Article:
Kanteh, A., Jallow, H.S., Manneh, J. et al. (25 more authors) (2023) Genomic epidemiology of SARS-CoV-2 infections in The Gambia: an analysis of routinely collected surveillance data between March, 2020, and January, 2022. The Lancet Global Health, 11 (3). e414-e424. ISSN 2214-109X

https://doi.org/10.1016/s2214-109x(22)00553-8

Genomic epidemiology of SARS-CoV-2 infections in The Gambia: an analysis of routinely collected surveillance data between March, 2020, and January, 2022

Abdoulie Kanteh, Haruna S Jallow, Jarra Manneh, Bakary Sanyang, Mariama A Kujabi, Sainabou Laye Ndure, Sheikh Jarju, Alhagie Papa Sey, Dabiri Damilare K, Yaya Bah, Sana Sambou, Gibril Jarju, Buba Manjang, Abubacarr Jagne, Sheikh Omar Bittaye, Mustapha Bittaye, Karen Forrest, Desta Alamerew Tiruneh, Ahmadou Lamin Samateh, Sheriffo Jagne, Stéphane Hué, Nuredin Mohammed, Alfred Amambua-Ngwa, Beate Kampmann, Umberto D'Alessandro, Thushan I de Silva, Anna Roca, Abdul Karim Sesay

Summary

Background COVID-19, caused by SARS-CoV-2, is one of the deadliest pandemics of the past 100 years. Genomic sequencing has an important role in monitoring of the evolution of the virus, including the detection of new viral variants. We aimed to describe the genomic epidemiology of SARS-CoV-2 infections in The Gambia.

Methods Nasopharyngeal or oropharyngeal swabs collected from people with suspected cases of COVID-19 and international travellers were tested for SARS-CoV-2 with standard RT-PCR methods. SARS-CoV-2-positive samples were sequenced according to standard library preparation and sequencing protocols. Bioinformatic analysis was done using ARTIC pipelines and Pangolin was used to assign lineages. To construct phylogenetic trees, sequences were first stratified into different COVID-19 waves (waves 1–4) and aligned. Clustering analysis was done and phylogenetic trees constructed.

Findings Between March, 2020, and January, 2022, 11 911 confirmed cases of COVID-19 were recorded in The Gambia, and 1638 SARS-CoV-2 genomes were sequenced. Cases were broadly distributed into four waves, with more cases during the waves that coincided with the rainy season (July–October). Each wave occurred after the introduction of new viral variants or lineages, or both, generally those already established in Europe or in other African countries. Local transmission was higher during the first and third waves (ie, those that corresponded with the rainy season), in which the B.1.416 lineage and delta (AY.34.1) were dominant, respectively. The second wave was driven by the alpha and eta variants and the B.1.1.420 lineage. The fourth wave was driven by the omicron variant and was predominantly associated with the BA.1.1 lineage.

Interpretation More cases of SARS-CoV-2 infection were recorded in The Gambia during peaks of the pandemic that coincided with the rainy season, in line with transmission patterns for other respiratory viruses. The introduction of new lineages or variants preceded epidemic waves, highlighting the importance of implementing well structured genomic surveillance at a national level to detect and monitor emerging and circulating variants.

Funding Medical Research Unit The Gambia at London School of Hygiene & Tropical Medicine, UK Research and Innovation, WHO.

Lancet Glob Health 2023; 11: e414–24

Medical Research Council Unit The Gambia at London School of Hygiene & Tropical Medicine, Banjul, The Gambia (A Kanteh MSc, J Manneh MSc, B Sanyang BSc, M A Kujabi MSc, S L Ndure BSc, S Jarju DVM, D Damilare K BSc, Y Bah BSc, K Forrest MD, N Mohammed PhD, A Amambua-Ngwa PhD, Prof B Kampmann PhD, Prof U D'Alessandro PhD, T I de Silva PhD, Prof A Roca PhD, A K Sesay PhD)**; National Public Health Reference Laboratory** (H S Jallow MSc, A P Sey HND, S Jagne PhD)**, Ministry of Health, Banjul, The Gambia** (S Sambou MSc, G Jarju MSc, B Manjang PhD, A Jagne MD, S O Bittaye MD, M Bittaye FWACS, A L Samateh MD)**; Centre for Mathematical Modelling of Infectious Diseases and Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK** (S Hué, PhD)**; The Florey Institute for Host–Pathogen Interactions & Department of Infection, Immunity and Cardiovascular Disease, Medical School, University of Sheffield, Sheffield, UK** (T I de Silva)**; World Health Organization, Banjul, The Gambia** (D A Tiruneh MPH)

Correspondence to:
Dr Abdul Karim Sesay, Medical Research Council Unit The Gambia at London School of Hygiene & Tropical Medicine, Banjul, The Gambia
aksesay@mrc.gm

For **WHO COVID-19 data** see https://covid19.who.int

Introduction

The COVID-19 pandemic caused by SARS-CoV-2 has claimed more than 6 million lives and infected more than 617 million people globally (as of September, 2022).[1] Only 2·0% of cases of, and 3·9% of deaths from, COVID-19 have been recorded in Africa, despite Africa accounting for 17% of the world's population. By contrast, 36% of COVID-19 cases and deaths have been recorded in Europe—more than 225 million infections. The pandemic has evolved differently around the globe, and started later in Africa than in other regions.[2] Despite the relatively low number of cases recorded in Africa, seroprevalence studies suggest that transmission has been much higher than reported.[3]

Despite the low mutation of SARS-CoV-2, widespread transmission and high case numbers increased the virus's diversity during the pandemic.[4] Prolonged infections in immunocompromised individuals[5] and a high frequency of reinfections increase the likelihood of new mutations, leading to the emergence of new variants. More importantly, new variants of concern reported in different countries are responsible for increased transmission, disease severity, and immune evasion.[6] For example, the alpha variant or B.1.1.7 lineage, which was first identified in the UK in November, 2020, was reported to be 43–90% more transmissible than the wild-type SARS-CoV-2.[7] The delta variant (B.1.617.2), which was first reported in India in October, 2020, is four times more likely to evade the immune system and 60% more transmissible than the alpha variant.[8] The emergence of the delta variant changed the dynamic of the epidemic in many African countries

Research in context

Evidence before this study

We searched PubMed with the terms: "genomics" AND "epidemiology" AND "SARS-CoV-2" AND "Africa" for articles published in English between Dec 1, 2019, and Feb 28, 2022. Our search returned 71 publications, only two of which described the genomic epidemiology of SARS-CoV-2 in a west African country, clearly showing the lack of genomic sequencing capacity in this region and hampering understanding of the evolution of the virus during the pandemic. Most of the data available focused on a few samples mainly collected during the early phases of the pandemic. Two reports were based on available African SARS-CoV-2 genomes in the Global Initiative on Sharing Avian Influenza Data, and showed disparities in how the SARS-CoV-2 virus and the COVID-19 pandemic have evolved on the continent. The results suggest that a robust and holistic approach is required to understand country-specific patterns, which could be useful for the monitoring and control of the pandemic worldwide.

Added value of this study

We present a comprehensive description of the genomic epidemiology of the SARS-CoV-2 pandemic in The Gambia informed by 1638 genomes sequenced over almost 2 years. We describe four waves of COVID-19, with the most intense waves recorded during the Gambian rainy seasons. All waves ended despite the lack of major restrictions to limit transmission. New waves happened when new variants or lineages, or both, entered the country, generally those already prevalent in Europe and other countries. Phylogenetic clustering showed high local transmission during the summer (rainy season), with most viral importations during the dry season.

Implications of all the available evidence

Proper surveillance, including strengthening entry points and screening of asymptomatic individuals, especially during the rainy seasons, would help to control potential future waves of COVID-19 in The Gambia and west Africa. In-depth analysis—including phylodynamic inference—could help to elucidate the transmission dynamics and importation patterns of SARS-CoV-2 over the course of this and future pandemics.

because it coincided with an increase in reported cases and deaths.[9] The omicron variant, first identified in South Africa in November, 2021, was described as the most transmissible variant, quickly spread across the world, and is highly evasive of antibody immunity.[10]

Effective strategies to quickly detect new variants, sources of infections, outbreaks, and transmission patterns in different geographical settings are essential to controlling the spread of SARS-CoV-2. Whole-genome sequencing enables detailed understanding of the evolution and population structure of pathogens.[11] High-resolution viral genome sequences linked with epidemiological data enabled identification of the origin of SARS-CoV-2[12] and monitoring of the virus's spread and evolution, both globally[13] and locally.[14] Furthermore, the information obtained from genomic sequencing is essential for guiding the design of the next generation of vaccines.[15] As of September, 2022, more than 13 million SARS-CoV-2 genomes have been sequenced globally and submitted to the Global Initiative on Sharing Avian Influenza Data (GISAID). Unfortunately, only 1% of these sequences are from African countries, which limits understanding of the local evolution of the virus and its effects on the continent.

For **GISAID** see https://www.gisaid.org/

The Gambia, a small country in west Africa, identified its first imported COVID-19 case from the UK[16] in March, 2020. By February, 2022, The Gambia had recorded almost 12 000 cases of COVID-19 and 365 deaths. More than 1600 SARS-CoV-2 genomes have already been sequenced and analysed, with the aim of providing real-time genomic data to monitor circulating variants within The Gambia. Moreover, this information is essential for monitoring local transmission and will substantially contribute to the global genomic dataset. We have already described the origin of the first five cases diagnosed in The Gambia during the early phase of the pandemic,[16] and used phylogenetic analysis to confirm SARS-CoV-2 reinfection in two healthy individuals.[17] In this study, we aimed to extensively describe the genomic epidemiology of the first four SARS-CoV-2 waves in The Gambia.

For **Gambian COVID-19 data** see https://moh.gov.gm

Methods

Study context

We did an epidemiological genomic analysis of SARS-CoV-2 detected by routine testing in The Gambia. The Gambia has a population of about 2·5 million people, who are predominantly Muslim. The median age is 17·8 years. More than half of the population live in urban areas, mainly on the coast. The Gambian climate is subtropical with two seasons: a rainy season between June and October (average temperature 23–33°C, >80% humidity) and a dry season from November to May (average temperature 18–30°C).[18]

The Gambian Government is the country's main health-care provider, and operates seven referral hospitals, eight major health centres, and 16 minor health centres across the country. Health-care delivery is mostly via the primary or local minor health centres. In addition, there are about 30 private and non-governmental organisation clinics across the country. Measures implemented to control the spread of COVID-19 in The Gambia have been described previously.[19] Briefly, international borders were closed in March, 2020, and a state of emergency was declared. Initial SARS-CoV-2 testing by PCR was focused on

identifying and tracing imported cases plus isolating case contacts, especially travellers from Senegal and their contacts (Senegal had already experienced a COVID-19 wave and borders The Gambia on three sides). This study was approved by the joint ethics committee of the Gambian Government and the Medical Research Council (MRC) Unit The Gambia at the London School of Hygiene & Tropical Medicine (LSHTM). The committee waived the need to obtain informed consent from participants.

Procedures

Samples analysed in this study were routinely collected by the Gambian Ministry of Health and MRC Unit The Gambia at LSHTM. There were no specific criteria for sample collection, but most (>85%) were collected as a pre-travel testing requirement (appendix p 1). Demographic data, including sex, age, address, travel history, and other relevant information were collected whenever possible (appendix p 3). Lineages and associated metadata collected for each sample were recorded in RedCap at the MRC Unit The Gambia at LSHTM diagnostic laboratory and DHIS2 at the Gambian Ministry of Health's National Public Health Laboratory.

FLOQSwabs (Copan Diagnostics, Murrieta, CA, USA) were used to obtain nasopharyngeal or oropharyngeal samples, or both. Swabs were placed in single tubes containing a refrigerated universal transport medium and delivered to the diagnostic laboratory within 24 h. Extracted RNA was kept at –80°C pending confirmation of the presence of SARS-CoV-2 by PCR. Samples that tested positive were subsequently sent to the genomics laboratory at MRC Unit The Gambia at LSHTM for sequencing. In most cases, samples were freeze-thawed once to preserve RNA integrity before sequencing. Samples without basic metadata such as date of collection, age, and sex were not processed. From March to June, 2020, all samples were processed and sequenced, irrespective of cycle threshold values. Because of the high proportion of failed sequencing among samples with cycle threshold values higher than 30, from June, 2020, only samples with a value of 30 or less were sequenced (except for in a few critical cases when sequencing was necessary for patient management). In addition, samples with low DNA concentration (≤4 ng/μL) after library preparation were also excluded.

Sample processing, RNA extraction, and real-time RT-PCR were done according to standard WHO guidelines.[19]

Genomic analysis

Complementary DNA conversion and multiplex PCR were prepared according to the ARTIC nCoV-2019 protocol. Updated primer schemes were always used to ensure optimum sequencing of emerging variants of concern. The list of different primers used throughout with start and end dates are listed in the appendix (p 80). Due to variable cycle threshold values, all RNA samples were run for 35 cycles. Pooled PCR products were either cleaned up using AMPure XP beads (Indianapolis, IN, USA) or library prepared directly without bead clean-up according to manufacturer's instructions. Libraries were prepared for sequencing on the Oxford Nanopore (Oxford, UK) platform.

Samples were prepared and sequenced in 96-well plates with one complementary-DNA and RNA-extraction negative control each, per plate. FASTQ files were analysed using ARTIC guidelines to generate consensus genomes. Lineages were assigned with Pangolin and quality of the genomes was assessed with Nextclade. A consensus sequence was defined as passing quality control if greater than 50% of the genome was covered by confident calls or there was at least one contiguous sequence of more than 10 000 bases and no evidence of contamination in the negative control. A confident call was defined as having 20 times depth of coverage. For Gambian sequences, if the coverage fell below these thresholds, the bases were masked with the character "N", indicating that the base at that position is unknown or not available. Low-quality single-nucleotide polymorphisms were also masked with Ns.

See **Online** for appendix

Clustering and phylogenetic analyses

For phylogenetic analysis, we constructed two sets of trees. The first tree consists of only Gambian sequences. Only sequences for which less than 10% of bases were ambiguous were included to build a robust phylogenetic tree. First we aligned the sequences using MAFFT (version 7.407) and then constructed a maximum likelihood tree using IQTREE (version 1.3.11.1) with ModelFinder, branch support, and bootstrap using the commands "iqtree -s gambia_seq.fasta -alrt 1000 -bb 1000". ModelFinder was used to establish the best model for our dataset. Thus, the Hasegawa-Kishino-Yano (HKY) model of nucleotide substitution fitted our dataset and was used to construct a maximum likelihood phylogenetic tree with branch supports estimated using ultra-fast bootstrapping with 1000 replicates. The output tree file was visualised and annotated with microreact.[20]

Clustering analysis was done in Uvaia (version 2.0.1). First, we stratified Gambian SARS-CoV-2 sequences into different pandemic waves (waves 1–4) on the basis of local transmission. For each wave, we used "uvaialign" in Uvaia to align each sequence against the Wuhan-Hu-1 MN908947.3 reference genome with the command "srun -p gpu -N 1 --cpus-per-task 16 uvaialign -r --reference= MN908947.3.fasta first_wave.fa". The alignment file from each wave was then used to search for the closest neighbouring sequences available in GISAID with the command "srun -p gpu -N 1 --cpus-per-task 16 uvaia -r peroba_align.210706_182403.aln.xz --trim 230 first_wave.fa -o first_wave". The last accessed date for the full database used was April 12, 2022. For each run, a csv file containing the closest sequences to Gambian samples was filtered and only the closest

For **Uvaia** see https://github.com/quadram-institute-bioscience/uvaia

For the **ARTIC Network** see https://artic.network

For the **nCoV-2019 protocol** see https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bp2l6n26rgqe/v3

matching sequences from GISAID to each Gambian sample (here denoted as rank 1) were used to construct each phylogenetic tree. For each wave, a maximum likelihood phylogenetic tree was constructed.

Statistical analysis

Summary statistics were prepared using proportions for categorical variables. 95% CIs were calculated according to exact binomial distribution, giving at least 5% absolute precision for proportions of major variants noted in most cases. All data management and statistical analyses were done in R (version 4.2.1) and RStudio (2022.7.1.554). The sample size was based on convenience sampling and determined by the total number of SARS-CoV-2 cases in The Gambia, as well as the proportion that were sent for sequencing from the National Public Health Laboratory and MRC Unit The Gambia laboratory services.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Between March 17, 2020, and Jan 31, 2022, 75 554 samples were tested for SARS-CoV-2 in The Gambia and 11 911 (16%) were positive. Among the 4479 samples received for sequencing, 1643 (37%) passed our inclusion criteria and quality controls and were included in this study. Five samples failed genotyping and were dropped for further analysis (figure 1). The proportion of cases successfully sequenced ranged from 4% to 100% (figure 2; appendix p 81), giving at least 5% absolute precision for the 95% CI estimate for the proportion of variants detected in most cases (appendix p 81). Most samples were collected in the coastal region of The Gambia (appendix p 3).

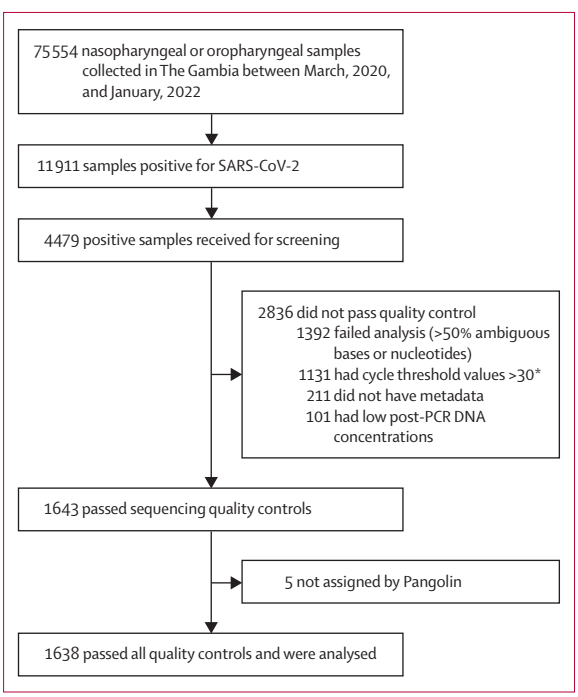

Figure 1: **Distribution of samples received for sequencing from March, 2020 to January, 2022**
*This threshold was introduced in July, 2020.

The Gambia has experienced four COVID-19 waves since the first identified case of SARS-CoV-2 infection (figure 3A). The first wave, which was predominantly associated with the B.1.416 lineage started in July, 2020, and lasted for 3 months. The second wave was from January to May, 2021, and and was milder (fewer official cases and deaths; figure 3A). Variants eta (B.1.525) and alpha (B.1.1.7) and lineage B.1.1.420 dominated during the second wave. The third wave (July–September, 2021) was associated with a higher number of cases and deaths than the previous two waves and was driven by the delta (B.1.617.2) variant and its subtypes. The fourth wave was between December, 2021, and January, 2022, and was dominated by the omicron variant and associated subtypes (figure 3B).

Between March, 2020, and January, 2022, 55 different lineages were identified (appendix p 83). The delta variant and it subtypes accounted for 492 (30% [95% CI 28–32]) of the 1638 sequenced samples (appendix p 82), and the B.1.416 lineage accounted for 191 (12% [10–13]). Eight lineages—delta (AY.34.1), B.1.416, B.1.617.2, B.1.1, eta (B.1.525), B.1.1.420, omicron (B.A.1.1), and alpha (B.1.1.7)—were each detected in more than 100 samples and thus represented the dominant lineages in The Gambia. 37 (67%) of the 55 lineages identified were present in fewer than ten samples (appendix p 81).

The frequency and distribution of lineages changed with time (figure 4). During the first wave, lineage B.1.416 represented 173 (49% [95% CI 44–54]) of the 354 samples sequenced and B.1.1 accounted for 57 (16% [12–20]; appendix p 82). Some lineages identified during the first wave disappeared during the second wave, and new lineages emerged: B.1.416, for example represented only eight (1%) of 555 lineages sequenced in the second wave, with the last sample identified in February, 2021 (figure 4A). In March, 2021, almost a year after the identification of the first case of COVID-19 in The Gambia, B.1.1.420 became the dominant lineage, accounting for 126 (23% [95% CI 19–26) of 555 sequenced samples. Eta (B.1.525), meanwhile, accounted for 124 (22% [19–26]) and alpha (B.1.1.7) for 119 (21% [18·1–25·1]; appendix p 82).

The delta (B.1.617.2) variant was first confirmed in The Gambia in early April, 2021, and became the dominant variant within a few weeks (figure 4C). Between July and August, 2021 (ie, the peak of the third wave), the delta variant subtype AY.34.1 accounted for 256 (54% [95% CI 50–59]) of 470 samples sequenced. Other prominent lineages included B.1.617.2, which

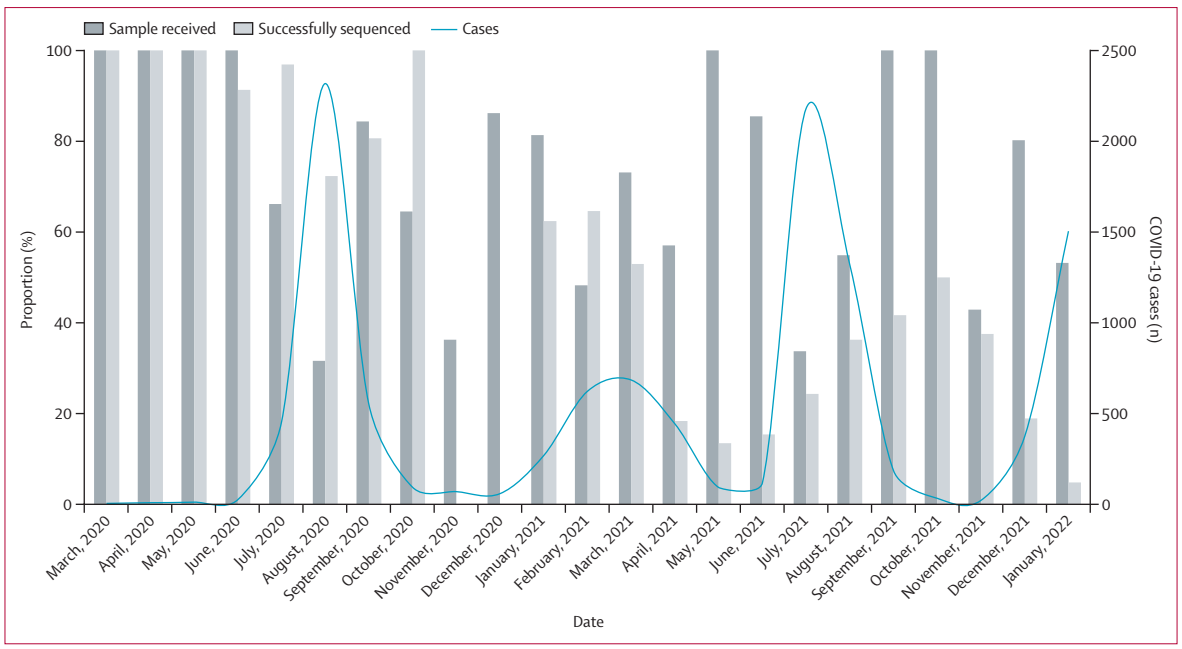

Figure 2: **Proportion of samples collected and successfully sequenced and COVID-19 case numbers in The Gambia**

accounted for 134 (34% [26–35]) samples, and AY.34, which accounted for 41 (9% [95% CI 6–11]; appendix p 82).

The omicron variant and its subtypes dominated the fourth wave (figure 4D). Subtype BA.1.1 accounted for 125 (79% [95% CI 72–86]) of 156 genomes sequenced (appendix p 82). BA.1 represented approximately seven (4% [2–9]) samples, and delta variant subtypes accounted for 12 (8%), with subtype AY.34 being the most common.

Phylogenetic analysis of Gambian samples revealed eight distinct, strongly supported clades, which clustered based on lineage types (figure 5). The delta variant formed two distinct clades, with lineage AY.34.1 forming a separate clade from B.1.617.2 and other delta lineages (figure 5). Phylogenetic clustering based on closest global sequences showed high local transmission events during the first and third waves (figure 6A, C). Lineage B.1.416 clustered with genomes from Senegal (B.1.416), suggesting the spread of infections between these two countries (figure 6A). Similarly, high clustering among Gambian delta samples during the third wave showed possible importation and subsequent spread via local transmission (figure 6C). Overall, cluster analysis showed possible importation of the virus, mainly from Europe and Africa (figure 6).

Discussion

In this Article, we have presented a comprehensive description of the genomic epidemiology of the SARS-CoV-2 epidemic in The Gambia. Since the first detected case in March, 2020, there have been four different waves, two of which fell during the rainy season and two during the dry season. All waves ended despite a lack of major behavioural restrictions to stop the epidemic. New waves occurred when new variants or lineages entered the country (generally those already prevalent in Europe and other countries at the time). Phylogenetic clustering revealed high local transmission during the rainy season (ie, the first and third waves), with more imported cases during the second and fourth waves. Lineage diversity was higher in the first and second waves than in the third and fourth. The high proportion of samples testing positive for SARS-CoV-2 enabled our study to provide insights that would not have been possible with low-density genomics surveillance.

Both the first and the third waves lasted less than 3 months but were intense. They occurred during the rainy seasons and started at almost the same time in 2020 (B.1.416 lineage) and 2021 (delta variant), when humidity is more than 80% and daily maximum temperatures are often higher than 33°C.[21] The effect of temperature and humidity on SARS-CoV-2 transmission is unclear,[21] but our findings are consistent with transmission dynamics described for other respiratory viral pathogens, for which the rainy season coincides with the period of highest incidence.[22] Even if virus infectivity is not modified by climate parameters, people are more likely to spend time indoors during heavy rains, which might increase the likelihood of household transmission.[18,19] The hypothesis that lower temperatures could be associated with increased SARS-CoV-2 transmission[23] as a result of increased susceptibility to infection due to irritation of nasal mucosae (especially in temperate regions) does not fit with the occurrence of the second wave in The Gambia. Despite the introduction of the very transmissible alpha and eta variants,[24] the second wave had a lower

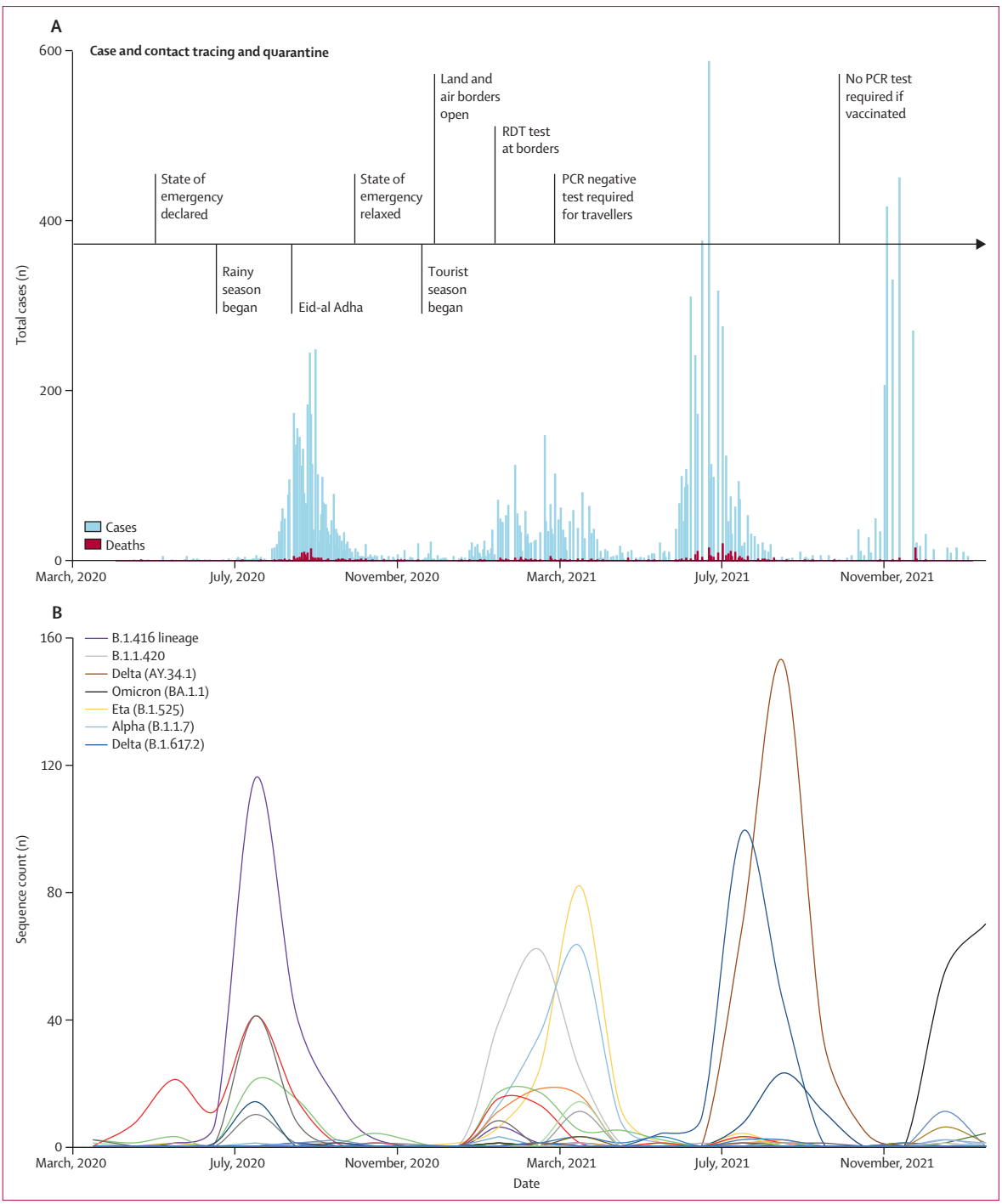

Figure 3: **COVID-19-associated cases and deaths relative to pandemic containment measures implemented (A) and trajectory of major SARS-CoV-2 lineages and variants (B) in The Gambia from March, 2020, to January, 2022**
RDT=rapid diagnostic (antigen) test.

epidemic peak and lasted longer than the first and third waves. Lineage diversity was also highest in this second wave, which followed the relaxation of lockdowns and the official opening of both air and land borders. Consequently, new lineages and variants were introduced to the country from different parts of the world, as was clearly shown by our clustering analysis.

By contrast with Senegal and Burkina Faso, where sharp increases in cases of COVID-19 occurred immediately after the first few confirmed cases,[19] the first epidemic

wave in The Gambia did not immediately follow the diagnosis of the first case. This lag was probably due to the declaration of a state of emergency, in which The Gambia's air and land entry points were closed. Although genomic analysis of the first five cases in The Gambia confirmed importation of the virus from Europe and Asia,[16] the close genomic relatedness of Gambian sequences during the first wave with viruses sampled in Senegal suggests that importation of the new strains occurred mainly by land because restrictions on air travel lasted until the end of the first wave. The closure of land borders is also more difficult to enforce than are air travel restrictions.

These genomic findings are epidemiologically probable. The sudden increase in cases in July, 2020, coincided with preparations for the major Muslim feast Eid-al Adha, when an inflow of undetected cases across the Senegal–Gambia border might have occurred. This timing aligns with the high prevalence of the Senegal–Gambia lineage B.1.416, which was the dominant lineage in the first COVID-19 wave in Senegal (March–August, 2020).[25] This lineage is characterised by few mutations across the genome, the most notable of which is the D614G on the spike protein gene, which is linked to increased viral load and high transmission but not to increased disease severity.[26] So far, The Gambia and Senegal have reported the highest number of cases caused by B.1.416. Other African countries, such as Burkina Faso and Morocco, have also reported this lineage, but per GISAID data, only two genomes have been recorded in each country. The low number of isolates of this lineage in other west African or neighbouring countries could be because of a lack of sequencing facilities in the region. Nonetheless, according to GISAID, South Africa sequenced more genomes than any other country in Africa and did not identify this lineage. Importantly, high clustering among Gambian sequences suggests high local transmission, mostly associated with the B.1.416 lineage. Consistent with our findings, importations of SARS-CoV-2 within Africa increased as more cases were reported on the continent due to high local transmission.[27] However, phylogenetic analysis of Gambian genomes along with global sequences also showed relatedness with samples from Europe, North America, and Asia—specifically the B.1 and B.1.1 lineages, which were dominant in most countries in the early phase of the pandemic.

The second wave of COVID-19 in The Gambia was probably precipitated by the reopening of borders and relaxation of lockdown measures in September, 2020, and the influx of tourists from Europe and Africa during and immediately after Christmas, 2020. The introduction of more lineages and variants and the clustering of Gambian sequences with those from different parts of the world further supports the role of travelling and tourism in encouraging transmission. This wave was predominantly caused by two variants, alpha and eta, and the B.1.1.420 lineage. By contrast with reports from the UK and elsewhere,[28] in The Gambia the alpha variant, although more transmissible than previous lineages, was associated with a slightly lower number of sequenced cases during the wave than the eta and B.1.1.420 lineage. Zhao and colleagues[29] used mathematical modelling to show that the eta variant seemed slightly more infectious than the alpha variant in Nigeria, a pattern also noted in other African countries.[30] Phylogenetic analysis showed clustering of most of the alpha Gambian samples with samples from Germany, Ghana, Spain, and the UK, while eta samples clustered with those from the UK, Germany, and Nigeria (figure 6). The B.1.1.420 lineage, which was reported to originate in Italy[25] and later became the dominant lineage in both Senegal[25] and The Gambia's second wave, clustered with sequences mainly from Senegal. Genomic analysis showed mutations in the spike protein associated with immune invasion (N440K)[31] and increased transmission (D614G).[26] Perez and colleagues[25] suggested that the B.1.1.420 lineage

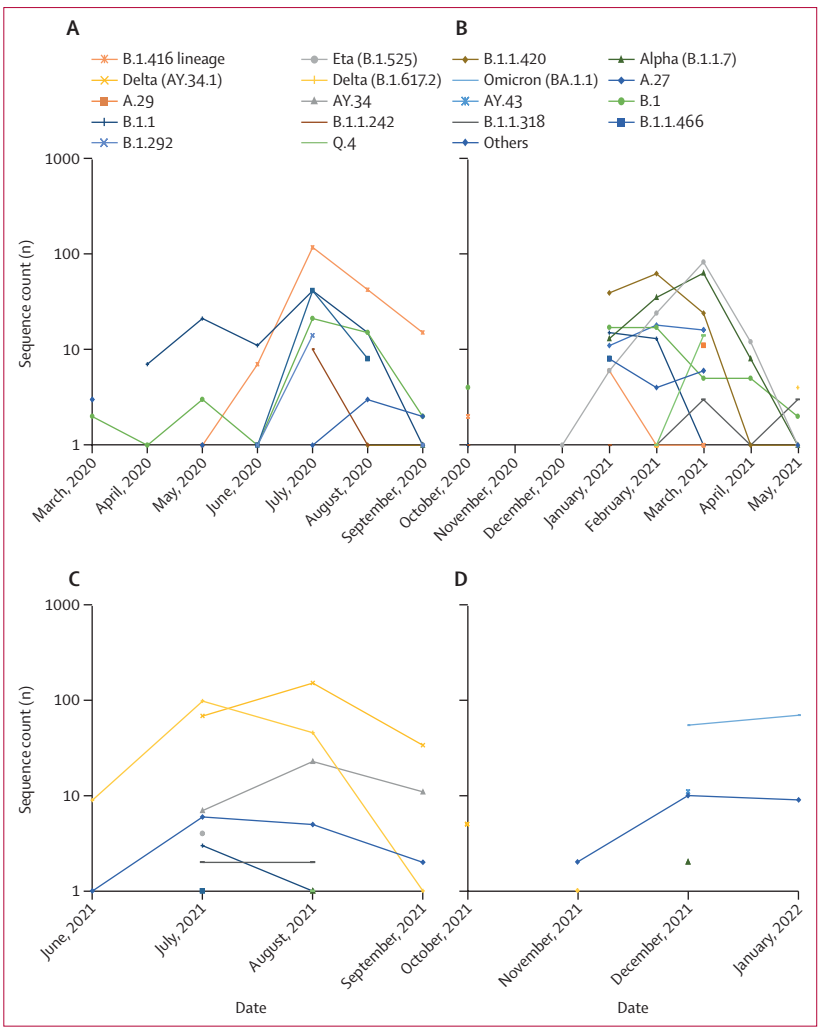

Figure 4: **SARS-CoV-2 lineages in the first (A), second (B), third (C), and fourth (D) waves of the COVID-19 pandemic in The Gambia**
"Others" includes lineages present in ten or fewer samples (appendix p 83). The y-axis is on a logarithmic scale.

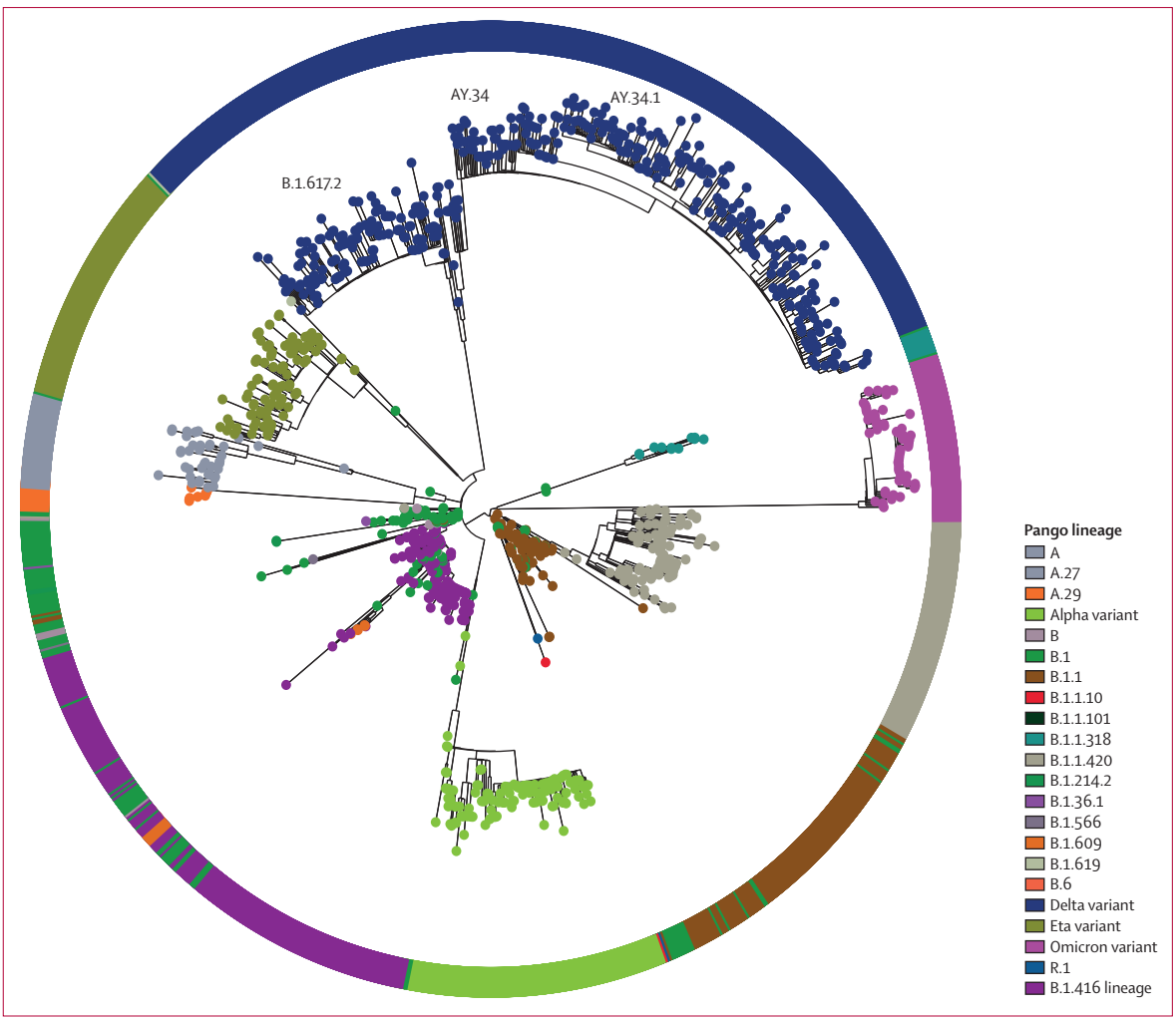

Figure 5: **Maximum likelihood tree of 1313 SARS-CoV-2 genomes sampled in The Gambia from March, 2020, to January, 2022**
Sequences in which less than 10% of bases were ambiguous were used to construct the phylogenetic tree.

was driven by both increased genetic fitness and other factors, such as relaxation of lockdowns in Senegal.

The third wave was caused by the highly transmissible delta variant and its subtypes, resulting in high numbers of daily cases. The first delta sequence recorded in The Gambia was isolated in April, 2021, from a traveller from Lebanon, a country where delta was the dominant strain between April and June, 2021.[32] In The Gambia, the peak of the third wave was in July–August, 2021, and was associated with increased mortality (as in other African countries), with deaths peaking during the week of July 19. According to our phylogenetic analysis, most of the Gambian delta sequences were highly related, suggesting high local transmission. A few genomes clustered with sequences from Europe, Africa (Senegal only), Asia, and North America, suggesting possible importation from these regions. By contrast with other circulating variants, delta evolved into two distinct clades, suggesting that the variant may have gone through positive selection due to the acquisition of many mutations. This divergence was driven by the AY.34.1 sublineage forming a single cluster associated with more than half (54%) of the samples sequenced in The Gambia during this wave. High prevalence of this lineage was reported in Senegal during the same period, with only a few cases reported in other parts of Africa, highlighting again the similarities of the COVID-19 pandemic in the two countries.

The fourth wave was caused by the omicron variant and characterised by the sharpest increase in cases, followed by a drastic decrease in less than a month. In South Africa, where omicron was first detected in November, 2021, daily cases peaked in December and quickly declined thereafter.[33] A similar trend was noted in many other countries, including in eastern and central Africa and Europe.[33] The rapid decline of cases might have been due to the high infectivity of omicron, leading to a high proportion of infection in the population, which

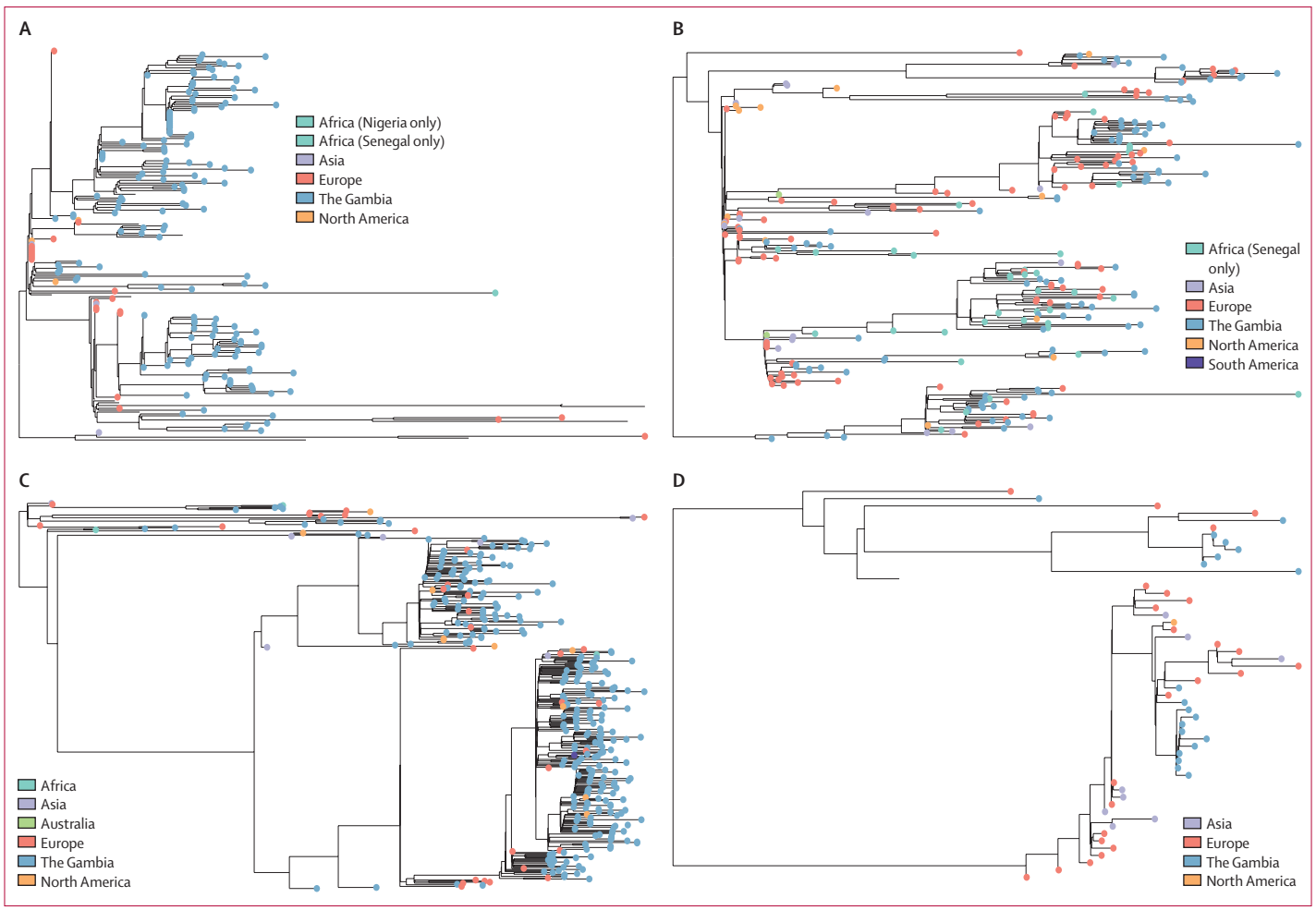

Figure 6: **Maximum likelihood tree of the closest SARS-CoV-2 genomes to those sampled in The Gambia in the first (A), second (B), third (C), and fourth (D) waves of the COVID-19 pandemic**
Data are from March, 2020, to January, 2022. Branch tips represent the continent where closest global sequences were sampled from. All trees were rooted on the SARS-CoV-2 reference genome (GenBank accession number MN908947.3).

subsequently led to a decrease in susceptible individuals. The fourth wave coincided with the Gambian tourist season (usually October–February), when fully vaccinated individuals were allowed into the country without needing a negative PCR result. As expected, phylogenetic analysis showed more Gambian omicron sequences clustering with those from Europe, the origin of more than half of all tourists.[34] In addition, BA.1.1 subtypes, which were associated with high infection rates in Europe and North America, were the most dominant subtypes in The Gambia. The omicron variant encodes more than 30 amino acid substitutions in the spike protein, enhancing its ability to evade the immune system.[21] It thus spreads faster and is more likely to reinfect previously infected or vaccinated individuals, or both, than previous variants. A substantial number of COVID-19 cases might have been missed as the requirement for a negative PCR for outgoing travellers (ie, people leaving The Gambia), who represent almost 90% of SARS-CoV-2 cases in this study, was waived. Delta variant subtypes were still circulating during the fourth wave, most notably the AY.43 sub-lineage, which has been reported in more than 130 countries.

Our study has some limitations, which should be considered when interpreting our results. Systematic testing of individuals with influenza-like symptoms in The Gambia has been limited throughout the COVID-19 pandemic, and therefore most sequences were obtained from samples collected from asymptomatic outgoing travellers. Thus, overall case numbers have probably been substantially underestimated, and some lineages circulating in the country might have been missed. Furthermore, not all positive samples were sent for sequencing, with a lower proportion sequenced during the peak of the first wave. As a result, there is probably an increased likelihood that lineages could have been missed during this first wave. Most samples were collected at the coast, and lineages circulating exclusively upcountry could have been missed. The inference of increased transmission and severity in the third wave,

mainly caused by the delta variant, although expected as per international data could have been biased by a potential change in behaviour towards testing sick individuals or the increased number of outgoing travellers coinciding with summer, 2021, when international flights were open. In addition, most positive samples failed our quality controls, which could have led to under-reporting of some lineages. Finally, given that SARS-CoV-2 sampling in many countries has been affected by sampling bias, the sequences deposited in GISAID might not have all the necessary metadata, and thus the findings of our clustering analysis should be interpreted with caution.

Our analysis of the SARS-CoV-2 epidemic in The Gambia showed higher peak waves during the rainy season, in line with transmission patterns for other respiratory viruses. Lineage diversity was highest when borders were opened. The introduction of new lineages and variants preceded epidemic waves, highlighting the importance of implementing well structured and active genomic surveillance at the national level to detect and monitor emerging and circulating variants. In-depth analysis, including phylodynamic inference, could help to elucidate the transmission dynamics and importation patterns of SARS-CoV-2 over the course of this and future pandemics.

Contributors

AK, AR, and AKS conceived the study and did the formal analysis. AK did the bioinformatics analysis. JM, BS, MAK, SLN, DDK and YB processed the samples (library preparation and sequencing). SJag, HSJ, APS, SJar, and KF were responsible for sample acquisition and diagnostic PCR. SS, GJ, BM, AJ, SJag, SOB, MB, ALS, SH, NM, AA-N, BK, UD'A, TIdS, AR, and AKS were responsible for leadership and supervision of different aspects of the work. DAT, TIdS, AR, BK, AKS and UD'A acquired funding. AK and AKS had full access to and verified all the data in the study. AK, AR, and AKS wrote the first draft of the Article. All authors reviewed the final version and supported the decision to submit it for publication.

Declaration of interests

We declare no competing interests.

Data sharing

All consensus sequences were deposited with GISAID and the US National Center for Biotechnology Information if they met minimum quality-control requirements. De-identified individual participant data including text, tables, figures, and supplementary material are available for public access.

Acknowledgments

We wish to acknowledge the support of all staff at the National Public Health Laboratory and the Directorate of Planning and Information and Epidemiology and Disease Control (Banjul, The Gambia). We thank the diagnostic laboratory team, high-power computing team, field workers, clinical staff, and data management team for their valuable support, all participants whose samples were used in this study, and all laboratories and institutions that deposited their sequences in GISAID for public access.

References

1 Worldometer. COVID-19 live coronavirus pandemic. 2022. https://www.worldometers.info/coronavirus/ (accessed Sept 20, 2022).
2 Martinez-Alvarez M, Jarde A, Usuf E, et al. COVID-19 pandemic in west Africa. *Lancet Glob Health* 2020; **8:** e631–32.
3 Uyoga S, Adetifa IMO, Karanja HK, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Kenyan blood donors. *Science* 2021; **371:** 79–82.
4 Chen J, Lu H. New challenges to fighting COVID-19: virus variants, potential vaccines, and development of antivirals. *Biosci Trends* 2021; **15:** 126–28.
5 Corey L, Beyrer C, Cohen MS, Michael NL, Bedford T, Rolland M. SARS-CoV-2 variants in patients with immunosuppression. *N Engl J Med* 2021; **385:** 562–66.
6 US Centers for Disease Control and Prevention. SARS-CoV-2 variant classifications and definitions. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html (accessed Nov 30, 2022).
7 Davies NG, Abbott S, Barnard RC, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 2021; **372:** eabg3055.
8 Planas D, Veyer D, Baidaliuk A, et al. Reduced sensitivity of SARS-CoV-2 variant delta to antibody neutralization. *Nature* 2021; **596:** 276–80.
9 WHO. COVID-19 variants prolong Africa's pandemic wave. 2021. https://www.afro.who.int/news/covid-19-variants-prolong-africas-pandemic-wave (accessed Feb 15, 2022).
10 Karim SSA, Karim QA. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *Lancet* 2021; **398:** 2126–28.
11 Bryant J, Chewapreecha C, Bentley SD. Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. *Future Microbiol* 2012; **7:** 1283–96.
12 Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020; **395:** 565–74.
13 Alm E, Broberg EK, Connor T, et al. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European region, January to June 2020. *Eurosurveillance* 2020; **25:** 2001410.
14 Page AJ, Mather AE, Le-Viet T, et al. Large-scale sequencing of SARS-CoV-2 genomes from one region allows detailed epidemiology and enables local outbreak management. *Microb Genom* 2021; **7:** 000589.
15 Seib KL, Dougan G, Rappuoli R. The key role of genomics in modern vaccine and drug design for emerging infectious diseases. *PLoS Genet* 2009; **5:** e1000612.
16 Kanteh A, Manneh J, Jabang S, et al. Origin of imported SARS-CoV-2 strains in The Gambia identified from whole genome sequences. *PLoS One* 2021; **16:** e0241942.
17 Sanyang B, Kanteh A, Usuf E, et al. COVID-19 reinfections in The Gambia by phylogenetically distinct SARS-CoV-2 variants—first two confirmed events in west Africa. *Lancet Glob Health* 2021; **9:** e905–07.
18 Access Gambia. Climate of Gambia. 2021. https://www.accessgambia.com/information/climate-weather.html (accessed March 8, 2022).
19 Abatan B, Agboghoroma O, Akemoke F, et al. Intense and mild first epidemic wave of coronavirus disease, The Gambia. *Emerg Infect Dis* 2021; **27:** 2064–72.
20 Argimón S, Abudahab K, Goater RJE, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* 2016; **2:** e000093.
21 Mecenas P, da Rosa Moreira Bastos RT, Rosário Vallinoto AC, Normando D. Effects of temperature and humidity on the spread of COVID-19: a systematic review. *PLoS One* 2020; **15:** e0238339.
22 Jarju S, Greenhalgh K, Wathuo M, et al. Viral etiology, clinical features and antibiotic use in children <5 years of age in The Gambia presenting with influenza-like illness. *Pediatr Infect Dis J* 2020; **39:** 925–30.
23 Moriyama M, Hugentobler WJ, Iwasaki A. Annual review of virology seasonality of respiratory viral infections. *Annu Rev Virol* 2020; **7:** 83–101.
24 Zhao Y, Huang J, Zhang L, Chen S, Gao J, Jiao H. The global transmission of new coronavirus variants. *Environ Res* 2022; **206:** 112240.
25 Perez LJ, Orf GS, Berg MG, et al. The early SARS-CoV-2 epidemic in Senegal was driven by the local emergence of B.1.416 and the introduction of B.1.1.420 from Europe. *Virus Evol* 2022; **8:** 1–12.
26 Korber B, Fischer WM, Gnanakaran S, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 2020; **182:** 812–27.
27 Wilkinson E, Giovanetti M, Tegally H, et al. A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. *Science* 2021; **374:** 423–31.

28 Angius F, Pala G, Manzin A. SARS-CoV-2 and its variants: the pandemic of unvaccinated. *Front Microbiol* 2021; **12:** 2695.

29 Zhao S, Musa SS, Chong MK, et al. Correspondence to: the co-circulating transmission dynamics of SARS-CoV-2 alpha and eta variants in Nigeria: a retrospective modeling study of COVID-19. *J Glob Health* 2021; **11:** 5028.

30 Tegally H, San JE, Cotten M, et al. The evolving SARS-CoV-2 epidemic in Africa: insights from rapidly expanding genomic surveillance. *Science* 2022; **378:** eabq5358.

31 Weisblum Y, Schmidt F, Zhang F, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* 2020; **9:** e61312.

32 Merhi G, Trotter AJ, de Oliveira Martins L, et al. Replacement of the alpha variant of SARS-CoV-2 by the delta variant in Lebanon between April and June 2021. *Microb Genom* 2022; **8:** 000838.

33 WHO. Omicron-fuelled COVID-19 surge in Africa plateaus. 2022. https://www.afro.who.int/news/omicron-fuelled-covid-19-surge-africa-plateaus (accessed March 10, 2022).

34 Access Gambia. Tourist statistics for Gambia. 2022. https://www.accessgambia.com/information/tourism-statistics.html (accessed May 24, 2022).