

SARS-CoV-2 genomes report for WHO Africa Region

06/07/2020

Produced by the TIBA COVID-19 Pandemic Response Unit



Table of Contents

Summary	3
1 Sequence data.....	4
2 Lineage diversity.....	5
3 D614G mutation.....	7
4 Phylogenetic analysis.....	8
4.1. Approximate time-scaled tree of Africa sequences on global backbones	8
4.2 Nextstrain Analysis.....	11
5 Estimated number of introductions into African countries.....	14
Appendix.....	15
Distribution of gender.....	15
Distribution of age	15
Distribution of patient status.....	15

Summary

- SARS-CoV-2 genomes from nine countries in WHO Africa Region (out of 47) are available. These represent ~1% of publicly available sequences globally.
- The majority of genomes are from South Africa (34%), Democratic Republic of the Congo (DRC) (27%) and Kenya (22%).
- 20 different lineages have been identified. 86% of all African SARS-CoV-2 genomes are assigned to the B.1 lineage. B.1 is a large lineage that emerged from the epidemic in Italy.
- There have been multiple separate introductions (currently estimated as at least 82 across the nine countries) into Africa from other continents. Of these, about 66% were from Europe.
- Most of the introductions do not appear to have spread between countries in Africa. However, 11 introductions from within Africa were tentatively identified.
- 83% 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 537 sequences from nine WHO Africa countries (out of 47) were downloaded from GISAID on 25 June 2020 (Figure 1). These represent ~1% of publicly available global sequences. The majority are from South Africa, Democratic Republic of the Congo (DRC) and Kenya, representing 34%, 27% and 22% of all Africa sequences, respectively. 520 full-length sequences with complete collection date information were further investigated using phylogenetic analysis. The earliest sequence collected in Africa was from Nigeria on 27th February 2020 and the most recent one is from South Africa on 4th June 2020.

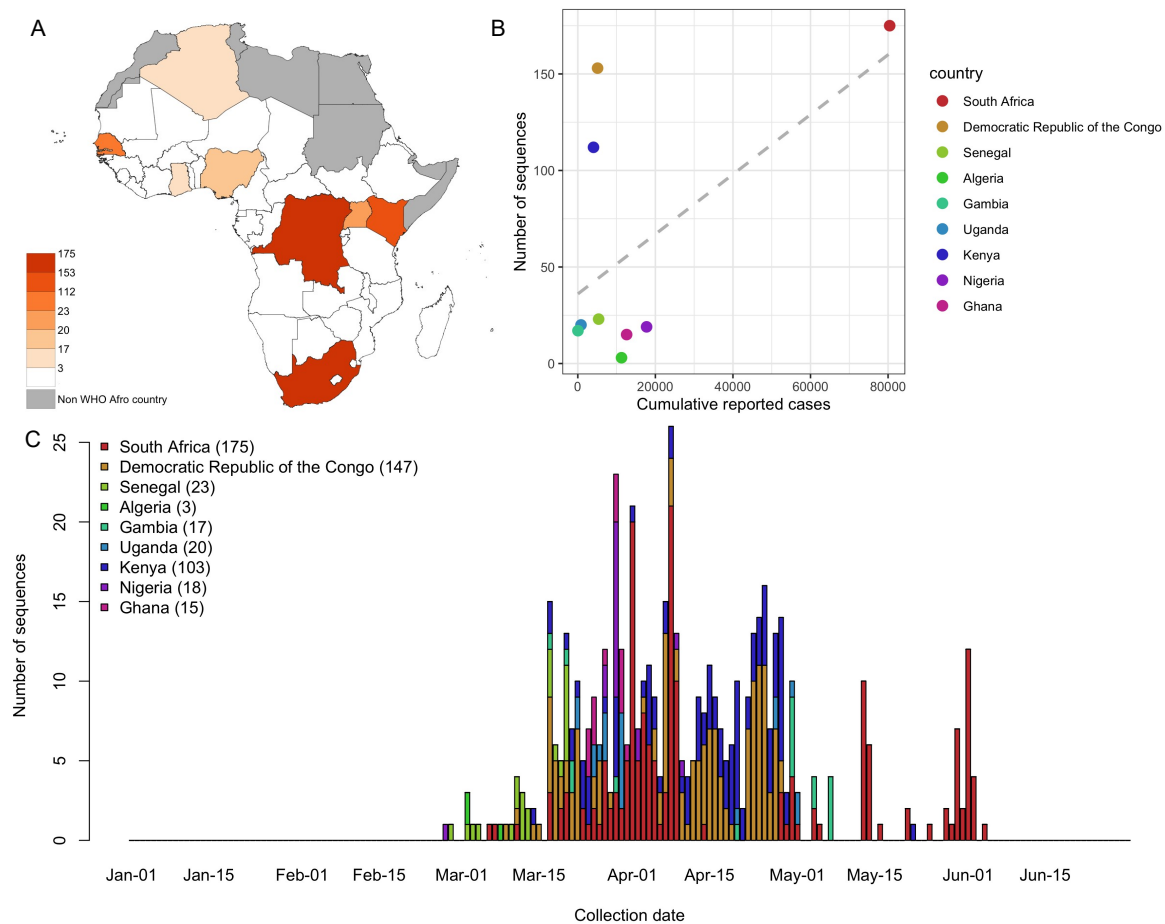


Figure 1 Number of sequences in WHO Africa countries. (A) SARS-CoV-2 genomes available on GISAID by country. (B) Comparison of cumulative reported cases and number of sequences available. (C) SARS-CoV-2 genomes available on GISAID by collection date. 6 genomes from the Democratic Republic of the Congo, 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 was summarized below; the distribution of patient gender and age as well as patient status are summarized in 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 2 June 2020, a total of 102 lineages have been identified globally. There is no significant difference between the properties of these lineages in terms of transmissibility and virulence at the present time.

Within Africa, 537 sequences were assigned to 20 lineages, including 18 sub-lineages with bootstrap support in the range of 52 and 100 (Table 1 and Figure 2). 5.2% of sequences from 6 countries (except South Africa, Algeria and Gambia) were assigned to lineage A; the remaining sequences were assigned to lineage B, and the majority of them (n=462) belong to sub-lineage B.1 (which represents a large European lineage that corresponds to the Italian outbreak) and its descendent sub-lineages. Sequences belonging to sub-lineage B.4 were identified in DRC, Kenya and Uganda, which originated in Iran.

Table 1 Number of sequences per lineage per country

LINAGE	ALGERIA	DRC	GAMBIA	GHANA	KENYA	NIGERIA	SENEGAL	SOUTH AFRICA	UGANDA	
A		0	5	0	5	2	9	1	0	4
A.2		0	0	0	0	0	0	1	0	0
A.P7		0	0	0	0	0	0	1	0	0
B		0	0	0	0	4	0	0	9	3
B.1		3	126	1	5	80	3	19	87	5
B.1.1		0	1	14	2	13	2	0	44	1
B.1.1.1		0	13	0	0	1	0	0	4	6
B.1.1.10		0	0	0	0	0	0	0	1	0
B.1.1.8		0	0	1	0	0	0	0	0	0
B.1.3		0	0	0	1	0	0	0	0	0
B.1.5		0	0	0	0	1	0	0	8	0
B.1.6		0	4	0	0	0	0	0	0	0
B.1.8		0	0	0	0	0	0	0	16	0
B.2		0	1	0	0	1	0	0	1	0
B.2.1		0	2	0	1	2	4	0	1	0
B.2.2		0	0	0	1	0	0	0	1	0
B.3		0	0	0	0	0	0	1	0	0
B.4		0	1	0	0	7	0	0	0	1
B.6		0	0	1	0	1	0	0	2	0
B.9		0	0	0	0	0	0	0	1	0
NONE		0	0	0	0	0	1*	0	0	0

*sequence failed to be assigned into a lineage due to short length (sequence length=1152 nt).

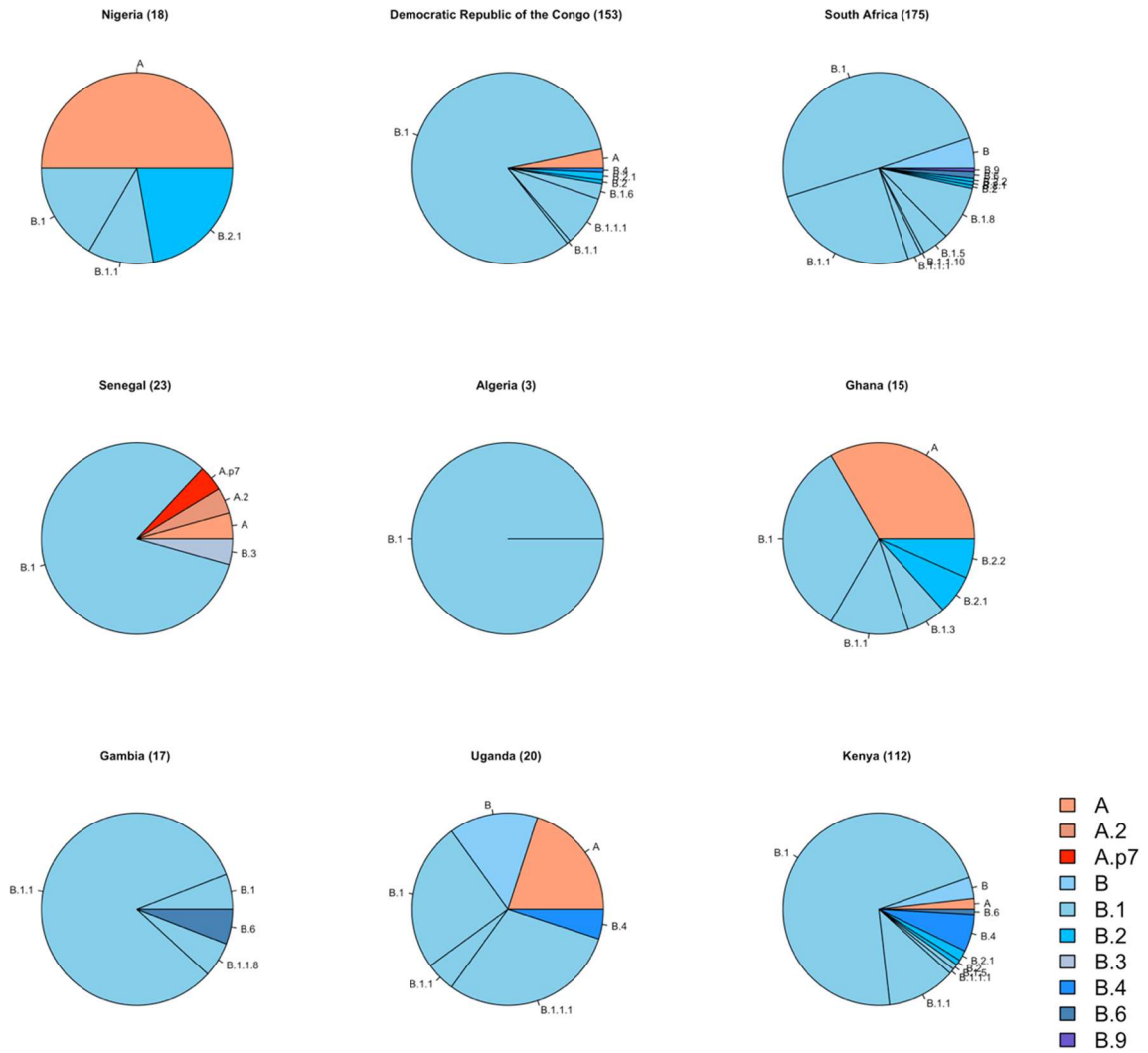


Figure 2 Lineage distribution per country. Lineage diversities are differentiated by red (lineage A and sub-lineages) and blue (lineage B and sub-lineages) categories; Number of sequences are shown beside country names.

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, 83% of them have such mutation. The distribution of genomes with such mutation in individual countries is summarized in Figure 3.

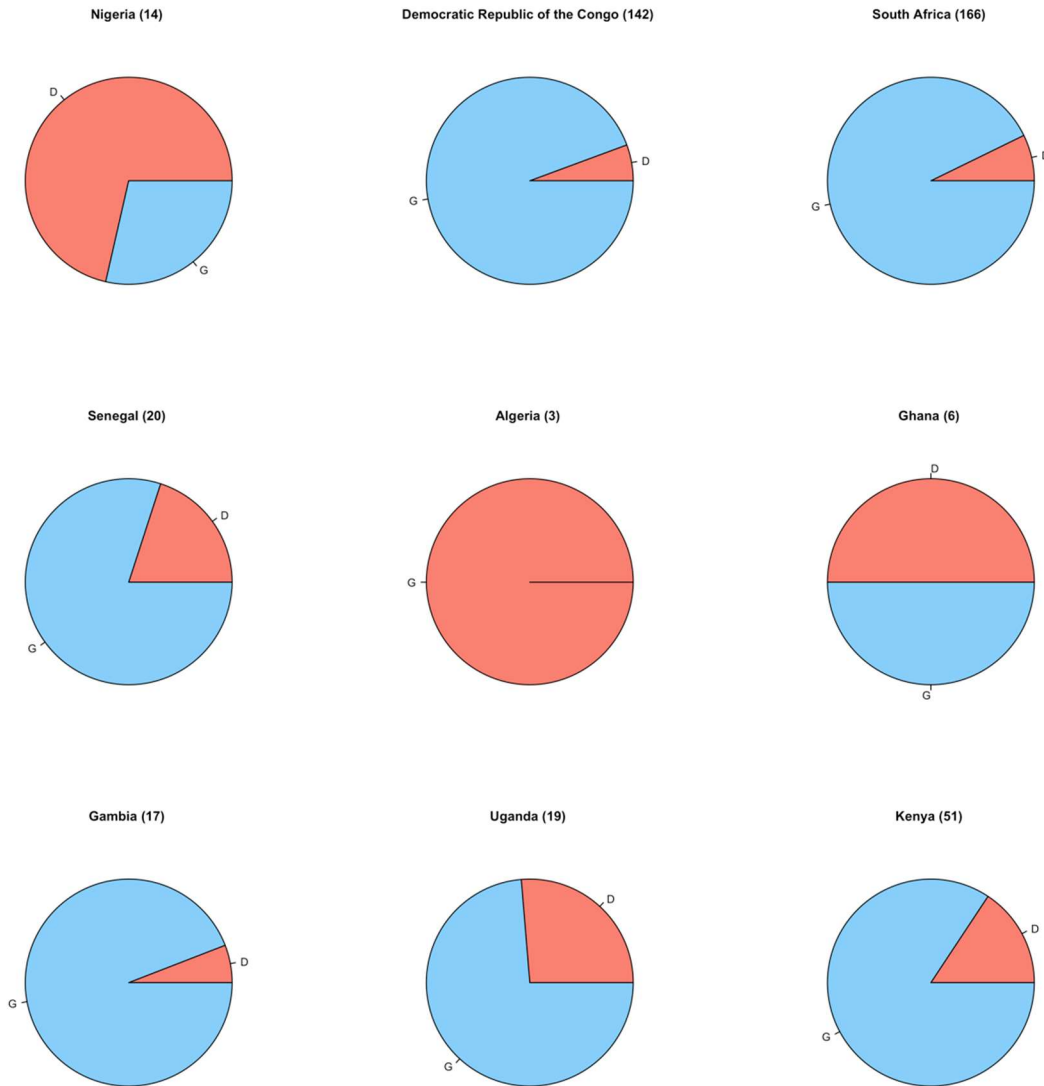


Figure 3 Distribution of D614G mutation per country. Lineage diversities are differentiated by red (S^{D614} and blue (S^{G614}) categories; Number of sequences with such information are shown beside country names.

4 Phylogenetic analysis

4.1. Approximate time-scaled tree of Africa sequences on global backbones

All available full-length SARS-CoV-2 genomes were retrieved from GISAID. Sequences from non-human hosts, as well as sequences with incomplete sampling date or >5% ambiguous bases were removed. Sequences were then down-sampled to 2388 ‘backbone sequences’ across the globe by keeping unique representative sequences per country, state and epiweek.

422 full length Africa sequences with complete sampling date were added to the global subsampled tree which was reconstructed by Neighbour joining method. The phylogeny was approximately time scaled using the R package treedater (Volz) with a strict clock setting and rooted on the oldest sequence, with symmetric discrete trait to inter transmission between continents.

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 (Figure 4). This trend can be observed within each individual African country except Algeria which have three sequences available. In addition, most of the introductions do not appear to have spread further, except those in DRC, Kenya and South Africa (Figure 5).

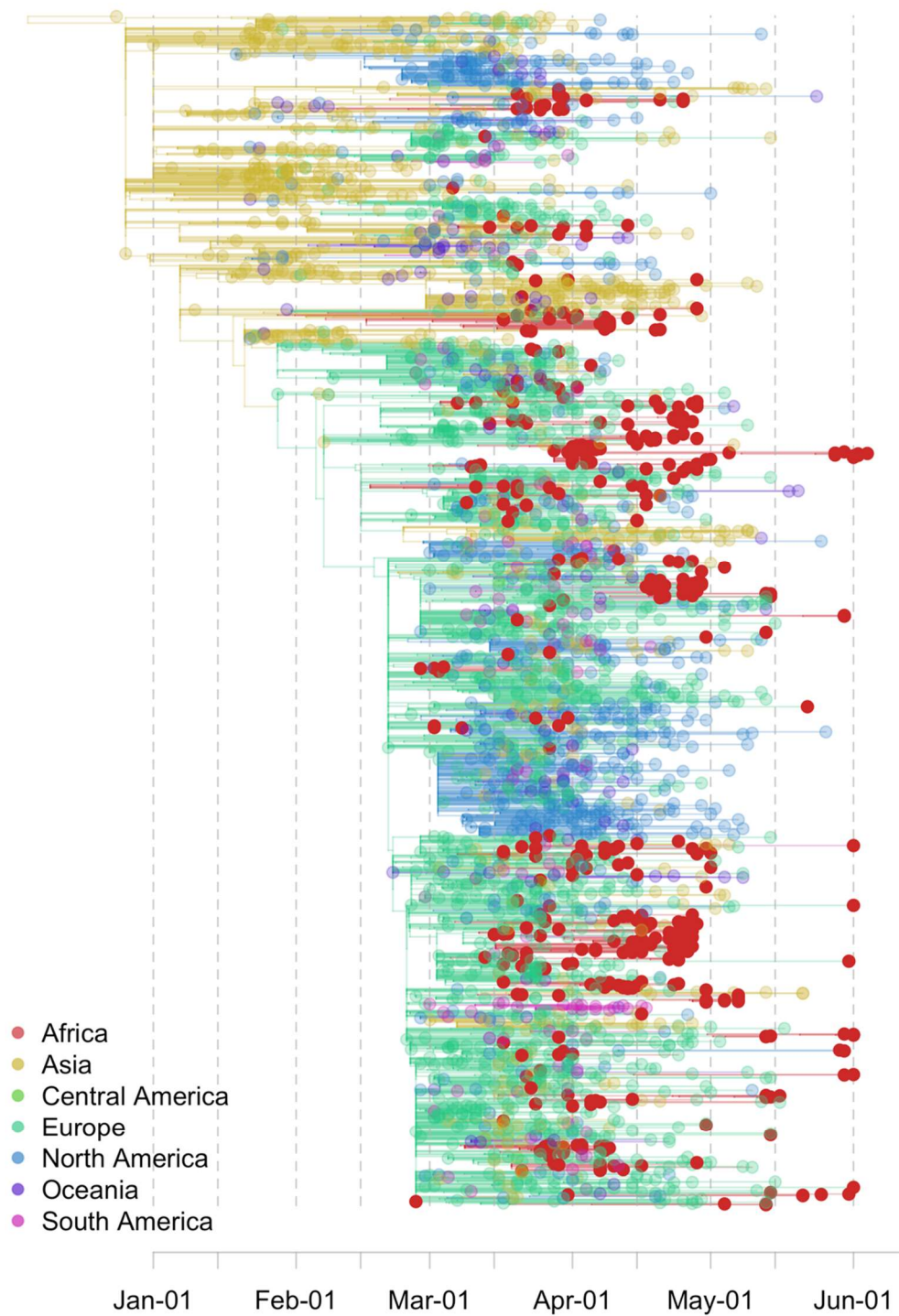


Figure 4 African genomes (n=382) highlighted on approximate time-scaled global subsampled tree (n=2388). Tips are coloured by continents; Nodes and Branches are coloured by indicated ancestral states using discrete trait method; All sequences from Africa are highlighted in red.

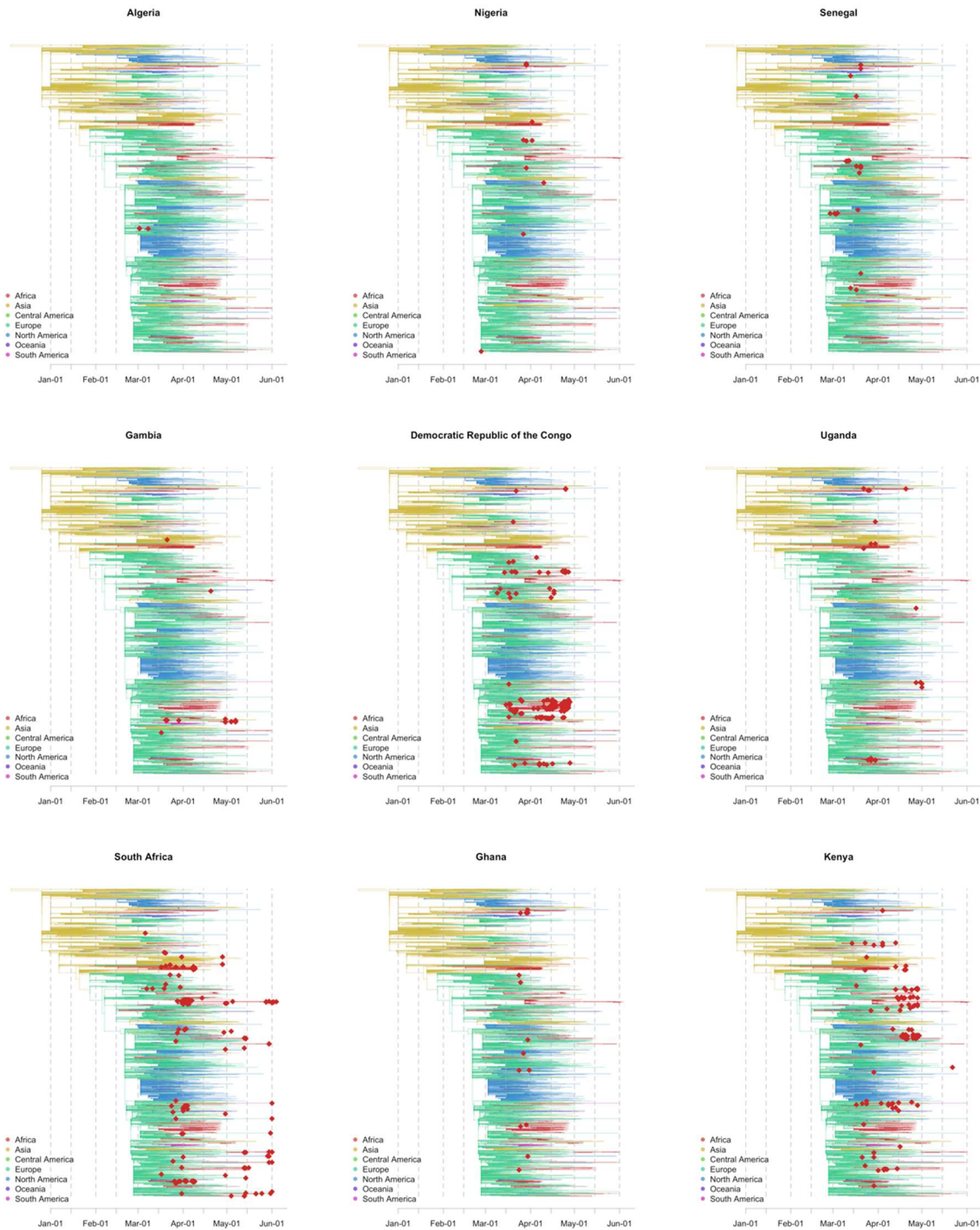


Figure 5 Genomes from 9 individual African countries highlighted on approximate time-scaled global subsampled tree. Sequences from each individual African country are highlighted on tips (in red) on the same tree as Figure 4. Nodes and Branches are coloured by indicated ancestral states using discrete trait method.

4.2 Nextstrain Analysis

The latest spatiotemporal phylogenetic analysis of Africa sequences is available on Nextstrain (<https://doi.org/10.1093/bioinformatics/bty407>), by using sequences from African countries (n= 422 between Feb 2020 and Jun 2020) and subsampled global sequences. 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 6). Possible transmissions between Africa countries were identified. There are multiple transmission events (belonging to lineage B.1) from Kenya to South Africa in March 2020 (Figure 7).

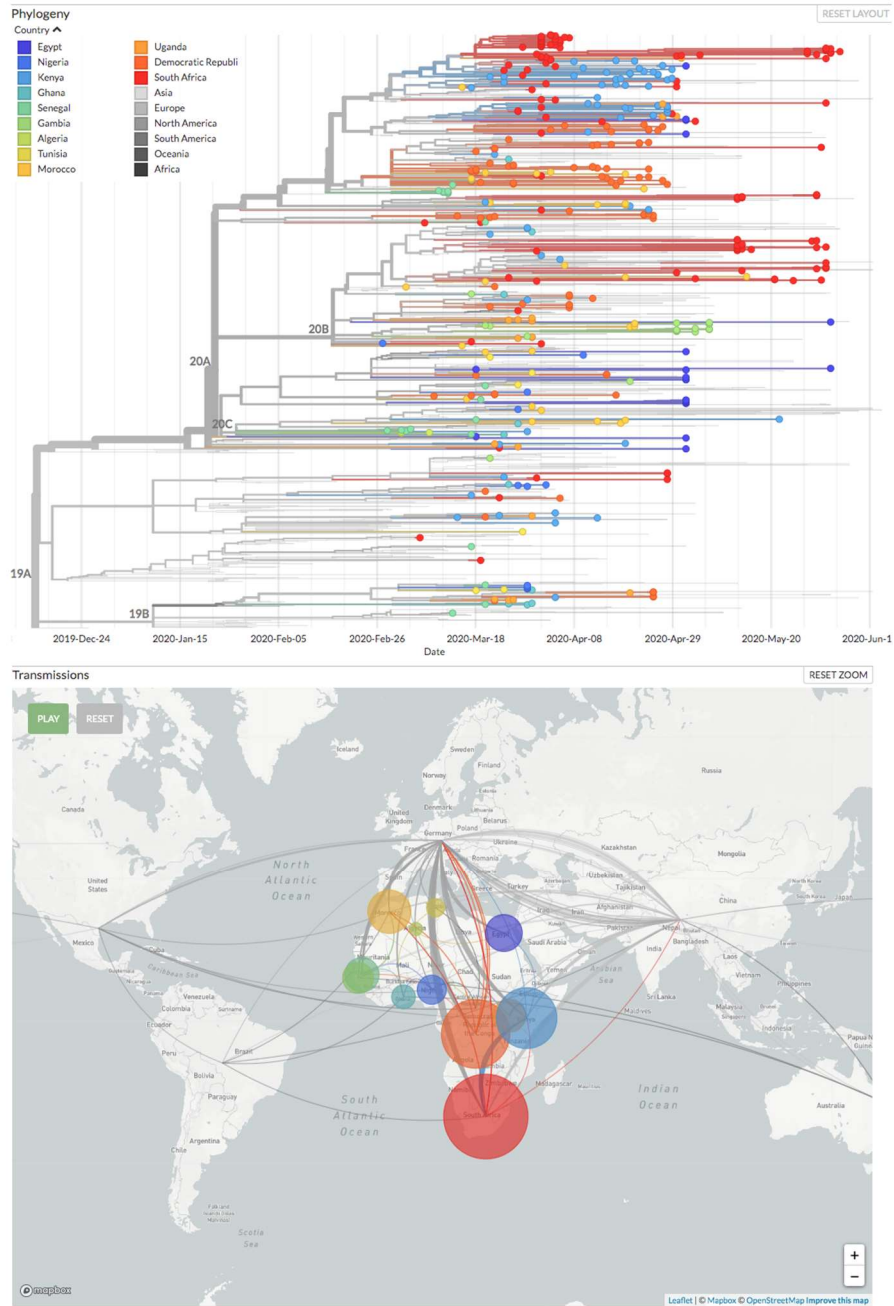


Figure 6 Genomic epidemiology of SARS-CoV-2 in Africa as of June 2020. The colours of internal nodes of the time-scaled tree indicates ancestral locations inferred by the discrete trait method. Transmissions were drawn as links between locations on the map below. The node size indicates number of sequences 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).

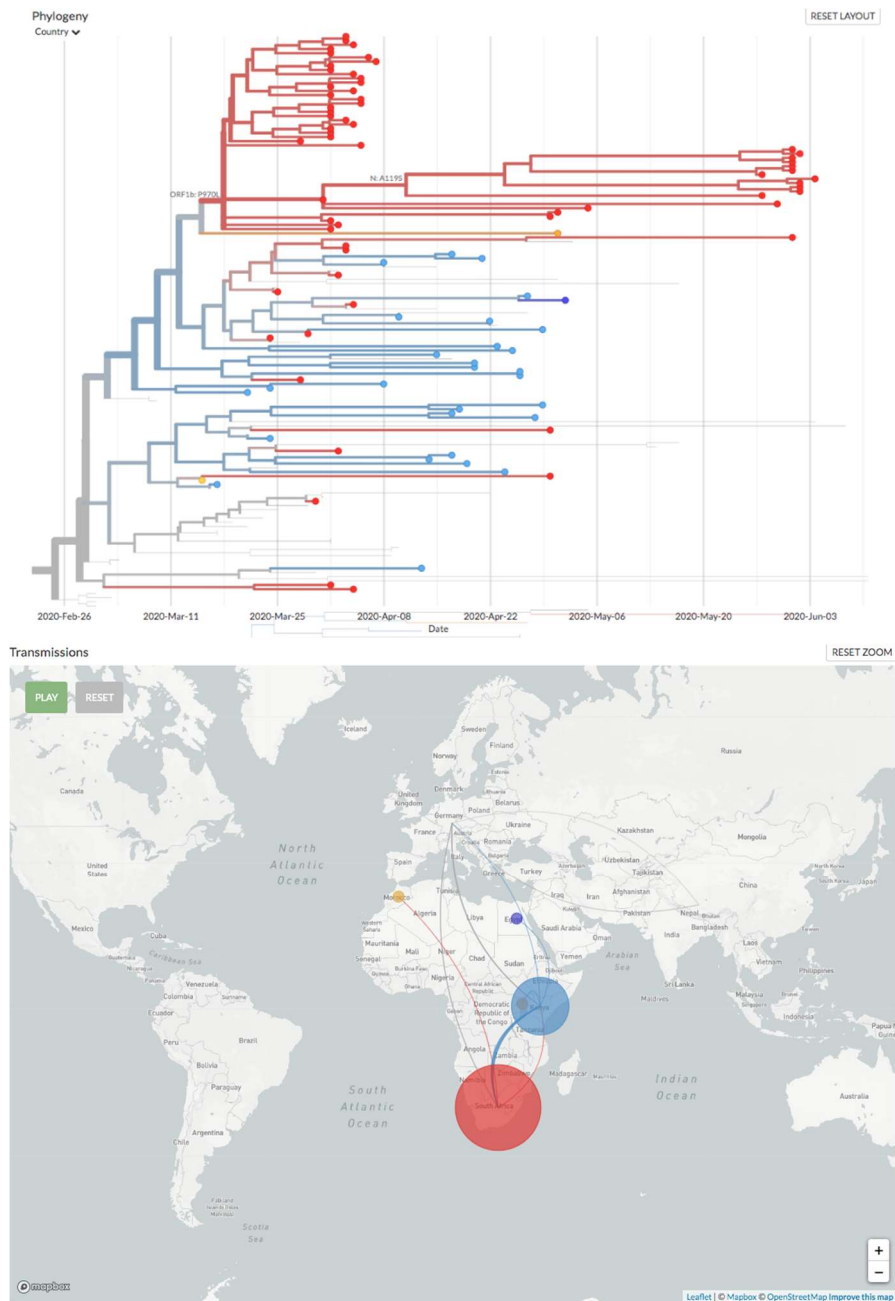


Figure 7 Highlighted transmissions between Kenya and South Africa in B.1 lineage. The transmission events between Kenya (in blue) and South Africa (in red) were reconstructed on partial B.1 lineage extracted from the phylogeny in Figure 7. On the map, the node size indicates number of sequences on the partial lineage; the colour of the line indicates the source country of transmission; the weight of the line indicates the number of possible transmissions. The thick blue line highlights transmission from Kenya to South Africa.

5 Estimated number of introductions into African countries

The numbers of introductions to African 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 Africa (Table 2), and the transitions to or from African countries (Table 3). Overall, we estimated approximately 82 introductions from other continents into Africa, 66% of them were from Europe. We currently estimate that there have been approximately 11 introductions between African countries.

Table 2 Introductions from other continents

FROM/TO	TOTAL	SOUTH AFRICA	DRC	SENEGAL	ALGERIA	GAMBIA	UGANDA	KENYA	NIGERIA	GHANA
EUROPE	54	16	8	7	2	5	4	8	2	2
ASIA	23	4	4	0	0	2	2	9	2	0
OCEANIA	1	0	0	0	0	0	0	0	0	1
NORTH AMERICA	2	0	1	0	0	0	0	1	0	0
SOUTH AMERICA	2	1	1	0	0	0	0	0	0	0
TOTAL OUTSIDE AFRICA	82	21	14	7	2	7	6	18	4	3

Table 3 Introductions from African countries

FROM/TO	TOTAL	SOUTH AFRICA	DRC	SENEGAL	ALGERIA	GAMBIA	UGANDA	KENYA	NIGERIA	GHANA
SOUTH AFRICA	2	-	0	0	0	0	1	1	0	0
DRC	1	0	-	1	0	0	0	0	0	0
SENEGAL	0	0	0	-	0	0	0	0	0	0
ALGERIA	1	0	0	0	-	0	0	1	0	1
GAMBIA	0	0	0	0	0	-	0	0	0	0
UGANDA	0	0	0	0	0	0	-	0	0	0
KENYA	5	4	0	0	0	0	1	-	0	0
NIGERIA	1	0	0	0	0	0	0	1	-	0
GHANA	1	0	0	1	0	0	0	0	0	-
TOTAL WITHIN AFRICA	11	4	0	2	0	0	2	3	0	0

Appendix

Distribution of gender

In six countries, a higher proportion of sequences were collected from males than females. South Africa is the only country to go against this trend. No gender data is available from Uganda.

GENDER	ALGERIA	DRC	GAMBIA	GHANA	KENYA	NIGERIA	SENEGAL	SOUTH AFRICA	UGANDA
FEMALE	1	65	4	5	58	4	8	87	0
MALE	2	88	13	10	51	12	15	44	0
UNKNOWN	0	0	0	0	3	3	0	44	20

Distribution of age

The majority of sequences were collected from patients in the 19 to 65 age group. DRC and Senegal are the only countries that have sequences available that are isolated from individuals under 5 years of age. From all the countries, there are relatively few sequences isolated from elderly patients (over 65). No age data is available for sequences isolated in Kenya or Uganda.

AGE (Y)	ALGERIA	DRC	GAMBIA	GHANA	KENYA	NIGERIA	SENEGAL	SOUTH AFRICA	UGANDA
<5	0	1	0	0	0	0	1	0	0
5 TO 18	0	11	2	0	0	2	1	6	0
19 TO 65	2	126	12	10	0	14	17	135	0
>65	1	11	3	1	0	0	4	17	0
UNKNOWN	0	4	0	4	112	3	0	17	20

Distribution of patient status

There is a lack of data regarding the status of the patients the sequences are sourced from. Senegal and Algeria appear to be the only countries recording this consistently. Some data is available from DRC, Gambia and Nigeria. The majority of the sequences fall into the “Unknown” category.

PATIENT STATUS	ALGERIA	DRC	GAMBIA	GHANA	KENYA	NIGERIA	SENEGAL	SOUTH AFRICA	UGANDA
HOSPITALIZED	3	7	0	0	0	0	2	0	0
NOT HOSPITALIZED	0	0	0	0	0	0	13	0	0
ISOLATION	0	0	0	0	0	0	1	0	0
STABLE IN QUARANTINE	0	1	0	0	0	0	0	0	0
STABLE	0	0	0	0	0	0	0	0	0
DECEASED	0	0	1	0	0	0	0	0	0
ALIVE	0	0	0	0	0	0	0	0	5
LIVE	0	0	0	0	0	0	0	2	4
RECOVERED	0	0	1	0	0	0	0	0	0
RELEASED	0	0	11	0	0	0	0	21	0
UNKNOWN	0	145	4	16	112	3	0	166	20