DOMINANCE OF THE SARS-COV-2 501Y.V2 LINEAGE IN GAUTENG

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

SOUTH AFRICA | 28 JANUARY 2021

Summary

The 501Y.V2 lineage has been recently been shown to predominate in the Eastern Cape, Western Cape and KwaZulu-Natal provinces of South Africa, all of which experienced major outbreaks of COVID-19. This variant of concern harbours mutations associated with increased transmissibility and neutralizing antibody resistance. Here we describe a preliminary analysis of 479 sequences from Gauteng, the country's economic hub, indicating that the 501Y.V2 lineage first appeared in November and by December accounted for 84% (62/74) of sequences. The Eastern Cape, Western Cape, KwaZulu-Natal and Gauteng data suggests that 501Y.V2 lineage may be predominant throughout South Africa.

Background

Coronavirus disease 2019 (COVID-19) is a respiratory disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease was first identified in Wuhan, China in December 2019 and has subsequently caused a global pandemic with over 99.8 million cases and 2.4 million deaths reported as of 26 January 20211. In South Africa, the first case was reported on 5 March 2020 and as of 26 January 2021, 1.42 million cases and 41,117 deaths have been documented2.

Whole genome sequencing (WGS) has been extensively employed in response to this pandemic. As of 26 January 2021, 337,106 genome sequences from around the world were available in the GISAID public database3. Genomic data has been shown to complement epidemiological findings and has demonstrated the ongoing evolution of the virus, which has contributed to guiding public health and infection control strategies related to SARS-CoV-24-8 [2-6].

Coronaviruses have the largest genomes of all RNA viruses with the SARS-CoV-2 genome consisting of ~30, 000 nucleotides, containing 13-15 open reading frames and encoding 12 proteins9. The virus contains 4 structural proteins, namely the spike (S), envelope (E), membrane (M) and nucleocapsid (N)9. The spike protein mediates host cell entry through receptor engagement and membrane fusion and is also the target for neutralizing antibodies9.

Since its emergence, SARS-CoV-2 has diversified into several different clades and lineages based on specific mutational signatures. This includes the Nextstrain10 and Pangolin11 nomenclatures. The Nextstrain nomenclatures provide a more general overview of the diversity of SARS-CoV-2 strains circulating globally, while the Pangolin nomenclature reflects a relatively higher resolution.

The Network for Genomics Surveillance in South Africa (NGS-SA)12, which includes the National Institute for Communicable Diseases (NICD), National Health Laboratory Services (NHLS), University of KwaZulu-Natal (UKZN), University of Cape Town (UCT), Stellenbosch University (SUN) and University of Free State(UFS), has been monitoring changes in the SARS-CoV-2 genome causing infections in South Africa, since March 2020.

Between March and September 2020, analysis of over 1,000 viral sequences generated by the NICD from specimens collected from across 7 of the 9 provinces in South Africa indicated that this virus mutated at a relatively slow rate. However, sequences from samples collected in October 2020 showed the emergence of a new SARS-CoV-2 lineage, or group of mutated viruses, in the Eastern Cape, Western Cape and KwaZulu-Natal Provinces, where it is now the predominant lineage as was recently reported6.

This lineage, named 501Y.V2, possesses 20-40 nucleotide mutations that were not previously seen in viruses from South Africa prior to September 20206. Of note were the accumulation of mutations in the spike gene, which encoded amino acid changes or deletions at L18F, D80A, D215G, Δ242-244, R246I, K417N, E484K and N501Y6.

In this report, we describe preliminary SARS-CoV-2 sequencing results from samples collected in Gauteng between April and December 2020.

Methods

SARS-CoV-2 samples

Clinical specimens that tested positive for SARS-CoV-2 by real-time polymerase chain reaction (PCR) (cycle threshold (Ct)-values less than 30) were obtained from both public and private sector laboratories in Gauteng, South Africa. For private laboratories, samples were randomly selected on a weekly basis from July through October 2020 for further characterisation. All samples collected from the Centre for Respiratory Diseases and Meningitis (CRDM) severe acute respiratory illness (SARI) surveillance programme sites at Helen Joseph and Rahima Moosa Hospitals in Johannesburg, collected between April through December 2020 were processed for sequencing. In addition, we received occasional samples selected by convenience from the public National Health Laboratory Service laboratories in Gauteng from April through December 2020.

Genome sequencing and analysis

RNA was extracted using the QIAamp® Viral RNA Mini kit (QIAGEN, Hilden, Germany) as per manufacturer's instruction with the following modifications; the input volume of sample was doubled (280µL instead of 140µL) and the purified RNA was eluted twice using 40µL elution buffer for each elution. The RNA extract was concentrated by vacuum centrifugation to a final volume to 22µL. cDNA synthesis, tilling PCR and PCR cleanup were performed using a protocol developed by the ARTIC network for SARS-CoV-213,14. Sequencing was performed on the Illumina MiSeq and NextSeq platforms (Illumina, San Diego, CA, USA). Raw sequencing reads were quality trimmed and aligned to SARS-CoV-2 Wuhan-Hu-1 reference strain (GenBank accession number MN908947). Clades and mutations were assigned using Nextclade (https://clades.nextstrain.org/, v0.7.6). Lineages were assigned using the Phylogenetic Assignment of Named Global Outbreak LINeages lineage assigner web tool (https://pangolin.cog-uk.io/).

Results

Demographic information of sequenced SARS-CoV-2 samples

Overall, 479 samples collected between April through December 2020, from patients presenting at private (N=310) and public (N=169) healthcare facilities in Gauteng, South Africa were sequenced. (Figure 1).

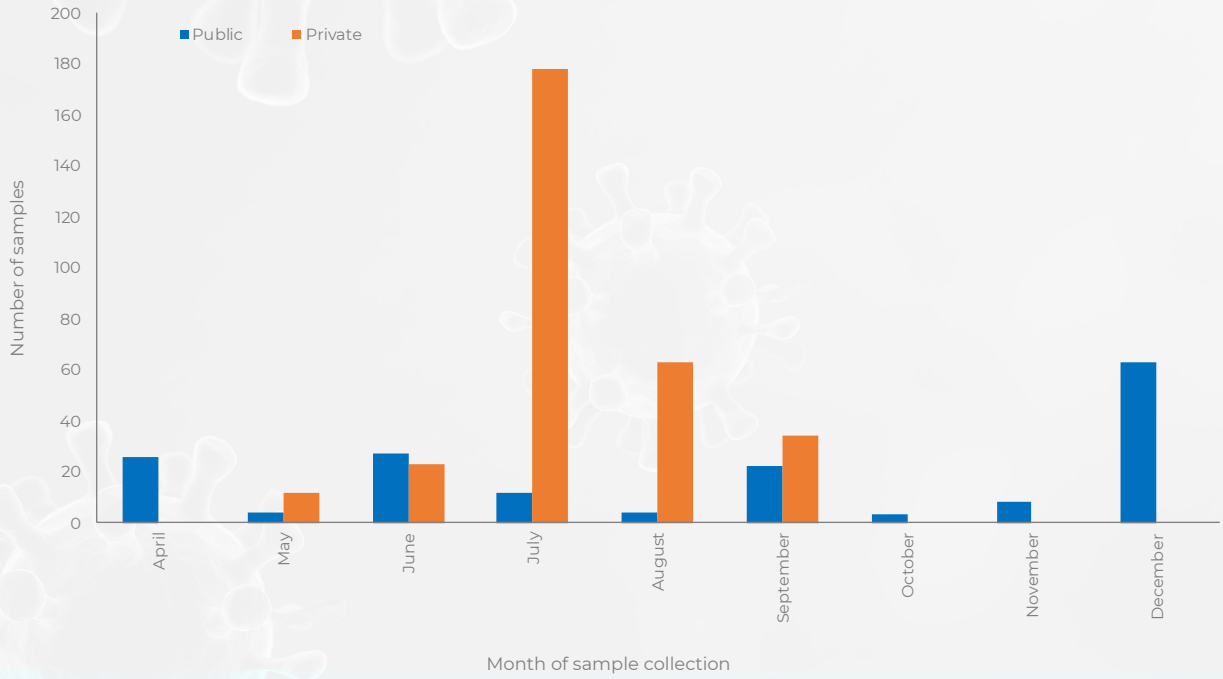

Figure 1. Source (public or private sector) of SARS-CoV-2 samples sequenced by month (April – December 2020), Gauteng, South Africa (n = 479).

Preliminary genomic analysis

Monthly numbers of samples processed for sequencing varied, with results for <10 samples in October and November currently available. In contrast, April through September and December had relatively good sequence representation. Samples were collected from 3 of the 12 districts in Gauteng, namely the City of Ekurhuleni, City of Johannesburg Metropolitan Municipality and Mogale City, which include the two most populated districts in Gauteng. We used the Pangolin classification to define lineages circulating in Gauteng and while 108 sequences were unassigned due to low coverage in lineage-defining regions, the lineage assignments for each of these months differed from each other in this preliminary analysis (Figure 2). Overall the B.1.1 (11%), B.1.1.52 (14%), B.1.1.54 (24%) and C1 (23%) lineages predominated prior to October 2020 in Gauteng. Following emergence of the B.1351 or 501Y.V2 lineage, first identified in this sample in November 2020, we saw an almost complete replacement (84%) of all other lineages in December 2020.

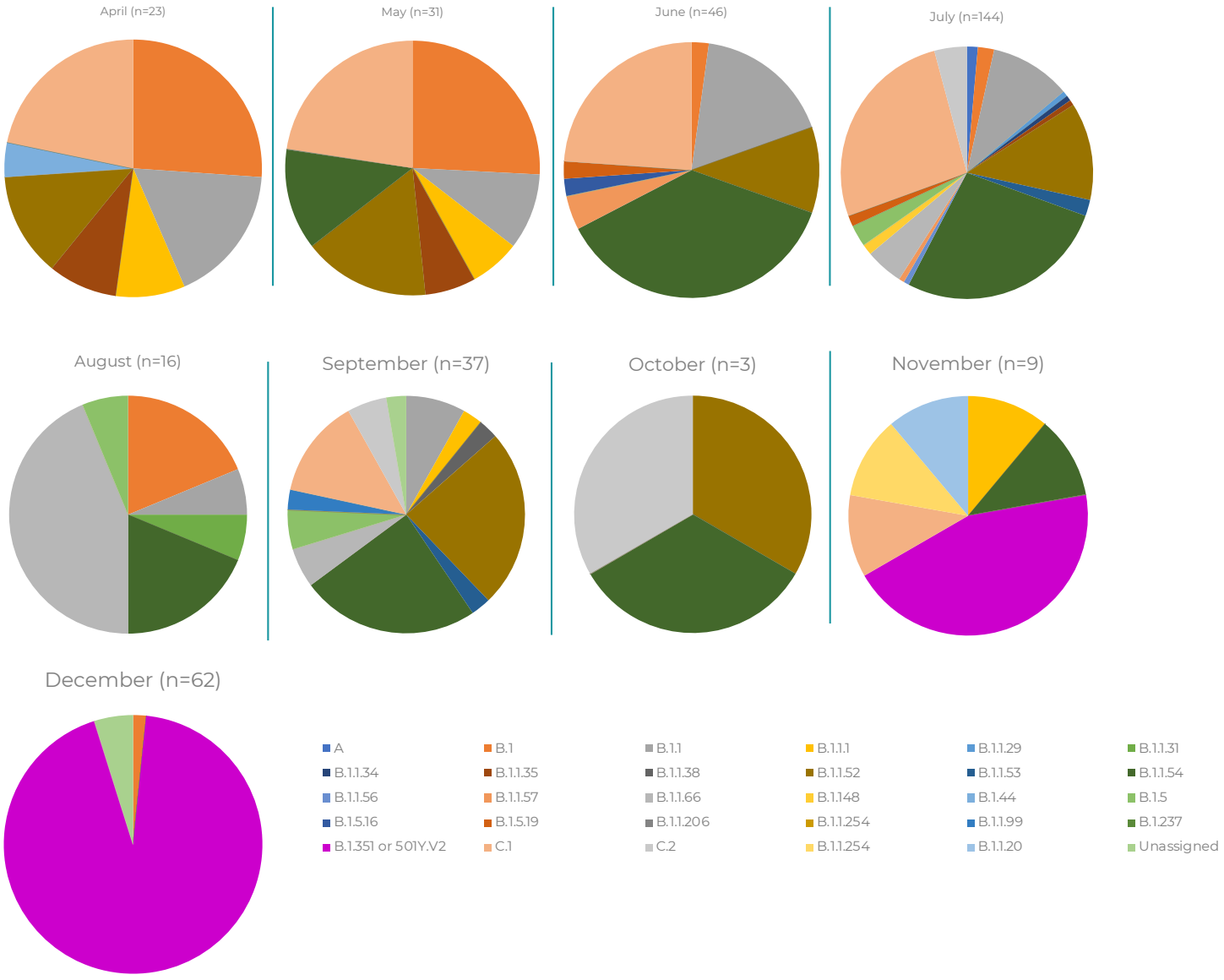

Figure 2. Overall SARS-CoV-2 lineage distribution in Gauteng, South Africa, per month. SARS-CoV-2 lineage distribution depicted as pie charts for 9 months of 2020 from which samples collected between April to December were sequenced (n=371).

Limitations

There are several limitations which must be considered when interpreting these results:

- Sample collection was not evenly distributed across Gauteng or according to month, which may bias the results.
- The sample submission from some laboratories was inconsistent, as many laboratories have been overwhelmed with the sheer number of SARS-CoV-2 diagnostic tests performed on a daily basis.
- Numbers of samples in some months are low, limiting our ability to accurately identify when the 501Y.V2 lineage was introduced into Gauteng, however sequencing of stored samples is ongoing.
- To understanding the emergence of the new lineage in Gauteng, further analyses investigating geographic location, characteristics/demographics of individuals infected, as well as sourcing additional specimens across the time of the local epidemic are efforts that are being prioritized. Updated data will also include specimens from January 2021.

Conclusions

It is critically important to continue monitoring the evolution and spread of this virus, particularly in the spike protein, given the increasing likelihood of immune selection pressure from individuals who may become re-infected with SARS-CoV-2 as well as those who have received a vaccine as part of current vaccine trials or through emergency authorisation use.

Review of preliminary sequences currently available, suggests that following the introduction of the 501Y.V2 lineage in Gauteng in late 2020, this lineage has rapidly spread replacing all previous lineages to become the dominant circulating lineage by December 2020.

Contributors

- Adriano Mendes
- Allison J. Glass
- Amelia Buys
- Andries Dreyer
- Anne von Gottberg
- Arshad Ismail
- Boitshoko Mahlangu
- Cardia Fourie
- Cheryl Cohen
- Christa Viljoen
- Daniel Amoaka
- Elias Bereda
- Florah Mnyameni
- Florette K. Treunicht
- Gloria Selabe
- Howard Newman
- Jackie Kleinhans
- Jeannette Wadula
- Jinal N. Bhiman
- Kathleen Subramoney
- Lia Rotherham
- Linda de Gouveia
- Lynn Morris
- Marianne Wolfaardt
- Marietjie Venter
- Michaela Davis
- Mushal Allam
- Mvuyo Mkhasi
- Nicole Wolter
- Noxolo Ntuli
- Phillip Senzo Mtshali
- Sibongile Walaza
- Simnikiwe Mayaphi
- Stanford Kwenda
- Thabo Mohale
- Warren Lowman
- Zamantungwa TH Khumalo
- Zinhle Makatini

Contact details

Jinal N. Bhiman – jinalb@nicd.ac.za - +27 11 386 6390
Daniel Amoako – danielam@nicd.ac.za
Thabo Mohale – thabom@nicd.ac.za

Acknowledgements

We would like to acknowledge all laboratories within the CRDM surveillance programmes, the NHLS, Lancet and Pathcare laboratories who contributed samples despite the pressures of the pandemic. We would like to acknowledge the SARI surveillance investigator group. We would like to acknowledge the NGS-SA consortium for support, advice and scientific discussion throughout this pandemic.
We encourage additional laboratories who have the capacity to send SARS-CoV-2 positive samples to the NICD for sequencing, so that we may continue to monitor the evolution of this virus with greater coverage for Gauteng and the country.

References

1. WHO Coronavirus Dashboard. https://covid19.who.int/ [Accessed: 26 January 2021].
2. DOH COVID-19 Online Resource & News Portal. https://sacoronavirus.co.za/ [Last Accessed 26 January 2021].
3. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 22(2017).
4. Rockett, R.J., et al. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. Nat. Med. 26, 1398-1404 (2020).
5. Tegally, H., et al. Major new lineages of SARS-CoV-2 emerge and spread in South Africa during lockdown. medRxiv (2020).
6. Tegally, H., et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. medRxiv (2020).
7. Volz, E., et al. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. medRxiv (2021).
8. Voloch, C.M., et al. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. medRxiv (2020).
9. Kim, D., et al. The Architecture of SARS-CoV-2 Transcriptome. Cell 181, 914-921 e910 (2020).
10. Hadfield, J., et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34, 4121-4123 (2018).
11. Rambaut, A., et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 5, 1403-1407 (2020).
12. Msomi, N., Mlisana, K., de Oliveira, T. & Network for Genomic Surveillance in South Africa writing, g. A genomics network established to respond rapidly to public health threats in South Africa. Lancet Microbe 1, e229-e230 (2020).
13. Quick, J. nCoV-2019 sequencing protocol v3 (LoCost) V.3. (Protocols.io, 2020).
14. Giandhari, J.e.a. NEBnext library construction and sequencing for SARS-CoV-2: Adapting COVID-19 ARTIC protocol. (2020).