UNIVERSIDAD
SAN FRANCISCO
DE QUITO

Training course in Bioinformatic tools

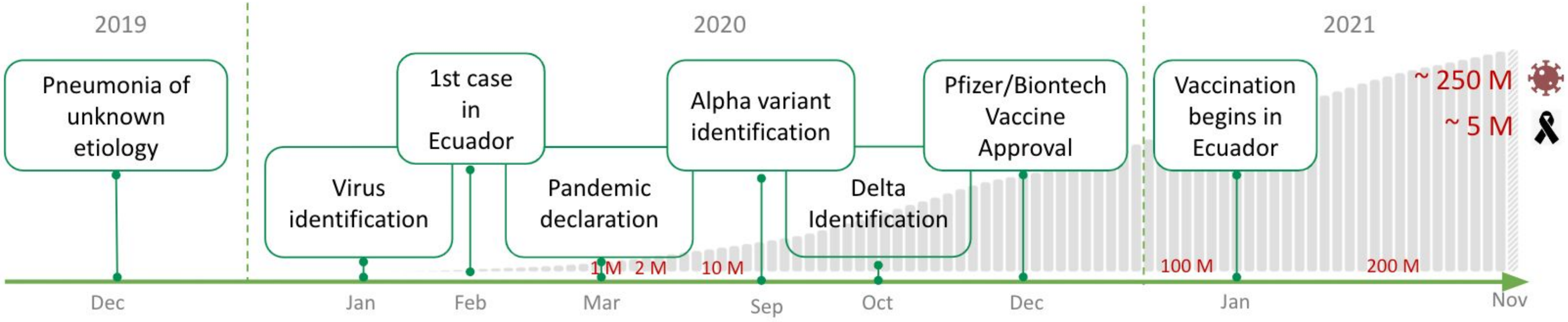
Instituto de Microbiología
Centro de Bioinformática

Virtual training

Module 3: SARS-CoV-2 analysis

May 2023

COVID-19 pandemics



Pneumonia of unknown etiology

Virus identification

1st case in Ecuador

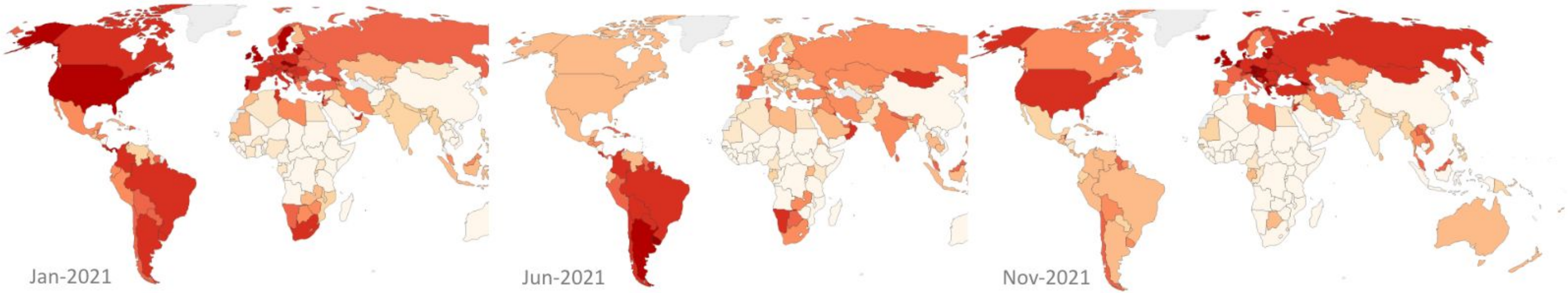
Pandemic declaration

Alpha variant identification

Delta Identification

Pfizer/Biontech Vaccine Approval

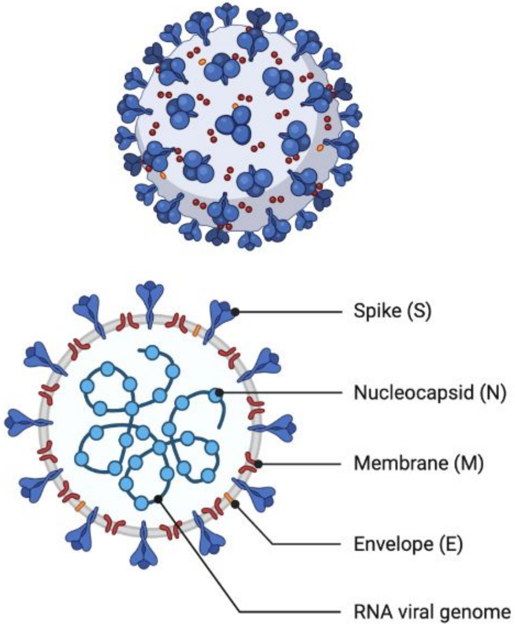
Vaccination begins in Ecuador



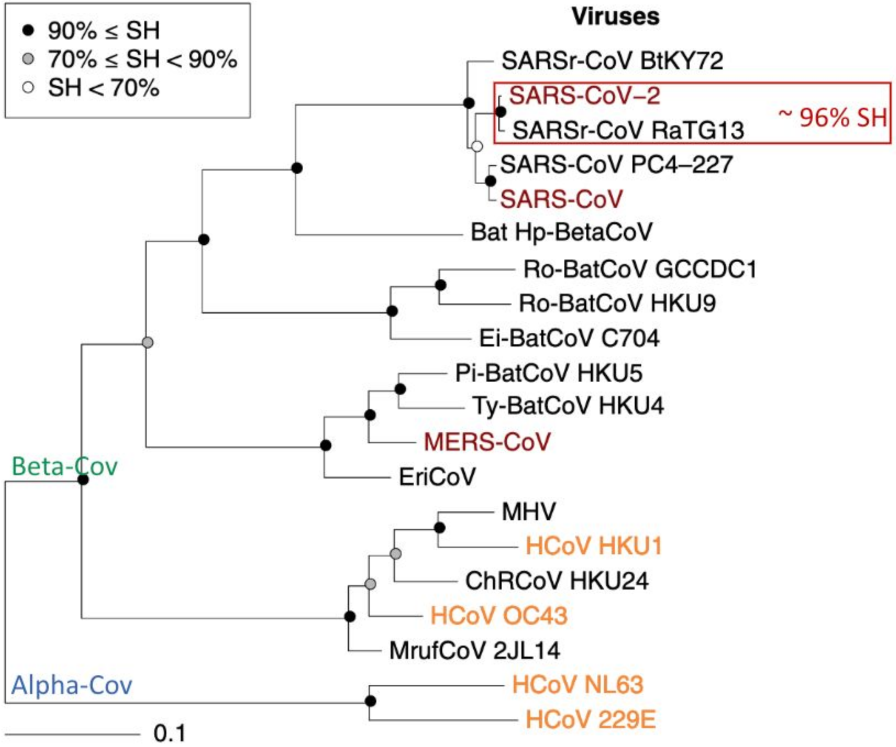
[1] American Society for Microbiology. (2020). *COVID-19 Resources*. [2] WHO. (2021). *COVID-19 Dashboard*. [3] Ritchie *et al.* (2021). *"Coronavirus Pandemic (COVID-19)" - OurWorldInData.org*.

SARS-CoV-2 virus

Family: *Coronaviridae*
Genre: *Betacoronavirus*
Subgenre: *Sarbecovirus*

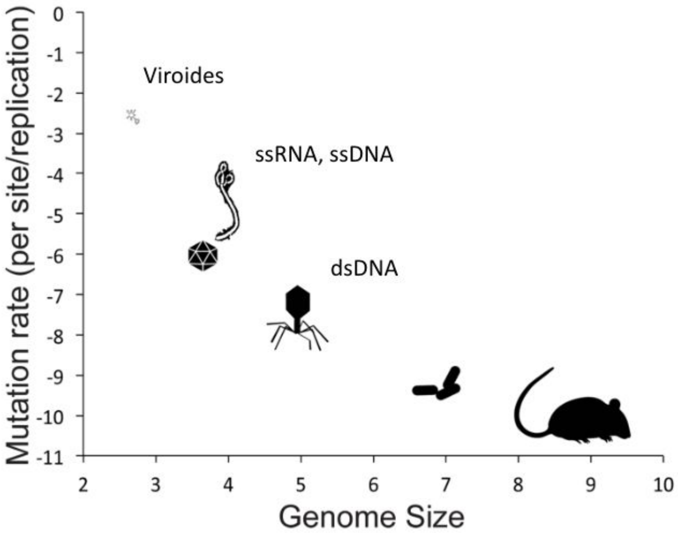


Source: BioRender.com



Source: Gorbalenya *et al.* (2020). *Nature Microbiology.*

SARS-CoV-2 virus

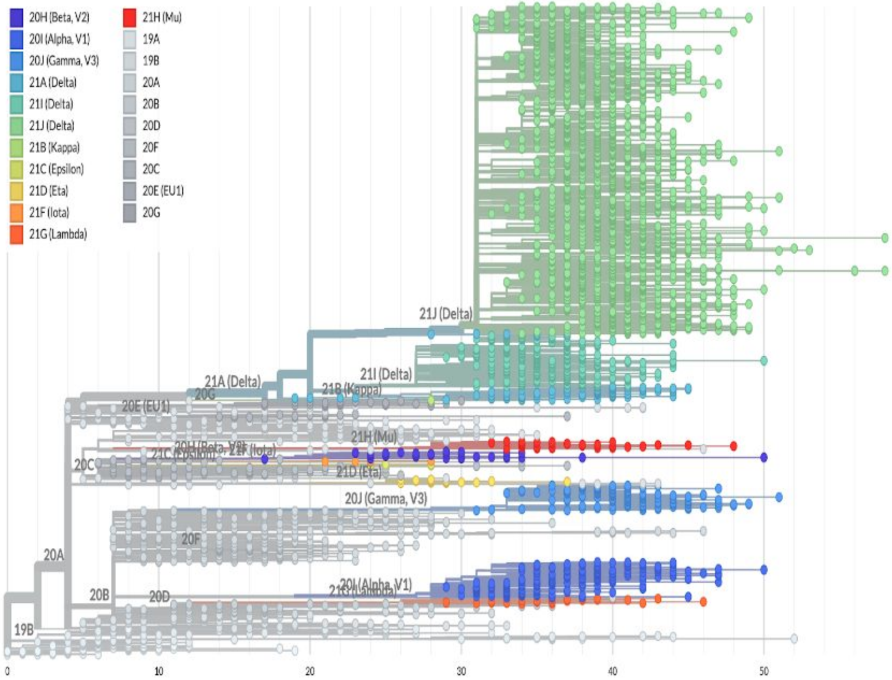


Mutation rate of SARS-CoV-2

3 x 10^{-6}/nt/cycle^2

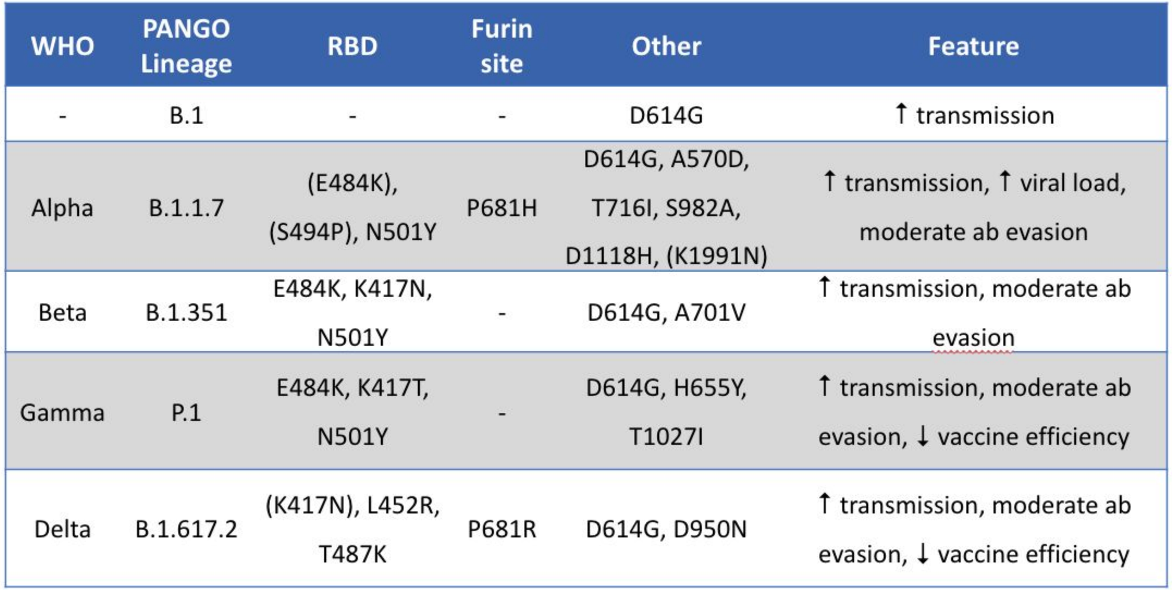
2 – 3 nt/month

Source: Duffy *et al.* (2018). *PLOS Biology;* ² Borges *et al.* (2021). *Evolutionary Biology*



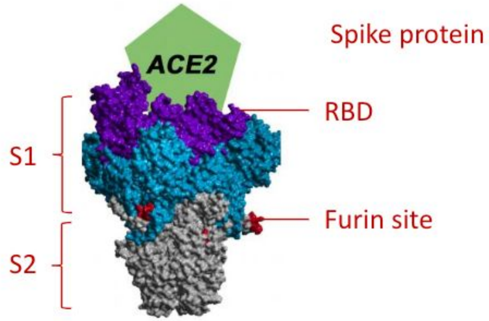
Source: Nextstrain. (2021). *Genomic epidemiology of novel coronavirus.*

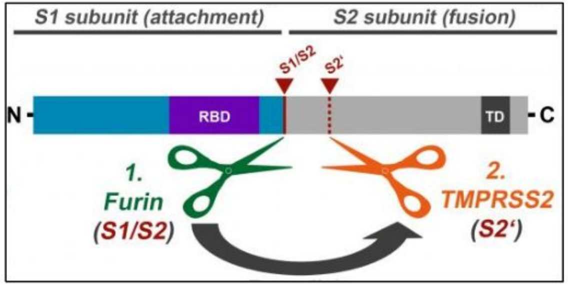
SARS-CoV-2 virus



WHO	PANGO Lineage	RBD	Furin site	Other	Feature
-	B.1	-	-	D614G	↑ transmission
Alpha	B.1.1.7	(E484K), (S494P), N501Y	P681H	D614G, A570D, T716I, S982A, D1118H, (K1991N)	↑ transmission, ↑ viral load, moderate ab evasion
Beta	B.1.351	E484K, K417N, N501Y	-	D614G, A701V	↑ transmission, moderate ab evasion
Gamma	P.1	E484K, K417T, N501Y	-	D614G, H655Y, T1027I	↑ transmission, moderate ab evasion, ↓ vaccine efficiency
Delta	B.1.617.2	(K417N), L452R, T487K	P681R	D614G, D950N	↑ transmission, moderate ab evasion, ↓ vaccine efficiency

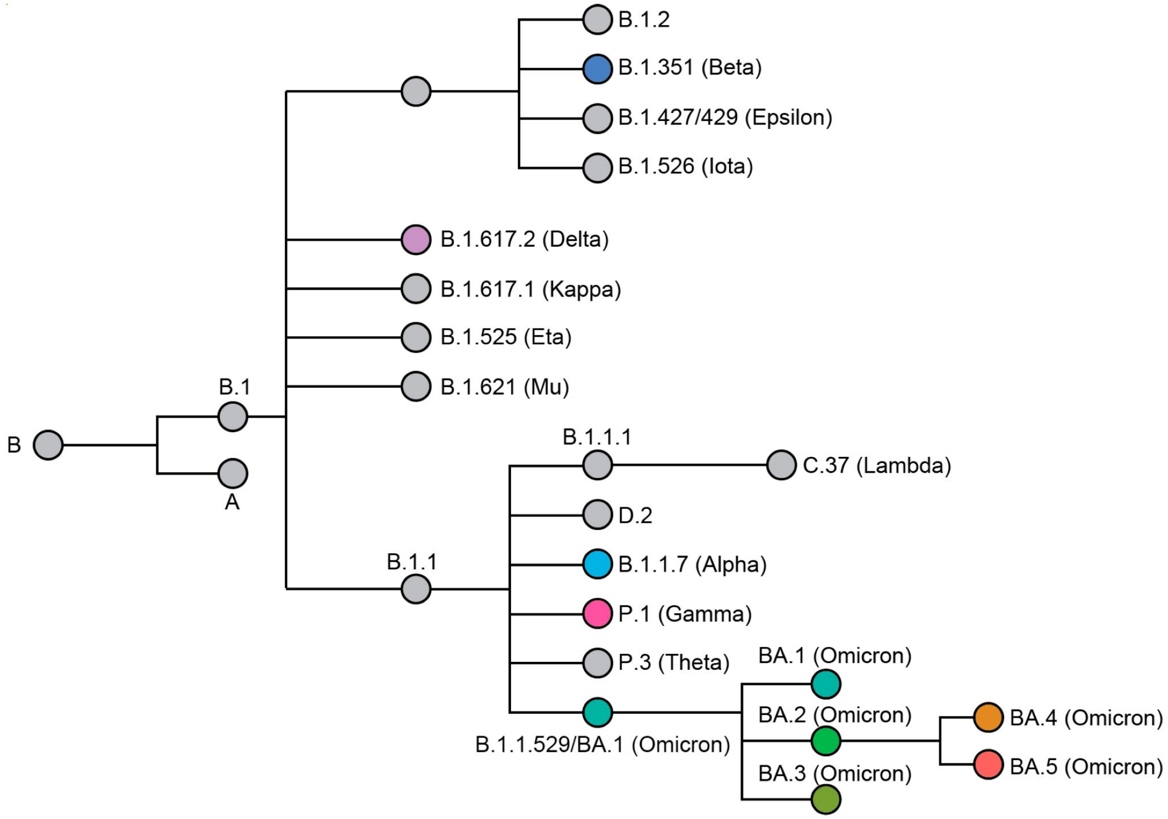
Source: Jackson *et al.* (2021). *Nature Reviews.*



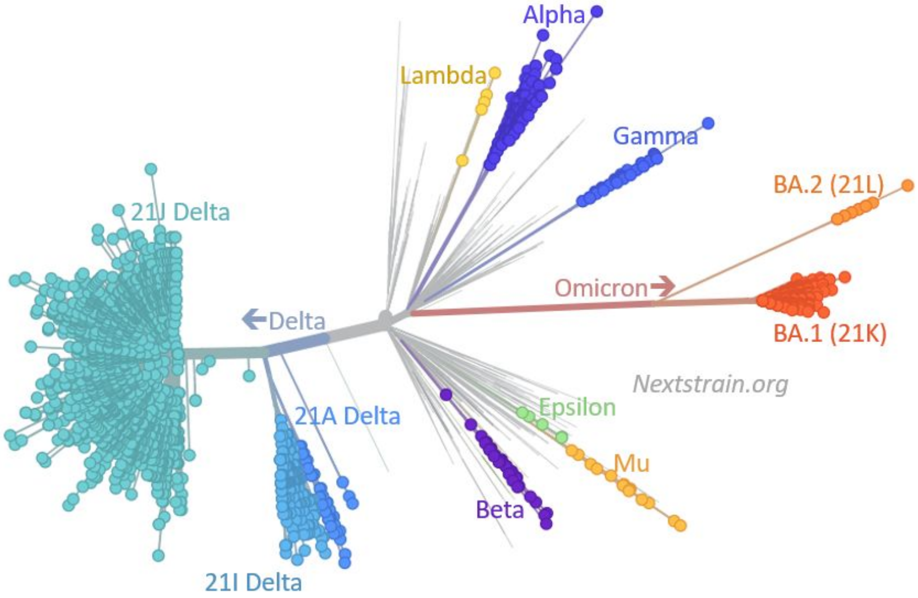


Source: Hoffman *et al.* (2020). *Molecular Cell.*

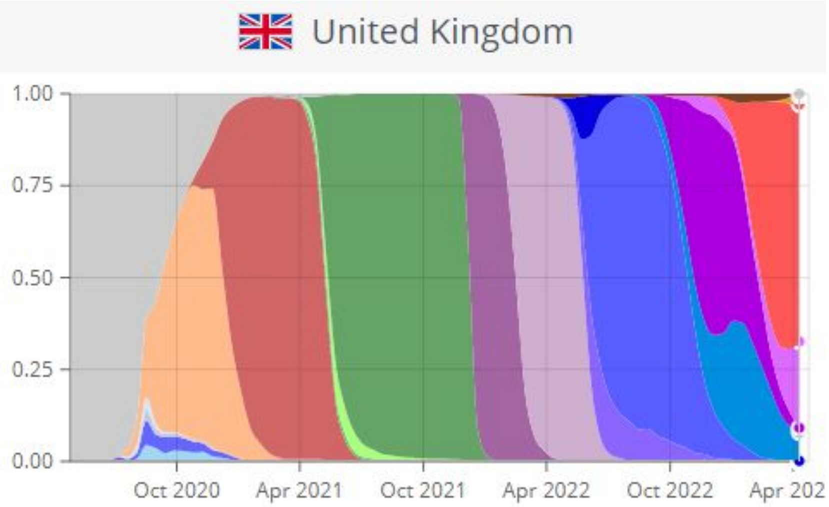
SARS-CoV-2 lineages and variants

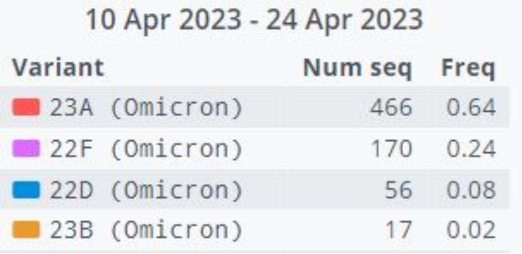


SARS-CoV-2 lineages and variants



Source: Hodcroft, 2022

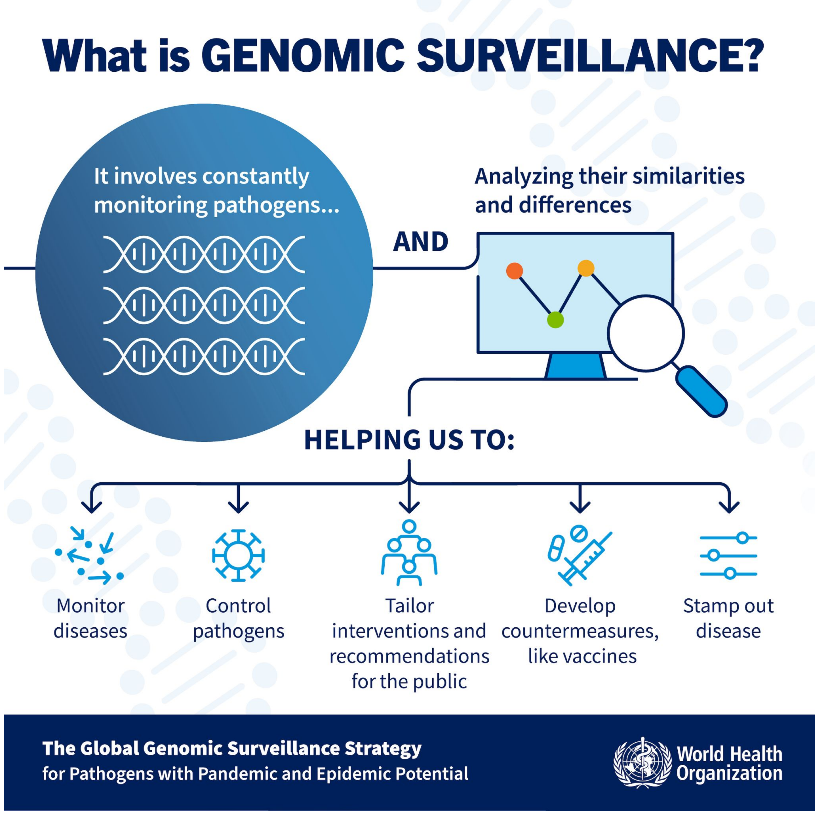




10 Apr 2023 - 24 Apr 2023

Variant	Num seq	Freq
🟥 23A (Omicron)	466	0.64
🟪 22F (Omicron)	170	0.24
🟦 22D (Omicron)	56	0.08
🟧 23B (Omicron)	17	0.02

Genomic surveillance



What is GENOMIC SURVEILLANCE?

It involves constantly monitoring pathogens…

AND

Analyzing their similarities and differences

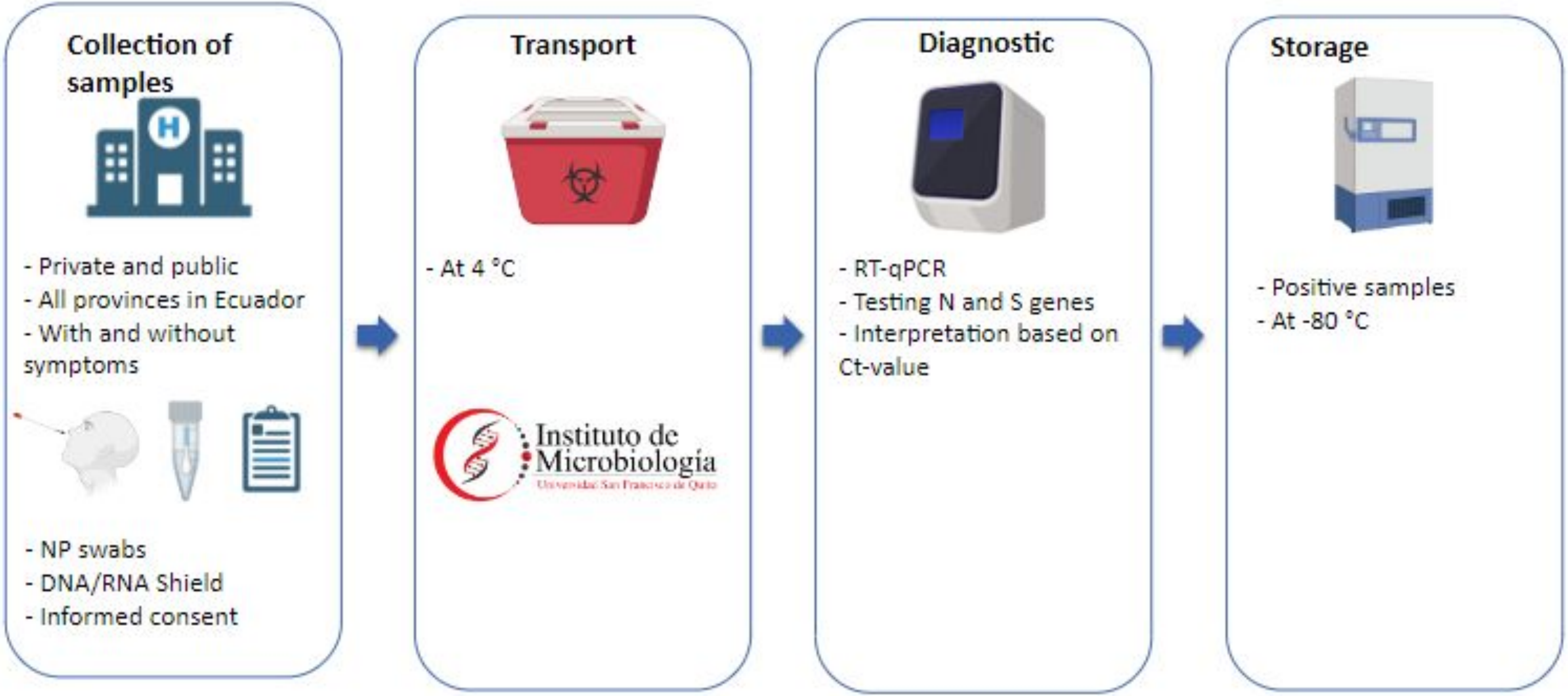
HELPING US TO:

- Monitor diseases
- Control pathogens
- Tailor interventions and recommendations for the public
- Develop countermeasures, like vaccines
- Stamp out disease

The Global Genomic Surveillance Strategy
for Pathogens with Pandemic and Epidemic Potential

World Health Organization

Workflow



Collection of samples

- Private and public
- All provinces in Ecuador
- With and without symptoms

- NP swabs
- DNA/RNA Shield
- Informed consent

Transport

- At 4 °C

Instituto de Microbiología
Universidad San Francisco de Quito

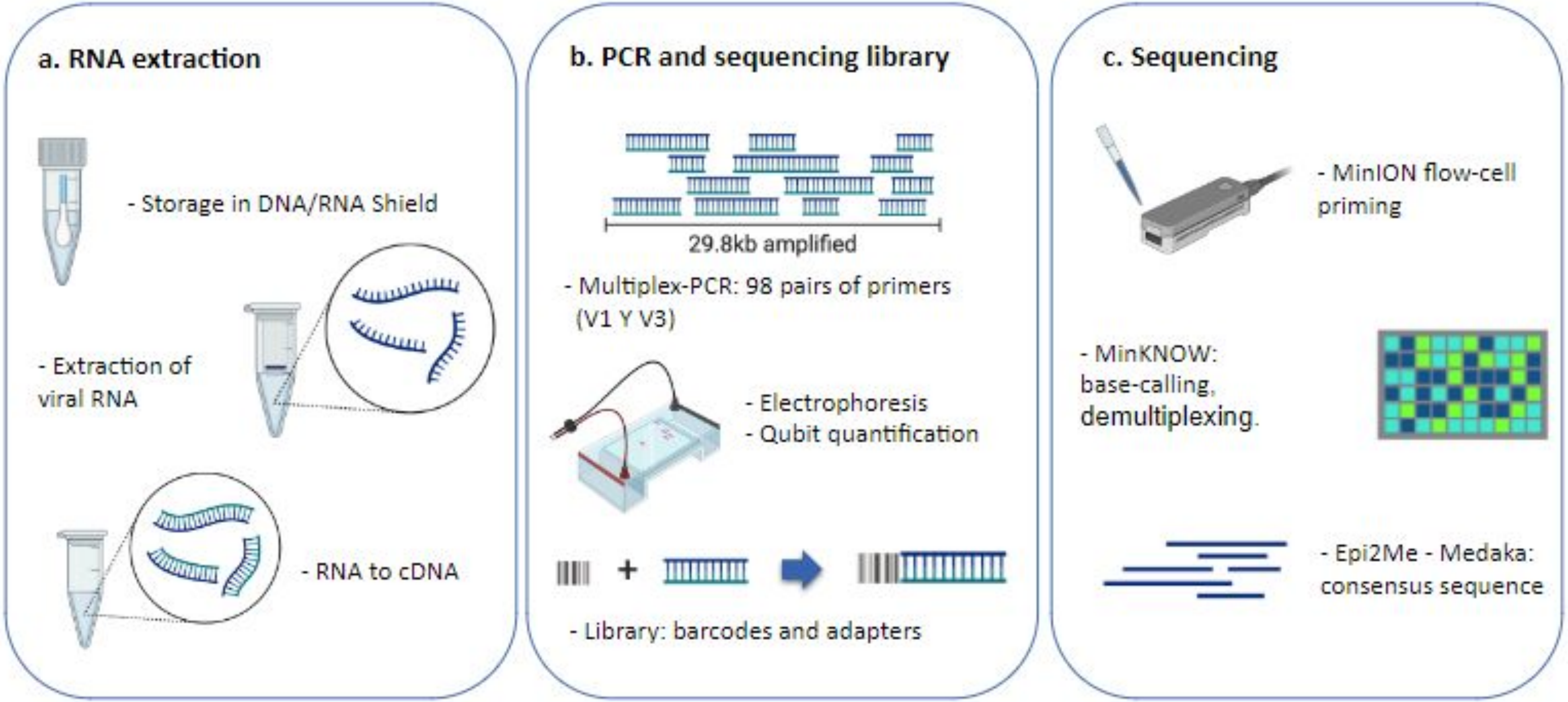
Diagnostic

- RT-qPCR
- Testing N and S genes
- Interpretation based on Ct-value

Storage

- Positive samples
- At -80 °C

Workflow



a. RNA extraction

- Storage in DNA/RNA Shield
- Extraction of viral RNA
- RNA to cDNA

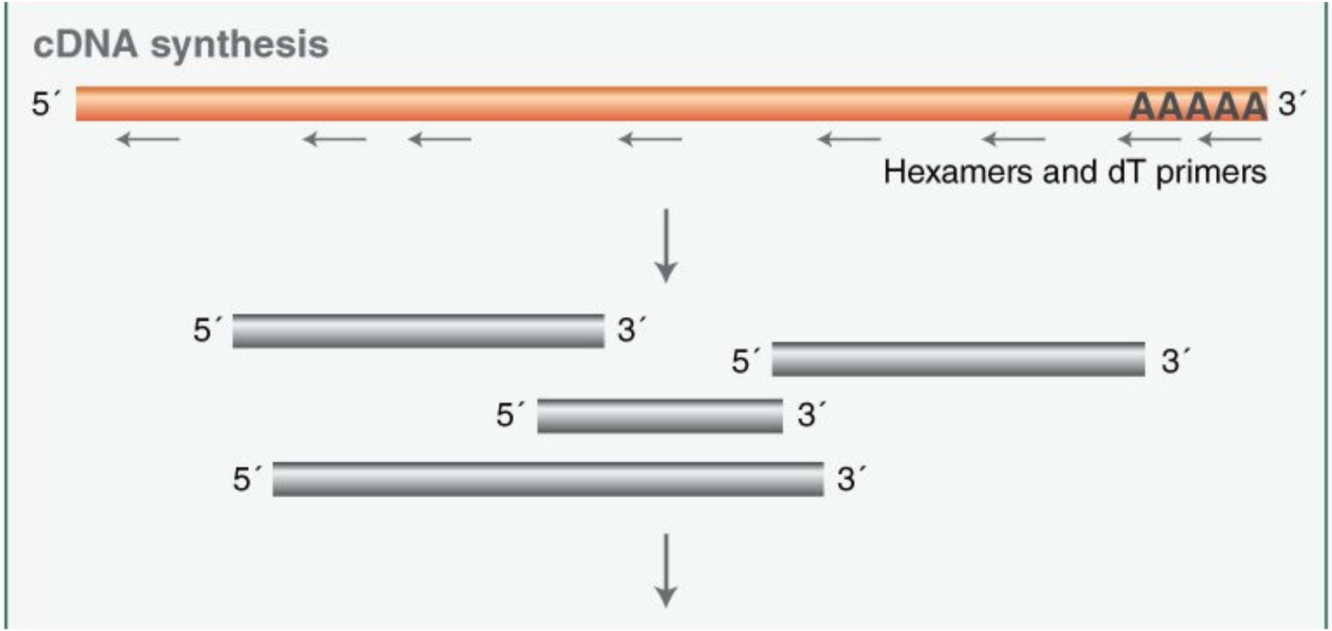
b. PCR and sequencing library

- Multiplex-PCR: 98 pairs of primers (V1 Y V3)
- Electrophoresis
- Qubit quantification
- Library: barcodes and adapters

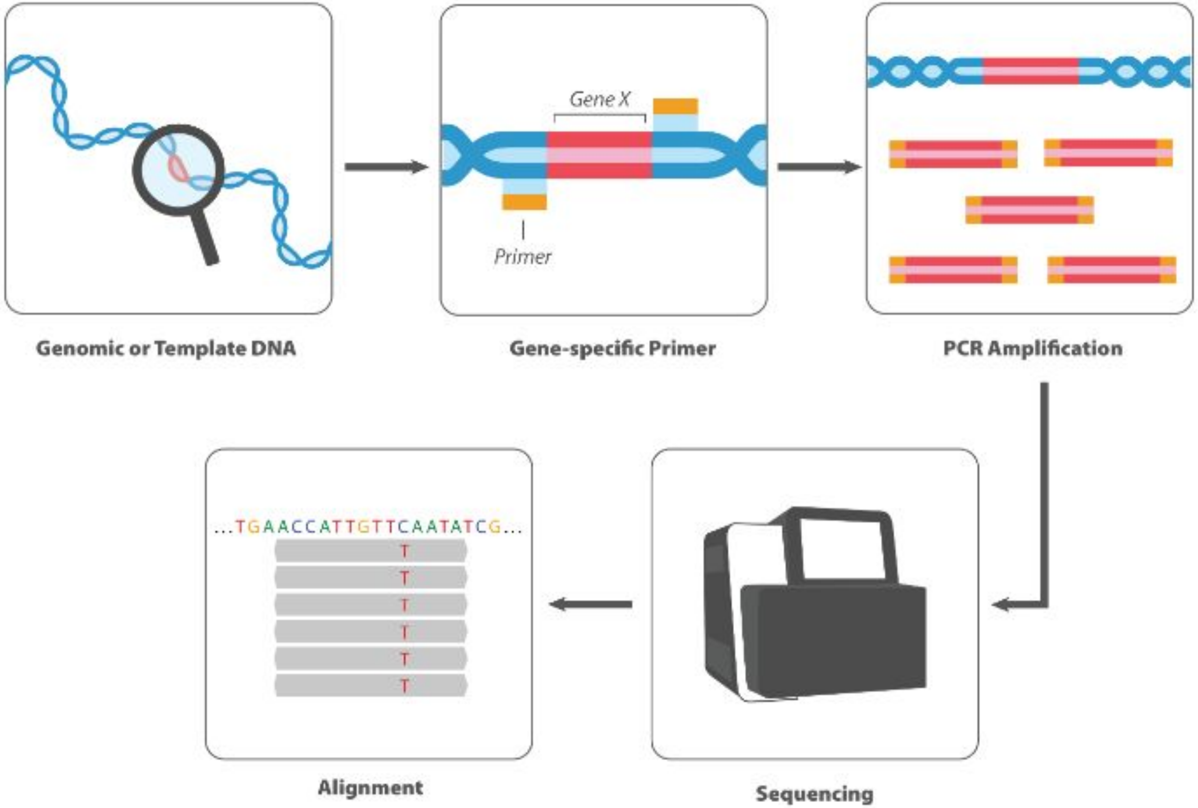
c. Sequencing

- MinION flow-cell priming
- MinKNOW: base-calling, demultiplexing.
- Epi2Me - Medaka: consensus sequence

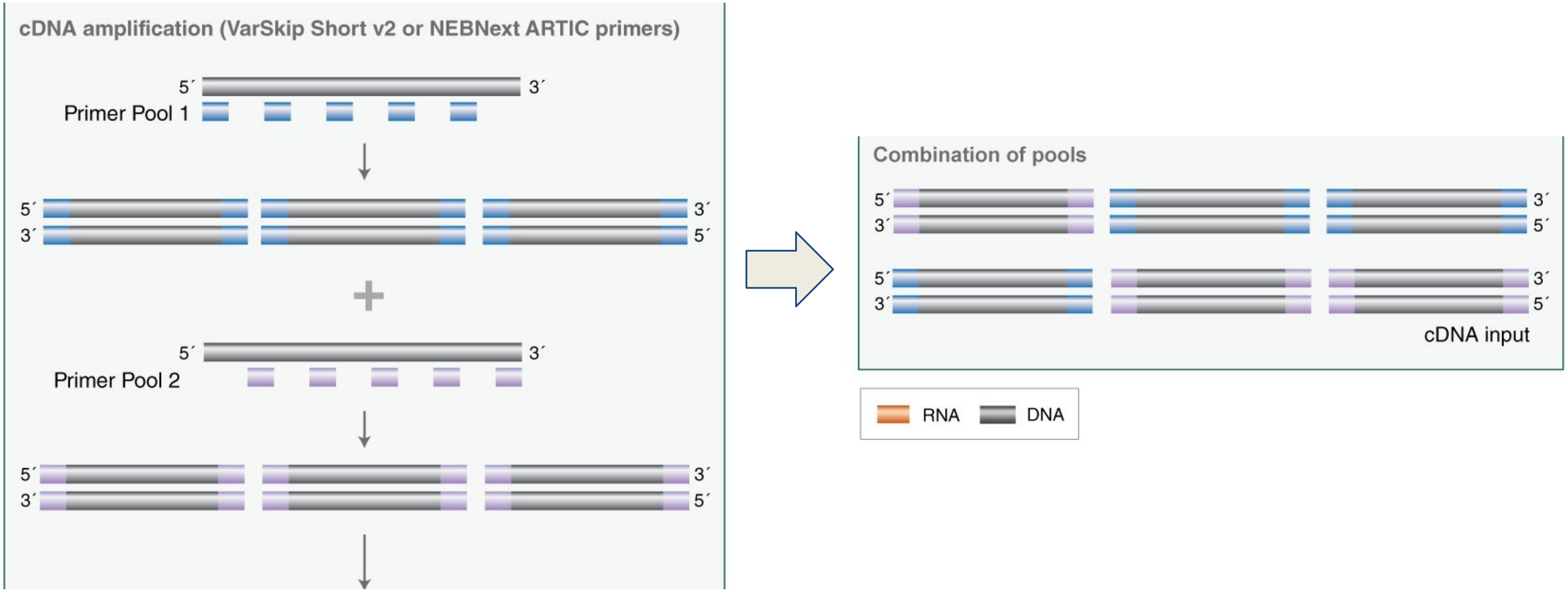
RNA to cDNA



Amplicon-based sequencing

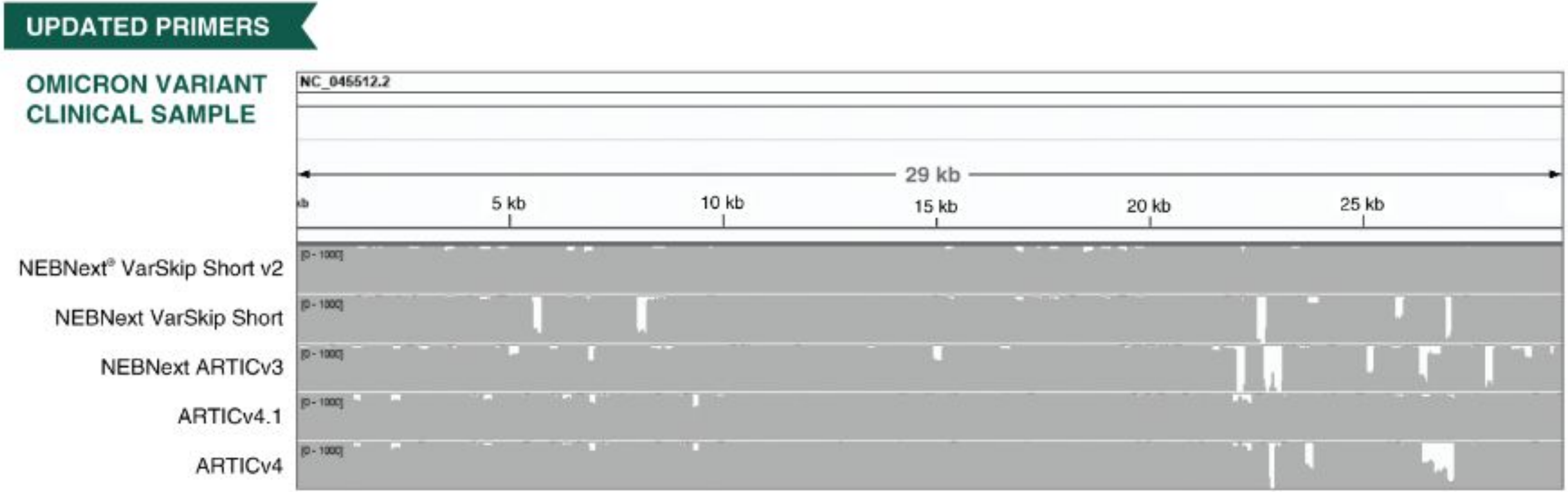


Amplicon-based sequencing



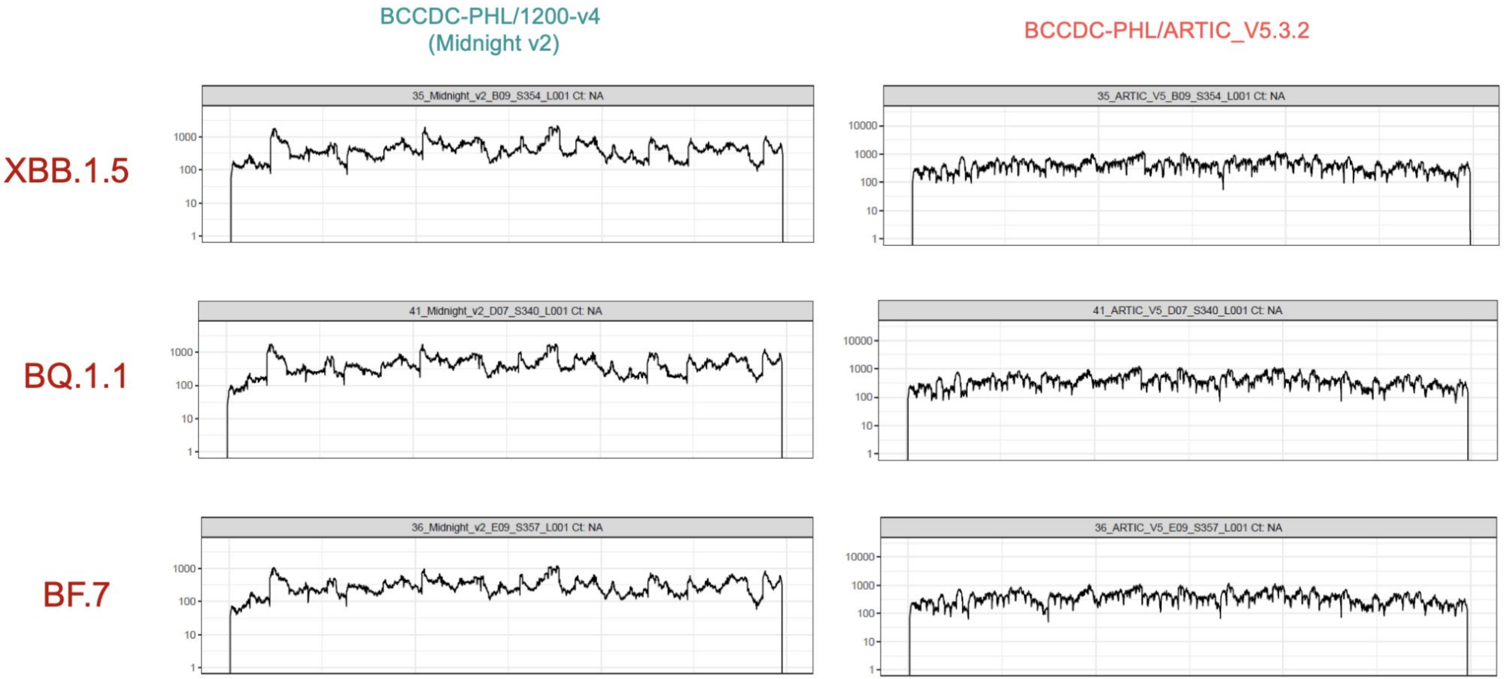
SARS-CoV-2 whole-genome primers

New England Biolabs



SARS-CoV-2 whole-genome primers

ARTIC network



BCCDC-PHL/1200-v4 (Midnight v2)

BCCDC-PHL/ARTIC_V5.3.2

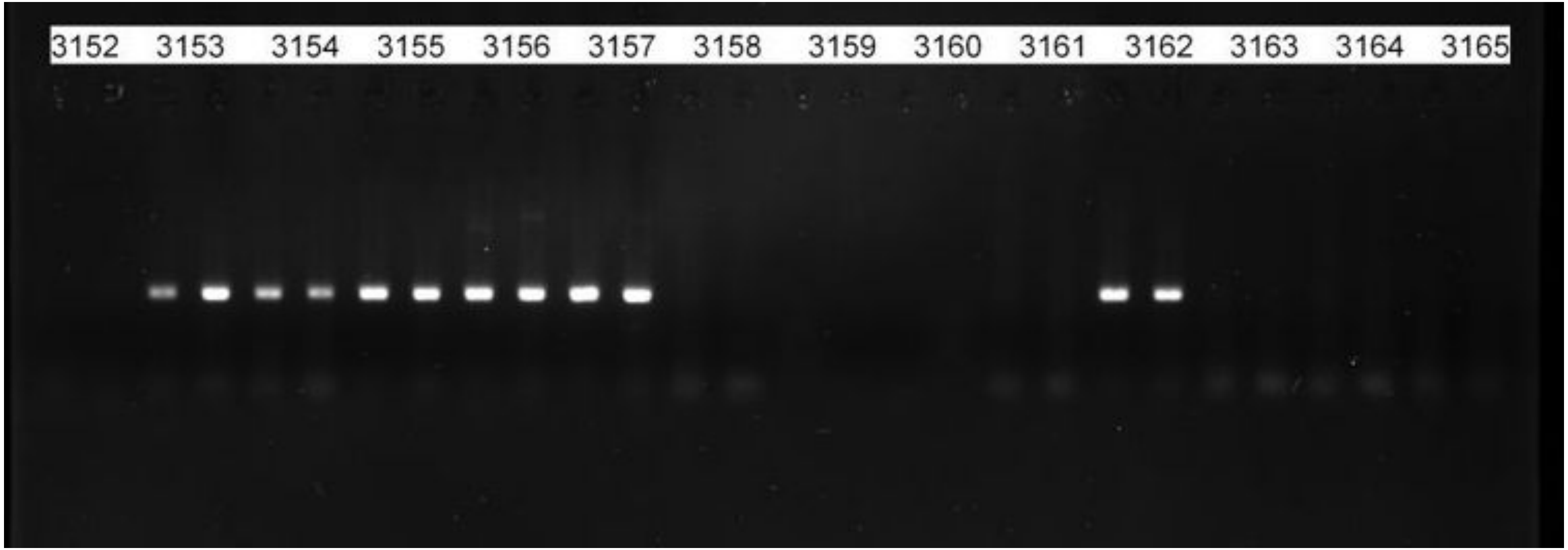
XBB.1.5

BQ.1.1

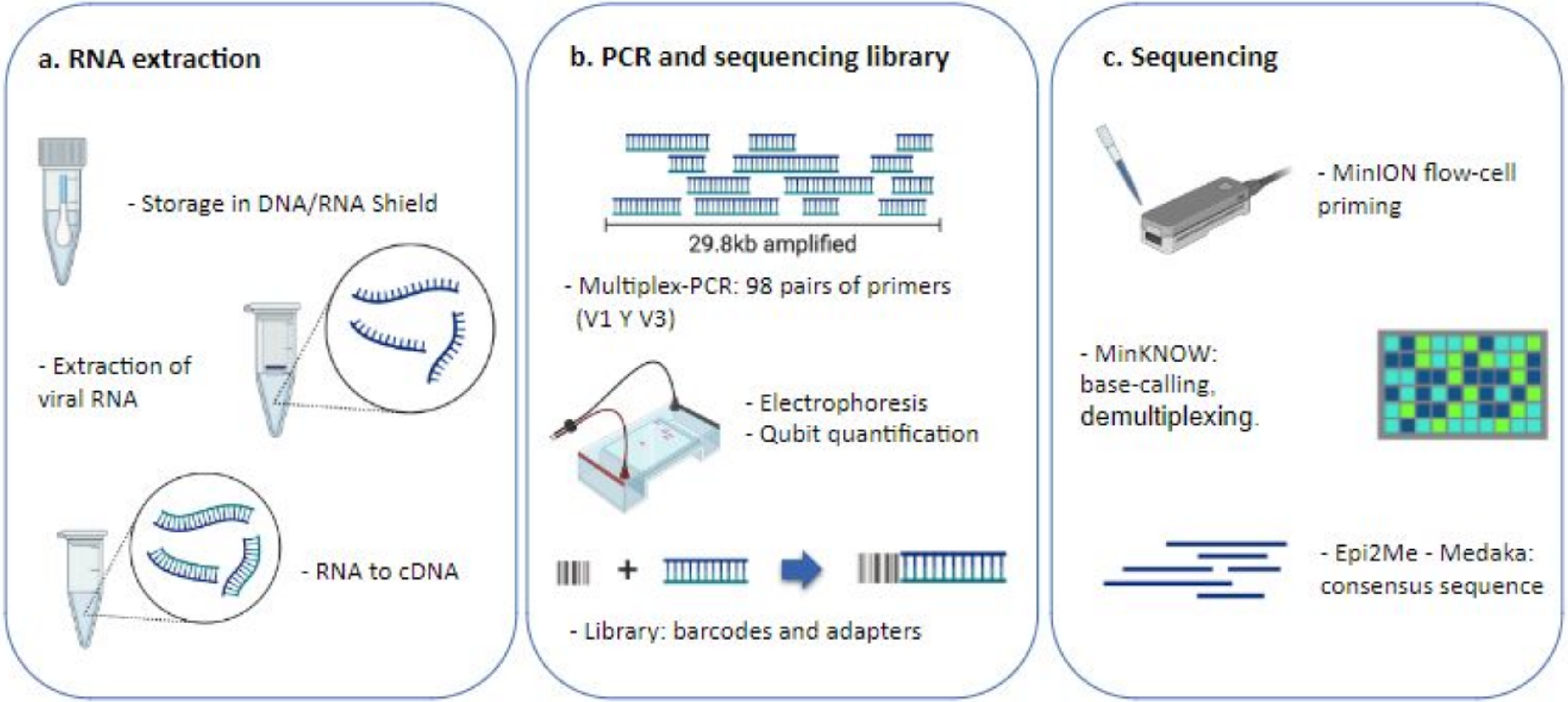
BF.7

Source: https://community.artic.network/t/sars-cov-2-version-5-3-2-scheme-release/462

Agarose gel electrophoresis



Workflow



a. RNA extraction

- Storage in DNA/RNA Shield

- Extraction of viral RNA

- RNA to cDNA

b. PCR and sequencing library

29.8kb amplified

- Multiplex-PCR: 98 pairs of primers (V1 Y V3)

- Electrophoresis
- Qubit quantification

- Library: barcodes and adapters

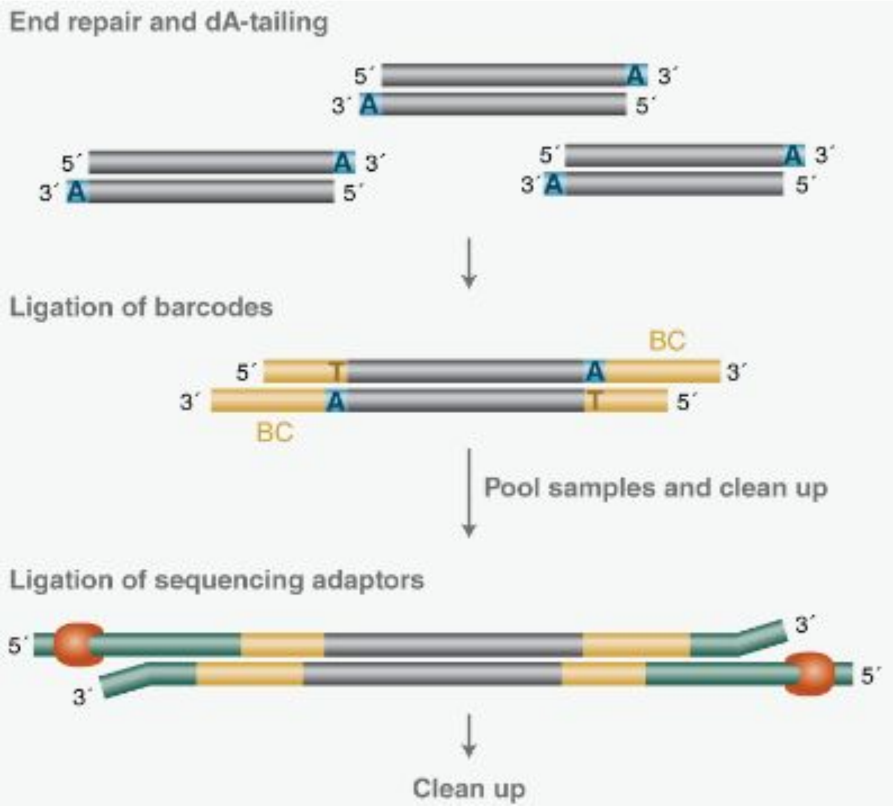
c. Sequencing

- MinION flow-cell priming

- MinKNOW: base-calling, demultiplexing.

- Epi2Me - Medaka: consensus sequence

Library preparation - Nanopore

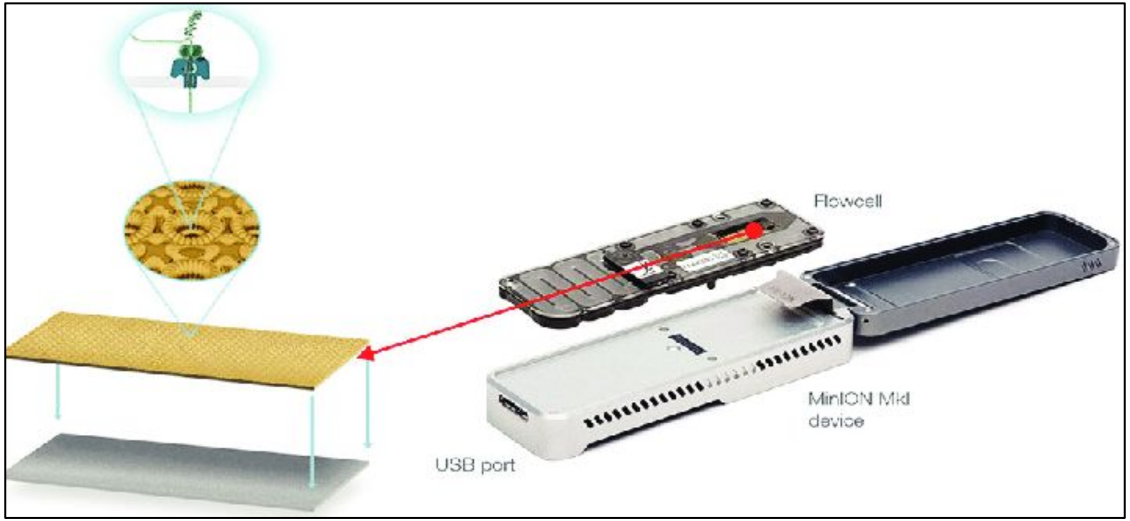

End repair and dA-tailing
5´
A
3´
3´
A
5´
Ligation of barcodes
BC
5´
T
A
3´
3´
A
T
5´
BC
Pool samples and clean up
Ligation of sequencing adaptors
5´
3´
3´
5´
Clean up

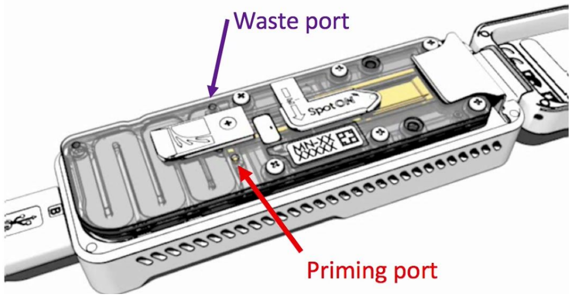
🟧 RNA	🟨 Barcode (BC)	● Motor protein
▬ DNA	🟩 Adaptor	

1	2	3	4	5	6	7	8	9	10	11	12	
BC01	BC02	BC03	BC04	BC05	BC06	BC07	BC08	BC09	BC10	BC11	BC12	A
BC13	BC14	BC15	BC16	BC17	BC18	BC19	BC20	BC21	BC22	BC23	BC24	B
BC25	BC26	BC27	BC28	BC29	BC30	BC31	BC32	BC33	BC34	BC35	BC36	C
BC37	BC38	BC39	BC40	BC41	BC42	BC43	BC44	BC45	BC46	BC47	BC48	D
BC49	BC50	BC51	BC52	BC53	BC54	BC55	BC56	BC57	BC58	BC59	BC60	E
BC61	BC62	BC63	BC64	BC65	BC66	BC67	BC68	BC69	BC70	BC71	BC72	F
BC73	BC74	BC75	BC76	BC77	BC78	BC79	BC80	BC81	BC82	BC83	BC84	G
BC85	BC86	BC87	BC88	BC89	BC90	BC91	BC92	BC93	BC94	BC95	BC96	H

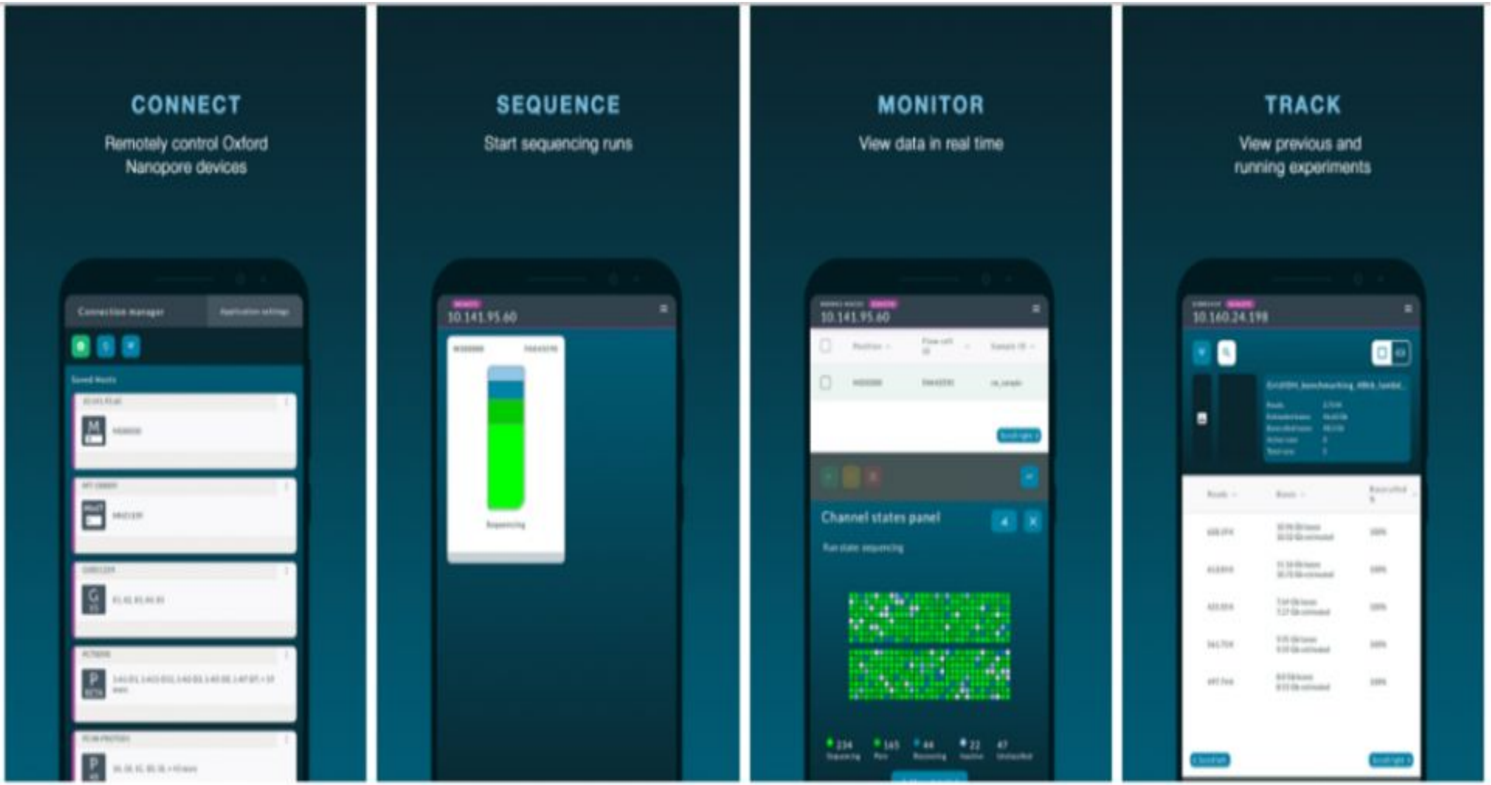
BCxx: Barcode xx

Flowcell loading

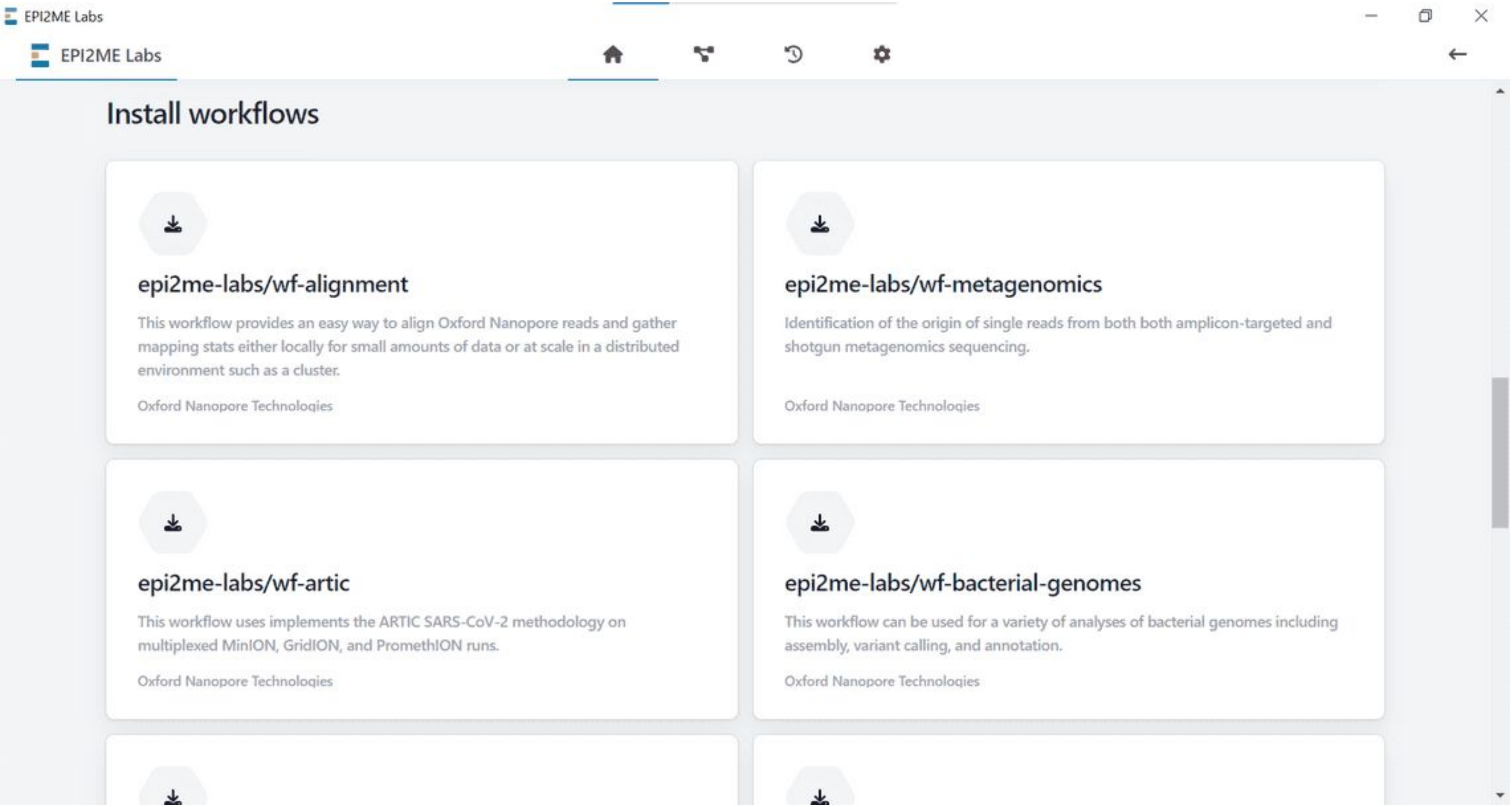




MinKNOW software



EPI2ME software



Install workflows

epi2me-labs/wf-alignment

This workflow provides an easy way to align Oxford Nanopore reads and gather mapping stats either locally for small amounts of data or at scale in a distributed environment such as a cluster.

Oxford Nanopore Technologies

epi2me-labs/wf-metagenomics

Identification of the origin of single reads from both both amplicon-targeted and shotgun metagenomics sequencing.

Oxford Nanopore Technologies

epi2me-labs/wf-artic

This workflow uses implements the ARTIC SARS-CoV-2 methodology on multiplexed MinION, GridION, and PromethION runs.

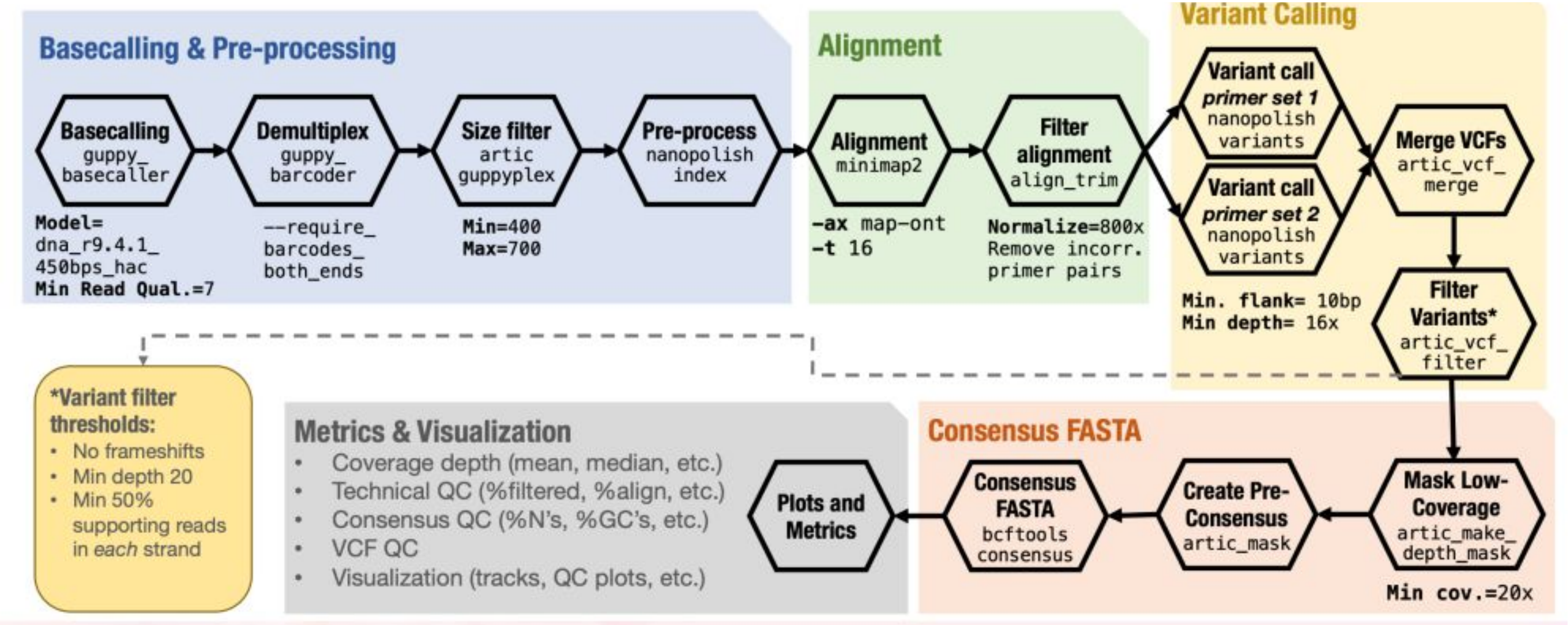
Oxford Nanopore Technologies

epi2me-labs/wf-bacterial-genomes

This workflow can be used for a variety of analyses of bacterial genomes including assembly, variant calling, and annotation.

Oxford Nanopore Technologies

ARTIC Pipeline



*Variant filter thresholds:

- No frameshifts
- Min depth 20
- Min 50% supporting reads in *each* strand

Metrics & Visualization

- Coverage depth (mean, median, etc.)
- Technical QC (%filtered, %align, etc.)
- Consensus QC (%N's, %GC's, etc.)
- VCF QC
- Visualization (tracks, QC plots, etc.)

Workflow

Lineage assignment

Assignment of clades and lineages

Nextclade

GISAID

Mutations and substitutions

Phylogeny

Alignment

MAFFT version 7

- Wuhan-Hu-1 reference genome
- Aliview: check alignment

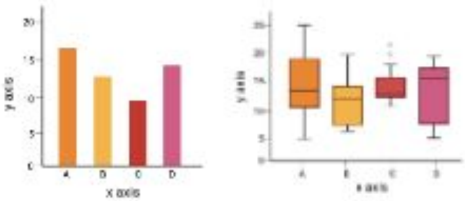
Phylogenetic tree construction

IQ-TREE

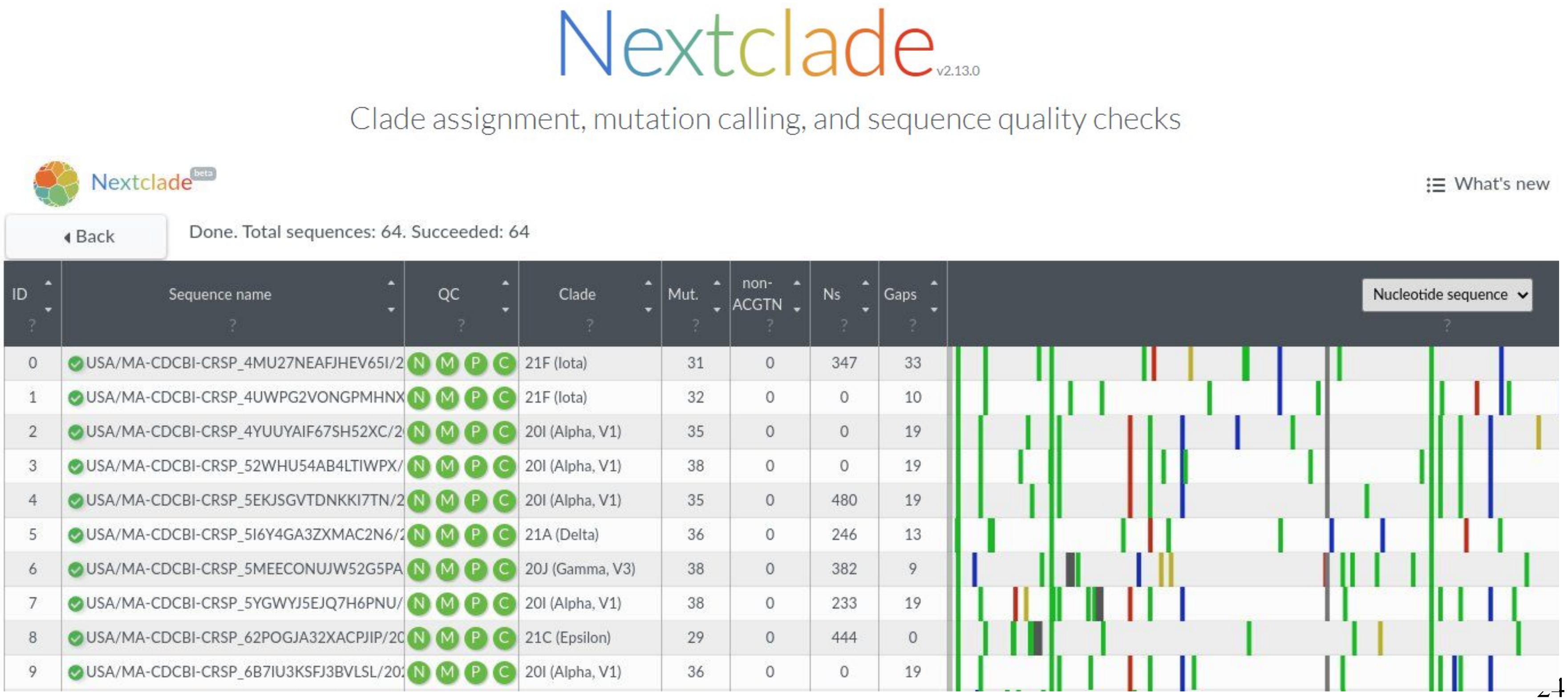
- Maximum Likelihood, GTR, 1000x
- Visualization in iTOL

Statistics

R Studio



Nextclade - Nextstrain



Nextclade v2.13.0

Clade assignment, mutation calling, and sequence quality checks

Nextclade beta

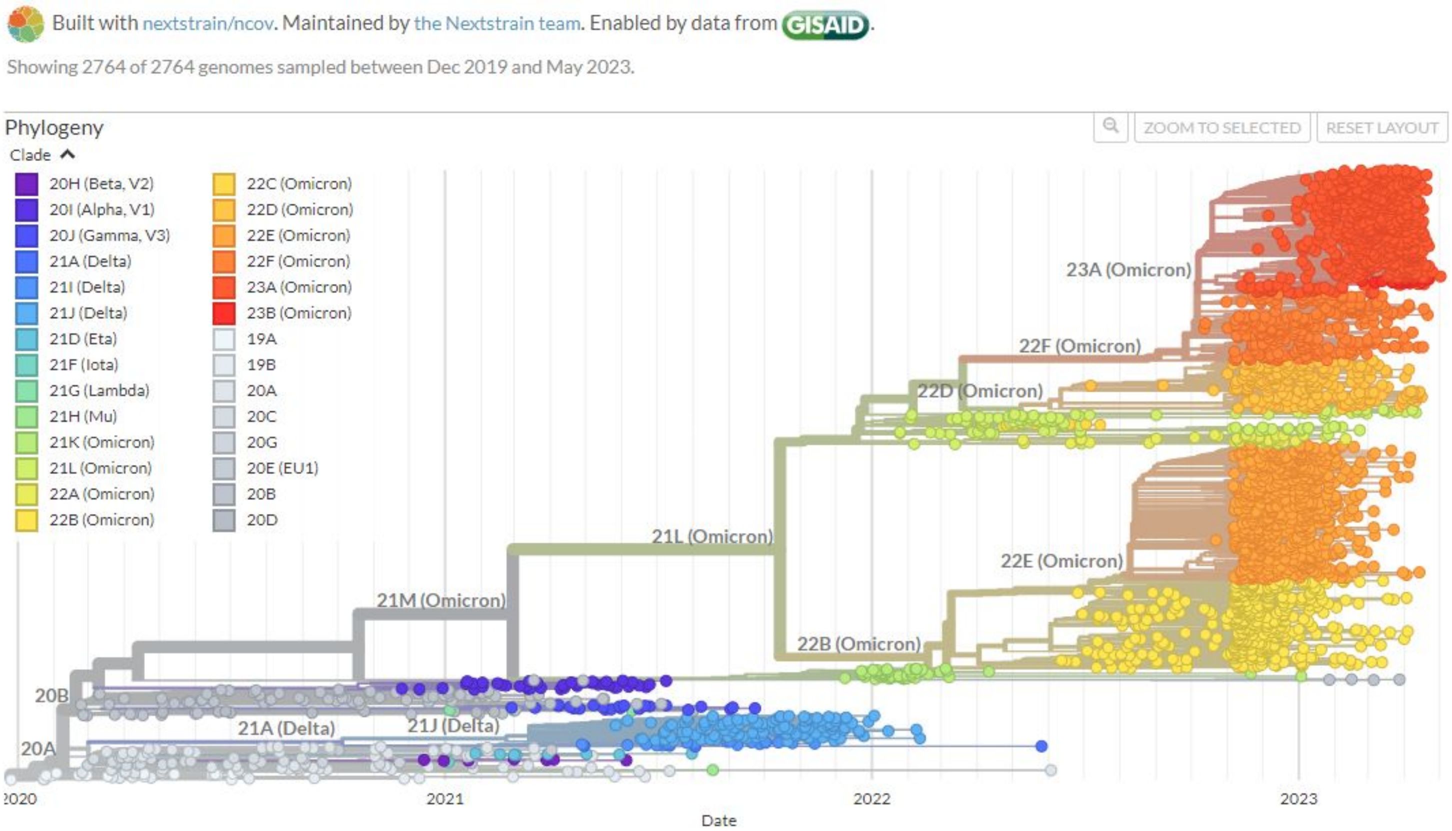
☰ What's new

◂ Back

Done. Total sequences: 64. Succeeded: 64

ID	Sequence name	QC	Clade	Mut.	non-ACGTN	Ns	Gaps	Nucleotide sequence
0	✓ USA/MA-CDCBI-CRSP_4MU27NEAFJHEV65I/2	N M P C	21F (Iota)	31	0	347	33	
1	✓ USA/MA-CDCBI-CRSP_4UWPG2VONGPMHNX	N M P C	21F (Iota)	32	0	0	10	
2	✓ USA/MA-CDCBI-CRSP_4YUUYAIF67SH52XC/2	N M P C	20I (Alpha, V1)	35	0	0	19	
3	✓ USA/MA-CDCBI-CRSP_52WHU54AB4LTIWPX/	N M P C	20I (Alpha, V1)	38	0	0	19	
4	✓ USA/MA-CDCBI-CRSP_5EKJSGVTDNKKI7TN/2	N M P C	20I (Alpha, V1)	35	0	480	19	
5	✓ USA/MA-CDCBI-CRSP_5I6Y4GA3ZXMAC2N6/2	N M P C	21A (Delta)	36	0	246	13	
6	✓ USA/MA-CDCBI-CRSP_5MEECONUJW52G5PA	N M P C	20J (Gamma, V3)	38	0	382	9	
7	✓ USA/MA-CDCBI-CRSP_5YGWYJ5EJQ7H6PNU/	N M P C	20I (Alpha, V1)	38	0	233	19	
8	✓ USA/MA-CDCBI-CRSP_62POGJA32XACPJIP/20	N M P C	21C (Epsilon)	29	0	444	0	
9	✓ USA/MA-CDCBI-CRSP_6B7IU3KSFJ3BVLSL/20	N M P C	20I (Alpha, V1)	36	0	0	19	

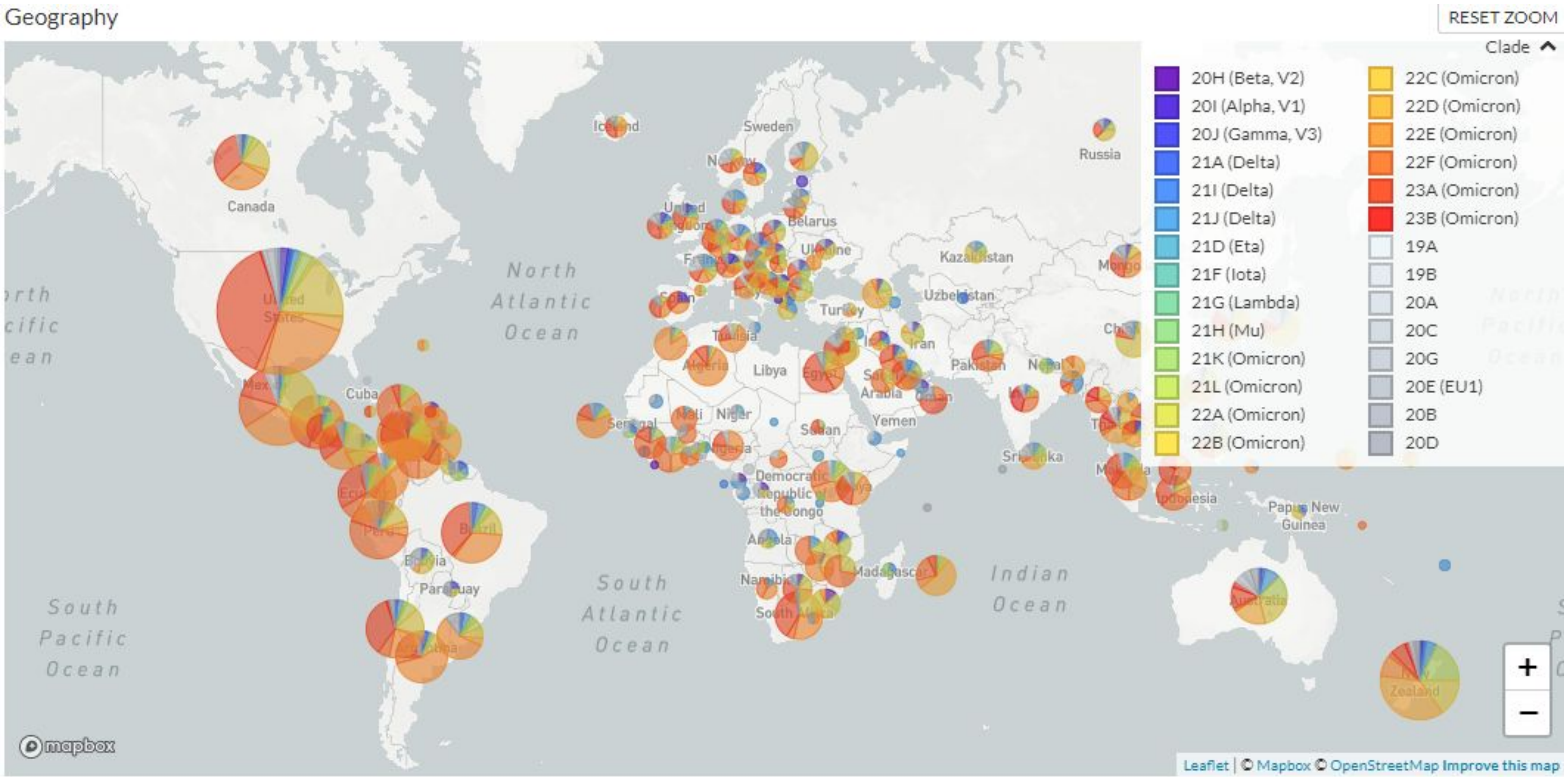
Nextclade



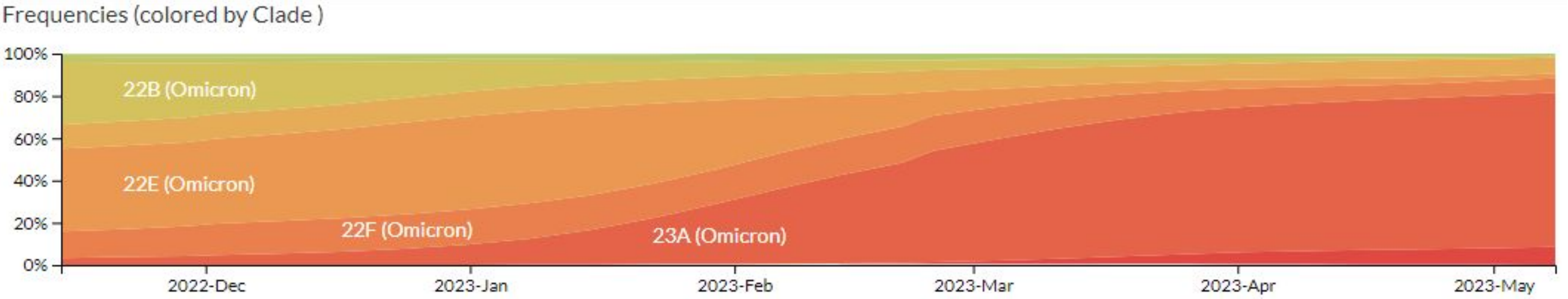
Built with nextstrain/ncov. Maintained by the Nextstrain team. Enabled by data from GISAID.

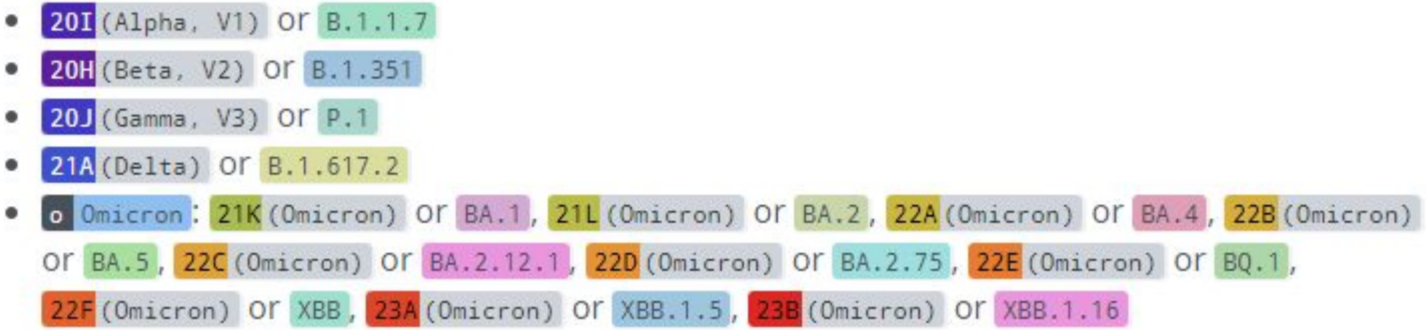
Showing 2764 of 2764 genomes sampled between Dec 2019 and May 2023.

Nextclade


Geography
RESET ZOOM
Clade ^
20H (Beta, V2)
20I (Alpha, V1)
20J (Gamma, V3)
21A (Delta)
21I (Delta)
21J (Delta)
21D (Eta)
21F (Iota)
21G (Lambda)
21H (Mu)
21K (Omicron)
21L (Omicron)
22A (Omicron)
22B (Omicron)
22C (Omicron)
22D (Omicron)
22E (Omicron)
22F (Omicron)
23A (Omicron)
23B (Omicron)
19A
19B
20A
20C
20G
20E (EU1)
20B
20D
Canada
United States
Mexico
Cuba
North Atlantic Ocean
South Atlantic Ocean
South Pacific Ocean
Indian Ocean
Iceland
Sweden
Norway
United Kingdom
France
Spain
Belarus
Ukraine
Russia
Kazakhstan
Uzbekistan
Mongolia
China
Turkey
Tunisia
Algeria
Libya
Egypt
Saudi Arabia
Yemen
Iran
Oman
Pakistan
Nepal
Sri Lanka
Malaysia
Indonesia
Papua New Guinea
Australia
New Zealand
Senegal
Mali
Niger
Nigeria
Sudan
Democratic Republic of the Congo
Angola
Namibia
South Africa
Madagascar
Ecuador
Peru
Brazil
Bolivia
Paraguay
Argentina
mapbox
+
−
Leaflet | © Mapbox © OpenStreetMap Improve this map

Nextclade

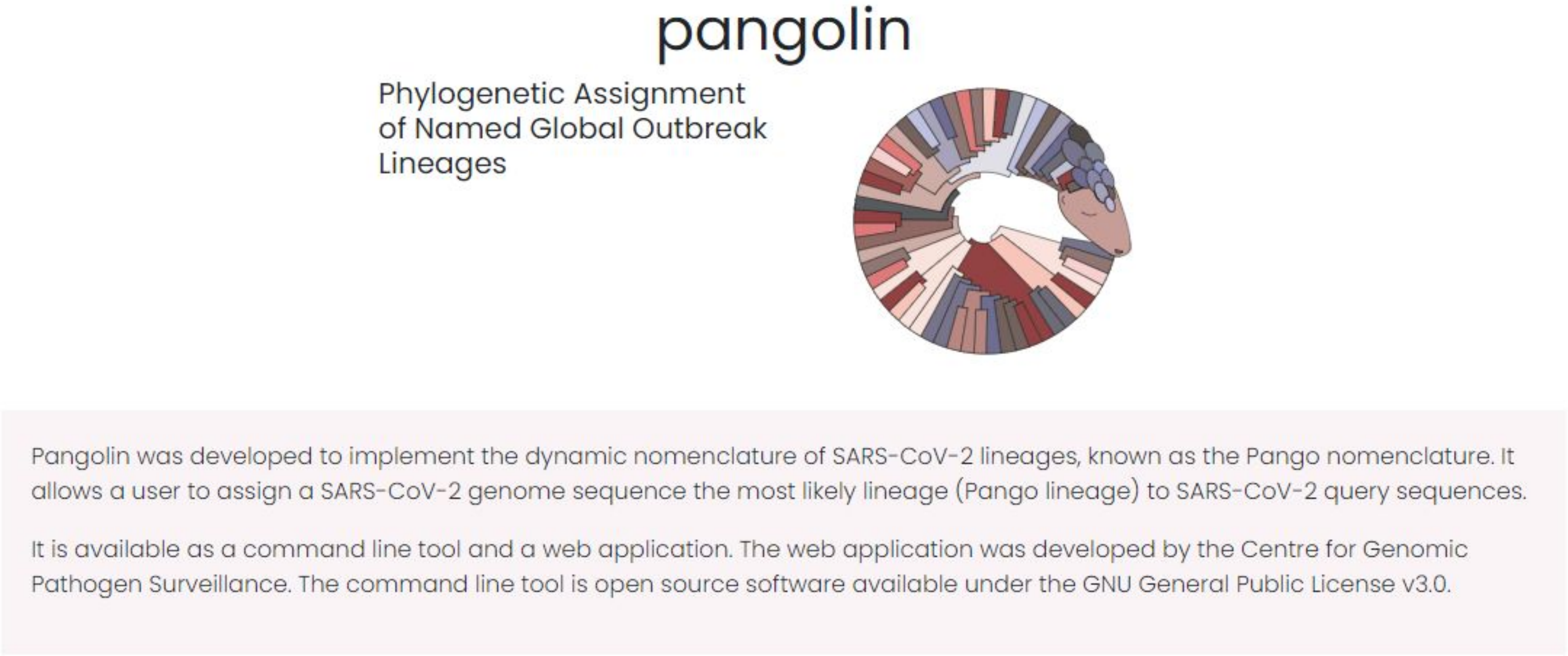




- 20I (Alpha, V1) or B.1.1.7
- 20H (Beta, V2) or B.1.351
- 20J (Gamma, V3) or P.1
- 21A (Delta) or B.1.617.2
- o Omicron: 21K (Omicron) or BA.1, 21L (Omicron) or BA.2, 22A (Omicron) or BA.4, 22B (Omicron) or BA.5, 22C (Omicron) or BA.2.12.1, 22D (Omicron) or BA.2.75, 22E (Omicron) or BQ.1, 22F (Omicron) or XBB, 23A (Omicron) or XBB.1.5, 23B (Omicron) or XBB.1.16

https://covariants.org/variants/

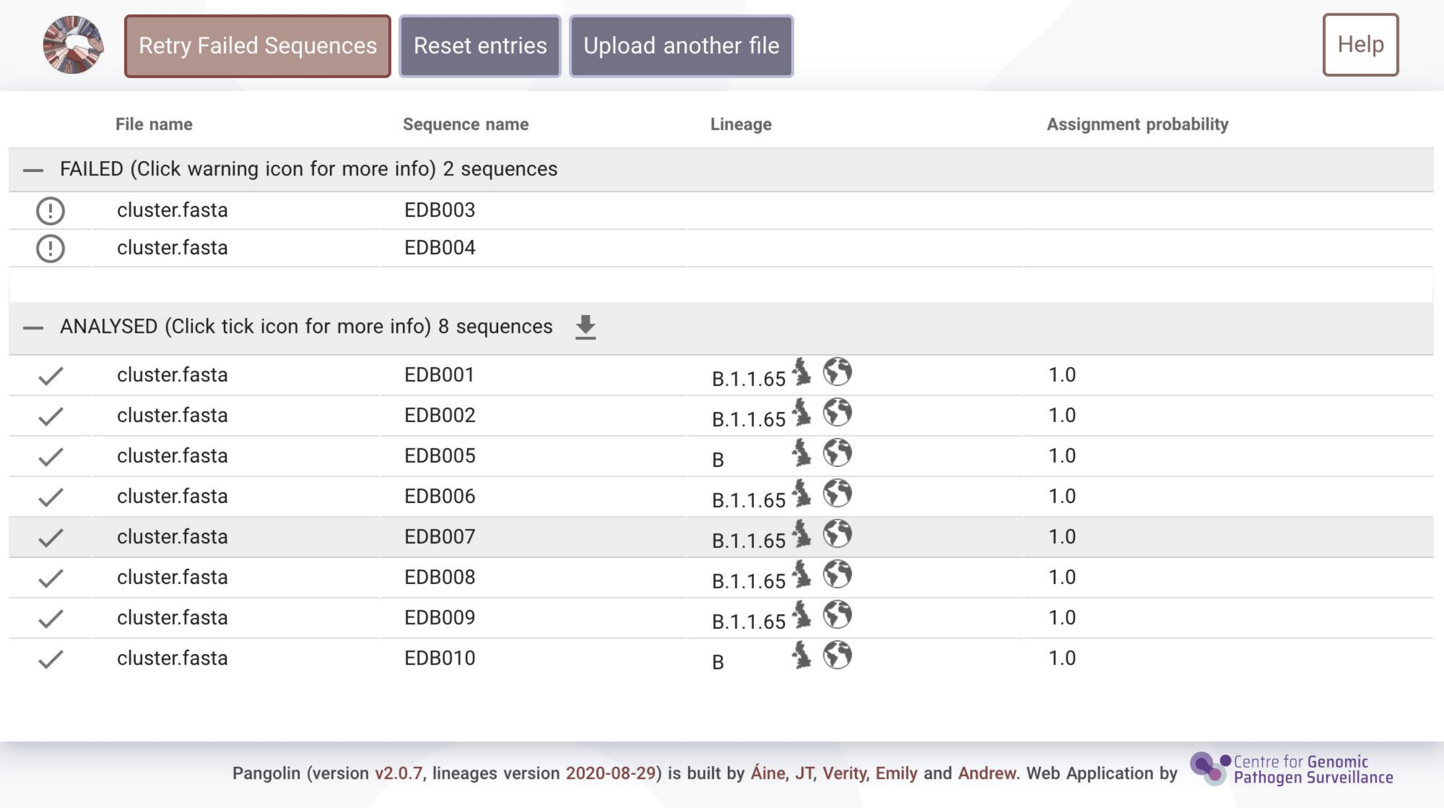
Pangolin COVID-19 Lineage assigner



Pangolin was developed to implement the dynamic nomenclature of SARS-CoV-2 lineages, known as the Pango nomenclature. It allows a user to assign a SARS-CoV-2 genome sequence the most likely lineage (Pango lineage) to SARS-CoV-2 query sequences.

It is available as a command line tool and a web application. The web application was developed by the Centre for Genomic Pathogen Surveillance. The command line tool is open source software available under the GNU General Public License v3.0.

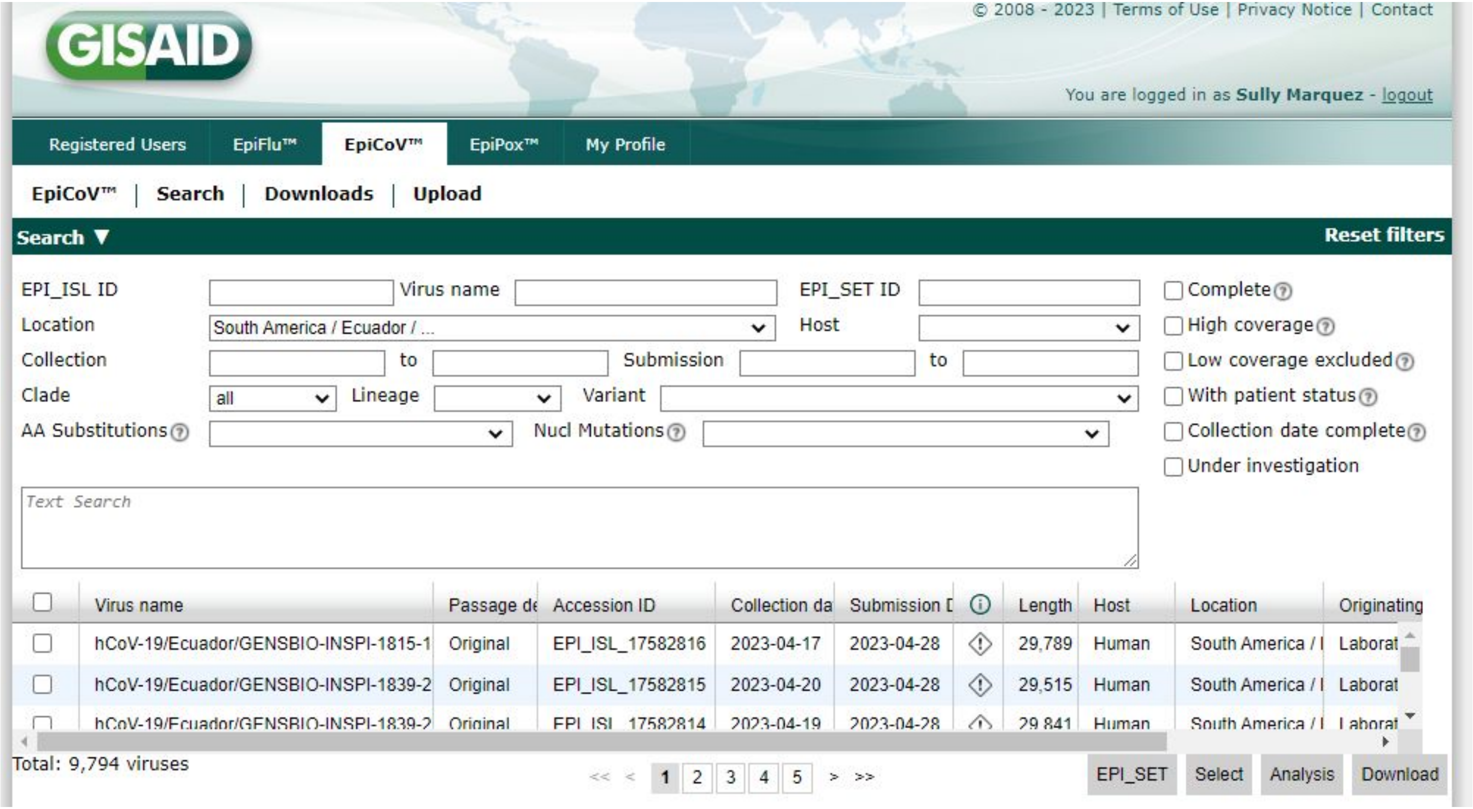
Pangolin COVID-19 Lineage assigner



	File name	Sequence name	Lineage	Assignment probability
—	FAILED (Click warning icon for more info) 2 sequences			
!	cluster.fasta	EDB003		
!	cluster.fasta	EDB004		
—	ANALYSED (Click tick icon for more info) 8 sequences			
✓	cluster.fasta	EDB001	B.1.1.65	1.0
✓	cluster.fasta	EDB002	B.1.1.65	1.0
✓	cluster.fasta	EDB005	B	1.0
✓	cluster.fasta	EDB006	B.1.1.65	1.0
✓	cluster.fasta	EDB007	B.1.1.65	1.0
✓	cluster.fasta	EDB008	B.1.1.65	1.0
✓	cluster.fasta	EDB009	B.1.1.65	1.0
✓	cluster.fasta	EDB010	B	1.0

Pangolin (version v2.0.7, lineages version 2020-08-29) is built by Áine, JT, Verity, Emily and Andrew. Web Application by Centre for Genomic Pathogen Surveillance

GISAID database



You are logged in as **Sully Marquez** - logout

EPI_ISL ID		Virus name		EPI_SET ID	
Location	South America / Ecuador / ...			Host	
Collection		to		Submission	to
Clade	all	Lineage		Variant	
AA Substitutions		Nucl Mutations			

- □ Complete
- □ High coverage
- □ Low coverage excluded
- □ With patient status
- □ Collection date complete
- □ Under investigation

□	Virus name	Passage de	Accession ID	Collection da	Submission D	ⓘ	Length	Host	Location	Originating
□	hCoV-19/Ecuador/GENSBIO-INSPI-1815-1	Original	EPI_ISL_17582816	2023-04-17	2023-04-28	⚠	29,789	Human	South America / I	Laborat
□	hCoV-19/Ecuador/GENSBIO-INSPI-1839-2	Original	EPI_ISL_17582815	2023-04-20	2023-04-28	⚠	29,515	Human	South America / I	Laborat
□	hCoV-19/Ecuador/GENSBIO-INSPI-1839-2	Original	EPI_ISL_17582814	2023-04-19	2023-04-28	⚠	29,841	Human	South America / I	Laborat

Total: 9,794 viruses