Doherty Institute | The University of Melbourne | The Royal Melbourne Hospital

A joint venture between The University of Melbourne and The Royal Melbourne Hospital

Lecture – Introduction to Nextclade for Lineage Assignment and QC

Dr Kristy Horan

Objectives

- Use Nextclade to perform quality control on our consensus sequences
- Understand and interpret Nextclade outputs
- Download Nextclade results

Nextclade: analysis of viral genetic sequences

Nextclade is an open-source project for viral genome alignment, mutation calling, clade assignment, quality checks and phylogenetic placement.

Nextclade consists of a set of related tools:

- Nextclade Web - a web application available online at clades.nextstrain.org
- Nextclade CLI - a command-line tool

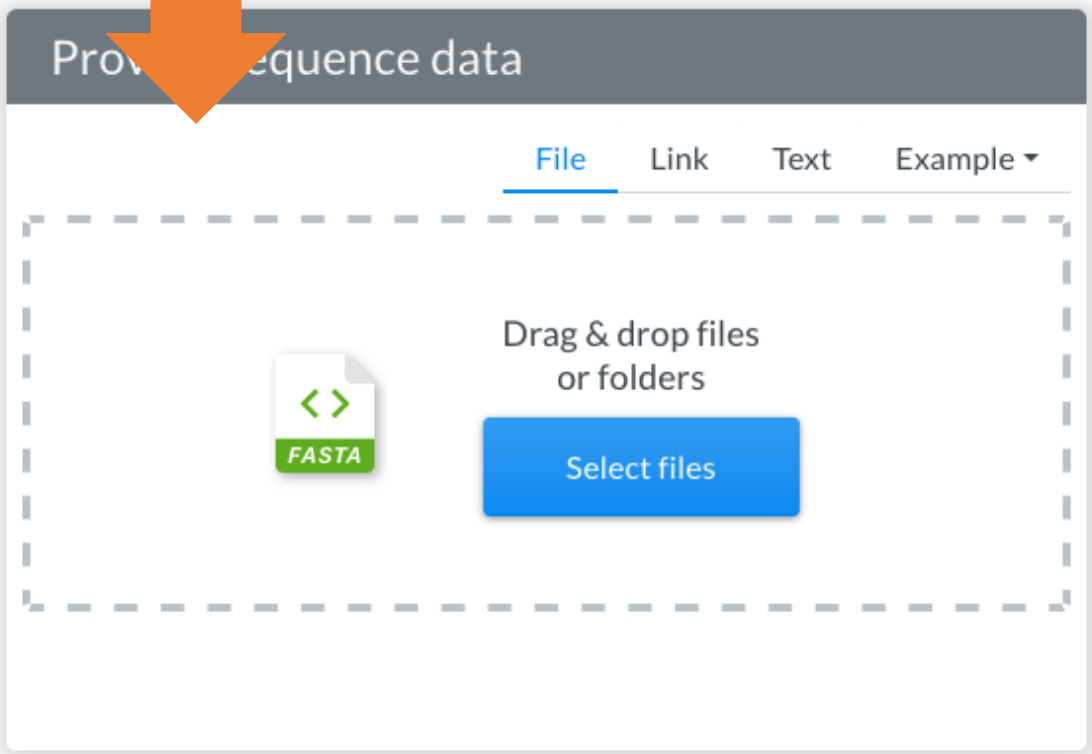
Both tools are powered by the same algorithms, they consume the same inputs and produce the same outputs, but they differ in the user interface, the features included, and the degree of customization. It is recommended to start with Nextclade Web and later proceed to CLI tools if you have more advanced use-cases (for example, repeated batch processing, bioinformatics pipelines).

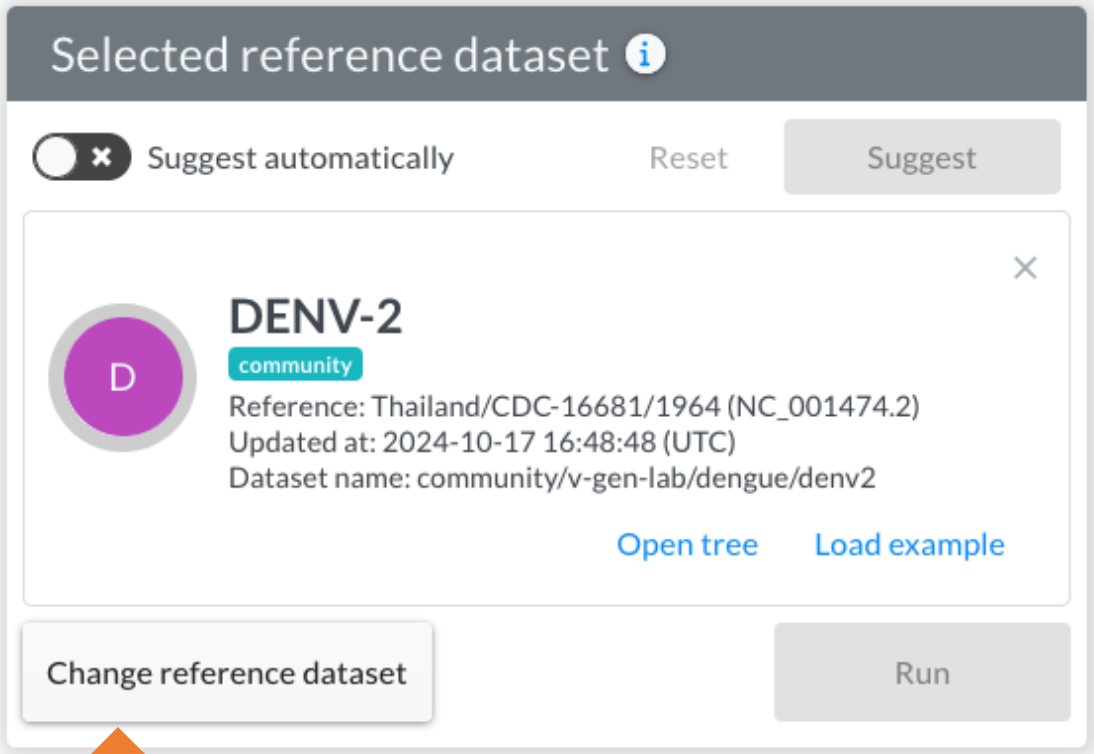
Tutorial – How to use Nextclade to analyse and QC your data

Data input here

Nextclade v3.10.0

Clade assignment, mutation calling, and sequence quality checks

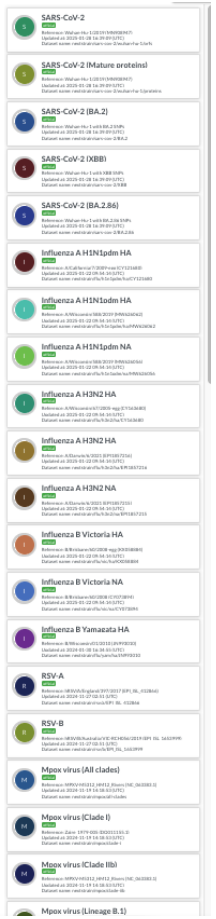




Select reference dataset here

Nextclade supports many viruses

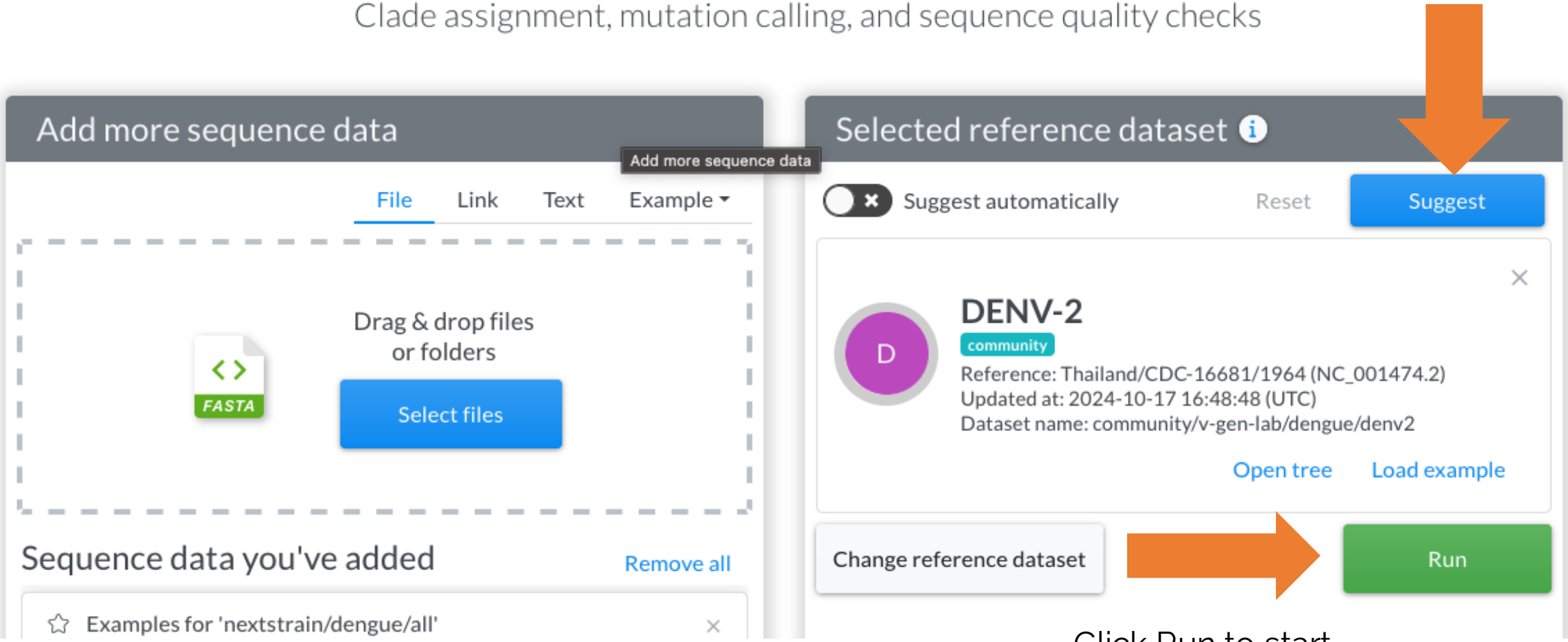
Change reference dataset



Nextclade v3.10.0

Clade assignment, mutation calling, and sequence quality checks

If unsure which ref dataset, click 'Suggest'



Click Run to start analysis

Nextclade | Start | Dataset | Results | Tree | Export | Done. Total sequences: 165. Succeeded: 165 | Settings | About | Citation | Docs | CLI | EN

Show tree

Download results

Filter results

Select gene

Bad sequence

Stretches of Ns

Frameshift

#	i	Sequence name	QC	Clade	Pangolin lineage (Nextclade)	Unaliased	Immune escape	ACE2 binding	Mut.	non-ACGTN	Ns	Cov.	Gaps	Ins.	FS	SC
0	1	USA/CA-LACPHL-AY03266/2023	N M P C F S	23F	EG.5.1.13	XBB.1.9.2.5.1.	0.39	-0.27	30	0	1229	95.1%	39	0	0	0 (1)
1	4	USA/CA-LACPHL-AY03267/2023	N M P C F S	23F	EG.5.1.16	XBB.1.9.2.5.1	0.32	-0.22	20	0	0	99.3%	30	0	0	0 (1)
2	0	OY754687								2	160	99.5%	56	0		
3	3	OY754651	N M P C F S	23F	EG.5.1	XBB.1.9.2.5.1	0.32	-0.22	22	2	3400	88.6%	9	0		
4	2	OY754528	N M P C F S	23E	GE.1	XBB.2.3.10.1	0.05	0.46	13	14	11269	62.3%	0	0	0	0
5	6	USA/WA-UW-23[illegible]46451/2023	N M P C F S	recombinant	XCR	XCR	0.37	-0.77	24	0	0	98.8%	30	0	0	0 (1)
6	7	USA/WA-UW-23102330989/2023	N M P C F S	23F	HV.1	XBB.1.9.2.5.1.	0.42	-0.26	22	0	0	98.8%	30	0	0	0 (1)
7	8	USA/CA-LACI[illegible]	[illegible]G		XBB.1.5.70	XBB.1.5.70	0.37	-0.77	18	0	756	96.7%	30	0	0	0 (1)
8	5	OY754526	[illegible]23E		GJ.2	XBB.2.3.3.2	0.04	0.55	19	0	1944	93.5%	9	0	0	0
9	9	OY754632	N M P C F S	23I	JN.3	BA.2.86.1.3	0.64	-0.09	67	2	185	99.4%	3	0	0	0
10	12	OY754626	N M P C F S	23F	EG.5.1.3	XBB.1.9.2.5.1.	0.32	-0.22	20	2	143	99.5%	9	0	0	0 (1)
11	10	USA/CA-LACPHL-AY03247/2023	N M P C F S	23B	XBB.1.16.6	XBB.1.16.6	0.34	-0.37	25	0	129	98.8%	32	0	0 (1)	0 (1)
12	11	USA/CA-LACPHL-AY03264/2023	N M P C F S	23A	HR.1.1	XBB.1.5.77.1.1	0.05	0.46	26	0	1	99.2%	30	0	0	0 (1)
13	14	OY754681	N M P C F S	23B	JF.1	XBB.1.16.6.1	0.39	-0.92	25	2	83	99.7%	56	0	0	0 (1)
14	13	USA/CA-LACPHL-AY03246/2023	N M P C F S	23C	DV.7.1.4	BA.2.75.3.4.1.	0.80	-1.46	50	0	1398	94.6%	27	0	0	0
15	15	OY754673	N M P C F S	recombinant	XCH.1	XCH.1	0.37	-0.77	27	5	142	99.5%	9	0	0	0 (1)
16	16	USA/LA-EVTL20384/2023	N M P C F S	23A	JD.1.1	XBB.1.5.102.1	0.39	-1.49	24	0	0	99.9%	21	0	0	0 (1)
17	18	OY754523	N M P C F S	23F	EG.5.1.3	XBB.1.9.2.5.1.	0.32	-0.22	21	2	143	99.5%	9	0	0	0 (1)
18	19	USA/CA-LACPHL-AY03248/2023	N M P C F S	23D	FL.2.5	XBB.1.9.1.2.5	0.03	0.61	22	0	1	99.2%	30	0	0	0 (1)
19	17	USA/NY-PBRI-NYC2O026/2023	N M P C F S	23F	EG.5.1.4	XBB.1.9.2.5.1.	0.58	0.00	25	0	0	99.6%	56	0	0	0 (1)
20	20	OY754683	N M P C F S	23I	BA.2.86.1	BA.2.86.1	0.64	-0.09	61	2	182	99.4%	3	0	0	0
21	21	USA/CA-LACPHL-AY03262/2023	N M P C F S	23F	HK.31	XBB.1.9.2.5.1.	0.32	-0.22	24	0	0	99.3%	30	0	0	0 (1)
22	22	OY754698	N M P C F S	23D	EG.9.1	XBB.1.9.2.9.1	0.10	-0.01	21	1	331	98.9%	9	0	0	0 (2)
23	23	USA/MN-MDH-37211/2023	N M P C F S	23B	HF.1	XBB.1.16.13.1	0.05	0.46	27	0	0	99.5%	56	0	0	0 (1)
24	24	OY754610	N M P C F S	23F	EG.5.1.6	XBB.1.9.2.5.1.	0.32	-0.22	20	3	142	99.5%	9	0	0	0 (1)
25	26	OY754688	N M P C F S	23I	JN.3.1	BA.2.86.1.3.1	-0.00	0.00	69	2	79	99.7%	68	0	1 (2)	0
26	25	OY754532	N M P C F S	23B	XBB.1.16.15	XBB.1.16.15	0.06	0.40	21	0	1596	94.7%	9	0	0	0 (1)
27	29	USA/CA-LACPHL-AY03275/2023	N M P C F S	23G	GK.1.10	XBB.1.5.70.1.1	0.47	-1.38	32	0	4	99.2%	27	0	0	0 (1)
28	27	USA/NY-PBRI-NYC2O027/2023	N M P C F S	23F	HV.1.6	XBB.1.9.2.5.1	0.69	-0.19	29	0	19	99.6%	56	0	0	0 (1)

Genome annotation ?

Nucleotide sequence

5000 10000 15000 20000 25000

version 3.0.0-alpha.0 (commit: 4deeba5, branch: master)

Sequence name	QC	Clade	Mut.	non-ACGTN	Ns	Cov.	Gaps	Ins.	FS	SC
?	?	?	?	?	?	?	?	?	?	?

- “Mut.”: number of mutations with respect to the reference sequence
- “non-ACGTN”: number of ambiguous nucleotides that are not *N*
- “Ns”: number of missing nucleotides indicated by *N*
- “Gaps”: number of nucleotides that are deleted with respect to the reference sequence
- “Ins.”: number of nucleotides that are inserted with respect to the reference sequence
- “FS”: Number of uncommon frame shifts (total number, including common frame shifts are in parentheses)
- “SC”: Number of uncommon premature stop codons (total number, including common premature stops are in parentheses)

QC Field

Hover mouse over highlighted sample to show popup.

Warning

Overall QC score: 56
Overall QC status: mediocre
Detailed QC assessment:

- **Missing Data**: good
  No issues
- **Private Mutations**: good
  No issues
- **Frame shifts**: mediocre
  Unexpected 1 frame shift(s) detected: NS5:404-900. QC score: 75
- **Stop codons**: good
  No issues

Failed

Overall QC score: 773
Overall QC status: bad
Detailed QC assessment:

- **Missing Data**: bad
  Too much missing data found. Total Ns: 2880 (1100 allowed). QC score: 278
- **Private Mutations**: good
  No issues
- **Frame shifts**: good
- **Stop codons**: good
  No issues

N	P	F	S	2III_C.1.1	508	0	2880	68.0%	0	0	0	0

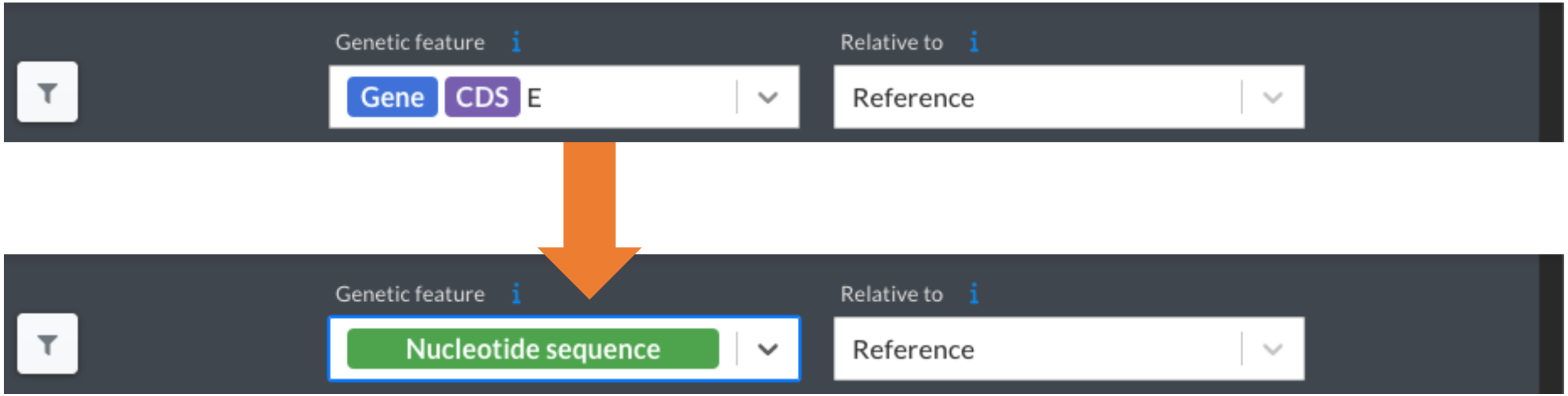
Why was it marked as fail? Too much missing data (N or

Nextclade QC Metrics

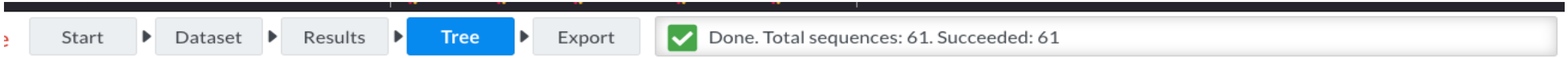
- N Missing Data – threshold 20,000
- P Private Mutations – cutoff 50, typical 5
- F Frame Shift
- S Stop Codon
- Min length: 10,000

- Nextclade implements a variety of quality control metrics to quickly spot problems in your sequencing/assembly pipeline.
- Bad sequences are colored red, mediocre ones yellow and good ones white. You can view detailed results of the QC metrics by hovering your mouse over a sequences QC entry:

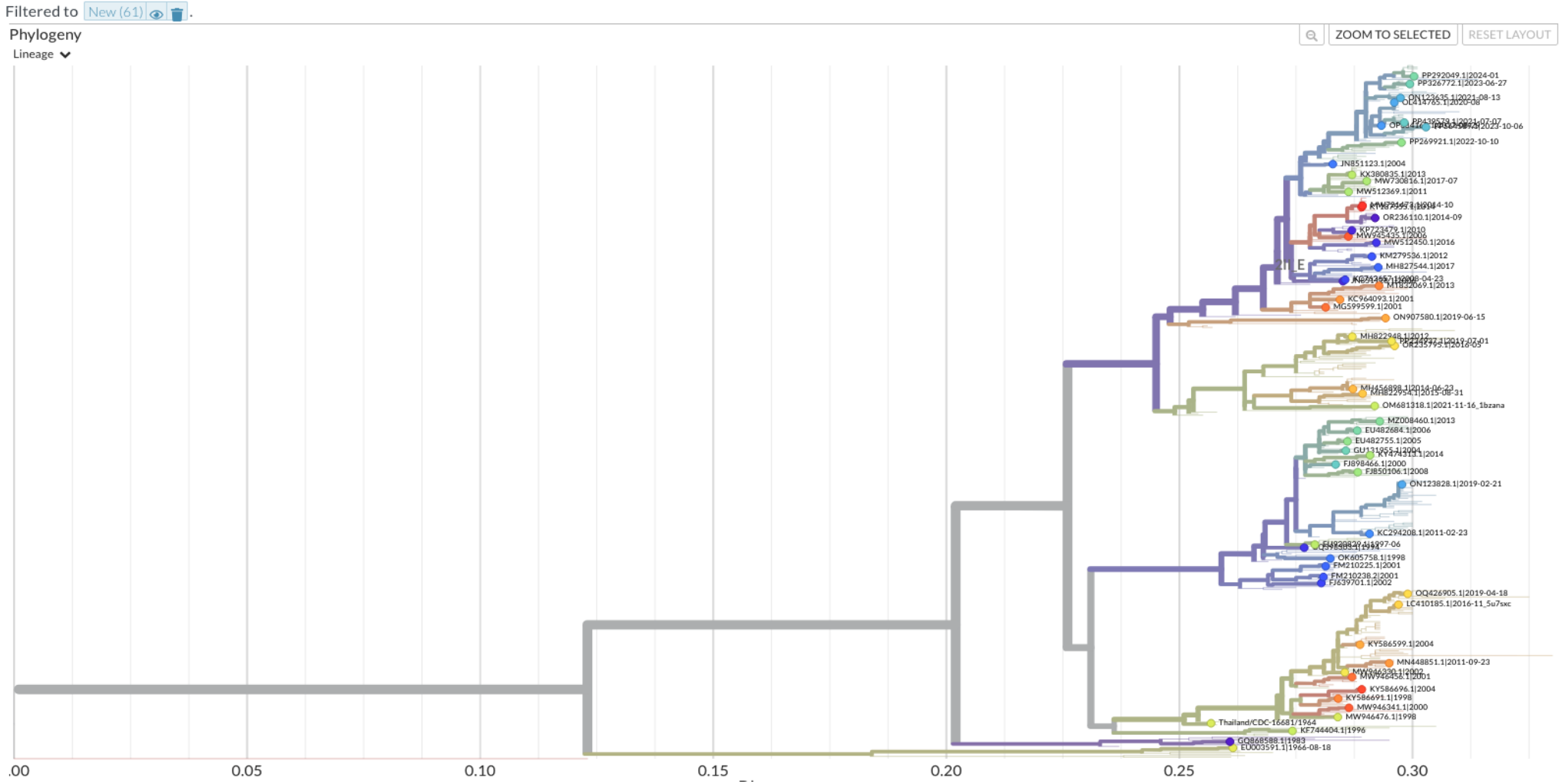
Change Genetic Feature to Nucleotide to show full genome



Note: In Dengue the E gene is routinely sequenced as it's genetically distinct between serotypes and provides sufficient resolution for clade information and possible lineage information.



Click on the "Tree" button to see where your sequences have been placed.
The tree is nearly identical to Next*strain* tree, so interact with it.

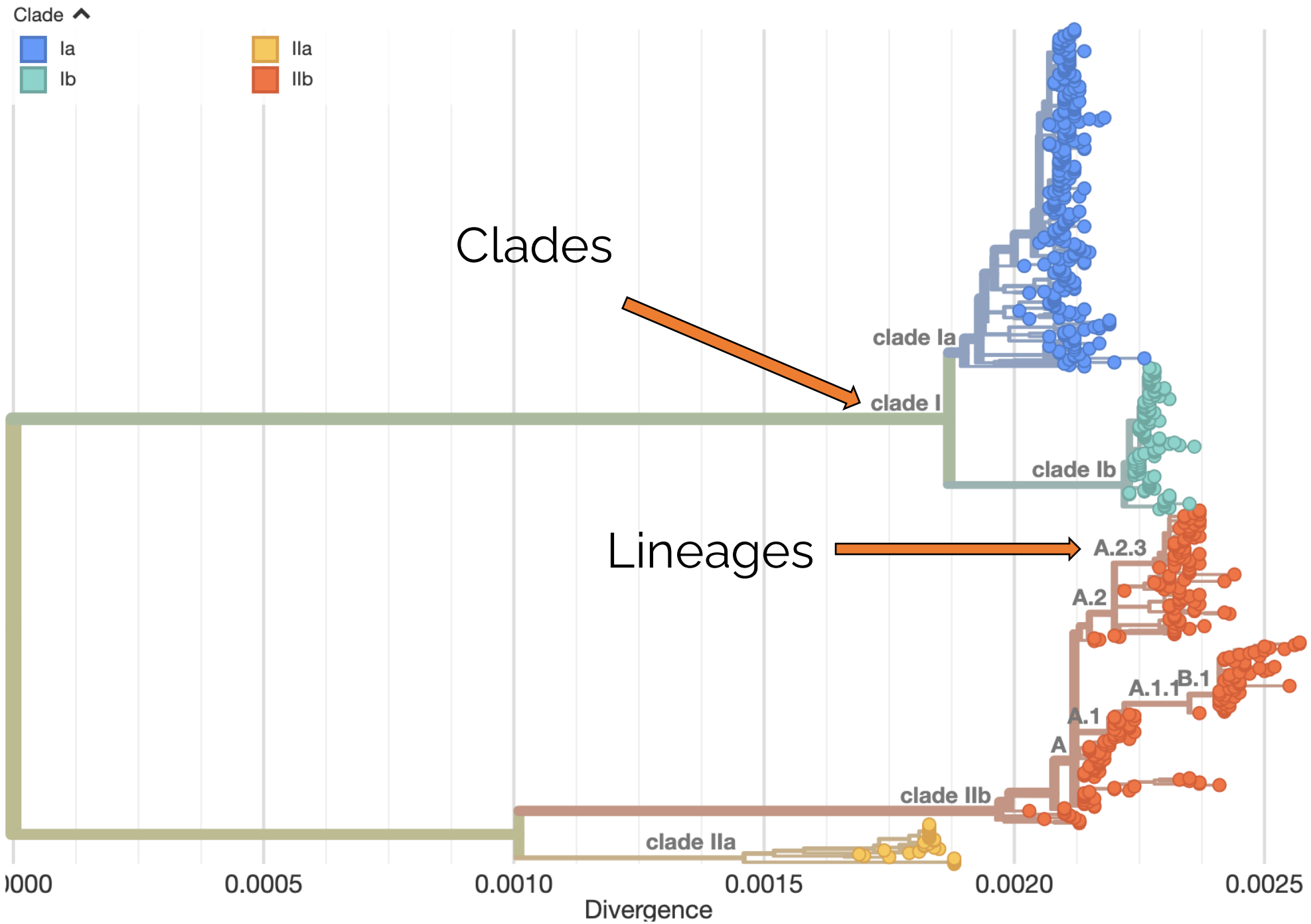


MPXV Clades and Lineages

The subtypes of mpox virus were also renamed; the clade formerly known as "Congo Basin (Central African)" was renamed **clade I**, and the clade formerly known as "West African" was renamed **clade II.**

MPXV is classifed into 2 main clades, clade I and clade II, with each further subdivided into clade Ia, the newly identifed clade Ib, clade IIa, and clade IIb. Clade I mpox is generally associated with higher CFRs (1.4% to ~10%) compared with clade II (0.1% to 3.6%).

Lineages are fine grained, higher resolution groupings. Current lineages are designed by Nextstrain


Clade
Ia
Ib
IIa
IIb
Clades
Lineages
clade Ia
clade I
clade Ib
A.2.3
A.2
A.1.1
B.1
A.1
A
clade IIb
clade IIa
0.0005
0.0010
0.0015
0.0020
0.0025
Divergence

Questions? + Resources

- Nextclade website: https://clades.nextstrain.org
- Documentation: https://docs.nextstrain.org/projects/nextclade
- Nextclade is also a CLI tool and is avaliable in Galaxy.
  - https://docs.nextstrain.org/projects/nextclade/en/stable/user/nextclade-cli/index.html
  - The web interface is excellent, provides all the features of the CLI and even more.

- Differences between clade I and Clade II MPXV
  - https://publichealth.jhu.edu/sites/default/files/2024-06/mpox-clad-i-vs-ii.pdf