VIRUS EVOLUTION

Virus Evolution, 2025, **11(1)**, veaf083
DOI: https://doi.org/10.1093/ve/veaf083
Advance Access Publication 3 November 2025
Research Article

Emergence of the BA.2.87.1 lineage of SARS-CoV-2 in South Africa, a highly diverged BA.2-related lineage

Dikeledi Kekana[1,†,*], Buhle Ntozini[1,†], Ryan Hisner[2], Mukhlid Yousif[3,4], Phindile Ntuli[3], Nkosenhle Ndlovu[3], Kerrigan McCarthy[3,4,5], Anele Mnguni[1], Boitshoko Mahlangu[1], Ayanda Nzimande[1], Nadine Stock[1], Houriiyah Tegally[6], Mary-Ann Davis[7], Monika Moir[6], Eduan Wilkinson[6], Cheryl Baxter[6], Jinal Bhiman[8,9], Cheryl Cohen[1,4], Sibongile Walaza[1,4], Anne von Gottberg[1,4], Tulio de Oliveira[6,10], Nicole Wolter[1,4], Darren Martin[2]

[1]Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases of the National Health Laboratory Service, 1 Modderfontein Road, Sandringham, Johannesburg, 2192, South Africa
[2]Department of Integrative Biomedical Sciences, Division of Computational Biology, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Anzio Road, Observatory, Cape Town, 7925, South Africa
[3]Centre for Vaccines and Immunology, National Institute for Communicable Diseases of the National Health Laboratory Service, 1 Modderfontein Road, Sandringham, Johannesburg, 2192, South Africa
[4]School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, 7 York Road, Parktown, Johannesburg, 2193, South Africa
[5]School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, 7 York Road, Parktown, Johannesburg, 2193, South Africa
[6]School for Data Science and Computational Thinking, Stellenbosch University, Centre for Vaccines and Immunology, 44 Banghoek Road, Stellenbosch, 7599, South Africa
[7]Western Cape Government Health and Wellness and School of Public Health, University of Cape Town, Robert Sobukwe Road, Bellville, Cape Town, 7535, South Africa
[8]SAMRC Antibody Immunity Research Unit, School of Pathology, University of the Witwatersrand, 7 York Road, Parktown, Johannesburg, 2193, South Africa
[9]Centre for HIV and STIs, National Institute for Communicable Diseases of the National Health Laboratory Service, 1 Modderfontein Road, Sandringham, Johannesburg, 2192, South Africa
[10]KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), Nelson R. Mandela School of Medicine, University of Kwa-Zulu Natal, 719 Umbilo Road, Durban, 4001, South Africa

*Corresponding author. Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases of the National Health Laboratory Service, 1 Modderfontein Road, Sandringham, Johannesburg, 2192, South Africa. E-mail: DikelediK@nicd.ac.za
†These authors contributed equally to this work

Abstract

The emergence of various SARS-CoV-2 lineages with adaptive mutations is of significant concern worldwide, especially when these mutations enhance the virus's ability to either evade immune responses or transmit more efficiently. Between September and December 2023, a highly diverged BA.2-related lineage, designated BA.2.87.1, was detected through diagnostic testing, syndromic surveillance, and wastewater surveillance in the Limpopo, Mpumalanga, Western Cape, Eastern Cape, and Gauteng provinces of South Africa. This lineage harbours 20 amino acid substitutions in Spike protein relative to baseline BA.2, including at antigenic sites of the receptor-binding domain (including N417T, K444N, V445G, L452M, N460K, K478T, N481K, and R493Q) and, most strikingly, large deletions of the N-terminal domain (NTD) residues 15–26 and 136–146. Such large NTD deletions have never been observed in circulating lineages but do sometimes occur in highly mutated sequences originating in chronic infections. Phylodynamic analysis supports the possibility that the BA.2.87.1 lineage originated in a chronic infection in that the nearest known ancestor of this lineage last circulated at least 18 months prior to its first detection. Although BA.2.87.1 had immune evasion and/or transmission potential, its detection was not associated with a surge of infections and it was displaced by the globally dominant BA.2.86 lineage, JN.1, in the last few weeks of 2023. Our findings further strengthen the case for genomic surveillance through clinical and wastewater surveillance systems. SARS-CoV-2 continues to circulate and evolve within the global population. Multiple divergent Omicron lineages such as BA.1, BA.2, BA.3, BA.4, and BA.5 that have emerged from the southern African region have had a major impact on the epidemiology of the virus worldwide. This is likely driven by the large population of immunocompromised individuals due to the high burden of HIV/AIDS and TB in the region that facilitates long-term chronic infections. This article provides insights into the emergence of the BA.2.87.1 lineage, which briefly circulated in South Africa. The lineage displayed a unique mutational profile, including major substitutions in the receptor-binding domain and N-terminal domain deletions. The study also highlights the critical role of syndromic and wastewater surveillance in monitoring the circulation and evolution of SARS-CoV-2.

Keywords: virus evolution; SARS-CoV-2; within-host evolution; variants of concern; South Africa; omicron; BA.2.87.1

Introduction

Following the COVID-19 pandemic, SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) continues to circulate globally with persistently high rates of viral evolution (Pollard et al. 2020, Sohrabi et al. 2020, Faghy et al. 2022). Within months of the November/December 2019 emergence of SARS-CoV-2 in Wuhan, China (Zhu et al. 2020), various potentially concerning viral lineages had emerged: referred to by the World Health Organization as variants of concern (VOCs), variants of interest (VOIs), or variants under monitoring (VUMs). These lineages contained constellations of mutations that provided them with combinations of increased transmissibility and immune evasiveness. In some cases, viruses of the most rapidly spreading lineages, VOCs such as the Delta variant (Vulturar et al. 2024) also displayed increased potential to cause severe disease (Markov et al. 2023, World Health Organization 2024).

The emergence of the Omicron VOC in South Africa in October 2021 (Viana et al. 2022) resulted in the largest global wave of SARS-CoV-2 infections of the pandemic (Karim and Karim 2021, Si et al. 2023). As with some other VOCs such as Alpha, Beta, and Gamma, Omicron was highly mutated relative to previously known lineages (Karim and Karim 2021). Within 4 weeks of the detection of three genetically distinct Omicron lineages in November 2021 (BA.1, BA.2, and BA.3), the BA.1 lineage effectively displaced the Delta VOC that was dominating global circulation at the time (Viana et al. 2022, WHO 2022). However, it was ultimately BA.2 lineages that dominated the global SARS-CoV-2 landscape: effectively displacing BA.1 by approximately April 2022 (Tiecco et al. 2022, WHO 2022). The rapid rise of BA.2 to dominance was attributed to its greater transmissibility compared to BA.1, which was due to its Spike protein having a higher ACE2 receptor binding affinity than BA.1 (Li et al. 2022) and its ability to evade population-wide immunity acquired through vaccination and/or prior infection (Wang et al. 2023). The BA.4 and BA.5 lineages were first detected in January 2022 (Tegally et al. 2022); both were most closely related to BA.2 but were genetically distinct and did not directly descend from BA.2. Although both lineages spread globally and dominated SARS-CoV-2 infections throughout the world, they were largely displaced in early 2023 by a resurgent BA.2 lineage XBB: a recombinant of two divergent BA.2 lineages first described in September 2022 (Tamura et al. 2023). The global dominance of XBB ended in December 2023 after the rise of another highly divergent BA.2 lineage: BA.2.86. Although first detected in Israel and Denmark, BA.2.86 likely also originated in the same part of southern Africa as BA.1, BA.2, BA.3, BA.4, and BA.5 lineages (Khan et al. 2023).

Here, we report the detection and genome analysis of BA.2.87.1, another BA.2-related lineage that is highly divergent relative to all previously detected Omicron lineages. Only 10 infections with this lineage have been confirmed by whole genome sequencing. All specimens from these infections were sampled between 20 September and 12 December 2023 from three South African provinces (Limpopo, Gauteng, and Mpumalanga) through the national SARS-CoV-2 laboratory-based genomic surveillance program and the sentinel syndromic respiratory illness surveillance programmes of the South African National Institute for Communicable Diseases (NICD). During this period, evidence of the presence of this lineage was also detected in 31 wastewater samples from Gauteng Province. One partial, contaminated sequence from a traveller to the USA in October 2023 contained signs of BA.2.87.1 (EPI_ISL_18421941), and three spike-only wastewater BA.2.87.1 sequences from Thailand were detected in Dec 2023/Jan 2024 (EPI_ISL_18968820-18 968 822). Here, we genetically characterize this new lineage, we use wastewater sequencing data to track its rise in prevalence, and lastly, we explore some of its unusual genomic features.

Materials and methods

Ethics approval

The SARI and SRI protocols were approved by the University of the Witwatersrand Human Research Ethics Committee (HREC) (M081042) and the University of KwaZulu-Natal Human Biomedical Research Ethics Committee (BREC) (BF157/08). The ILI protocol was approved by HREC and BREC protocol numbers M120133 and BF080/12, respectively. Ethical approval for ILI-VW was obtained from the University of the Witwatersrand Research Ethics Committee. All participants in the surveillance program gave written informed consent to participate. Patient information was anonymized and de-identified prior to analysis.

The SARS-CoV-2 genomic surveillance efforts in South Africa have been approved by Research Ethics Committees at the University of KwaZulu-Natal (BREC/00001510/2020), the University of the Witwatersrand (M180832), Stellenbosch University (N20/04/008_COVID-19), the University of Cape Town (383/2020), the University of Pretoria (H101/17) and the University of the Free State (UFS-HSD2020/1860/2710). Individual participant consent was not required for genomic surveillance. This requirement was waived by the respective research ethics committees. Wastewater testing did not involve human participants. An ethics waiver was obtained from the Human Research Ethics Committee of the University of the Witwatersrand (number R14/49).

Clinical specimens

SARS-CoV-2 specimens were obtained through syndromic surveillance programmes in South Africa for ILI among outpatients and severe respiratory illness (SRI) among hospitalized patients (NICD Weekly respiratory pathogens surveillance report, 2024), as well as from diagnostic specimens collected through SARS-CoV-2 genomic surveillance performed by the Network for Genomics Surveillance in South Africa (NGS-SA) (Msomi et al. 2020). Samples were nasal or nasopharyngeal swabs in a viral transport medium.

Wastewater samples

As part of routine national wastewater sampling 1 L 'grab' samples are collected weekly from 28 national sentinel sampling sites around South Africa (Iwu-Jaja et al. 2023) commencing in 2021, and twice weekly from 21 sub-catchment areas within the districts of Gauteng Province, commencing in 2023.

Whole genome sequencing

Clinical and wastewater samples underwent Illumina-based whole-genome sequencing with clinical sequences being confirmed using the Ion Torrent platform.

Illumina-based whole genome sequencing

RNA extraction for clinical specimens was performed on a Chemagic 360 using a CMG-1049 kit (PerkinElmer, Massachusetts, USA). Sequencing was performed using the Illumina COVIDSeq protocol (Illumina Inc., San Diego, CA, USA), an amplicon-based next-generation sequencing approach, using ARTIC version 4 SARS-CoV-2 primer pools. Pooled polymerase chain reaction (PCR) -amplification products were processed for tagmentation and adapter ligation using Integrated DNA Technologies (IDT)

for Illumina indexes (Illumina Inc.). Libraries were quantified using a Qubit 4.01 fluorometer (Invitrogen, Waltham, MA, USA) and the Qubit dsDNA High Sensitivity assay according to the manufacturer's instructions. Fragment sizes were analysed using a TapeStation 4200 (Invitrogen). Libraries were pooled and normalised to 1 nM concentration with a 10% PhiX spike-in. Libraries were loaded onto a 300-cycle NextSeq P1/P2 reagent cartridge and run on an Illumina NextSeq 1000/2000 instrument (Illumina Inc.).

Ion torrent whole genome sequencing

Automated extraction of viral RNA from clinical specimens was performed on a MagNA Pure 96 system (Roche Diagnostics, Rotkreuz, Switzerland), using the MagNA Pure DNA and Viral Nucleic Acid kit according to the manufacturer's instructions. Extracts were screened by quantitative PCR and mean cycle threshold (Ct) values for the SARS-CoV-2 N and ORF1ab genes using the TaqMan 2019-nCoV assay kit v1 (Thermo Fisher Scientific, Waltham, MA, USA) on the ViiA7 Real-time PCR system (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. Extracts were sequenced using the low viral titre protocol. Sequencing was performed using the Ion AmpliSeq SARS-CoV-2 Research Panel on the Ion Torrent Genexus Integrated Sequencer (Thermo Fisher Scientific, Waltham, MA, USA), which combines automated complementary DNA (cDNA) synthesis, library preparation, and template preparation. The panel consists of two primer pools targeting 237 amplicons tiled across the ~30 Kb SARS-CoV-2 genome providing >99% genome coverage, together with an additional five primer pairs targeting human expression controls. The SARS-CoV-2 amplicons ranged from 125 to 275 bp in length.

Illumina sequencing of wastewater samples

Sample concentration, nucleic acid extraction, and sequencing were performed according to the protocols described previously (Iwu-Jaja et al. 2023). Briefly, viruses were concentrated from samples by centrifuging 250 ml aliquots at 4650×*g* at 4°C for 30 min. The pellet was resuspended in 70 ml of water and filtered through a centricon (Merck, Germany) filter for ultrafiltration. RNA was extracted from the concentrated and filtered sample using the MagMAX Wastewater Ultra Nucleic Acid Isolation Kit (LTC, Austin, TX, USA) on a KingFisher Flex instrument. SARS-CoV-2 was detected by digital PCR using QIAcuity OneStep Advance Probe kit (Qiagen, Germany). Positive samples were then sequenced using ARTIC version 4 primers and the COVIDSeq protocol (Illumina Inc.).

Genome assembly, quality control, lineage, and clade calling

Raw Illumina sequencing reads from clinical specimens were assembled using the Exatype NGS SARS-CoV-2 pipeline v4.2.5 (https://sars-cov-2.exatype.com/), with default parameter settings (10% minimum prevalence to report nucleotide sequence variants, 80% minimum prevalence to include a nucleotide sequence variant in a consensus sequence). The pipeline performs quality control on reads and then maps the reads to a SARS-CoV-2 reference genome (NC_045512.2, accession number MN908947.3) using Examap.

Samples sequenced with the Ion Torrent method were assembled using the SARS-CoV-2 RECoVERY (Reconstruction of Coronavirus Genomes & Rapid Analysis) pipeline implemented in the Galaxy instance ARIES (https://aries.iss.it). Quality control was performed using Qualimap, and all consensus sequences were manually inspected and edited using Aliview v.1.28 (Larsson 2014) (http://ormbunkar.se/aliview/).

Nextclade (https://clades.nextstrain.org) (v3.1.0) (Aksamentov et al. 2021) and Pangolin (v4.3.1) (O'Toole et al. 2021) were used for clade and lineage assignments, respectively. Nextclade was also used for visualization of the sequences and the identification of mutations. Unexpected frameshifts that were plausibly attributable to sequencing errors were manually corrected using Aliview (v1.28). Additionally, the Gensplore genome browser (https://github.com/theosanderson/gensplore), was also used to visualize the sequences.

Phylogenetic analyses

Ten BA.2.87.1 sequences were aligned with 182 Omicron sequences from the Nextstrain SARS-CoV-2 workflow (https://github.com/nextstrain/ncov) reference dataset and 59 additional BA.2.87 sequences using MAFFT (total $n = 251$) (Katoh 2002) (Supplementary Table 1). The alignment was analysed for recombination using Recombination Detection Program (RDP5) (Martin et al. 2021) with default settings, except that sequences were treated as linear. Window sizes for the Maxchi, Chimaera, and RDP5 methods were set to five variable nucleotide sites and the window sizes for the SiScan and Recscan methods were set to 2000 alignment sites. A maximum likelihood phylogenetic tree was constructed using IQ-TREE (https://iqtree.github.io/), under the TIM + F + R2 model as selected by the automatic model selection. To assess branch support, an approximate likelihood-ratio test based on the Shimodaira–Hasegawa procedure (SH-aLRT) with 1000 replicates was used. For the interactive nextstrain tree using the same dataset, genomes were aligned with Nextalign (v 2.14.0) against Wuhan-Hu-1 (MN908947.3) and we applied the Nextstrain SARS-CoV-2 masking scheme (mask 5′/3′ ends and known problematic/homoplasic sites) prior to tree inference (default SARS-CoV-2 masking). The build produced was visualized using Auspice (v 2.66.0).

A regression plot of root-to-tip genetic distances along this tree versus sampling dates (day, month, and year) was constructed using TempEst version 1.5.3 (Rambaut et al. 2016). This regression was used to assess the strength of the temporal signal within the data. Bayesian phylogenies were inferred using BEAST version 1.8.4 (Drummond and Rambaut 2007), applying a HKY + I + G nucleotide substitution model under a strict molecular clock with BEAST's default CTMC scale prior (1/x) on the mean clock rate; no informative constraints were imposed on the rate (parameter clock.rate, initialized at 1.0). Furthermore, a coalescent exponential population prior distribution was applied and specimen collection dates (day, month, year) were used as tip dates. The length of the Markov Chain was set at 100 000 000 steps, and parameters were logged every 10 000 steps. Tracer version 1.7.1 (Rambaut et al. 2018) was used to visualize the log files. Once the run had completed, the effective sample size for the joint statistic was >700, indicating that the Markov chain Monte Carlo (MCMC) runs had converged. A maximum clade credibility tree was generated from the posterior distribution of sampled trees using TreeAnnotator version 1.8.4 (https://beast.community/treeannotator) after discarding the first 10% tree samples as burn-in. The maximum clade credibility (MCC) tree was visualized using ggtree in RStudio (G. Yu et al. 2017). Mutation frequency data and lineage prevalence statistics were obtained using covSPECTRUM (https://cov-spectrum.org/). For associations between the occurrence of C66T and C28256T mutations, covSPECTRUM searches were designed to detect the number of sequences with coverage at both nucleotide sites 66 and 28 256 and neither C66T nor C28256T, with both C66T and C28256T, with C66T but not C28256T and with C28256T but not C66T (Table 3). These four numbers were then used to perform 2 × 2 chi-square tests. It must be emphasized that these

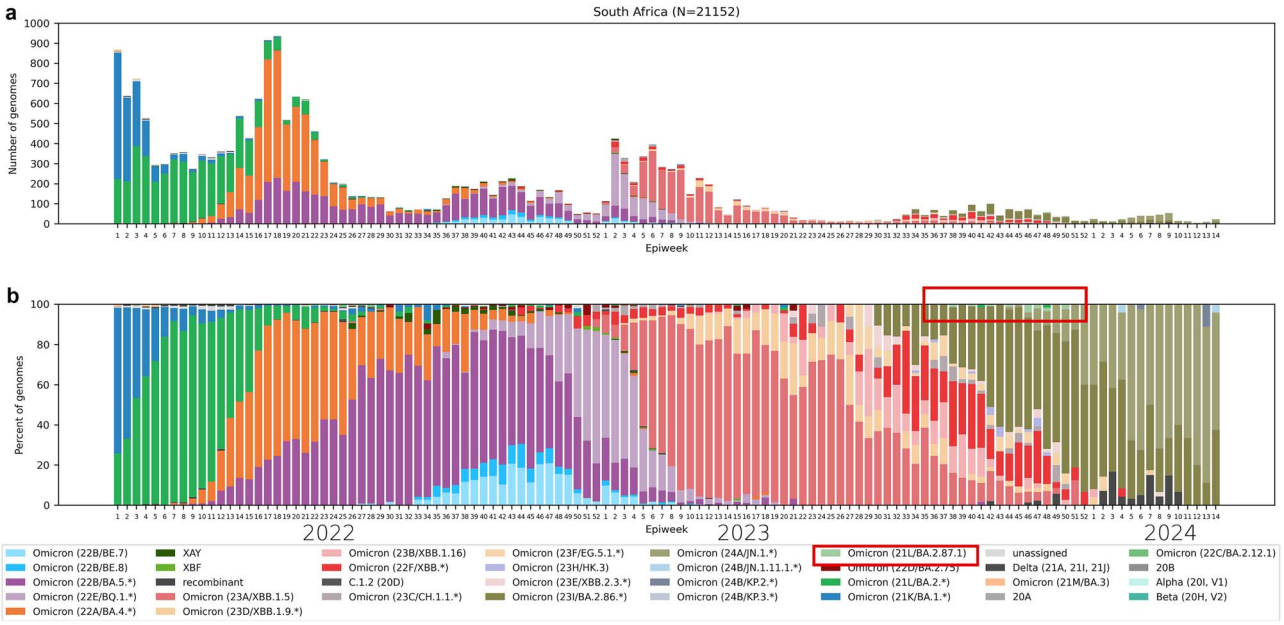

Figure 1. Prevalence of SARS-CoV-2 lineages between January 2022 and March 2024 in South Africa. The BA.2.87.1 lineage is indicated by the block. (a) Number of genomes and (b) proportions of lineages per epiweek.

covSPECTRUM searches yielded the total numbers of sequences with or without these mutations and not the numbers of independent occurrences of C66T and C28256T mutation events.

Wastewater sequence analysis

FASTQ files were trimmed, filtered based on sequence quality, assembled, and mapped to the reference genome (NC_045512.2) according to published criteria using the web-based Exatype tool (https://sars-cov-2.exatype.com/). Only samples with a minimum of 1 000 000 reads, a Phred score of 30 or more, and a length of >100 bp were processed for mutational analyses. To capture the dynamics of virus evolution and spread, we used Freyja (https://github.com/andersen-lab/Freyja) to estimate the relative abundance of virus lineages present in wastewater. Freyja uses a 'barcode' library of lineage-defining mutations to uniquely define all known SARS-CoV-2 lineages and solves for lineage abundance using a depth-weighted least absolute deviation regression approach. Wastewater samples with at least 50% genome coverage were included in Freyja analyses. Additionally, an amino acid variation file was generated using Exatype, which was then used to analyse Spike protein mutations evident within each sequenced sample. A custom R script (https://github.com/setshabaTaukobong/Mutational-Dotplot) was used to generate a mutation profile report.

Results and discussion

Description of cases

During routine genomic surveillance between September and December 2023, the NICD identified 10 divergent SARS-CoV-2 genomic sequences (1.62% (10/615) of all NICD sequences on GISAID during the same period) (Fig. 1) from the Limpopo ($n = 3$), Gauteng ($n = 5$), and Mpumalanga ($n = 2$) provinces of South Africa (Fig. 2a). The first detected case was in an individual aged 90–99 years from Gauteng (Table 1). The sequences exhibited a genetic profile distinct from the dominant lineages of Omicron (BA.2.86 and JN.1) that were circulating in South Africa in the latter half of 2023 (Fig. 3). They were initially determined by the PANGO lineage designation team to be most closely related to the BA.2.87 lineage last detected in South Africa in June 2022, prompting assignment of the sequences to the newly designated lineage, BA.2.87.1 (Pangolin issue #2484; https://github.com/cov-lineages/pango-designation/issues/2484). To date (21 June 2025), no other genome sequence belonging to this lineage has been detected in South Africa. Demographic and clinical characteristics of the individuals from whom the 10 BA.2.87.1 specimens were collected are shown in Table 1.

Epidemiological backdrop of the emergence and disappearance of BA.2.87.1

Since the fifth pandemic wave, attributed to BA.4/BA.5 lineages, SARS-CoV-2 cases remained relatively stable in the public sector surveillance programmes up until January 2025. In 2023, there was an increase in SARS-CoV-2 test positivity rates (6%–20%) in public sector influenza-like illness (ILI) surveillance between Weeks 1 and 6, coinciding with the detection of BQ.1 and XBB.1.5 variants. Thereafter, the positivity rate remained low and stable until an uptick in cases around Week 33 of 2023 in both private and public sector ILI surveillance, peaking in Week 43 of 2023 in private sector ILI surveillance with a positivity rate of 31%. This increase in positivity rate was likely due to the rising dominance of BA.2.86 and its descendant lineages (National Institute for Communicable Diseases 2023). In 2024, low overall numbers of positive specimens were reported with an average positivity rate of <10% observed in the first 11 weeks, following the detection of BA.2.87.1 (National Institute for Communicable Diseases 2024) (Supplementary Figure 1).

Mutational profile of BA.2.87.1

Relative to the Wuhan-Hu-1 reference genome, the BA.2.87.1 sequences have up to 123 mutations (Fig. 4a), with 43 located in the Spike gene (Fig. 4b). These include two large deletions in the antigenic supersite of the NTD (Δ15–26 and Δ136–146), mutations at various important antigenic sites in the receptor-binding

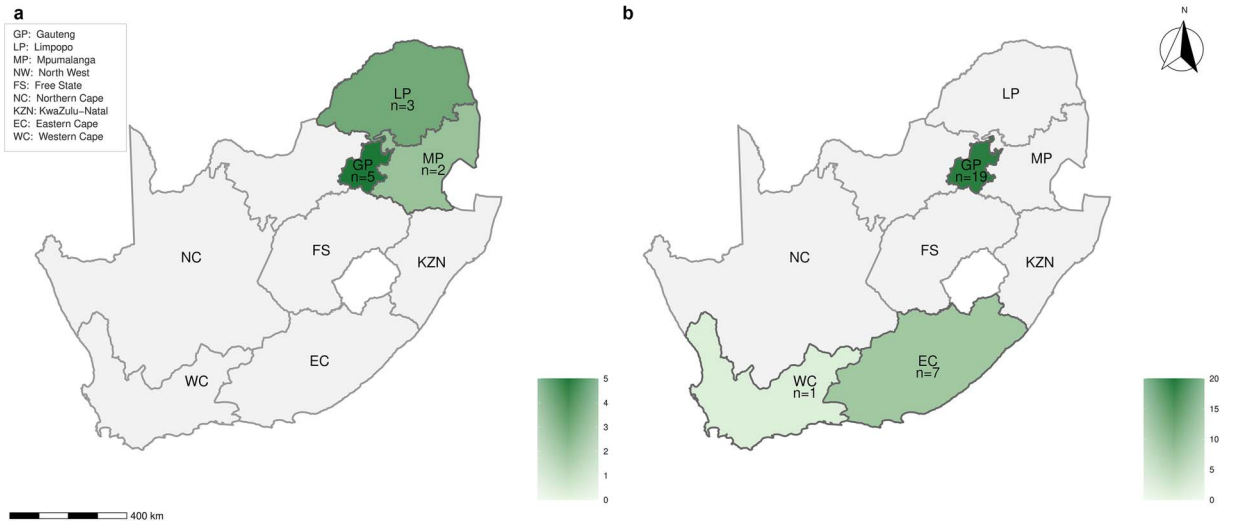

Figure 2. Map of South Africa highlighting the detection of BA.2.87.1 lineage in (a) clinical cases and (b) wastewater samples between August and December 2023. Administrative boundaries: Natural Earth (public domain) accessed *via* rnaturalearth (R); projected to Africa Albers Equal Area.

Table 1. Demographic and clinical characteristics of individuals with BA.2.87.1 infection, South Africa, September–December 2023

	Age group (years)	Sex	Collection date (epidemiological week)	Province	Lab sector	Underlying conditions	SARS-Cov-2 vaccination	Inpatient/outpatient
Sample 1	90–99	F	20-Sep-23 (38)	Gauteng	Private	Unknown	Unknown	Unknown
Sample 2	50–59	F	07-Oct-23 (40)	Gauteng	Private	Unknown	Unknown	Unknown
Sample 3	60–69	F	21-Oct-23 (42)	Gauteng	Private	Unknown	Unknown	Unknown
Sample 4	0–9	M	02-Nov-23 (44)	Limpopo	Private	Unknown	Not vaccinated	Unknown
Sample 5	0–9	M	12-Nov-23 (46)	Limpopo	Private	Unknown	Not vaccinated	Unknown
Sample 6	20–29	M	13-Nov-23 (46)	Gauteng	Private	Unknown	Unknown	Unknown
Sample 7	0–9	M	14-Nov-23 (46)	Mpumalanga	Public	None	Not vaccinated	Inpatient
Sample 8	50–59	F	21-Nov-23 (47)	Limpopo	Private	Unknown	Unknown	Unknown
Sample 9	30–39	F	29-Nov-23 (48)	Gauteng	Private	None	2 doses, both Pfizer	Outpatient
Sample 10	20–29	F	12-Dec-23 (50)	Mpumalanga	Public	HIV-infected on ART	Not vaccinated	Inpatient

Abbreviations: F, female; M, male; ART, antiretroviral therapy.

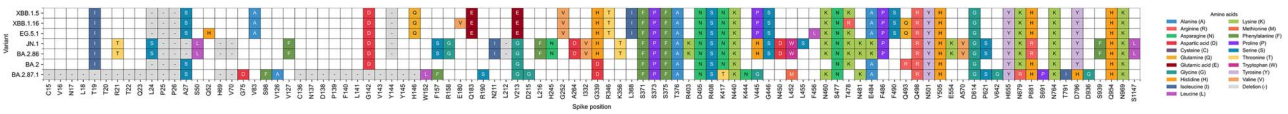

Figure 3. Spike amino acid mutations in the SARS-CoV-2 variants circulating between August 2023 and December 2023 in South Africa. XBB.1.5, XBB.1.16, *E.G.5.1*, JN.1, BA.2.86, BA.2, and BA.2.87.1 compared to the Wuhan-Hu-1 reference genome (MN908947).

domain (including lineage defining mutations K417T, K444N, V445G, L452M, N460K, and N481K), mutations close to the furin cleavage site (including lineage defining mutations N679R and S691P), and two lineage defining mutations close to the S2' cleavage site (T791I and Y796H). Relative to the BA.2 reference genome, BA.2.87.1 also carries reversions at Spike T478K and R493Q. For a more detailed discussion of the potential adaptive value of some of these Spike mutations, please consult the supplementary material. Outside of Spike, BA.2.87.1 has additional mutations in the ORF1a, ORF1b, ORF9b, E, M, and N genes (Fig. 4a).

Wastewater sequencing tracks the emergence and disappearance of BA.2.87.1

To better understand the rise and fall of BA.2.87.1 detection, we searched wastewater sequencing datasets from 2022 to 2024 generated from a routine national and sub-catchment wastewater surveillance network (Iwu-Jaja et al. 2023) for evidence of its presence. We identified samples with sequence reads carrying the two Spike NTD deletion events characteristic of BA.2.87.1 (Δ15–26 and Δ136–146), as well as samples in which the Freyja tool (https://github.com/andersen-lab/Freyja) determined that BA.2.87.1 was present. Between August and the end of November 2023, 31/720 (4.3%) wastewater samples collected across South Africa were identified that had evidence of the presence of BA.2.87.1 (Table 2; Fig. 5) (NTD deletions, Δ15–26 and Δ136–146, were identified in five samples); therefore, the lineage was detected in wastewater a month before clinical detection in September. All but two samples were collected in Gauteng Province, of which 65.5% (19/29) were collected from inspection holes and the balance from wastewater treatment plant influent. No samples with evidence of BA.2.87.1 were found in Limpopo Province, where only two wastewater treatment plants submit surveillance samples. However, single wastewater samples from the Western and Eastern Cape provinces had

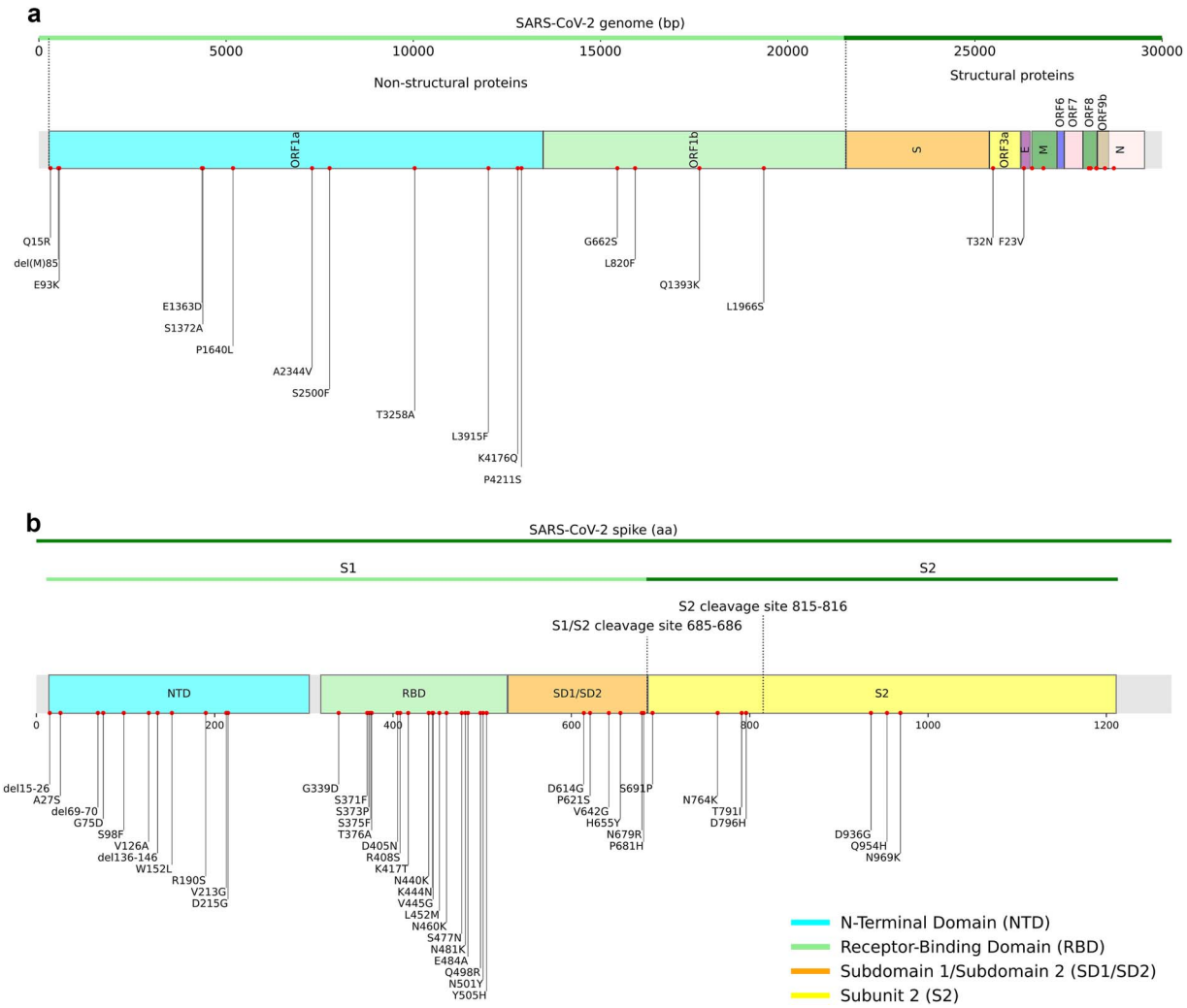

Figure 4. Genome-wide and Spike amino-acid substitutions and deletions in SARS-CoV-2 BA.2.87.1 relative to the Wuhan-Hu-1 reference genome. (a) Amino-acid substitutions and deletions in regions outside Spike. (b) Amino-acid substitutions and deletions in the Spike protein with domain boundaries indicated. Abbreviations: NTD, N-terminal domain; RBD, receptor-binding domain; SD1, subdomain 1; SD2, subdomain 2; S1, subunit 1; S2, subunit 2.

evidence of BA.2.87.1, despite no sequenced clinical samples from these provinces revealing this unique variant (Fig. 2b). In Gauteng Province, samples were predominantly found in a single metropolitan area, namely, not only in Tshwane (19 samples) but also in Ekurhuleni (8 samples) and Johannesburg (8 samples). BA.2.87.1 was not detected in any wastewater samples from national wastewater surveillance after November 2023. Amongst all positive wastewater samples, SARS-CoV-2 N gene concentration ranged from 2.16–579.64 genome copies per reaction, whilst the read frequency of BA.2.87.1 compatible strains ranged from <1% (three samples) to 100% (four samples).

It was firmly established globally and locally in South Africa (Iwu-Jaja et al. 2023) during the COVID-19 pandemic, that wastewater surveillance reflects the burden of clinical SARS-CoV-2 infections, and displays the same contemporaneous (or broader) genomic diversity of variants prevalent in sequenced isolates obtained from clinical strains. This is perennially true, regardless of the peak/wave/stage of the pandemic, by virtue of the fact that infected persons will always shed SARS-CoV-2 into wastewater.

The identification of the BA 2.87.1 variant in wastewater, as highlighted in this paper, provides supportive evidence that this variant was circulating in communities in a more widespread distribution than clinical surveillance findings suggested. It is plausible that the variant may have transiently spread to other regions, especially given the nature of population movement between provinces. The lack of corresponding clinical samples may be due to limited clinical sequencing coverage or sample availability during that period. Furthermore, the detection of these sequences in wastewater but not among sequenced strains from these geographical locations points to the sensitivity of wastewater as a surveillance tool. Our wastewater and clinical findings are indeed linked—in that the same variant BA.2.87.1 was detected by both surveillance modalities, pointing to the fact that transmission of this variant was occurring in a broader time and geographical range than suggested by the 10 clinical samples bearing BA.2.87.1 virus.

BA.2.87.1 mutations likely alter its gene expression

A noteworthy feature of the BA.2.87.1 genome is a rare C66T substitution in the 5′ untranslated region (UTR; only found in 0.13% of globally sequenced genomes as of 23 March 2024), which may alter the relative and/or absolute production of the different SARS-CoV-2 subgenomic RNAs (Fig. 6).

Table 2. Collection date, location, and SARS-CoV-2 sequence coverage of wastewater samples identified by Freyja tool or single-nucleotide polymorphism analysis with mutational profiles suggestive of BA.2.87.1

Sample ID	Collection date (epidemiological week)	Geographic location	Description of sampling point	Province	Population size upstream of sampling site	SARS-CoV-2 concentration (gc/ul)	SARS-CoV-2 genome coverage	BA.2.87.1 deletion	% of reads with BA.2.87.1 traits
23-0706	2023/08/22 (34)	Tshwane MM	IH, higher socio-economic area	Gauteng	17 041	4,25	81	[a]	11
23-0707	2023/08/22 (34)	Tshwane MM	IH, higher socio-economic area	Gauteng	21 918	12,24	94	[a]	20
23-0828	2023/09/15 (37)	Tshwane MM	WWTP influent	Gauteng	245 915	42,73	98	[a]	10
23-0831	2023/09/15 (37)	Tshwane MM	IH, lower socio-economic area	Gauteng	1210	30,61	50	[a]	100
23-0835	2023/09/18 (38)	Ekurhuleni MM	IH, lower socio-economic area	Gauteng	6639	117,11	96	[a]	62
23-0840	2023/09/19 (38)	Tshwane MM	WWTP influent	Gauteng	245 915	33,77	97	[a]	2
23-0846	2023/09/12 (38)	Buffalo City MM	WWTP influent	Eastern Cape	154 576	65,57	57	[a]	1
23-0865	2023/09/22 (38)	Tshwane MM	WWTP influent	Gauteng	245 915	4,26	98	[a]	14
23-0872	2023/09/26 (39)	Tshwane MM	WWTP influent	Gauteng	245 915	89,84	99	[a]	7
23-0881	2023/09/27 (39)	Johannesburg MM	IH, lower socio-economic area	Gauteng	unk	84,38	96	[a]	2
23-0913	2023/10/03 (40)	Tshwane MM	IH, higher socio-economic area	Gauteng	21 918	10,95	98	[a]	0
23-0939	2023/10/06 (40)	Tshwane MM	Tertiary hospital effluent	Gauteng	3000	32,65	78	[a]	100
23-0951	2023/10/10 (41)	Tshwane MM	WWTP influent	Gauteng	245 915	14,74	96	[a]	8
23-0953	2023/10/10 (41)	Tshwane MM	Tertiary hospital effluent	Gauteng	3000	87	92	[a]	10
23-0954	2023/10/10 (41)	Tshwane MM	IH, lower socio-economic area	Gauteng	7048	6,08	95	[a]	22
23-0960	2023/09/10 (37)	Cape Town MM	WWTP influent	Western Cape	460 000	11,83	99	[a]	4
23-1 051	2023/10/25 (43)	Johannesburg MM	IH, lower socio-economic area	Gauteng	unk	4,22	95	Δ15–26	<1
23-1 061	2023/10/26 (43)	Ekurhuleni MM	IH, higher socio-economic area	Gauteng	14 285	27,25	95	Δ136–146	3
23-1 062	2023/10/26 (43)	Ekurhuleni MM	IH, lower socio-economic area	Gauteng	23 934	151,66	94	[a]	1
23-1 063	2023/10/26 (43)	Ekurhuleni MM	IH, lower socio-economic area	Gauteng	79 144	90,59	66	[a]	2
23-1 079	2023/10/30 (44)	Ekurhuleni MM	IH, lower socio-economic area	Gauteng	6639	2,16	97	[a]	0
23-1 085	2023/10/31 (44)	Tshwane MM	Tertiary hospital effluent	Gauteng	3000	10,54	98	Δ15–26 and Δ136–146	<1
23-1087	2023/10/31 (44)	Tshwane MM	IH, higher socio-economic area	Gauteng	21 918	579,64	79	[a]	<1
23-1114	2023/11/06 (45)	Ekurhuleni MM	IH, higher socio-economic area	Gauteng	14 285	14,11	98	[a]	2
23-1115	2023/11/06 (45)	Ekurhuleni MM	WWTP influent	Gauteng	905 996	69,48	98	[a]	6
23-1124	2023/11/07 (45)	Tshwane MM	WWTP influent	Gauteng	245 915	22,36	81	[a]	2
23-1127	2023/11/07 (45)	Ekurhuleni MM	WWTP influent	Gauteng	unk	44,37	98	[a]	4
23-1146	2023/11/10 (45)	Tshwane MM	Institution effluent	Gauteng	7337	173,57	98	Δ15–26 and Δ136–146	100
23-1147	2023/11/10 (45)	Tshwane MM	IH, lower socio-economic area	Gauteng	7048	not recorded	98	[a]	100
23-1148	2023/11/10 (45)	Tshwane MM	IH, lower socio-economic area	Gauteng	1210	2,17	82	Δ15–26 and Δ136–146	1
23-1156	2023/11/14 (46)	Tshwane MM	WWTP influent	Gauteng	245 915	40,69	97	[a]	10

Freyja has used these nucleotide changes for BA.2.87.1 classification: ['A10037G', 'A11782G', 'A12791C', 'A22206G', 'A22812C', 'A23598G', 'A24369G', 'A28715G', 'C10702T', 'C12008T', 'C12896T', 'C15925T', 'C17644A', 'C19698A', 'C21855T', 'C22916A', 'C23423T', 'C23934T', 'C2536T', 'C25487A', 'C28045T', 'C28256T', 'C4012T', 'C5184T', 'C541T', 'C7296T', 'C7764T', 'G21604A', 'G21786A', 'G22017T', 'G22132T', 'G22894T', 'G23761A', 'G23948C', 'G26529A', 'G4354T', 'G542A', 'T19364C', 'T21939C', 'T22552C', 'T22896G', 'T22942G', 'T23005G', 'T23487G', 'T26311G', 'T28474C', 'T4379G', 'T7480C']. [a]Deletion not detected.

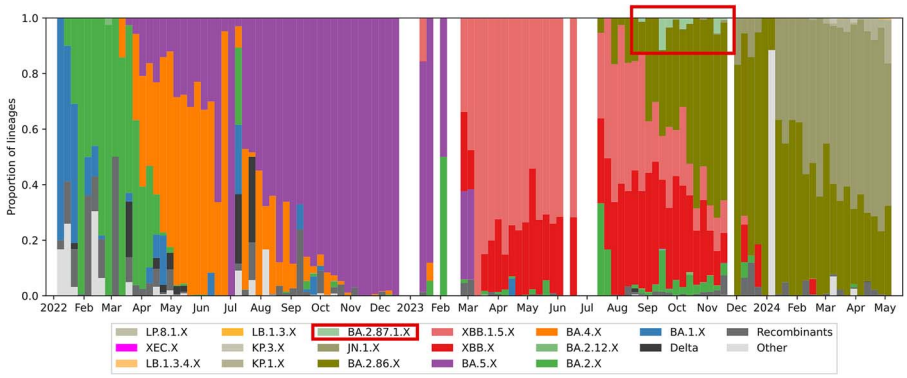

Figure 5. Nationwide wastewater genomic surveillance in South Africa. The proportion of lineages by month from wastewater samples, from December 2022 to March 2024 estimated using the Freyja tool. Only samples with sequence coverage of >50% were included. The BA.2.87.1 lineage is indicated by the block.

Table 3. covSPECTRUM search conditions for co-occurrence and exclusivity of C66T and C28256T mutations

Category	covSPECTRUM search conditions
Coverage at both sites, neither C66T nor C28256T	[2-of: T65T, T67T] & [1-of: C66C, C66A, C66G] & [2-of: T28255T, T28257T] & [1-of: C28256C, C28256A, C28256G]
Both C66T and C28256T	[3-of: T65T, C66T, T67T] & [3-of: T28255T, C28256T, T28257T]
C66T but not C28256T	[3-of: T65T, C66T, T67T] & [2-of: T28255T, T28257T] & [1-of: C28256C, C28256A, C28256G]
C28256T but not C66T	[2-of: T65T, T67T] & [1-of: C66C, C66A, C66G] & [3-of: T28255T, C28256T, T28257T]

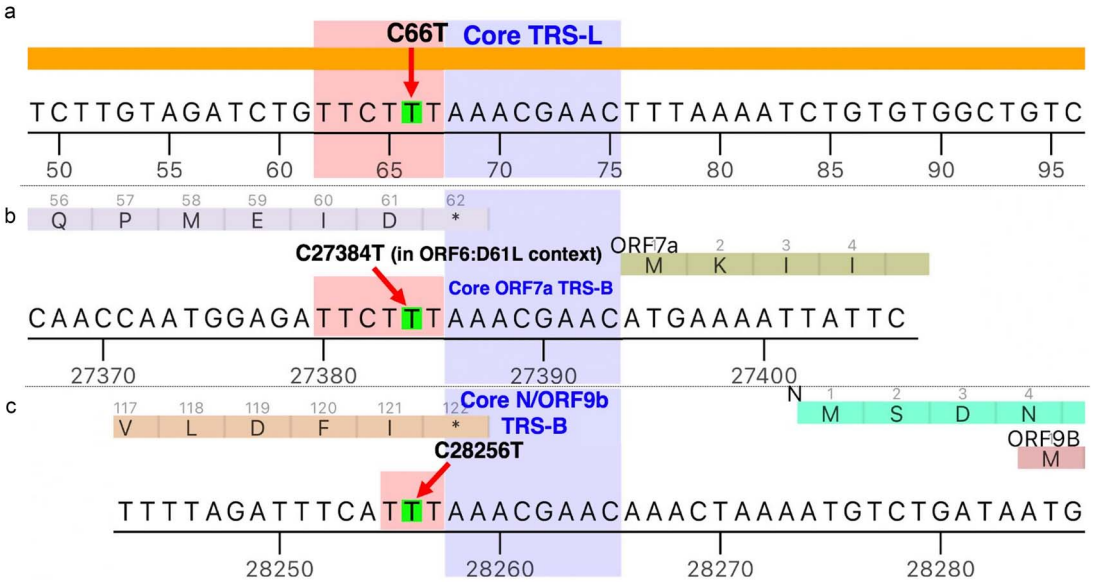

Figure 6. Nucleotide substitutions that may impact the gene expression of BA.2.87.1 by creating a mismatch in the TRS-L and TRS-B sequences of Spike, M, ORF7a (in lineages with ORF6:D61L), ORF8, and N/ORF9b. (a) 5′ leader (TRS-L) context. The C66T substitution falls within the core TRS-L motif and reduces complementarity to downstream TRS-B cores. Arrows mark the mutated base(s); . (b) ORF7a TRS-B region (shown in lineages carrying ORF6:D61L context). The C27384T change introduces a mismatch within the ORF7a TRS-B core. (c) N/ORF9b TRS-B region. The C28256T change creates a mismatch in the shared N/ORF9b TRS-B site.

All coronaviruses produce subgenomic RNAs (sgRNAs) that code for the four structural proteins (S, E, M, and N) and all accessory proteins (of which SARS-CoV-2 has at least six), located in the 3′-most third of the genome (Sola et al. 2015). These sgRNAs are produced through a discontinuous transcription mechanism. The genomic leader, starting at nt 68 in the SARS-CoV-2 genome contains the core ACGAAC transcription regulatory sequence leader (TRS-L). Upstream of each subgenomic ORF is a TRS-B (body) sequence that matches the TRS-L sequence (V'kovski et al. 2021). The more extensive the match between the TRS-B and the TRS-L regions, the more frequently the RdRp switches templates to produce corresponding sgRNA molecules during negative strand synthesis (Sola et al. 2005). In all BA.2 lineages, five of the eight TRS-Bs, in addition to matching the core TRS-L sequence (AAACGAAC), also match at least two nucleotides immediately upstream of the core TRS-L (**CT**AAACGAAC). The TRS-Bs that match this stretch of 10 TRS-L nucleotides belong to Spike, M, ORF7ab, ORF8, and N/ORF9b.

In BA.2.87.1, the C66T mutation alters the nucleotide sequence of the extended TRS-L from **C**TAAACGAAC to **T**TAAACGAAC, which should result in a less extensive match between the TRS-L and the TRS-Bs for S, M, ORF7ab, ORF8, and N/ORF9b. However, in addition to C66T, BA.2.87.1 also has C27384T and C28256T, which change the corresponding TRS-Bs for ORF7ab and N/ORF9b so that they match the new TRS-L sequence. Since Spike, M, and ORF8 no longer match the extended TRS-L at that position, their expression would be expected to decrease somewhat relative to that of ORF7ab and N/ORF9, although it remains uncertain how significant this effect might be. If Spike expression is decreased, this could conceivably be an immune evasion tactic.

C66T and C28256T substitutions appear together in circulating lineages far more frequently than would be expected by chance (chi-square with Yates correction = 1 671 912; $P < .00001$), consistent with the importance of maintaining optimal ORF9b and N expression, as has been implied by the evolutionary trend for increased expression of N and ORF9b in SARS-CoV-2 since the start of the pandemic (Parker et al. 2022, Bouhaddou et al. 2023). Though less common than C28256T, C66T sequences with the M TRS-B-associated C26469T ($P < .00001$) and the ORF7a TRS-B-associated C27384T ($P < .00001$) also occur more frequently than expected by chance. The corresponding nucleotide in the ORF8 TRS-B, C27884T, has only occurred together with C66T in sampled sequences once in the entire pandemic and never since the beginning of 2022, as one would expect given the long-term trend towards decreased expression and/or deletion of ORF8 (Hisner et al. 2023). It is also noteworthy that the C66T-C28256T combination has shown up in numerous chronic infection sequences, including a closely documented B.1.517 infection lasting more than 471 days (Chaguza et al. 2023).

Considering the phylogenetic placement of the C66T, C27384T, and C28256T mutations in the global SARS-CoV-2 USHER tree (from 23 March 2024) (Hinrichs et al. 2023), it is suggested that the order in which these mutations tend to arise is C66T first, then C28256T (although sometimes mapping to the same branch of the tree as C66T), and then C27384T. This implies that whenever the TRS-L associated C66T mutation occurs, it likely prompts the N gene–associated extended TRS-R change, followed by that of ORF7a-associated extended TRS-R change (Supplementary Figure 2). Besides indicating that this is the likely order in which the mutations arose during the genesis of the BA.2.87.1 lineage, the convergence of these mutations in multiple lineages strongly suggests that any changes in the gene expression program that they cause are likely to be adaptive in at least some situations.

The assumption that all three mutations may need to occur together to achieve any substantial adaptive advantage is further supported by the fact that, whenever C66T mutations do rarely occur, they tend to occur together with C27384T far more frequently than would be expected by chance ($P < .00001$). For more detailed discussion on the potential adaptive value of these and similar TRS-associated mutations occurring in other SARS-CoV-2 lineages, consult the supplementary material.

Phylogenetic evidence for prolonged evolution of BA.2.87.1 during chronic human infection

Given our prior detections of the complex XAY and XBA Omicron-Delta recombinants in the same regions where BA.2.87.1 was sampled (Roemer et al. 2023), we attempted but were unable to detect evidence of recombination within the BA.2.87.1 genomes using the RDP5 tool (Martin et al. 2021).

We inferred a maximum-likelihood (ML) phylogeny with IQ-TREE using ModelFinder and assessing node support with 1000 ultrafast bootstrap replicates. The dataset comprised 10 BA.2.87.1 study genomes, 182 representative Omicron genomes from the Nextstrain ncov reference set, and 59 BA.2.87 genomes to test BA.2.87.1's placement relative to BA.2.87 (total $n = 251$).

The ML phylogenetic tree (Supplementary Figure 3) yielded evidence of a strong temporal signal consistent with clock-like evolution of the Omicron lineages (correlation coefficient = 0.95, $R^2 = 0.90$; Fig. 7), thereby justifying the use of a strict molecular clock model for Bayesian phylogenetic inference. The MCC tree places BA.2.87.1 near the BA.2 root with high posterior support (posterior probabilities = [1.0]; Fig. 8) and concordant ML support (bootstrap = 100%). In neither tree does it cluster with its presumed parental lineage, BA.2.87, suggesting misassignment, although homoplasy at recurrent sites and phylogenetic model/sampling uncertainty remain alternative explanations (https://github.com/cov-lineages/pango-designation/issues/2484).

The time-scaled MCC tree suggests that BA.2.87.1 diverged from the other BA.2 lineages between 20 December 2021 and 25 March 2022 (95% highest posterior probability density (HPD) with 02 February 2022 having the highest posterior probability). It is conceivable, therefore, that the progenitor of the BA.2.87.1 branch diverged from the trunk of the BA.2 tree during the initial wave of BA.2 infections in southern Africa, at approximately the same time as the BA.4 and BA.5 lineages arose (Tegally et al. 2022). The estimated date when the most recent common ancestor of the 10 sampled BA.2.87.1 sequences existed is June 2023 (95% HPD: 06 May 2023–22 July 2023 (Fig. 8): approximately the time when BA.2.86 was discovered (Khan et al. 2023). It is apparent, therefore, in the ∼16 months between February 2022 and June 2023, the BA.2.87.1 progenitor evolved without spawning any detected intermediate lineages: a pattern of evolution that is consistent with it having originated from a single long-term chronic infection (Kemp et al. 2021, Landis et al. 2023).

Alternative explanations for the origin of BA.2.87.1 include prolonged evolution in a nonhuman host, cryptic circulation in an isolated population, and treatment-induced mutagenesis. Sequences from nonhuman hosts, such as mink and deer, have distinct mutational signatures. In the case of mink, S:Y453F, S:F486L, S:N501T, ORF1a:G4177E, ORF3a:H182Y, and ORF3a:T229I each independently emerged in separate outbreaks (Naderi et al. 2023), whilst deer sequences are marked by an overwhelming predominance of C → T mutations, mostly synonymous (Marques et al. 2025). The mutations in BA.2.87.1 do not resemble those seen in sequences from mink, deer, or any other animal. It must be conceded, however, that there have been no concerted efforts to either detect evidence of SARS-CoV-2 circulating in southern African wildlife or determine mutation signatures that might be associated with such infections. The possibility of a wildlife reservoir cannot, therefore, be excluded.

There are no known examples of prolonged (>3 months) circulation of a variant within an isolated human population with no detection outside of that population. It is therefore implausible that BA.2.87.1 continually circulated for over 16 months without being detected.

Treatment with mutagenic drugs, such as molnupiravir (MOV), is known to induce a large number of mutations, and there are numerous examples of onward transmission of such hypermutated viruses (Sanderson et al. 2023). However, the markings of such sequences, in particular a high proportion of G → A mutations and a preference for specific nucleotide contexts, are entirely lacking in the branch leading to BA.2.87.1, on which just 6 of 53 mutations are G → A, and the estimated un-normalised probability of MOV-like contexts, determined

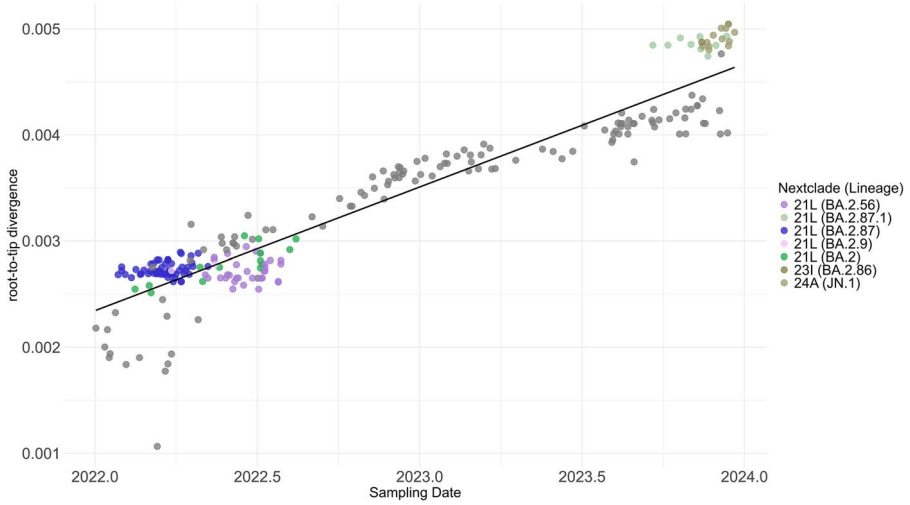

Figure 7. Root-to-tip regression plot using genetic distances against sampling dates for sequences belonging to lineage BA.2.87.1 ($n = 10$) sampled in South Africa and other Omicron lineages from a custom Nextstrain build ($n = 241$), indicating that the BA.2.87.1 sequences evolved in a clock-like manner (correlation coefficient = 0.9487, R2 = 0.9). Maximum likelihood tree constructed using IQ-TREE was used to infer genetic distances. Coloured nodes represent BA.2.* lineages, whilst grey nodes represent the rest of omicron lineages.

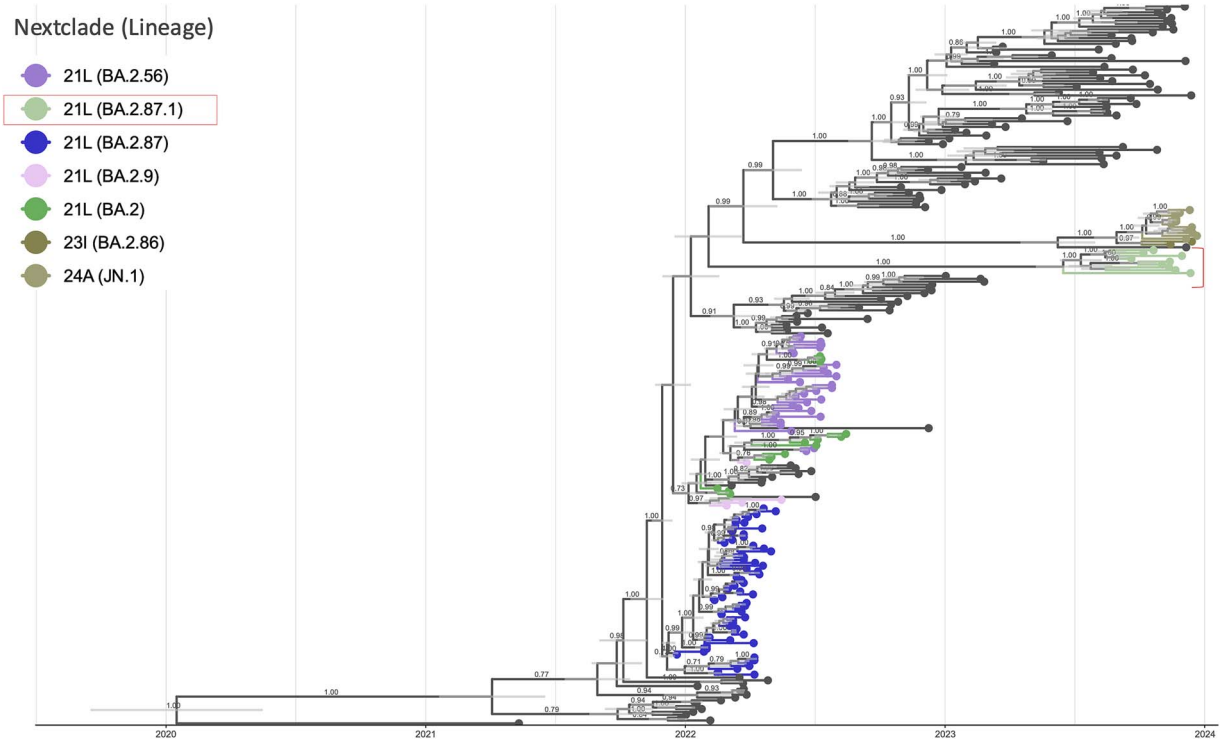

Figure 8. Maximum clade credibility tree of lineage $n = 10$ BA.2.87.1 sampled in South Africa and other Omicron lineages ($n = 241$) estimated using a BEAST strict clock and exponential coalescent model (sampled between 12 May 2021 and 21 December 2023). Numbers at internal nodes are posterior probabilities (rounded to two decimals; values < 0.70 not shown). Grey horizontal bars represent the 95% highest posterior density (HPD) interval. Coloured nodes represent BA.2.* lineages, whilst grey nodes represent the rest of Omicron lineages.

using the MOVBranchCApp tool (https://movbranchapp.streamlit.app/) is approximately zero (summed log likelihood −11.969).

The mutation profile of BA.2.87.1 is also suggestive of an origin during a chronic infection

Perhaps the most distinctive aspect of BA.2.87.1 that is consistent with an origin during a chronic infection is its large Spike NTD deletions—Δ15–26 and Δ136–146. Our examination of the available sequences indicated that such large NTD deletions seem to only occur in rarely sampled, highly mutated, highly divergent sequences originating in likely chronic infections (Kemp et al. 2021). Until BA.2.87.1 was detected, there had been no circulating lineages with such deletions.

The NTD contains five loops projecting outward from the Spike surface. Loops N1, N3, and N5 compose the NTD antigenic supersite, the primary NTD target for neutralizing antibodies. Deletions in these loops have been shown to evade neutralizing antibodies, both in other lineages (Cerutti et al. 2021; McCallum et al. 2021b; McCarthy et al. 2021), and in BA.2.87.1 (Wang et al. 2024; Yang et al. 2024).

Compared to other sarbecoviruses, SARS-CoV-2 has unusually long NTD loops (Cantoni et al. 2022). Experiments with virus-like particles suggest that shorter NTD loops not only increase infectivity and the ability to fuse with cell membranes but also lead to Spike instability and premature S1 shedding (Qing et al. 2021). This comports well with the phylogenetic evidence, which suggests that extensive NTD deletions can be advantageous within a chronically infected host but are likely to reduce transmissibility, which requires sustained viability in the inhospitable extracellular environment, particularly in aerosols, in which a variety of dissolved substances are capable of inactivating virions (Lin et al. 2020).

Remarkably, the NTD deletions in BA.2.87.1 eliminate both cysteines that form the S:C15–C136 disulfide bond. Alterations to this disulfide bond have been rare (only 25 previous sequences have had both S:C15 and S:C136 deleted, and only 0.08% of sequenced genomes since 1 January 2022 have had any C15–C136-disulfide bond–altering mutations), but they often occur in highly mutated sequences deriving from chronic infections.

Several mutations at the N-terminal end of Spike, including P9L, S12P, and S13I, have been shown to alter the location of signal peptide cleavage, resulting in the loss of S:C15 (McCallum et al. 2021a; Yu et al. 2023). Most notably, the Epsilon VOI (B.1.427/429), which spread widely in North America in early 2021, possessed S:S13I as well as S:W152C, resulting in the replacement of the C15–C136 disulfide by a new C136–C152 disulfide bond. This caused a structural rearrangement of the Epsilon NTD, enabling evasion of NTD-targeting antibodies (McCallum et al. 2021a).

A few previous minor designated lineages have, like BA.2.87.1, featured the total loss of the C15-C136 disulfide, including C.1.2 (P9L, C136F), B.1.640.1/2 (P9L, Δ136–144), AT.1 (P9L, Δ136–144), and B.1.630 (P9L, C136F). However, none of these lineages had deletions at both cysteine locations. Remarkably, of the 34 sequences that have ever had both C15 and C136 deleted, 21 have been collected in South Africa.

Though no major lineages have had NTD deletions nearly as large as those in BA.2.87.1, a large number of highly mutated sequences (also presumably originating in chronic infections) have featured such deletions. Yu et al. (2023) isolated and examined a virus with deletions similar to BA.2.87.1 (Δ14–22, Δ136–144) and found that it resulted in extensive restructuring of the NTD as well as near-complete evasion of NTD-targeting antibodies from convalescent sera. They also found that the virus harbouring these mutations displayed an increase in the 1-RBD-up conformation. This comports well with previous work with SARS-CoV-2 VLPs by Qing et al. (2021), which found that shorter NTD loops increased the frequency of the 1-RBD-up conformation.

In addition to its large NTD deletions and C15-C136 disulfide elimination, BA.2.87.1 possesses many other mutations characteristic of sequences from chronically infected individuals. Some of these mutations, such as S:N460K, S:R493Q, and ORF1b:G662S, first appeared in chronic infection sequences and subsequently—months or years later—became widespread or even universal in circulating lineages. Other BA.2.87.1 mutations, however, are almost exclusively found in highly mutated known/presumed chronic infection sequences possessing dozens of mutations relative to their nearest circulating relatives and that derive from variants that vanished from circulation months or years prior.

S:N417T is one such mutation. S:K417T was in Gamma (P.1) and briefly appeared in a few small circulating BA.2 lineages from March to July 2022. Since then, T417 has been almost exclusively confined to known or presumed chronic infection sequences (Cele et al. 2022, Yépez et al. 2022). Notably, 90% of published cryptic wastewater lineages in the USA, which are strongly suspected to originate in chronic infections, possess T417, though the reasons for this remain obscure (Gregory et al. 2022, Smyth et al. 2022). Since the beginning of 2023, there have been fewer than 150 sequences, out of a total of ∼1.5 million sequences, that have had S:T417.

BA.2.87.1 possesses several other uncommon mutations that are disproportionately common in highly mutated sequences resulting from known/presumed chronic infections. S:V126A, for example, has only been found in 117 sequences between 2022 and 2023, and 10 of those sequences are from lineages that had ceased circulating months or years before they were collected, indicating that they likely originate in chronic infections. One is a B.1.1 from Russia collected in October 2022 that has ∼90 unique mutations (Nabieva et al. 2023). Another 2 of these 10 sequences come from a BA.5.2.1 with ∼45 private mutations. Remarkably, these two BA.5.2.1 sequences possess the exact same NTD deletions as BA.2.87.1, as well as S:Y796H, M:A104V, ORF1a:L3915F—all in BA.2.87.1—and substitutions (with different amino acid residues) at S:P621 and S:D936.

Conclusions

BA.2.87.1 is a highly divergent Omicron-derived SARS-CoV-2 lineage that is likely to have arisen and circulated for 3–4 months in close proximity to the northeastern provinces of South Africa as demonstrated by clinical and wastewater genomic surveillance. The number of sequences from South Africa that have had both ends of the Spike C15–C136 disulfide bond deleted exemplifies the disproportionate share of highly mutated, likely chronic infection–derived sequences in this region. Even though BA.2.87.1 did not ultimately represent a substantial antigenic shift relative to these other lineages, its genomic peculiarities are an important snapshot preview of the types of genomic modifications that can emerge in SARS-CoV-2 during its ongoing evolutionary adaptation to infecting humans and other hosts.

The convergence of two different SARS-CoV-2 surveillance modalities, namely, clinical and wastewater, each utilising genomic and phylogenetic data, facilitated both the detection and corroboration of the presence and prevalence of this variant at a population or community level through triangulation of data. Ongoing genomic surveillance of variants will enable timely updates to vaccine composition. Further, real-time data sharing and discussions between teams working on clinical and wastewater surveillance programmes facilitated interpretation of the results and appreciation of the significance of these findings for understanding SARS-CoV-2 evolution.

Supplementary data

Supplementary data are available at *VEVOLU Journal* online.

Conflict of interest: N.W. and A.v.G. have received grant funding from the Bill and Melinda Gates Foundation, the US Centers for Disease Control and Prevention and Sanofi. Cheryl Cohen has received grants to the institution from the US Centers for Disease Control and Prevention, Bill and Melinda Gates Foundation, Taskforce for Global Health and Sanofi Pasteur unrelated to the work under consideration.

Funding

This work was supported in part by a Fogarty International Centre Global Infectious Disease research training grant from the National Institutes of Health to the University of Pittsburgh and

National Institute for Communicable Diseases (D43TW011255). Sequencing activities for NICD are supported by a conditional grant from the South African National Department of Health as part of the emergency COVID-19 response; a cooperative agreement between the National Institute for Communicable Diseases of the National Health Laboratory Service and the United States Centers for Disease Control and Prevention (FAIN# U01IP001048; NU51IP000930); the South African Medical Research Council (SAMRC, project number 96838); the African Society of Laboratory Medicine (ASLM) and Africa Centers for Disease Control and Prevention through a sub-award from the Bill and Melinda Gates Foundation grant number INV-018978, INV-049272, and INV-050051; the UK Foreign, Commonwealth and Development Office and Wellcome (Grant no 221003/Z/20/Z); This work was partly funded by the SEQAFRICA project which is funded by the UK Department of Health and Social Care's Fleming Fund using UK aid. NICD sequencing was also supported by The Coronavirus Aid, Relief, and Economic Security Act (CARES ACT) through the Centers for Disease Control and Prevention (CDC) and the COVID International Task Force (ITF) funds through the CDC under the terms of a subcontract with the African Field Epidemiology Network (AFENET) AF-NICD-001/2021. Hyrax Biosciences' Exatype platform, used for the assembly of SARS-CoV-2 genomes, was supported by the South African Medical Research Council (SAMRC) with funds received from the Department of Science and Innovation. Additional funds for NGS-SA were also routed through the University of KwaZulu-Natal from the SAMRC with funds received from the South African Department of Science and Innovation. The content and findings reported/illustrated are the sole deduction, view, and responsibility of the researcher and do not reflect the official position and sentiments of the funders.

Data availability

All 10 clinical sequences were deposited in GISAID (EPI_SET ID: EPI_SET_250305mp, doi:http://dx.doi.org/10.55876/gis8.250305mp) and the GISAID accession identifiers are included in Supplementary Table 1 along with accession numbers of reference sequences used for the phylogenetic trees. Raw sequencing data from the wastewater surveillance and clinical samples are available on SRA (BioProjects: PRJNA1279593 and PRJNA1280570, respectively). An interactive version of the phylogenetic tree is available at https://nextstrain.org/groups/NICD-CRDM/ncov/BA.2.87.1.

References

Aksamentov I, Roemer C, Hodcroft E *et al.* Nextclade: Clade assignment, mutation calling and quality control for viral genomes. *J Open Source Softw* 2021;**6**:3773. https://doi.org/10.21105/joss.03773

Bouhaddou M, Reuschl A-K, Polacco BJ *et al.* SARS-CoV-2 variants evolve convergent strategies to remodel the host response. *Cell* 2023;**186**:4597–4614.e26. https://doi.org/10.1016/j.cell.2023.08.026

Cantoni D, Murray MJ, Kalemera MD *et al.* Evolutionary remodelling of N-terminal domain loops fine-tunes SARS-CoV-2 spike. *EMBO Rep* 2022;**23**:e54322. https://doi.org/10.15252/embr.202154322

Cele S, Karim F, Lustig G *et al.* SARS-CoV-2 prolonged infection during advanced HIV disease evolves extensive immune escape. *Cell Host Microbe* 2022;**30**:154–162.e5. https://doi.org/10.1016/j.chom.2022.01.005

Cerutti G, Guo Y, Zhou T *et al.* Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* 2021;**29**:819–833.e7. https://doi.org/10.1016/j.chom.2021.03.005

Chaguza C, Hahn AM, Petrone ME *et al.* Accelerated SARS-CoV-2 intrahost evolution leading to distinct genotypes during chronic infection. *Cell Rep Med* 2023;**4**:100943. https://doi.org/10.1016/j.xcrm.2023.100943

Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007;**7**:1–8. https://doi.org/10.1186/1471-2148-7-214/TABLES/1

Faghy MA, Owen R, Thomas C *et al.* Is long COVID the next global health crisis? *J Glob Health* 2022;**12**:03067. https://doi.org/10.7189/jogh.12.03067

Gregory DA, Trujillo M, Rushford C *et al.* Genetic diversity and evolutionary convergence of cryptic SARS- CoV-2 lineages detected via wastewater sequencing. *PLoS Pathog* 2022;**18**:e1010636. https://doi.org/10.1371/journal.ppat.1010636

Hinrichs A, Ye C, Turakhia Y *et al.* The ongoing evolution of UShER during the SARS-CoV-2 pandemic. *Nat Genet* 2023;**56**:4–7. https://doi.org/10.1038/s41588-023-01622-5

Hisner R, Gueli F, Peacock T. *Repeated Loss of ORF8 Expression in Circulating SARS-CoV-2 Lineages*. 2023. Virological.org. https://Virological.Org/t/Repeated-Loss-of-Orf8-Expression-in-Circulating-Sars-Cov-2-Lineages/931

Iwu-Jaja C, Ndlovu NL, Rachida S *et al.* The role of wastewater-based epidemiology for SARS-CoV-2 in developing countries: Cumulative evidence from South Africa supports sentinel site surveillance to guide public health decision-making. *Sci Total Environ* 2023;**903**:165817. https://doi.org/10.1016/j.scitotenv.2023.165817

Karim SSA, Karim QA. Omicron SARS-CoV-2 variant: A new chapter in the COVID-19 pandemic. *Lancet* 2021;**398**:2126–8. https://doi.org/10.1016/S0140-6736(21)02758-6

Katoh K. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;**30**:3059–66. https://doi.org/10.1093/nar/gkf436

Kemp SA, Collier DA, Datir RP *et al.* SARS-CoV-2 evolution during treatment of chronic infection. *Nature* 2021;**592**:277–82. https://doi.org/10.1038/s41586-021-03291-y

Khan K, Lustig G, Römer C *et al.* Evolution and neutralization escape of the SARS-CoV-2 BA.2.86 subvariant. *Nat Commun* 2023;**14**:1–9. https://doi.org/10.1038/s41467-023-43703-3

Landis JT, Moorad R, Pluta LJ *et al.* Intra-host evolution provides for the continuous emergence of SARS-CoV-2 variants. *MBio* 2023;**14**:e0344822. https://doi.org/10.1128/mbio.03448-22

Larsson A. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 2014;**30**:3276–8. https://doi.org/10.1093/BIOINFORMATICS/BTU531

Li L, Liao H, Meng Y *et al.* Structural basis of human ACE2 higher binding affinity to currently circulating omicron SARS-CoV-2 subvariants BA.2 and BA.1.1. *Cell* 2022;**185**:2952–2960.e10. https://doi.org/10.1016/j.cell.2022.06.023

Lin K, Schulte CR, Marr LC. Survival of MS2 and Φ6 viruses in droplets as a function of relative humidity, pH, and salt, protein, and surfactant concentrations. *PLoS One* 2020;**15**:e0243505. https://doi.org/10.1371/journal.pone.0243505

Markov PV, Ghafari M, Beer M *et al.* The evolution of SARS-CoV-2. *Nat Rev Microbiol* 2023;**21**:361–79. https://doi.org/10.1038/s41579-023-00878-2

Marques AD, Hogenauer M, Bauer N *et al.* Evolution of SARS-CoV-2 in white-tailed deer in Pennsylvania 2021-2024. *PLoS Pathog* 2025;**21**:e1012883. https://doi.org/10.1371/journal.ppat.1012883

Martin DP, Varsani A, Roumagnac P *et al.* RDP5: A computer program for analyzing recombination in, and removing signals of

recombination from, nucleotide sequence datasets. *Virus Evol* 2021;**7**:veaa087. https://doi.org/10.1093/ve/veaa087

McCallum M, Bassi J, De Marco A *et al.* SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. *Science* 2021a;**373**:648–54. https://doi.org/10.1126/science.abi7994

McCallum M, De Marco A, Lempp FA *et al.* N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* 2021b;**184**:2332–2347.e16. https://doi.org/10.1016/j.cell.2021.03.028

McCarthy KR, Rennick LJ, Nambulli S *et al.* Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* 2021;**371**:1139–42. https://doi.org/10.1126/science.abf6950

Msomi N, Mlisana K, de Oliveira T *et al.* A genomics network established to respond rapidly to public health threats in South Africa. *Lancet Microbe* 2020;**1**:e229–30. https://doi.org/10.1016/S2666-5247(20)30116-6

Nabieva E, Komissarov AB, Klink GV *et al.* *A Highly Divergent SARS-CoV-2 Lineage B.1.1 Sample in a Patient with Long-Term COVID-19*. medRxiv 2023.09.14.23295379. https://doi.org/10.1101/2023.09.14.23295379

Naderi S, Chen PE, Murall CL *et al.* Zooanthroponotic transmission of SARS-CoV-2 and host-specific viral mutations revealed by genome-wide phylogenetic analysis. *elife* 2023;**12**:e83685. https://doi.org/10.7554/eLife.83685

National Health Laboratory Service. *Weekly Respiratory Pathogens Surveillance Report*. Johannesburg, South Africa: NICD, 2023. https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/surveillance-reports/weekly-respiratory-pathogens-surveillance-report-week/

National Health Laboratory Service. *Weekly Respiratory Pathogens Surveillance Report*. Johannesburg, South Africa: NICD, 2024. https://www.nicd.ac.za/diseases-a-z-index/disease-index-covid-19/surveillance-reports/weekly-respiratory-pathogens-surveillance-report-week/

O'Toole Á, Scher E, Underwood A *et al.* Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol* 2021;**7**:veab064. https://doi.org/10.1093/ve/veab064

Parker MD, Stewart H, Shehata OM *et al.* Altered subgenomic RNA abundance provides unique insight into SARS-CoV-2 B.1.1.7/alpha variant infections. *Commun Biol* 2022;**5**:666. https://doi.org/10.1038/s42003-022-03565-9

Pollard CA, Morran MP, Nestor-Kalinoski AL. The COVID-19 pandemic: A global health crisis. *Physiol Genomics* 2020;**52**:549. https://doi.org/10.1152/PHYSIOLGENOMICS.00089.2020

Qing E, Kicmal T, Kumar B *et al.* Dynamics of SARS-CoV-2 spike proteins in cell entry: Control elements in the amino-terminal domains. *MBio* 2021;**12**:e0159021. https://doi.org/10.1128/mBio.01590-21

Rambaut A, Lam TT, Carvalho LM *et al.* Exploring the temporal structure of heterochronous sequences using TempEst (formerly path-O-gen). *Virus Evol* 2016;**2**:vew007. https://doi.org/10.1093/VE/VEW007

Rambaut A, Drummond AJ, Xie D *et al.* Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol* 2018;**67**:901–4. https://doi.org/10.1093/sysbio/syy032

Roemer C, Sheward DJ, Hisner R *et al.* SARS-CoV-2 evolution in the omicron era. *Nat Microbiol* 2023;**8**:1952–9. https://doi.org/10.1038/s41564-023-01504-w

Sanderson T, Hisner R, Donovan-Banfield I *et al.* A molnupiravir-associated mutational signature in global SARS-CoV-2 genomes. *Nature* 2023;**623**:594–600. https://doi.org/10.1038/s41586-023-06649-6

Si Y, Wu W, Xue X *et al.* The evolution of SARS-CoV-2 and the COVID-19 pandemic. *PeerJ* 2023;**11**:e15990. https://doi.org/10.7717/peerj.15990

Smyth DS, Trujillo M, Gregory DA *et al.* Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater. *Nat Commun* 2022;**13**:635. https://doi.org/10.1038/s41467-022-28246-3

Sohrabi C, Alsafi Z, O'Neill N *et al.* World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int J Surg (London, England)* 2020;**76**:71. https://doi.org/10.1016/J.IJSU.2020.02.034

Sola I, Moreno JL, Zúñiga S *et al.* Role of nucleotides immediately flanking the transcription-regulating sequence Core in coronavirus subgenomic mRNA synthesis. *J Virol* 2005;**79**:2506–16. https://doi.org/10.1128/jvi.79.4.2506-2516.2005

Sola I, Almazán F, Zúñiga S *et al.* Continuous and discontinuous RNA synthesis in coronaviruses. *Annu Rev Virol* 2015;**2**:265–88. https://doi.org/10.1146/annurev-virology-100114-055218

Tamura T, Ito J, Uriu K *et al.* Virological characteristics of the SARS-CoV-2 XBB variant derived from recombination of two omicron subvariants. *Nat Commun* 2023;**14**:2800. https://doi.org/10.1038/s41467-023-38435-3

Tegally H, Moir M, Everatt J *et al.* Emergence of SARS-CoV-2 omicron lineages BA.4 and BA.5 in South Africa. *Nat Med* 2022;**28**:1785–90. https://doi.org/10.1038/s41591-022-01911-2

Tiecco G, Storti S, Arsuffi S *et al.* Omicron BA.2 lineage, the "stealth" variant: Is it truly a silent epidemic? A literature review. *Int J Mol Sci* 2022;**23**:7315. https://doi.org/10.3390/ijms23137315

V'kovski P, Kratzel A, Steiner S *et al.* Coronavirus biology and replication: Implications for SARS-CoV-2. *Nat Rev Microbiol* 2021;**19**:155–70. https://doi.org/10.1038/s41579-020-00468-6

Viana R, Moyo S, Amoako DG *et al.* Rapid epidemic expansion of the SARS-CoV-2 omicron variant in southern Africa. *Nature* 2022;**603**:679–86. https://doi.org/10.1038/s41586-022-04411-y

Vulturar D-M, Moacă L-Ș, Neag MA *et al.* Delta variant in the COVID-19 pandemic: A comparative study on clinical outcomes based on vaccination status. *J Pers Med* 2024;**14**:358. https://doi.org/10.3390/jpm14040358

Wang L, Møhlenberg M, Wang P *et al.* Immune evasion of neutralizing antibodies by SARS-CoV-2 omicron. *Cytokine Growth Factor Rev* 2023;**70**:13–25. https://doi.org/10.1016/j.cytogfr.2023.03.001

Wang X, Jiang S, Ma W *et al.* Robust neutralization of SARS-CoV-2 variants including JN.1 and BA.2.87.1 by trivalent XBB vaccine-induced antibodies. *Sig Transduct Target Ther* 2024;**9**:123. https://doi.org/10.1038/s41392-024-01849-6

World Health Organization (WHO). *One Year since the Emergence of COVID-19 Virus Variant Omicron*. Geneva, Switzerland. 2022. https://Www.Who.Int/News-Room/Feature-Stories/Detail/One-Year-since-the-Emergence-of-Omicron

World Health Organization (WHO). *Tracking SARS-CoV-2 Variants*. Internet. Geneva, Switzerland, 2024. https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/

Yang S, Yu Y, Jian F *et al.* Antigenicity assessment of SARS-CoV-2 saltation variant BA.2.87.1. Sig Transduct Target Ther 2024;**13**:2343909. https://doi.org/10.1038/s41392-024-01849-6

Yépez Y, Marcano-Ruiz M, Bezerra RS *et al.* Evolutionary history of the SARS-CoV-2 gamma variant of concern (P.1): A perfect storm. *Genet Mol Biol* 2022;**45**. https://doi.org/10.1590/1678-4685-gmb-2021-0309

Yu G, Smith DK, Zhu H *et al.* Ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017;**8**:28–36. https://doi.org/10.1111/2041-210X.12628

Yu X, Juraszek J, Rutten L *et al.* Convergence of immune escape strategies highlights plasticity of SARS-CoV-2 spike. *PLoS Pathog* 2023;**19**:e1011308. https://doi.org/10.1371/journal.ppat.1011308.

Zhu H, Wei L, Niu P. The novel coronavirus outbreak in Wuhan, China. *Glob Health Res Policy* 2020;**5**:6. https://doi.org/10.1186/s41256-020-00135-6

Virus Evolution, 2025, **11**, 1–14
https://doi.org/10.1093/ve/veaf083
Research Article