pass.named.vcf.gz.tbi	30/04/2023 10:00 AM	TBI File	1 KB
MakeUp_sample_6.primertrimmed.rg.sor...	30/04/2023 9:24 AM	BAM File	3,215 KB
MakeUp_sample_6.primertrimmed.rg.sor...	30/04/2023 9:24 AM	BAI File	1 KB
MakeUp_sample_7.pass.named.vcf.gz	30/04/2023 10:00 AM	GZ File	3 KB
MakeUp_sample_7.pass.named.vcf.gz.tbi	30/04/2023 10:02 AM	TBI File	1 KB
MakeUp_sample_7.primertrimmed.rg.sor...	30/04/2023 9:25 AM	BAM File	5,988 KB
MakeUp_sample_7.primertrimmed.rg.sor...	30/04/2023 9:25 AM	BAI File	1 KB
MakeUp_sample_8.pass.named.vcf.gz	30/04/2023 10:02 AM	GZ File	3 KB
MakeUp_sample_8.pass.named.vcf.gz.tbi	30/04/2023 10:02 AM	TBI File	1 KB
MakeUp_sample_8.primertrimmed.rg.sor...	30/04/2023 9:25 AM	BAM File	7,603 KB
MakeUp_sample_8.primertrimmed.rg.sor...	30/04/2023 9:25 AM	BAI File	1 KB
MakeUp_sample_9.pass.named.vcf.gz	30/04/2023 10:02 AM	GZ File	3 KB
MakeUp_sample_9.pass.named.vcf.gz.tbi	30/04/2023 10:02 AM	TBI File	1 KB
MakeUp_sample_9.primertrimmed.rg.sor...	30/04/2023 9:25 AM	BAM File	9,473 KB

Now rename both files with unique and sensible identifiers for your group and share it with all the RG members.

20 The "all_consensus.fasta" contains your consensus genomes for each sample and the "wf-artic-report.html" looks like this. We will go over both in more detail during the prac.



SARS-CoV-2 ARTIC Sequencing report

Results generated through the wf-artic Nextflow workflow provided by Oxford Nanopore Technologies.

N.B. The versions of pangolin and nextclade are indicated in the footer of this report. Because of the fast moving nature of the pandemic these versions may not be the most recent, but we check daily for new versions. More details are provided here.

Read Quality control

This section displays basic QC metrics indicating read data quality.

Read length distribution.

Read quality score

Number of reads per barcode (filtered by 150 < length < 1200)

Artic Analysis status

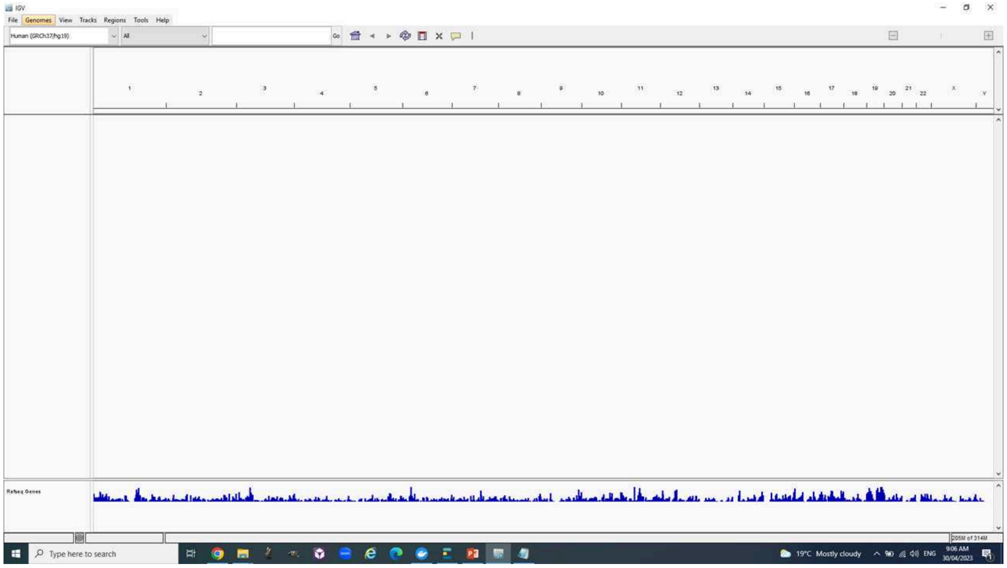
Snap shot of html summary report

21 We will use the "all_consensus.fasta" for most downstream analyses. It contains the consensus fasta sequences of your genomes.

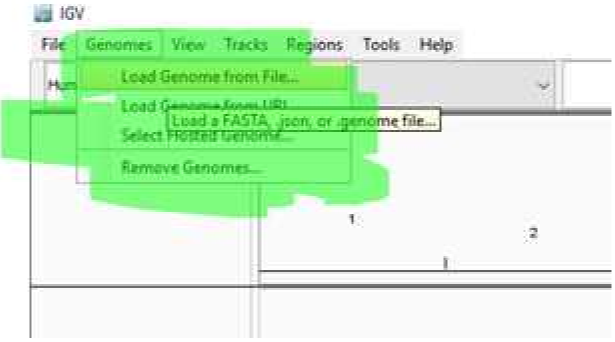
Second II: Visualization whole genome alignment and principle steps 2-7 of section I

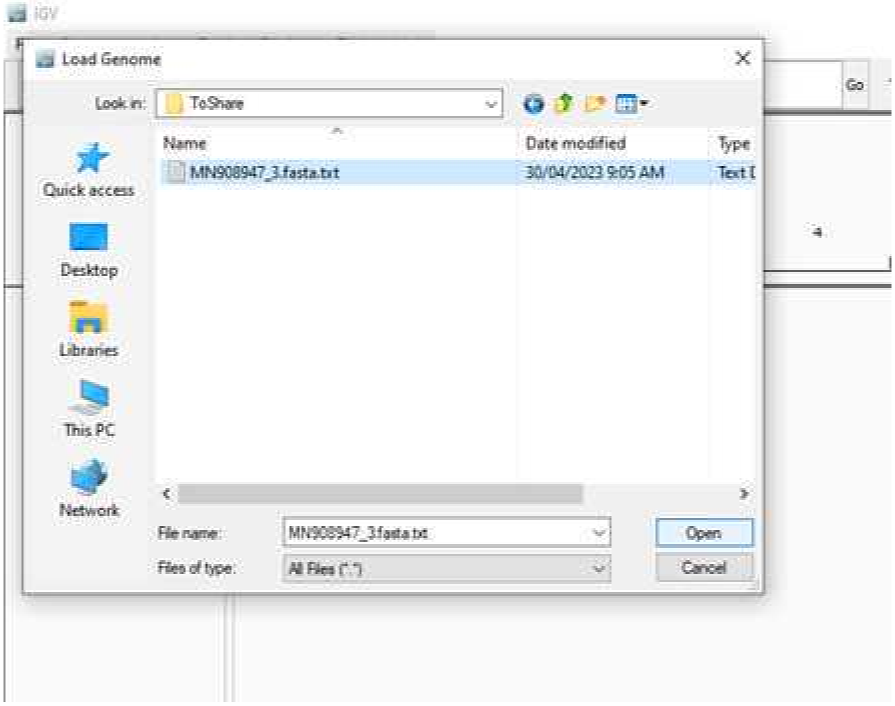
22 You will use IGV to look at one of your read sets mapped against the reference genome to get a better understanding of the "raw" data.

23 Open IGV.


IGV
File Genomes View Tracks Regions Tools Help
Human (GRCh37/hg19)
All
Go
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
RefSeq Genes
Type here to search
19°C Mostly cloudy
9:06 AM
30/04/2023

24 Select "Genomes" > "Load Genomes from File.." and add the "MN908947_3.fasta" reference shared **here** (ANU only).





Look in: ToShare

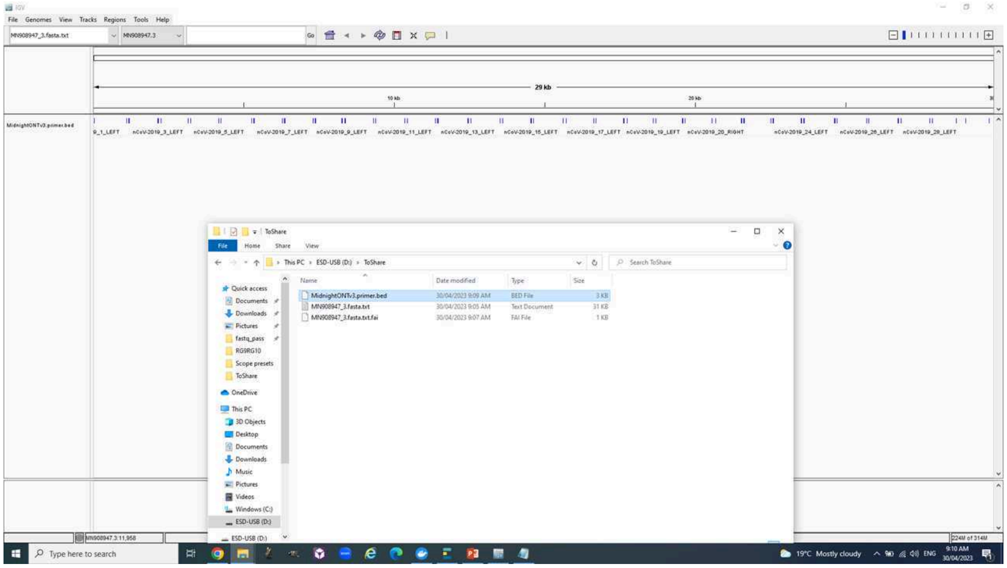
Name	Date modified	Type
MN908947_3.fasta.txt	30/04/2023 9:05 AM	Text I

File name: MN908947_3fasta.txt

Files of type: All Files (*.*)

The file can end in ".txt" or better ".fasta"

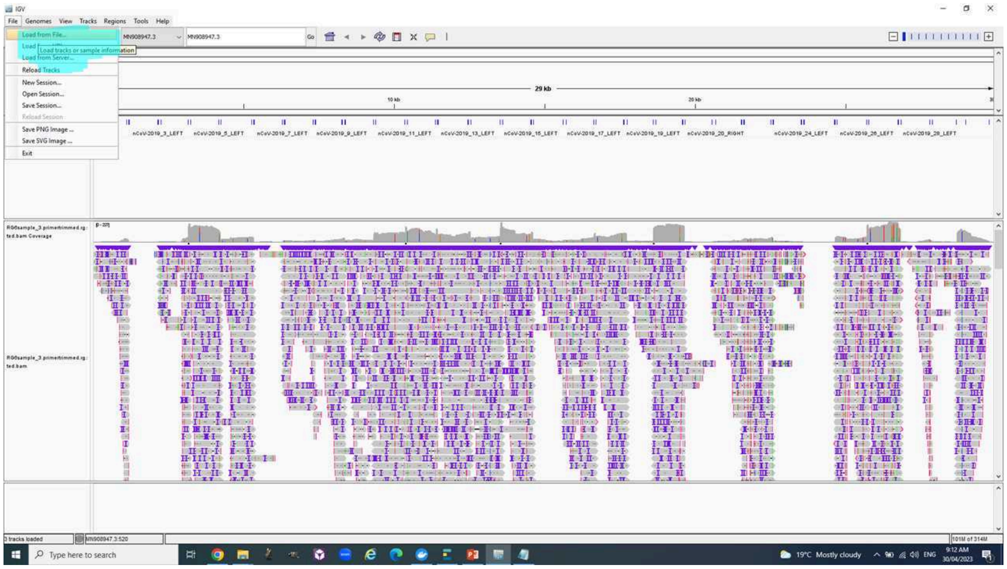
25 Drag and drop the "MidnightONTv3.primer.bed" into IGV to illustrate where your primer sequences bind.



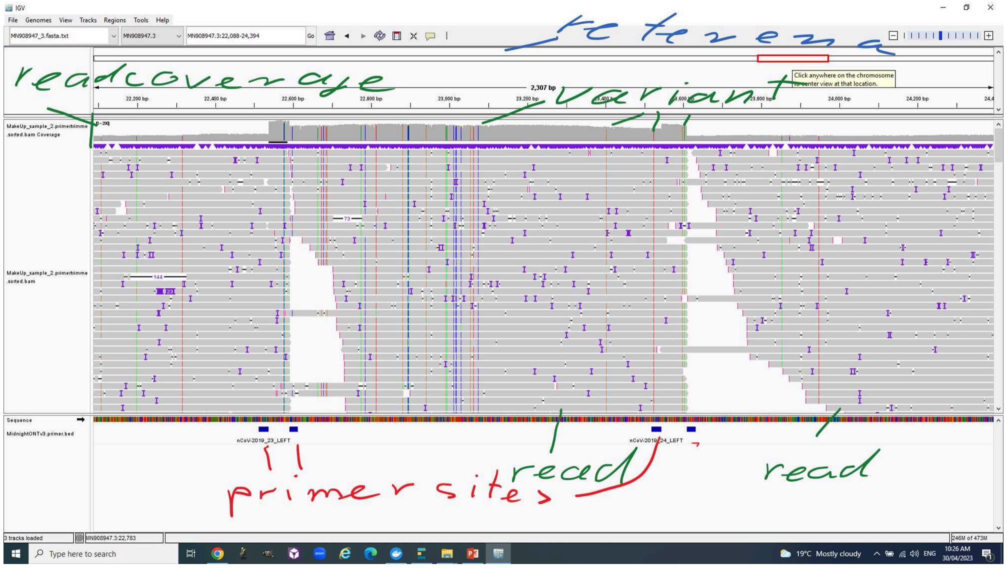
This PC › ESD-USB (D:) › ToShare

Name	Date modified	Type	Size
MidnightONTv3.primer.bed	30/04/2023 9:09 AM	BED File	3 KB
MN908947_3.fasta.txt	30/04/2023 9:05 AM	Text Document	31 KB
MN908947_3.fasta.txt.fai	30/04/2023 9:07 AM	FAI File	1 KB

26 Load one of the "XXX.primertrimmed.rg.sorted.bam" files from the "output" folder of Epi2Me **here** (ANU only) or from your own dataset.



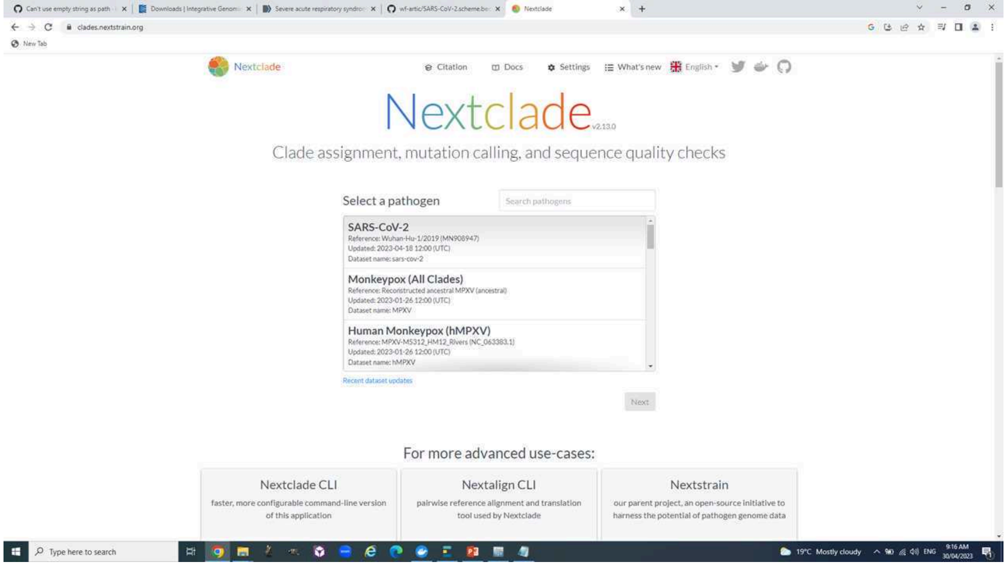
27 Now you see the alignments of all the reads as pileup against the reference, the read coverage and variants.



You can zoom into certain regions e.g. the S-gene 21563-25384 to see how the reads are different to the reference via the different colored bars in the grey area labelled with variants in the picture above. There is loads more to explore. Ask if you have any questions.

Section III: Nextclade for lineage calling, QC measures, amino acid substitutions identification and so much more

28 We will use Nexclade **https://clades.nextstrain.org/** to get some quality assessment metrics, lineages calling, number of mutations, and much more.



Nextclade v2.13.0

Clade assignment, mutation calling, and sequence quality checks

Select a pathogen

Search pathogens

SARS-CoV-2
Reference: Wuhan-Hu-1/2019 (MN908947)
Updated: 2023-04-18 12:00 (UTC)
Dataset name: sars-cov-2

Monkeypox (All Clades)
Reference: Reconstructed ancestral MPXV (ancestral)
Updated: 2023-01-26 12:00 (UTC)
Dataset name: MPXV

Human Monkeypox (hMPXV)
Reference: MPXV-M5312_HM12_Rivers (NC_063383.1)
Updated: 2023-01-26 12:00 (UTC)
Dataset name: hMPXV

Recent dataset updates

Next

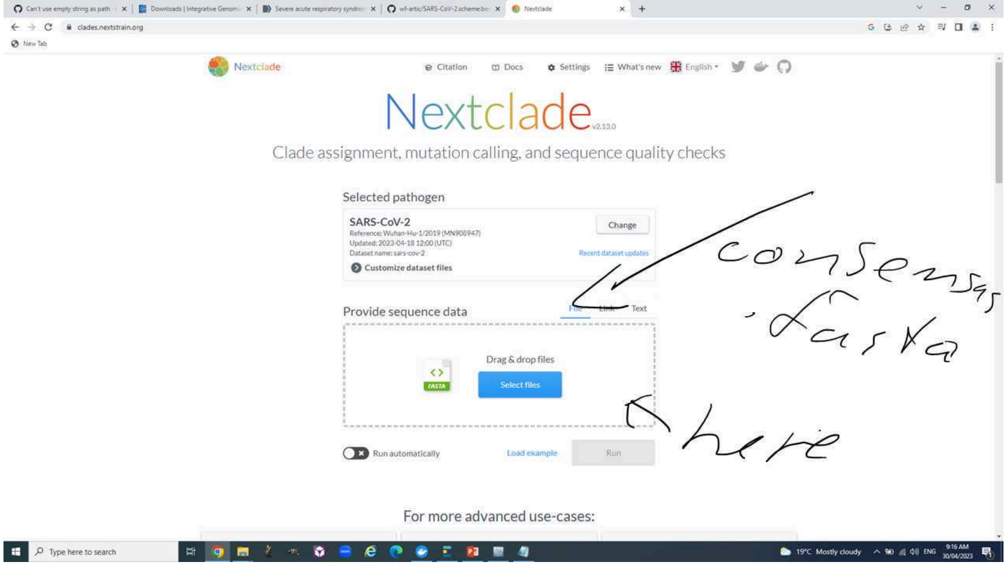
For more advanced use-cases:

Nextclade CLI
faster, more configurable command-line version of this application

Nextalign CLI
pairwise reference alignment and translation tool used by Nextclade

Nextstrain
our parent project, an open-source initiative to harness the potential of pathogen genome data

29 Open the webpage **https://clades.nextstrain.org/**, select SARS-CoV-2 and drag and drop your consensus sequences into it.



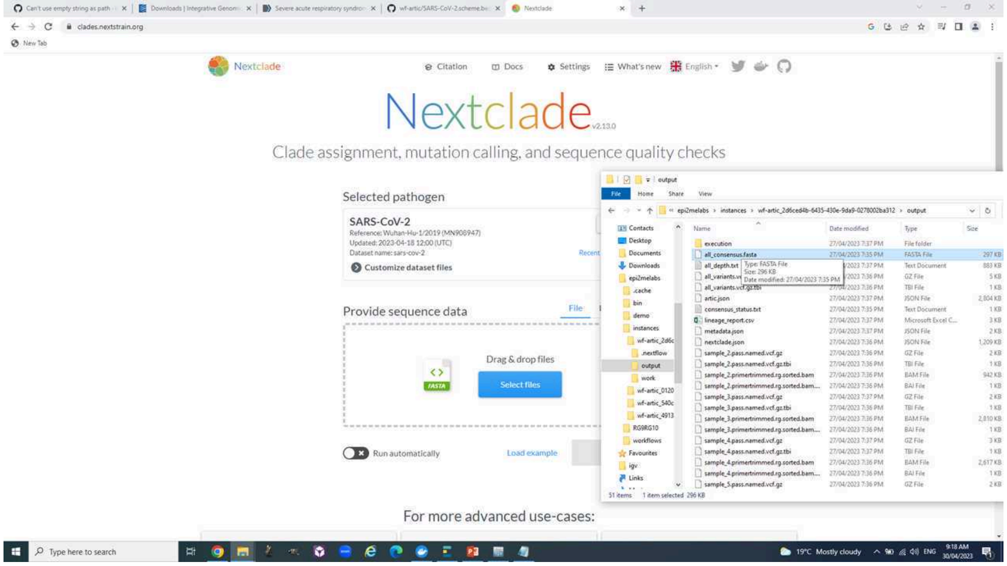
Selected pathogen

SARS-CoV-2
Reference: Wuhan-Hu-1/2019 (MN908947)
Updated: 2023-04-18 12:00 (UTC)
Dataset name: sars-cov-2

Provide sequence data

For more advanced use-cases:

Pretty please!



Selected pathogen

SARS-CoV-2
Reference: Wuhan-Hu-1/2019 (MN908947)
Updated: 2023-04-18 12:00 (UTC)
Dataset name: sars-cov-2

Provide sequence data

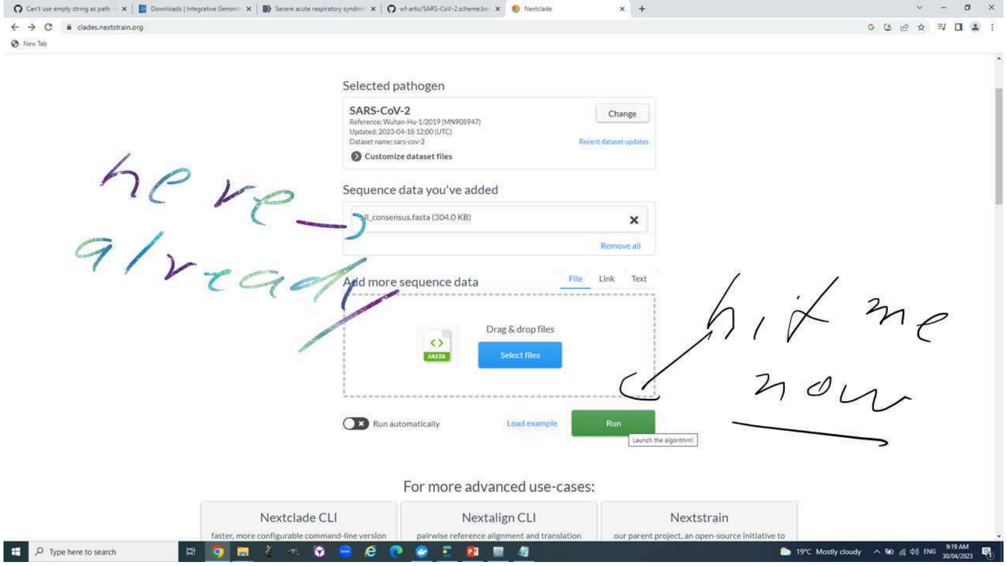
For more advanced use-cases:

epi2melabs › instances › wf-artic_2d6ced4b-6435-430e-9da9-0278002ba312 › output

Name	Date modified	Type	Size
execution	27/04/2023 7:37 PM	File folder	
all_consensus.fasta	27/04/2023 7:35 PM	FASTA File	297 KB
all_depth.txt	/2023 7:37 PM	Text Document	883 KB
all_variants.v	/2023 7:36 PM	GZ File	5 KB
all_variants.vcf.gz.tbi	27/04/2023 7:36 PM	TBI File	1 KB
artic.json	27/04/2023 7:37 PM	JSON File	2,804 KB
consensus_status.txt	27/04/2023 7:35 PM	Text Document	1 KB
lineage_report.csv	27/04/2023 7:37 PM	Microsoft Excel C...	3 KB
metadata.json	27/04/2023 7:37 PM	JSON File	2 KB
nextclade.json	27/04/2023 7:36 PM	JSON File	1,209 KB
sample_2.pass.named.vcf.gz	27/04/2023 7:36 PM	GZ File	2 KB
sample_2.pass.named.vcf.gz.tbi	27/04/2023 7:36 PM	TBI File	1 KB
sample_2.primertrimmed.rg.sorted.bam	27/04/2023 7:36 PM	BAM File	942 KB
sample_2.primertrimmed.rg.sorted.bam....	27/04/2023 7:36 PM	BAI File	1 KB
sample_3.pass.named.vcf.gz	27/04/2023 7:37 PM	GZ File	2 KB
sample_3.pass.named.vcf.gz.tbi	27/04/2023 7:36 PM	TBI File	1 KB
sample_3.primertrimmed.rg.sorted.bam	27/04/2023 7:36 PM	BAM File	2,810 KB
sample_3.primertrimmed.rg.sorted.bam....	27/04/2023 7:36 PM	BAI File	1 KB
sample_4.pass.named.vcf.gz	27/04/2023 7:37 PM	GZ File	3 KB
sample_4.pass.named.vcf.gz.tbi	27/04/2023 7:37 PM	TBI File	1 KB
sample_4.primertrimmed.rg.sorted.bam	27/04/2023 7:36 PM	BAM File	2,617 KB
sample_4.primertrimmed.rg.sorted.bam....	27/04/2023 7:36 PM	BAI File	1 KB
sample_5.pass.named.vcf.gz	27/04/2023 7:36 PM	GZ File	2 KB

Type: FASTA File
Size: 296 KB
Date modified: 27/04/2023 7:35 PM

51 items 1 item selected 296 KB



Selected pathogen

SARS-CoV-2
Reference: Wuhan-Hu-1/2019 (MN908947)
Updated: 2023-04-18 12:00 (UTC)
Dataset name: sars-cov-2

Sequence data you've added

all_consensus.fasta (304.0 KB)

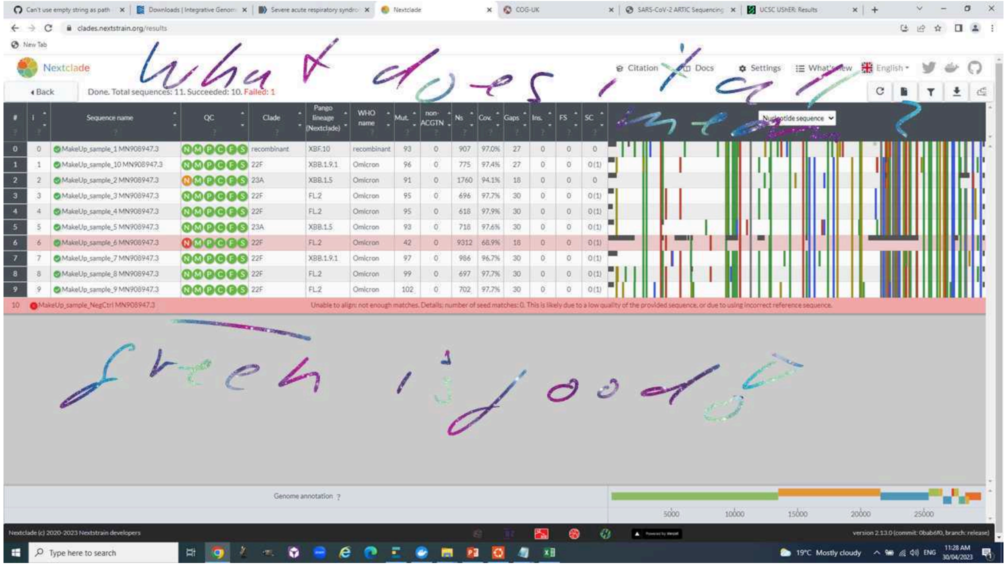
Add more sequence data

Run automatically

For more advanced use-cases:

Thanks!

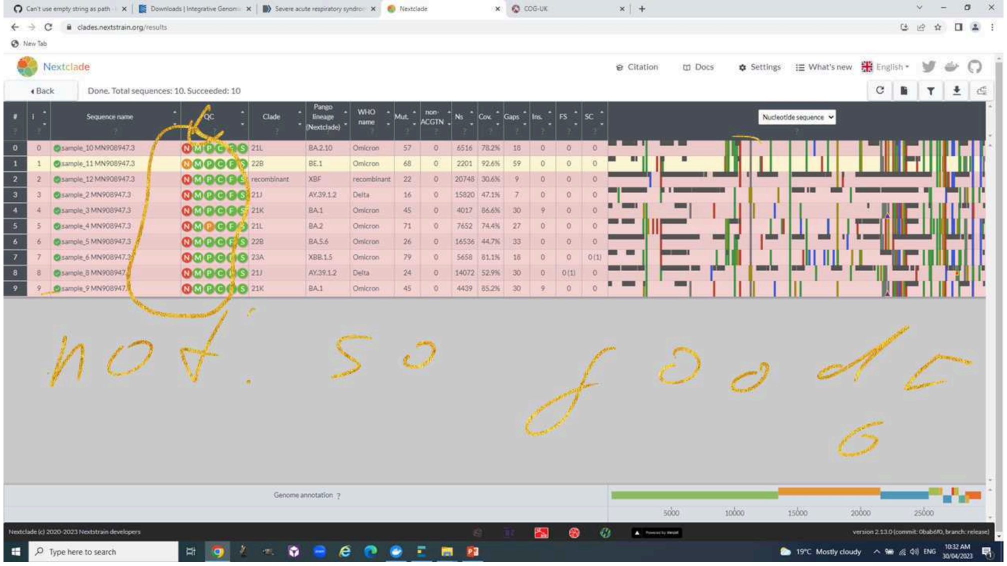
30 This is an example of a good Midnight consensus sequence set.



Done. Total sequences: 11. Succeeded: 10. Failed: 1

#	i	Sequence name	QC	Clade	Pango lineage (Nextclade)	WHO name	Mut.	non-ACGTN	Ns	Cov.	Gaps	Ins.	FS	SC
0	0	MakeUp_sample_1 MN908947.3	N M P C F S	recombinant	XBF.10	recombinant	93	0	907	97.0%	27	0	0	0
1	1	MakeUp_sample_10 MN908947.3	N M P C F S	22F	XBB.1.9.1	Omicron	96	0	775	97.4%	27	0	0	0 (1)
2	2	MakeUp_sample_2 MN908947.3	N M P C F S	23A	XBB.1.5	Omicron	91	0	1760	94.1%	18	0	0	0
3	3	MakeUp_sample_3 MN908947.3	N M P C F S	22F	FL.2	Omicron	95	0	696	97.7%	30	0	0	0 (1)
4	4	MakeUp_sample_4 MN908947.3	N M P C F S	22F	FL.2	Omicron	95	0	618	97.9%	30	0	0	0 (1)
5	5	MakeUp_sample_5 MN908947.3	N M P C F S	23A	XBB.1.5	Omicron	93	0	718	97.6%	30	0	0	0 (1)
6	6	MakeUp_sample_6 MN908947.3	N M P C F S	22F	FL.2	Omicron	42	0	9312	68.9%	18	0	0	0 (1)
7	7	MakeUp_sample_7 MN908947.3	N M P C F S	22F	XBB.1.9.1	Omicron	97	0	986	96.7%	30	0	0	0 (1)
8	8	MakeUp_sample_8 MN908947.3	N M P C F S	22F	FL.2	Omicron	99	0	697	97.7%	30	0	0	0 (1)
9	9	MakeUp_sample_9 MN908947.3	N M P C F S	22F	FL.2	Omicron	102	0	702	97.7%	30	0	0	0 (1)
10		MakeUp_sample_NegCtrl MN908947.3	Unable to align: not enough matches. Details: number of seed matches: 0. This is likely due to a low quality of the provided sequence, or due to using incorrect reference sequence.											

31 This is an example of a not so good Midnight consensus sequence set.



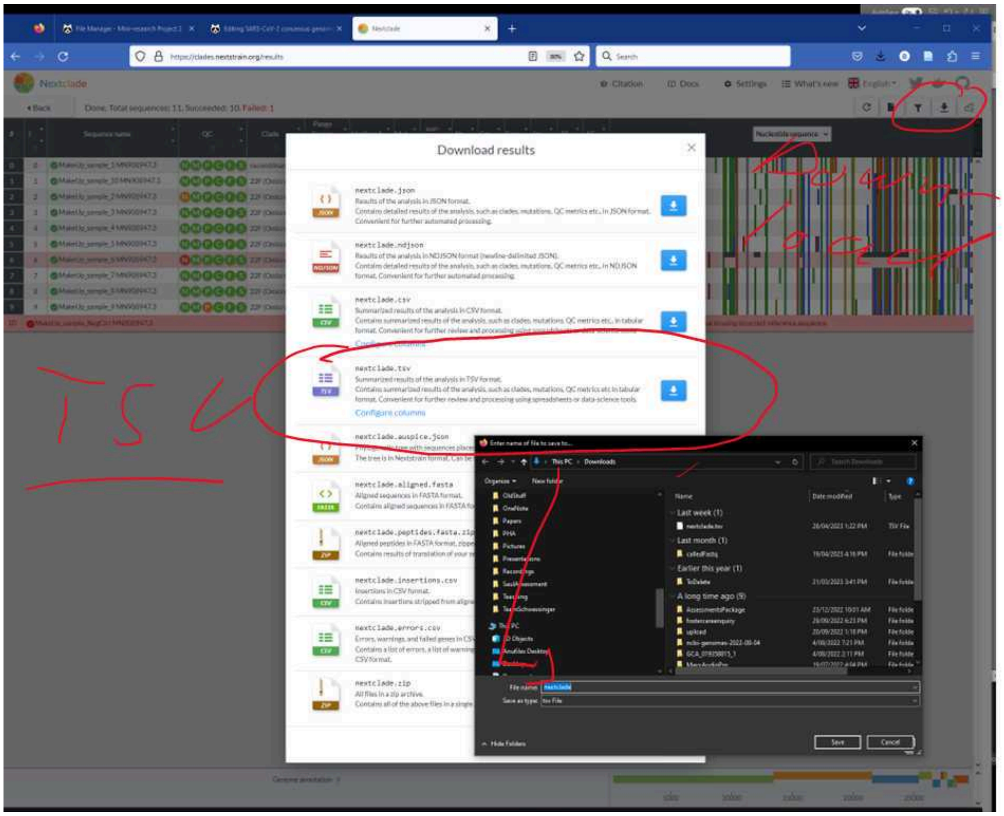
‹ Back Done. Total sequences: 10. Succeeded: 10

#	I	Sequence name	QC	Clade	Pango lineage (Nextclade)	WHO name	Mut.	non-ACGTN	Ns	Cov.	Gaps	Ins.	FS	SC
0	0	sample_10 MN908947.3	N M P C F S	21L	BA.2.10	Omicron	57	0	6516	78.2%	18	0	0	0
1	1	sample_11 MN908947.3	N M P C F S	22B	BE.1	Omicron	68	0	2201	92.6%	59	0	0	0
2	2	sample_12 MN908947.3	N M P C F S	recombinant	XBF	recombinant	22	0	20748	30.6%	9	0	0	0
3	3	sample_2 MN908947.3	N M P C F S	21J	AY.39.1.2	Delta	16	0	15820	47.1%	7	0	0	0
4	4	sample_3 MN908947.3	N M P C F S	21K	BA.1	Omicron	45	0	4017	86.6%	30	9	0	0
5	5	sample_4 MN908947.3	N M P C F S	21L	BA.2	Omicron	71	0	7652	74.4%	27	0	0	0
6	6	sample_5 MN908947.3	N M P C F S	22B	BA.5.6	Omicron	26	0	16536	44.7%	33	0	0	0
7	7	sample_6 MN908947.3	N M P C F S	23A	XBB.1.5	Omicron	79	0	5658	81.1%	18	0	0	0 (1)
8	8	sample_8 MN908947	N M P C F S	21J	AY.39.1.2	Delta	24	0	14072	52.9%	30	0	0 (1)	0
9	9	sample_9 MN908947	N M P C F S	21K	BA.1	Omicron	45	0	4439	85.2%	30	9	0	0

32 Explore the results page to get to the following:

1. Pango lineage assignment.
2. Genome coverage.
3. Number of mutations relative to reference.
4. Number of Ns.
5. Gaps.
6. Changes to the amino acid sequence of the S protein.

33 Make sure to download your samples via the Download dialog in the "TSV" (tab separated values) format. You can import this into Excel later.



Download results

nextclade.json
Results of the analysis in JSON format.
Contains detailed results of the analysis, such as clades, mutations, QC metrics etc., in JSON format. Convenient for further automated processing.

nextclade.ndjson
Results of the analysis in NDJSON format (newline-delimited JSON).
Contains detailed results of the analysis, such as clades, mutations, QC metrics etc., in NDJSON format. Convenient for further automated processing.

nextclade.csv
Summarized results of the analysis in CSV format.
Contains summarized results of the analysis, such as clades, mutations, QC metrics etc., in tabular format. Convenient for further review and processing

nextclade.tsv
Summarized results of the analysis in TSV format.
Contains summarized results of the analysis, such as clades, mutations, QC metrics etc in tabular format. Convenient for further review and processing using spreadsheets or data-science tools.
Configure columns

nextclade.auspice.json
The tree is in Nextstrain format. Can be

nextclade.aligned.fasta
Aligned sequences in FASTA format.
Contains aligned sequences in FASTA fo

nextclade.peptides.fasta.zip
Aligned peptides in FASTA format, zippe
Contains results of translation of your

nextclade.insertions.csv
Insertions in CSV format.
Contains insertions stripped from aligr

nextclade.errors.csv
Errors, warnings, and failed genes in CS
Contains a list of errors, a list of warnin
CSV format.

nextclade.zip
All files in a zip archive.
Contains all of the above files in a single

Enter name of file to save to...

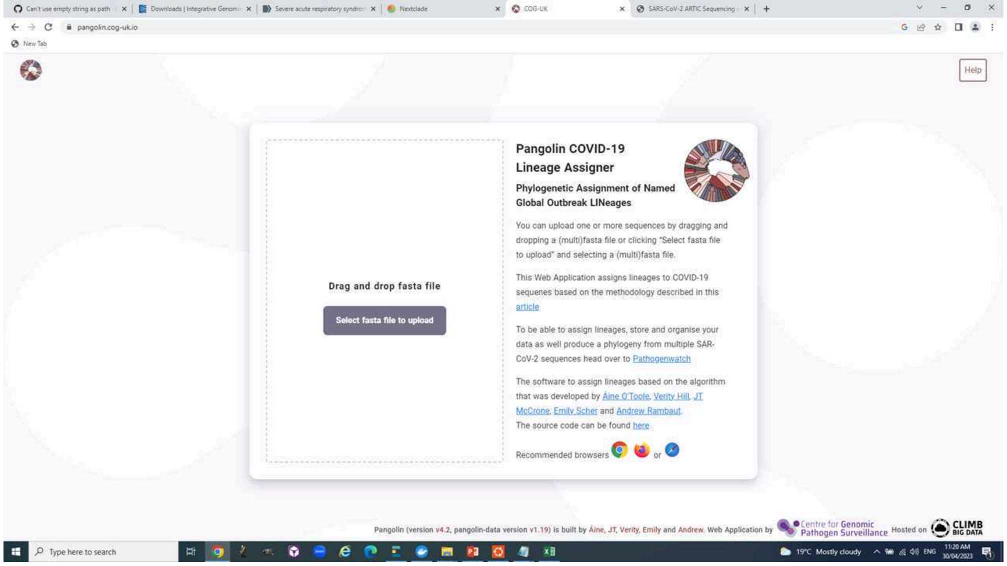
File name: nextclade
Save as type: tsv File

Ask question during the prac for anything that is unclear.

Section IV: Pangolin lineage assignment

34 Let's assign the latest lineage designation with Pangolin.

Navigate to **https://pangolin.cog-uk.io/**.



Help

Drag and drop fasta file

Select fasta file to upload

Pangolin COVID-19 Lineage Assigner

Phylogenetic Assignment of Named Global Outbreak LINeages

You can upload one or more sequences by dragging and dropping a (multi)fasta file or clicking "Select fasta file to upload" and selecting a (multi)fasta file.

This Web Application assigns lineages to COVID-19 sequenes based on the methodology described in this article

To be able to assign lineages, store and organise your data as well produce a phylogeny from multiple SAR-CoV-2 sequences head over to Pathogenwatch

The software to assign lineages based on the algorithm that was developed by Áine O'Toole, Verity Hill, JT McCrone, Emily Scher and Andrew Rambaut.
The source code can be found here

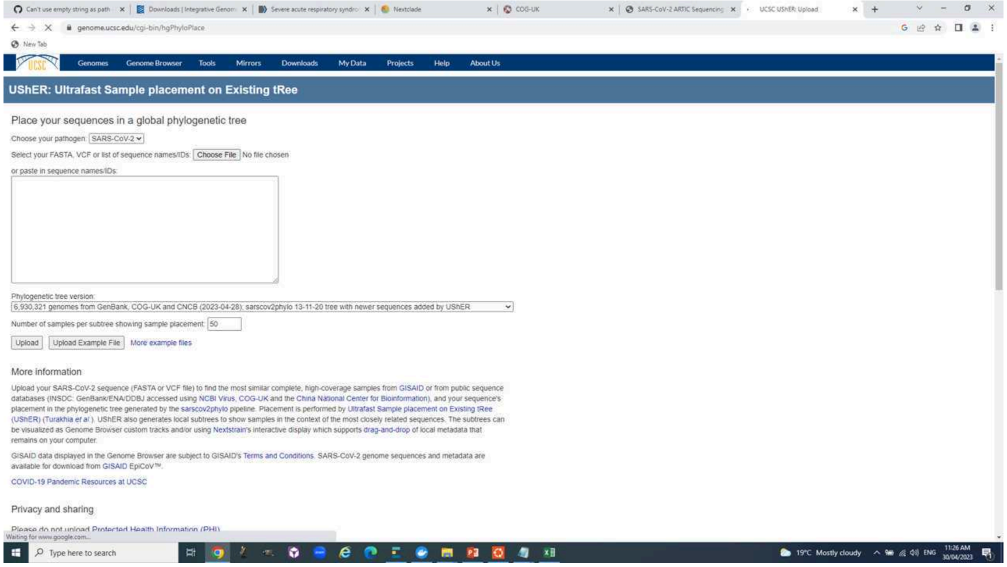
Recommended browsers or

Pangolin (version v4.2, pangolin-data version v1.19) is built by Áine, JT, Verity, Emily and Andrew. Web Application by Centre for Genomic Pathogen Surveillance Hosted on CLIMB BIG DATA

Drag and drop your sequences, start analysis, and hope for the best.

Step V: Use Usher to place your samples into a global context and see its closest public neighbor.

35 Navigate to https://genome.ucsc.edu/cgi-bin/hgPhyloPlace.



genome.ucsc.edu/cgi-bin/hgPhyloPlace

Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About Us

UShER: Ultrafast Sample placement on Existing tRee

Place your sequences in a global phylogenetic tree

Choose your pathogen: SARS-CoV-2

Select your FASTA, VCF or list of sequence names/IDs: Choose File No file chosen

or paste in sequence names/IDs:

Phylogenetic tree version:
6,930,321 genomes from GenBank, COG-UK and CNCB (2023-04-28); sarscov2phylo 13-11-20 tree with newer sequences added by UShER

Number of samples per subtree showing sample placement: 50

Upload Upload Example File More example files

More information

Upload your SARS-CoV-2 sequence (FASTA or VCF file) to find the most similar complete, high-coverage samples from GISAID or from public sequence databases (INSDC: GenBank/ENA/DDBJ accessed using NCBI Virus, COG-UK and the China National Center for Bioinformation), and your sequence's placement in the phylogenetic tree generated by the sarscov2phylo pipeline. Placement is performed by Ultrafast Sample placement on Existing tRee (UShER) (Turakhia et al.). UShER also generates local subtrees to show samples in the context of the most closely related sequences. The subtrees can be visualized as Genome Browser custom tracks and/or using Nextstrain's interactive display which supports drag-and-drop of local metadata that remains on your computer.

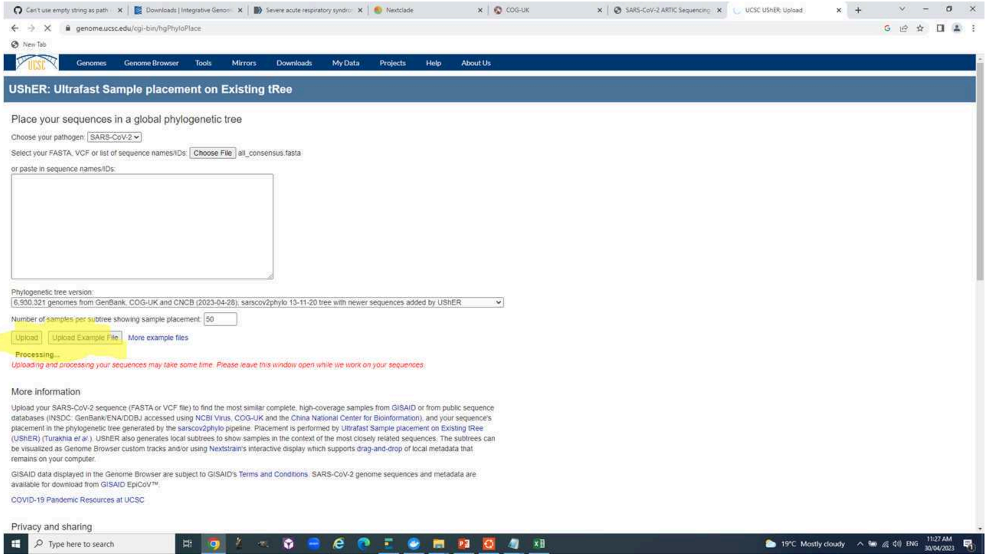
GISAID data displayed in the Genome Browser are subject to GISAID's Terms and Conditions. SARS-CoV-2 genome sequences and metadata are available for download from GISAID EpiCoV™.

COVID-19 Pandemic Resources at UCSC

Privacy and sharing

Please do not upload Protected Health Information (PHI)

36 Upload your sequences



UShER: Ultrafast Sample placement on Existing tRee

Place your sequences in a global phylogenetic tree

Choose your pathogen: SARS-CoV-2

Select your FASTA, VCF or list of sequence names/IDs: Choose File all_consensus.fasta

or paste in sequence names/IDs:

Phylogenetic tree version:
6,930,321 genomes from GenBank, COG-UK and CNCB (2023-04-28); sarscov2phylo 13-11-20 tree with newer sequences added by UShER

Number of samples per subtree showing sample placement: 50

Upload | Upload Example File | More example files

Processing...

Uploading and processing your sequences may take some time. Please leave this window open while we work on your sequences.

More information

Upload your SARS-CoV-2 sequence (FASTA or VCF file) to find the most similar complete, high-coverage samples from GISAID or from public sequence databases (INSDC: GenBank/ENA/DDBJ accessed using NCBI Virus, COG-UK and the China National Center for Bioinformation), and your sequence's placement in the phylogenetic tree generated by the sarscov2phylo pipeline. Placement is performed by Ultrafast Sample placement on Existing tRee (UShER) (Turakhia *et al.*). UShER also generates local subtrees to show samples in the context of the most closely related sequences. The subtrees can be visualized as Genome Browser custom tracks and/or using Nextstrain's interactive display which supports drag-and-drop of local metadata that remains on your computer.

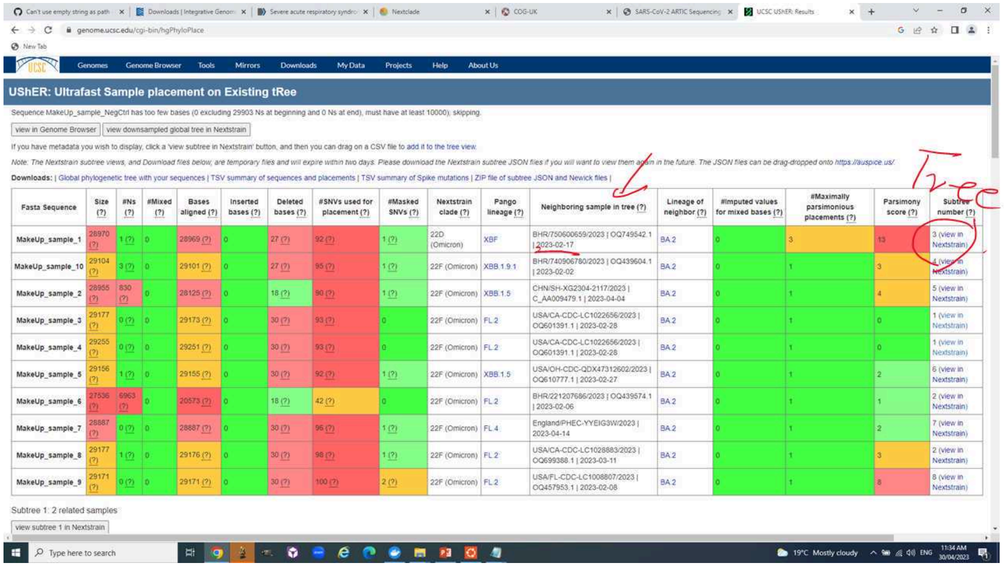
GISAID data displayed in the Genome Browser are subject to GISAID's Terms and Conditions. SARS-CoV-2 genome sequences and metadata are available for download from GISAID EpiCoV™.

COVID-19 Pandemic Resources at UCSC

Privacy and sharing

37 Explore the webpage for results including:

1. Neighboring sample in Tree.
2. That ^^ samples deposition date.
3. Subtree of each of your sequences.



UShER: Ultrafast Sample placement on Existing tRee

Sequence MakeUp_sample_NegCtrl has too few bases (0 excluding 29903 Ns at beginning and 0 Ns at end), must have at least 10000); skipping.

view in Genome Browser | view downsampled global tree in Nextstrain

If you have metadata you wish to display, click a 'view subtree in Nextstrain' button, and then you can drag on a CSV file to add it to the tree view.

Note: The Nextstrain subtree views, and Download files below, are temporary files and will expire within two days. Please download the Nextstrain subtree JSON files if you will want to view them again in the future. The JSON files can be drag-dropped onto https://auspice.us/.

Downloads: | Global phylogenetic tree with your sequences | TSV summary of sequences and placements | TSV summary of Spike mutations | ZIP file of subtree JSON and Newick files |

Fasta Sequence	Size (?)	#Ns (?)	#Mixed (?)	Bases aligned (?)	Inserted bases (?)	Deleted bases (?)	#SNVs used for placement (?)	#Masked SNVs (?)	Nextstrain clade (?)	Pango lineage (?)	Neighboring sample in tree (?)	Lineage of neighbor (?)	#Imputed values for mixed bases (?)	#Maximally parsimonious placements (?)	Parsimony score (?)	Subtree number (?)
MakeUp_sample_1	28970 (?)	1 (?)	0	28969 (?)	0	27 (?)	92 (?)	1 (?)	22D (Omicron)	XBF	BHR/750600659/2023 | OQ749542.1 | 2023-02-17	BA.2	0	3	13	3 (view in Nextstrain)
MakeUp_sample_10	29104 (?)	3 (?)	0	29101 (?)	0	27 (?)	95 (?)	1 (?)	22F (Omicron)	XBB.1.9.1	BHR/740906780/2023 | OQ439604.1 | 2023-02-02	BA.2	0	1	3	4 (view in Nextstrain)
MakeUp_sample_2	28955 (?)	830 (?)	0	28125 (?)	0	18 (?)	90 (?)	1 (?)	22F (Omicron)	XBB.1.5	CHN/SH-XG2304-2117/2023 | C_AA009479.1 | 2023-04-04	BA.2	0	1	4	5 (view in Nextstrain)
MakeUp_sample_3	29177 (?)	0 (?)	0	29173 (?)	0	30 (?)	93 (?)	0	22F (Omicron)	FL.2	USA/CA-CDC-LC1022656/2023 | OQ601391.1 | 2023-02-28	BA.2	0	1	0	1 (view in Nextstrain)
MakeUp_sample_4	29255 (?)	0 (?)	0	29251 (?)	0	30 (?)	93 (?)	0	22F (Omicron)	FL.2	USA/CA-CDC-LC1022656/2023 | OQ601391.1 | 2023-02-28	BA.2	0	1	0	1 (view in Nextstrain)
MakeUp_sample_5	29156 (?)	1 (?)	0	29155 (?)	0	30 (?)	92 (?)	1 (?)	22F (Omicron)	XBB.1.5	USA/OH-CDC-QDX47312602/2023 | OQ610777.1 | 2023-02-27	BA.2	0	1	2	6 (view in Nextstrain)
MakeUp_sample_6	27536 (?)	6963 (?)	0	20573 (?)	0	18 (?)	42 (?)	0	22F (Omicron)	FL.2	BHR/221207686/2023 | OQ439574.1 | 2023-02-06	BA.2	0	1	1	2 (view in Nextstrain)
MakeUp_sample_7	28887 (?)	0 (?)	0	28887 (?)	0	30 (?)	96 (?)	1 (?)	22F (Omicron)	FL.4	England/PHEC-YYEIG3W/2023 | 2023-04-14	BA.2	0	1	2	7 (view in Nextstrain)
MakeUp_sample_8	29177 (?)	1 (?)	0	29176 (?)	0	30 (?)	98 (?)	1 (?)	22F (Omicron)	FL.2	USA/CA-CDC-LC1028883/2023 | OQ699388.1 | 2023-03-11	BA.2	0	1	3	2 (view in Nextstrain)
MakeUp_sample_9	29171 (?)	0 (?)	0	29171 (?)	0	30 (?)	100 (?)	2 (?)	22F (Omicron)	FL.2	USA/FL-CDC-LC1008807/2023 | OQ457953.1 | 2023-02-08	BA.2	0	1	8	8 (view in Nextstrain)

Subtree 1: 2 related samples

view subtree 1 in Nextstrain

38 Ask questions in during the prac.