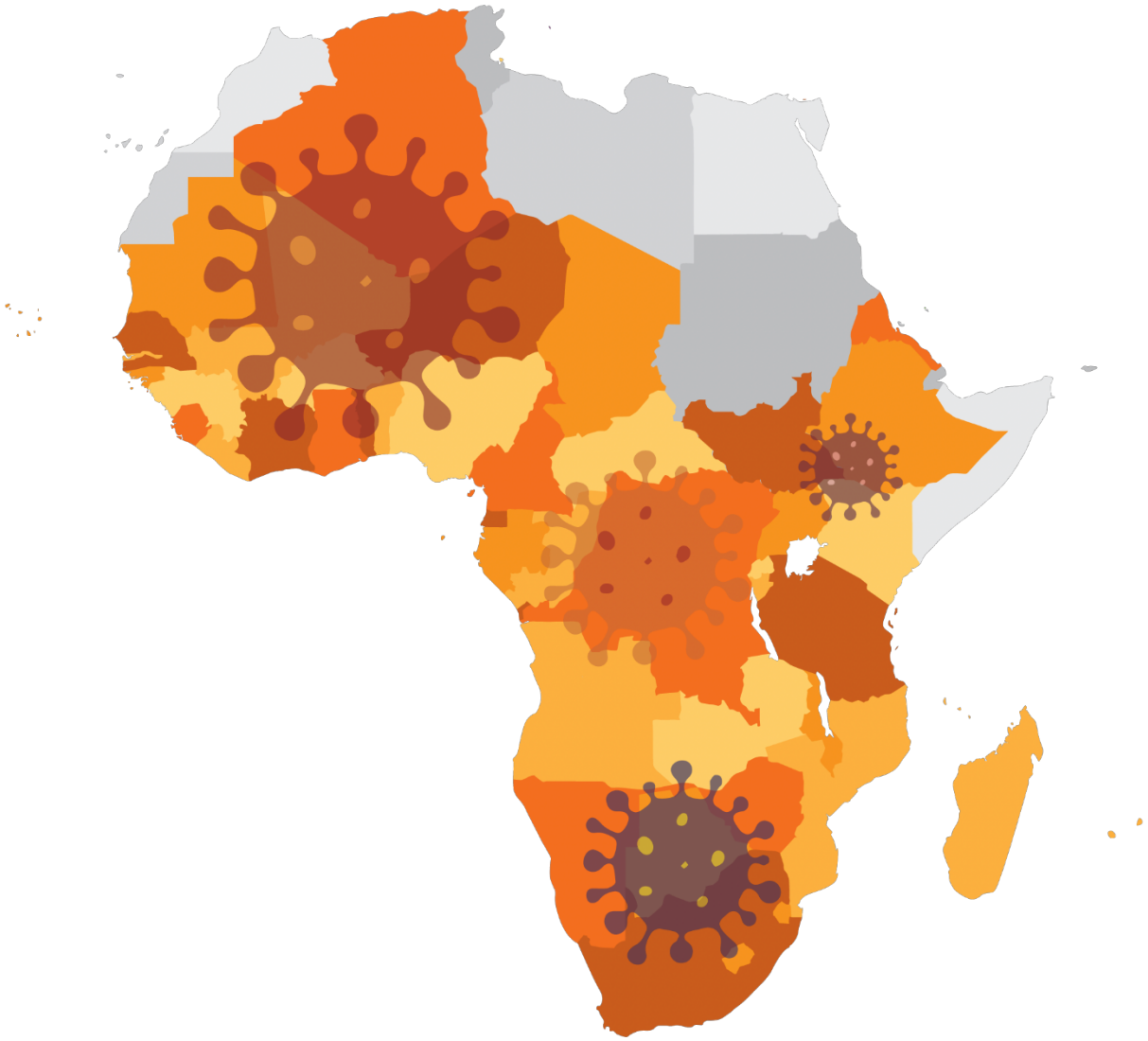
SARS-CoV-2 genomes report for WHO Africa Region

21/08/2020

Produced by the TIBA COVID-19 Pandemic Response Unit



NIHR Global Health Research Unit Tackling Infections to Benefit Africa at the University of Edinburgh

Table of Contents

Summary 3
1 Sequence data 4
2 Lineage diversity 5
3 D614G mutation 9
4 Phylogenetic analysis 10
5 Estimated number of introductions into African countries 13
Appendix 14
Distribution of gender 14
Distribution of age 14
Distribution of patient status 14

Summary

- SARS-CoV-2 genomes (n=1340) from fifteen countries in WHO Africa region (out of 47) are available. These represent ~2% of publicly available sequences globally.

- The majority of genomes are from South Africa (51%), Democratic Republic of the Congo (DRC) (20%) and Senegal (10%).

- 25 different lineages have been identified. 93% of all African SARS-CoV-2 genomes are assigned to sub-lineage B.1 and its descendant sub-lineages; B.1.106 is a South Africa dominant sub-lineage.

- There have been multiple separate introductions (currently estimated as at least 123 across the fifteen countries) into Africa from other continents. Of these, about 72% were from Europe.

- Most of the introductions do not appear to have spread between countries in Africa. However, 14 introductions from within Africa were tentatively identified.

- 93% of African SARS-CoV-2 genomes have the D614G mutation in spike protein. It has been suggested that this mutation increases virus infectivity.

1 Sequence data

This report summarises data comprising of all SARS-CoV-2 genome sequences from WHO Africa region on GISAID (https://www.gisaid.org). Sequences from African countries not within the WHO Africa region were excluded.

In total 1340 sequences from fifteen WHO Africa countries (out of 47) were downloaded from GISAID on 17th August 2020 (Figure 1). These represent ~2% of publicly available global sequences. The majority are from South Africa, Democratic Republic of the Congo (DRC) and Senegal, representing 51%, 20% and 10% of all Africa sequences, respectively. The earliest sequence collected in Africa was from Nigeria on 27th February 2020 and the most recent one is from South Africa on 18th July 2020.

803 sequences were submitted to GISAID since the last report: 550 sequences from three countries (South Africa, Senegal, and Nigeria) were isolated after 4th June 2020. Sequences from other countries were isolated between March and May (Figure 1 C).

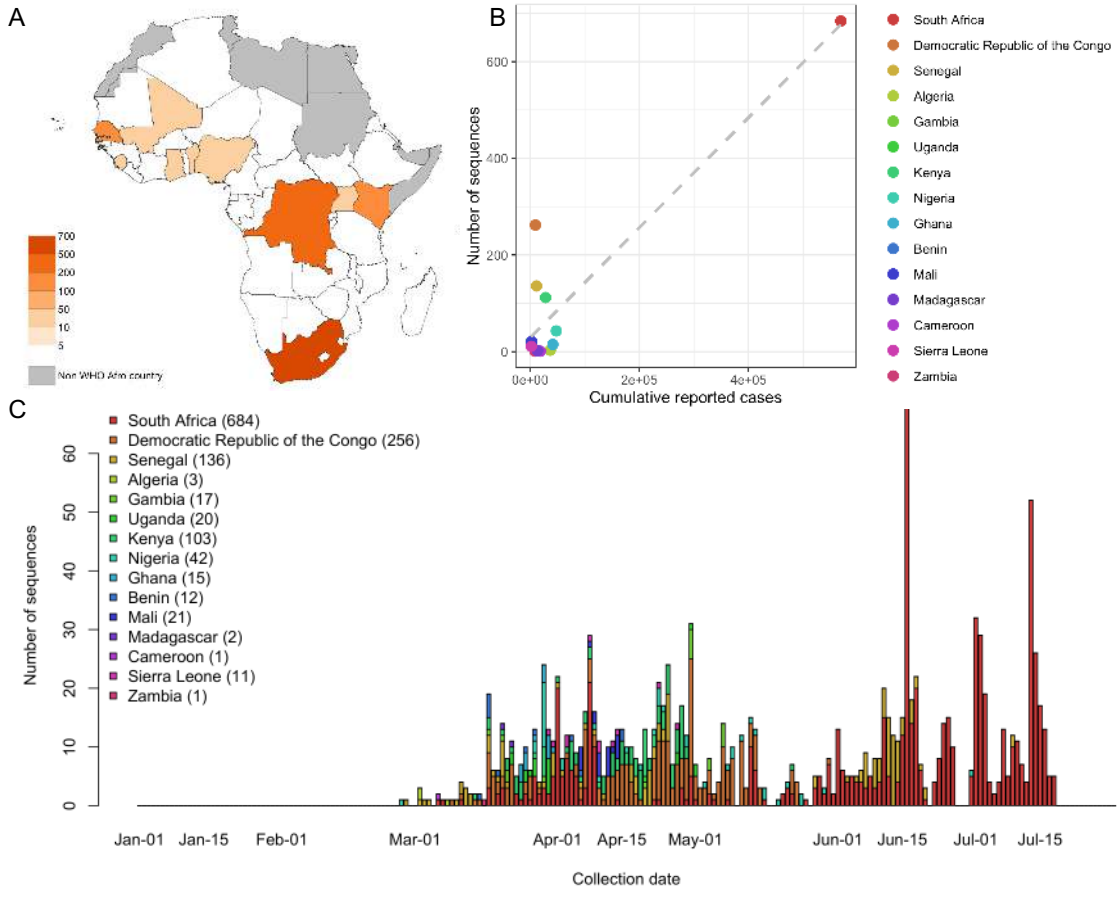


Figure 1 Number of sequences for WHO Africa countries. (A) SARS-CoV-2 genomes available on GISAID by country. (B) Comparison of cumulative reported cases and number of sequences available. (C) SARS-CoV-2 genomes available on GISAID by collection date. 6 genomes from the Democratic Republic of the Congo, 9 genomes from Kenya and 1 genome from Nigeria were removed due to incomplete collection dates.

2 Lineage diversity

The corresponding metadata for African sequences were obtained from GISAID. The lineage diversity for all African sequences is summarized below; the distribution of patient gender and age, as well as patient status, are summarized in the appendix.

All the sequences were screened and classified into lineages containing common mutations (https://pangolin.cog-uk.io and https://doi.org/10.1101/2020.04.17.046086). As of 17th August 2020, a total of 166 lineages have been identified globally; 136 of them are B.1 (which represents a large European lineage that corresponds to the Italian outbreak) and its descendant sub-lineages. There is no significant difference between the properties of these lineages in terms of transmissibility and virulence at the present time.

Within Africa, 1430 sequences were assigned to 25 lineages, including 23 sub-lineages (Table 1 and Figure 2). 3.4% of sequences from 10 countries were assigned to lineage A; the remaining sequences were assigned to lineage B, and the majority of them (93%) belong to sub-lineage B.1 and its descendant sub-lineages, 37% belong to sub-lineage B.1.1.

Compared to the last report:
6 descendant sub-lineages of B.1 were newly identified (B.1.9, B.1.106, B.1.84, B.1.102, B.1.2, B.1.36 and B.1.22); while 2 sub-lineages (A.p7 and B.1.1.8) were declassified due to being unobserved and inactive.

There is a South African dominant sub-lineage B.1.106: 88% (35 out of total 40) sequences available globally belonging to this lineage were from South Africa, others from UK and US), and both the earliest (28th March 2020) and latest (13th July 2020) sequence in this sub-lineage were isolated from South Africa (Figure 2 and Figure 3).

For the sequences isolated after 4th June, only 1 of them belongs to lineage A, all other sequences belong to lineage B.1 and its descendant sub-lineages B.1.1, B.1.1.1, B.1.5, B.1.36, B.1.102 and B.1.106 (Figure 3).

Table 1 Number of sequences per lineage per country

	Algeria	Benin	Cameroon	DRC	Gambia	Ghana	Kenya	Madagascar	Mali	Nigeria	Senegal	Sierra Leone	South Africa	Uganda	Zambia	TOTAL
A	0	4	0	6	0	2	2	0	13	8	3	5	1	4	0	48
A.2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
B	0	0	0	0	0	0	1	0	1	0	0	0	3	2	0	7
B.1	3	8	1	191	1	5	77	2	7	8	66	3	129	5	0	506
B.1.1	0	0	0	13	15	3	17	0	0	13	25	2	344	2	1	435
B.1.1.1	0	0	0	18	0	0	1	0	0	0	0	0	101	5	0	125
B.1.1.10	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
B.1.102	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
B.1.106	0	0	0	0	0	0	0	0	0	0	0	0	57	0	0	57
B.1.2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
B.1.22	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2
B.1.3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
B.1.36	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	4
B.1.5	0	0	0	0	0	0	4	0	0	1	38	0	24	0	0	67
B.1.6	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	5
B.1.8	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	16
B.1.84	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	22
B.1.9	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	3
B.2	0	0	0	1	0	3	1	0	0	0	0	0	1	1	0	7
B.2.1	0	0	0	2	0	1	3	0	0	3	0	0	1	0	0	10
B.2.2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
B.3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
B.4	0	0	0	1	0	0	5	0	0	0	0	0	0	1	0	7
B.6	0	0	0	0	1	0	1	0	0	0	1	1	4	0	0	8
B.9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
Unknown	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	3

*Unknown: sequences failed to be assigned into a lineage due to short length.

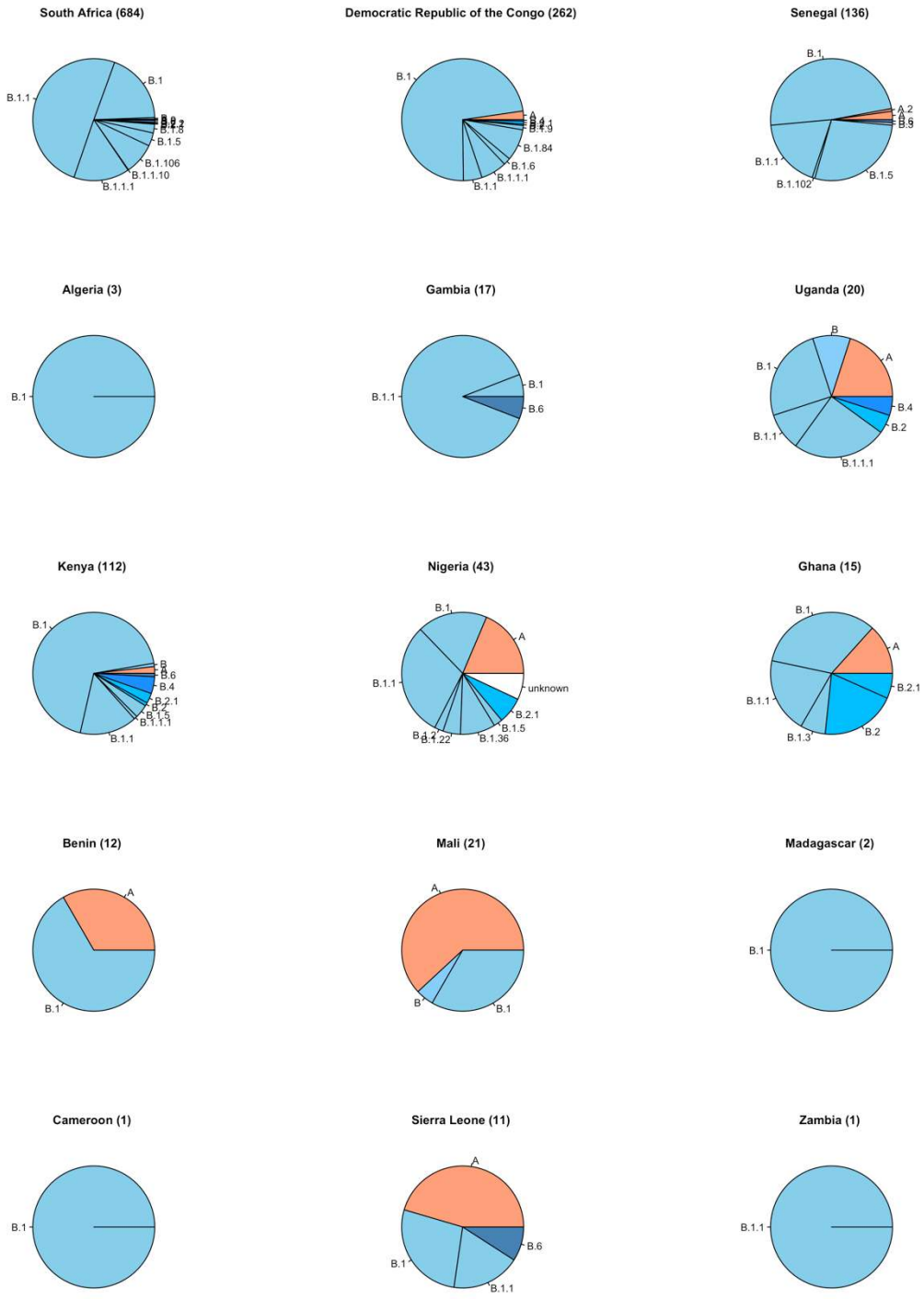


Figure 2 Lineage distribution per country. Lineage diversities are differentiated by red (lineage A and sub-lineages) and blue (lineage B and sub-lineages) categories; Number of sequences are shown beside country names.

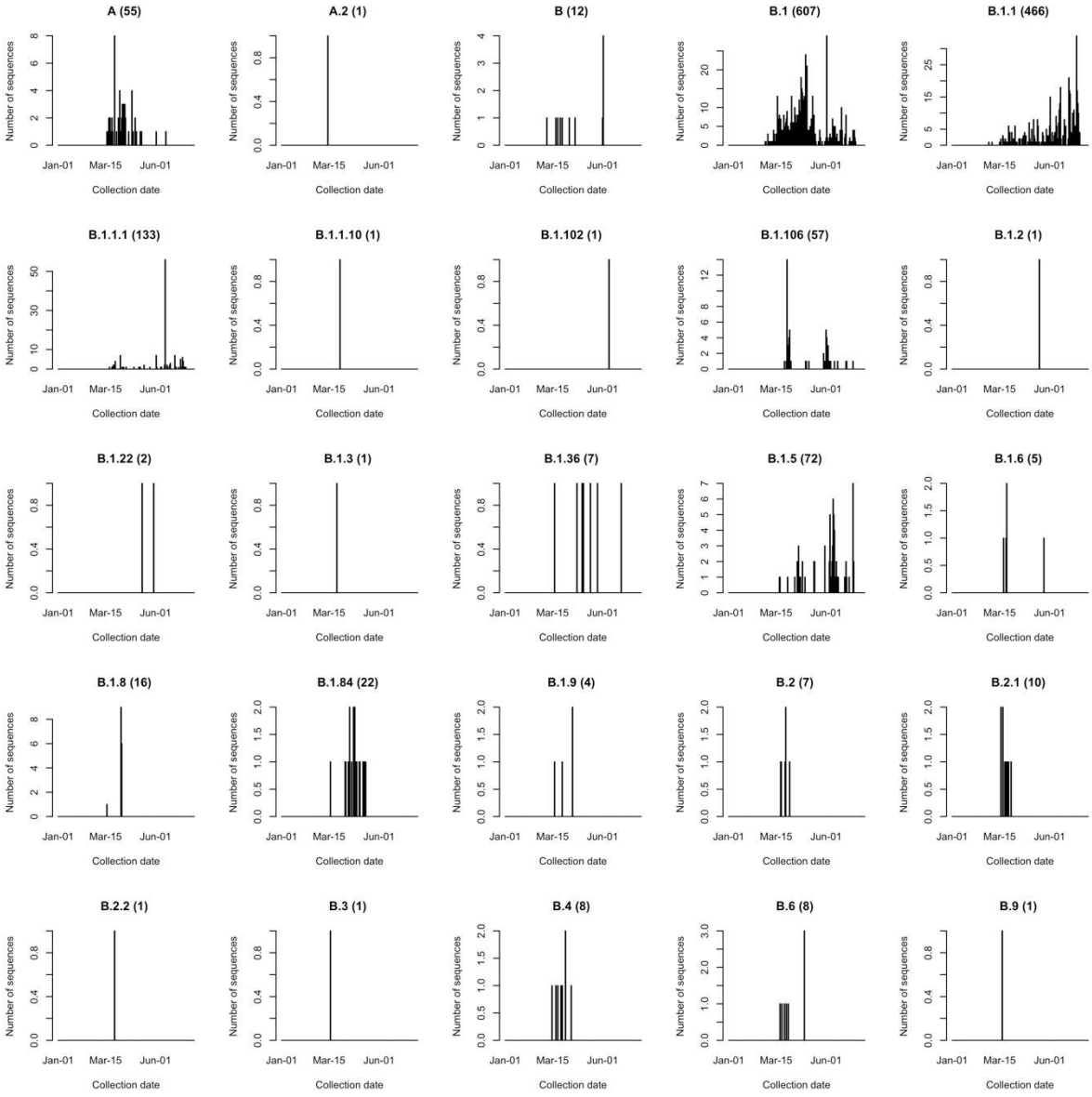


Figure 3 Lineage distribution via time. SARS-CoV-2 genomes within each lineage are shown by collection date. Number of sequences are shown beside country names.

3 D614G mutation

It is suggested the D614G mutation in the SARS-CoV-2 spike protein may increase the virus infectivity (https://www.biorxiv.org/content/10.1101/2020.06.12.148726v1). Among all African genomes, 1240 (93%) have the D614G mutation; 1234 (out of 1240) sequences belong to B.1 and descendant sub-lineages. The distribution of genomes with such mutation in individual countries is summarized in Figure 4.

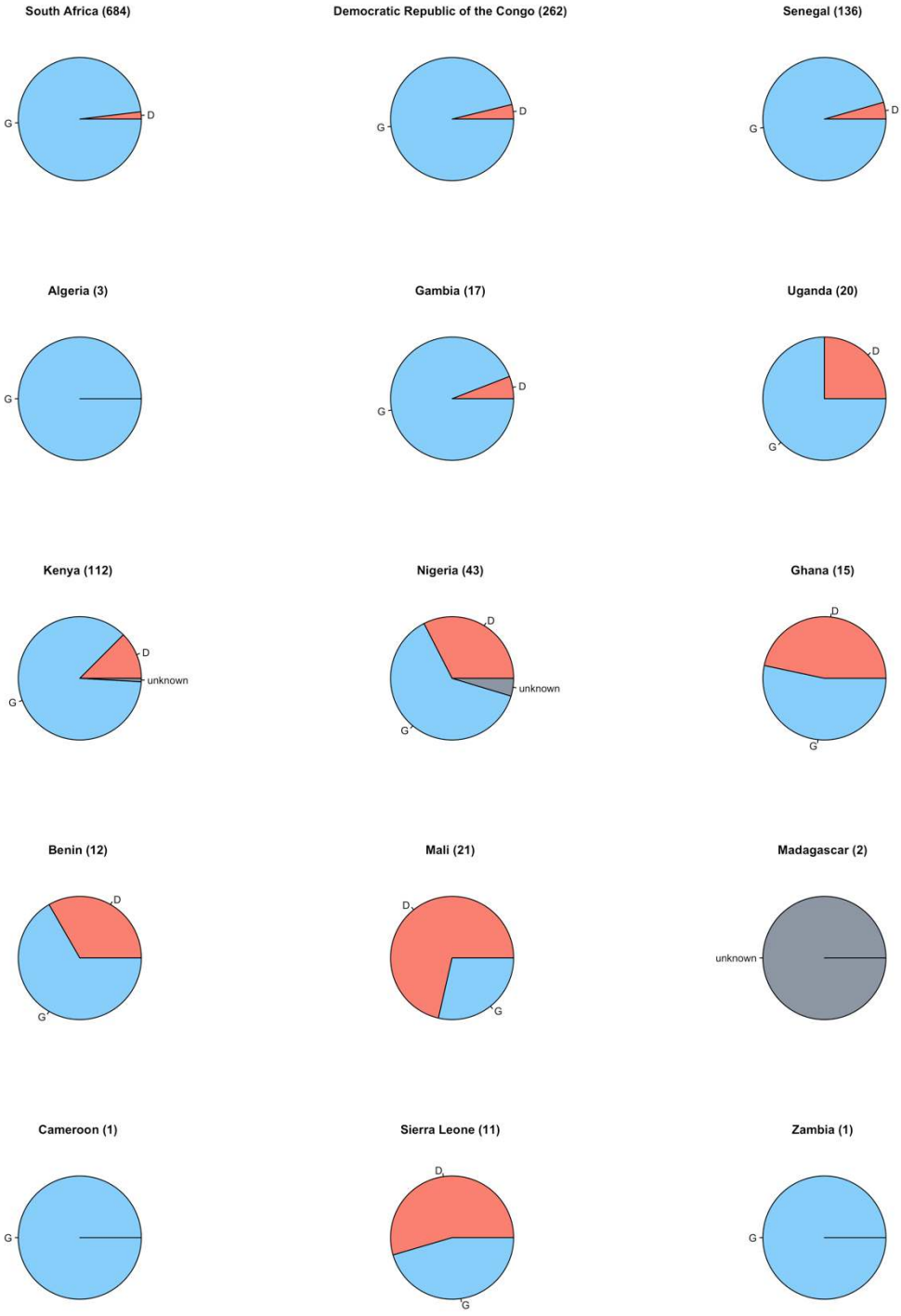


Figure 4 Distribution of D614G mutation per country. Lineage diversities are differentiated by red (S^{D614}) and blue (S^{G614}) categories; Number of sequences with such information are shown beside country names. Unknown means spike protein sequences have gaps at this position.

4 Phylogenetic analysis

The latest spatiotemporal phylogenetic analysis of Africa sequences is available on Nextstrain (https://doi.org/10.1093/bioinformatics/bty407), by using subsampled sequences from African countries and global sequences (between Dec 2019 and Aug 2020). Partial sequences, as well as sequences with incomplete sampling date or >5% ambiguous bases were removed. Over-representative sequences were down-sampled in both African countries (n= 965) and other continents (n=1147).

Internal node colours indicate ancestral state and shifts are drawn as links between demes on the map. This is coupled with an interactive time slider to see how the pathogen has evolved and spread over the course of the epidemic. Multiple views into different facets (discrete traits) of the data are presented and remain in sync as one interacts with the data (Figure 5). Possible transmissions between countries are identified.

We observed multiple separate introductions into Africa from other continents, mainly from Europe, which is consistent with the finding that the majority of sequences belong to the large European lineage B.1. In addition, most of the introductions do not appear to have spread further, except those within B.1 sub-lineages found in South Africa, DRC, Senegal, and Kenya (Figure 6).

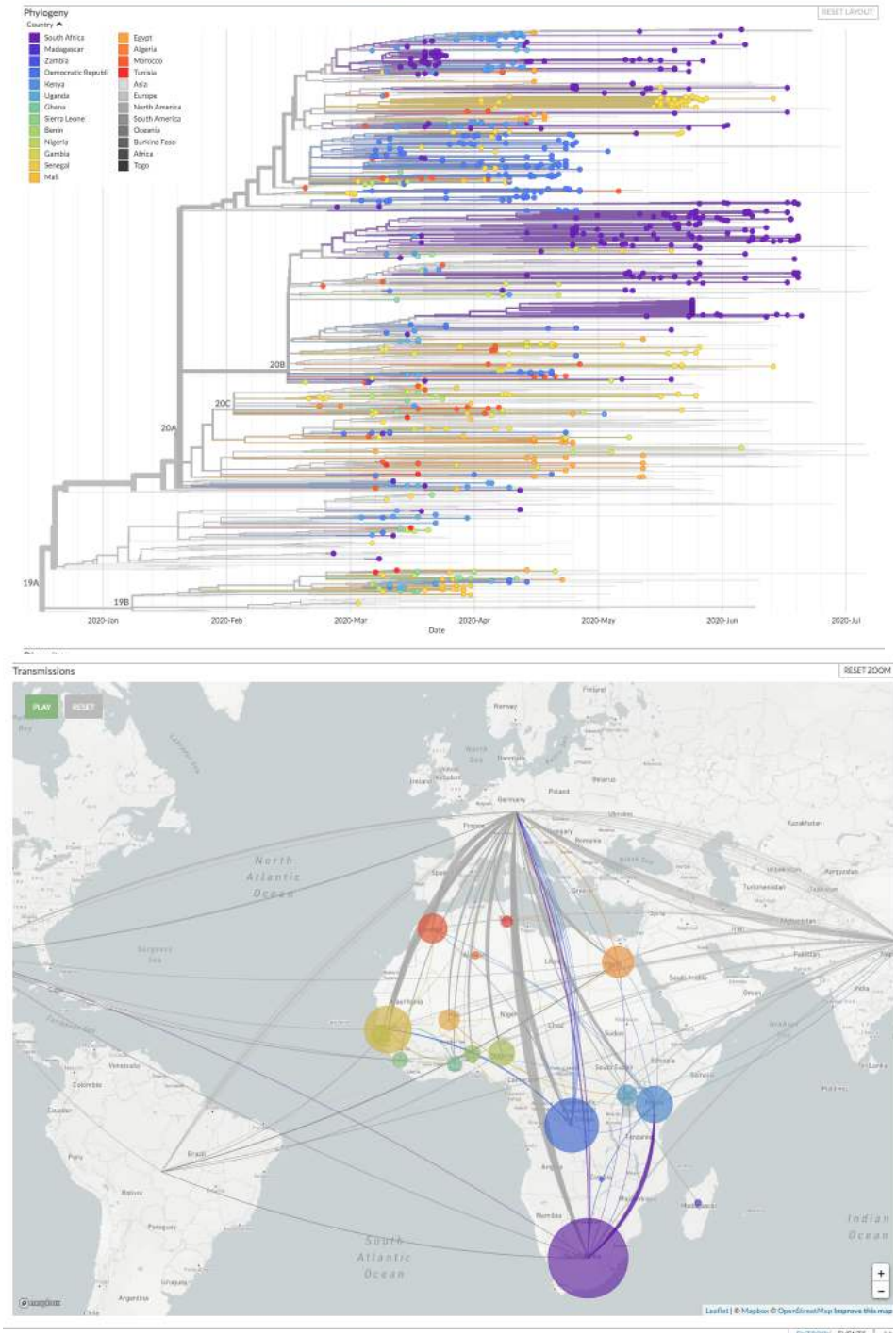


Figure 5 Genomic epidemiology of SARS-CoV-2 in Africa as of 17th August 2020. The colours of internal nodes of the time-scaled tree indicates ancestral locations inferred by the discrete trait method. Transmissions were drawn as links between locations on the map below. The node size indicates number of sequences on the tree; the colour of the line indicates the source country of transmission; the weight of the line indicates the number of possible transmissions. Live display at (https://nextstrain.org/ncov/africa?f_region=Africa).

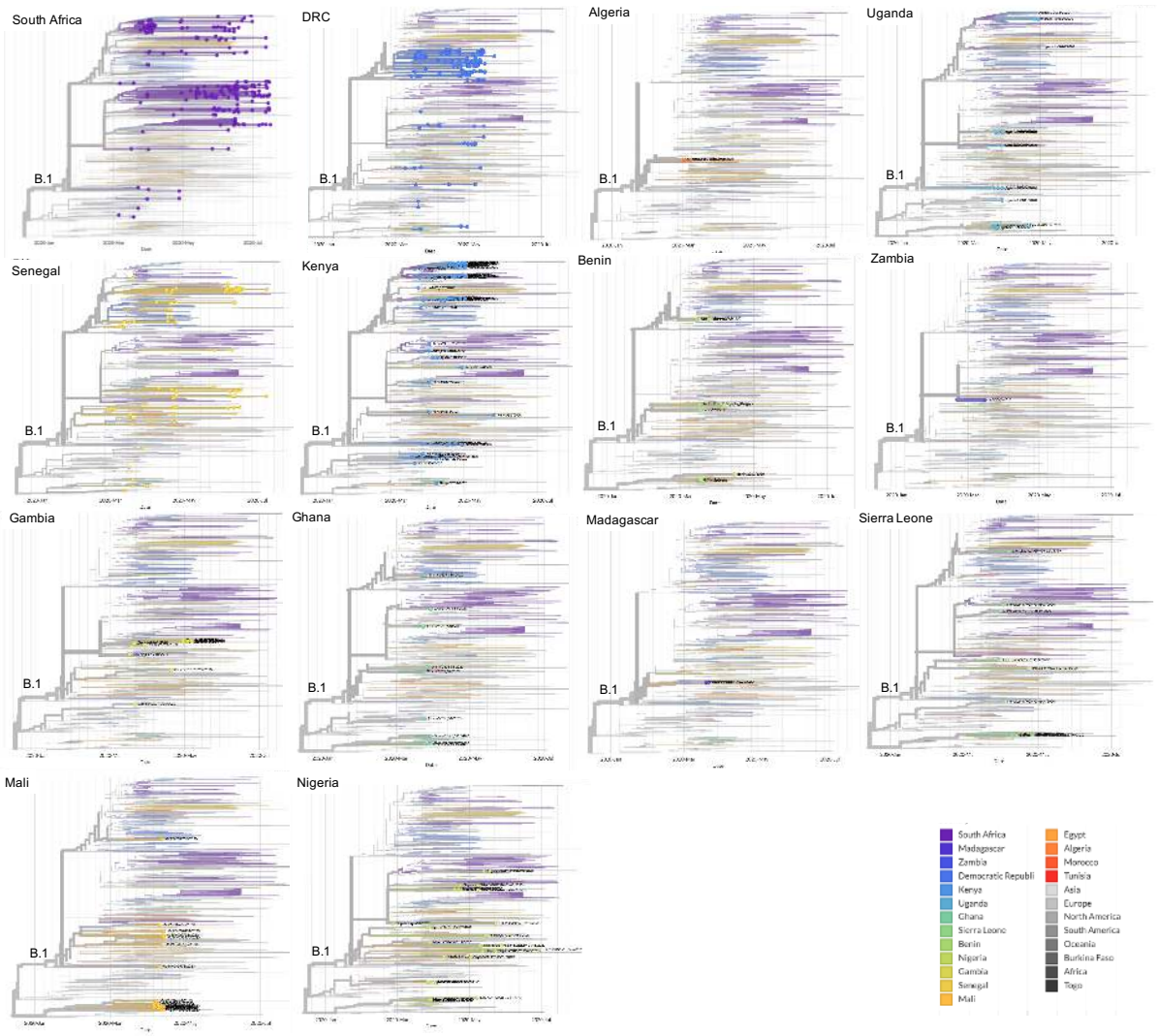


Figure 6 Genomes from individual African countries highlighted on approximate time-scaled global subsampled tree. Sequences from different African countries are highlighted on the same tree as Figure 5. The colours of internal nodes of the time-scaled tree indicates ancestral locations inferred by the discrete trait method. Lineage B.1 are indicated on the node. Sequence (n=1) from Cameroon was excluded from the phylogeny due to high proportion (14%) of ambitious bases.

5 Estimated number of introductions into African countries

The numbers of introductions to the fifteen WHO Africa countries were explored by using likelihood reconstruction methods on the phylogenetic tree in Mesquite (http://www.mesquiteproject.org). The numbers are tentative and are likely to change as more data from the same countries and/or from other countries become available. The two tables below show the counts of transitions from other continents to the fifteen WHO Africa countries (Table2), and the transitions from African countries (include those not within the WHO Africa region) to the fifteen WHO Africa countries (Table 3). Overall, we estimated approximately 123 introductions from other continents into Africa, 72% of them were from Europe. We currently estimate that there have been approximately 14 introductions between African countries.

Table 2 Introductions from other continents

FROM/TO	South Africa	DRC	Senegal	Algeria	Gambia	Uganda	Kenya	Nigeria	Ghana	Benin	Mali	Madagascar	Sierra Leone	Zambia	Cameroon	TOTAL
Europe	21	20	16	1	3	3	6	6	1	3	4	1	2	1	1	89
Asia	5	2	1	0	1	2	5	2	1	1	2	0	1	0	0	23
South America	4	2	0	0	0	0	1	0	0	0	0	0	0	0	0	7
Oceania	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
North America	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	4

Table 3 Introductions from African countries

FROM/TO	South Africa	DRC	Senegal	Algeria	Gambia	Uganda	Kenya	Nigeria	Ghana	Benin	Mali	Madagascar	Sierra Leone	Zambia	Cameroon	TOTAL
South Africa	-	0	0	0	0	1	3	0	0	0	0	0	0	0	0	4
DRC	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Senegal	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0
Algeria	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0
Gambia	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0
Uganda	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0
Kenya	1	0	0	0	0	1	-	0	0	0	1	0	0	0	0	3
Nigeria	0	0	0	0	0	0	1	-	0	0	0	0	0	0	0	1
Ghana	0	0	2	0	0	0	0	0	-	0	0	0	0	0	0	2
Benin	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0
Mali	0	0	0	0	0	0	1	0	0	0	-	0	0	0	0	1
Madagascar	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0
Sierra Leone	0	0	0	0	0	0	1	0	0	0	0	0	-	0	0	1
Zambia	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0
Cameroon	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0
Egypt	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Morocco	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Tunisia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix

Distribution of gender

In six countries, a higher proportion of sequences were collected from males then females. No gender data is available from Uganda, Sierra Leone, and Zambia.

	Algeria	Benin	Cameroon	DRC	Gambia	Ghana	Kenya	Madagascar	Mali	Nigeria	Senegal	Sierra Leone	South Africa	Uganda	Zambia
FeMale	1	6	0	98	4	5	58	2	6	9	54	0	430	0	0
Male	2	6	1	164	13	10	51	0	12	20	82	0	187	0	0
Unknown	0	0	0	0	0	0	3	0	3	14	0	11	67	20	1

Distribution of age

The majority of sequences were collected from patients in the 19 to 65 age group. DRC, Senegal and South Africa are the only countries that have sequences available that are isolated from individuals under 5 years of age. There are relatively more sequences isolated from elderly patients (over 65) than patients age under 18. No age data is available for sequences isolated in Kenya, Uganda, and Zambia.

	Algeria	Benin	Cameroon	DRC	Gambia	Ghana	Kenya	Madagascar	Mali	Nigeria	Senegal	Sierra Leone	South Africa	Uganda	Zambia
<5	0	0	0	1	0	0	0	0	0	0	5	0	10	0	0
5 TO 18	0	0	0	14	2	0	0	0	1	3	6	0	38	0	0
19 TO 65	2	12	1	218	12	10	0	1	12	20	107	0	560	0	0
>65	1	0	0	21	3	1	0	1	4	1	17	0	42	0	0
Unknown	0	0	0	8	0	4	112	0	4	19	1	11	34	20	1

Distribution of patient status

There is a lack of data regarding the status of the patients the sequences are sourced from. The majority of the sequences fall into the “Unknown” category.

	Algeria	Benin	Cameroon	DRC	Gambia	Ghana	Kenya	Madagascar	Mali	Nigeria	Senegal	Sierra Leone	South Africa	Uganda	Zambia
Alive	0	0	1	0	0	0	0	0	0	0	2	0	399	0	0
Deceased	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Hospitalized	3	0	0	7	0	0	0	0	0	8	0	0	0	0	0
Not Hospitalized	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0
Recovered	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Released	0	0	0	0	11	0	0	2	0	0	134	0	0	0	0
Stable in quarantine	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Unknown	0	12	0	254	4	15	112	0	21	22	0	11	285	20	1