



SARS-CoV-2 genomic surveillance

Dikeledi Kekana

Centre for Respiratory Diseases and Meningitis,
National Institute for Communicable Diseases,
A division of the National Health Laboratory Service

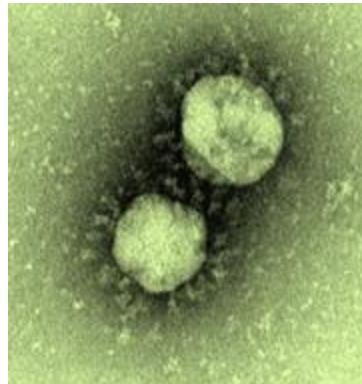


Image: Monica Birkhead,
CEZPD, NICD

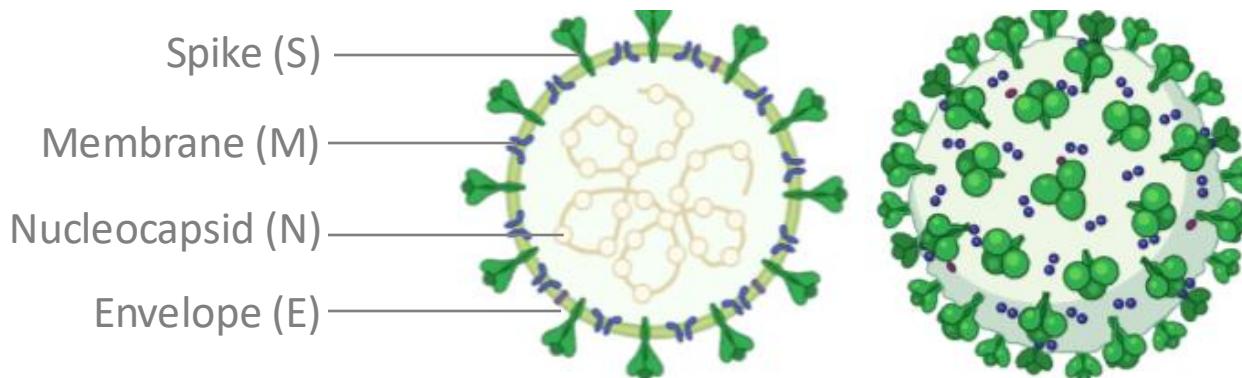
NATIONAL INSTITUTE FOR
COMMUNICABLE DISEASES

Division of the National Health Laboratory Service

- SARS-CoV-2 is a positive sense single-stranded RNA virus with a **~30 000bp genome**
- Encodes **12 functional open reading frames**, containing **11 coding regions** that make **12 proteins**
- Mutations occur along the length of the genome
 - Track genome to **identify mutations**
 - Track where mutations happen to **predict function**
 - Track mutations to select sequences for **experiments**

SARS-CoV-2 genome

- SARS-CoV-2 genome encodes 29 viral components
 - 4 structural proteins, 16 non-structural proteins, 9 accessory proteins





Viruses are constantly evolving

- A virus's 'aim' or 'goal' is to infect a host, replicate, and move on to infect other hosts
- Viral infection → **host immune response**
 - Host learns to recognize this virus → stronger immune response if re-infected with same virus
- **DNA/RNA replication machinery makes mistakes** → accumulation of mutations (**antigenic drift**)
- **Recombination** of different viruses → new strain with little to no immune protection (**antigenic shift**)
- The virus is therefore able to **accumulate mutations**
 - Mutations with a **fitness advantage** → more likely to survive and will be **selected for** (**positive selection**)
 - Mutations with a **fitness detriment** → less likely to survive and will be **selected against** (**negative selection**)
 - Mutations with **no effect** → neutral, **no selection**



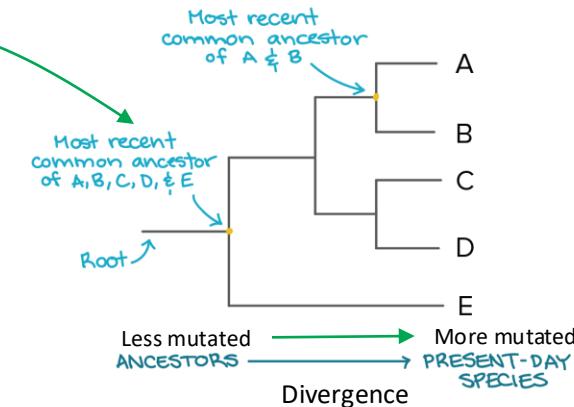
Mutations can have different impacts

- Majority of mutations have no impact (synonymous changes)
- Impacts mostly from non-synonymous mutations (i.e. change in nucleotide
→ change in amino acid)
- Increased **transmissibility** → infect more people
- Increased **replication fitness** → higher viral load → more infectious
- Increased **immune escape** → infect previously-infected or vaccinated people
- Increased **severity** → cause more serious disease in infected people
- Not all mutations will have an important impact, but some will
 - Need to monitor to detect these as early as possible

Viruses mutate at different rates

- **Molecular clock** hypothesis: nucleic acid **mutates** at a **relatively constant rate** over time

- Rate **differs between species**
- Often used to determine TMRCA of two species
- Can also be used to determine **how quickly mutations should accumulate**

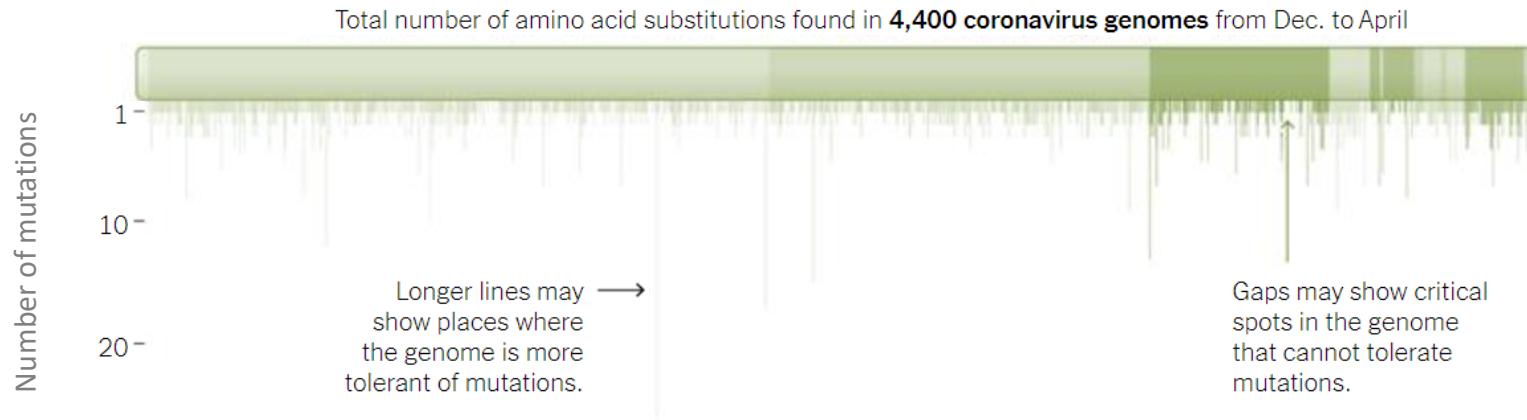


<https://www.khanacademy.org/science/ap-biology/natural-selection/phylogeny/a/phylogenetic-trees>

- SARS-CoV-2 molecular clock initially estimated to be **8.0×10^{-4} substitutions/site/year** ~
24 substitutions/year
 - Shown to be **increased ~4-fold** along branches leading to **emergence VOCs** (Tay *et al.*,
<https://doi.org/10.1093/molbev/msac013>)

SARS-CoV-2 evolution

- SARS-CoV-2 is in a continual state of evolution



- Mutations occur **regularly** during the viral replicative cycle
 - 1 change per week
 - Substitution rate of 0.00084 per site per year (half that of influenza A and a quarter that of HIV-1)
- Most of the changes have **not conferred a selective advantage**
- Mutations » viral diversification » classification

Making sense of substitutions

1

Variant calling



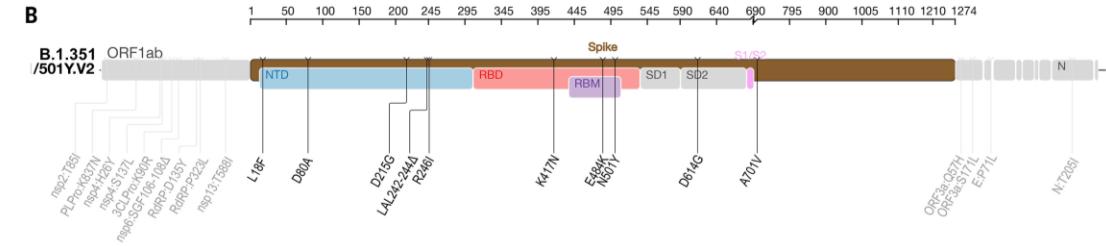
2

Consensus fasta file generation



3

Clade assignments and identification of mutations



<https://covdb.stanford.edu/sierra/sars2/by-sequences/>



<https://outbreak.info>



Nextstrain and GISAID

All the hype about the D614G mutation

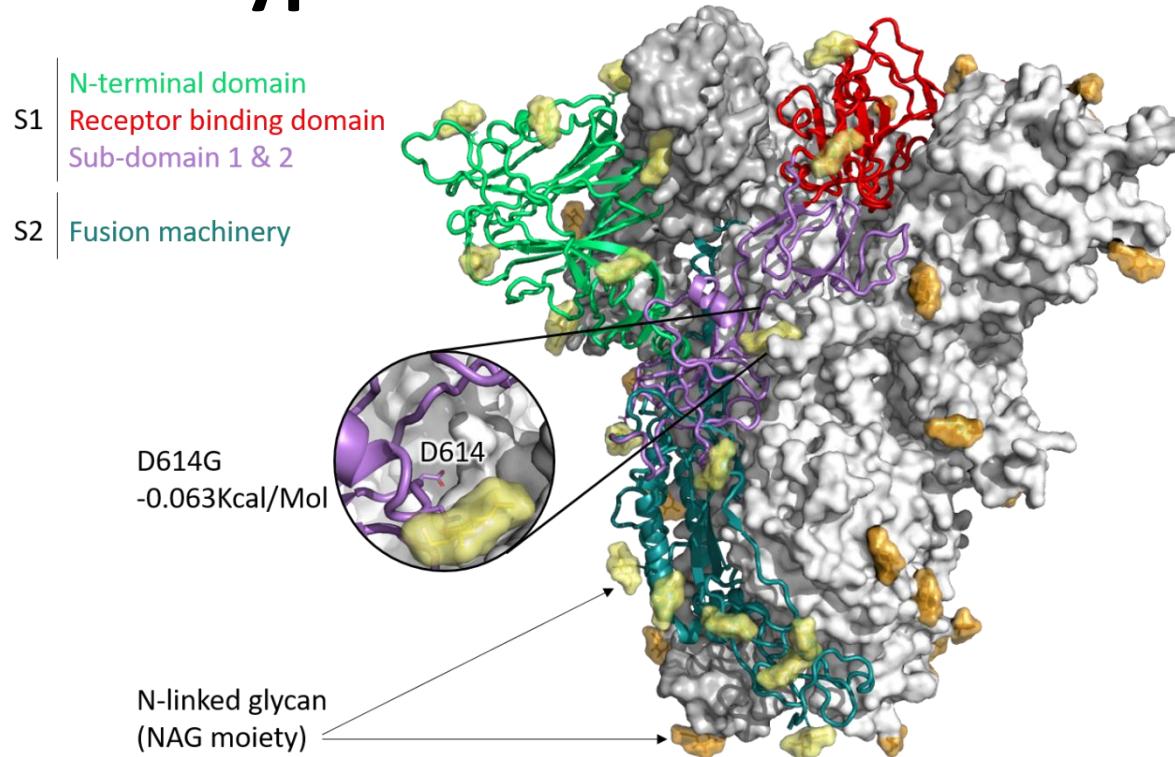
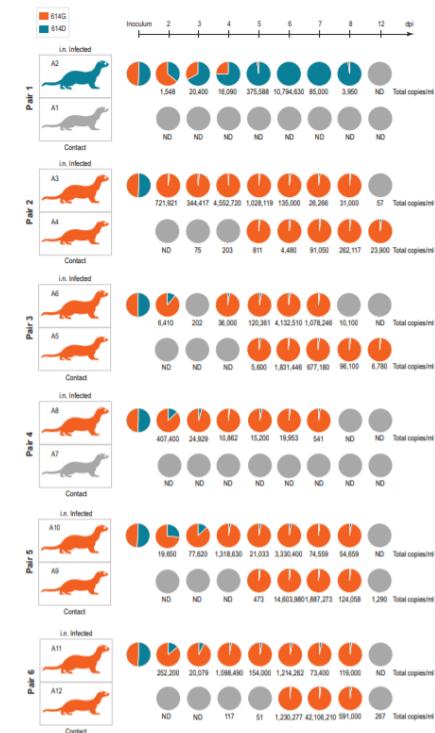
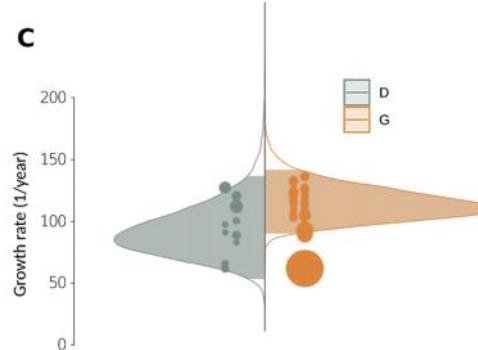
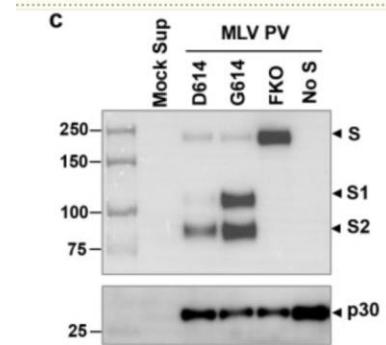
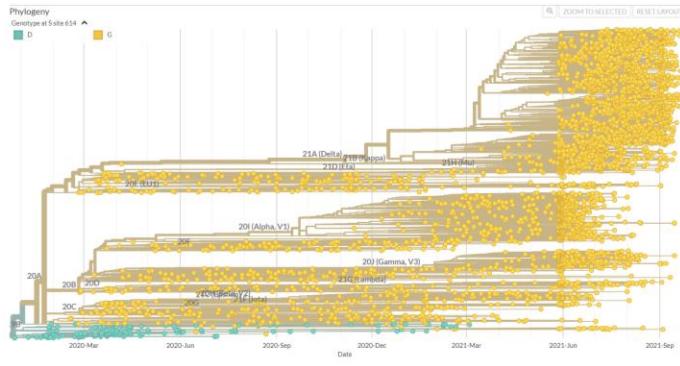


Image: Kurt Wibmer
NICD

Spike D614G mutation country frequency



Replacement and continued dominance of a single spike mutation, which increased viral replicative fitness, but not pathogenesis

N501Y variants have increased ACE2 binding and transmissibility



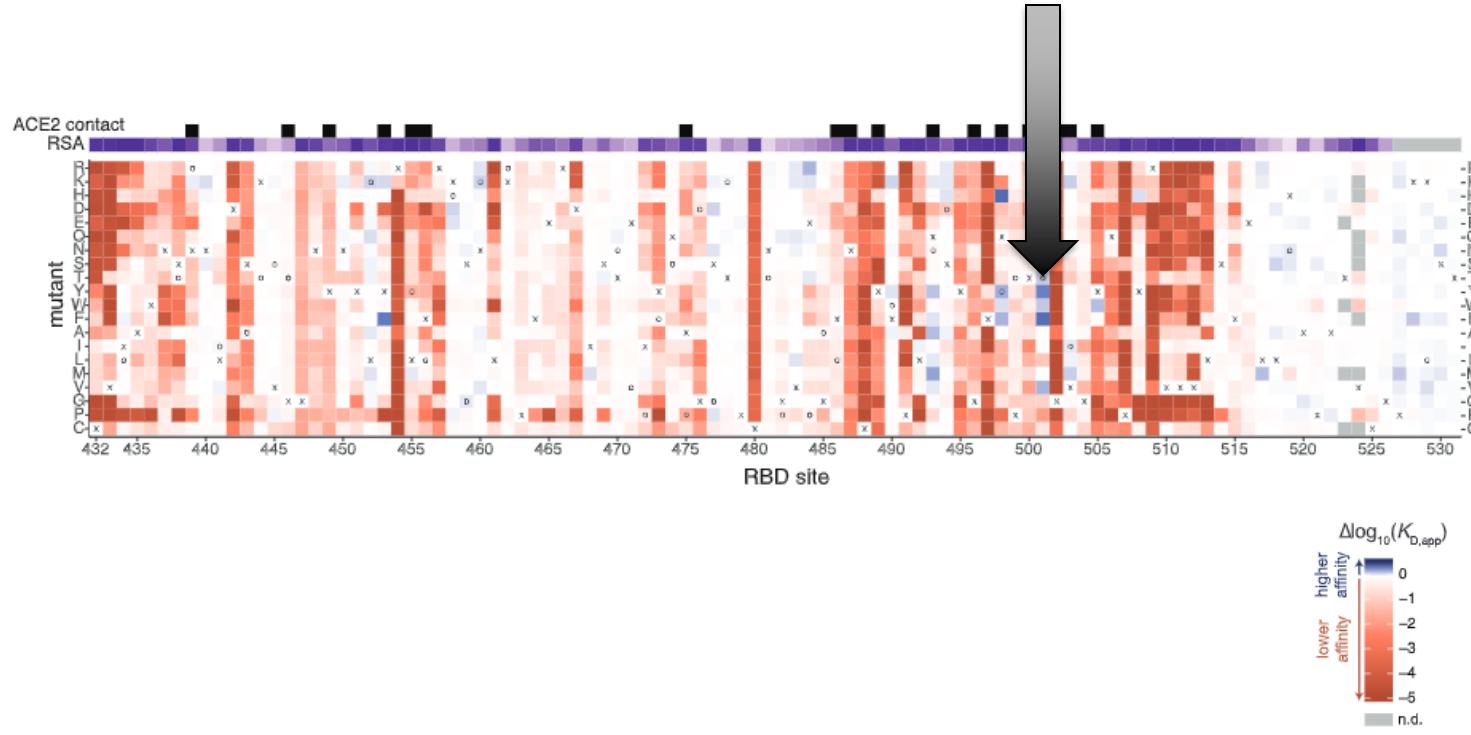
20I
B.1.1.7
Alpha



20H
B.1.351
Beta



20J
P.1
Gamma



<https://www.sciencedirect.com/science/article/pii/S0092867420310035>



The detection of a divergent BA.3-related lineage

07 March 2025

Centre for Respiratory Diseases and Meningitis,
National Institute for Communicable Diseases,
A division of the National Health Laboratory Service



Overview

- In March 2025, NICD has detected a divergent BA.3-related lineage. The lineage has not yet been designated by pangolin.
- The BA.3-related sequences have been detected in three specimens collected in Gauteng (n=2) and Kwa-Zulu Natal (n=1) provinces between November 2024 and January 2025
- The genomes were sequenced at the NICD from specimens collected for diagnostics (n=2) and influenza-like illness (ILI) surveillance (n=1)
- As of 06/03/2025, no similar sequences have yet been identified outside of South Africa

Metadata of three BA.3-related sequences with private mutations, 5 March 2025

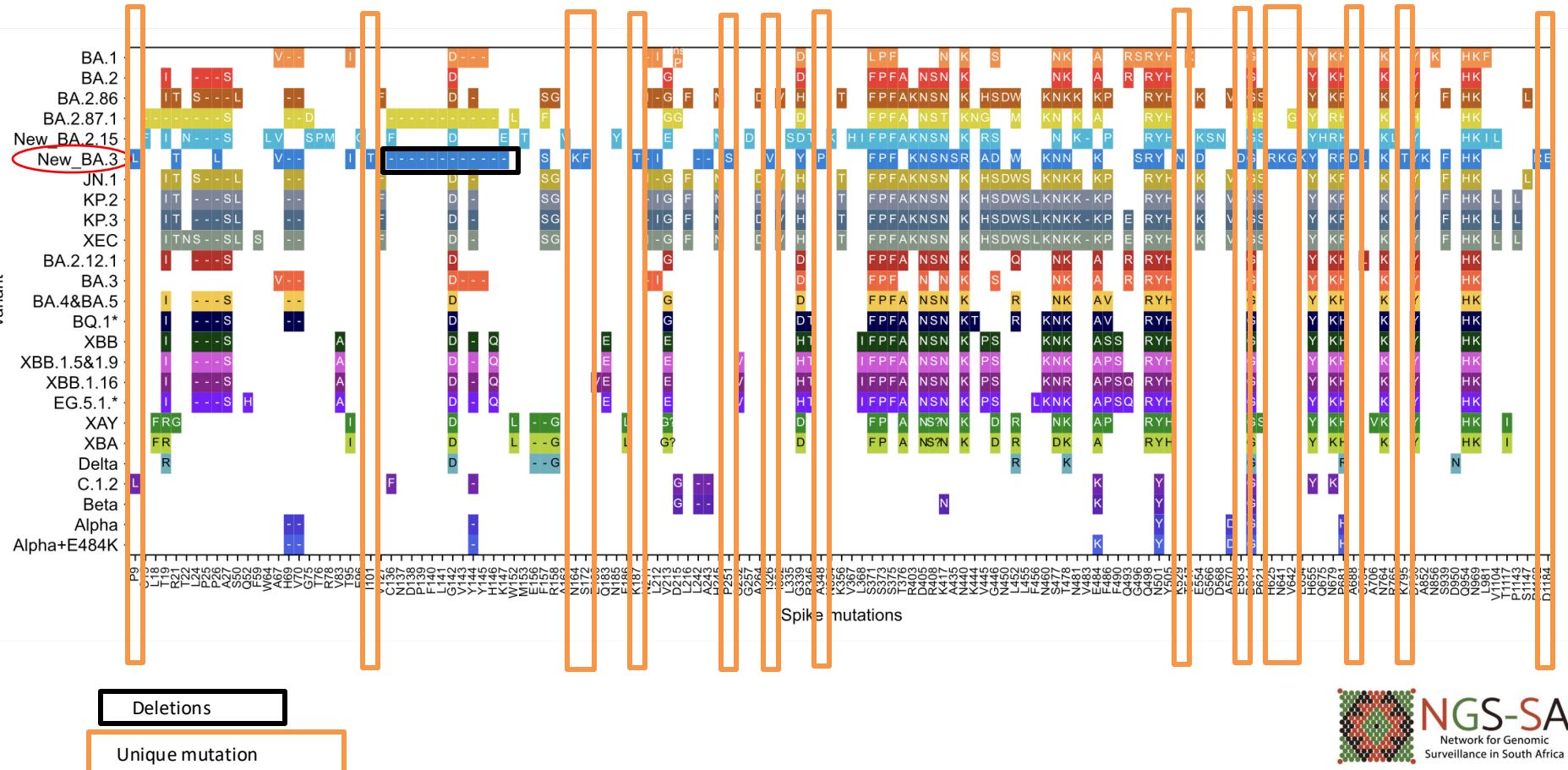
Collection date	Age group (years)	Sex	Location	Lab	Surveillance	N gene Ct	Rnase Ct	S gene Ct	Initial lineage designation
22-Nov-24	5-9	M	GAUTENG	AMPATH	Diagnostics	unavailable	unavailable	unavailable	BA.3
24-Nov-24	50-59	F	GAUTENG	AMPATH	Diagnostics	unavailable	unavailable	unavailable	BA.3
10-Jan-25	40-49	F	KwaZulu-Natal	NICD	ILI	23.2	26.3	22.0	BA.3



Genomic profile: BA.3-related lineage

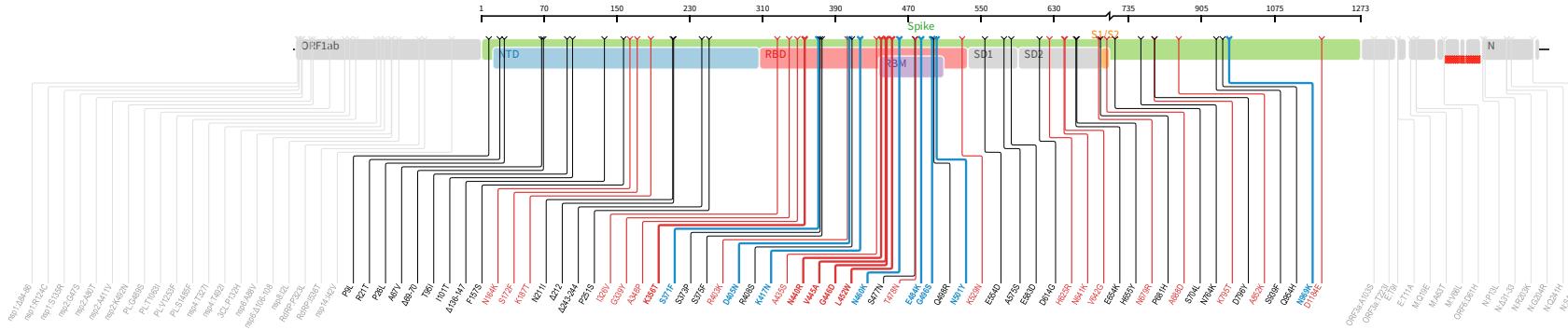
- New BA.3-related variant viruses (n=3) have 119 mutations relative to the Wuhan strain including 89 non-synonymous substitutions (concentrated in spike: 57/89) and 7 deletions (4/7 in spike)
- Several mutations are concentrated in important regions of the spike protein:
 - One large deletion (36 nt) in the N-terminal domain, Δ136-147 (one nucleotide longer than that found in BA.2.87.1)
 - Multiple substitutions at important antigenic sites in the receptor-binding domain (e.g. I326V, G339Y, A348P, K356T, R403K, A435S, N440R, V445A, G446D, L452W, T478N, K529N)
 - Mutations near the furin cleavage site (N679R (also in BA.2.87.1), A688D)

Spike protein mutation profile





Mutational profile of the spike protein among new BA.3-related variant (n=3)



- Substitutions indicated in red: <0.1% global prevalence (N164K, S172F, K187T, I326V, G339Y, A348P, K356T, R403K, A435S, N440R, V445A, G446D, L452W, T478N, K529N, H625R, N641K, V642G, N679R, A688D, K795T, A852K, D1184E)
- Substitutions indicated in blue: Monoclonal antibody resistance mutations





New BA.3-related variant has four deletions in spike NTD

- $\Delta 69-70$, $\Delta 212$, $\Delta 243-244$ have been previously reported in multiple variants (BA.1, BA.2, BA.5 and Beta). In addition, the new BA.3-related variant carries a unique large deletion $\Delta 136-147$ (12 amino acids).
- A similar deletion with one less amino acid ($\Delta 136-146$) was previously detected in BA.2.87.1
- NTD deletions between positions 141-146 have been previously described in the Alpha and Omicron BA.1 and BA.2 variants (Li et al; 2020)
- A deletion of this size, situated with the antigenic supersite of the NTD, may affect binding of antibodies targeting the NTD. Additionally, as the NTD helps with stabilizing the RBD, the deletion may alter the spike conformation (Yu et al; 2023)



Phylogeny

Clade ▾

- █ 21L
- █ 24B
- █ 22A
- █ 24C
- █ 22B
- █ 24E
- █ 22C
- █ 24G
- █ 22D
- █ 24H
- █ 22E
- █ 24I
- █ 23C
- █ 23I
- █ 21M
- █ 24A

The BA.3-related variant branches off the BA.3 root with long distance

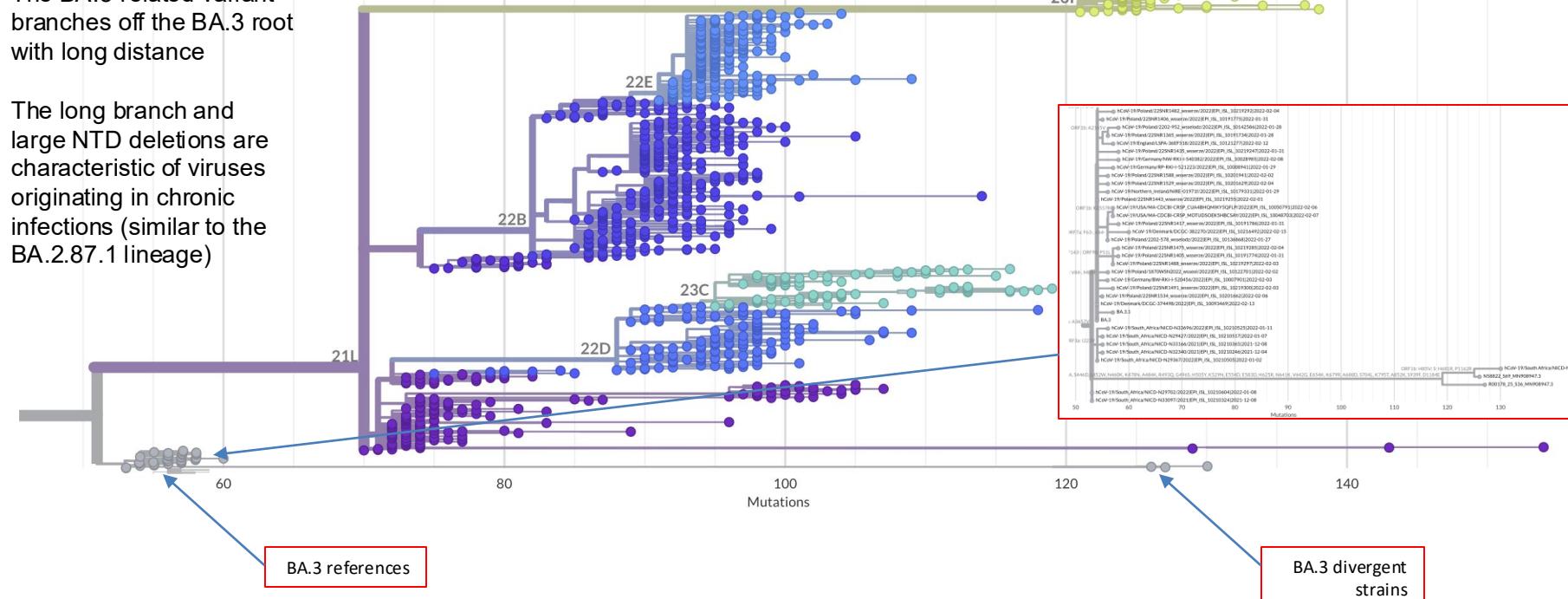
The long branch and large NTD deletions are characteristic of viruses originating in chronic infections (similar to the BA.2.87.1 lineage)

Nextstrain tree shows the new variant as a BA.3 basal taxon



ZOOM TO SELECTED

RESET LAYOUT





BA.3 saltation with 57 spike AA mutations, South Africa #2909

Closed



dikeledik opened on Mar 10 · edited by dikeledik

Edits

...

Three divergent sequences of a BA.3 related lineage were detected in South Africa collected in the Gauteng (n=2) and Kwa-Zulu Natal (n=1) provinces between November 2024 and January 2025. Samples were sampled through the lab-based national surveillance (n=2) and ILI surveillance (n=1). The new BA.3-related variant viruses (n=3) have 120 nucleotide substitutions relative to the Wuhan strain and 89 non-synonymous substitutions (concentrated in spike (57/89) and 7 deletions (4/7 in spike (Δ 69-70, **Δ 136-147**, Δ 212, Δ 243-244)).

Accession numbers:

EPI_ISL_19771108, EPI_ISL_19771107, EPI_ISL_19771105

Mutations observed on branch

Nt

Unique (23):G2281T, G6476T, A12095C, T13375C, T15047C, C21639T, T22054A, C22077T, A22538G, G22577T, A22786T, A22881G, C23075T, A23436G, C23625A, A23946C, G24116A, C24117A, C25114A, G25234A, G25699T, A26771T, A28996T

Homoplasies (44):C21T, C193T, C635T, C2037T, T4579A, C9886T, C10747T, C11620T, C14649T, C21588T, G21624C, T21864C, T22032C, A22122C, C22313T, G22604C, G22770A, G22865T, T22896C, A22898G, G22899A, C22916T, T22917G, T22942G, A22996C, G23012A, C23013A, G23040A, G23048A, G23149T, G23224T, G23311T, T23485A, T23487G, G23522A, A23598G, C23673T, C24378T, C25803T,

A
N
—
L
N
—
T
N
—
P
N
—
R

5 new BA.2 sequences with private mutations (Ampath – national surveillance)

Collection date	Age group (years)	Sex	Location	Info	N gene Ct	ORF1ab Ct	S gene Ct
20-Sep-23	93	F	Groenkloof, Gauteng	unavailable	unavailable	unavailable	unavailable
07-Oct-23	57	F	Wilgers, Gauteng	unavailable	unavailable	unavailable	unavailable
02-Nov-23	2	M	Polokwane, Limpopo	unavailable	unavailable	unavailable	unavailable
13-Nov-23	29	M	Roseacress, Gauteng	unavailable	unavailable	unavailable	unavailable
12-Nov-23	1	M	Pholoso, Limpopo	unavailable	unavailable	unavailable	unavailable

- NextClade assigns to clade 21L
- Pangolin assigns to lineage BA.2
- Spike NTD mutations shared with BA.2.86.* , but has some unique mutations and reversions to the ancestral BA.2 lineage.
- Sequences have two long and unique deletions on the spike (del_15-26 and del_136-146)
- Phylogeny shows that the closest related lineage is BA.2.15 (with sequence divergence, cumulative distance tree from the root, of 74 while divergence of the new BA.2 = 136). The new BA.2 branches off directly from the ancestral BA.2, as is the case with the BA.2.86.*

Nextclade



◀ Back

Done. Total sequences: 181. Succeeded: 181

New_BA.2

Citation

Docs

Settings

What's new

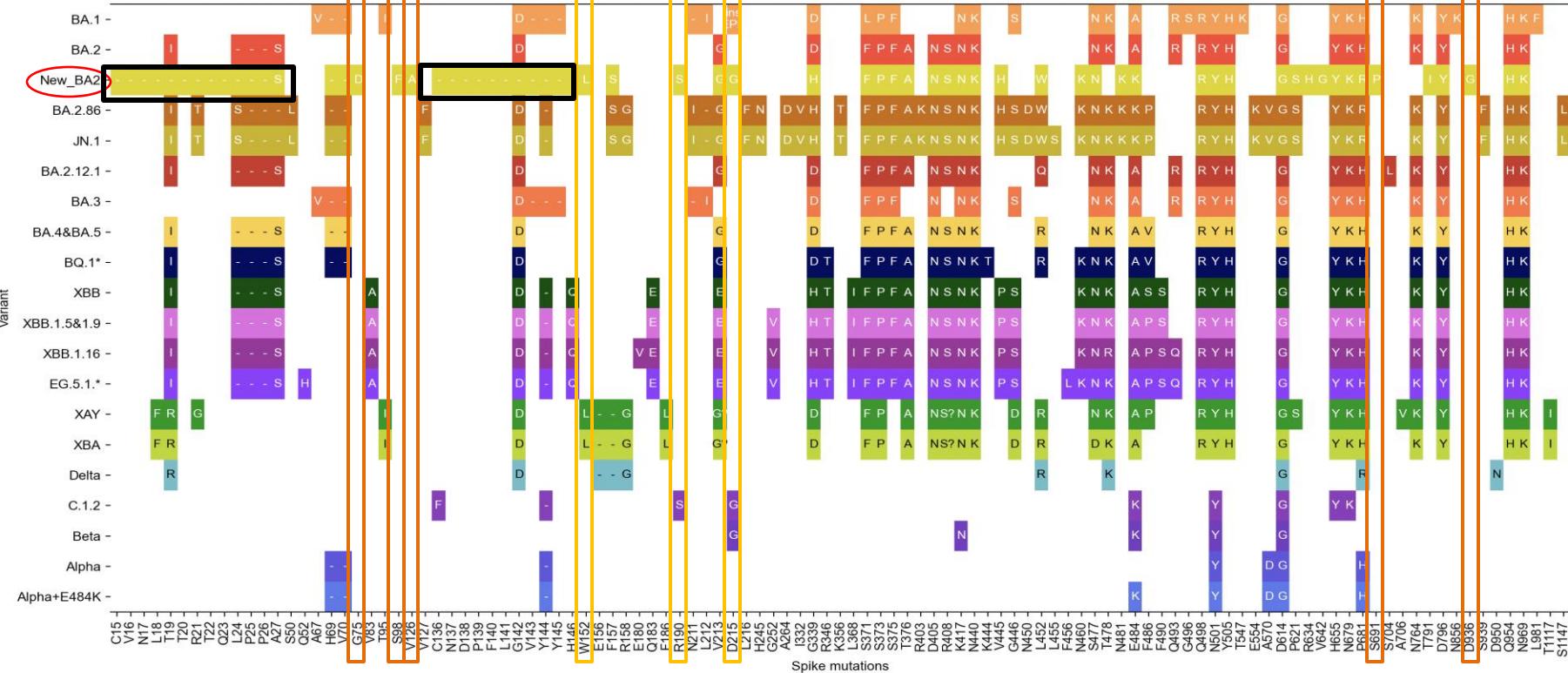
English



#	i	Sequence name	QC	Clade	Pango lineage (Nextclade)	WHO name	Mut.	non-ACGTN	Ns	Cov.	Gaps	Ins.	FS	SC	Gene S
?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
0	0	N56614	N M P C F S	21L	BA.2	Omicron	118	0	579	97.4%	136	0	0 (1)	0	
1	1	N56836	N M P C F S	21L	BA.2	Omicron	122	0	523	97.8%	136	0	0 (1)	0	
2	2	hCoV-19/South Africa/NICD-N57176/2	N M P C F S	21L	BA.2	Omicron	104	0	2048	91.5%	136	0	0 (1)	0	
3	3	hCoV-19/South Africa/NICD-N57208/2	N M P C F S	21L	BA.2	Omicron	119	0	189	98.9%	136	0	0 (1)	0	
4	4	hCoV-19/South Africa/NICD-N57216/2	N M P C F S	21L	BA.2	Omicron	112	0	1312	94.2%	136	0	0 (1)	0	
5	7	hCoV-19/South Africa/NICD-N56224/2	N M P C F S	23I	BA.2.86	Omicron	103	0	238	98.7%	65	0	0	0	
6	18	hCoV-19/South Africa/NICD-N57015/2	N M P C F S	23I	BA.2.86	Omicron	96	0	696	96.3%	65	0	0	0	
7	20	hCoV-19/South Africa/NICD-N57017/2	N M P C F S	23I	BA.2.86	Omicron	101	0	120	99.2%	65	0	0	0	
8	21	hCoV-19/South Africa/NICD-N57018/2	N M P C F S	23I	BA.2.86	Omicron	112	0	11	99.5%	65	0	0	0	
9	23	hCoV-19/South Africa/NICD-N57020/2	N M P C F S	23I	BA.2.86	Omicron	109	0	14	99.5%	65	0	0	0	
10	27	hCoV-19/South Africa/NICD-N57025/2	N M P C F S	23I	BA.2.86	Omicron	106	0	275	98.7%	65	0	0	0	
11	28	hCoV-19/South Africa/NICD-N57026/2	N M P C F S	23I	BA.2.86	Omicron	100	0	276	98.5%	80	0	0	0	

BA.2.86

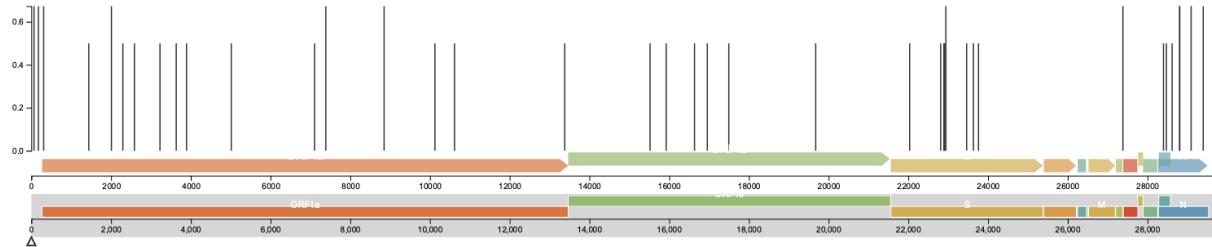
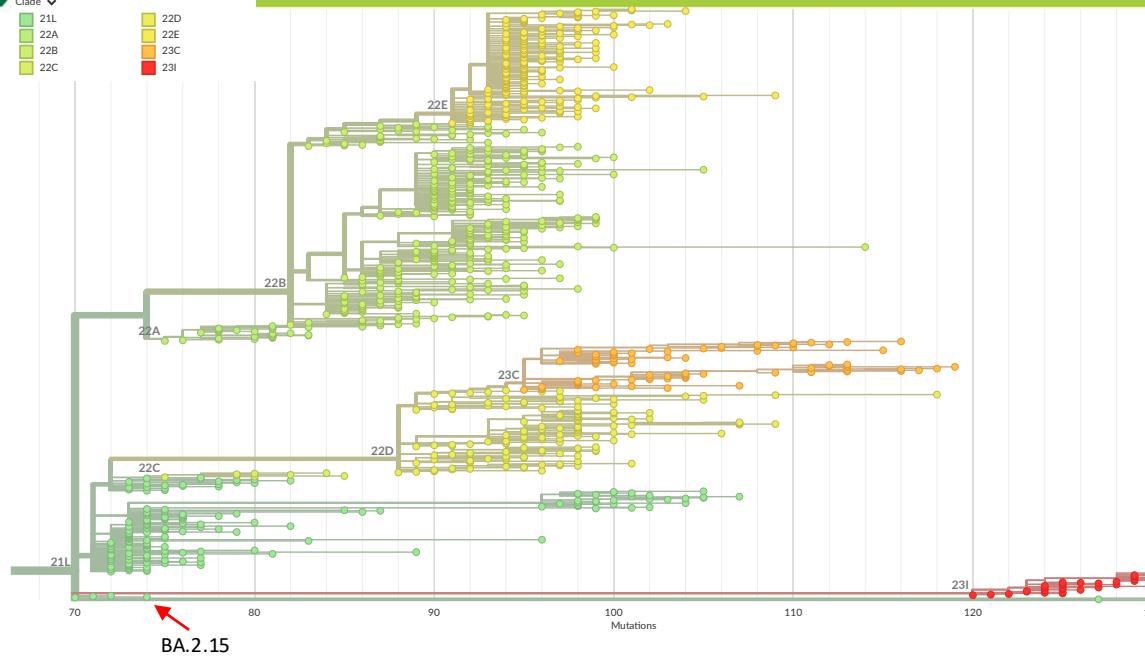
Spike protein mutation profile



Deletions

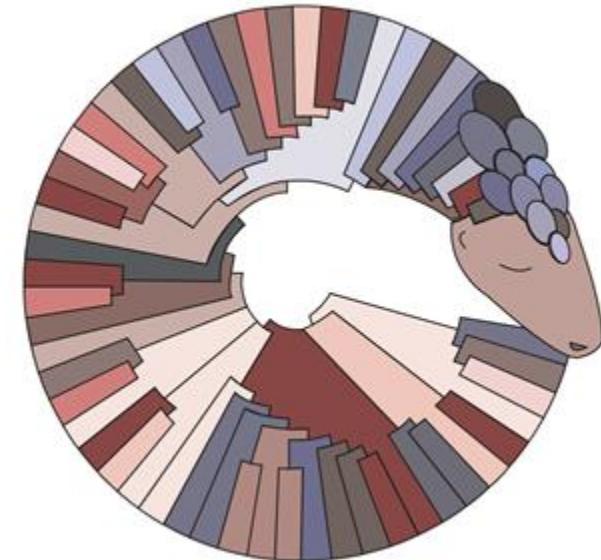
Private mutations

Mutations shared
with one or two
lineages



How to go from SARS-CoV-2 sequence to variant?

- Three main naming systems:
 1. GISAID clades
 2. Nextstrain clades
 3. Pangolin lineages



Clades vs lineages

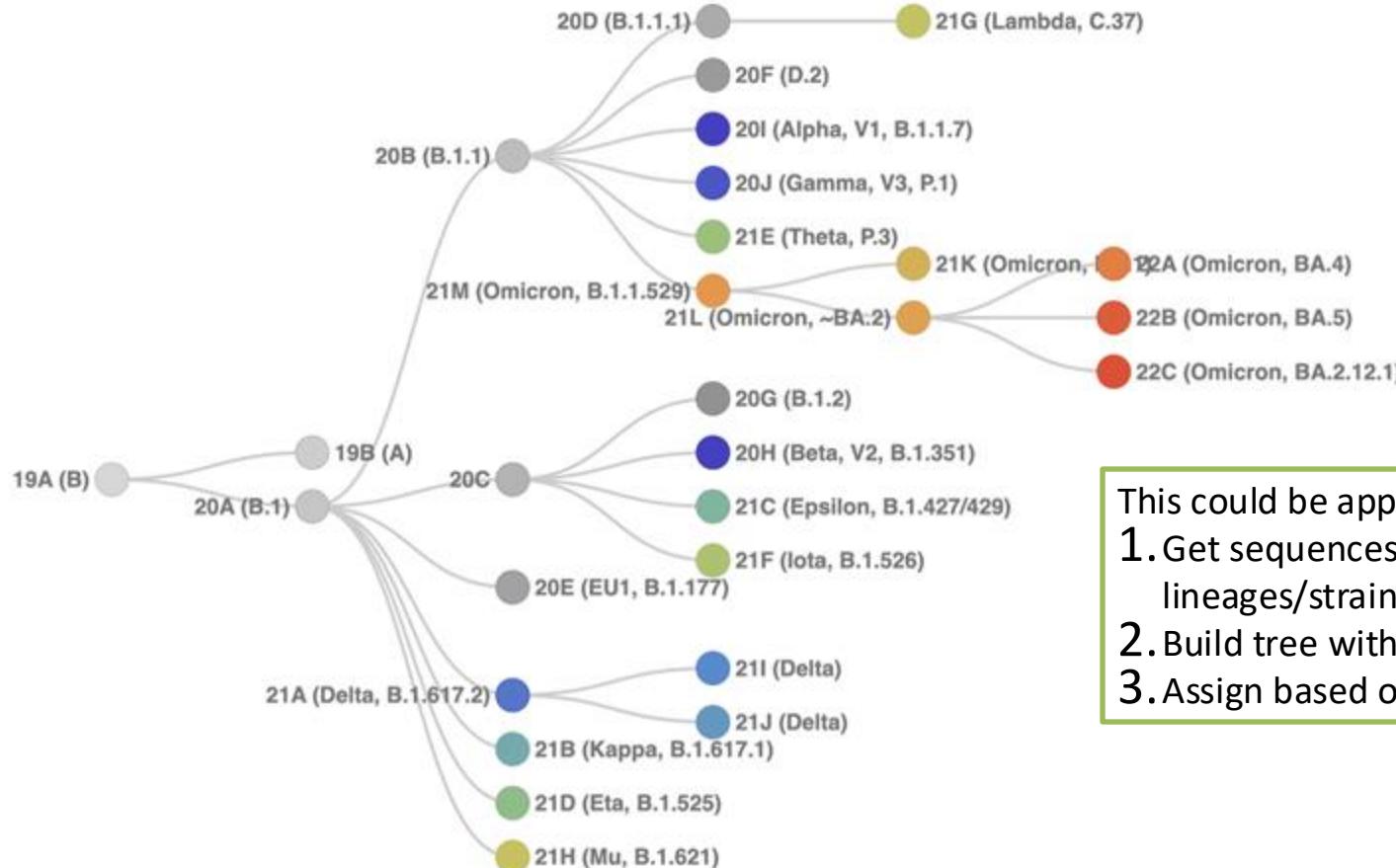
Nextstrain Clades	Pangolin lineages
Assigns based on phylogenetic analysis	Assigns based on phylogenetic analysis (accuracy) or machine learning (speed)
Macro-level (larger groups per clade)	More detailed (smaller groups per lineage)
Requires shared ancestry, mutations, epidemiological significance	Requires shared ancestry, mutations, epidemiological significance
Naming: number-letter system (year of designation + next available letter), e.g. 21A	Naming: hierarchical alphanumeric system, dot implies ancestry; e.g. B.1 is the first descendant of B



Nextclade



Nextstrain clades



This could be applied to any pathogen

1. Get sequences of known lineages/strains/clades
2. Build tree with your sequences
3. Assign based on where they fall



Nextclade



How to assign clades and lineages

- Tool called Nextclade, two options:
 1. Command line tool
 2. **Web interface:** <https://clades.nextstrain.org/>
- Input: fasta file containing your sequence/s of interest
 - Works on full genomes or fragments (≥ 100 bp)
 - Assigns something unless it can't align the sequences – fast but **not always reliable**
- Output: webpage containing clades, genome visualisation, basic phylogenetic tree, downloads

Nextclade

v3.2.0

Clade assignment, mutation calling, and sequence quality checks

Add more sequence data

[File](#) [Link](#) [Text](#) [Example ▾](#)

Drag & drop files or folders



FASTA

Select files

Sequence data you've added

[Remove all](#)

Eswatini_240109.fasta (2.93 MB)



Selected reference dataset [i](#)



Suggest automatically

[Reset](#)

[Suggest](#)

A new version of this dataset is available.
[What's new?](#)

[Update](#)



SARS-CoV-2

official

Reference: Wuhan-Hu-1/2019 (MN908947)

Updated at: 2024-01-16 20:31:02 (UTC)

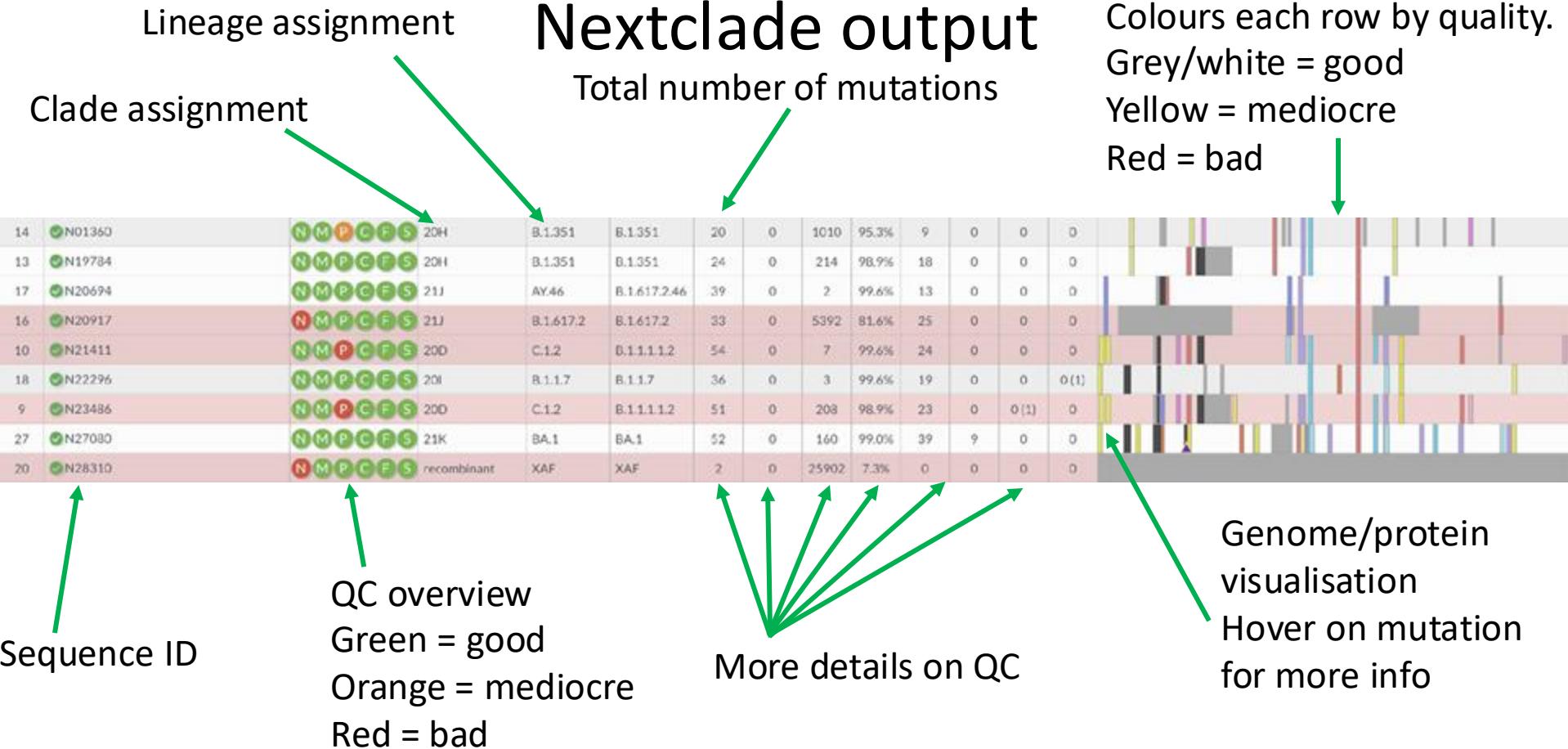
Dataset name: nextstrain/sars-cov-2/wuhan-hu-1/orfs

[Load example](#)

[Change reference dataset](#)

[Run](#)

Nextclade output



Nextclade download

Clade assignment

Variant assignment

Assigned quality

index	seqName	clade	Nextclade_pango	partiallyAliased	clade_nextstrain	clade_who	clade_legacy	qc.overallStatus	totalSubstitutions
14	N01360	20H	B.1.351	B.1.351	20H	Beta	20H (Beta, V2)	good	20
13	N19784	20H	B.1.351	B.1.351	20H	Beta	20H (Beta, V2)	good	24
17	N20694	21J	AY.46	B.1.617.2.46	21J	Delta	21J (Delta)	good	39
16	N20917	21J	B.1.617.2	B.1.617.2	21J	Delta	21J (Delta)	bad	33
10	N21411	20D	C.1.2	B.1.1.1.1.2	20D	unassigned	20D	bad	54
18	N22296	20I	B.1.1.7	B.1.1.7	20I	Alpha	20I (Alpha, V1)	good	36
9	N23486	20D	C.1.2	B.1.1.1.1.2	20D	unassigned	20D	bad	51
27	N27080	21K	BA.1	BA.1	21K	Omicron	21K (Omicron)	good	52
20	N28310	recombinant	XAF	XAF	recombinant	recombinant	recombinant	bad	2

Sequence ID

Lineage assignment

Total number of mutations

Lots of columns
not pictured...

Variant under monitoring (VUM)

Working Definition of “SARS-CoV-2 variant under monitoring”

- A SARS-CoV-2 variant with **genetic changes that are suspected to affect virus characteristics** with some indication that it may pose a future risk, but **evidence of phenotypic or epidemiological impact is currently unclear**, requiring enhanced monitoring and repeat assessment pending new evidence.
- Note: It is expected that our understanding of the impacts of these variants may fast evolve, and designated Variants under Monitoring may be readily added/removed; therefore, WHO labels will not be assigned at this time. Former VOIs/VOCs may, however, be monitored for an extended period under this category, and will maintain their assigned WHO label until further notice.
- **Currently 15 VUMs (including C.1.2)**

Variant of interest (VOI)

Working Definition of “SARS-CoV-2 Variant of Interest”

- A SARS-CoV-2 isolate is a variant of interest (VOI) if it is **phenotypically changed compared to a reference isolate** or has a genome with mutations that lead to amino acid changes **associated with established or suspected phenotypic implications**;

AND

- has been identified to cause **community transmission/multiple COVID-19 cases/clusters, or has been detected in multiple countries**;

OR

- is otherwise assessed to be a VOI by WHO in consultation with the WHO SARS-CoV-2 Virus Evolution Working Group.
- Currently 2 VOI (Lambda and Mu)

<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>

Variant of concern (VOC)

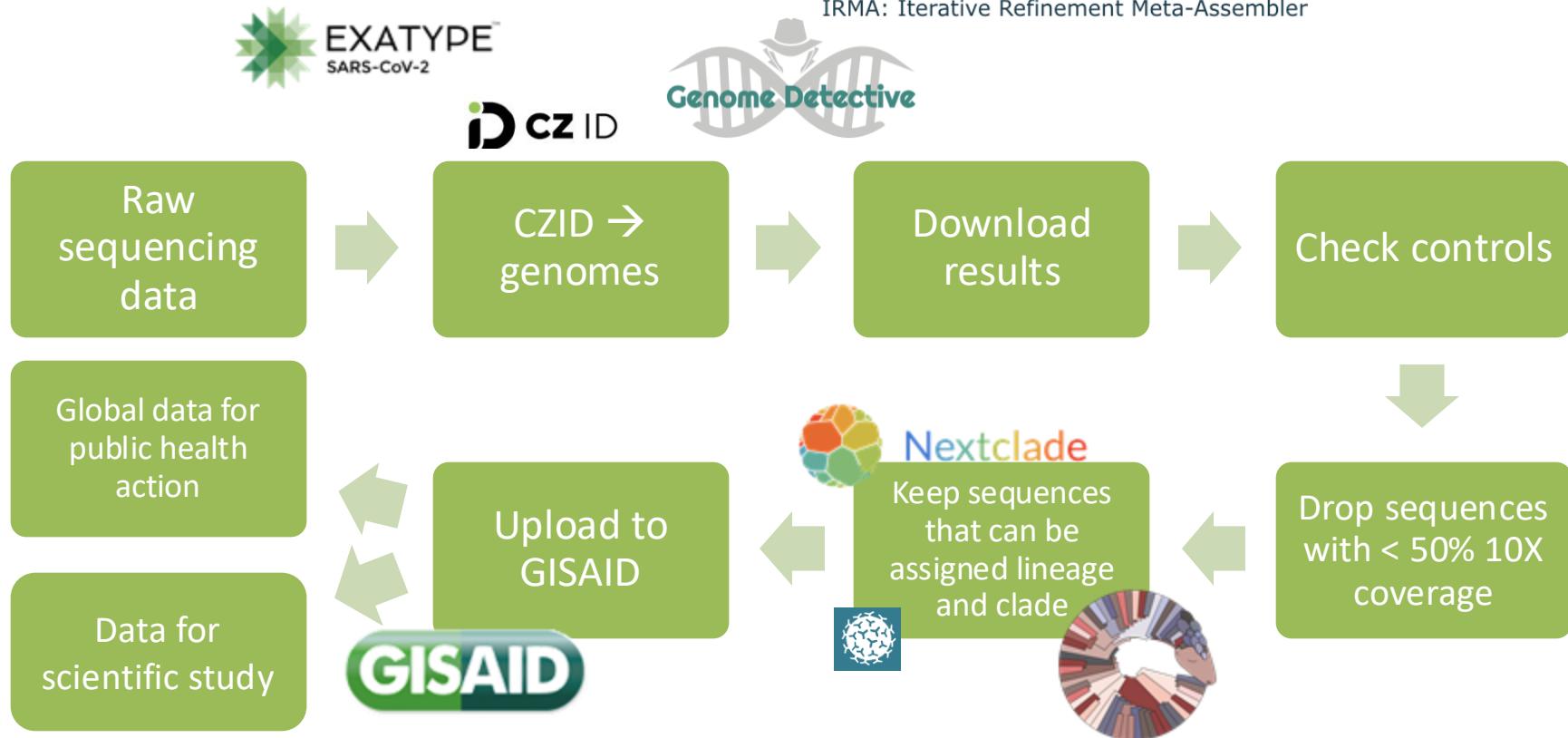
Working Definition of “SARS-CoV-2 Variant of Concern”

- A VOI (as defined above) is a variant of concern (VOC) if, through a **comparative assessment**, it has been demonstrated to be associated with
 - Increase in transmissibility or detrimental change in COVID-19 epidemiology;
 - Increase in virulence or change in clinical disease presentation; or
 - Decrease in effectiveness of public health and social measures or available diagnostics, vaccines, therapeutics.

OR

- assessed to be a VOC by WHO in consultation with the WHO SARS-CoV-2 Virus Evolution Working Group.
- Currently 4 VOC (Alpha, Beta, Gamma and Delta)

How we process our genomes



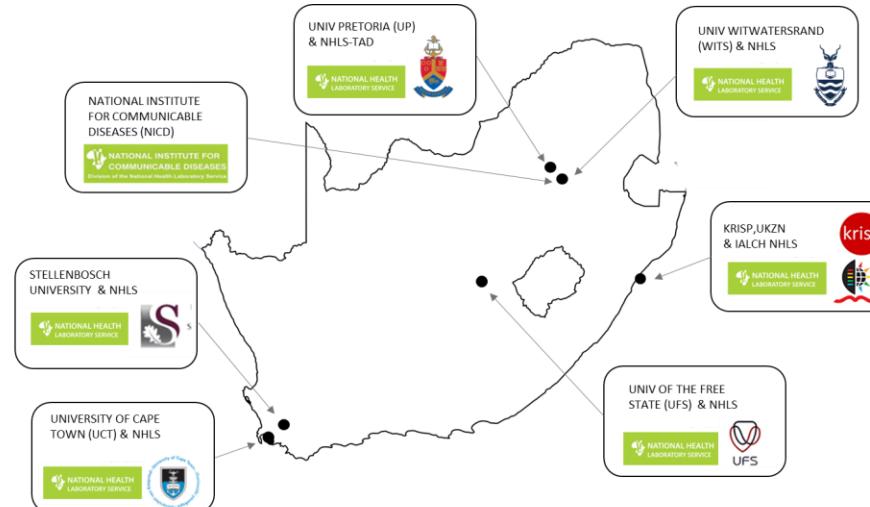


Data sharing (reporting)

- Facilitates sharing between labs
- Facilitates **global public health efforts**
- There are a number of public databases – each have pros and cons
- We currently share to GISAID
 - Limited open access
 - Protects rights of data sharers
 - Millions of genomes available
 - Has some built-in analysis tools

SARS-CoV-2 Sequencing Update

07 Aug 2025



Supported by the DSI and the SA MRC

Msom N, Mlisana K, et al. Lancet Microbe 2020

Number of South African genomes deposited on GISAID, by specimen collection week, 2020 – 2025 (N=56 128*)

Total genomes: 56 067

2020 genomes: 7 082

2021 genomes: 26 862

2022 genomes: 15 721

2023 genomes: 5 177

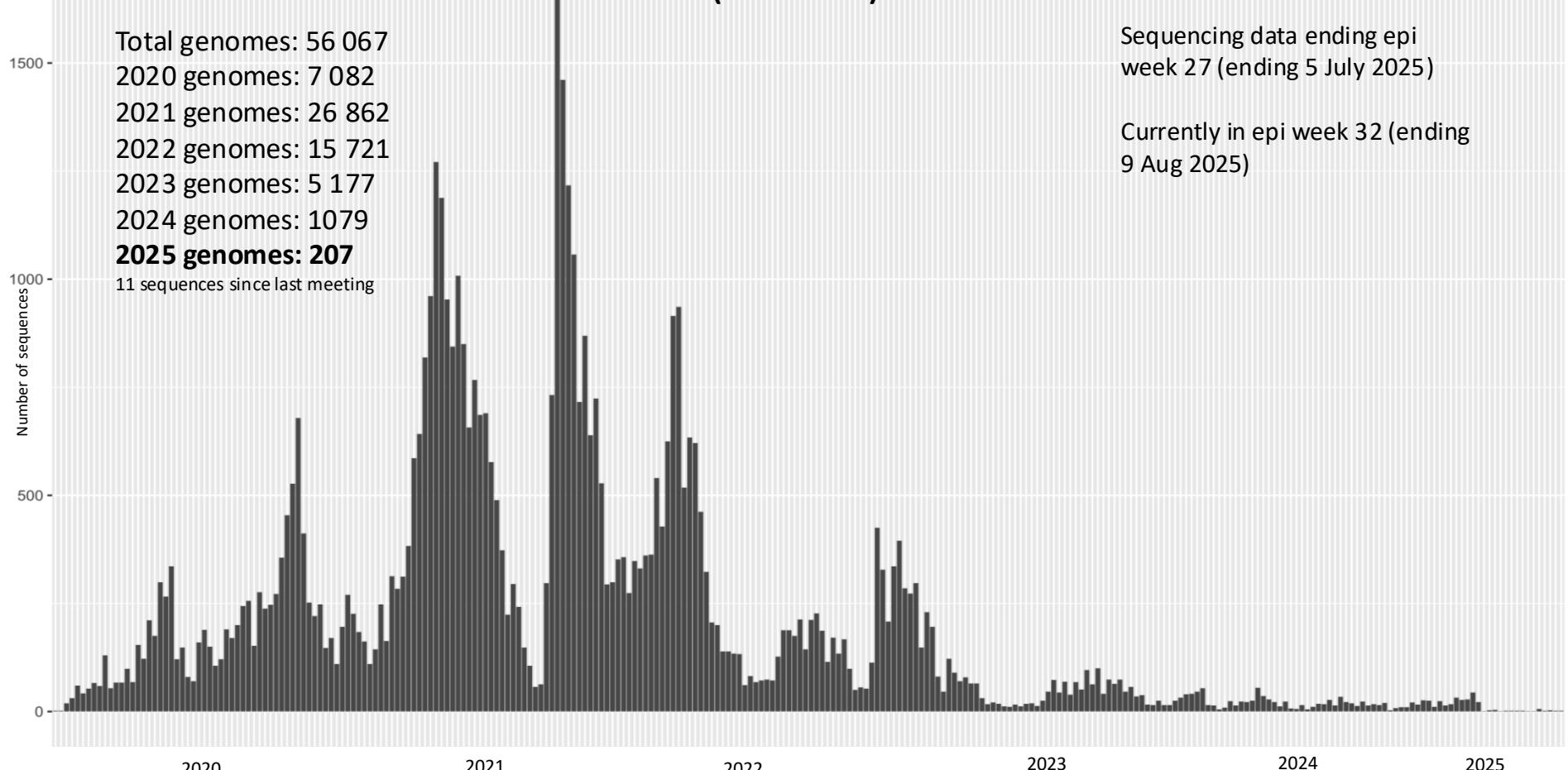
2024 genomes: 1 079

2025 genomes: 207

11 sequences since last meeting

Sequencing data ending epi
week 27 (ending 5 July 2025)

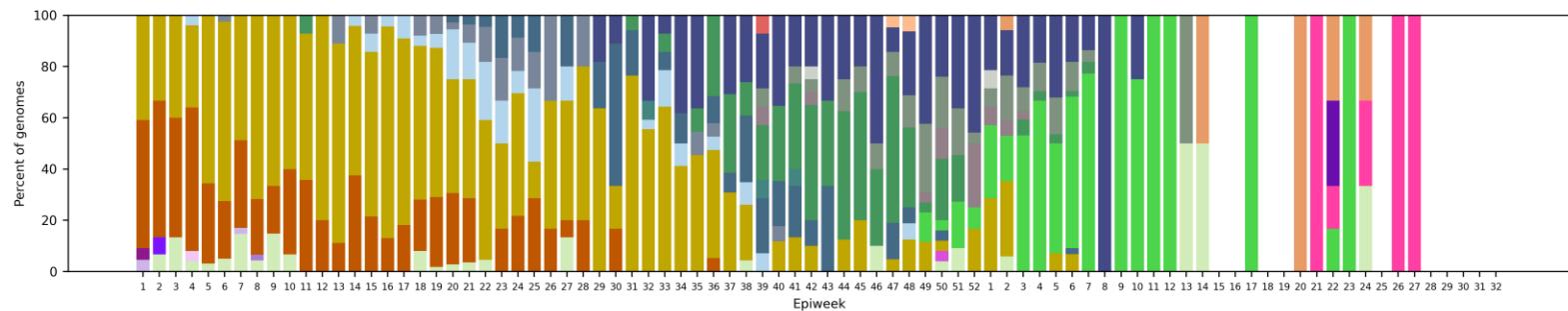
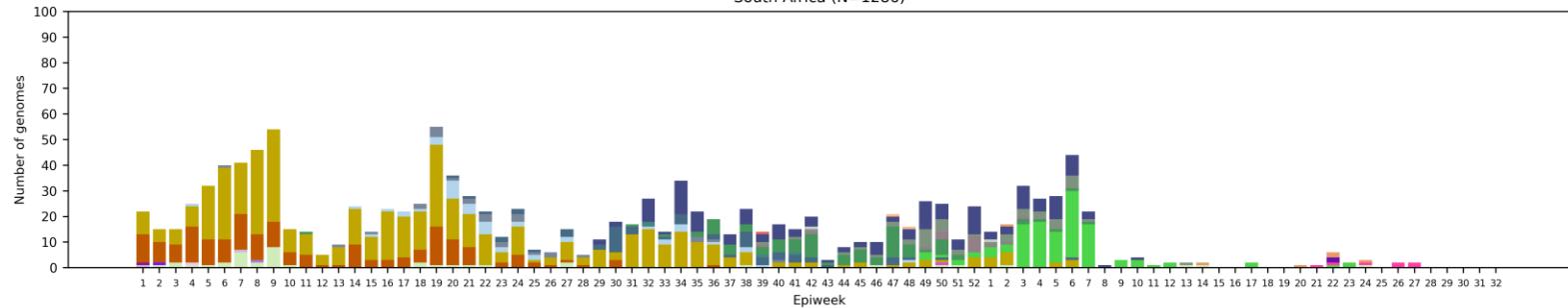
Currently in epi week 32 (ending
9 Aug 2025)



*This represents the cleaned, de-duplicated dataset of unique National and Pneumonia Surveillance sequences. This dataset will be used for all further figures.

South Africa, 2024-2025, N=1286*

South Africa (N=1286)



Clade key (bar graph)



*Excludes sequences missing collection dates. Lineages of particular interest (mainly WHO Omicron subvariants under monitoring) are separate from the main clade groupings, one sequence does not have province data and will be excluded in the plots below.

[#]Recombinants include all recombinant lineages (viruses consisting of segments of two different lineages) detected in South Africa at low levels. Currently it consists of XT, XAS, XAZ, XBA, XBF.



Acknowledgments



Thabo
Mohale



Arshad
Ismail



Cathrine
Scheepers



Zamantungwa
Khumalo



Anne
Von Gottberg



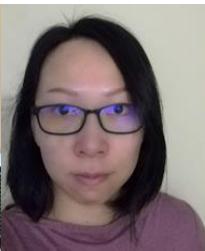
Anele
Mnguni



Daniel
Amoako



Morne
Du Plessis



Annie
Chan



Noxolo
Ntuli



Josie
Everatt



Boitshoko
Mahlangu



Jinal
Bhiman

Acknowledgments

Centre for Respiratory Diseases & Meningitis

Anne von Gottberg

Thabo Mohale

Daniel Amoako

Josie Everett

Boitshoko Mahlangu

Noxolo Ntuli

Anele Mnguni

Amelia Buys

Cardia Fourie

Noluthando Duma

Linda de Gouveia

Jackie Kleynhans

Nicole Wolter

Sibongile Walaza

Mignon du Plessis

Stefano Tempia

Mvuyo Makhosi

Cheryl Cohen

Centre for HIV and STIs

Cathrine Scheepers

Constantinos Kurt Wibmer

Thandeka Moyo

Tandile Hermanus

Frances Ayres

Zanele Molaudzi

Bronwen Lambson

Tandile Hermanus

Mashudu Madzivhandila

Prudence Kgagudi

Brent Oosthuysen

Lynn Morris

Penny Moore

Sequencing Core Facility

Zamantungwa Khumalo

Annie Chan

Morne du Plessis

Stanford Kwenda

Phillip Senzo Mtshali

Mushal Allam

Florah Mnyameni

Arshad Ismail

Centre for Emerging, Zoonotic & Parasitic Diseases

Jaqueline Weyer

NICD Groups

NICD COVID-19 response team

NICD SARS-CoV-2 Sequencing Group

NICD leadership

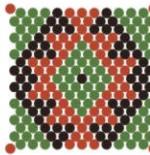
Michelle Groome

Adrian Puren

Harry Moultrie

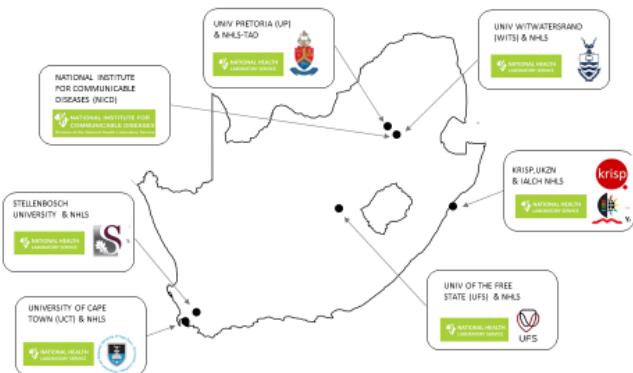
Kerrigan McCarthy





NGS-SA

Network for Genomic
Surveillance in South Africa



Supported by the DSI and the SA MRC



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY



science & innovation
Department:
Science and Innovation
REPUBLIC OF SOUTH AFRICA



NATIONAL INSTITUTE FOR
COMMUNICABLE DISEASES
Division of the National Health Laboratory Service



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

WITS
UNIVERSITY



UNIVERSITY OF
KWAZULU-NATAL
INYUVESI
YAKWAZULU-NATAL

NATIONAL HEALTH
LABORATORY SERVICE

National sequencing collaborators and funders

Adriano Mendes
Allison J. Glass
Andries Dreyer
Christa Viljoen
Elias Bereda
Eugenie Elliott
Florah Mnyameni
Florette K. Treurnicht
Gloria Selabe
Howard Newman
Jeannette Wadula
Kathleen Subramoney
Lia Rotherham

Marianne Wolfaardt
Marietjie Venter
Michaela Davis
Simnikiwe Mayaphi
Warren Lowman
Zinhle Makatini



Department:
Health
REPUBLIC OF SOUTH AFRICA



Department:
Science and Technology
REPUBLIC OF SOUTH AFRICA



The
**Fleming
Fund**

