A SARS-CoV-2 lineage A variant (A.23.1) with altered spike has emerged and is dominating the current Uganda epidemic

Daniel Lule Bugembe[1*], My V.T.Phan[1,5*], Isaac Ssewanyana[2], Patrick Semanda[2], Hellen Nansumba[2], Beatrice Dhaala[1], Susan Nabadda[2] , Áine Niamh O'Toole[3], Andrew Rambaut[3], Pontiano Kaleebu[1,4], Matthew Cotten[1,6]

Affiliations

1. MRC/UVRI & LSHTM Uganda Research Unit, Entebbe, Uganda
2. Central Public Health Laboratories of the Republic of Uganda, Kampala, Uganda
3. Institute for Evolutionary Biology, University of Edinburgh
4. Uganda Virus Research Institute, Entebbe, Uganda
5. EMC, Rotterdam
6. MRC-University of Glasgow Centre for Virus Research, Glasgow, UK

*These authors contribute equally.

Correspondence: Matthew Cotten (Matthew.Cotten@lshtm.ac.uk)

Keywords: SARS-CoV-2, new variant, A.23; A.23.1, spike protein changes

Abstract

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was first detected in March 2020 in Uganda. Recently the epidemic showed a shift of SARS-CoV-2 variant distribution and we report here newly emerging A sub-lineages, A.23 and A.23.1, encoding replacements in the spike protein, nsp6, ORF8 and ORF9, with A.23.1 the major virus lineage now observed in Kampala. Although the clinical impact of the A.23.1 variant is not yet clear it is essential to continue careful monitoring of this variant, as well as rapid assessment of the consequences of the spike protein changes for vaccine efficacy.

28 **Main Text**

29 The novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)(1) and
30 the associated disease Coronavirus Disease 2019 (COVID-19)(2)(3) continue to spread
31 throughout the world, causing >100 million infections and >2 million deaths (Johns Hopkins
32 COVID-19 Dashboard). Genomic surveillance has played a key role in the responses to the
33 pandemic; particularly, sequence data from SARS-CoV-2 provide information on the
34 transmission patterns and the evolution of the virus as it enters new regions and spreads. As
35 COVID-19 vaccines become available, monitoring SARS-CoV-2 genetic changes, especially
36 changes at epitopes with implications for immune escape is crucial. A detailed classification
37 system has been defined to help monitor SARS-CoV-2 as it evolves(4) and novel virus
38 sequences first classified into 2 main phylogenetic lineages A and B representing the
39 earliest divergence of SARS-CoV-2 in the pandemic and then into sub-lineages within these.
40 Several Variants of Concern (VOCs) have emerged and spread with implications to
41 compromise vaccine efficacy and/or therapeutic antibody treatments. These VOCs include
42 lineage B.1.1.7 first identified in the UK (5), B.1.351 in South Africa (6) and lineage P.1
43 (B.1.1.28.1) in Brazil (7). The novel sub-lineage A (A.23.1) reported here encodes multiple
44 spike, nsp6, ORF8 and ORF9 protein changes, and some of the replacements are predicted
45 to be functionally similar to those observed in lineage B VOCs.

46

47 **Status of the SARS-CoV-2 epidemic in Uganda**

48 The first SARS-CoV-2 infection in Uganda was detected in March 2020. Initially the
49 virus was detected among international travellers until passenger flights were stopped in late
50 March 2020, a second route of virus entry was with truck drivers entering from adjacent
51 countries (8). Since August 2020, community transmission has dominated the Uganda case
52 numbers. By end of January 2021 detected case count in Uganda is 39,000, with 318 deaths
53 attributed to SARS-CoV-2 infection. The SARS-CoV-2 infection has spread throughout the

54 country, however the Kampala area is a major centre of virus infection where 60-80% of the
55 daily new cases have been identified during the months of June 2020 to January 2021
56 (Uganda COVID-19 Daily Situation Report). We have continued our efforts to generate
57 SARS-CoV-2 genomic sequence data to monitor virus movement and changes (8).
58
59 **Changes in prevalence of lineage A viruses**
60 The genomes were classified into Pango lineages(4) using the Pangolin module
61 pangoLEARN https://github.com/cov-lineages/pangolin) and into NextStrain clades using
62 NextClade (9) (https://clades.nextstrain.org/). Across the entire epidemic, 80 (39%) strains
63 belonged to the major lineage B whereas 123 (61%) strains were classified as lineage A
64 (Supplementary Table 1). This distribution of lineages changed dramatically over the course
65 of the year. A clear feature of the earlier COVID-19 epidemic in Uganda was the diversity of
66 viruses found throughout the country attributed to frequent flights into Uganda from Europe,
67 UK and Dubai (with origins further east); this is reflected in the range of lineages seen from
68 March to May 2020 in the Kampala region (Figure 1, left panel). After passenger flights were
69 limited in March, the virus was still able to enter via land travel, primarily with truck drivers.
70 Uganda is landlocked country, characterised by its important geographical position, i.e. the
71 crossing of two main routes of the Trans-Africa Highway in East Africa. The essential nature
72 of produce and goods transport, therefore, resulted in potential virus movement from/to
73 Kenya, South Sudan, DRC, Rwanda and Tanzania. In the period of June to October 2020
74 characterised by truck driver movement and no flights, lineage B.1 strains predominated,
75 similar to pattern observed in Kenya(10) (Figure 1 middle panel). Given the diversity of virus
76 lineages found in the country from March until October 2020, it was unexpected that by late
77 December 2020 to January 2021 almost exclusively lineage A viruses (N=49) were found in
78 Kampala with only one B.1 (Figure 1 right panel).In all time periods, the SARS-CoV-2
79 positive sample were obtained from 9 or more clinical locations throughout the Kampala

region indicating that the differences are unlikely to be due to sampling different subpopulations in the city at different times.

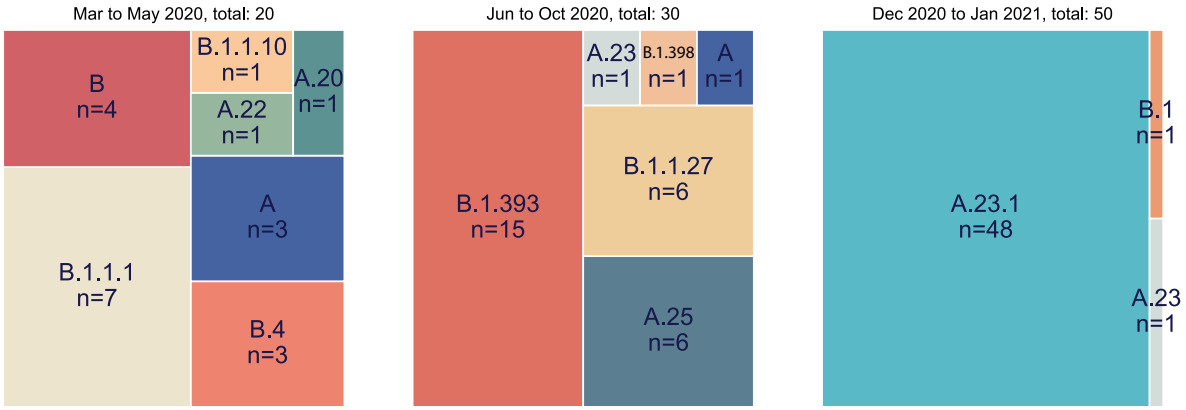


Figure 1 SARS-CoV-2 lineage diversity Kampala. All full genomes from the Kampala area were lineage typed using the pangolin resource (https://github.com/cov-lineages/pangolin) Lineage counts were stratified into three collection periods (March-May 2020, June-October 202 and December 2020 to January 2021). The percentage of each lineage within each set was plotted as a treemap using squarify (https://github.com/laserson/squarify)with the size of each sector proportional to the number of genomes, genomes numbers are listed with "n=".

Virus sequence diversity including fatal cases

To monitor the epidemic in more detail, full genome SARS-CoV-2 sequences were generated from SARS-CoV-2 positive samples in Uganda using 1500bp amplicon followed by MinION sequencing using modified 1500 bp amplicon method(11). All newly and previously generated genomes that are complete and high-coverage (N=203) were used to construct a maximum-likelihood phylogenetic tree (Figure 2). Important details to note: the genome sequences from 6 lethal Ugandan cases belonged to two lineages A.25 and

B.1.393 (Figure 2).

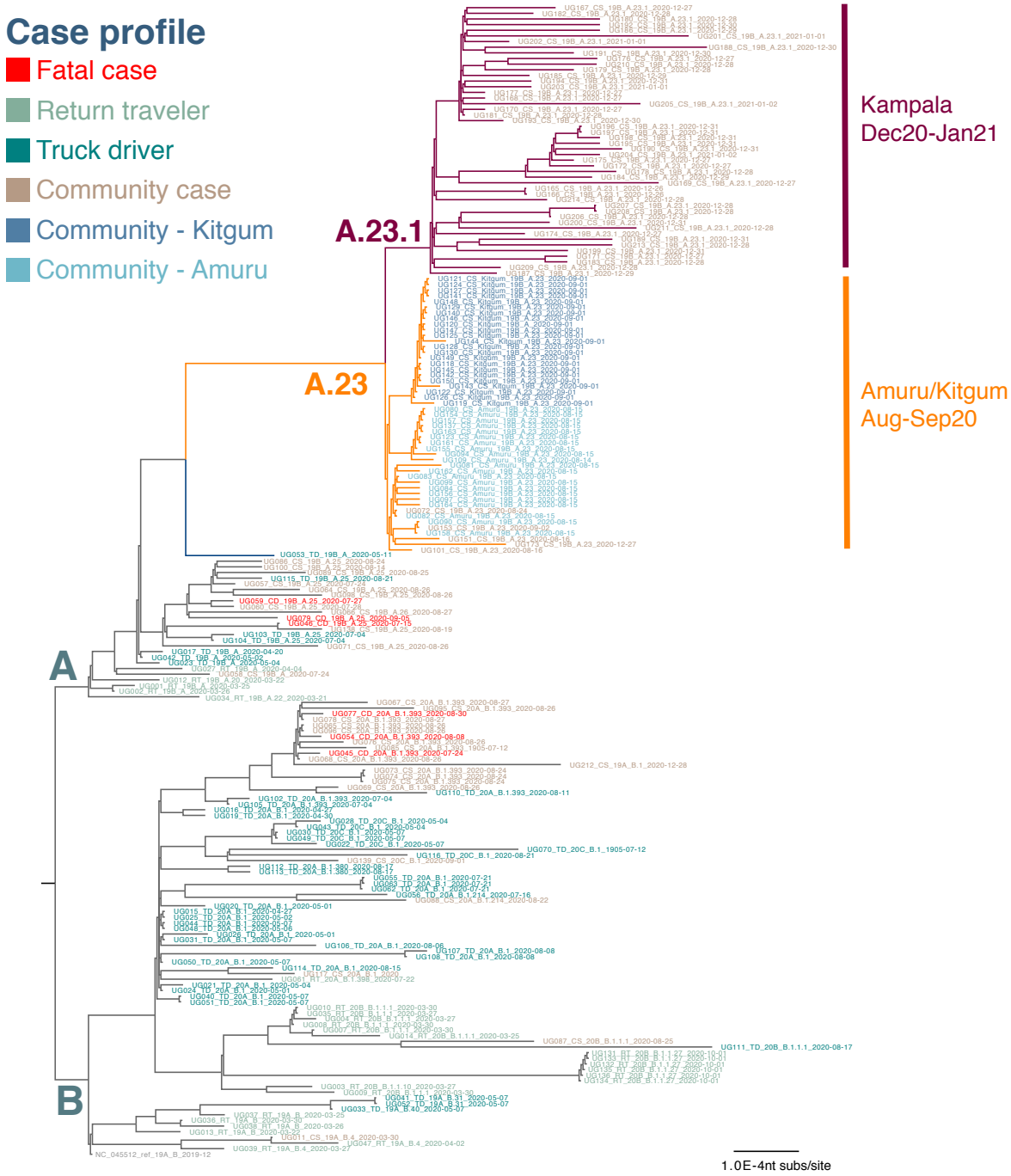


Figure 2. Maximum-likelihood phylogenetic tree comparing all available complete and high-coverage Uganda sequences (N=203). Strain names are coloured according to the case profile (community: light brown, truck driver: dark green, return traveller: light green, lethal cases: red). The clusters from the Amuru (light blue), Kitgum (dark blue) and the recent community Kampala genomes are indicated. The tree was rooted where lineages A and B were split. The branch length is drawn to the scale of number of nucleotide substitutions per site, indicated in lower right.

Of note, the SARS-CoV-2 lineage A is far less prevalent than lineage B in Europe, UK and USA as compared to Asia. The presence of lineage A viruses from lethal and community cases throughout Uganda indicates that this lineage is circulating in Uganda and capable of producing a severe infection. Several variant lineages were observed at low frequencies and only briefly and may have undergone apparent extinction, similar to patterns observed in the UK (12) and Scotland (13).

A genome identified from a truck driver is often observed basal to each cluster (Figure 2), suggesting the importance of this route in the introduction and spread of the virus into Uganda. Additionally, genomes identified from truck drivers could provide important information, especially those truck drivers coming from ports of entries (POEs) bordering countries with limited genomic data on contemporary SARS-CoV-2 circulation, such as South Sudan and Tanzania (data not shown). As mentioned earlier, most of genomes identified from truck drivers coming from POEs bordering Kenya belonged to lineage B.1, consistent with the pattern reported in Kenyan cases (Supplementary Table 1). On the other hand, genomes identified from truck drivers from Tanzania, albeit small numbers, belonged to both A and B lineages (data not shown). From the small number of genomes from the Elegu POE bordering South Sudan, the viruses belonged to lineage A and B.1; the lineage A strain from this truck driver (strain UG053) was basal to the newly emergent A.23 variant (discussed below). Continued monitoring of all truck drivers coming in and out of the Uganda is therefore very important and would help us to better understand the inland cross-country import and export and circulation of strains in this part of world, where (large scale) genomic surveillance is scarce and resource is limited.

SARS-CoV-2 outbreaks in prisons

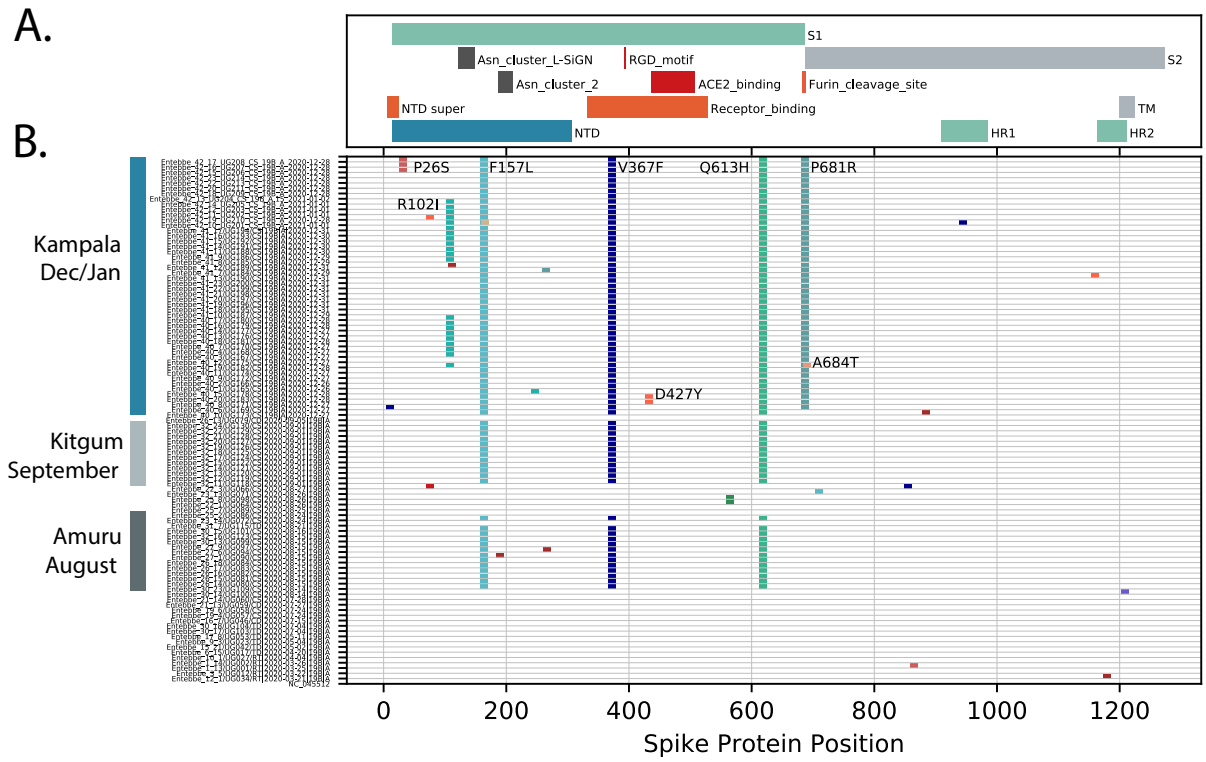
Outbreaks of SARS-CoV-2 infections were reported in the Amuru and Kitgum prisons in August 2020(14),(15). The SARS-CoV-2 genome sequences identified from individuals

133 from Amuru and Kitgum prisons belonged exclusively to a lineage A (Figure 2) with three
134 amino acid (aa) changes encoded in the spike protein (F157L, V367F and Q613H, Figure 3)
135 that now define lineage A.23 (see below). Some individuals from Amuru prison were
136 transferred to Kitgum prison, potentially facilitating virus movement between these prisons.
137 By October 2020, lineage A.23 viruses were also found outside of the prisons in a
138 community sample from Lira (a town 140 km from Amuru), and in two samples from the
139 Kitgum hospital and in a community sample from Kampala. Lineage A viruses contributed to
140 only 25% of the viruses in the Kampala region (Figure 1) in the period June to October 2020,
141 which was consistent with the variety of virus entering Uganda and Kampala via international
142 travellers and truck drivers seen in the initial period of the epidemic(8),(16). By December
143 2020, 49 of the 50 sequenced samples from the Kampala region (from 9 clinical sites)
144 belonged to the new A.23.1 lineage (Figure 1 and 2) The pattern of mutations in these virus
145 sequences was consistent with their evolution from the original A.23 viruses observed in
146 Amuru/Kitgum cluster (Figure 2). A plot of nucleotide changes over time for Ugandan lineage
147 A viruses (results not shown) showed a consistent evolutionary rate of roughly 2 nucleotide
148 change per month that has been observed for SARS-CoV-2 throughout the pandemic
149 (17),(18).

150 **Important changes observed in the spike protein**

151 The spike protein is crucial for virus entry into host cells, for tropism, and is a critical
152 component of COVID-19 vaccine development and monitoring. The changes in spike protein
153 observed in Uganda and global A.23 and A.23.1 viruses are shown in Figure 3. Many amino
154 acid (aa) changes were single events with no apparent transmission observed. However, the
155 initial lineage A.23 genomes from Amuru and Kitgum encoded three amino acid changes in
156 the exposed S1 domain of spike, including F157L, V367F and Q613H (Figure 3). The V367F
157 change is reported to modestly increase infectivity(19). Importantly, the Q613H change may
158 have similar consequences as the D614G change observed in the B.1 lineage found

159 predominantly in Europe and USA; in particular, D614G was reported to increase infectivity,
160 spike trimer stability and furin cleavage (19),(20),(21),(22). These changes were not
161 encoded by the closest known related genome (strain UG053) from a truck driver entering
162 from South Sudan (Figure 2) and were not observed in previously reported genomes from
163 Uganda (8).



164

165 **Figure 3. Spike protein changes in lineage A.23 and A.23.1 relative to the SARS-CoV-2**
166 **reference strain (NC_045512) encoded protein are documented. Lower panel:** Each line
167 represents the encoded spike protein sequence from a single genome, ordered by date of
168 samples collection (bottom earliest, top most recent). Sequences from Amuru in August
169 2020, Kitgum in September 2020 and Kampala December 2020/January 2021 are indicated.
170 Markers indicating the positions of amino acid (aa) differences from the reference strain,
171 changes observed in multiple genomes are annotated with the annotation (original aa
172 position new aa). Upper panel: The locations of important spike protein features are
173 indicated. NTD: N-terminal domain, RBD: receptor-binding domain, S1: spike 1, S1: Spike 2,
174 TM: transmembrane domain, HR1: helical repeat 1, HR2: helical repeat 2, NTD super: N-
175 terminal domain supersite.
176

177 Of concern, the recent Kampala and global A.23.1 virus sequences from December
178 2020-January 2021 now encoded 4 or 5 amino acid changes in the spike protein (now
179 defining lineage A.23.1, see below) plus additional protein changes in nsp3, nsp6, ORF8

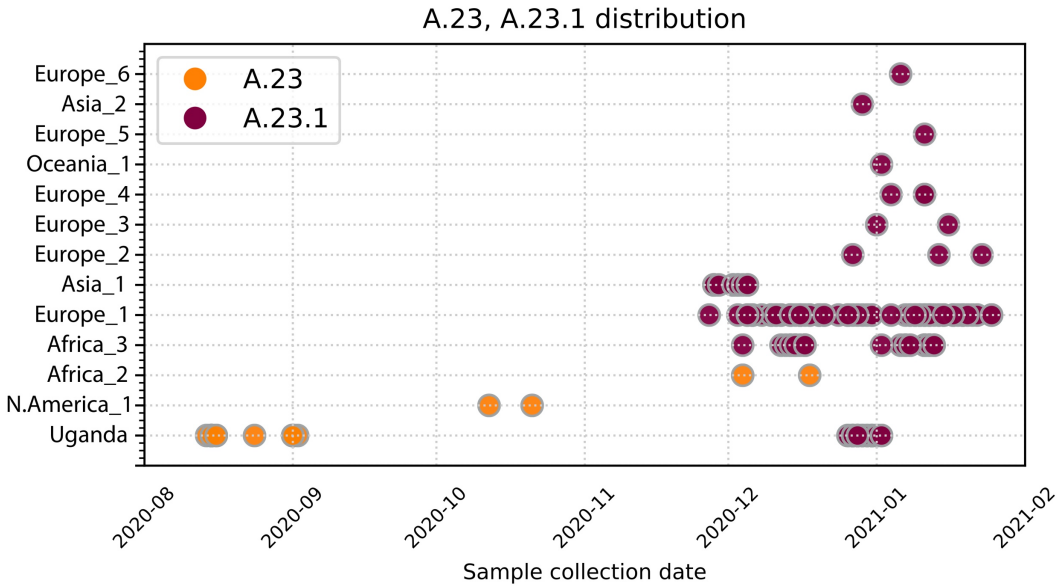
180 and ORF9 (Figure 3, 4). The P681R spike change encoded by all recent Kampala genomes
181 since December 2020 adds a basic amino acid adjacent to the spike furin cleavage site. This
182 same change has been shown *in vitro* to enhance the fusion activity of the SARS-CoV-2
183 spike protein, likely due to increased cleavage by the cellular furin protease (23);
184 importantly, a similar change (P681H) is encoded by the recently emerging VOC B.1.1.7 that
185 is now spreading globally across 75 countries as of 5 February 2021 (5) (24). There are also
186 changes in the spike N-terminal domain (NTD), a known target of immune selection,
187 observed in samples from Kampala A.23.1 lineage, including P26S and R102I plus 8
188 additional singleton changes (observed in only one genome, Figure 3).

189 **New lineage A designations.**

190 The viruses detected in Amuru and Kitgum met the criteria for a new SARS-CoV-2
191 lineage (25)(26) by clustering together on a global phylogenetic tree, sharing epidemiological
192 history and source from a single geographical origin, and encoding multiple defining SNPs.
193 These features including especially the three spike changes F157L, Q613H and V367F
194 define the new lineage A.23. Continued circulation and evolution of A.23 in Uganda was
195 observed and two additional changes in spike R102I and P681R were observed in
196 December 2020 in Kampala; these SNPs define the sub-lineage A.23.1. Additional changes
197 in non-spike regions also define the A.23 and A.23.1, including nsp3: E95K, nsp6: M86I,
198 L98F, ORF 8: L84S, E92K and ORF9 N: S202N, Q418H. These new lineages can be
199 assigned since pangolin version v2.1.10 and pangoLEARN data release 2021-02-01.

200 Screening SARS-CoV-2 genomic data from GISAID, viruses from A.23 and A.23.1
201 are now found in 12 countries outside Uganda (from Africa, Asia, Europe, North America
202 and Oceania) indicating global movement of the newly emerging variants (Figure 4). In a
203 screen of all available lineage A viruses in GISAID (Feb 4, 2021), the A.23 variant was first
204 observed in Uganda in August 2020 and then in a country in North America (denotated as
205 N.America_1) in October and in country in Africa (denotated as Africa_2) in December

206 (Figure 4). The variant A.23.1 was first seen in December in 2020 although we have a 2-
207 month gap in Uganda sequence data from October/November 2020. Outside of Uganda,
208 A.23.1 was found in another country in Africa (Africa_3) from the end of November in 9
209 different countries across Europe (6 countries), Asia (2 countries) and Oceania (1 country).
210 Of note, the international flights out of Uganda were restarted on 1 October 2020 with
211 frequent flights to Europe and US overlaying via a country in Africa or Asia. Phylogenetic
212 analysis supports the close evolution of A.23 to A.23.1 (Supplementary Figure 1).



213

214 **Figure 4. Current global distribution of A.23 and A.23.1.** All available SARS-CoV-2
215 complete genomes annotated as complete and lineage A from GISAID were retrieved on
216 Feb 4 2021 and lineage typed using Pangolin(27). and confirmed as A.23 and A.23.1 by
217 extracting examining the encoded spike protein. A.23 and A.23.1 genomes were plotted by
218 country and sample collection date. Countries were anonymized by continent.
219

220 **Additional changes in Ugandan A.23 and A.23.1 genomes compared to other VOC**
221 **genomes**

222 There are changes in other genomic regions of the virus accompanying the
223 adaptation. We employed profile Hidden Markov Models (pHMMs) prepared from 44 amino
224 acid peptides across the SARS-CoV-2 proteome to detect and visualize protein changes
225 from the early lineage B reference strain NC_045512. Measuring the identity score (bit-

226 scores) of each pHMMs across a query genome provides a measure of protein changes (in
227 44 amino acid steps) across the viral genome (Figure 5A). Applying this method to the most
228 recent lineage A.23.1 genome sequences the changes in spike (discussed above) as well as
229 changes in the transmembrane protein nsp6 and the interferon modulators ORF8 and 9
230 (Figure 5A). Modest changes were also observed in nsp13.

231

232 **Figure 5. All protein changes across lineage new variants.** All forward open reading
233 frames from the 35 early lineage B SARS-CoV-2 genomes were translated, and processed
234 into 44 aa peptides (with 22 aa overlap), clustered at 0.65 identity using Uclust (28), aligned
235 with MAAFT (29) and converted into pHMMs using HMMER-3(30). The presence of each
236 domains and its bit-score (a measure of the similarity between the query sequence and the
237 sequences used for the pHMM(30)) was sought in each set of SARS-CoV-2 VOC genomes
238 and the 1-mean of the normalized domain bit-scores was plotted across the genome (e.g. 1 -
239 the similarity of the identified query domain to the reference lineage B SARS-CoV-2
240 domain). Domains were coloured by the proteins from which they were derived with the
241 colour code indicated below the figure. **Panel A.** Query set are 49 most Uganda lineage
242 A.23.1 genomes. **Panel B.** All B.1.1.7 full genomes lacking Ns deposited in GISAID on Jan

26 2021, **Panel C.** All B.1.351 full genomes lacking Ns present in GISAID on Jan 26 2021, **Panel D.** All P.1 full genomes lacking Ns present in GISAID on Jan 26 2021.

We asked if a similar pattern of evolution was appearing in VOCs as SARS-CoV-2 adapted to human infection. We applied the same pHMM analysis to compare set of VOC or A.23.1 lineage viruses to the original SARS-CoV-2 lineage B genome sequences observed in December 2019/January 2020. The plots in Figure 5 show the difference in 44 amino acid steps across the viral proteome (see Methods for further details). We prepared sets of SARS-CoV-2 genomes annotated as B.1.1.7, B.1.351 or P.1 in GISAID and the A.23.1 genomes from this study. Both the A.23.1 and the B.1.1.7 lineage encode nsp6, spike, ORF8 and 9 changes (Figure 5B). Lineage B.1.351 encodes nsp6, spike and ORF6 changes (Figure 5C) and lineage P.1 encodes nsp6, spike and ORF6 changes (Figure 5D). Although the exact amino acid and positions of change within the proteins can differ in each lineage, there are some striking similarities in the common proteins that have been altered. Of interest, the nsp6 change present in B.1.1.7, B.1.351 and P.1 is a 3 amino acid deletion (106, 107 and 108) in a protein loop of nsp6 predicted to be on exterior of the autophagy vesicles on which the protein accumulates(31).The three amino acid nsp6 changes of lineage A.23.1 are L98F in the same exterior loop region, and the M86I and M183I changes predicted to be in intramembrane regions but adjacent to where the protein exits the membrane(31).

The spike changes affect the immunogenic N-terminal domain, the receptor binding domain, positions 613/614 (two lineages) and the furin cleavage site. A compilation of the amino acid changes in these lineages is found in Supplementary Table 2 with proteins that are altered in all 4 lineages marked in red.

Discussion

We report the emergence and spread of a new SARS-CoV-2 variant of the A lineage (A.23.1) with multiple protein changes throughout the viral genome. A similar phenomenon recently occurred with the B.1.1.7 lineage, detected first in the southeast of England (5) and now globally and with the B.1.351 lineage in South Africa(6), and P.1 lineage in Brazil(7) suggesting that local evolution and spread may be a common feature of SARS-CoV-2. Importantly, the lineage A.23.1 variant shares features found in the lineage B VOCs: alternation in key spike protein regions, especially the furin cleavage site and the 613/614 change that may increase spike multimer formation. The VOCs and A.23.1 strains also encode changes in similar region of the nsp6 protein which may be important for altering cellular autophagy pathways that promote replication. Changes or disruption of ORF7,8 and 9 are also present in the VOCs and A.23.1. The ORF8 changes or deletion probably indicates this protein is unnecessary for human replication, similar deletions accompanied SARS-CoV-2 adaption to humans(32),(33).

We suspect that all the emerging SARS-CoV-2 may be adjusting to infection and replication in humans and it is notable that the VOCs and lineage A.23.1 share a common patterns in their evolution. The spike changes are best understood due to the massive global effort to define the receptor and develop vaccines against the infection. The analysis reported in Figure 5 reveals common functions of SARS-CoV-2 that have been altered in all four variants, especially nsp6 and the ORF 7, 8 and 9. The functional consequences of the additional non-spike changes warrant additional studies and the current analysis may focus efforts of the proteins that are commonly changed in the variant lineages.

Methods

Sample collection, whole genome MinION sequencing and genome assembly

Residual nucleic acid extract from SARS-CoV-2 RT-PCR positive samples were obtained from Central Public Health Laboratory (Kampala, Uganda). The nucleic acid was

296 converted to cDNA and amplified using SARS-CoV specific 1500bp-amplicon spanning the
297 entire genome as previously described(11).The resulting DNA amplicons were used to
298 prepare sequencing libraries, barcoded individually and then pooled to sequence on MinION
299 R.9.4.1 flowcells, following the standard manufacturer's protocol.
300 The genome assembies were performed as previously described (8). Briefly, reads
301 from fast5 files were basecalled and demultiplexed using Guppy 3.6 running on the UMIC
302 HPC. Adapters and primers sequences were removed using Porechop
303 (https://github.com/rrwick/Porechop) and the resulting reads were mapped to the reference
304 genome Wuhan-1 (GenBank NC_045512.2) using minimap2(34) and consensus genomes
305 were generated in Geneious (Biomatters Ltd). Genome polishing was performed in Medaka,
306 and SNPs and mismatches were checked and resolved by consulting raw reads.
307
308 **Phylogenetic analyses**
309 For the local Uganda virus comparison, all newly and previously generated genomes
310 from Uganda (N=203, excluding 3 low coverage genomes) were aligned using MAFFT(29)
311 and manually checked in AliView(35). The 5' and 3' untranslated regions (UTRs) were
312 trimmed. Maximum-likelihood (ML) phylogenetic tree was constructed in IQTREE(36) under
313 the GTR+F+R2 model as best-fitted substitution model according to Akaike Information
314 Criterion (AIC) determined by ModelFinder in IQTREE and run for 100 pseudo-replicates.
315 Resulting tree was visualised in Figtree(37) and rooted at the point of splitting lineage And B.
316 For phylogenetic analyses of Uganda lineage A.23 and A.23.1 strains comparing to
317 global A.23 and A.23.1 strains, the global SARS-CoV-2 lineage A.23 and A.23.1 genomes
318 were retrieved from GISAID on 4 February 2021 (N = 156). These global lineage A.23 and
319 A.23.1 genomes combining with Ugandan A.23 and A.23.1 genomes (N = 97) and an
320 outgroup of this A.23 lineage (Uganda strain UG053) were aligned using MAFFT and
321 manually checked in AliView, followed by trimming 5' and 3' UTRs. From this alignment, the

322 spike protein region was extracted for phylogenetic construction. The global and Ugandan
323 A.23 and A.23.1 genomes were used to construct a ML tree under the GTR+F+I model as
324 best-fitted substitution model according to AIC determined by ModelFinder and run for 100
325 pseudo-replicates in IQTREE. A ML tree was constructed in IQTREE for the nucleotide
326 sequences encoding the spike protein from the global and Ugandan A.23 and A.23.1 strains,
327 under the TPM2+F+I model of evolution as best-fitted model estimated by ModelFinder AIC
328 and run for 100 pseudo-replicates. Resulting trees were visualised in Figtree and rooted
329 using the strain UG053.
330
331 **Ethical approvals**
332 This study was approved by the Uganda Virus Research Institute- Research and
333 Ethics Committee (UVRI-REC Federalwide Assurance [FWA] FWA No. 00001354, study
334 reference. GC/127/20/04/771) and by the Uganda National Council for Science and
335 Technology, reference number HS936ES. The novel reported SARS-CoV-2 genomes are
336 available on GISAID (https://www.gisaid.org/) under the accession numbers
337 EPI_ISL_954226-EPI_ISL_954300 and EPI_ISL_955136.
338
339
340 **Acknowledgements**
341 We thank all global SARS-CoV-2 sequencing groups for their open and rapid sharing
342 of sequence data and GISAID for providing an effective platform for making these data
343 available. We are grateful to the Oxford Nanopore Technologies and the ARTIC Network for
344 their support and we thank Pope Moseley for his constructive comments on the manuscript.
345 The SARS-CoV2 diagnostic and sequencing award is jointly funded by the UK Medical
346 Research Council (MRC) and the UK Department for International Development (DFID)
347 under the MRC/DFID Concordat agreement (grant agreement number NC_PC_19060) and

is also part of the EDCTP2 programme supported by the European Union. The UMIC high performance computer was supported by MRC (grant number MC_EX_MR/L016273/1) to PK. A.R. acknowledges the support of the Wellcome Trust (Collaborators Award 206298/Z/17/Z ARTIC network) and the European Research Council (grant agreement no. 725422 – ReservoirDOCS). The study is additionally funded by the Wellcome, DFID - Wellcome Epidemic Preparedness – Coronavirus (grant agreement number 220977/Z/20/Z) awarded to MC.

Author contribution statement

All authors contributed to the work presented in this paper.

Competing Interests statement

The authors declare no competing interests.

References

1. Edward C. Holmes, Yong-Zhen Zhang EC. Initial genome release of novel coronavirus. Virological.org [Internet]. 2020 [cited 2021 Jan 24]; Available from: http://virological.org/t/319

2. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N Engl J Med. 2020 Jan 29;NEJMoa2001316.

3. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med. 2020 Feb;S2213260020300795.

4. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020 Nov;5(11):1403–7.

5. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data [Internet]. Infectious Diseases (except HIV/AIDS); 2021 Jan [cited 2021 Jan 29]. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.12.30.20249034

6. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, 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 [Internet]. Epidemiology; 2020 Dec [cited 2021 Jan 6]. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.12.21.20248640

7. Voloch CM, Silva F R da, de Almeida LGP, Cardoso CC, Brustolini OJ, Gerber AL, et al. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil [Internet]. Genetic and Genomic Medicine; 2020 Dec [cited 2021 Jan 30]. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.12.23.20248598

8. Bugembe DL, Kayiwa J, Phan MVT, Tushabe P, Balinandi S, Dhaala B, et al. Main Routes of Entry and Genomic Diversity of SARS-CoV-2, Uganda. Emerg Infect Dis. 2020 Oct;26(10):2411–5.

9. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Kelso J, editor. Bioinformatics. 2018 Dec 1;34(23):4121–3.

10. Githinji G, de Laurent ZR, Mohammed KS, Omuoyo DO, Macharia PM, Morobe JM, et al. Tracking the introduction and spread of SARS-CoV-2 in coastal Kenya [Internet]. Epidemiology; 2020 Oct [cited 2020 Dec 7]. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.10.05.20206730

11. Cotten M, Bugembe DL, Kaleebu P, Phan MVT. Alternate primers for whole-genome SARS-CoV-2 sequencing [Internet]. Genomics; 2020 Oct [cited 2020 Nov 30]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.10.12.335513

12. Page AJ, Mather AE, Le Viet T, Meader EJ, Alikhan N-FJ, Kay GL, et al. Large scale sequencing of SARS-CoV-2 genomes from one region allows detailed epidemiology and enables local outbreak management [Internet]. Epidemiology; 2020 Sep [cited 2020 Oct 9]. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.09.28.20201475

13. Filipe ADS, Shepherd J, Williams T, Hughes J, Aranday-Cortes E, Asamaphan P, et al. Genomic epidemiology of SARS-CoV-2 spread in Scotland highlights the role of European travel in COVID-19 emergence [Internet]. Infectious Diseases (except HIV/AIDS); 2020 Jun [cited 2020 Dec 14]. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.06.08.20124834

14. Daily Monitor. Amuru prison closed as 153 test positive for Covid-19. 2020 Aug 23; Available from: https://www.monitor.co.ug/uganda/news/national/amuru-prison-closed-as-153-test-positive-for-covid-19-1924660

15. Penelope Nankunda. COVID-19: Uganda registers 318 new cases in a single day. MSN [Internet]. 2020 Aug 22; Available from: https://www.msn.com/en-xl/news/other/covid-19-uganda-registers-318-new-cases-in-a-single-day/ar-BB18gprA

16. Matthew Cotten et al. SARS-CoV-2 diversity in Uganda, December, 2020. Virological.org [Internet]. 2021; Available from: https://virological.org/t/sars-cov-2-diversity-in-uganda-december-2020/571

17. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. Virus Evol. 2020 Jul 1;6(2):veaa061.

18. Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB, et al. The emergence of SARS-CoV-2 in Europe and North America. Science. 2020 Oct 30;370(6516):564–70.

19. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. Cell. 2020 Sep 3;182(5):1284-1294.e9.

20. Nguyen HT, Zhang S, Wang Q, Anang S, Wang J, Ding H, et al. Spike glycoprotein and host cell determinants of SARS-CoV-2 entry and cytopathic effects. J Virol. 2020 Dec 11;

21. Gobeil SM-C, Janowska K, McDowell S, Mansouri K, Parks R, Manne K, et al. D614G Mutation Alters SARS-CoV-2 Spike Conformation and Enhances Protease Cleavage at the S1/S2 Junction. Cell Rep. 2021 Jan 12;34(2):108630.

22. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole Á, et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell. 2020 Nov;S0092867420315373.

23. 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. Mol Cell. 2020 May;78(4):779-784.e5.

24. Áine O'Toole et al. B.1.1.7 report 2021-02-05. 2021; Available from: https://cov-lineages.org/global_report_B.1.1.7.html

25. Áine O'Toole, JT McCrone, Verity Hill and Andrew Rambaut. Pangolin COVID-19 Lineage Assigner. Available from: https://pangolin.cog-uk.io/

26. Rambaut A, Holmes EC, Hill V, O'Toole Á, McCrone J, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology [Internet]. Microbiology; 2020 Apr [cited 2020 Apr 27]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.04.17.046086

27. Áine O'Toole et al. Phylogenetic Assignment of Named Global Outbreak LINeages (PANGOLIN). 2020; Available from: https://github.com/cov-lineages/pangolin

28. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010 Oct 1;26(19):2460–1.

29. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013 Apr 1;30(4):772–80.

30. Eddy SR. Accelerated Profile HMM Searches. PLOS Comput Biol. 2011 Oct 20;7(10):e1002195.

31. Benvenuto D, Angeletti S, Giovanetti M, Bianchi M, Pascarella S, Cauda R, et al. Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. J Infect. 2020 Jul;81(1):e24–7.

32. Su YCF, Anderson DE, Young BE, Linster M, Zhu F, Jayakumar J, et al. Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. Schultz-Cherry S, editor. mBio. 2020 Jul 21;11(4):e01610-20, /mbio/11/4/mBio.01610-20.atom.

33. The Chinese SARS Molecular Epidemiology Consortium. Molecular Evolution of the SARS Coronavirus During the Course of the SARS Epidemic in China. Science. 2004 Mar 12;303(5664):1666–9.

34. Li H. Minimap2: pairwise alignment for nucleotide sequences. Birol I, editor. Bioinformatics. 2018 Sep 15;34(18):3094–100.

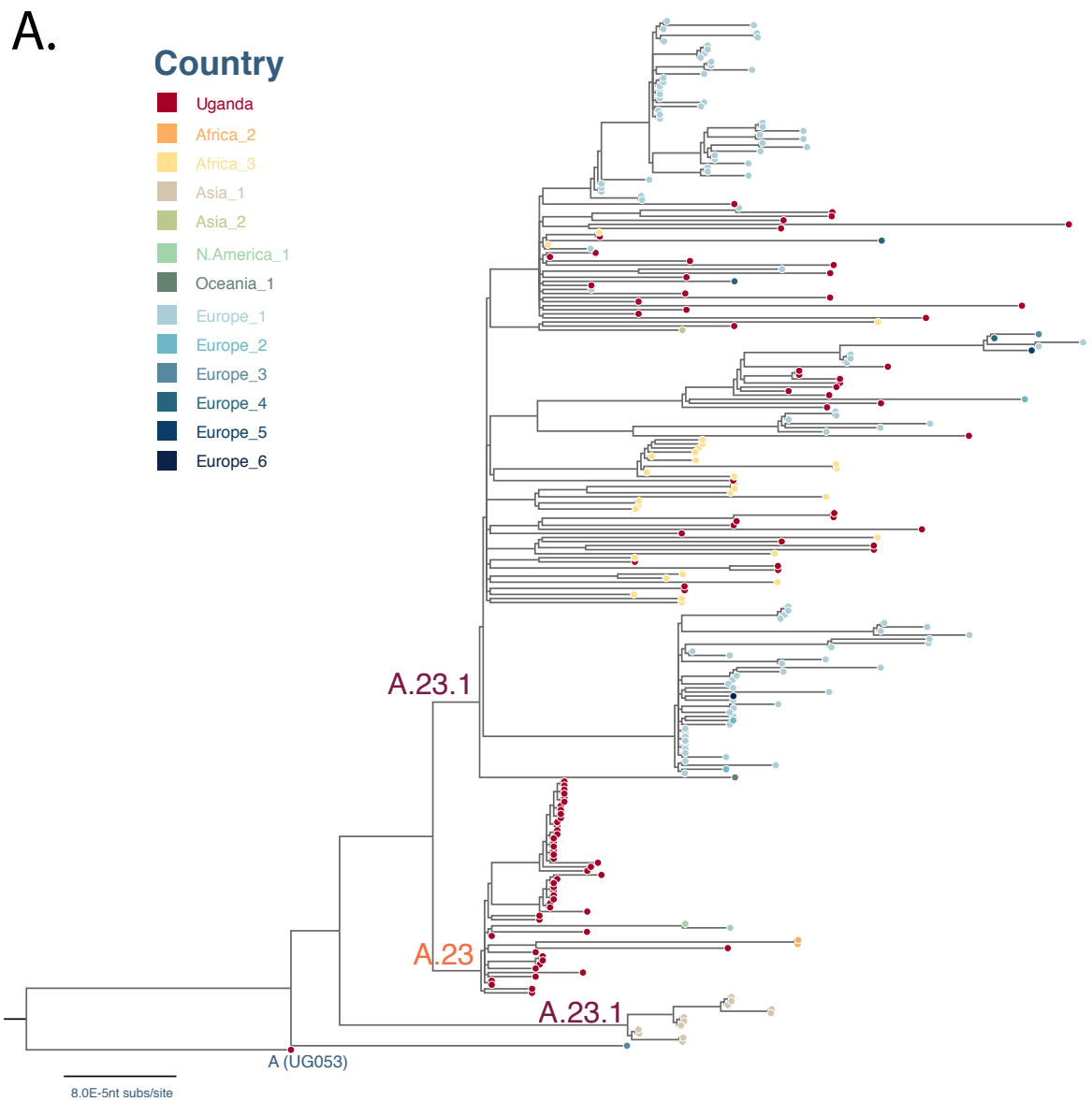
35. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014 Nov 15;30(22):3276–8.

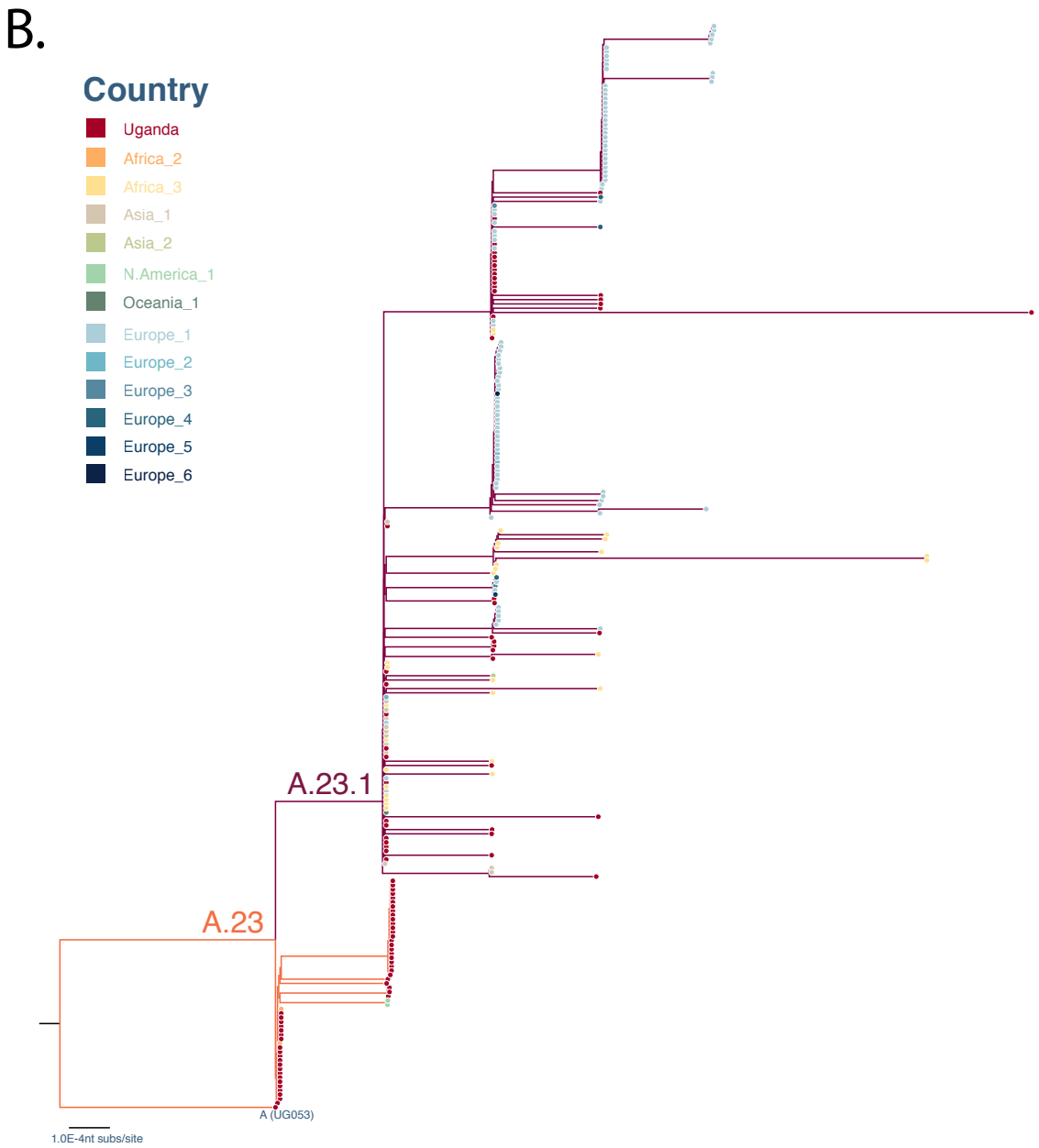
36. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015 Jan;32(1):268–74.

37. Rambaut A. FigTree http://tree.bio.ed.ac.uk/software/figtree. 2019;

38. Josh B. Singer, Gifford R, Cotten M, Robertson DL. CoV-GLUE project. 2020; Available from: http://cov-glue.cvr.gla.ac.uk/

Supplementary Material





Supplemental Figure 1. Maximum-likelihood phylogenetic tree comparing Uganda lineage A.23 and A.23.1 strains to global lineage A.23 and A.23.1. Panel A: A maximum-likelihood (ML) phylogenetic tree comparing Ugandan A.23 and A.23.1 (N = 97) with the global A.23 and A.23.1 (N = 156). **Panel B**: A ML phylogenetic tree of the nucleotide sequence encoding the spike protein region, comparing Ugandan A.23 and A.23.1 (N = 97) with the global A.23 and A.23.1 (N = 156). In both panels, the tree was rooted by the outgroup strain UG053 and strains were coloured according to the countries where they were identified. Branch length was drawn to the scale of number of nucleotide substitutions per site and trees were visualised in Figtree (37).

Supplementary Material Table 1. Lineage distribution in Uganda

Lineage[1]	Count of lineage
A	8
A.20	1
A.22	1
A.23	49
A.23.1	48
A.25	16
B	4
B.1	31
B.1.1.1	9
B.1.1.10	1
B.1.1.27	6
B.1.214	2
B.1.380	2
B.1.393	18
B.1.398	1
B.31	3
B.4	3
All lineage A and sub-A	123
All lineage B and sub-B	80
Total	203

Footnote.

1. All available genomes from Uganda were classified into Pango lineages using pangolin (27).

Supplementary Material Table 2. Summary of replacements in the A.23.1 and 3 VOC lineages[1].

Lineage	nsp2	nsp3	nsp5	nsp6	nsp12	spike	ORF3a	ORF4 E	ORF8	ORF9 N
A.23.1		nsp3: E95K		nsp6: M86I		S: R102I			ORF 8: L84S	N: S202N
A.23.1				nsp6: L98F		S: F157L			ORF 8: E92K	N: Q418H
A.23.1				nsp6: M183I		S: V367F				
A.23.1						S: Q613H				
A.23.1						S: P681R				
B.1.1.7		nsp3_T183I		nsp6_del_11288_11296	nsp12_P323L	S_N501Y			ORF 8_Q27*	N_D3L
B.1.1.7		nsp3_A890D				S_A570D			ORF 8_R52I	N_G204R
B.1.1.7		nsp3_I1412T				S_D614G			ORF 8_Y73C	N_R203K
B.1.1.7						S_P681H				N_S235F
B.1.1.7						S_T716I				
B.1.1.7						S_S982A				
B.1.1.7						S_D1118H				
B.1.1.7						S_del_21765_21770				
B.1.1.7						S_del_21991_21993				
B.1.351	nsp2_T85I	nsp3_K837N	nsp5_K90R	nsp6_del_11288_11296	nsp12_P323L	S_D80A	ORF 3a_Q57H	E_P71L		N_T205I
B.1.351						S_D215G				
B.1.351						S_K417N				
B.1.351						S_E484K				
B.1.351						S_N501Y				
B.1.351						S_D614G				
B.1.351						S_A701V				
B.1.351						S_non_cod_del_22281_22289				
B.1.281.1		nsp3_S370L		nsp6_del_11288_11296	nsp12_P323L	S_L18F	ORF 3a_S253P		ORF 8_E92K	N_P80R
B.1.281.1		nsp3_K977Q				S_T20N				N_R203K
B.1.281.1						S_P26S				N_G204R
B.1.281.1						S_D138Y				
B.1.281.1						S_R190S				
B.1.281.1						S_K417T				
B.1.281.1						S_E484K				
B.1.281.1						S_N501Y				
B.1.281.1						S_D614G				
B.1.281.1						S_H655Y				
B.1.281.1						S_T1027I				
B.1.281.1						S_V1176F				

1. Representative sets of each lineage (either all available gap-free full genomes (B.1.351 and B.1.281.1) or all genomes gap-free full genomes annotated as B.1.1.7 were analyzed using CoV-GLUE (38)to identify frequent amino acid replacements associated with each lineage.