Assembly and Characterization

2

Oct 19, 2021

DRAGEN COVID Lineage App SARS-CoV-2 Strain Characterization on the Illumina BaseSpace Platform V.2

1

Technical Outreach and Assistance for States Team[1]

[1]Centers for Disease Control and Prevention

protocol .

TOAST_public
Tech. support email: **toast@cdc.gov**

ycm6

This protocol provides instructions on how to run the DRAGON COVID Lineage app on the Illumina BaseSpace Sequence Hub. The DRAGON COVID Lineage app reads in fastq sequence files and produces Lineage, Clade, and Kmer Detection reports for each run. The information provided in these reports include Pangolin lineage classification, variant calls, and the fraction of human and SARS-CoV-2 kmers detected. The DRAGON COVID Lineage app also assembles a consensus sequence fasta file for each sample. This document applies to all whole-genome sequencing runs on the Illumina platform and downstream bioinformatics for public health laboratories.

For technical assistance, please contact: **TOAST@cdc.gov**

Technical Outreach and Assistance for States Team 2021. DRAGEN COVID Lineage App SARS-CoV-2 Strain Characterization on the Illumina BaseSpace Platform. **protocols.io** https://protocols.io/view/dragen-covid-lineage-app-sars-cov-2-strain-charact-by78pzrw
Technical Outreach and Assistance for States Team

(ILLUMINA MENU) Protocols for SARS-CoV-2 Library Prep with ARTIC Primers, Bioinformatic Analysis, and Database Submission

protocol ,

Oct 19, 2021

Oct 19, 2021

Oct 19, 2021 ycm6

54240

Part of collection
(ILLUMINA MENU) Protocols for SARS-CoV-2 Library Prep with ARTIC Primers, Bioinformatic Analysis, and Database Submission

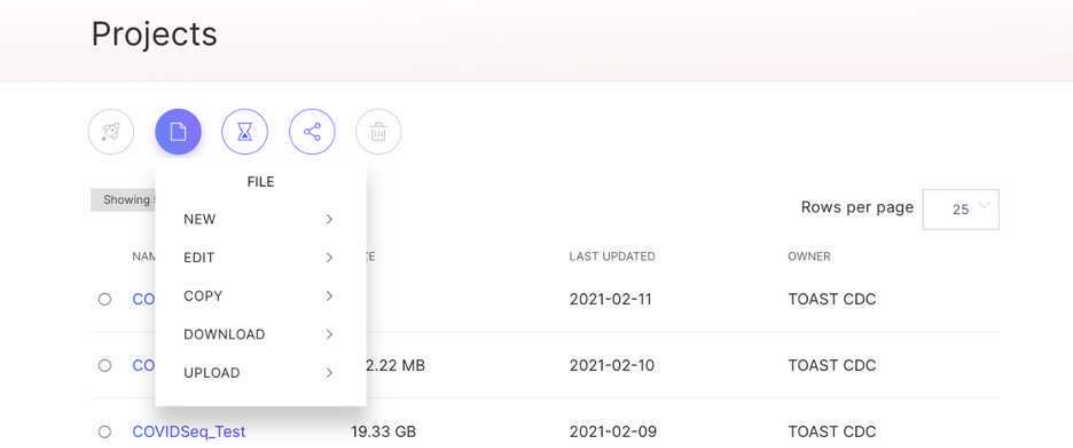
:

Create Basespace Project

1 Login to the Illumina BaseSpace Platform and create a new BaseSpace project

Login to the BaseSpace Sequence Hub at:
https://basespace.illumina.com/

Click on the 'Projects' tab on the top of the Sequence Hub homepage and create a new project directory by going to FILE -> NEW -> PROJECT:



The Projects panel in the BaseSpace Sequence Hub

Add Fastq files to BaseSpace Project

2 Add Fastq files to the new project directory. This can be done by either uploading fastq files from a local directory or by importing sequences using their Sequence Read Archive (SRA) accession numbers from NCBI using the 'SRA Import' app.

The step case format does not translate well in a pdf file. Be aware the pdf will present each step case sequentially which may be confusing when following the protocol.

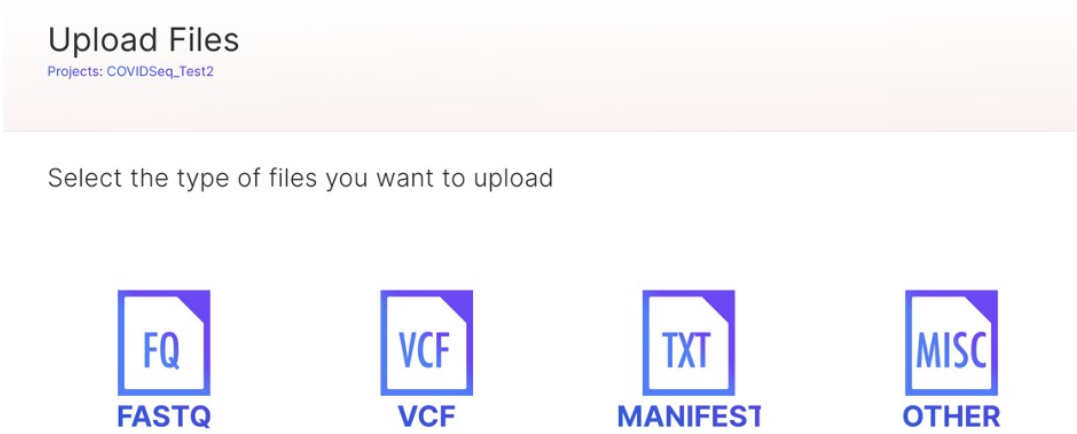
Step 2 includes a Step case.
Upload Fastq files from local directory
Import fastq files using the 'SRA Import' app

Run DRAGEN COVID Lineage App

step case

Upload Fastq files from local directory

Go to FILE --> UPLOAD --> FILES and it will bring you to the 'Upload Files' page



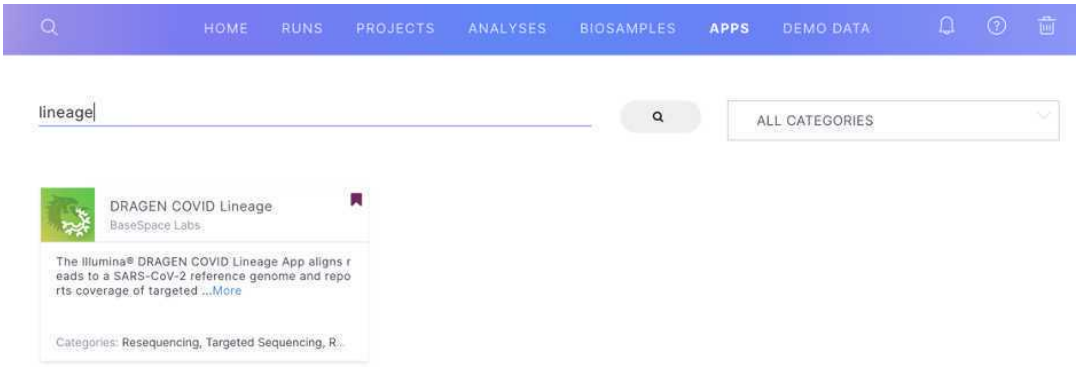
The Basespace Projects 'Upload Files' page

Click on the FASTQ tile and add the selected fastq files for upload. It should be noted that BaseSpace has specific upload requirements for fastq files including file naming schemes and header formats. They are described here: https://support.illumina.com/help/BaseSpace_Sequence_Hub/Source/Informatics/BS/UploadFastqReq_swBS.htm?Highlight=fastq

3

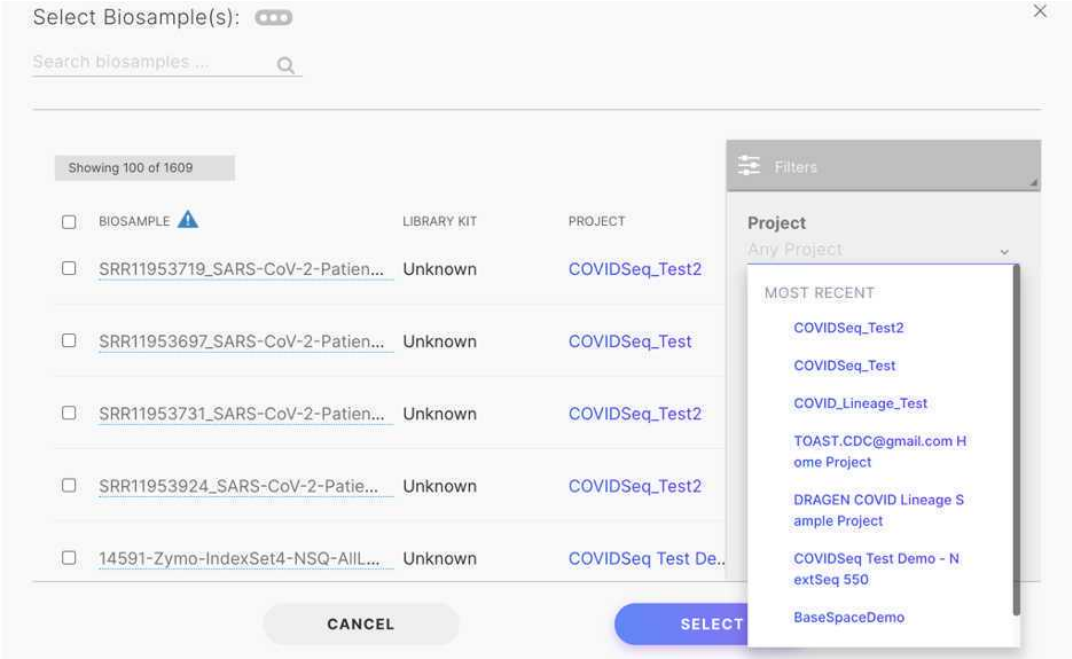
Run the DRAGEN COVID Lineage App

Go to the Apps panel and search 'lineage'. Click on the DRAGEN COVID Lineage app.



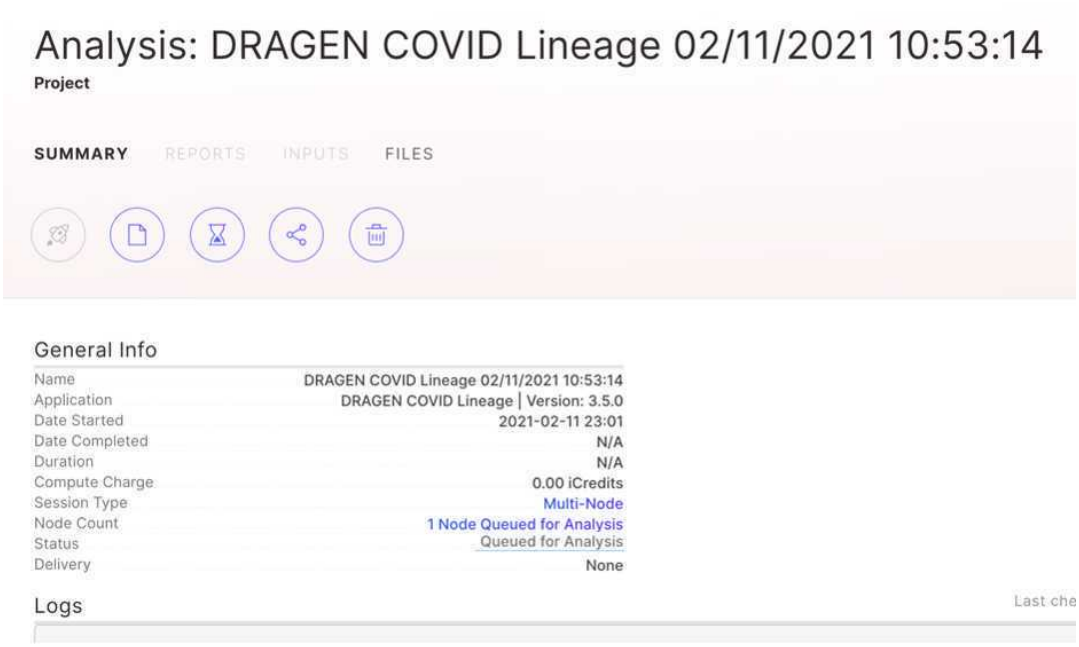
The 'DRAGEN COVID Lineage' BaseSpace app

Click ‘LAUNCH APPLICATION’ and then click ‘SELECT PROJECT’ and select the new project directory. Click ‘SELECT BIOSAMPLE(S)’. In the ‘SELECT BIOSAMPLE(S)’ window click the ‘Filters’ button on the right. Choose the new project directory



The ‘SELECT BIOSAMPLE(S)’ window

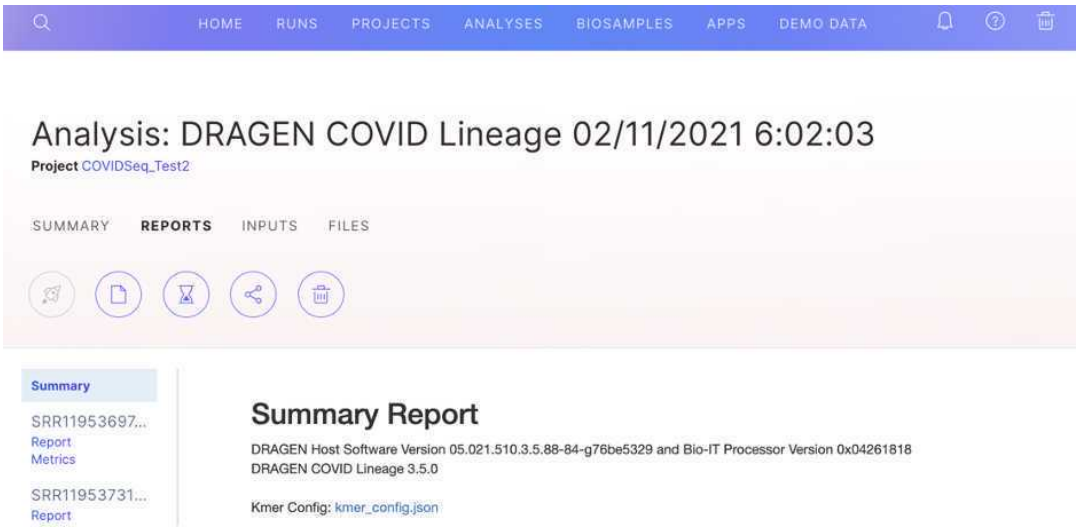
Select the samples to be used in the analysis and click the ‘SELECT’ button. At the bottom of the application page click ‘LAUNCH APPLICATION’. This will bring you to a new ‘ANALYSIS’ page similar to the following



The Analysis panel with a queued DRAGEN COVID Lineage run

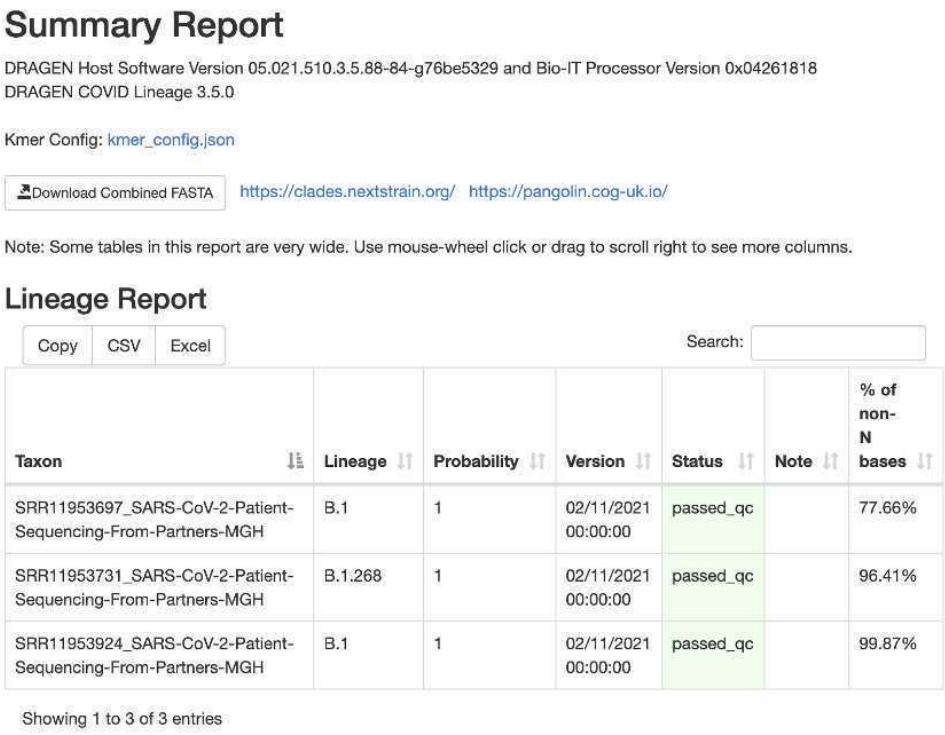
4 Accessing the DRAGEN COVID Lineage app summary report

The DRAGEN COVID Lineage App is finished when the status line says ‘Complete’. To view the ‘Summary Report,’ click on your analysis, navigate to the ‘REPORTS’ tab, and then click ‘Summary’ on the left side of the page.



The Analysis panel showing the DRAGEN COVID Lineage app Summary Report

The ‘Summary Report’ compiles together three individual reports: ‘Lineage Report’, ‘Clade Report’ and ‘Kmer Detection Report’. The ‘Lineage Report’, which provides Pangolin lineage assignments is shown below



Summary Report

DRAGEN Host Software Version 05.021.510.3.5.88-84-g76be5329 and Bio-IT Processor Version 0x04261818
DRAGEN COVID Lineage 3.5.0

Kmer Config: kmer_config.json

Download Combined FASTA https://clades.nextstrain.org/ https://pangolin.cog-uk.io/

Note: Some tables in this report are very wide. Use mouse-wheel click or drag to scroll right to see more columns.

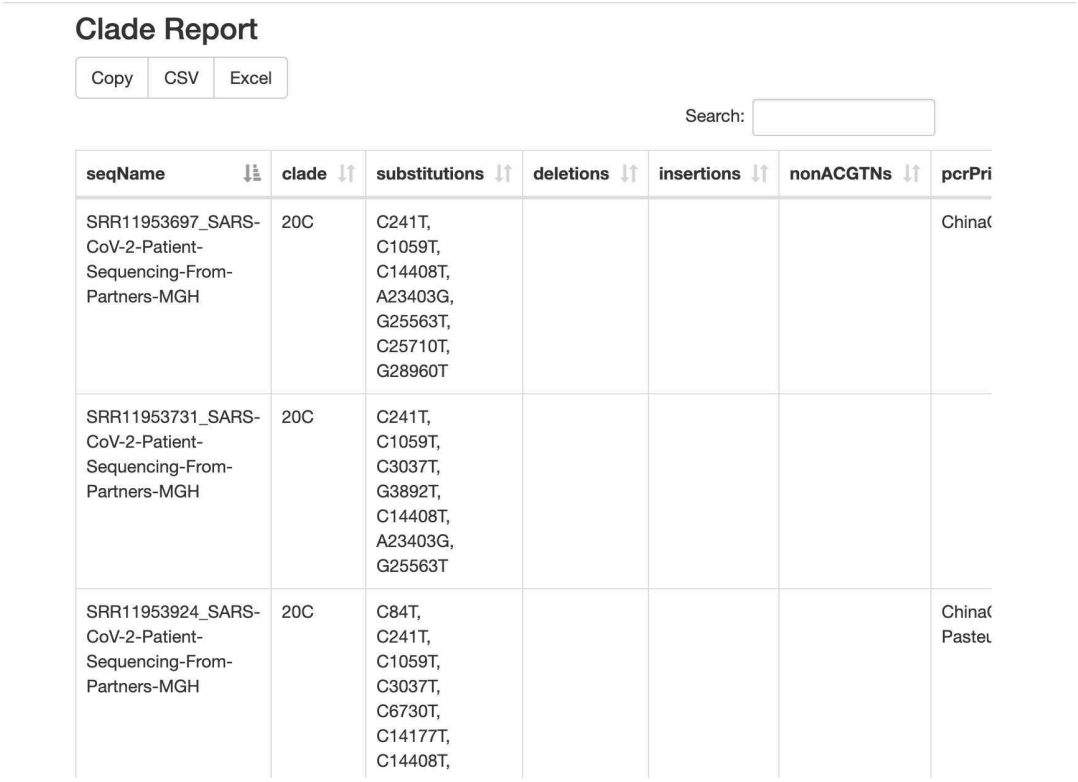
Lineage Report

Copy CSV Excel Search:

Taxon	Lineage	Probability	Version	Status	Note	% of non-N bases
SRR11953697_SARS-CoV-2-Patient-Sequencing-From-Partners-MGH	B.1	1	02/11/2021 00:00:00	passed_qc		77.66%
SRR11953731_SARS-CoV-2-Patient-Sequencing-From-Partners-MGH	B.1.268	1	02/11/2021 00:00:00	passed_qc		96.41%
SRR11953924_SARS-CoV-2-Patient-Sequencing-From-Partners-MGH	B.1	1	02/11/2021 00:00:00	passed_qc		99.87%

Showing 1 to 3 of 3 entries

The lineage section of the DRAGEN COVID Lineage Summary Report showing Pangolin lineage assignments



Clade Report

Copy CSV Excel

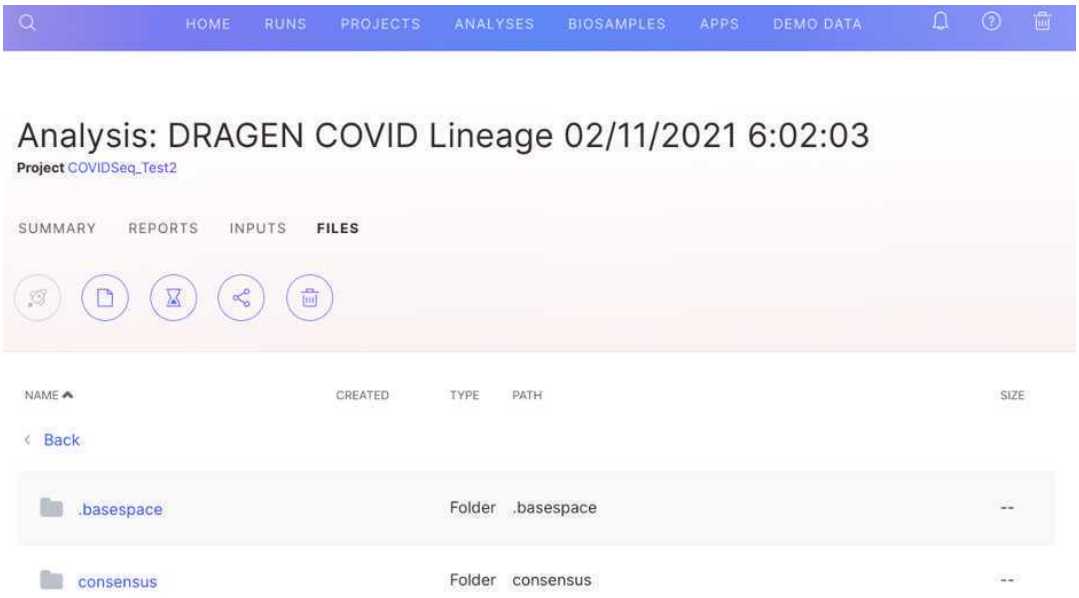
Search:

seqName	clade	substitutions	deletions	insertions	nonACGTNs	pcrPri
SRR11953697_SARS-CoV-2-Patient-Sequencing-From-Partners-MGH	20C	C241T, C1059T, C14408T, A23403G, G25563T, C25710T, G28960T				China(
SRR11953731_SARS-CoV-2-Patient-Sequencing-From-Partners-MGH	20C	C241T, C1059T, C3037T, G3892T, C14408T, A23403G, G25563T				
SRR11953924_SARS-CoV-2-Patient-Sequencing-From-Partners-MGH	20C	C84T, C241T, C1059T, C3037T, C6730T, C14177T, C14408T,				China(Pasteu

The Clade section of the DRAGEN COVID Lineage Summary Report showing NextClade clade assignments

5 Downloading consensus sequences generated by the DRAGEN COVID Lineage app

To download the SARS-CoV-2 consensus sequences produced by the DRAGEN COVID Lineage program, click on the 'FILES' tab and then click on one of the samples. This will bring you to a new 'FILES' page similar to the following:



The FILES tab of the Analysis panel for this DRAGEN COVID Lineage app run

Click on the 'consensus' directory and scroll down to the first file that has 'TYPE' 'fa'. Clicking on this sequence will pull up a new window the looks similar to the following:

Windows displaying the hard masked SARS-CoV-2 consensus sequence

This filename should end with 'consensus_hard_masked_sequence.fa'. Click 'DOWNLOAD'

to save the consensus sequence to a local directory. Note there is an additional TYPE 'fa' file that ends with a 'consensus_soft_masked_sequence.fa'. This consensus sequence should not be uploaded to a public database like Genbank or Gisaid.

Additional documentation for the DRAGEN COVID Lineage app is available here:

DRAGEN COVID Lineage (illumina.com)

Before submitting the resulting SARS-CoV-2 consensus sequence assemblies to public repositories, such as NCBI GenBank or GISAID, refer to the following documentation describing submission criteria and minimum quality control thresholds:

GenBank Submission Criteria: About GenBank Submission (nih.gov)
Gisaid Submission Criteria: **Gisaid inclusion criteria.pdf**

Alternative Lineage Assignment

6 The SARS-CoV-2 consensus sequence assembly generated by the CDRAGEN COVID Lineage app can also be uploaded to other lineage assignment software.

6.1 Upload the consensus sequence for each sample to the **Pangolin COVID-19 Lineage Assigner** at:

https://pangolin.cog-uk.io/

Click the 'Start analysis' button:

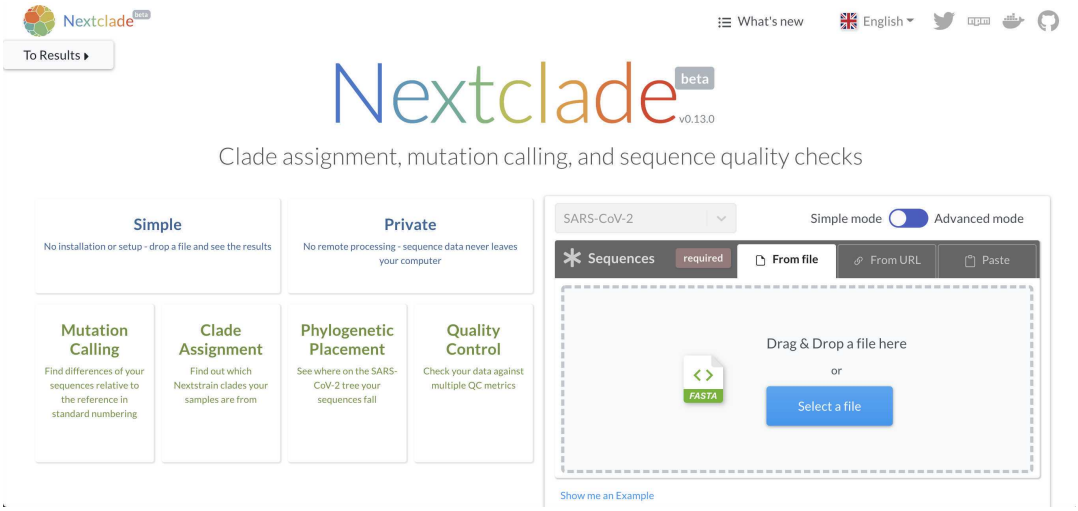
Pangolin COVID-19 Lineage Assigner example

The Pangolin COVID-19 Lineage Assigner returns the lineage classification and assignment probability:

Pangolin COVID-19 Lineage Assigner output

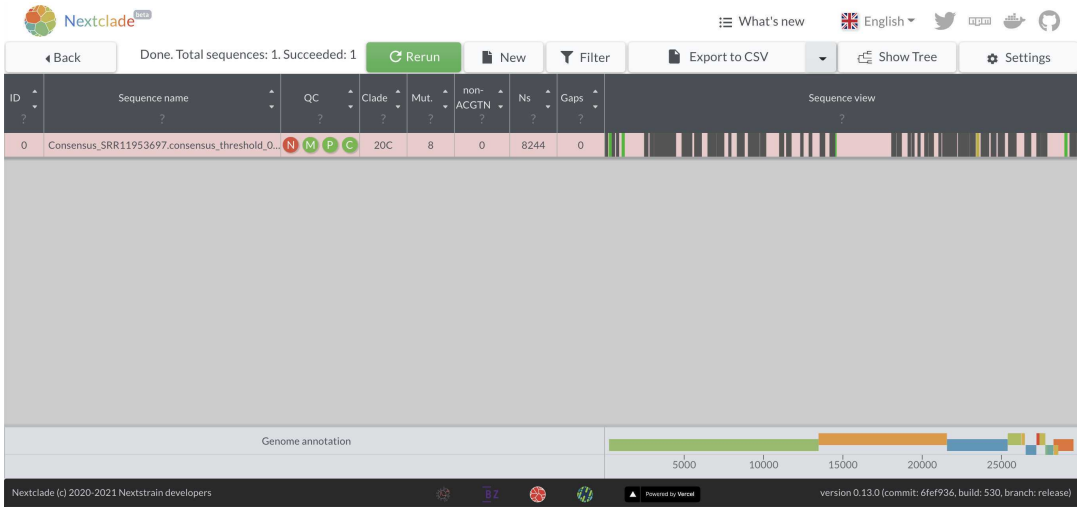
6.2 Or upload the consensus sequence for each sample to the **Nextclade** clade assignment web portal at:

https://clades.nextstrain.org/



NextClade assignment web portal

The Nextclade server provides clade classification as well as QC metrics and a list of amino acid substitutions. A summary output file can be downloaded with the 'Export to CSV' button.



Nextclade clade assignment output