

What are SARS-CoV-2 genomes from the WHO Africa region member states telling us?

Lu Lu ¹, Samantha Lycett ², Jordan Ashworth ¹, Francisca Mutapi ^{3,4}, Mark Woolhouse ^{1,4}

To cite: Lu L, Lycett S, Ashworth J, *et al*. What are SARS-CoV-2 genomes from the WHO Africa region member states telling us? *BMJ Global Health* 2021;**6**:e004408. doi:10.1136/bmjgh-2020-004408

Handling editor Seye Abimbola

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjgh-2020-004408>).

Received 9 November 2020
Revised 7 December 2020
Accepted 9 December 2020



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Usher Institute, Ashworth Laboratories, Kings Buildings, The University of Edinburgh, Edinburgh, UK

²The University of Edinburgh The Roslin Institute, Roslin, UK

³Institute of Immunology & Infection Research, The University of Edinburgh School of Biological Sciences, Edinburgh, UK

⁴NIHR Global Health Research Unit Tackling Infections to Benefit Africa (TIBA), The University of Edinburgh, Edinburgh, UK

Correspondence to Francisca Mutapi; f.mutapi@ed.ac.uk

INTRODUCTION

Pathogen genome sequencing can inform control measures including diagnosis, identifying infection sources and patterns of infection and disease prognosis. The first case of SARS-CoV-2 was reported on 31 December 2019, and on 10 January 2020 the first sequences of the virus were made publicly available allowing the development and standardisation of the real-time PCR diagnostics for the virus. The reagents for the PCR diagnostic test were designed using the full spectrum of the SARS-CoV-2 reference genome collected on 5 January 2020, in Wuhan.¹ A recent analysis has shown that the majority of current PCR diagnostic targets have undergone mutations with the nucleocapsid (N) gene primers and probes having undergone the most mutations.²

In Africa, analysing virus genome sequences revealed transmission patterns within individual countries. For example, analyses of sequences in Kenya during the early phase of the epidemic identified both imported and local community transmission, demonstrating transmission from Nairobi to the coastal regions.³ Similarly, South Africa was able to identify nosocomial transmission using genome sequencing.⁴

The course of infection and disease prognosis can vary depending on the virulence of the pathogen. In addition, viruses may be able to adapt to the environment and evade recognition by the immune system, so it is crucial to understand the molecular basis for any variation. Several studies have already reported some heterogeneity in the circulating strains of SARS-CoV-2.⁵ However, the important aspect is to relate molecular differences with phenotypic traits that have health implications. For example, a recent study in Singapore comparing the clinical and immunological indicators in patients infected with

Summary box

- To date, 23 countries from the WHO Africa region have deposited a total of 3995 SARS-CoV-2 sequences to publicly available databases, with the majority of the genomes being from South Africa (56%).
- Eight-four different lineages have been identified from these countries with 86% genomes belonging to the B.1 sublineage and its descendants.
- There have been multiple separate introductions of SARS-CoV-2 infections into Africa, with approximately 43% of these coming from Europe.
- 95% of African SARS-CoV-2 genomes have the D614G mutation in spike protein thought to be associated with higher infectivity but lower disease severity.

the wild-type SARS-CoV-2 and those infected with the Δ 382 variant of SARS-CoV-2 which has a deletion in its ORF8 of 382-nucleotide showed that the Δ 382 variant of SARS-CoV-2 was associated with a milder infection.⁶

THE STATE OF GENOME SEQUENCES AVAILABLE PUBLICLY FROM WHO AFRO COUNTRIES

Overall, there has been a much-increased global effort and speed in pathogen sequencing and availing the data in public depositories during this SARS CoV-2 pandemic. Average lag between sample collection and sequence availability has been estimated to be 3 months. We have calculated that for other previous human infective viruses, there was an average of over two and a half years, between sample collection and public sequence availability in 2015. In 2019, this had reduced to an average of about 1 year.

During the SARS-CoV-2 pandemic, several African countries have published in-country sequences faster than the average 3 months. For example, 3 days after the confirmation of

Nigeria's first COVID-19 case, the genome sequencing results of the SARS-CoV-2 specimen from the country were announced on 1 March 2020.⁷ However, to date, there have been no published analyses of SARS-CoV-2 genomes at a continent level in Africa to summarise the patterns on the continent. We have therefore conducted an analysis of all the publicly available sequences from the WHO Africa Region member countries to summarise the currently available information from SARS-CoV-2 genomes.

Data comprising all SARS-CoV-2 genome sequences and the corresponding metadata from WHO Africa region (AFRO) on the GISAID sequence data repository were downloaded on 30 November 2020 (<https://www.gisaid.org>). Spatiotemporal phylogenetic analyses were conducted using the complete genomes available at the time, from African countries and other continents. Over representative sequences were subsampled. The number of introductions to WHO AFRO member countries were estimated by inferring the ancestral states on the phylogenetic tree using the likelihood reconstruction method.

As of 30 November 2020, 3995 SARS-CoV-2 genomes had been submitted to GISAID from 23 countries in the WHO Africa Region (out of 47). These represented ~2% of publicly available sequences globally. The majority of genomes were from South Africa (56%), Democratic Republic of the Congo (9%) and Gambia (9%) (figure 1A,B).

SARS COV-2 IN WHO AFRICA COUNTRIES

Eighty-four lineages were identified in the sequences from WHO AFRO countries. Currently, 266 lineages (including A, B, C and D lineages and all their descendent sublineages) that have so far contributed to active spread globally have been identified by the PANGOLIN phylogenetic classification.⁸ Knowledge of circulating lineages is important for multiple reasons, including determining potential efficacy of vaccines, as has been demonstrated by the challenges for vaccine development for other viruses such as HIV-1, influenza or Dengue driven by viral diversity. A study analysing 18514 SARS-CoV-2 sequences sampled since December 2019 concluded that although there are some reported differences in SARS-CoV-2 genomes, viral diversity is not sufficient to reduce the efficacy of vaccines currently under development.⁹

As of 30 November, the most prevalent lineage globally was B.1 (figure 1C, D). Our analyses show that 86% of all African SARS-CoV-2 genomes belong to this B.1 lineage which represents a large European lineage that corresponds to the Italian outbreak in February, and its descendant sublineages. In South Africa, B.1.106 was identified in April 2020 as the dominant sublineage, but reports now indicate that this lineage became extinct after the effective control of nosocomial outbreaks.¹⁰ These include the novel C.1 lineage, which has 16 mutations compared with the original Wuhan sequence. The C.1 lineage has been reported as the most geographically

widespread lineage in South Africa.¹⁰ The implications of these changes, if any, in terms of transmission and disease outcome are still to be determined.

Our analyses indicate that there have been multiple separate introductions of SARS-CoV-2 into Africa. We estimated that based on currently available sequences, there have been at least 211 separate introductions of the virus into WHO AFRO countries from other continents and African countries outside the WHO AFRO region (figure 1D). Of these, 43% were introduced from Europe, supporting the epidemiological case tracing findings. Most of these introductions do not appear to have spread between countries in Africa. However, 80 introductions from within Africa were tentatively identified, but the small number of sequences currently available from the African countries means that directionality of these introductions cannot as yet, be determined.

WHAT STILL NEEDS TO BE DONE

There is still need for sequencing of the virus in other African countries and for the sequences to be deposited in publicly available data bases. Our analyses indicate that we are still missing sequences from 24 WHO AFRO countries and, furthermore, some of the countries with sequence data have submitted a total of less than 10 sequences to date. The launch of a network of laboratories in September 2020 to reinforce genome sequencing of SARS-CoV-2 in Africa by the WHO World and the Africa Centres for Disease Control and Prevention (Africa CDC) is a welcome development.¹¹ This strengthens the capacity of the member states already provided by other partners on the continent including the African Center of Excellence for Genomics of Infectious Diseases (ACEGID), Nigerian Center for Disease Control (Nigeria CDC) and Tackling Infections to Benefit Africa (TIBA).¹²

There is also need to ensure that the sequencing is science-led or purpose-led so that limited resources are not wasted on sequencing for the sake of sequencing. Finally, there is need to ensure that supporting data are collected and well curated so that patient meta-data including clinical, immunological and prognosis can easily be related to the sequence data. This will ensure that the impact of molecular changes on phenotypic attributes can be more readily identified. For example, when SARS-CoV-2 emerged in Wuhan, it had a D residue (aspartate) at position 614 in the sequence of the virus's spike protein. By June, the D residue had been replaced by G (glycine) on most continents. This mutation is thought to be associated with higher infectivity but lower disease severity.^{6,13} Our analyses have shown that 94% of African SARS-CoV-2 genomes have the D614G mutation in spike protein.

CONCLUSION

For a global pandemic, with potential global solutions, there is need for comprehensive levels of molecular information from SARS-CoV-2 viruses circulating throughout

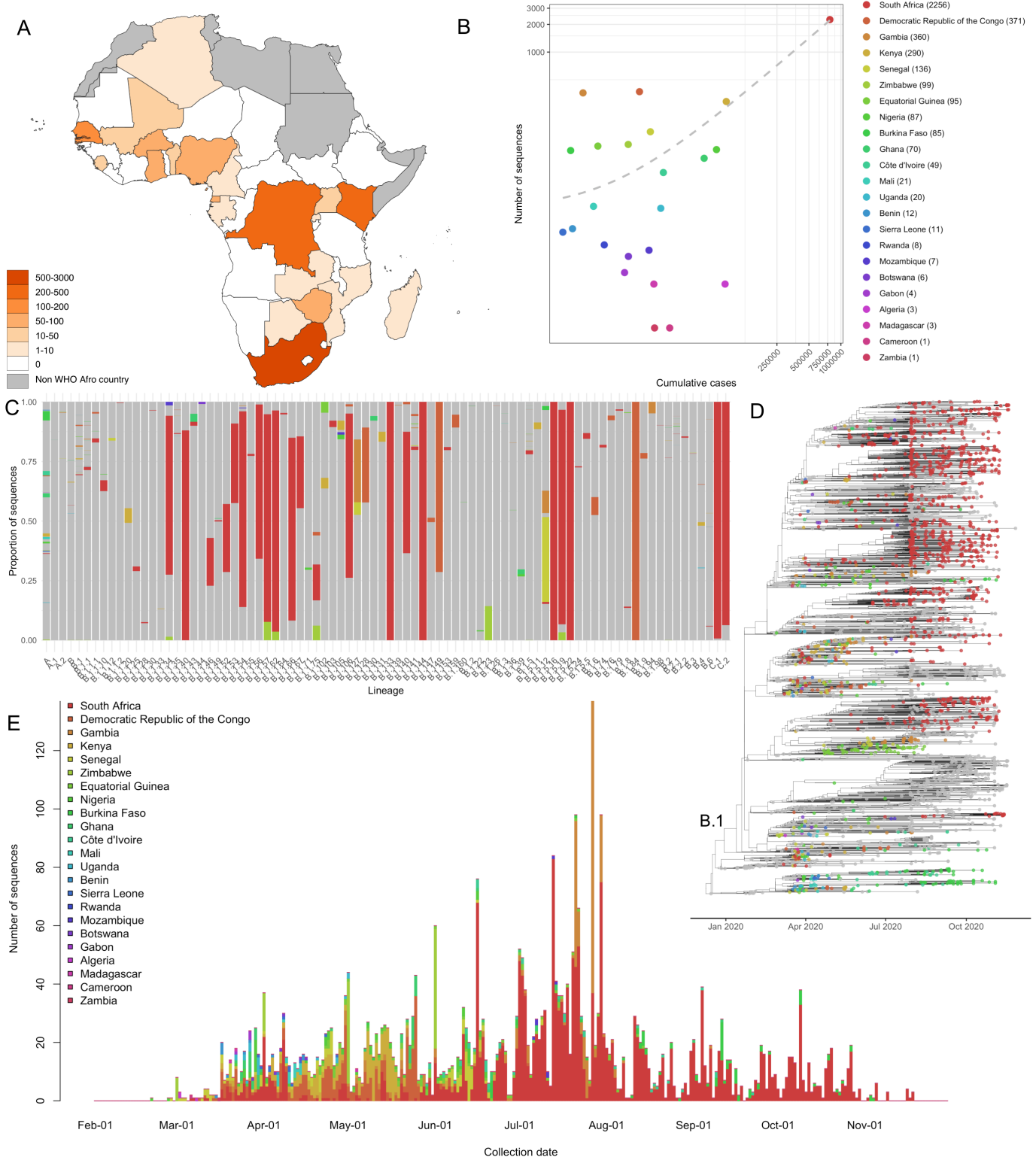


Figure 1 Summary of sequence data for WHO AFRO countries. (A) Heat map of SARS-CoV-2 genomes available on GISAID by African country. (B) Comparison of cumulative reported cases and number of sequences available per country, with the variables on the axes, number of sequences and cumulative cases log transformed. Different colours represent different countries. (C) Global lineages in African countries (colours present different African countries as in B) compared with other regions in the world (in grey). (D) Genomes from individual countries highlighted on approximate time-scaled global subsampled tree ($n=3266$), with tips coloured by the sampling locations as per country-colour key in (C). Lineage B.1 is indicated on the node. (E) SARS-CoV-2 genomes available on GISAID by collection date. Colours for WHO AFRO countries are consistent in (B) to (E), other world regions are in grey.

all the continents. It is hoped that the low number of sequences submitted from Africa (figure 1E) will be accelerated to generate the much-needed sequence information from the continent.

Twitter Jordan Ashworth @_jordanash and Francisca Mutapi @PIG_Edinburgh

Acknowledgements We gratefully acknowledge all SARS-CoV-2 genomes authors, originating and laboratories in Africa submitting data to the GISAID's EpiCov Database on which this study was based, the complete list is in the online supplemental table 1). All submitters of data may be contacted directly via the GISAID website (www.gisaid.org).

Contributors LL, SL and JA conducted the data analysis and were involved in the manuscript preparation. FM collated the information and lead the manuscript preparation. MW conceived and supervised the work and was involved in the manuscript preparation.

Funding This work was commissioned by the National Institute for Health Research (NIHR) Global Health Research programme (16/136/33) using UK aid from the UK Government. The study also received funding from the Scottish Funding Council GCRF COVID-19 grant at the University of Edinburgh.

Disclaimer The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Map disclaimer The depiction of boundaries on this map does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. This map is provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is

properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Lu Lu <https://orcid.org/0000-0002-9330-7022>

Samantha Lycett <http://orcid.org/0000-0003-3159-596X>

Jordan Ashworth <http://orcid.org/0000-0003-1740-7475>

Francisca Mutapi <http://orcid.org/0000-0002-2537-3466>

Mark Woolhouse <https://orcid.org/0000-0003-3765-8167>

REFERENCES

- 1 Wu F, Zhao S, Yu B, *et al*. A new coronavirus associated with human respiratory disease in China. *Nature* 2020;579:265–9.
- 2 Wang R, Hozumi Y, Yin C, *et al*. Mutations on COVID-19 diagnostic targets. *Genomics* 2020;112:5204–13.
- 3 Githinji G. Introduction and local transmission of SARS-CoV-2 cases in Kenya, 2020. Available: <https://virological.org/t/introduction-and-local-transmission-of-sars-cov-2-cases-in-kenya/497>
- 4 Giandhari J, Pillay S, Wilkinson E. Early transmission of SARS-CoV-2 in South Africa: an epidemiological and phylogenetic report. *medRxiv* 2020. doi:10.1101/2020.05.29.20116376
- 5 Islam MR, Hoque MN, Rahman MS, *et al*. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Sci Rep* 2020;10:14004.
- 6 Young BE, Fong S-W, Chan Y-H, *et al*. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet* 2020;396:603–11.
- 7 Makoni M. Africa contributes SARS-CoV-2 sequencing to COVID-19 tracking, 2020. Available: <https://www.the-scientist.com/news-opinion/africa-contributes-sars-cov-2-sequencing-to-covid-19-tracking-67348>
- 8 Rambaut A, Holmes EC, O'Toole Áine, *et al*. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 2020;5:1403–7.
- 9 Dearlove B, Lewitus E, Bai H, *et al*. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. *Proc Natl Acad Sci U S A* 2020;117:23652–62.
- 10 Tegally H, Wilkinson E, Lessells RR. Major new lineages of SARS-CoV-2 emerge and spread in South Africa during lockdown. *medRxiv* 2020. doi:10.1101/2020.10.28.20221143
- 11 Organization WH. COVID-19 genome sequencing laboratory network launches in Africa, 2020. Available: <https://www.afro.who.int/news/covid-19-genome-sequencing-laboratory-network-launches-africa>
- 12 Tessema SK, Inzaule SC, Christoffels A, *et al*. Accelerating genomics-based surveillance for COVID-19 response in Africa. *Lancet Microbe* 2020;1:e227–8.
- 13 Zhang L, Jackson CB, Mou H, *et al*. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv* 2020. doi:10.1101/2020.06.12.148726. [Epub ahead of print: 12 Jun 2020].