SARS-CoV-2 Variants, South Sudan, January–March 2021

Daniel Lule Bugembe, My V.T. Phan, Abe G. Abias, James Ayei, Lul Lojok Deng, Richard Lino Loro Lako, John Rumunu, Pontiano Kaleebu, Joseph Francis Wamala, Juma John HM, Dennis Kenyi Lodiongo, Sudhir Bunga, Matthew Cotten

As the coronavirus pandemic continues, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) sequence data are required to inform vaccine efforts. We provide SARS-CoV-2 sequence data from South Sudan and document the dominance of SARS-CoV-2 lineage B.1.525 (Eta variant) during the country's second wave of infection.

As of August 2021, coronavirus disease (COVID-19) had caused >199 million cases and >4.2 million deaths worldwide (*1*). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes COVID-19, is being sequenced to document virus evolution and to inform vaccine efforts. In South Sudan, the COVID-19 index case was confirmed on April 4, 2020 (*2*); it was followed by 2 infection waves, in May–July 2020 and in February–March 2021 (Appendix Figure 1, https://wwwnc.cdc.gov/EID/article/27/12/21-1488-App1.pdf). As of August 3, 2021, a total of 11,063 cases and 119 deaths (*1*) had been reported in South Sudan. An earlier study from South Sudan reported that, after the second wave, 28% of the population showed serologic evidence of infection (*3*).

Chronic underdevelopment caused by prolonged conflicts has left South Sudan with a weak health system and population displacement. Large populations live in camps that may promote rapid spread and amplification of SARS-CoV-2, and poor socioeconomic conditions limit community-based COVID-19 prevention efforts. Monitoring the circulating viral genomic lineages in South Sudan is crucial, especially as vaccination is implemented and novel virus variants appear globally.

Author affiliations: Medical Research Council/Uganda Virus Research Institute and London School of Hygiene & Tropical Medicine Uganda Research Unit, Entebbe, Uganda (D.L. Bugembe, M.V.T. Phan, P. Kaleebu, M. Cotten); National Public Health Laboratory—Ministry of Health, Juba, South Sudan (A.G. Abias, J. Ayei, L.L. Deng, R.L.L. Lako, J. Rumunu); World Health Organization, Juba (J.F. Wamala, J.J. HM); US Centers for Disease Control and Prevention, Juba, South Sudan (D.K. Lodiongo, S. Bunga); University of Glasgow Centre for Virus Research, Glasgow, Scotland, UK (M. Cotten)

DOI: https://doi.org/10.3201/eid2712.211488

The Study

As part of South Sudan COVID-19 surveillance, samples were collected from community surveillance, point-of-entry screening, and sentinel site surveillance and tested for SARS-CoV-2 by real-time reverse transcription PCR (RT-PCR) at the National Public Health Laboratory (Juba, South Sudan) (*4*). During the second COVID-19 wave in February–March 2021 (Appendix Figure 1), we tested 56,014 samples for SARS-CoV-2; 6,645 samples tested positive (12% positivity). We selected a set of 70 (1%) of these positive samples for genomic sequencing with these inclusion criteria: diagnostic RT-PCR cycle threshold (C_t) values <31, from multiple locations (Figure 1), from new arrivals, from death cases, and from sites showing community transmission. We extracted nucleic acid from swab material and generated SARS-CoV-2 genome as previously described (*5*). A total of 45 complete genomes generated from samples collected in January–March 2021 showed a prevalence of 2 lineages: B.1.525 (Eta) and A.23.1 (Figure 1). The A.23.1 lineage, which was observed in October 2020 in Uganda (*6*) and has now spread globally to 26 countries, was one we observed in Juba and Nimule in early January 2021. The A.23.1 case-patients in Nimule, a South Sudan town on the border with Uganda (Figure 1), were travelers returning from Uganda. In Juba, the earliest-reported A23.1 case was in a traveler returning to South Sudan from Uganda. We detected A23.1 for only a short period; from the end of January to the end of March, we detected only B.1.525 genomes (Appendix Table). The B.1.525 lineage, reported earliest in

the United Kingdom and Nigeria, has spread to 44 countries and is considered a variant of interest (*7*).

Phylogenetic analyses of the South Sudan genomes combined with the available global A.23.1 or B.1.525 genomes were performed to gain insight into the virus movement. The maximum-likelihood trees of both A.23.1 and B.1.525 genome sequences (Figure 2) suggested multiple importations of the strains into South Sudan; South Sudan strains belonged to several sublineages, rather than a single sublineage.

Both A.23.1 and B.1.525 lineages encoded changes in their spike protein (Appendix Figure 2) as well as other parts of the genome and substitutions or deletions in the nonstructural protein 6, open reading frame 3a and 8, and nucleocapsid genes (data not shown), which might be associated with higher transmission or immune evasion. Especially relevant, the A.23.1 genomes encoded spike P681R, which is adjacent to the small (S) 1/S2 furin cleavage site and is also present in the variant of concern B.1.617.2 (Delta) lineage, which is spreading in India and globally and may increase S1/S2 cleavage (*10,11*; B. Lubinski et al., unpub. data. http://biorxiv.org/lookup/doi/10.1101/2021.06.30.450632). A related P681H substitution is present in variants of concern B.1.1.7 (Alpha) and P.1 (Beta). The South Sudan B.1.525 genomes encoded a deletion in the N-terminal domain (NTD) at spike positions 69 and 79, which is also present in B.1.1.7 and many other global variants, and a deletion in the spike NTD in positions 141–146, which may help in evasion of host immune responses. The spike D614G substitution may alter the spike protein conformation; the Q677H substitution is near the furin cleavage site and may alter spike processing.

Conclusions

We describe the patterns of SARS-CoV-2 virus genomics in South Sudan in the second wave of infections during February–March 2021, showing circulation of B.1.525 (Eta) as well as the variant A.23.1. South Sudan faced high transmission of SARS-CoV-2 during this reporting period; our data suggest that the B.1.525 lineage spread widely and progressively increased in frequency in the country during the period. Data from Uganda and Rwanda retrieved from GISAID (https://www.gisaid.org) also showed the appearance of B.1.525 at this time.

A limitation of our study is that sample numbers are low and were limited by the challenges of procurement, shipment, and testing in a harsh and

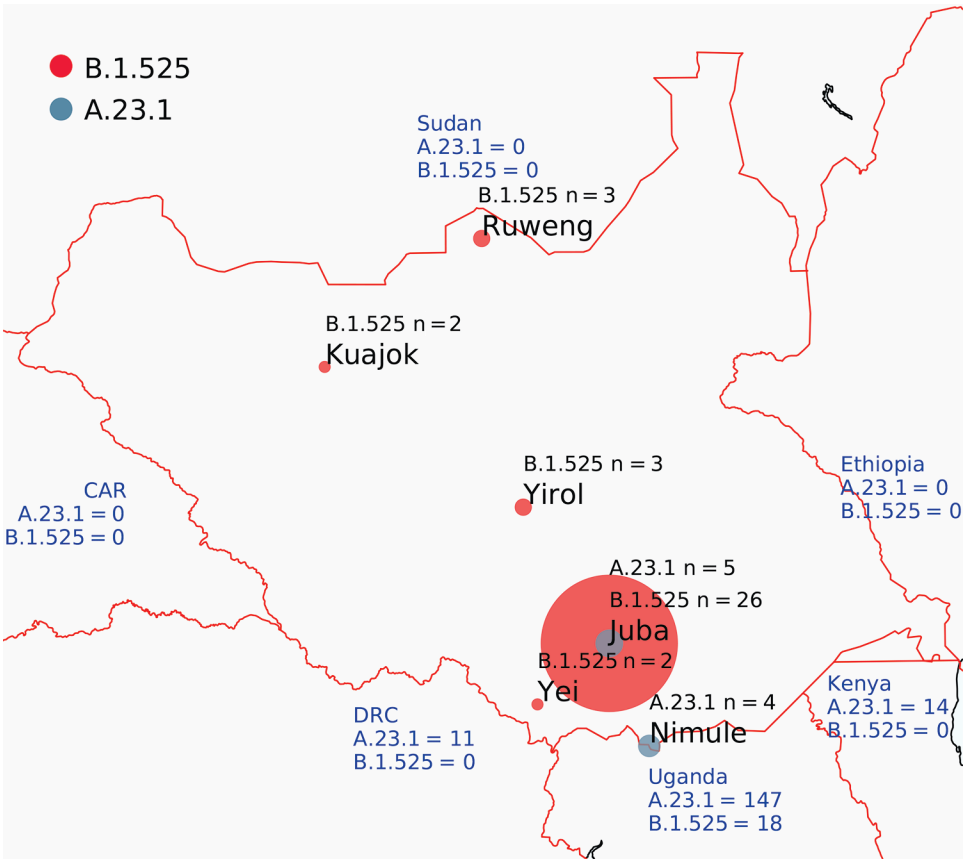


Figure 1. Locations of severe acute respiratory syndrome coronavirus 2 infection case-patients from whom genomes were isolated, South Sudan. Red circles indicate viruses of lineage B.1.525; dark gray circles indicate lineage A.23.1. Circle size is proportional to number of genomes. Blue text shows the number of A.23.1 and B.1.525 genomes reported from neighboring countries. CAR, Central African Republic; DRC, Democratic Republic of the Congo.

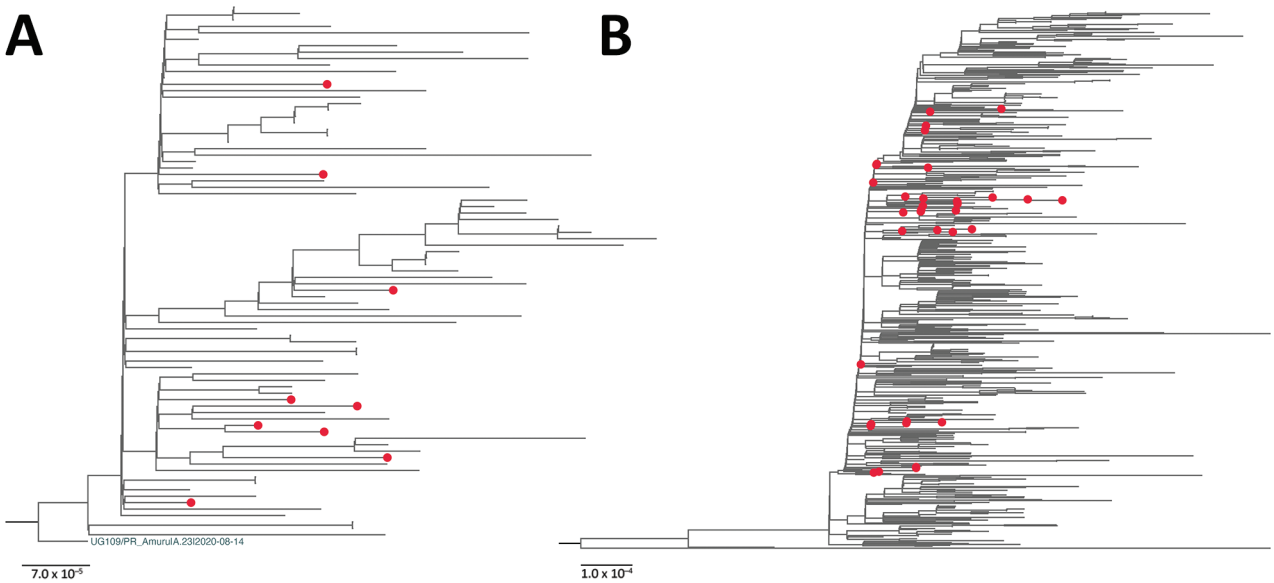


Figure 2. Maximum-likelihood phylogenetic tree of severe acute respiratory syndrome coronavirus 2 viruses from South Sudan (red dots) and reference sequences. A) Lineage A.23.1. All sequences from South Sudan were combined with a subset of all available global A.23.1 genomes, algorithmically thinned. All available global A.23.1 genomes were retrieved from GISAID (https://www.gisaid.org) and aligned, and for the first genome, all genomes closer than 5 hamming distance were removed. This process was continued until the entire set was thinned. This global, thinned A.23.1 set was combined with all South Sudan A.23.1 genomes and used to infer the A.23.1 maximum-likelihood tree. The tree was rooted with the A.23 strain (UG109/PR_Amuru|A.23|2020–08–14). B) Lineage B.1.525. The B.1.525 genome sequences were prepared in the same manner as those for A.23.1 except the hamming distance of 20. Maximum-likelihood phylogenetic trees were constructed in RaxML-NG (*8*) under the general time reversible plus gamma 4 plus invariate sites model as the best-fit model of substitution according to the Akaike information criterion determined by modeltestNG (*9*) and run for 100 pseudoreplicates and visualized using FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree). For B.1.525, the tree was midpoint rooted for clarity. Scale bar indicates nucleotide substitutions per site.

resource-poor environment. Careful sample selection was performed to provide an unbiased description of the epidemic; however, not all positive samples yielded genome sequences. This lack of data could introduce bias in the reported genomes. Nonetheless, the study accurately describes SARS-CoV-2 lineages during the second wave of epidemic in South Sudan.

Substantial land-based traffic with neighboring countries makes it imperative to document the viruses circulating in this region. Careful monitoring of locally circulating viruses as vaccination becomes widespread is essential for interpreting vaccine function and for informing the healthcare systems whether the current vaccines are still a good match for the circulating viruses. We recommend continued genomic surveillance in South Sudan to help with public health responses, especially as new waves of infections come to the country and continent.

D.L.B., M.V.T.P., and M.C. were funded by the Medical Research Council/United Kingdom Research and Innovation (grant no. NC_PC_19060) and Wellcome Trust (grant no. 220977/Z/20/Z).

About the Author

Dr. Bugembe is a scientist at the Medical Research Council/Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine Research Unit in Entebbe, Uganda. His research interests are virus evolution and the computational analysis of viral genetic data to predict infectious disease trends.

References

1. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. COVID-19 dashboard. 2020 [cited 2021 Jan 24]. https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6
2. World Health Organization. South Sudan confirms first case of COVID-19. 2020 [cited 2021 Sep 3]. https://www.afro.who.int/news/south-sudan-confirms-first-case-covid-19
3. Wiens KE, Mawien PN, Rumunu J, Slater D, Jones FK, Moheed S, et al. Seroprevalence of severe acute respiratory syndrome coronavirus 2 IgG in Juba, South Sudan, 2020. Emerg Infect Dis. 2021;27:1598–1606. https://doi.org/10.3201/eid2706.210568
4. World Health Organization. South Sudan joint external evaluation report 2018. 2018 [cited 2021 Sep 3]. https://www.afro.who.int/publications/south-sudan-joint-external-evaluation-report2018

5. Cotten M, Bugembe DL, Kaleebu P, Phan, MVT. Alternate primers for whole-genome SARS-CoV-2 sequencing. Virus Evol. 2021; 27:veab006. https://doi.org/10.1093/ve/veab006
6. Bugembe DL, Phan MVT, Ssewanyana I, Semanda P, Nansumba H, Dhaala B, et al. Emergence and spread of a SARS-CoV-2 lineage A variant (A.23.1) with altered spike protein in Uganda. Nat Micronbiol. 2021; 6:1094–1101. https://doi.org/10.1038/s41564-021-00933-9
7. World Health Organization. Tracking SARS-CoV-2 variants. 2021 [cited 2021 Sep 3]. https://www.who.int/en/activities/tracking-SARS-CoV-2-variants
8. Kozlov, AM, Darriba D, Flouri, T, Morel, B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum-likelihood phylogenetic inference. Bioinformatics. 2019;35:4453–55. https://doi.org/10.1093/bioinformatics/btz305
9. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Molecular Biology and Evolution. 2020; 37:291–294. https://doi.org/10.1093/molbev/msz189
10. Johnson BA, Xie X, Bailey AL, Kalveram B, Lokugamage KG, Muruato A, et al. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. Nature. 2021;591:293–9 https://doi.org/10.1038/s41586-021-03237-4
11. Hoffmann M, Kleine-Weber H, Pöhlmann S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. https://doi.org/10.1016/j.molcel.2020.04.022

Address for correspondence: Matthew Cotten, MRC/UVRI & LSHTM Uganda Research Unit, PO Box 49, Plot 51-59 Nakiwogo Rd, Entebbe, Uganda; email: matthew.cotten@lshtm.ac.uk

EID journal

Want to stay updated on the latest news in *Emerging Infectious Diseases*? Let us connect you to the world of global health. Discover groundbreaking research studies, pictures, podcasts, and more by following us on Twitter at @CDC_EIDjournal.

Article DOI: https://doi.org/10.3201/eid2712.211488

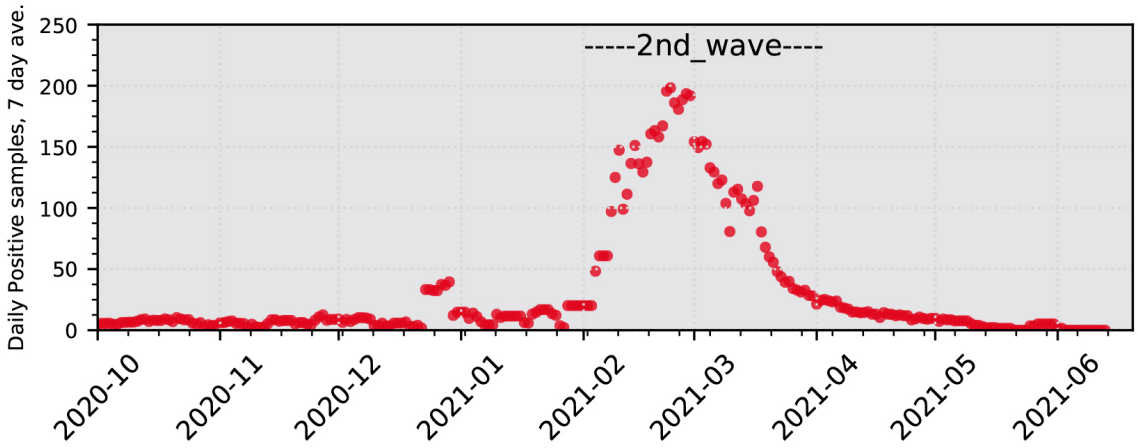
SARS-CoV-2 Variants, South Sudan, January–March 2021

Appendix

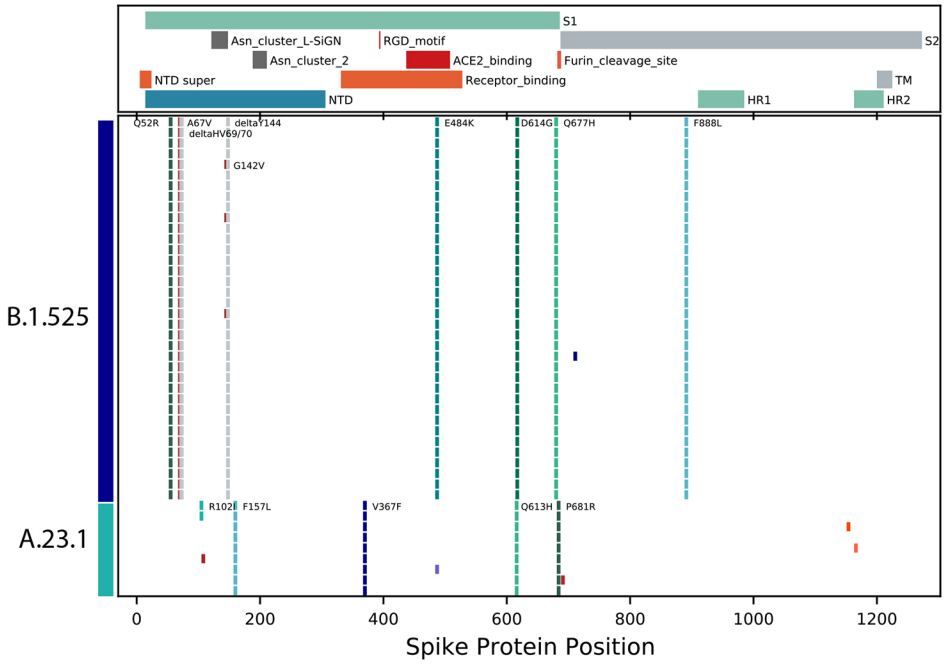
Appendix Table. Summary of samples and resulting genomes from study of SARS-CoV-2 variants, South Sudan, January–March 2021.

Genome	Lineage	C_t	Category	Patient outcome	Local address	Sample date
1	A.23.1	26.3	Screening_4Travel	Survived	Nimule	2021 Jan 8
2	A.23.1	24.6	Screening_4Travel	Survived	Nimule	2021 Jan 8
3	A.23.1	23.9	Screening_4Travel	Survived	Nimule	2021 Jan 13
4	A.23.1	NA	NA	NA	Juba	NA
5	A.23.1	26.34	Screening_4Travel	Survived	Nimule	2021 Jan 18
6	A.23.1	17.3	Screening_4Travel	Survived	Juba	2021 Jan 20
7	A.23.1	28	Screening_4Travel	Survived	Juba	2021 Jan 25
8	A.23.1	28.67	Screening_4Travel	Survived	Juba	2021 Jan 26
9	A.23.1	26.3	Screening_4Travel	Survived	Juba	2021 Jan 27
10	B.1.525	23.2	asymptomatic follow up	Survived	Juba	2021 Jan 23
11	B.1.525	23	Screening_4Travel	Survived	Juba	2021 Jan 23
12	B.1.525	22.83	Screening_4Travel	Survived	Juba	2021 Jan 26
13	B.1.525	24.56	Alert/susp_WithSymp	Survived	Yirol	2021 Jan 27
14	B.1.525	27.3	Screening_4Travel	Survived	Juba	2021 Jan 28
15	B.1.525	27.7	Screening_4Travel	Survived	Juba	2021 Jan 28
16	B.1.525	26.1	Screening_4Travel	Survived	Juba	2021 Jan 28
17	B.1.525	21.48	Alert/susp_WithSymp	Survived	Yirol	2021 Jan 28
18	B.1.525	16.3	Alert/susp_WithSymp	Survived	Juba	2021 Feb 5
19	B.1.525	17	Alert/susp_WithSymp	Survived	Juba	2021 Feb 6
20	B.1.525	18.49	Alert/susp_WithSymp	Survived	Yirol	2021 Feb 9
21	B.1.525	23.01	Screening_4Travel	Survived	Kuajok	2021 Mar 2
22	B.1.525	27.56	Screening_4Travel	Survived	Kuajok	2021 Mar 2
23	B.1.525	24.25	Screening_4Travel	Survived	Juba	2021 Mar 2
24	B.1.525	29.19	Alert/susp_WithSymp	Survived	Juba	2021 Mar 3
25	B.1.525	27.48	Screening_4Travel	Survived	Juba	2021 Mar 3
26	B.1.525	20.27	Cont_Without sympt	Survived	Juba	2021 Mar 4
27	B.1.525	24.34	Screening_4Travel	Survived	Juba	2021 Mar 4
28	B.1.525	23.8	Alert/susp_WithSymp	Survived	Juba	2021 Mar 4
29	B.1.525	22.02	Alert/susp_WithSymp	Survived	Juba	2021 Mar 5
30	B.1.525	22.49	Alert/susp_WithSymp	Survived	Ruweng	2021 Mar 5
31	B.1.525	19.33	Screening_4Travel	Survived	Gudele	2021 Mar 7
32	B.1.525	21.17	Screening_4Travel	Survived	Juba	2021 Mar 8
33	B.1.525	27.44	Screening_4Travel	Survived	Juba	2021 Mar 8
34	B.1.525	26.71	Screening_4Travel	Survived	Juba	2021 Mar 8
35	B.1.525	21.52	Screening_4Travel	Survived	Atlabara	2021 Mar 9
36	B.1.525	22.6	Alert/susp_WithSymp	Survived	Ruweng	2021 Mar 9
37	B.1.525	24.99	Screening_4Travel	Survived	Juba	2021 Mar 10
38	B.1.525	30.59	Alert/susp_WithSymp	Survived	Yei	2021 Mar 11
39	B.1.525	22.41	Alert/susp_WithSymp	Survived	Ruweng	2021 Mar 11
40	B.1.525	25.77	Postmortem_withsymp	Died	Yei	2021 Mar 12
41	B.1.525	25.76	Screening_4Travel	Survived	Juba	2021 Mar 15
42	B.1.525	15.81	Screening_4Travel	Survived	Juba	2021 Mar 15
43	B.1.525	20.98	Cont_Without sympt	Survived	Juba	2021 Mar 18
44	B.1.525	26.9	Alert/susp_WithSymp	Survived	Juba	2021 Mar 18
45	B.1.525	23.82	Alert/susp_WithSymp	Survived	Juba	2021 Mar 20

*Complete genomes were lineage typed using the Pango tool (*1*). NA, data not available.



Appendix Figure 1. Reported severe acute respiratory syndrome coronavirus 2 infections in South Sudan, shown as the 7-day average of daily reported SARS-CoV-2 infections (*2*).



Appendix Figure 2. Changes in the spike protein compared to the NC_045512 reference strain of severe acute respiratory syndrome coronavirus. Each line represents the encoded spike protein sequence from a single genome; sequences are ordered by lineage and then by sample collection date. Markers indicate the positions of amino acid (aa) differences from the reference strain; only changes observed in multiple genomes were annotated with the annotation (original aa position new aa) located either to the left or right of the most recent instance of the substitution. Important spike protein features are indicated above the graph. HR1, helical repeat 1; HR2, helical repeat 2; NTD, N terminal domain; NTD super, N terminal domain supersite; RBD, receptor-binding domain; S1, spike 1; S2: Spike 2; TM, transmembrane domain.

References

1. O'Toole Á, Hill V, McCrone JT, Scher E, Rambaut A. Pangolin COVID-19 lineage assigner. 2020 [cited 2021 Sep 3]. https://pangolin.cog-uk.io/

2. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. COVID-19 dashboard. 2020 [cited 2021 Jan 24]. https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6