

# Geneious workflow for SARS-CoV-2 genome assembly - Download and installation instructions

<b>1 Introduction</b>	<b>2</b>
<b>2 Files, Software &amp; Set-up</b>	<b>2</b>
2.1 Software requirements	2
2.2 Setting up the workflows	3
2.2.1 Download the files from GitHub	3
2.2.2 Import data to Geneious	3
2.2.3 Import the workflow to Geneious	4
2.3 Add the plugins for the “SARS-CoV-2_Assembly_WrapperPlugins.geneiousWorkflow”	10
2.3.1 Import the plugins to Geneious	10
2.3.2 Add the plugins to the workflow	11
<b>3 Running the workflow in Geneious</b>	<b>13</b>
3.1 Run the workflow	13
3.2 Output data	13
References	15

# 1 Introduction

This is a workflow for [Geneious Prime](#) that automates the bioinformatics analysis of SARS-CoV-2 genomes from Nanopore data. The workflow runs on FASTQ files and outputs SARS-CoV-2 consensus genome sequences. For instructions on how to perform basecalling on Nanopore data and generate FASTQ files see [Nanopore's instructions](#).

There are two workflows available:

- “SARS-CoV-2\_Assembly\_Basic” performs basic analysis including trimming, mapping, consensus generation and annotation.
- “SARS-CoV-2\_Assembly\_WrapperPlugins” also includes pango-lineage assignment, using pangolin (O’Toole *et al.*, 2021) and clade assignment, mutation calling and quality checks using Nextclade (Hadfield *et al.*, 2018) as wrapper plugins in the workflow.

Both workflows can be used with the ARTIC primers (Quick, 2020) or midnight primers (Freed *et al.*, 2020). The workflows are designed to work on Windows, Linux and Mac OS.

## 2 Files, Software & Set-up

### 2.1 Software requirements

Software requirements are presented in Table 1. Some software is required regardless of workflow, while some software is only required for “SARS-CoV-2\_Assembly\_WrapperPlugins”.

*Table 1. A list of software required for a functioning workflow.*

Software	Link	Workflow
Geneious Prime	<a href="#">geneious.com</a>	Both
Python	<a href="#">python.org</a>	SARS-CoV-2_Assembly_WrapperPlugins
Docker Desktop	<a href="#">docker.com</a>	SARS-CoV-2_Assembly_WrapperPlugins
Nextclade CLI	<a href="#">Nextclade-cli</a>	SARS-CoV-2_Assembly_WrapperPlugins

## 2.2 Setting up the workflows

### 2.2.1 Download the files from GitHub

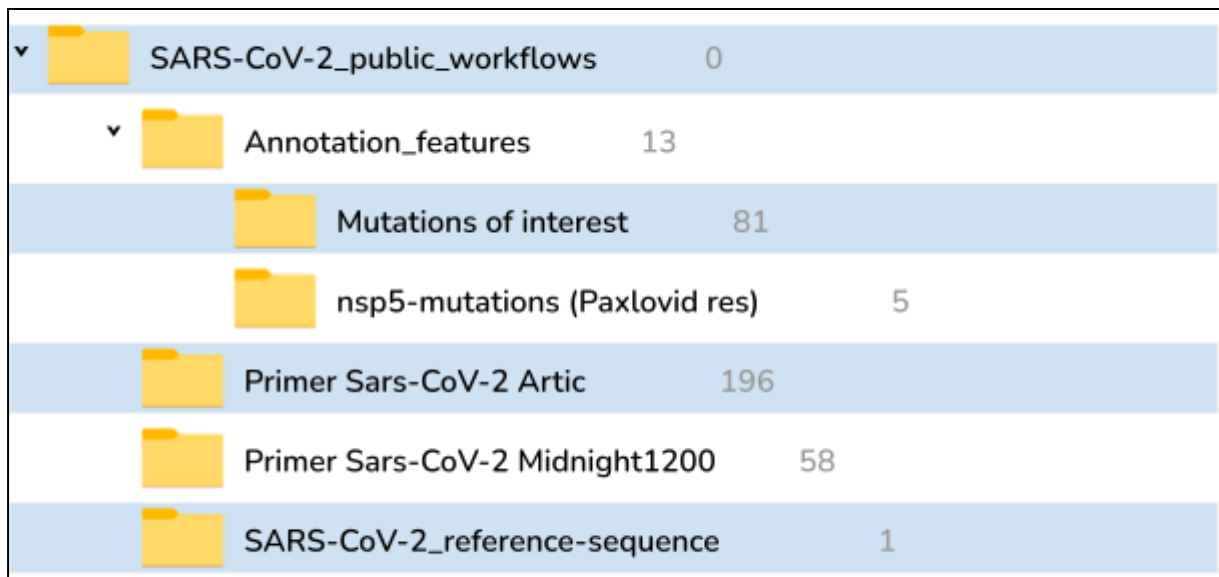
Download the GitHub repository by pressing '<> Code' and 'Download ZIP'

Or use git clone:

```
git clone https://github.com/clinical-genomics-uppsala/Geneious_SARS-CoV-2.git
```

### 2.2.2 Import data to Geneious

The workflow is designed to work with the folder structure presented in Figure 1. At the left side panel in Geneious Prime, select **Local** (or select a subfolder of your choice). Press **Add > Import Folder...** and select the folder "SARS-CoV-2\_public\_workflows" in the folder downloaded from GitHub.



▼	SARS-CoV-2_public_workflows	0
▼	Annotation_features	13
	Mutations of interest	81
	nsp5-mutations (Paxlovid res)	5
	Primer Sars-CoV-2 Artic	196
	Primer Sars-CoV-2 Midnight1200	58
	SARS-CoV-2_reference-sequence	1

Figure 1. An illustration on what the folder structure will look like when imported into Geneious.

### 2.2.3 Import the workflow to Geneious

**Select Tools > \*Workflows > Manage Workflows...**

Click on **Import** and then select the *.geneiousWorkflow* file that you want to use, "SARS-CoV-2\_Assembly\_Basic" or "SARS-CoV-2\_Assembly\_WrapperPlugins". Once imported, the workflow should appear at the bottom of the list of existing workflows.



## Trim using BBDuk

\*Note: If your version of Geneious Prime is older than 2023.0, you may have to install BBDuk by downloading the [plugin from Geneious's website](#), and add it to your Geneious Prime software.

Double-click “Trim using BBDuk”. In the window “Edit Trim using BBDuk”, select **Choose...** under **Trim Adapters > Adapters**. Select the folder “Primer\_Sars-CoV-2\_Artic” or “Primer\_Sars-CoV-2\_Midnight1200” depending on which set of primers you plan to use. Select **Apply & Close**.

Options to expose to user when workflow is run

- ☒ Expose no options
- ☐ Expose all options
- ☐ Expose some options

Optionally label exposed options as:  ☐ Access exposed options via button 

Expose:  With Alternative Label:

All Operation Options (those not exposed to workflow user and default values for options that are exposed)

*BBDuk Adapter/Quality Trimming Version 38.84 by Brian Bushnell*

☒ Trim Adapters

Adapters:   

Trim:

Kmer Length:

Maximum Substitutions:

Maximum Substitutions + INDELS:

☐ Trim partial adapters from ends with kmer length:

☒ Trim Low Quality

Trim:

Minimum Quality:

☐ Trim adapters based on paired read overhangs

Minimum Overlap:

☒ Discard Short Reads

Minimum Length:  bp

All Operation Options (those not exposed to workflow user and default values for options that are exposed)

☐ Trim Low Complexity (Entropy)

Minimum Entropy:

Entropy Window Size:

Entropy Kmer Size:

☐ Keep original order (slower, but ensures results are reproducible)

Maximum memory to use:  MB

Custom BBDuk Options:  

Note: The full command line and output from BBDuk is available from the 'Info' tab of the results

\*Note: If you plan to use both sets of primers and don't want to manually select the primers set everytime you switch between them, it is possible to create two workflows where each workflow has selected one set of primers each. One way to do this is to go to **Manage Workflows**, select the imported workflow and click **Copy**. You can edit the name of the copied workflow and choose a different set of primers for this workflow.

## Align/Assemble -> Map to Reference

The alignment step uses Minimap2. If it is not already installed, go to Tools -> Plugins and under "Available plugins", the plugin named Minimap should be available for installation. It is also possible to download and install Minimap into Geneious Prime at [geneious.com/plugins/minimap2](https://geneious.com/plugins/minimap2).

Double-click "Align/Assemble -> Map to Reference". Ensure that the reference sequence has been selected. If it says *automatic*, select **Choose...** and add the reference sequence *NC\_045512.fasta*. Ensure that under "Method", the "Mapper:" has Minimap2 selected. Select "Data Type". Select **Apply & Close**.

Options to expose to user when workflow is run

- ☒ Expose no options
- ☐ Expose all options
- ☐ Expose some options

Optionally label exposed options as:  ☐ Access exposed options via button ?

Expose:  With Alternative Label:  + -

All Operation Options (those not exposed to workflow user and default values for options that are exposed)

Data

- ☒ Dissolve contigs and re-assemble
- Reference Sequence:   ?
- ☐ Assemble by:  part of name, separated by
- ☒ Assemble each sequence list separately

Method

- Mapper:  ?
- Minimap 2.24*
- Data Type:  ?
- ☐ Override K-Mer Length:  ?

☒ Include Secondary Alignments

Maximum secondary alignments per read:  ?

Minimum secondary to primary alignment score ratio:

Note: Paired reads can be set up or changed using *Sequence > Set Paired Reads*

Trim Before Mapping

- ☐ Use existing trim regions
- ☒ Remove existing trim regions from sequences
- ☐ Re-trim sequences
- ☐ Do not trim (discard trim annotations)

Results

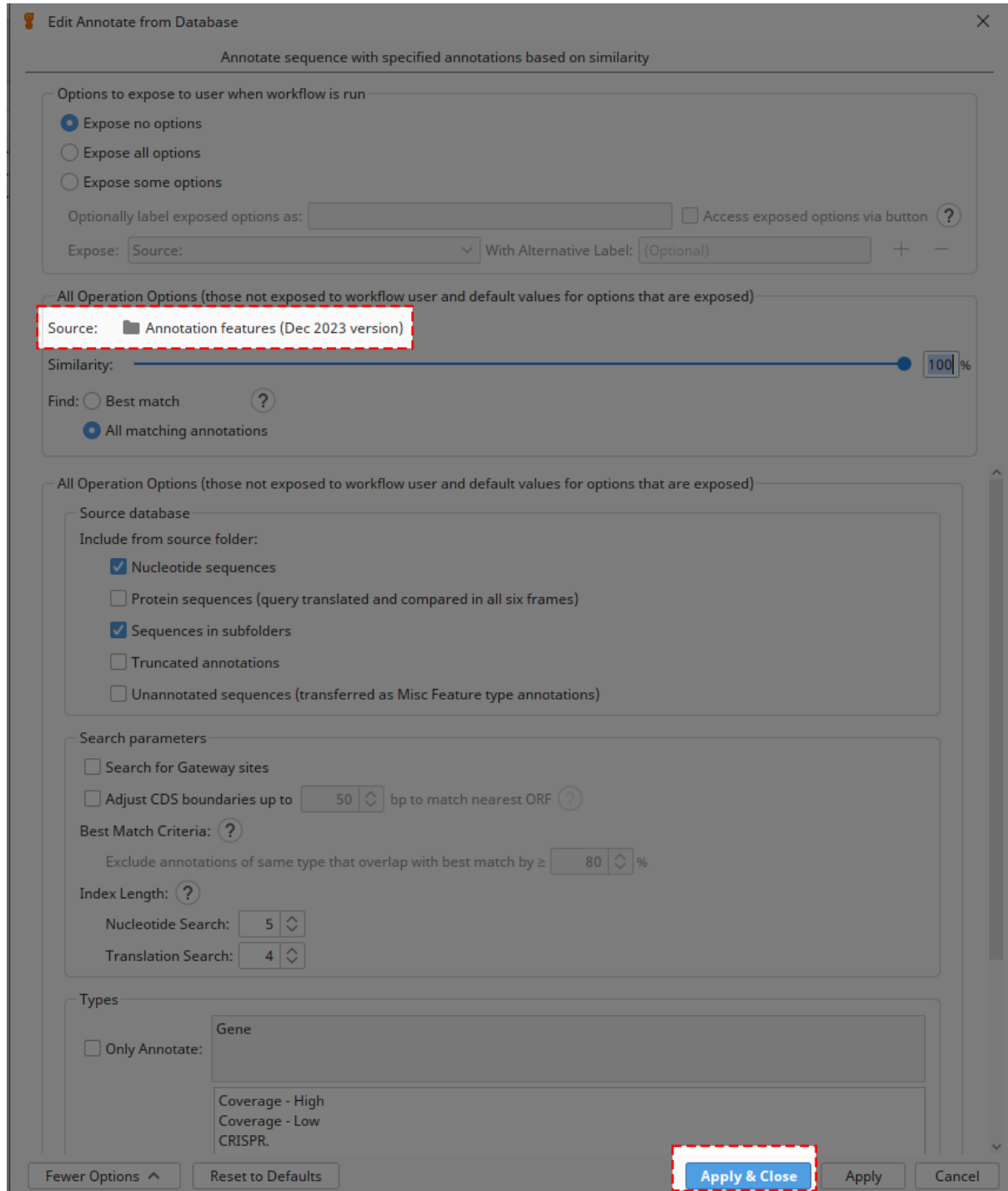
Assembly Name:  ...

- ☐ Save assembly report
- ☐ Save list of unused reads
- ☐ Save list of used reads ☐ Include mates
- ☐ Save in sub-folder
- ☒ Save contigs
- ☐ Save consensus sequences

More Options

## Annotate from Database (1)

Double-click the first out of the two steps called “Annotate from Database”. In the window “Edit Annotate from Database”, click on **Reference Features** (located right of **Source:**). Select the folder *Annotation features*. Select **Apply & Close**.



Edit Annotate from Database

Annotate sequence with specified annotations based on similarity

Options to expose to user when workflow is run

☒ Expose no options  
☐ Expose all options  
☐ Expose some options

Optionally label exposed options as:  ☐ Access exposed options via button ?

Expose:  With Alternative Label:  + -

All Operation Options (those not exposed to workflow user and default values for options that are exposed)

Source:

Similarity:  100%

Find: ☐ Best match ?  
☒ All matching annotations

All Operation Options (those not exposed to workflow user and default values for options that are exposed)

Source database

Include from source folder:

☒ Nucleotide sequences  
☐ Protein sequences (query translated and compared in all six frames)  
☒ Sequences in subfolders  
☐ Truncated annotations  
☐ Unannotated sequences (transferred as Misc Feature type annotations)

Search parameters

☐ Search for Gateway sites  
☐ Adjust CDS boundaries up to  bp to match nearest ORF ?

Best Match Criteria: ?

Exclude annotations of same type that overlap with best match by ≥  %

Index Length: ?

Nucleotide Search:   
Translation Search:

Types

☐ Only Annotate:

Gene  
Coverage - High  
Coverage - Low  
CRISPR.

Fewer Options ^ Reset to Defaults **Apply & Close** Apply Cancel

## Annotate from Database (2)

Double-click the second out of the two steps called “Annotate from Database”. In the window “Edit Annotate from Database”, click on **Reference Features** (located right of **Source:**). Select the folder *SARS-CoV-2 reference sequence*. Select **Apply & Close**.

Edit Annotate from Database

Annotate sequence with specified annotations based on similarity

Options to expose to user when workflow is run

☒ Expose no options  
☐ Expose all options  
☐ Expose some options

Optionally label exposed options as:  ☐ Access exposed options via button ?

Expose:  Source:  With Alternative Label:  (Optional) + -

All Operation Options (those not exposed to workflow user and default values for options that are exposed)

Source:  SARS-CoV-2 reference sequence

Similarity:  80 %

Find: ☐ Best match ?  
☒ All matching annotations

All Operation Options (those not exposed to workflow user and default values for options that are exposed)

Source database

Include from source folder:

☒ Nucleotide sequences  
☐ Protein sequences (query translated and compared in all six frames)  
☒ Sequences in subfolders  
☒ Truncated annotations  
☐ Unannotated sequences (transferred as Misc Feature type annotations)

Search parameters

☐ Search for Gateway sites  
☐ Adjust CDS boundaries up to  50  bp to match nearest ORF ?

Best Match Criteria: ?

Exclude annotations of same type that overlap with best match by ≥  75  %

Index Length: ?

Nucleotide Search:  10

Translation Search:  4

Types

☐ Only Annotate:  Gene  
Coverage - High  
Coverage - Low  
CRISPR.

Fewer Options ^ Reset to Defaults **Apply & Close** Apply Cancel

## Use your own annotation features

It is possible to use your own annotation features and select those instead for the annotation step. If you are only using one folder with annotation features, you can delete one annotation step from the workflow (select the step and press “Delete step”).



## Export (last step)

Double click the step “Export” located at the end of the workflow. Select **Browse** next to the option **Export to Folder**. Select a folder on your computer where you want your final results of the workflow to be located. Note that “SARS-CoV-2\_Assembly\_WrapperPlugins” has 3 Export steps to be edited. Select **Apply & Close**.

Batch Export each document to a separate file

Options to expose to user when workflow is run

☒ Expose no options  
☐ Expose all options  
☐ Expose some options

Optionally label exposed options as:  ☐ Access exposed options via button ?

Expose: File Format:  With Alternative Label:  + -

All Operation Options (those not exposed to workflow user and default values for options that are exposed)

File Format: FASTA sequences/alignment Options

Export to Folder: {DEFAULT\_LOCATION} Browse

Extension: .fasta

☐ Export all documents to a single file called: results

☐ If this workflow step fails, silently continue using only successful data sets

Fewer Options ^ Reset to Defaults Apply & Close Apply Cancel

## 2.3 Add the plugins for the “SARS-CoV-2\_Assembly\_WrapperPlugins.geneiousWorkflow”

If you are using the basic workflow (*SARS-CoV-2\_Assembly\_Basic.geneiousWorkflow*), you can skip ahead to section “3 Running the workflow in Geneious”.

If you want to use the workflow called “SARS-CoV-2\_Assembly\_WrapperPlugins.geneiousWorkflow” some additional steps are necessary.

### 2.3.1 Import the plugins to Geneious

The plugins with installation instructions are available at these Github repositories:

- [Geneious pangolin wrapper](#)
- [Geneious Nextclade wrapper](#)

Select the relevant plugins depending on your operating system. Ensure that the plugins are successfully imported into Geneious. They should be visible under **Tools**.

### 2.3.2 Add the plugins to the workflow

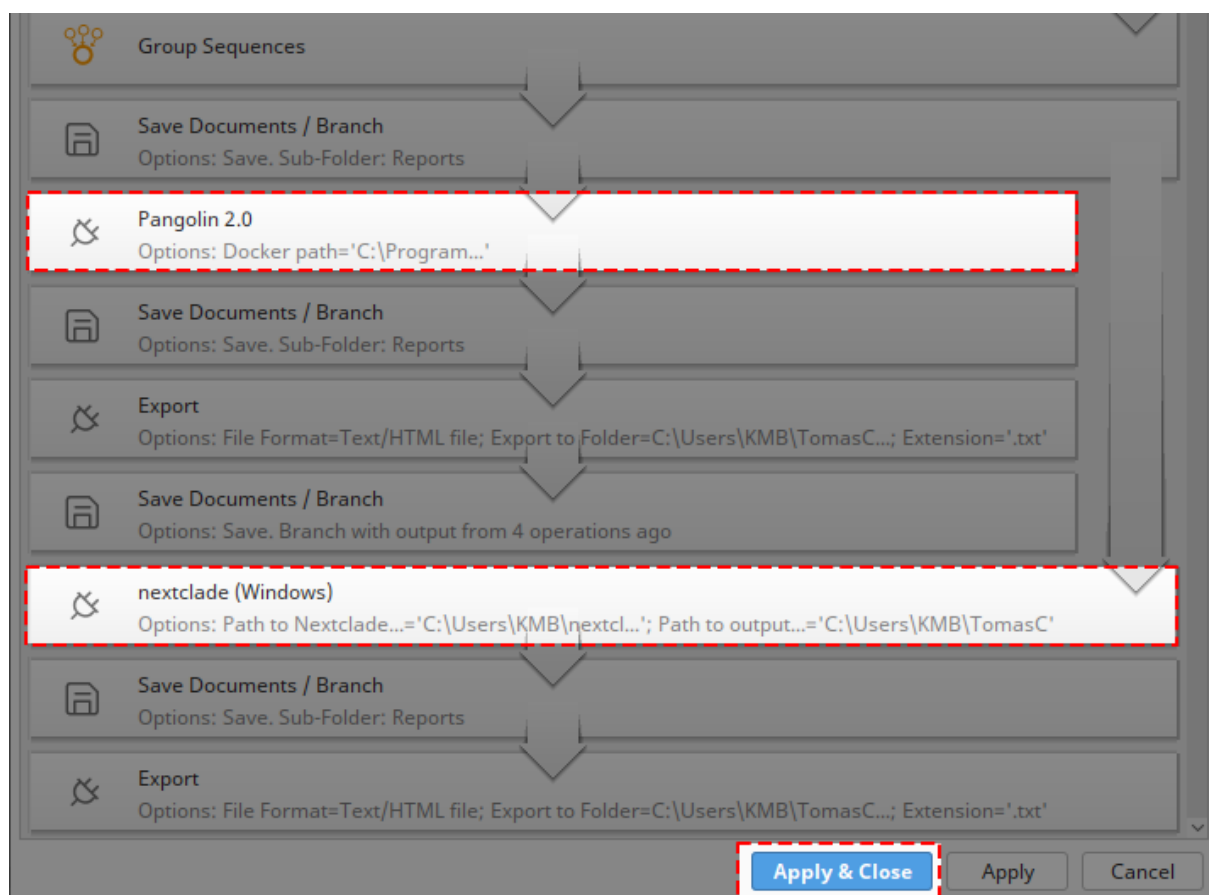
Select **Tools > Workflows > Manage Workflows...**

Double-click the workflow “*SARS-CoV-2\_Assembly\_WrapperPlugins.geneiousWorkflow*”.

Select the step that’s below to where the plugin should be (Figure 3) and press **+ Add Step**.

The plugins can be found if selecting **Add Operation (from x available)...**

The two plugin steps should now be located in the correct part of the workflow (Figure 3). If the steps are positioned in the wrong order, you can drag them to the correct position.



*Figure 3. A screenshot of the Editing mode of “SARS-CoV-2\_Assembly\_WrapperPlugins.geneiousWorkflow”. The locations of the plugin steps have been highlighted.*

The plugins have some user options that need to be specified. This is done by double clicking the plugin step and a new box will be displayed.

### For the pangolin plugin:

**Docker path:** The path to docker executable.

Options to expose to user when workflow is run

☒ Expose no options  
☐ Expose all options  
☐ Expose some options

Optionally label exposed options as:  ☐ Access exposed options via button (?)

Expose: options.Docker path With Alternative Label: (Optional) + -

All Operation Options (those not exposed to workflow user and default values for options that are exposed)

Docker path C:\Program Files\Docker\Docker\resources\bin\docker.exe

More Options ▾ Apply & Close Apply Cancel

### For the Nextclade plugin:

**Nextclade path:** The path to where nextclade (or nextclade.exe on Windows) is located on your system.

**Path for output files:** This is a path of your choice, where you'd like the Nextclade output files to be located.

Options to expose to user when workflow is run

☒ Expose no options  
☐ Expose all options  
☐ Expose some options

Optionally label exposed options as:  ☐ Access exposed options via button (?)

Expose: options.Nextclade path With Alternative Label: (Optional) + -

All Operation Options (those not exposed to workflow user and default values for options that are exposed)

Nextclade path /home/ete03/miniconda3/bin/nextclade  
Path for output files /home/ete03/Documents/

More Options ▾ Apply & Close Apply Cancel

### Export (If not already done)

Finally, a folder where to export the reports from the two plugins should be specified. Two positions under the pangolin step, there is a step called “Export”. Double click on it. In the window “Edit Export”, go to the line “Export to Folder” and select **Browse** to specify the output folder of your results. Repeat the same procedure with the step “Export”, located two positions under the Nextclade step.

## 3 Running the workflow in Geneious

We provide a test data set to validate the workflows. They are 18 fastq (Accession PRJNA1048178) files of different SARS-S-CoV-2 lineages. The data is available at [NCBI](#).

### 3.1 Run the workflow

Import the FASTQ file(s) that are to be analysed into Geneious. One way to do this is to click “Add”, followed by “Import Files...” or “Import Folder...” and choose the FASTQ file(s). The imported files should then be visible in the left-hand panel in Geneious.

Select the FASTQ file(s) to analyse. Then proceed to **Tools > Workflows > “name of the workflow”** and press **OK**.

### 3.2 Output data

Once the workflow has completed running, new output files should have been generated.

Subfolders have now been created from where the input FASTQ-files were located:

- Read Mappings
- Consensus
- Reports (only for the wrapper plugins workflow)

The consensus file(s) in fasta-format will also have been exported to the folder selected in the Export step of the workflow.

For the “SARS-CoV-2\_Assembly\_WrapperPlugins.geneiousWorkflow”, there should also be:

- X Sequences Pangolin.txt
- X Sequences Nextclade.txt
- a folder called “nextclade\_analysis\_<today’s date>” containing Nextclade’s output files according to [Nextclade's output files](#).

## References

- Freed, N.E. *et al.* (2020) 'Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding', *Biology Methods & Protocols*, 5(1), p. bpaa014. Available at: <https://doi.org/10.1093/biomethods/bpaa014>.
- Hadfield, J. *et al.* (2018) 'Nextstrain: real-time tracking of pathogen evolution', *Bioinformatics (Oxford, England)*, 34(23), pp. 4121–4123. Available at: <https://doi.org/10.1093/bioinformatics/bty407>.
- O'Toole, Á. *et al.* (2021) 'Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool', *Virus Evolution*, 7(2), p. veab064. Available at: <https://doi.org/10.1093/ve/veab064>.
- Quick, J. (2020) 'nCoV-2019 sequencing protocol v3 (LoCost)'. Available at: <https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye> (Accessed: 29 March 2023).