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

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	geneious.com	Both
Python	python.org	SARS-CoV-2_Assembly_WrapperPlugins
Docker Desktop	docker.com	SARS-CoV-2_Assembly_WrapperPlugins
Nextclade CLI	Nextclade-cli	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.

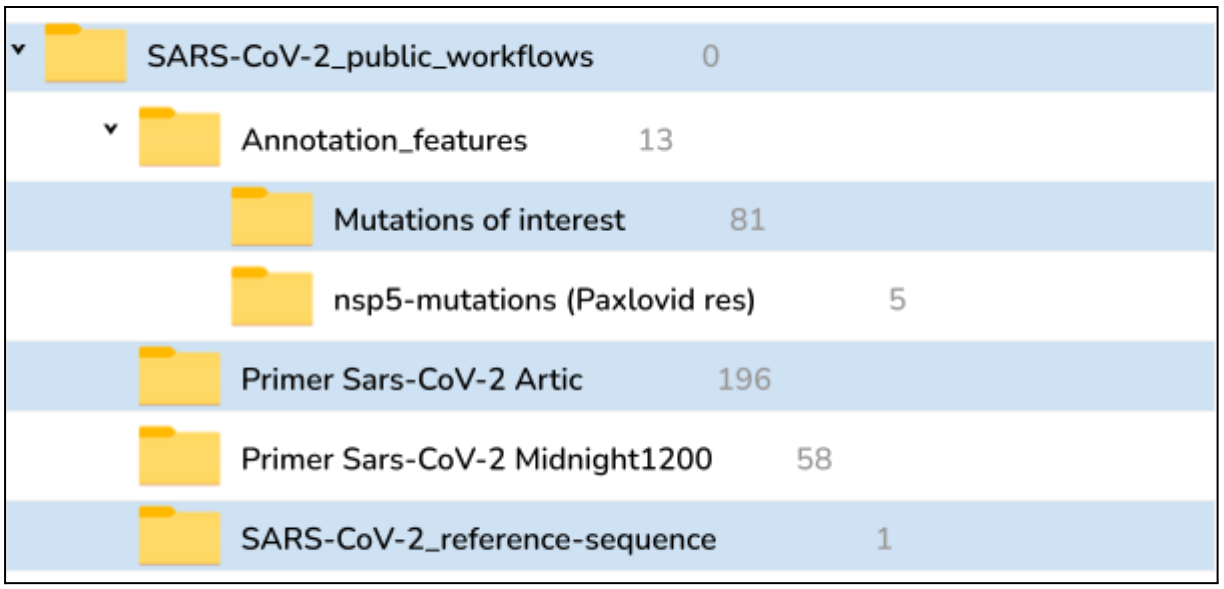


Figure 1. An illustration on what the folder structure will look like when imported into Geneious.

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.

Next ensure that the different steps of the workflow are linked to the files imported in section 2.2.2

Double-click on the workflow and the window “Edit Workflow” should appear (Figure 2). The steps that may need manual editing are:

- Trim using BBDuk
- Align/Assemble -> Map to Reference
- Annotate from Database (1)
- Annotate from Database (2)
- Export (“SARS-CoV-2_Assembly_WrapperPlugins” has 3 Export steps)

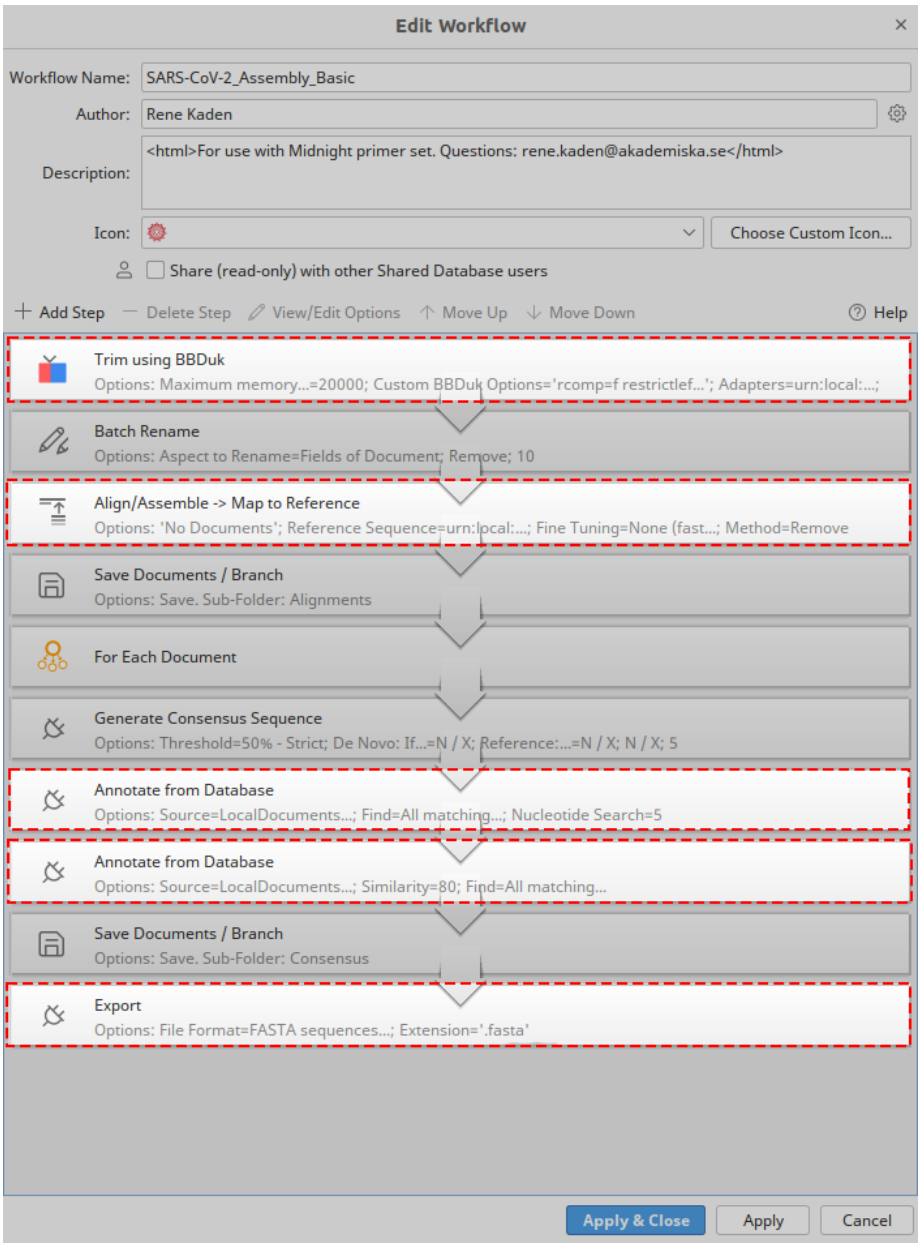
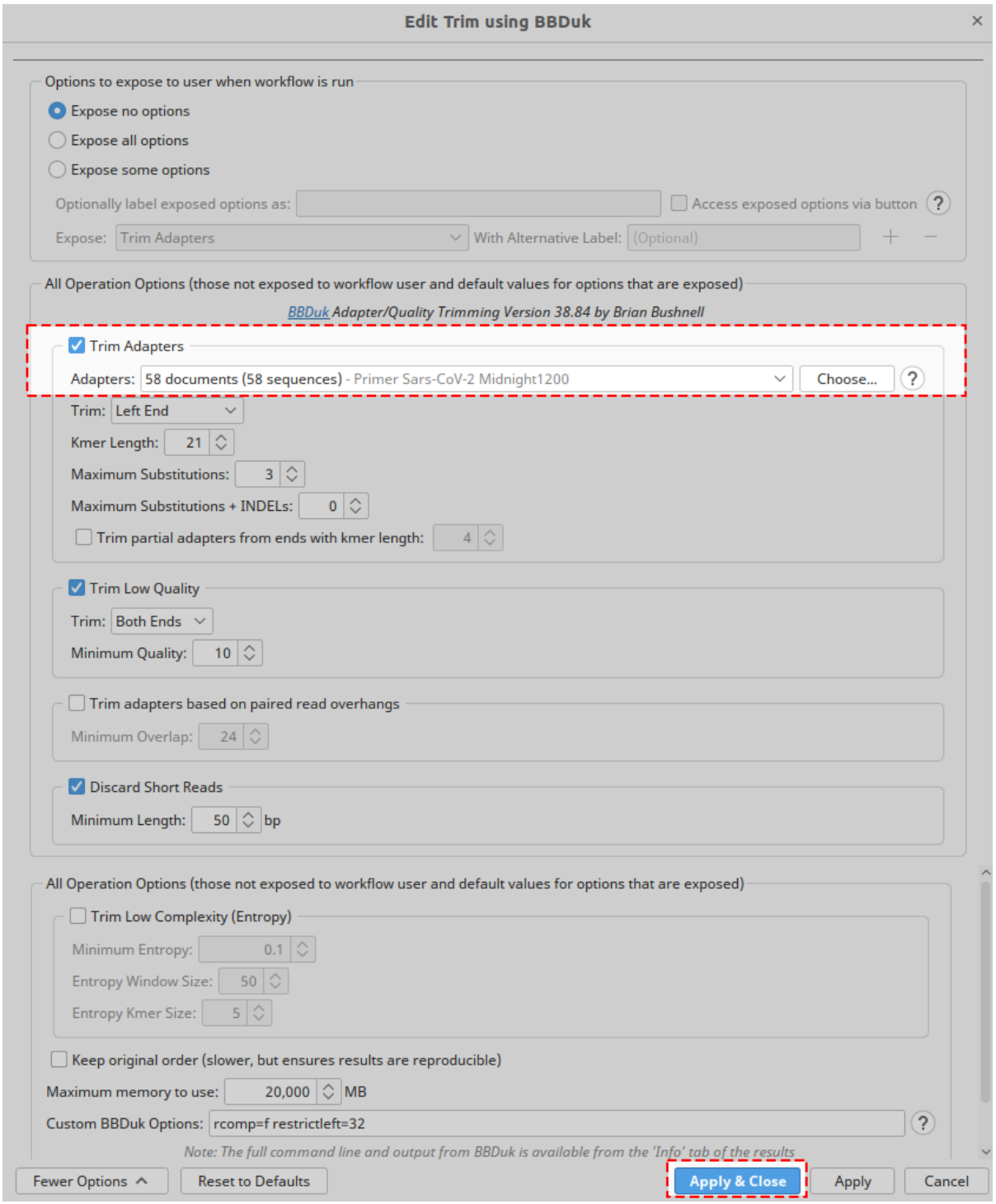


Figure 2. The steps that may require manual modifications have been highlighted. Note that “SARS-CoV-2_Assembly_WrapperPlugins” looks slightly different but contains the same steps that need the same type of modifications.

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**.

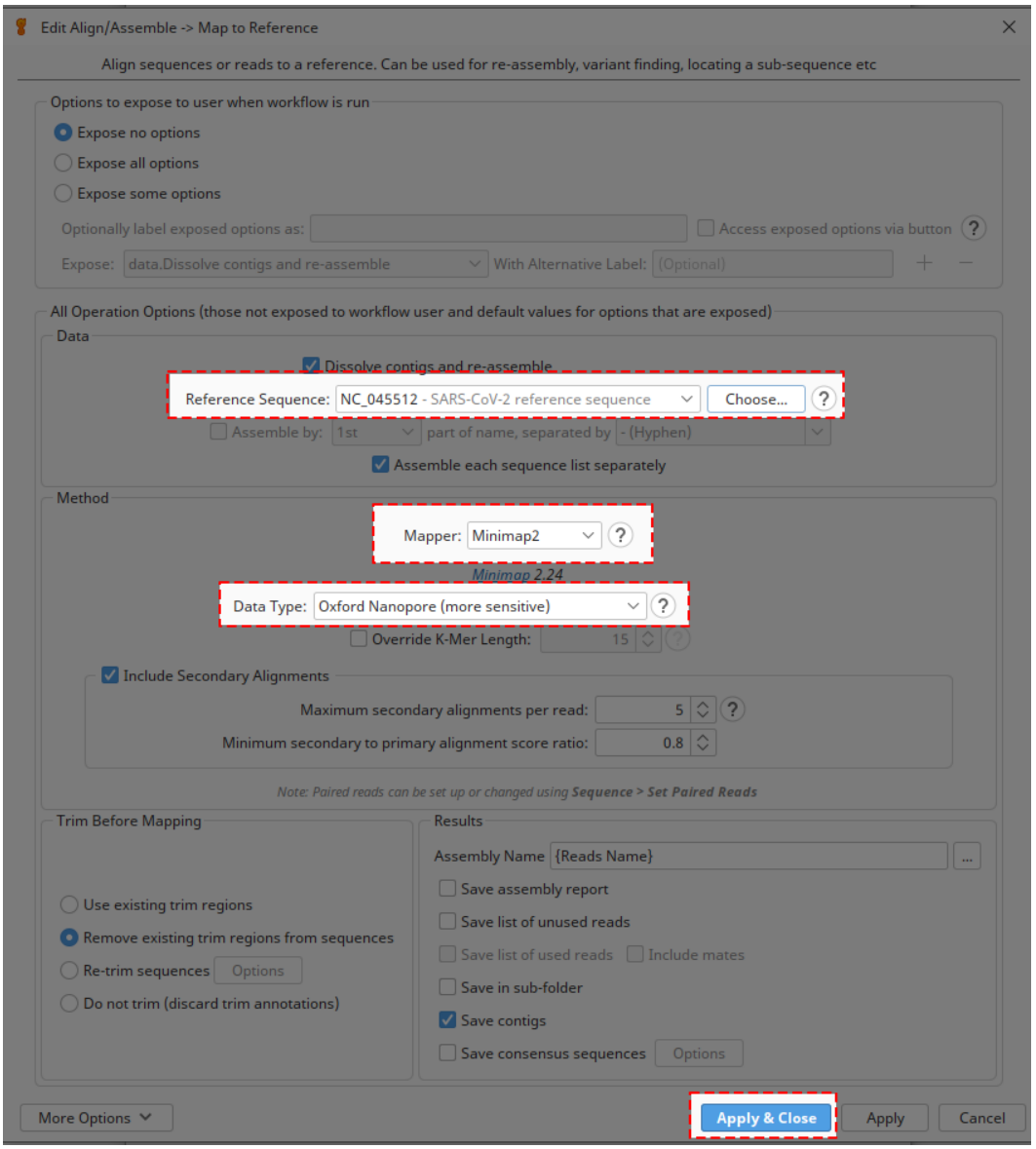


*Note: If you plan to use both sets of primers and don't want to manually select the primers set everytime you switch between them, it is possible to create two workflows where each workflow has selected one set of primers each. One way to do this is to go to **Manage Workflows**, select the imported 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

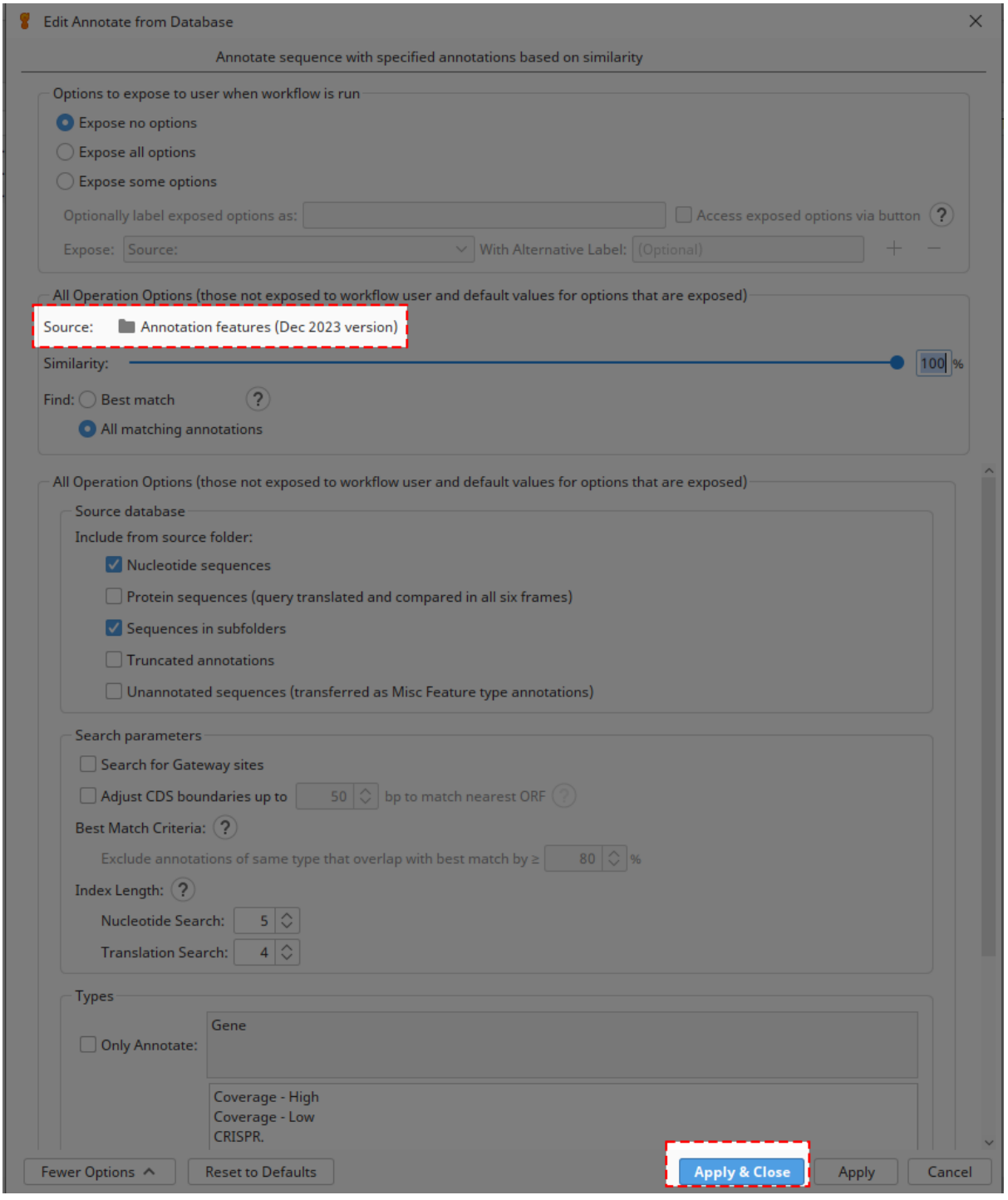
The alignment step uses Minimap2. If it is not already installed, go to Tools -> Plugins and under "Available plugins", the plugin named Minimap should be available for installation. It is also possible to download and install Minimap into Geneious Prime at geneious.com/plugins/minimap2.

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**.



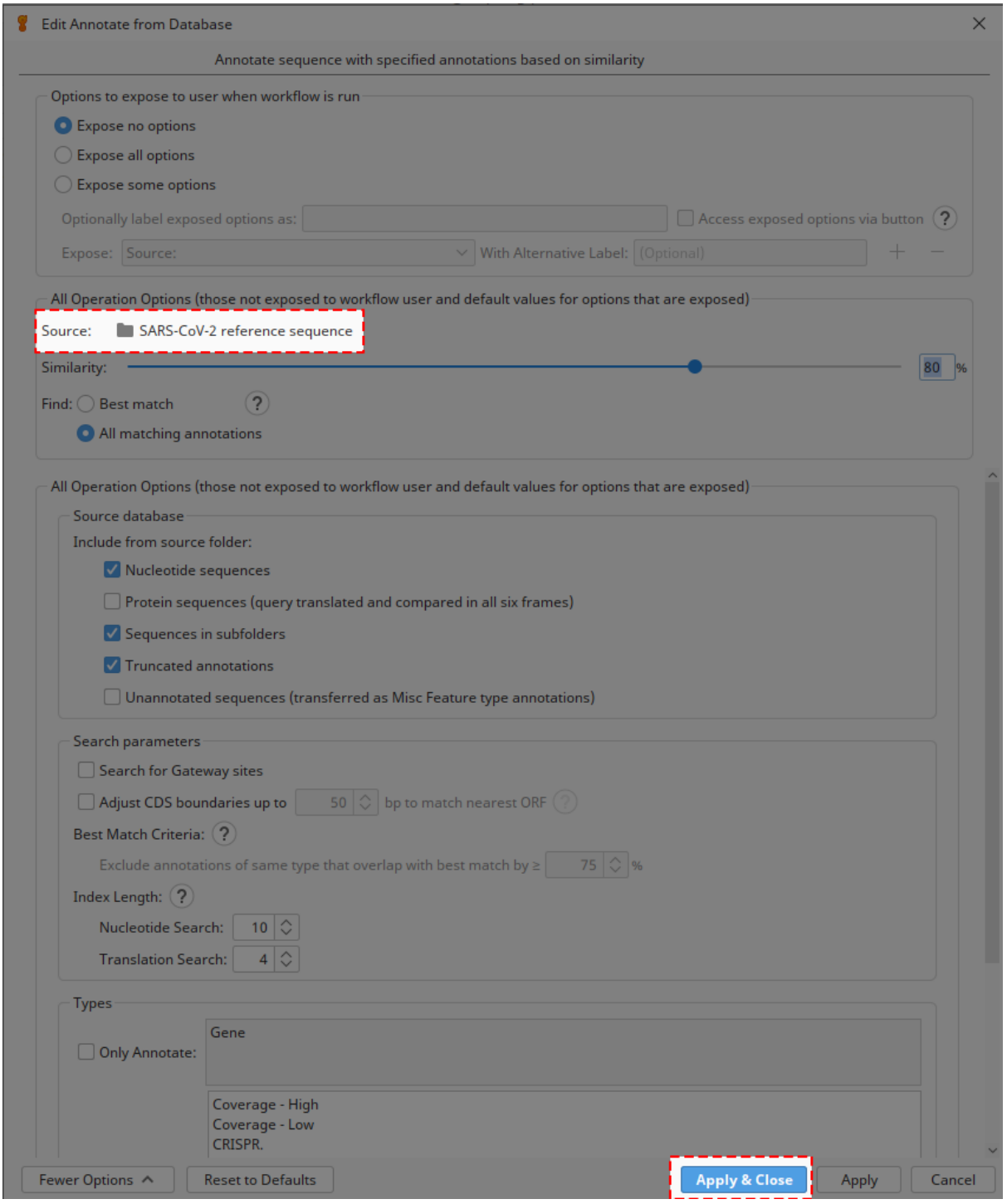
Annotate from Database (1)

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**.



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**.

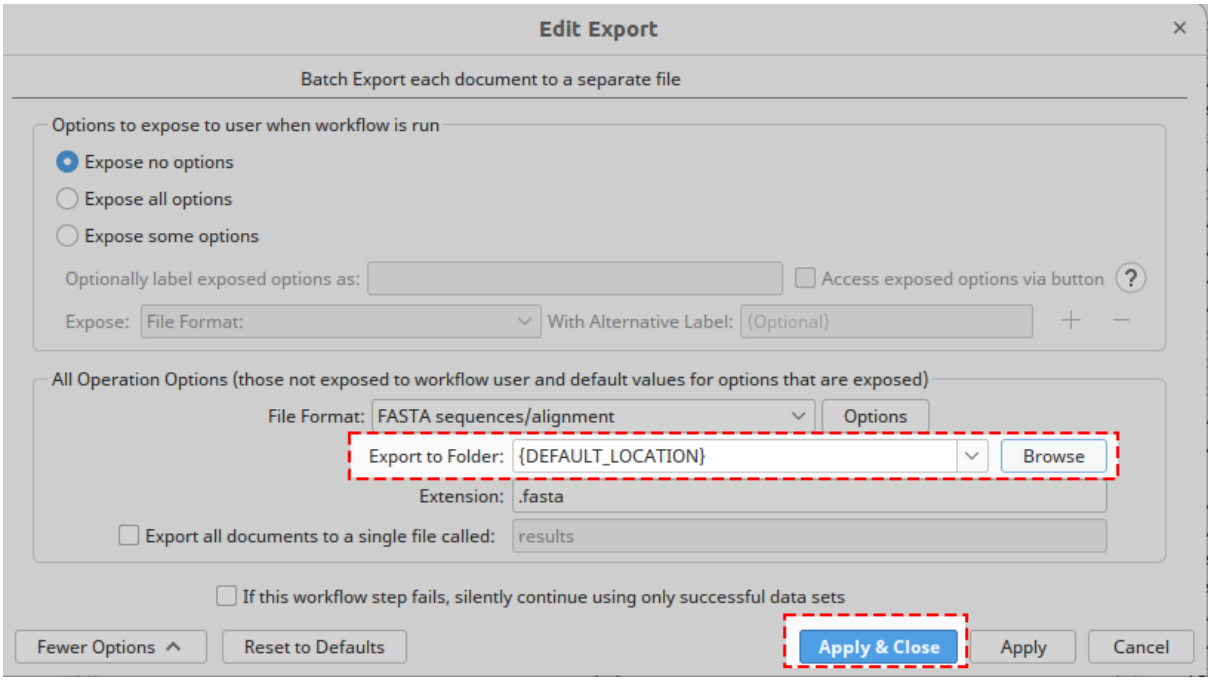


Use your own annotation features

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**.



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.

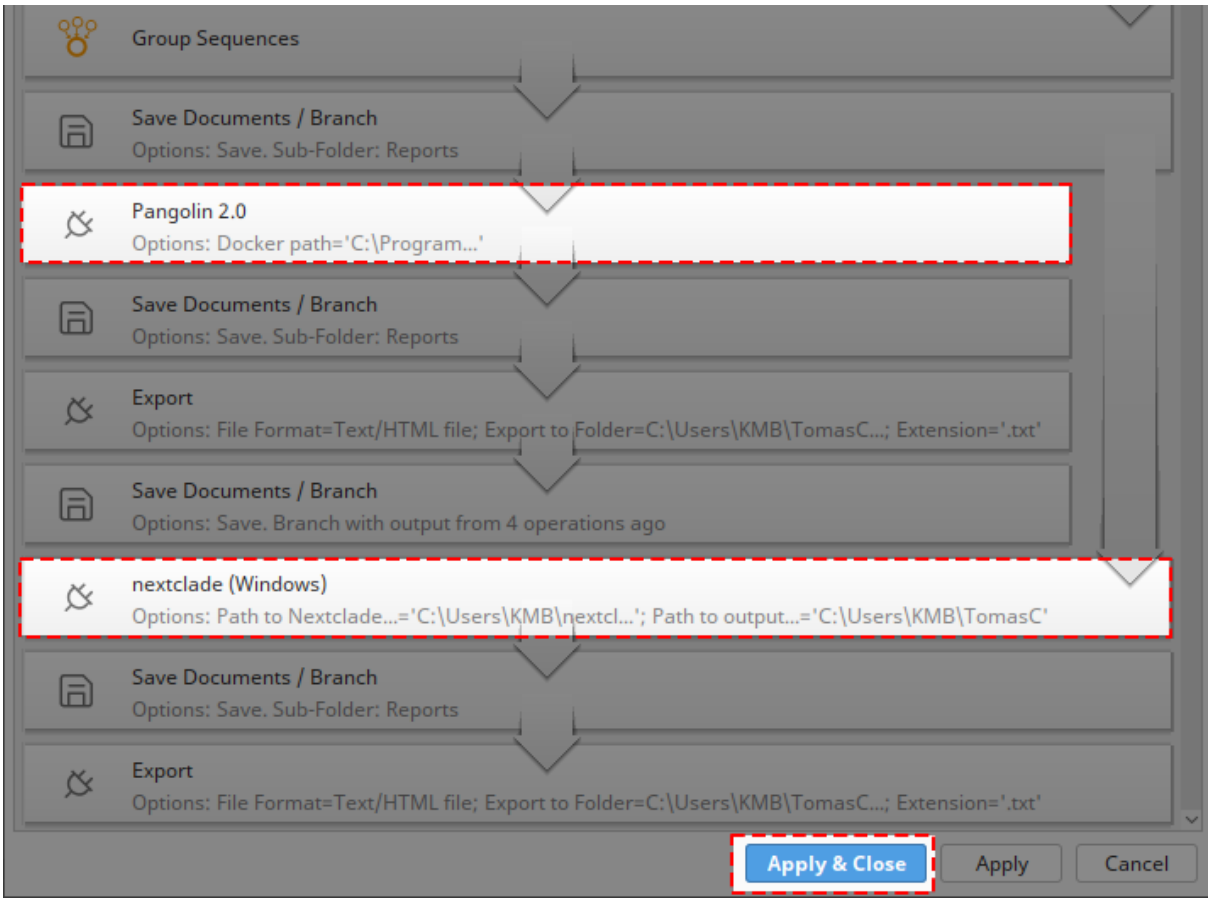
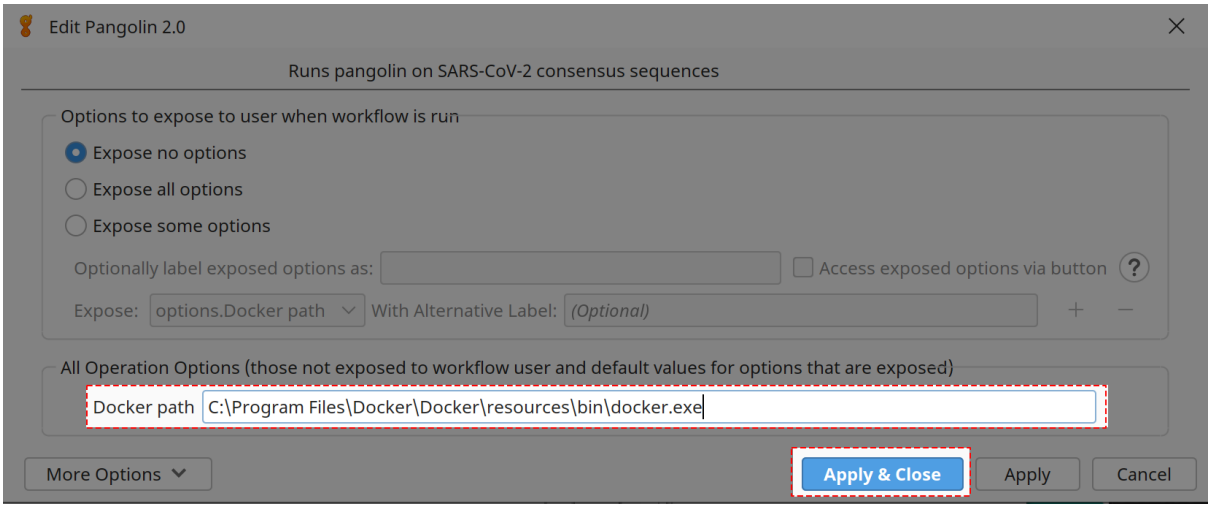


Figure 3. A screenshot of the Editing mode of "SARS-CoV-2_Assembly_WrapperPlugins.geneiousWorkflow". The locations of the plugin steps have been highlighted.

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:

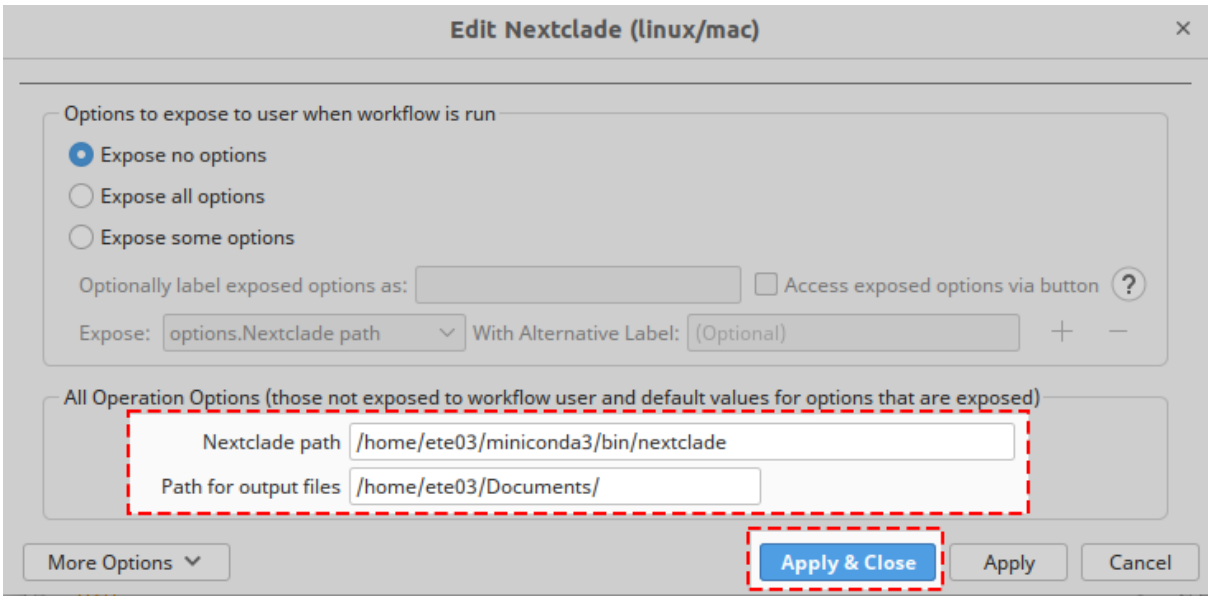
Docker path: The path to docker executable.



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.



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).