A genomics dissection of Kenya's COVID-19 waves: temporal lineage replacements and dominance of imported variants of concern

Gathii Kimita
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya
https://orcid.org/0000-0003-4555-4480

Josphat Nyataya
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya

Esther Omuseni
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya

Faith Sigei
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya
https://orcid.org/0000-0001-7582-3296

Allan Lemtudo
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya
https://orcid.org/0000-0002-8204-7560

Eric Muthanje
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya

Brian Andika
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya

Rehema Liyai
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya

Rachel Githii
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya

Clement Masakhwe
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya
https://orcid.org/0000-0003-3379-7346

Stephen Ochola
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya

George Awinda
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya

Carol Kifude
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya

Beth Mutai

Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya
https://orcid.org/0000-0002-7396-3903

John Waitumbi (john.waitumbi@usamru-k.org)
Kenya Medical Research Institute/US Army Medical Research Directorate-Africa, Kenya
https://orcid.org/0000-0003-2540-4397

Article

Keywords: wave, lineage, kenya, lineages, delta, variants

Posted Date: October 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-942627/v1

Version of Record: A version of this preprint was published at Communications Medicine on August 17th, 2022. See the published version at https://doi.org/10.1038/s43856-022-00167-8.

Abstract

Kenya's COVID-19 epidemic was slow to peak. It was seeded early in March 2020, and did not peak until late-July 2020 (wave 1), mid-November 2020 (wave 2) and late-March 2021 (wave 3). Here we present SARS-CoV-2 lineages associated with the three COVID-19 waves through analysis of 483 genomes, which included 167 Alpha (B.1.1.7), 57 Delta (B.1.617.2) and 12 Beta (B.1.351) variants of concerns (VOC) that dominated the third wave. In total, 35 lineages were identified. The early European lineages B.1 and B.1.1 were the first to be seeded in Kenya. The B.1 lineage continued to expand and remained the most dominant lineage accounting for 55.8% and 56.3% in waves 1 and 2 respectively. The alpha (B.1.1.7), delta (B.1.167.2) and beta (B.1.351) VOCs dominated in wave 3 at 59.0%, 20.1% and 4.2% respectively. Eventually, the delta variant took over at the tail end of wave 3 and at the time of going to press, it had become the major lineage in the whole country. Phylogenetic analysis suggested multiple introductions of variants from outside Kenya especially during the first and third wave. Phylogeny also highlighted local lineage diversification as local transmission events supervened. The data highlights the importance of genome surveillance in determining circulating variants to aid in public health interventions.

Background

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has literally been the 2020/21 blockbuster virus. It reached every part of the globe in less than 9 months, and at the time of writing this report, it had infected 199+ million and killed 4.2+ million people globally. The virus is the etiological agent of coronavirus disease 2019 (COVID-19), a mysterious severe respiratory illness that first appeared in Wuhan, Hubei Province, China, in December 2019 (Wu et al., 2020). SARS-CoV-2 is believed to have been transmitted to humans as a result of a zoonotic spillover event believed to have been from a progenitor bat coronavirus and civet cats as intermediates (Zhou et al., 2020). To date, COVID-19 has overshadowed all other human health calamities, ravaged global economies and disrupted human social interactions.

In Kenya, the first confirmed case of COVID-19 was on 13th March 2020 from a Kenyan citizen returning home from the USA via London, UK (Ministry of Health, 2020). Within two weeks, 31 cases traceable to the index case and other international travelers were identified. As a result, the citizenry panicked, and the government instituted a series of countermeasures that included border closures, mandatory quarantine on returning travelers, night curfews, ban on gatherings, and mandatory mask use while in public spaces. While these measures slowed the spread of the disease, the virus still managed to infiltrate into the community, and new infections were associated with local transmission events. COVID-19 cases continued to increase, reaching a peak of approximately 1000 cases by late July 2020 and then declined steadily to very low numbers by mid-September 2020, marking the "end" of the first wave.

Buoyed by the reduction in COVID-19 numbers, most of the COVID-19 restrictions were lifted to ease pressure on a slumping economy. The next one and half months were characterized by low infections, but this short-lived lull was interrupted by a spike in infection rates that rose steadily, peaking at 1554

cases by mid-November 2020, triggering another round of lockdowns. The reopening of schools was postponed, and public gatherings, especially political rallies, stood banned. These measures progressively reduced infections, and by January 2021, the second wave burned out, after which a third wave of COVID-19 ensued (Ritchie et al., 2020).

Towards the end of 2020, SARS-CoV-2 variants of concern (VOCs) and variants of interest (VOIs) began to emerge independently from different parts of the globe. These variants were associated with increased transmissibility, virulence, clinical presentation and decreased effectiveness to diagnostics and therapeutics, including vaccines (WHO, 2021). Kenya has so far detected 3/4 currently classified VOCs, i.e. the Alpha variant (B.1.1.7/ 20I/S:501Y. V1), Beta (B.1.351, 20H/S:501Y. V2) and Delta (B.1.617.2, 21A/S:478K). The country has also recorded two variants of interest, i.e. Eta (B.1.525, 20A/S484K) and the A.23.1 lineage. VOCs have higher transmissibility than other lineages, as witnessed by the rapid transmission and global expansion of the alpha and delta variants. The initial fear that some of these mutants and/or other future variants could negatively impact vaccine efficacy and constitute postvaccination “antigenic escape” is already being witnessed with the Delta variant (Planas et al., 2021).

In this report, we use genomic surveillance to dissect the three COVID-19 waves that have occurred in Kenya since the beginning of the outbreak and provide data that show temporal dominance and eventual replacement of lineages by the more transmissible VOCs.

Methods

Ethics statement

This study was performed as part of public health surveillance approved by the Kenya government through the Ministry of Health (MOH) as part of response to the COVID-19 pandemic. The Scientific and Ethical Research Unit (SERU) of Kenya Medical Research Institute (KEMRI) approved a protocol to allow whole genome sequencing (SERU 4035).

Sample acquisition

Multiple laboratories, including the Basic Science Laboratory (BSL), were designated by the MOH as COVID-19 testing centers. BSL started supporting COVID-19 mass testing and whole genome sequencing in March 2020, and by May 2021, the laboratory had screened 45,689 respiratory samples for COVID-19 by RT-qPCR. Samples came from different parts of the country, and a few came from individuals coming from neighboring countries (Uganda, Congo). Nucleic acid isolation was performed using MagMAX Pathogen Kits with the KingFisher Flex particle purification system (Thermo Fisher Scientific, CA, USA). Of the 45,689 tested, 9.0% (n=4,109) were positive for SARS-CoV-2 at varying cycle thresholds (Ct). Of the 4,109 positive samples, 483 with Cts <33 were selected for whole genome sequencing.

Whole genome sequencing

In brief, cDNA was synthesized from RNA using random primers with the Superscript IV one step reverse transcriptase kit (Thermo Fisher Scientific, CA, USA). The cDNA was then used for tiled multiplex PCR using the ARTIC v3 primers as described in the associated protocol (Sevinsky et al., 2020). The amplicons were cleaned with AmpureXP beads (Beckman Coulter, USA) and then used to create sequence libraries using the NexteraXT (Illumina, USA) and Collibri ES (ThermoFisher Scientific, CA, USA) library preparation kits, as per the manufacturer's instructions. The libraries were assessed on D1000 HS screen tape on a Tape Station 4200 (Agilent, CA, USA) to determine their size distribution and concentration. A 12 pM library spiked with 10% Phix genome was then sequenced on the MiSeq benchtop sequencer (Illumina, CA, USA) using 600 V3 paired end chemistry.

Genome assembly, global data acquisition and quality filtering

Read demultiplexing was conducted onboard the MiSeq using the MiSeq reporter v2.6. The reads were quality filtered and assembled against SARS-CoV-2 Wuhan 1 as a reference (GenBank accession number: NC_045512) using the ngs mapper v 1.5 pipeline (Tyghe et al., 2016) and CLC genomics workbench v 8.5 (Qiagen, MD, USA). A minimum Phred base quality score of 30 and minimum depth of coverage of 5 were used to call the consensus sequence. The consensus sequences were further curated using Nextclade Web v1.5.1 (Aksamentov et al., 2020).

To compare the genome sequences of the current study to global isolates, SARS-CoV-2 genomes were sampled from the Global Initiative on Sharing All Influenza Data (GISAID) sequence database (Shu & McCauley, 2017). Only sequences flagged as "complete (>29,000 bp)", "high coverage only", and "low coverage excl" were downloaded and included sequences as of May 22nd 2021.

Lineage and clade assignment

Lineage assignment was performed on each consensus sequence using PANGOLIN v3 (Phylogenetic Assignment of named Global Outbreak LINeages) (Rambaut et al., 2020), which offers a hierarchical dynamic nomenclature describing a lineage as a cluster of sequences observed in a geographically distinct region with evidence of transmission in that region. Clades were assigned to each consensus sequence using Nextclade Web v 1.5.1. The Nextstrain clade system (Hadfield et al., 2018) uses a year-letter nomenclature on a clade exceeding 20% global representation and >2 positional differences from its parent clade while considering clade persistence with time as well as the extent of its geographical spread.

Phylogenetic analysis

Three phylogenies were constructed, one consisting of 113 context genomes sampled from around the globe and 328 samples from this work. The other tree was constructed to focus on the alpha variant and consisted of alpha variants from Kenya (n=151), the rest of Africa (n=872) and a sample of the earliest reported alpha variants (n=44), mostly from England. The last tree was constructed with a focus on the delta variant and consisted of delta variants from Kenya (n=33) and Africa (n=812) and a sample of the

earliest reported delta variants mostly from India. All trees were reconstructed with augur and visualized with auspice as implemented in the Nextstrain pipeline version 3.0.3 (Hadfield et al., 2018). Within Nextstrain, a random subsampling method was used to cap the maximum number of contextual sequences.

Data Availability

Assembled SARS-CoV-2 genomes in this study were uploaded to www.gisaid.org/ as FASTA files (gisaid epi isl numbers 2779282-2779550, 3031380-30314433)]

Results

Demographic data of subjects who contributed genome data are shown in Table 1. Of those with age data, most samples came from age groups 21-30 years (n=125) and 31-50 (n=178). 94 people did not indicate age. More sequences were from males (55.5%, n=268) than females (24.8%, n=120), notwithstanding 95 individuals (19.7%) who did not indicate gender.

Table 1. Demographic data for subjects who contributed genome sequences used in this study.

		n	%
Age distribution	<20 year	43	8.9
	21-30 yr	125	25.9
	31-50 yr	178	36.9
	51-60 yr	29	6.0
	>61 yr	14	2.9
	missing data	94	19.5
Median age	33		
Sex	Male	268	55.5
	Female	120	24.8
	?	95	19.7
Nationality	Kenyan	475	99.0
	Ugandan	6	1.2
	Congolese	2	0.4
Infection Wave	Wave 1	120	24.8
	Wave 2	80	16.6
	Wave 3	283	58.6

Expansion and displacement of SARS-CoV-2 lineages and eventual dominance of VOCs

A total of 483 viral genomes collected from persons who tested positive for SARS-CoV-2 between March 2020 and June 2021 were used to monitor the prevalence and evolution of SARS-CoV-2 lineages across the three COVID-19 waves (Figure 1). Each wave was preceded by low infection rates, probably as variants competed through narrow transmission bottlenecks that selected the fittest variants, some of them to eventually become the dominant variants in succeeding waves (Lythgoe et al., 2021). Nine Nextstrain clades were determined from the data, including 20I–Alpha V1 (34.2%), 20C (20.5%), 20A (12.6%), 21A-Delta (12%), 20B (7%), 19B (4.8%), 21D-Eta (4.3%), 20H-Beta, V2 (2.5%) and 20D (0.4%). 1.7% of the genomes could not be reliably grouped.

Thirty-five distinct Pangolin lineages were identified (Figure 1 and Supplementary Table 1), 15 of which were linked to wave 1 (collected from May 2020 to mid-September 2020), 11 to wave 2 (collected from late September 2020 to mid-January 2021), and 16 to wave 3 (collected between late January and June 2021).

The 15 lineages collected prior to and during wave 1 were derived from 120/483 genomes. The single most dominant lineage was B.1, which accounted for 55.8% of the genomes, followed by B.1.1 (22.5%) and A.25 (6.7%). Other minor lineages included B.1.243 (2.5%), B.1.549, B.1.446, B.1.393 and C.11 at 1.7%, while A, B.1.1.33, B.1.166, B.1.36.10, B.1.397, B.1.511 and B.1.612 were at <1%. In the 11 lineages identified in wave 2 (80/483 genomes), all but B.1, B.1.549 and B.1.446 were new (i.e. not detected in the first wave). B.1 remained as the dominant lineage (56.3%). B.1.549, which was a minor lineage during wave 1, was now the second most dominant lineage at 12.5%, followed by B.1.530 (10.0%) and B.1.596.1 (6.3%). Other lineages included A.23.1 (3.8%), A.23, B.1.1.254 and B.1.446 at 2.5% and B.1.384, B.1.428, N.8 at 1.3%. During wave 3, 16 lineages were observed from 283/483 genomes. During wave 3, three VOCs (B.1.1.7, B.1.351 and B.1.167.2) and one VOI (B.1.525) emerged. Collectively, the VOC/VOI rapidly expanded, replacing all Wave 2 lineages to account for 90.8% of all lineages in Wave 3. B.1.1.7 (alpha variant) was observed at a frequency of 59.0%, B.1.167.2 (delta variant) at 20.1%, B.1.525 (Eta variant) at 7.4% and 4.2% for B.1.351 (beta variant). Other variants observed during this wave included A.23.1 (3.5%) and B.1 (1.8%) and B.1.1.301, B.1.530, B.1.1, B.1.221, B.1.36.16, B.1.367, B.1.466, B.1.466.1, B.1.551 and B.4.4 at <1%.

Figure 2 shows the relationships between Pangolin lineages across the three COVID-19 waves. While each wave had characteristic lineages, some lineages were shared across the waves. In our dataset, B.1 was the only lineage present in all three waves. B.1.549 and B.1.446 were present in waves 1 and 2, A.23.1 and B.1.530 were present in waves 2 and 3, and lineage B.1.1 was present in waves 1 and 3.

Time scaled phylogenetic tree of Kenyan samples

A time-scaled phylogenetic tree including 112 genomes sampled from around the globe and 323 from this study is shown Figure 3. Kenyan samples branched into multiple lineages, suggesting multiple seeding events, and formed monophyletic clusters with notable intercluster divergence, indicating local transmission diversification.

Emergence and dominance of alpha (B.1.1.7) and delta (B.1.167.2) VOCs

A time-scaled phylogeny involving 1,067 B.1.1.7 genomes rooted against the Wuhan/WHO1/2019 reference is shown in Figure 4. The tree includes alpha variants from Kenya (n=151) shown as red circular branch tips; date range: 01 February 2021 to 20 May 2020, other parts of Africa (n=872); date range: 25 November 2020 to 29 June 2021 and samples of the earliest reported alpha variants (n=44); date range: 08 July 2020 to 30 November 2020, 88.6% of which were from England. Kenyan samples branched from different parts of the tree, indicating multiple independent alpha variant seeding events. No local transmission of the B.1.1.7 lineage was recorded prior to February 2021. The first recorded major case involved samples brought into the laboratory in February 2021 from an outbreak cluster (n=61) that occurred in Nanyuki, a small town in Laikipia County at the foothills of Mount Kenya. The suspected source was British soldiers returning from the United Kingdom. All the associated samples were of the B.1.1.7 lineage. The samples branched at three different locations on the tree (Figure 4, black arrows), indicating that the outbreak had three likely sources, with the majority of the cases (n=23) coming from a

single source. From our analysis, the earliest introduction of the alpha variant in Kenya outside the Nanyuki outbreak was from two samples (MOHK-TMP-PMK-4273 and KDH-234) collected on 01 February 2021 and 05 February 2021, respectively, both from a common source (Figure 4, shown as encircled red dots).

A time-scaled delta variant phylogenetic tree involving 893 B.1.617.2 genomes rooted with the Wuhan/WHO1/2019 reference is shown in Figure 5. The tree included Kenyan samples (n=31 from the current study and 2 from the coast deposited in GISAID; date range: 29 April 2021 to 02 June 2021), other parts of Africa (n=812; date range: 10 March 2020 to 09 July 2021) and 46 samples of the earliest reported delta variants (date range: 05 February 2021 and 31 March 2021), all traceable to India. In the tree, Kenyan samples branched at six locations, indicative of multiple independent introduction events (Figure 5). The earliest introductions were from four samples (Figure 5, red stars) from Nairobi collected in late April 2021. The other two delta introduction events included two samples from Kisumu that branched with a clade dominated by samples from Uganda and Rwanda (Figure 5, blue star) and a large cluster dominated by samples from Kisumu (Figure 5, purple star).

Discussion

In this study, we performed genomic surveillance of COVID-19 in the periods preceding and during the three waves that occurred in Kenya to understand the dynamics of SARS-CoV-2 lineages that were driving the waves. We show that the waves were driven by characteristic lineages (Figures 1 and 2). Each wave was preceded by low infection rates, probably as variants competed through narrow transmission bottlenecks that selected the fittest variants (Lythgoe et al., 2021), some of which eventually became the dominant variants. As shown in Figure 3, Kenyan samples into multiple lineages, illustrating multiple introduction events, and thereafter formed monophyletic clusters with notable intercluster divergence indicating ongoing local transmission. VoCs did not emerge in Kenya until after February 2021.

The earliest SARS-CoV-2 samples sequenced at our laboratory were collected in May 2020. This was 2 months after the confirmation of the first Kenyan case of COVID-19 on 13th March 2021 (Ministry of Health, 2020). Based on the genome sequences, the early SARS-CoV-2 was seeded from the early European lineages (B.1 and B.1.1), and the A.25 Ugandan lineage (Figure 1). The B.1 lineage dominated and had a countrywide distribution, having been identified in the counties of Kisumu, Kiambu, Busia, Nairobi, Uasin Gishu, Bungoma, Baringo, Mombasa and Wajir (data not shown). By August 2020, the B.1 lineage was the third most prevalent lineage globally, with 82,672 sequences deposited in the GISAID (n=82,672). This European lineage was first detected on ^{24} January 2020 and was most reported in North America and Europe. Its origin roughly corresponds to the Northern Italian outbreak early in 2020 (O'Toole et al., 2021). In our dataset (Figure 1, Figure 2 and Table S2), the B.1 lineage was maintained across the three COVID-19 waves, and while it was the most dominant lineage during the first (55.8%) and second (56.3%) waves, its dominance waned considerably by the third wave (1.8%). Other core lineages defining wave 1 were B.1.1 (European lineage that emerged in early February 2020), which occurred at a frequency

of 22.5%, and A.25 (Ugandan lineage), which occurred at a frequency of 6.7%. Within the A.25 lineage, 6/8 samples were collected from transborder truck drivers at Busia, a border town of Kenya and Uganda.

B.1.549, which was the fifth most prevalent lineage (1.7%) during wave 1, became the second most prevalent lineage at wave 2, at 12.5% (Supplementary Table 1). This lineage is mostly associated with Kenyan sequences and likely emerged from local transmission events. The majority of samples in our dataset from this lineage were from the Kenyan coast. The lineage was, however, not detected in the third wave, probably having been outcompeted to extinction by the more easily transmissible VOCs. The last global report of the B.1.549 lineage in GISAID was on January 29^{th,} 2021, Ohio, USA. Other local lineages that were present during wave 2 included B.1.530 (10.1%), B.1.596.1 (6.3%), N.8 (1.3%), B.1.428, B1.384 and the Ugandan lineage A.23. It is interesting to note that during the first and second waves, the local lineages persisted amidst the more dominant B.1 (Supplementary Table 1). The travel restrictions instituted early in these outbreaks may have allowed maintenance of local transmission events in the absence of external introductions.

The long interval between waves 2 and 3 allowed complacency in COVID-19 control practices, thus allowing introduction and displacements of local lineages by the VOCs. The only local lineages that survived passed wave 2, albeit at low frequencies, were B.1, B.1.1, B.530 and A.23.1 (Figure 2). The dominance of the B.1 lineage in the previous 2 waves was replaced by the alpha variant (B.1.1.7 lineage) that became the dominant lineage, accounting for 59.0% of all detected lineages in the early part of wave 3. Our earliest sample with the B.1.1.7 was on 1st February 2021 from two samples that came from Thika, Kiambu County, Kenya. Later that week, we detected the VOC in an outbreak that occurred in Nanyuki, Laikipia County and was linked to the British Army Training Unit in Kenya. Between December 2020 and January 2021, the variant was rampant in the UK (NERVTAG, 2020). Phylogenetic analysis (Figure 4) estimated this outbreak to be the first major introduction of the alpha variant into the country. The outbreak seemed to have been well contained, as there were no indications (with our data and Kenyan data deposited in GISAID) of out-branching from this outbreak cluster (Figure 4, black arrows). Other alpha variant clusters appeared to have been introduced multiple times (Figure 4, red circular branch tips) from independent sources. In less than 3 months after its detection, the alpha variant became the most dominant lineage and was the major cause of COVID-19 infections during the early part of the third wave. The alpha variant possesses several nonsynonymous mutations of immunological importance (Andrew et al., 2020) that are thought to confer increased transmissibility (Volz et al., 2021).

The delta variant (B.1,167.2 lineage), originally identified in India in October 2020, (Cherian et al., 2021; Edara et al., 2021), was first identified in Kisumu, the third largest city on the shores of Lake Victoria and was linked to travelers returning from India (Wasike, 2021). By the end of May 2021, delta had become the dominant variant in Western Kenya (David, 2021). Driven by its high transmissibility, estimated to be 60% more than the alpha variant (Planas et al., 2021), the variant soon extended its grip to the rest of the country. By August 2021, Kenya entered the 4th wave that was fueled by the delta variant. Our phylogenetic analysis corroborates initial reports (Wasike, 2021) that the delta variant was introduced in

Kenya through Kisumu. As shown in Figure 5, there is a large outbreak cluster dominated by samples from Kisumu (Figure 5, purple star) that branched with an Indian sample at the base of the clade.

In January 2021, the first introduction of the beta variant (B.1.351 lineage) in Kenya was reported in Kilifi County (KEMRI, 2021). Though not as highly transmissible as the alpha and delta variants, it has immune escape mutations (Harvey et al., 2021; Wibmer et al., 2021), which could potentially compromise COVID-19 vaccines. Similar to the other two VOCs, this lineage emerged during the third wave, and in our dataset, it was the fourth most dominant lineage at 4.2% (Supplementary Table 1). Of all the beta variants deposited in GISAID from Kenya (n=184: Date accessed 5 August 2021), 84.2% of the B.1.351 lineages were from coastal Kenya, including Kilifi County, Kwale and Mombasa County. The overrepresentation of the B.1.351 lineage on the Kenyan Coast points to its possible introduction through the southern border with Tanzania and to the two tourists from South Africa.

In addition to the three VOCs, wave 3 had two VOIs. The B.1.525 lineage (Eta variant), which has E484K, Q677H, F888L and a deletion suite similar to B.1.1.7 (O'Toole et al., 2021), was the third most prevalent at 7.4%. This variant had a countrywide distribution, including Western Kenya (Busia, Kisumu, Migori and Nyamira Counties), Coastal Kenya (Mombasa and Kwale), Rift Valley (Nandi, Uasin Gishu Counties), Northern Kenya (Garissa) and Eastern Kenya (Maukeni) and Nairobi County (data not shown). The other VOI was the A.23.1 lineage, an international lineage with variants of potential biological concern (O'Toole et al., 2021). This variant contains a constellation of mutations, including E484K, that could reduce COVID-19 vaccine effectiveness (Bugembe et al., 2020, 2021). The variant was dominant in the period between September and November 2020 in Uganda (Bugembe et al., 2020). Most of the samples with this lineage came from Busia, a major border town of Kenya and Uganda. It is likely that the lineage was seeded into Kenya from Uganda during cross-border trade and movement of people.

Conclusion

Three COVID-19 waves occurred in Kenya, and by the time of pressing, a 4^{th} wave had emerged. The waves were fueled by different core sets of lineages. Wave one was seeded by imported lineages, mainly of European origin. The second wave had a mix of European and local lineages, the latter arising from local transmission and diversification. The third wave was dominated by imported VOCs that totally displaced lineages identified in waves 1 and 2. Going forward, genomic surveillance will play a critical role in generating SARS-CoV-2 lineage intelligence and especially cataloguing those associated with disease severity and vaccine breakthrough events.

Declarations

Acknowledgement

Kenya's MOH provided extraction and PCR kits. SARS-CoV-2 sequencing and personnel cost were funded by the Armed Forces Health Surveillance Division (AFHSD) and its Global Emerging Infections

Surveillance and Research Branch (ProMIS P0095_21_KY, 2021). The funders had no role in data collection and analysis, decision to publish, or preparation of the manuscript.

We thank all of the authors who have contributed genome data in GISAID. Author credits for specific GISAID contributions can be found on www.gisaid.org/.

References

Aksamentov, I., Neher, R., Hodcroft, E., Sibley, T., Sanderson, Th., Stroud, N., Rubinsteyn, A., Bottoms, C., & Maguire, F. (2020). *Nextclade: Viral genome alignment, mutation calling, clade assignment, quality checks and phylogenetic placement*. https://github.com/nextstrain/nextclade

Andrew, R., Nick, L., Oliver, P., Wendy, B., Jeff, B., Alesandro, C., Connor, T., Peacock, T., Robertson, D. L., & Volz, E. (2020). Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *Virological.Org*. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563

Bugembe, D. L., Kayiwa, J., Phan, M. V. T., Tushabe, P., Balinandi, S., Dhaala, B., Lexow, J., Mwebesa, H., Aceng, J., Kyobe, H., Ssemwanga, D., Lutwama, J., Kaleebu, P., & Cotten, M. (2020). Main Routes of Entry and Genomic Diversity of SARS-CoV-2, Uganda. *Emerging Infectious Diseases*, *26*(10), 2411–2415. https://doi.org/10.3201/eid2610.202575

Bugembe, D. L., Phan, M. V. T., Ssewanyana, I., Semanda, P., Nansumba, H., Dhaala, B., Nabadda, S., O'Toole, Á. N., Rambaut, A., Kaleebu, P., & Cotten, M. (2021). Emergence and spread of a SARS-CoV-2 lineage A variant (A.23.1) with altered spike protein in Uganda. *Nature Microbiology 2021 6:8*, *6*(8), 1094–1101. https://doi.org/10.1038/s41564-021-00933-9

Cherian, S., Potdar, V., Jadhav, S., Yadav, P., Gupta, N., Das, M., Rakshit, P., Singh, S., Abraham, P., & Panda, S. (2021). Sars-cov-2 spike mutations, l452r, t478k, e484q and p681r, in the second wave of covid-19 in Maharashtra, India. *Microorganisms*, *9*(7), 2021.04.22.440932. https://doi.org/10.3390/microorganisms9071542

David, H. (2021). *Delta Variant May Spark 'Catastrophic' Covid Wave in Kenya in Weeks*. Bloomberg. https://www.bloomberg.com/news/articles/2021-06-29/delta-variant-may-spark-catastrophic-wave-in-kenya-in-weeks

Edara, V.-V., Pinsky, B. A., Suthar, M. S., Lai, L., Davis-Gardner, M. E., Floyd, K., Flowers, M. W., Wrammert, J., Hussaini, L., Ciric, C. R., Bechnak, S., Stephens, K., Graham, B. S., Bayat Mokhtari, E., Mudvari, P., Boritz, E., Creanga, A., Pegu, A., Derrien-Colemyn, A., … Fabrizio, T. P. (2021). Infection and Vaccine-Induced Neutralizing-Antibody Responses to the SARS-CoV-2 B.1.617 Variants. *New England Journal of Medicine*, NEJMc2107799. https://doi.org/10.1056/nejmc2107799

Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., & Neher, R. A. (2018). NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/bty407

Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S. J., & Robertson, D. L. (2021). SARS-CoV-2 variants, spike mutations and immune escape. In *Nature Reviews Microbiology* (Vol. 19, Issue 7, pp. 409–424). Nature Publishing Group. https://doi.org/10.1038/s41579-021-00573-0

KEMRI. (2021). *Detection of SARS-CoV-2 Variant 501Y . V2 (South African origin) in Coastal Kenya* (Vol. 2, Issue JANUARY).

Lythgoe, K. A., Hall, M., Ferretti, L., de Cesare, M., MacIntyre-Cockett, G., Trebes, A., Andersson, M., Otecko, N., Wise, E. L., Moore, N., Lynch, J., Kidd, S., Cortes, N., Mori, M., Williams, R., Vernet, G., Justice, A., Green, A., Nicholls, S. M., … Golubchik, T. (2021). SARS-CoV-2 within-host diversity and transmission. *Science*. https://doi.org/10.1126/SCIENCE.ABG0821

Ministry of Health. (2020). *First Case of Coronavirus Disease Confirmed in Kenya*. https://www.health.go.ke/first-case-of-coronavirus-disease-confirmed-in-kenya/

NERVTAG. (2020). *NERVTAG/SPI-M: Extraordinary meeting on SARS-CoV-2 variant of concern 202012/01 (variant B.1.1.7), 21 December 2020 - GOV.UK*. https://www.gov.uk/government/publications/nervtagspi-m-extraordinary-meeting-on-sars-cov-2-variant-of-concern-20201201-variant-b117-21-december-2020

O'Toole, Á., Hill, V., Pybus, O. G., Watts, A., Bogoch, I. I., Khan, K., Messina, J. P., Tegally, H., Lessells, R. R., Giandhari, J., Pillay, S., Tumedi, K. A., Nyepetsi, G., Kebabonye, M., Matsheka, M., Mine, M., Tokajian, S., Hassan, H., Salloum, T., … Kraemer, M. U. G. (2021). Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2. *Wellcome Open Research*, *6*(121), 121. https://doi.org/10.12688/wellcomeopenres.16661.1

Planas, D., Veyer, D., Baidaliuk, A., Staropoli, I., Guivel-Benhassine, F., Rajah, M. M., Planchais, C., Porrot, F., Robillard, N., Puech, J., Prot, M., Gallais, F., Gantner, P., Velay, A., Le Guen, J., Kassis-Chikhani, N., Edriss, D., Belec, L., Seve, A., … Schwartz, O. (2021). Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature 2021*, 1–5. https://doi.org/10.1038/s41586-021-03777-9

Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*. https://doi.org/10.1038/s41564-020-0770-5

Ritchie, H., Ortiz-Ospina, E., Beltekian, D., Mathieu, E., Hasell, J., Macdonald, B., Giattino, C., Appel, C., Rodés-Guirao, L., & Roser, M. (2020). *Coronavirus Pandemic (COVID-19)*. OurWorldInData.Org. https://ourworldindata.org/coronavirus

Sevinsky, J., Nassiri, A., Young, E., Blankenship, H., Libuit, K., Oakeson, K., Turner, L., & Consortium, S.-B. (2020). *SARS-CoV-2 Sequencing on Illumina MiSeq Using ARTIC Protocol: Part 1-Tiling PCR V.1 Coronavirus Method Development Community StaPH-B 1 more workspace*. 0–7. https://dx.doi.org/10.17504/protocols.io.bfefjjbn

Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. In *Eurosurveillance* (Vol. 22, Issue 13). https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494

Tyghe, V., Panciera, M., & Melendrez, M. (2016). *ngs_mapper*. https://doi.org/https://10.5281/zenodo.46716

Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., Hinsley, W. R., Laydon, D. J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C. V., Boyd, O., Loman, N. J., McCrone, J. T., Gonçalves, S., … Ferguson, N. M. (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*. https://doi.org/10.1038/s41586-021-03470-x

Wasike, A. (2021). *Kenya confirms 5 cases of Indian COVID-19 variant*. https://www.aa.com.tr/en/africa/kenya-confirms-5-cases-of-indian-covid-19-variant/2230554

WHO. (2021). *Tracking SARS-CoV-2 variants*. https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/

Wibmer, C. K., Ayres, F., Hermanus, T., Madzivhandila, M., Kgagudi, P., Oosthuysen, B., Lambson, B. E., de Oliveira, T., Vermeulen, M., van der Berg, K., Rossouw, T., Boswell, M., Ueckermann, V., Meiring, S., von Gottberg, A., Cohen, C., Morris, L., Bhiman, J. N., & Moore, P. L. (2021). SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nature Medicine*, *27*(4). https://doi.org/10.1038/s41591-021-01285-x

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., … Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270–273. https://doi.org/10.1038/s41586-020-2012-7

Figures

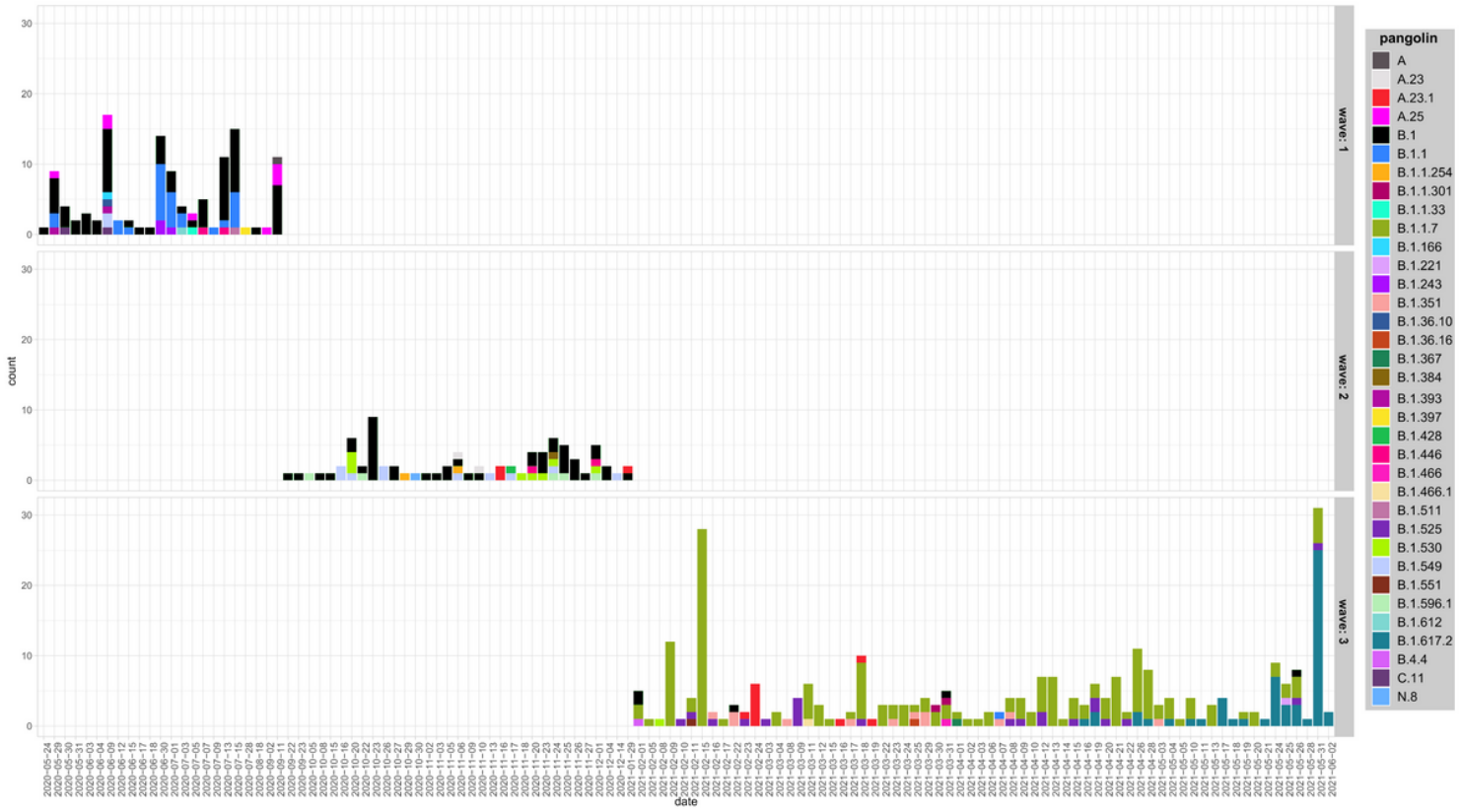

Figure 1

SARS-CoV-2 lineages circulating across the three COVID-19 waves in Kenya. Pango lineages detected during wave 1, wave 2 and wave 3.

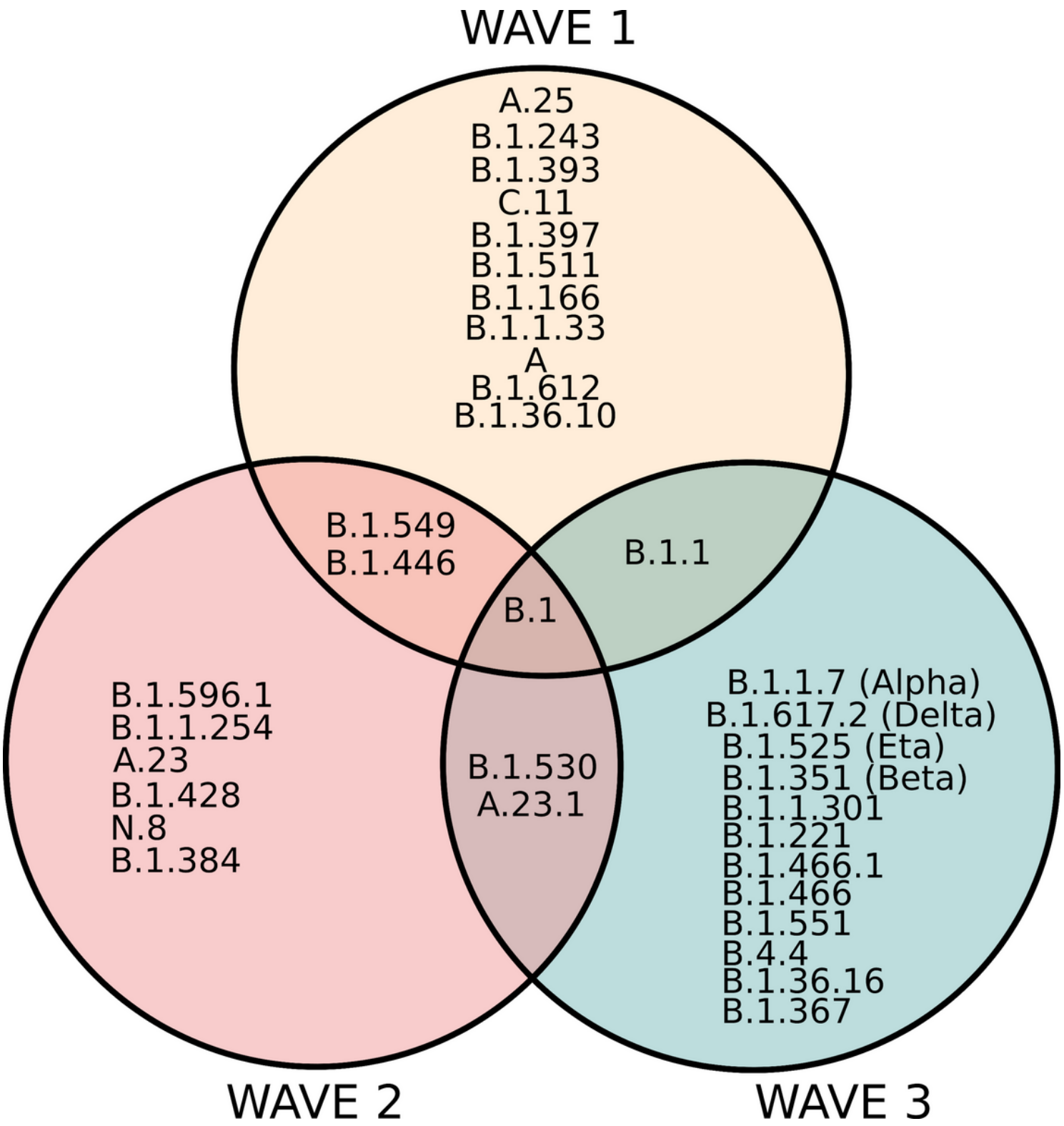

Figure 2

Venn plot showing unique and shared lineages across the three COVID-19 waves.

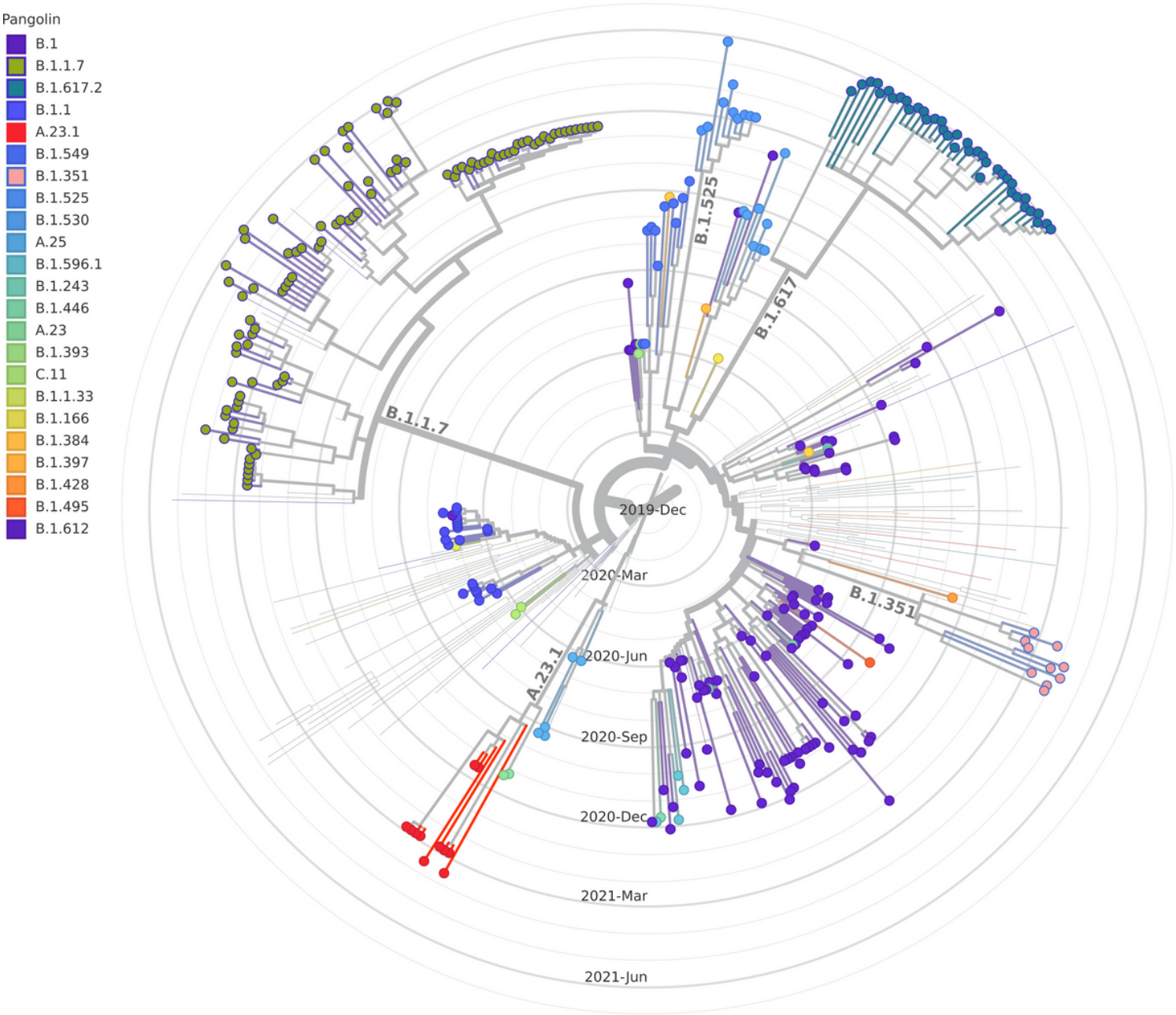

Figure 3

Time-scaled phylogenetic tree of Kenyan samples against global isolates. The tree was constructed with 112 genomes sampled from GISAID and 323 genomes from this study. Thin lines represent context global samples, while thick lines represent Kenyan samples. The different colors on circular tips of branches represent the Pango lineages.

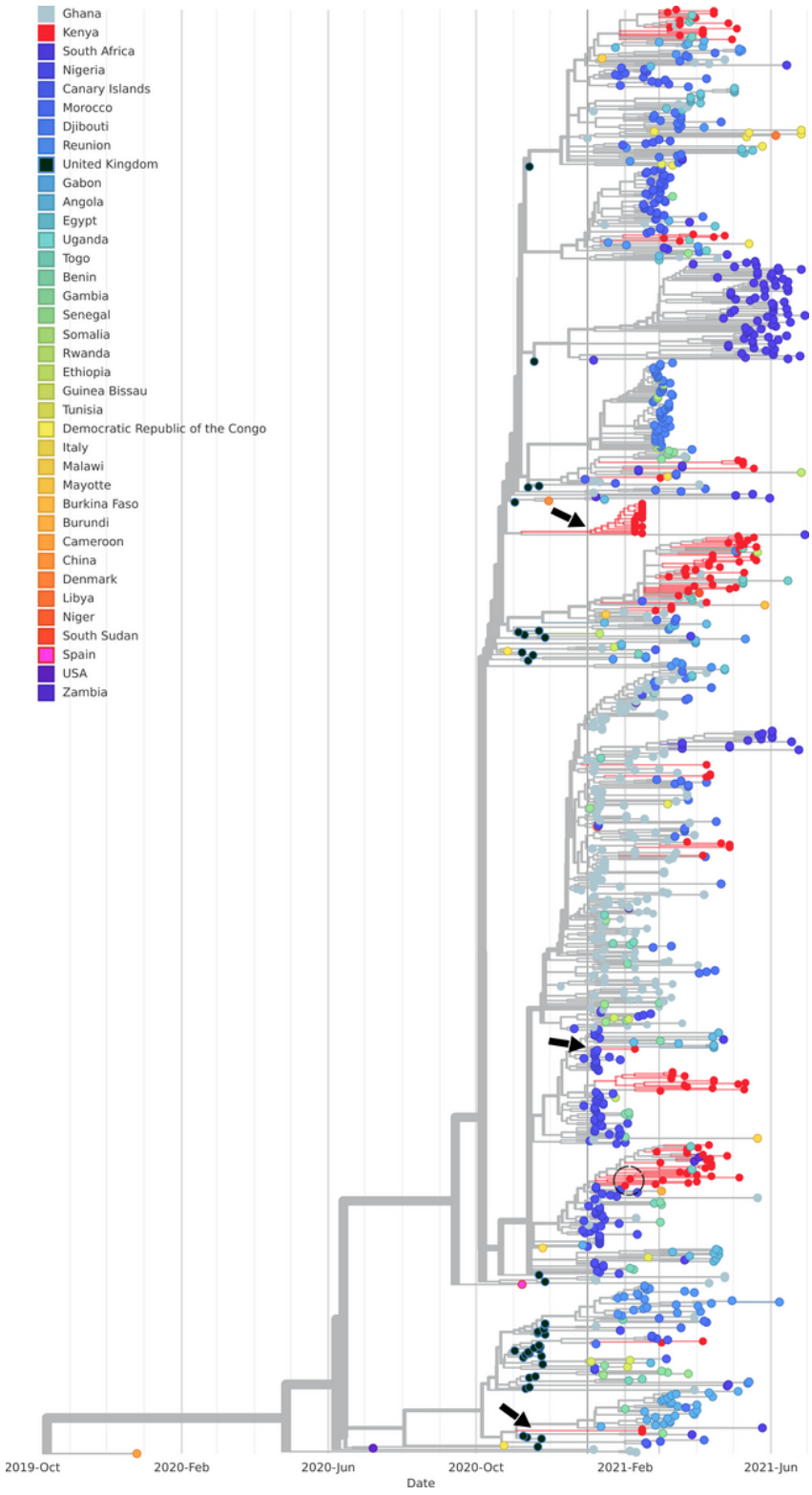

Figure 4

Phylogenetic tree of the B.1.1.7 lineage from our samples and those from across Africa. The tree was constructed with 1,068 genomes, including those from Kenya (n=151), those from Africa (n=872), and early B.1.1.7 lineages (n=44), and rooted with the Wuhan/WHO1/2019 reference. Kenyan samples are shown in red circular branch tips. Kenyan samples branched from different parts of the tree, indicating multiple independent alpha variant seeding events. Black arrows show samples from one of the first

major introductions of the B.1.1.7 lineage in Kenya following an outbreak in Nanyuki, Kenya. Samples in encircled red dots show earlier introduction (01 February 2021 and 05 February 2021) of the alpha variant outside the Nanyuki outbreak, both from a common source.

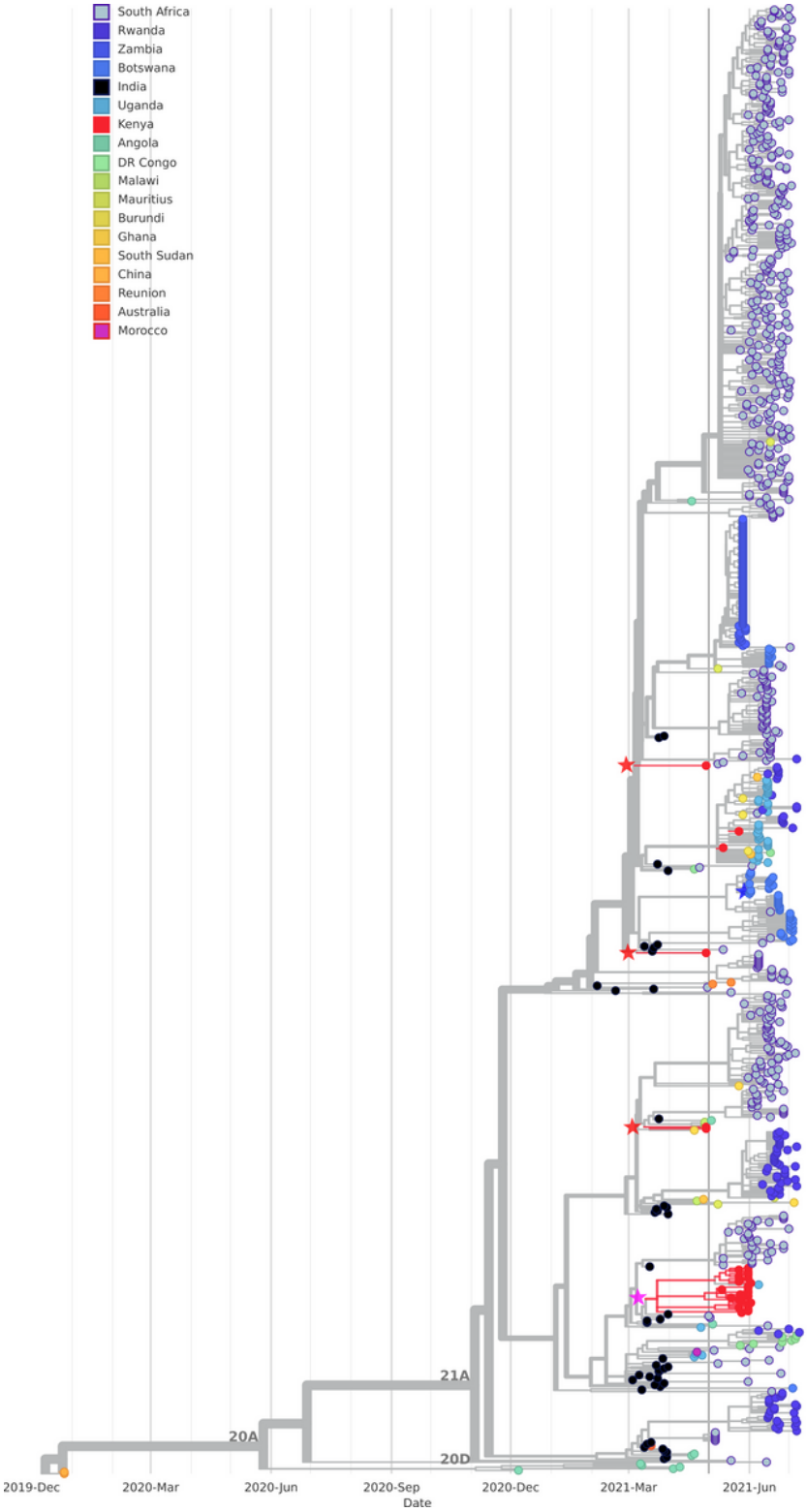

Figure 5

Phylogenetic tree of the B.1.617.2 lineage from our study samples and those from across Africa. The tree was constructed with 893 genomes, including those from Kenya (n=33), those from other parts of Africa

(n=812) and early B.1.617.2 lineages (n=46) traceable to India. Kenyan samples are shown as circular red branch tips. The red stars show the earliest delta variant introduction in late April 2021 from Nairobi samples, while samples from Kisumu (the county that had the first major delta variant outbreak) are contained in clades represented by blue and purple stars.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- KimitaetalSupplementalInformation.docx