BMC Medical Genomics

RESEARCH

Open Access

Genomic surveillance and phylo-evolutionary analysis of SARS-CoV-2 variants in Osun state, Nigeria during the second wave of COVID-19 pandemic

Sunday Babatunde Akinde[1,2], Omotayo Opemipo Oyedara[3], Taiwo Samson Olumakinde[2], Oluwaseyi Paul Olaniyan[2], Folasade Muibat Adeyemi[1], Rahman Ayodele Bolarinwa[4,5], Omolola Yinka Adeagbo[6], Timilehin Emmanuel Oluokun[2], Waidi Folorunso Sule[1*], Olabisi Olaniyi Ojo[7], Bamidele Abiodun Iwalokun[8], Emmanuel Sunday Fajoyegbe[1], Elijah Kolawole Oladipo[9,10], Daniel Oladimeji Oluwayelu[11], Abideen Akinkunmi Wahab[1], Hazeez Adebayo Durosomo[12], Temitope Fasunloye Ajani[1], Adetoun Adebanke Adebunmi[1], Olayiwola Olayode[13], Omokaro Obire[14] and Janet Olubukola Olaitan[1]

Abstract

Background SARS-CoV-2 genome sequencing, genomic characterization, and global data sharing are recommended to facilitate countermeasures against COVID-19. This study involved whole genome sequencing and phylo-evolutionay analysis of SARS-CoV-2 variants circulating in Osun State, Nigeria between January and June, 2021.

Methods In the present retrospective, molecular epidemiologic study, a total of 60 nasopharyngeal samples from RT-qPCR-positive COVID-19 patients originating from eight study locations were analyzed. This was done using RNA extraction application of the McKinsey Global Institute's (MGI) DNBSEQ-G50RS high-throughput genome Sequencing Technology and Bioinformatics Analysis Pipeline. Quantitative data were analyzed with descriptive and inferential statistics.

Results Forty-five SARS-CoV-2 (45/60) whole genome sequences (WGSs) were successfully analyzed with participants being mostly male adults (64.4%) and a median age of 44 years. Five PANGO lineages including: B (35.6%), B.1.525 (31.1%), B.1.1.7 (28.9%), B.1 (2.2%), and L.3 (2.2%) were identified, with Eta VOI (14/45) and Alpha VOC (13/45) as the dominant variants. A total of 29/45 of these genomes exhibited amino acid (aa) substitutions in the viral spike (S) protein, totaling 275 substitutions. Six of these genome sequences carried key mutations, including H69del, V70del, Y144del, D614G, N501Y, A570D, P681H, T716I, and Q677H. These mutations are known to influence SARS-CoV-2 transmissibility, virulence, and potential resistance to neutralization by vaccine-induced antibodies/convalescent sera. Phylo-evolutionary analysis revealed three distinct clusters, 16 sequences from Lineage B clustered closely with the reference Wuhan-Hu-1 strain (NC_045512.2). Notably, 19 of our WGSs showed genomic relatedness to SARS-CoV-2

*Correspondence:
Waidi Folorunso Sule
waidifolorunso@uniosun.edu.ng

Full list of author information is available at the end of the article

strains from other regions of Nigeria, as well as, from West/other African countries, including Egypt, Senegal, Morocco and South Africa. The single L.3 lineage reported clustered with strains from Nigeria and Benin.

Conclusions These evidences suggest inbound or outbound transmission of the dominant Eta and Alpha variants. The study highlights the dominance of Eta VOIs and Alpha VOCs during the second COVID-19 wave in Osun State, Nigeria, suggesting potential inbound and outbound spread of the virus between other states in Nigeria and other African countries. The general absence of severe illness (except for four participants with shortness of breath) among study participants may indicate a protective effect of post-exposure immune response; however, underreporting of severe cases cannot be ruled out as a potential contributing factor.

Keywords SARS-CoV-2, Whole genome sequencing, Phylo-evolutionary analysis, VOI/VOC, Osun state

Background

In December 2019, a cluster of pneumonia cases of unknown aetiology among humans emerged in Wuhan, Hubei Province, China, linked to a seafood market. The causative agent, a novel virus, was subsequently designated as 2019 novel Coronavirus (2019 nCoV) [1- 3]. Following the availability of complete genome sequence [1], the Coronavirus Study Group (CSG) of the International Committee on Taxonomy of Viruses(ICTV) classified the virus as severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) due to its genomic similarity to SARS-CoV from 2002 outbreak [4, 5].

SARS-CoV-2 is a positive-sense, single-stranded RNA virus in the family *Coronaviridae*, genus *Betacoronavirus* in the subgenus: *Sarbecovirus* [1, 6, 7]. The reference strain of SARS-CoV-2 (GenBank ID: MN908947 or GISAID ID: EPI_ISL_402124) isolated from a patient sample in Wuhan (China) consists of 29,903 nucleotides and features a gene order of 5'-replicase ORF1ab-S-E-M-N-3', along with accessory genes typical of coronaviruses [8]. Like the previous two emerging coronaviruses, SARS-CoV-1 of 2002 and MERS-CoV of 2012, SARS-CoV-2 causes respiratory disease, spreading through close physical contact and respiratory droplets [9, 10]. Infection occurs via the viral spike (S) protein binding to angiotensin-converting enzyme 2 (ACE2) in human respiratory tract [11]. By November 10 2024, SARS-CoV-2 had spread to over 234 countries, areas and territories, causing 776.8 million confirmed infections, with over 7 million deaths [12]. In Nigeria, by December 2023, 267,173 cases and 3,155 deaths from COVID-19 were reported [13].

Like other RNA viruses, SARS-CoV-2 undergoes rapid genomic evolution during replication in human host cells leading to alterations in its phenotype [14–16]. As the virus replicates in infected individuals, it accumulates mutations across its genome, leading to over 2,000 genetic variants derived from the ancestral Wuhan strain [17, 18]. Mutations that confer a competitive advantage such as increased transmission, replication efficiency and immune evasion, may become fixed in the viral population through natural selection, impacting viral phenotype [15]. An example is the D614G mutation in the spike protein which enhances infectivity and stability of the spike protein which was first identified in the Alpha variant [19]. Some of the factors responsible for the evolutionary process of the virus include viral factors (mutation and genetic drift in the spike protein, runaway evolution due to novel selective pressure); host factors (e.g., genetic susceptibility and immune system dynamics); environmental factors (reservoir and host interaction, climate and geography) [16, 20, 21]. Though there was a period (about 8 months) of evolutionary stasis following the emergence of SARS-CoV-2, when the virus exhibited limited apparent genomic evolution; the virus thereafter underwent rapid mutations. The mutation, in the spike protein leading to increased virus transmissibility, immune evasion and severe disease impacts alongside the application of phylogenetic analysis documented during the pandemic led to identification and designation of SARS-CoV-2 variants/lineages [15, 22]. Variants with increased fitness and transmissibility contributed to successive COVID-19 pandemic waves [23].

The SARS-CoV-2 genomic evolution possess significant public health risks spanning across five main areas including transmissibility, pathogenicity, immunogenicity/vaccination, diagnostics and therapeutics [12, 22, 24, 25]. These challenges necessitated national genomic surveillance, with the WHO/CDC classifying lineages into variant of concern (VOC), variant of interest (VOI) and variant under monitoring (VUM) to prioritize public health responses and research [22, 26]). In addition, SARS-CoV-2 lineages are classified for epidemiological identification and tracking using platforms such as GISAID, Nextrain, PANGOLIN and Marseille [27]). PANGOLIN serves as the primary source of SARS-CoV-2 lineage nomenclature [28].

Of the 16 SARS-CoV-2 emerging lineages, four were earlier identified as VOC: Alpha (B.1.1.7) from the UK, Beta (B.1.351) from South Africa, Gamma (P.1/B.1.1.28) from Brazil, Delta (B.1.617.2) from India and Omicron (B.1.1.529) from South Africa [12, 29, 30]. The remaining lineages were classified as VOIs including Eta (B.1.525), Zeta (P.2), Epsilon (B.1.427), Theta (P.3), Iota (B.1.526),

Kappa (B.1.617.1), Lambda (C.37) and Mu (B.1.621) [22, 23, 31].

Molecular epidemiology has facilitated the identification of specific variant-defining mutations unique to VOCs and VOIs. For example, the Alpha VOC (B.1.1.7 lineage) which emerged in September 2020 in UK became the dominant variant for the period and was exported to other countries [32]. It harboured 17 mutations including 14 non-synonymous point mutations and three deletions. Eight mutations (Δ69–70 deletion, Δ144 deletion, N501Y, A570D, P681H, T716I, S982A, and D1118H) were located in the spike protein [33]. Of these, The N501Y mutation in the receptor-binding domain (RBD) enhanced ACE2 binding affinity [33, 34], while the P681H mutation adjacent to the furin cleavage site (FCS) increased transmissibility [35]. The H69/V70 deletion was associated with immune escape and diagnostic test kit failure [36]; similar variant-defining mutations were observed in other VOCs [18].

The application of Next Generation sequencing (NGS) technology and bioinformatics have provided more opportunities for rapid genome sequencing and of analysis of SARS-CoV-2. Platforms such as Illumina (MiSeq/HiSeq), PacBio, Ion Torrent, Oxford Nanopore and MGI sequencing have been widely used [37, 38]. The MGI platform, coupled with ATOPlex design and standardized library preparation kits, offers cost effective, high-quality and efficient sequencing [38]. Genome-scale sequence data have transformed molecular phylogenetics into phylogenomics, enabling the reconstruction of SARS-CoV-2 evolutionary history using WGS data deposited in GISAID database [26, 31, 37].

This study involved whole-genome sequencing and phylogenomics of SARS-CoV-2 in Osun state, Nigeria during the first half of 2021 to identify and characterize circulating SARS-COV-2 variants. We used MGI's DNBSEQ® sequencing platform and ATOPlex technology followed by bioinformatics analysis of sequence data. We aimed to determine the genomic relatedness of local variants to global reference strains, trace the sources and trajectories of virus introduction to inform public health response.

Methods

Study location, ethical approval, study design, and sample collection

This retrospective, molecular epidemiologic study was conducted in Osun state, southwest Nigeria (7.56289640°N, 4.51995930°E), predominantly populated by Yorubas, with Osogbo as its capital. Osogbo's population was 731,000 in 2021 [39]. The 60 samples were collected from eight locations (Ede, Gbongan, Igbajo, Ikire, Ile-Ife, Ilesa, Moro, and Osogbo) across 11 of the state's 30 Local Government Areas (LGAs) (Fig. 1), with 76.7% of samples from Osogbo. Ethical approval was obtained from the Health Research Ethics Committee, UNIOSUN (UNIOSUNHREC 2021/006B) and Osun State Health Research Ethical Committee [OSHREC], Sept., 2022). Informed consent/consent to participate was not applicable as the clinical samples were collected as part of the State's routine surveillance and management of symptomatic patients and asymptomatic individuals intending to travel abroad during the COVID-19 pandemic. The study was conducted in accordance with the Helsinki Declaration on use of human subjects in research. Confidentiality was ensured by using coded identifiers instead of sample names, with data securely stored and accessible only to authorized researchers. Samples (n = 60) for this study were randomly selected from archives (-80℃) being originally collected between January and June 2021 during routine surveillance and clinical evaluation of patients by the Epidemiology Unit of the Ministry of Health, Osun State during the second wave of the pandemic [40–42]. To ensure adequate sequencing reads and genome quality were recovered, only positive samples with Ct values ˂ 30 for *E*, *RdRp* and *N* genes were included. When a selected sample number in the Excel spreadsheet of positive samples happened to have higher than 30 Ct value, we de-selected and re-selected in no specific order until we had 60 positive samples. All samples were collected in line with NCDC/WHO approved protocol. Briefly, nasopharyngeal and nasal swabs were collected by State's Epidemiology Unit into viral transport medium and transported under a cold chain to the NCDC-accredited Molecular Diagnostics Unit of the Multidisciplinary Research Laboratory (MRL), Osun State University, Osogbo, Nigeria. RT-qPCR was conducted on each of the samples to further confirm presence of SARS-CoV-2 in the samples [43]. The Epidemiology Unit also collected each patient's clinical data using Case Report Form.

Viral RNA extraction from clinical samples

The archived confirmed COVID-19 nasopharyngeal samples were retrieved from − 80℃ storage and allowed to thaw on ice under a Biosafety Level II plus cabinet using appropriate PPE. A 200 µl aliquot from each sample was processed for total RNA extraction using QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. The viral RNA extracts were subsequently used for whole genome sequencing/molecular analyses (converted to cognate cDNA for MGI - DNBSEQ-G50RS Sequence Technology) in National Institute for Medical Research (NIMR), Yaba, Lagos, Nigeria as shown below.

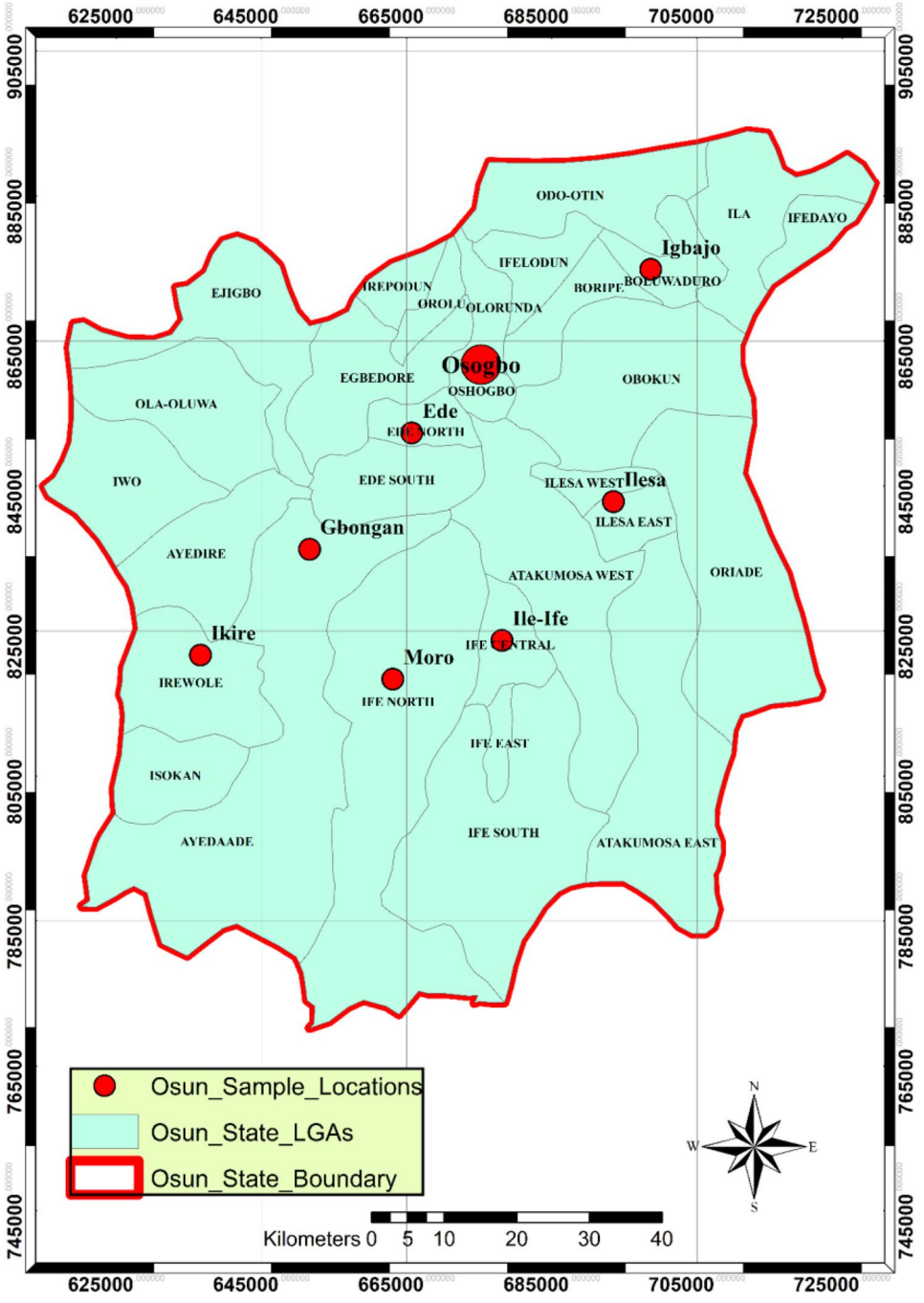

Fig. 1 Map of Osun State showing the sample locations; Osogbo with the biggest red sphere had the largest sample size

Reverse transcription and PCR amplification

Sixty viral RNA extracts were transcribed to corresponding cDNA using the MGI ATOPlex RNA Universal Library Prep Module [38]. The protocol was conducted in three steps using three separate designated laboratory areas to prevent aerosol and cross-contamination: Firstly, RNA (10 µl) extract were subjected to reverse transcription according to the manufacturer's protocol. Reaction conditions included: 25 °C/10 mins; 42 °C/30 mins; 70 °C/15 mins and held at 4 °C; first PCR mix was prepared on ice. Secondly, PCR amplification targeting the SARS-CoV-2 genome was performed using MGIEasy RNA Library Prep Kits following the manufacturer's instruction (PCR conditions: 1 cycle of 37 °C/5 mins and 95 °C/10 mins; 13 cycles of 95 °C/15 s, 64 °C/1 min, 60/1 min, 72 °C/30 s and hold at 4 °C); PCR products were cleaned, and second PCR mix was prepared on ice. This was followed by the introduction of dual barcoded

adaptors under these conditions: 1 cycle of 37 °C/5 mins and 95 °C/10 mins; 27 cycles of 95 °C/15 s, 64 °C/1 min, 60/1 min, 72 °C/30 s and held at 4 °C; The PCR products were cleaned up immediately for sample pooling and preparation of single stranded circular DNA (ssCirDNA).

Library Preparation (Generation of DNA nanoball for nucleotide Sequencing)

Quantitation of purified second PCR products was performed using Qubit® dsDNA HS Assay kit (ThermoFisher Scientific). Samples were pooled to a total of 400 ng concentration in a 48 µl volume, denatured (95 °C for 3 min and held at same temperature) and used for ssCirDNA synthesis with MGIEasy dual barcode circularization kit according to DNBSEQ-G50RS instructions (37 °C for 30 min, hold at 4 °C). The ssCirDNA libraries underwent clean-up step and quantitation using the Qubit® ssDNA Assay kit (ThermoFisher Scientific). DNA Nanoballs (i.e., final sequence libraries) were prepared using DNBSEQ-G50RS High-throughput sequencing reagents in a 3:1 ratio of sample DNA Nanoball to balanced library DNA Nanoball in a 100 µl volume before nucleotide sequencing.

Whole SARS-CoV-2 genome sequencing on MGI DNBSEQ-G50RS platform

Full-length viral genomes were sequenced using the DNBSEQ-G50RS platform with 100 bp paired-end sequencing (PE100). DNA Nanoballs were loaded onto a large flow cell (FCL) and sequenced with DNBSEQ-G50RS High-throughput PE100 reagents and Barcode Primer 3 according to manufacturer's instructions.

Bioinformatics analysis and phylogenomics

A bioinformatics analysis was performed using the MGI ATOPlex SARS-CoV-2 pipeline (https://github.com/MGI-tech-bioinformatics/SARS-CoV-2_Multi-PCR_v1.0). Reads generated from sequencing were demultiplexed and trimmed using the Trimmomatic algorithm, which is built into the MGI ATOPlex SARS-CoV-2 bioinformatics pipeline. Sequence quality was assessed using FASTQC tool. In the MGI ATOPlex SARS-CoV-2 bioinformatics pipeline, reads underwent referenced-based assembly by aligning them with the SARS-CoV-2 reference genome (Wuhan-Hu-1, MN908947.3) to generate a consensus sequences used for further analysis [38]. Phylogenomics analysis included 45 SARS-CoV-2 whole genome sequences (WGS) sequenced in this study (Supplementary File Table 1), reference genome (Wuhan-Hu-1; NC_045512.) and 710 high-coverage African SARS-CoV-2 WGS retrieved from GISAID. Sequences were aligned using MAFFT in Galaxy web tool (Galaxy Version 7.526) [44]. Phylogenomic relationships of nucleotide sequences were inferred using PhyML (PHYlogenetic inferences using Maximum Likelihood) and visualized using interactive Tree of Life (iTOL) v6.9.1 [45, 46]. Metadata for the 45 sequences are provided in Supplementary Table 1.

Data analysis

Quantitative data obtained from the study are presented with descriptive statistics (mode, median, interquartile range [IQR], percentage and odds ratio with 95% confidence interval [95% CI]). Inferential statistics to establish significant differences between variables was done with Mann–Whitney U-test.

Results

Study Location, participants' demographics and clinical data

The study involved 60 SARS-CoV-2 positive individuals residing in Osun State during the pre-COVID-19 vaccination era, coinciding with the second wave of the pandemic. Whole-genome sequencing of SARS-CoV-2 was, however, successful for 45 of the 60 nasopharyngeal samples. While metadata is provided in Supplementary File Table 1, the whole genome sequence data had been deposited in NCBI GenBank (Accession numbers shown in Supplementary File Table 2).

Participants aged 16–76 years (median: 44 years; interquartile range [IQR]: 61–32 years) included 29 males (median: 40 years; IQR: 59–34 years) and 16 females (median: 48.5 years; IQR: 64.5–24 years) who were comparable in median age ($p = 0.673$). Of these, 12 participants (7 males, 5 females) were aged 60–76 years (median: 65 years). None had a history of travel, contact with COVID-19 cases, or hospitalization. Clinical symptoms were reported in some participants: cough (10.0%), muscle aches (5.0%), shortness of breath (6.7%), nausea (3.3%), headache (5.0%), runny nose (3.3%) and sore throat (1.7%). One (1.7%) participant had diabetes mellitus; none (0.0%) had chronic lung, cardiovascular, renal, or liver disease, nor immunocompromised conditions. No mortality (0.0%) was recorded among the 60 participants (Table 1).

Lineage assignment and distribution of the study WGSs

The genomes represented five different GISAID clades: G ($n = 17$), GR ($n = 1$), GRY ($n = 9$), L ($n = 16$) and O ($n = 2$); or five PANGO lineages. Specifically, the G clade included B.1.525 ($n = 13$), B.1.1.7 ($n = 3$) and L.3 ($n = 1$); the GR was B.1.525; GRY comprised of B.1.1.7 ($n = 9$); L ($n = 16$) represented B lineage without sub-lineages; and the O included B.1 ($n = 1$) and B.1.1.7 ($n = 1$) (Fig. 2). Notably, ancestral lineages B and B.1 persisted from the first to the second wave. The dominant variants observed were Eta (14 B.1.525, VOI) and Alpha (13 B.1.1.7, VOC) with prevalence rates of 31.1% (95% CI: 17.6–44.6%)

Table 1 Demographic/Clinical information of the study participants

Demographic/Clinical information	No. involved	No. positive (%)
Travel history	60	0.0 (0.0)
Hospitalization	60	0.0 (0.0)
Contact with COVID-19 cases	60	0.0 (0.0)
Cough	60	6.0 (10.0)
Muscle ache	60	3.0 (5.0)
Runny nose	60	2.0 (3.3)
Sore throat	60	1.0 (1.7)
Shortness of breath	60	4.0 (6.7)
Nausea	60	2.0 (3.3)
Headache	60	3.0 (5.0)
Chronic lung disease	60	0.0 (0.0)
Chronic renal disease	60	0.0 (0.0)
Chronic liver disease	60	0.0 (0.0)
Cardiovascular disease	60	0.0 (0.0)
Diabetes mellitus	60	1.0 (1.7)
Immunocompromised state	60	0.0 (0.0)

and 28.9% (95% CI: 15.6–42.1%), respectively. Twelve of the 13 Alpha VOCs were from Osogbo; while 14 Eta VOIs originated from Osogbo (9 sequences), Ilesa (2 sequences) and one each from Ede, Ikire and Moro.

An analysis of the Eta VOI (B.1.525) and Alpha VOC (B.1.1.7) in relation to gender, age, symptoms, and viral load showed that among females, 8 out of 16 were infected with Eta VOI, While 4 out of 16 had Alpha VOC. For males, 6 out of 29 had Eta VOI, and 9 out of 29 had Alpha VOC. Of the 12 older individuals aged 60–76 years, 4 were infected with Eta VOI, and 4 with Alpha VOC. Among the 33 individuals aged 16–58 years, 10 (30.3%) were infected with Eta VOI, and 9 (27.3%) with Alpha VOC. Among the 15 symptomatic individuals, 6 (40.0%) were infected with Eta VOI, and 3 (20.0%) with Alpha VOC. Regarding viral load, among the 14 individuals with Ct values ≤ 21.00 for the E gene, 6 had Alpha VOC, and 2 had Eta VOI; of the 13 individuals with Ct values ≤ 21.00 for the RdRp gene, 5 had Alpha VOC, and 3 had Eta VOI; and among the 15 individuals with Ct values ≤ 21.58 for the N gene, 6 had Alpha VOC, and 3 had Eta VOI.

Mutation and variant analyses

Amino acid (AA) substitution data were available for 29 (64.4%) of the WGSs with genome sizes ranging from 29.823 to 29.888 Kb. Mutations occurred over the entire genome length of the WGSs (Fig. 3), and analysis of the S proteins showed a total of 275 AA substitutions (range of 2–57, with 4 as median AA substitutions); nine of the sequences had 8–57 AA substitutions (Fig. 4). Of these, six sequences (3 Alpha VOC and 3 Eta VOI) had defining mutations (Table 2). Five variants had Y144del, and four had H69del and V70del, but no mutation was shared between Eta and Alpha variants. The 16 B lineage genomes (Clade L) lacked observable mutations, except for one L.3 lineage. No significant difference in the median AA substitutions was found between Eta VOI (median = 4) and Alpha VOC (median = 3, $p = 0.62$).

Phylogenomics of the study SARS-CoV-2

Phylogenomics of the 45 SARS-CoV-2 WGSs identified three distinct clusters (Fig. 5). The B lineages (Clade Ls, 87.5% sampled from Osogbo) clustered with the reference Wuhan-Hu-1 strain (NC_045512.2) and three other lineages (1 B.1, 1 B.1.1.7 and 1 B.1.525, also from Osogbo]). Then a strain B.1.1.7 (Clade GRY [hCoV-19/Nigeria/MRL-OS-1259/2021 sampled from Osogbo]) formed a separate divergence as second cluster; while the third cluster comprised 25 strains sampled from Osogbo (majority), Ile-Ife, Ilesa, Ede, Ikire, Moro and Gbongan.

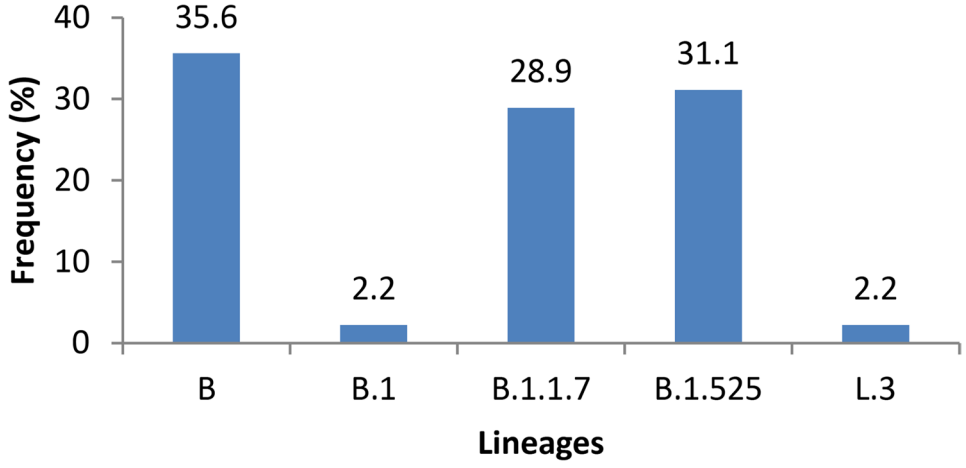

Fig. 2 Frequency distribution of PANGO lineages of SARS-CoV-2 identified from 45 genome sequences in Osun State, Nigeria, 2021. The chart depicts the prevalence of major lineages (e.g., B.1.1.7 and B.1.525) and their proportions within the dataset

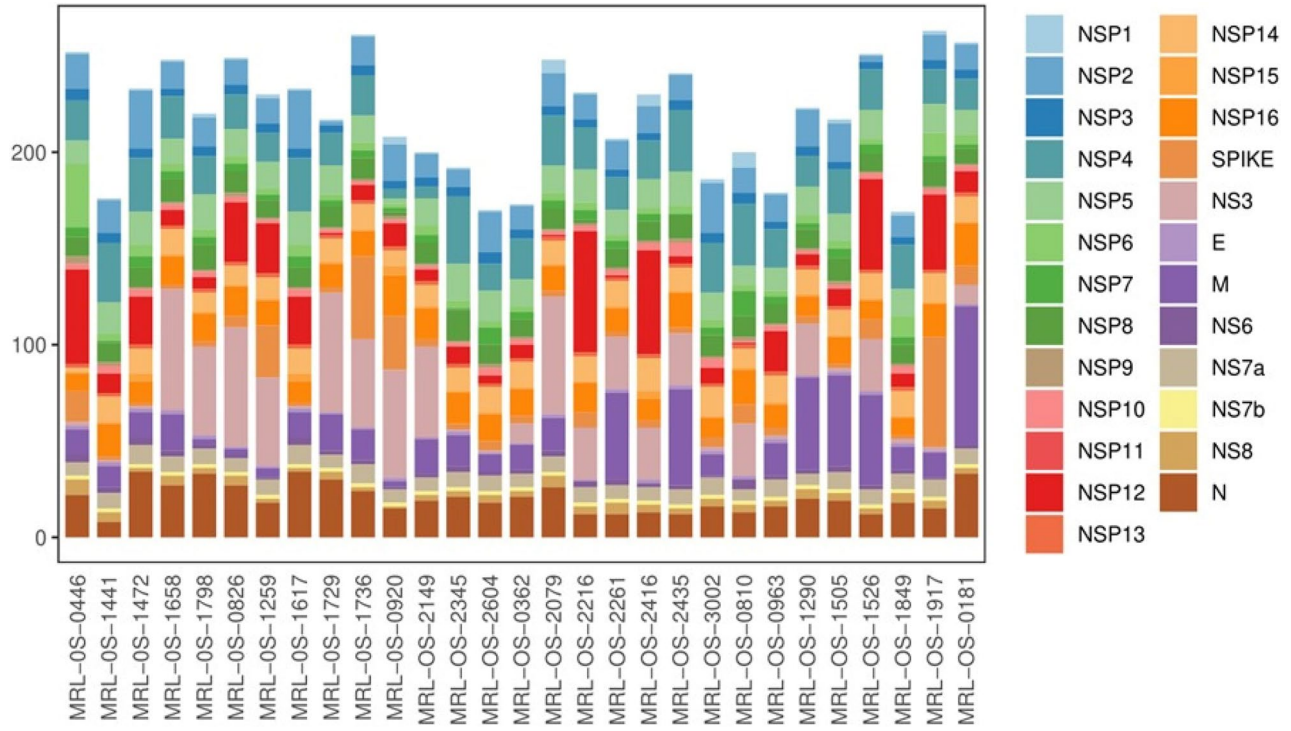

Fig. 3 Observable genome-length mutations in the NSP1-16, structural and accessory proteins of the study SARS-CoV-2 ($n = 29$)

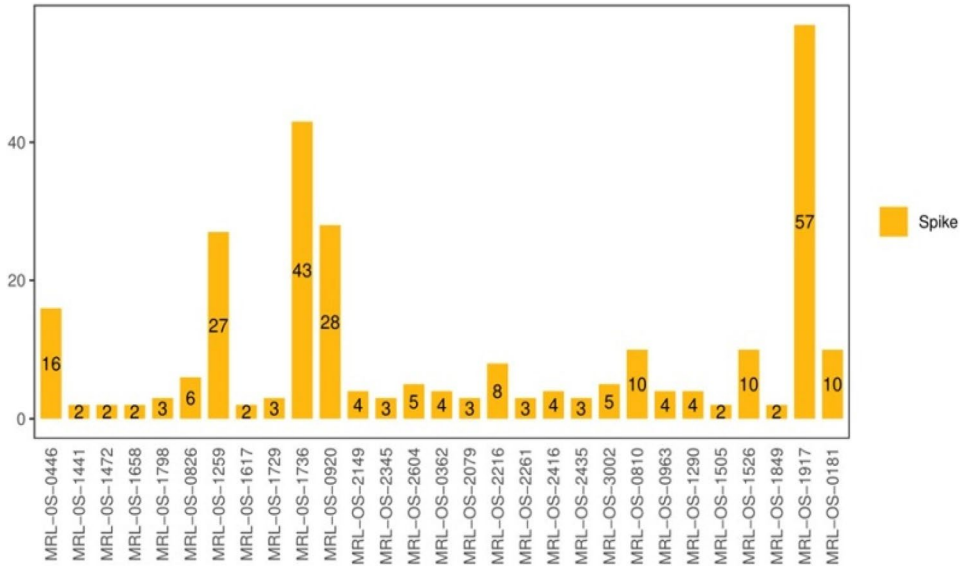

Fig. 4 Number of amino acid substitutions observed in the S protein of the study SARS-CoV-2 ($n = 29$)

Table 2 VOCs/VOIs - defining mutations in the S protein of the study SARS-CoV-2

S/No	Strains	Lineages	Mutations
1.	MRL-0 S-0446	B.1.1.7	Y144del
2.	MRL-0 S-1259	B.1.1.7	H69del, V70del, Y144del
3.	MRL-OS-1917	B.1.1.7	H69del, V70del, Y144del, D614G, E484K, Q677H, Q52R
4.	MRL-0 S-1736	B.1.525	H69del, V70del, Y144del, D614G, N501Y, A570D, P681H, T716I
5.	MRL-0 S-0920	B.1.525	H69del, V70del, Y144del
6.	MRL-OS-2416	B.1.525	Q52R

When the 45 WGSs were included with 710 other African strains, they formed three separate clusters (Fig. 6 and Supplementary Figure S1). The 16 B lineage strains (i.e., Clade Ls) expectedly showed relationship with the reference Wuhan-Hu-1 strain (NC_045512.2) and other strains from Senegal, Nigeria, Morocco, South Africa and Egypt. There were three strains of ours (hCov-19/Nigeria/MRL-OS-0446/2021; hCov-19/Nigeria/MRL-OS-0920/2021 and hCov-19/Nigeria/MRL-OS-0810/2021) sampled from Osogbo that clustered with Egypt, Morocco and Senegal. The remaining strains formed a separate cluster.

It is noteworthy that only one L.3 lineage (EPI_ISL_5367954 [hCoV19/Nigeria/MRL-OS-0181/2021]), sampled from Osogbo occurred in our WGSs; a phylogenomic analysis (Fig. 7) of this with 130 Nigerian and other African sequences showed that it clustered with 7 other Nigerian and two Benin strains.

The Phylogenetic tree was constructed with the neighbor-joining algorithm with evolutionary distance computed using the Kimura two-parameter model to recover their clustering pattern. Multiple sequence alignment was performed with MUSCLE in MEGA v11. The numbers at each node represent bootstrap values for 1000 replicates.

Multiple sequence alignment was performed with MAFFT in the Galaxy webtool. Phylogenomic relationships of nucleotide sequences were inferred using PhyML (PHYlogenetic inferences using Maximum Likelihood) and visualized using interactive Tree of Life (iTOL) v6.9.1.

Phylogenetic tree of the study's L.3 SARS-CoV-2 (red ID) in relation to other African L.3 strains. The Phylogenetic tree was constructed with the neighbour-joining algorithm with evolutionary distance computed using the Kimura two-parameter model to recover their clustering pattern. Multiple sequence alignment was performed with MAFFT in the Galaxy webtool. The numbers at each node represent bootstrap values for 1000 replicates. This analysis involved 106 nucleotide sequences.

Discussion

Whole genome sequencing has been recognized as a pivotal tool in clinical research and epidemiology just before the COVID-19 pandemic, enhancing our understanding of viral evolution, transmission and vaccine development [47]. This study identified circulating SARS-CoV-2 variants during Nigeria's second wave of the COVID-19 pandemic in Osun State, characterized the mutations and assessed genomic relatedness with strains from within and beyond Nigeria.

From 60 samples across eight different locations in Osun State, 45 were successfully sequenced, revealing a predominance of Eta (B.1.525) and Alpha (B.1.1.7) lineages, consistent with previous studies [42, 48–52]. The ancestral B lineage (35.6%) and the rare L.3 lineage (2.2%) were also detected, with the latter restricted in transmission. While Eta originated locally and expanded geographically, Alpha likely arrived via international transmission before 2021 [28, 53]. Consistent with our findings, several studies from Nigeria and other African regions (e.g., Benin, Burkina Faso, Côte d'Ivoire, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Sen egal, Togo, Zambia) have reported that the second pandemic wave was driven predominantly by Alpha and Eta variants [42, 48–51]. In addition, similar ancestral lineages (B and B.1) were reported during Nigeria's first pandemic wave [42], while the L.3 lineage, also reported by Olawoye et al. [42]. and Olorunfemi et al. [52], was detected once in our study, clustering with other strains in multiple locations. However, its apparent limited transmission highlights the need for further investigation into its virological and clinical characteristics.

Despite 64.4% of participants being male and 26.7% aged 60 – 76 years, including cases with comorbidities, no severe illness or mortality was observed, possibly due to widespread occurrence of convalescent sera [12, 54]. Although the sample size in this study was limited, the observed higher burden among males aligns with findings reported by Taboada et al. [55] in Mexico and Yadav et al. [56] in India. However, this contrasts with the study by Katowa et al. [50], which documented a higher disease burden among females in the Southern Province of Zambia. These discrepancies highlight the importance of regional, demographic, and possibly socio-cultural factors influencing disease dynamics. Despite the small sample size, the consistency of male predominance in multiple studies supports the robustness of our observations while underscoring the need for larger, more diverse cohorts to validate these findings.

Phylogenomic analyses revealed three distinct clusters. The ancestral B lineage clustered monophyletically with Wuhan-Hu-1, while B.1, B.1.1.7, and B.1.525 displayed varying relatedness. The L.3 lineage clustered with other Nigerian and Beninese strains, suggesting

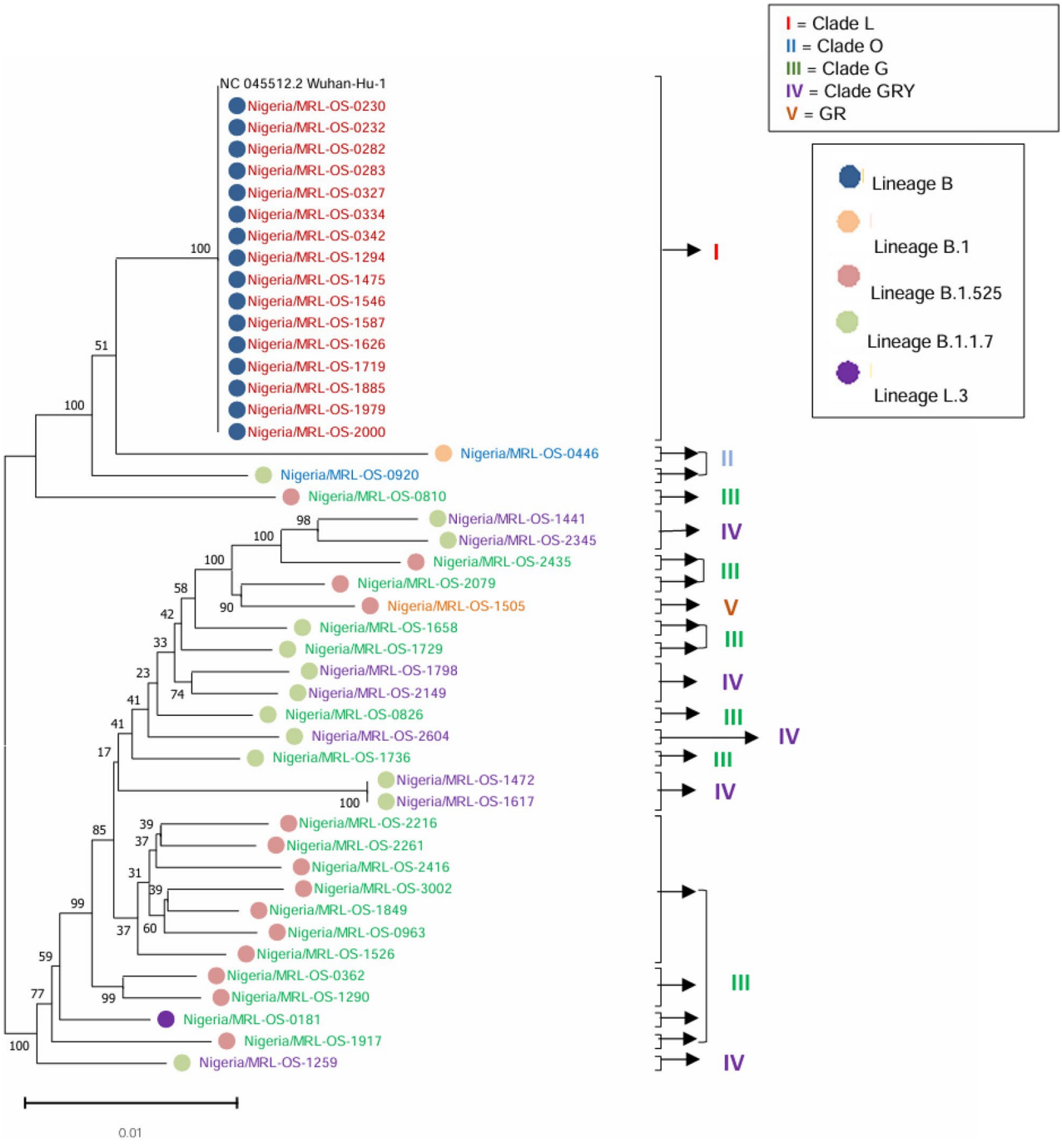

Fig. 5 Phylogenetic tree of the study's 45 SARS-CoV-2 strains

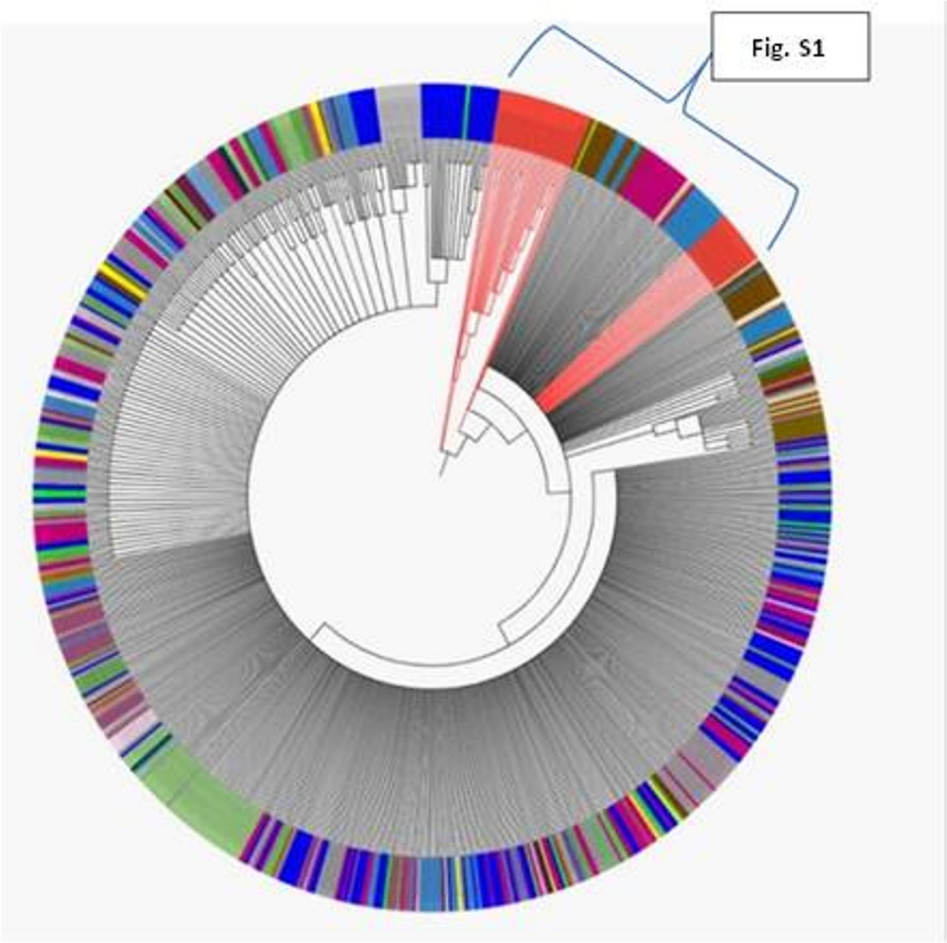

Fig. 6 Phylogenetics of the study's 45 SARS-CoV-2 in relation to other 710 African WGSs; Color codes show strains from our MRL and different geographical areas in Africa

limited cross-border transmission; similar observation that Nigerian L.3 clustered with strains from Russia, the Philippines, Australia, and Japan [52]. Continental relatedness was evident in our study, with strains clustering with samples from Senegal, Morocco, Egypt, and South Africa, pointing to potential travel-linked transmissions despite unclear participant travel histories.

Mutation analysis identified 275 AA substitutions in the Spike protein, many associated with increased infectivity, transmission, or immune evasion [34, 57]. Notable substitutions included D614G, N501Y, E484K, and Q677H, consistent with previous reports [35, 58]. Variants with Eta/Alpha-defining mutations (e.g., H69del, V70del) further underscore their role in the second wave's dynamics.

While it was expected that the ancestral B lineage clustered with the reference strain as there was no evidence of AA substitutions in the B lineage; that lineages B.1, B.1.1.7 and B.1.525 with 16, 28 and 10 AA substitutions, respectively in S protein clustered with the B lineage in Osogbo could not be explained. It was also surprising that a lineage B.1.1.7 with 27 AA substitutions in S protein, sampled in Osogbo formed a separate cluster as a divergence. The third cluster was also a monophyletic but comprised B.1.1.7, B.1.525 and surprisingly the only L.3 lineage having 10 AA substitutions in the S protein. These observations demand further studies with larger sample size in Osun State/Nigeria.

While we could not state with certainty the direction of spread of the strains with other countries, the phylogenomics clearly showed relatedness with SARS-CoV-2 circulating in other African countries (Fig. 6 and Supplementary Figure S1), especially the West Africa sub-region. The ancestral B lineage majorly sampled in Osogbo clustered with strains from Senegal, Morocco, South Africa and Egypt, as well as, some from Nigeria.

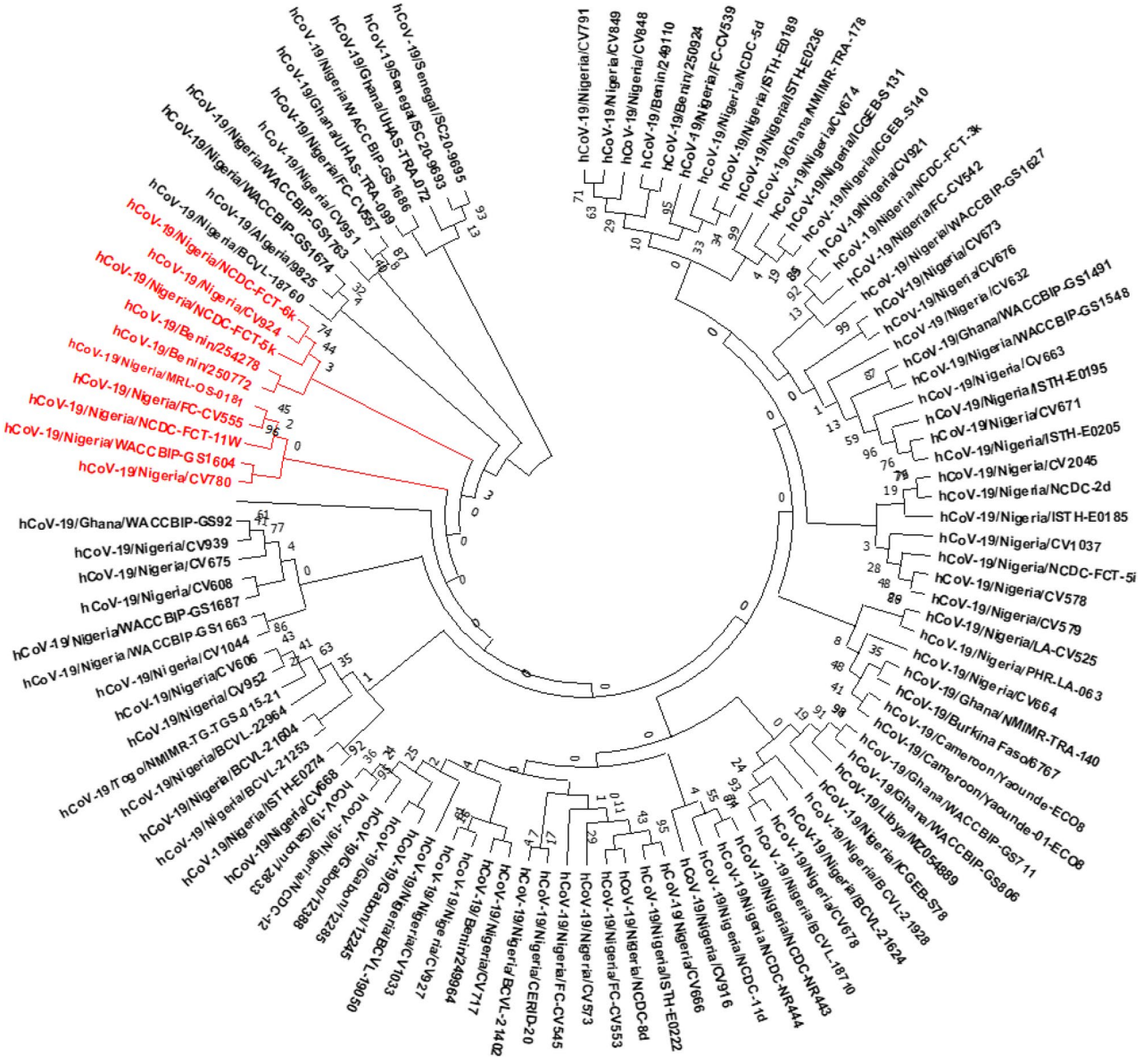

Fig. 7 Phylogenetic tree of the study's L.3 SARS-CoV-2 (red ID) in relation to seven other Nigerian and Benin strains

There were also three strains that clustered with some from Egypt, Morocco and Senegal; the occurrence of clustering of some of ours with those from Egypt, Morocco and Senegal pointed to either outbound or inbound travels/transmission, this however, remains a puzzle as there was no history of travel to those countries among the study participants; we could not state for certain either if there was history of travellers from the three countries into Osun State during the sampling period.

That the singular L.3 lineage sampled from Osogbo clustered with 7 other Nigerian strains and two Benin strains indicated possible cross-country transmission of the pandemic virus.

Conclusions

The study highlights the dominance of Eta VOIs and Alpha VOCs during the second COVID-19 wave in Osun State, Nigeria, suggesting potential inbound and outbound spread of the virus between other states in Nigeria and other African countries. The general absence of severe illness (except for four participants with shortness of breath) among study participants may indicate a protective effect of COVID-19 convalescent sera; however, underreporting of severe cases cannot be ruled out as a potential contributing factor. We recommend strengthening regional genomic data sharing among African countries to rapidly identify and control inbound/outbound SARS-CoV-2 variant spread.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-025-02286-2.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements
We appreciate and acknowledge the assistance of Multidisciplinary Research Laboratory, Osun state University; the Epidemiology Unit of Osun State's Ministry of Health and Central Sequencing Laboratory, NIMR, Yaba, Lagos, Nigeria.

Authors' contributions
Conceptualization and study design: SBA, RAB, WFS, OOO, BAI, HAD; Sample processing: TSO, OPO, OYA, TEO, BAI, TFA, AAA; Data collection and curation: OOO, FMA, OYA, OOO, BAI, EKO, DOO, AAA, JOO OO; Data and bioinformatics analyses: OOO, OO, TSO, OPO, TEO, BAI, ESF, TFA, AAA; Manuscript development: SBA, OOO, OPO, WFS, OOO, BAI, EKO. All authors reviewed the manuscript.

Funding
The study was self-funded.

Data availability
The viral nucleotide sequences generated were deposited to the NCBI GenBank (with accession numbers PX435292.1 - PX435307.1; and PX588385, PX588386 and PX588388) and GISAID.

Declarations

Ethics approval and consent to participate
Ethical approval to conduct the study was obtained from Health Research Ethics Committee, Osun State University, Osogbo, Osun State, Nigeria (UNIOSUNHREC 2021/006B) and Osun State Health Research Ethical Committee [OSHREC], Sept., 2022). The need for Informed consent/consent to participate was waived by Ethics Committee/IRB of Osun State Health Research Ethical Committee [OSHREC], Osun State Ministry of Health, Osun State, Nigeria. The study adheres to the tenets of Helsinki Declaration as applicable to use of human subjects in research.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Microbiology, Osun State University, Osogbo, Osun State, Nigeria
[2]Multidisciplinary Research Laboratory, Osun State University, Osogbo, Osun State, Nigeria
[3]Department of Biotechnology, Osun State University, Osogbo, Osun State, Nigeria
[4]Department of Haematology, Obafemi Awolowo University Teaching Hospital Complex, Ile-Ife, Osun State, Nigeria
[5]Department of Haematology and Immunology, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria
[6]Department of Public Health, Osun State Ministry of Health, Osogbo, Osun State, Nigeria
[7]Department of Natural Sciences, Albany State University, Albany, USA
[8]Division of Molecular Biology and Biotechnology, Nigerian Institute of Medical Research (NIMR), Yaba, Lagos, Lagos State, Nigeria
[9]Department of Microbiology, Laboratory of Molecular Biology, Immunology and Bioinformatics, Adeleke University, Ede, Osun State, Nigeria
[10]Division of Vaccine Design and Pharmacotherapies, Helix Biogen Institute, Ogbomoso, Oyo State, Nigeria
[11]Department of Veterinary Microbiology, University of Ibadan, Ibadan, Oyo State, Nigeria
[12]Biomedical and Biotechnology Workgroup, Alliance Global (AGBL Group), Lagos, Lagos State, Nigeria
[13]Department of Urban and Regional Planning, Osun State University, Osogbo, Osun State, Nigeria
[14]Department of Microbiology, Rivers State University, Port-Harcourt, Rivers State, Nigeria

Received: 2 May 2025 / Accepted: 1 December 2025
Published online: 13 December 2025

References

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265–9.
2. WHO. Coronavirus disease 2019 (COVID-19). Situation report 45. Geneva, Switzerland: 2020[a]. Available at: https://www.who.int/docs/default-source/coronaviruse/situationreports/20200305-sitrep-45-covid-19.pdf?sfvrsn=ed2ba78b_4.
3. WHO. Naming the coronavirus disease (COVID-19) and the virus that causes it. 2020[b]. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it.
4. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020. https://doi.org/10.1038/s41564-020-0695-z.
5. Gorbalenya AE, Baker SC, Baric RS, Groot RJ, de Drosten C, Gulyaeva AA et al. severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. bioRxiv. 2020; (p. 2020.02.07.937862). https://doi.org/10.1101/2020.02.07.937862.
6. Masters PS, Perlman S. Coronaviridae. In: Knipe DM and Howley PM, editors. Fields Virology, 6th Edn. Lippincott Williams & Wilkins, Philadelphia; 2013;825–858.
7. International Committee on Taxonomy of Viruses (ICTV). 2020. https://talk.ictvonline.org/.
8. WHO. Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health. Geneva. Licence: CC BY-NC-SA 3.0 IGO. 2021[a].
9. Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. Nat Rev Immunol. 2020;20(6):363–74.
10. WHO. Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases: interim guidance. 2020[c]. (Updated on March 19, 2020). https://www.who.int/publications-detail/laboratory-testing-for-2019-novel-coronavirus-in-suspected-human-cases-20200117.
11. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020;581(7807):221–4.
12. WHO. COVID-19 Epidemiological Update. Special Edition 174 published 24 December 2024. 2024[a]. Accessed 13 March 2025. https://www.who.int/publications/m/item/covid-19-epidemiological-update---24-december-2024.
13. WHO. The Current COVID-19 Situation. 2024[b]. https://www.who.int/countries/nga/2024. Accessed 14 March 2025.
14. Lauring AS, Hodcroft EB. Genetic variants of SARS-CoV-2 - What do they mean? JAMA. 2021;325(6):529–31. https://doi.org/10.1001/jama.2020.27124.
15. Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, et al. The evolution of SARS-CoV-2. Nat Rev Microbiol. 2023;21:361–9.
16. Manirambona E, Okesanya OJ, Olaleke NO, Oso TA, Lucero-Prisno IIIDE. Evolution and implications of SARS-CoV-2 variants in the post-pandemic era. Discover Public Health. 2024;21:16.
17. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of epidemiological lineages in an emerging pandemic using the Pangolin tool. Virus Evol. 2021;7(2):veab064. https://doi.org/10.1093/ve/veab064.
18. Choi JY, Smith DM. SARS-CoV-2 variants of concern. Yonsei Med J. 2021;62(11):961–8.
19. Public Health England. Investigation of novel SARS-CoV-2 variant: Variant of concern 202012/01: Technical Briefing 3. 2021; available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959360/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3.pdf.

20. Kerner G, Quintana-Murci L. The genetic and evolutionary determinants of COVID-19 susceptibility. Eur J Hum Genet. 2022. https://doi.org/10.1038/s41431-022-01141-7.
21. Guerrero BG, Vázquez JR, Ledezma JCR. Balance and genetic mechanisms in the COVID-19 process of mutation. Asian J Res Infect Dis. 2023. https://doi.org/10.9734/ajrid/2023/v13i4275.
22. WHO, Tracking. SARS-CoV-2 variants. 2021.Available at: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/.
23. Carattini YL. Characterization of Emerging Variants of SARS-CoV-2 and the Progressive Disappearance of the Parental Wuhan Strain. PhD Dissertation, University of Miami, Floria, USA. 2022.
24. WHO. SARS-CoV-2 genomic sequencing for public health goals. Interim Guidance. Geneva. 2021[b].
25. WHO. Guidance for surveillance of SARS-CoV-2 variants. Interim guidance. Geneva. 2021[c].
26. Zhao N, Zhou N, Fan H, Ding J, Xu X, Dong X, et al. Mutations and phylogenetic analyses of SARS-CoV-2 among imported COVID-19 from abroad in Nanjing. China Front Microbiol. 2022;13:851323.
27. Colson P, Fournier PE, Chaudet H, Delerce J, Giraud-Gatineau A, Houhamdi L, et al. Analysis of SARS-CoV-2 variants from 24,181 patients exemplifies the role of globalization and zoonosis in pandemics. Front Microbiol. 2022;12:786233.
28. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020;5(11):1403–07.
29. Wang L, Cheng G. Sequence analysis of the emerging SARS-CoV-2 variant Omicron in South Africa. J Med Virol. 2022;94:1728–3.
30. Fischer C, Maponga TG, Yadouleton A, Abílio N, Aboce E, Adewumi P, et al. Emergence and spread of the SARS-CoV-2 Omicron (BA.1) variant across africa: an observational study. Lancet Glob Health. 2025;13:e256–67.
31. Barona-Gómez F, Delaye L, Díaz-Valenzuela E, Plisson F, Cruz-Pérez A, Díaz-Sánchez M, et al. Phylogenomics and population genomics of SARS-CoV-2 in Mexico during the pre-vaccination stage reveals variants of interest B.1.1.28.4 and B.1.1.222 or B.1.1.519 and the nucleocapsid mutation S194L associated with symptoms. Microb Genomics. 2021;7(11):000684.
32. Chaillon A, Smith DM. Phylogenetic analyses of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) B.1.1.7 lineage suggest a single origin followed by multiple exportation events versus convergent evolution. Clin Infect Dis. 2021;73(12):2314–17.
33. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD et al. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science. 2021; 372(6538): eabg3055.
34. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. Cell. 2020;182(5):1295–e1020.
35. Peacock TP, Goldhill DH, Zhou J, Baillon L, Frise R, Swann OC, et al. The Furin cleavage site in the SARS-CoV-2 Spike protein is required for transmission in ferrets. Nat Microbiol. 2021;6(7):899–09.
36. Meng B, Kemp SA, Papa G, Datir R, Ferreira IATM, Marelli S, et al. Recurrent emergence of SARS-CoV-2 Spike deletion H69/V70 and its role in the alpha variant B.1.1.7. Cell Rep. 2021;35(13):109292. https://doi.org/10.1016/j.celrep.2021.109292.
37. Young AD, Gillung JP. Phylogenomics – principles, opportunities and pitfalls of big-data phylogenetics. Syst Entomol. 2020;45:225–47.
38. Nigerian Institute of Medical Research (NIMR) Central Sequencing Laboratory (CRL) Technical Team. SARS-CoV-2 Complete genome sequencing protocol. 2021; https://www.nimr.gov.ng.
39. National Population Commission of Nigeria. Nigeria's Population by State (2021 Estimates). National Population Commission of Nigeria. 2021.
40. Akande OW, Elimian KO, Igumbor E, Dunkwu L, Kaduru C, Olopha OO, et al. Epidemiological comparison of the first and second waves of the COVID-19 pandemic in Nigeria, February 2020 - April 2021. BMJ Global Health. 2021;6(11):e007076.
41. Tsueng G, Mullen JL, Alkuzweny M, Cano M, Rush B, Haag E, et al. Outbreak.info research library: a standardized, searchable platform to discover and explore COVID-19 resources. Nat Methods. 2023;20(4):536–40.
42. Olawoye IB, Oluniyi PE, Oguzie JU, Uwanibe JN, Kayode TA, Olumade TJ, et al. Emergence and spread of two SARS-CoV-2 variants of interest in Nigeria. Nat Commun. 2023;14(1):811. https://doi.org/10.1038/s41467-023-36449-5.
43. Shabani M, Nejati A, Yavarian J, Sadeghi K, Zadheidar S, Ahmadi AS, et al. The trend of phylogenetic and epitope variations of SARS-CoV-2 Omicron sublineages in Iran. Front Microbiol. 2025;15:1531712. https://doi.org/10.3389/fmicb.2024.1531712.
44. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
45. Guindon S, Lethiec F, Duroux P, Gascue O. 2005. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Research. 2005; 33: W557–W559.
46. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52(W1):W78–82.
47. Houldcroft CJ, Beale MA, Breuer J. 2017. Clinical and biological insights from viral genome sequencing. Nature Reviews Microbiology. 2017; 15(3): 183–2.
48. Sander AL, Yadouleton A, Filho de O, Tchibozo EF, Hounkanrin C, Badou G, et al. Mutations associated with SARS-CoV-2 variants of Concern, Benin, early 2021. Emerg Infect Dis. 2021;27(11):2889–03.
49. Ozer EA, Simons LM, Adewumi OM, Fowotade AA, Omoruyi EC, Adeniji JA, et al. Multiple expansions of globally uncommon SARS-CoV-2 lineages in Nigeria. Nat Commun. 2022;13(1):688. https://doi.org/10.1038/s41467-022-28317-5.
50. Katowa B, Kalonda A, Mubemba B, Matoba J, Shempela DM, Sikalima J, et al. Genomic surveillance of SARS-CoV-2 in the Southern Province of zambia: detection and characterization of Alpha, Beta, Delta, and Omicron variants of concern. Viruses. 2022;14(9):1865.
51. Anoh EA, Wayoro O, Monemo P, Belarbi E, Sachse A, Wilkinson E, et al. Subregional origins of emerging SARS-CoV-2 variants during the second pandemic wave in Côte d'ivoire. Virus Genes. 2023;59(3):370–6.
52. Olorunfemi AB, Suliman SAR, Tran TT, Ayorinde B, Fowora MA, Iwalokun BA, et al. Whole genome sequencing and phylogenetic analysis of SARS-CoV-2 strains isolated during the COVID-19 pandemic in Nigeria. IJID. 2024;10:174–8.
53. Wilkinson E, Giovanetti M, Tegally H, San JE, Lessells R, Cuadros D, et al. A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. Science. 2021;374(6566):423–41.
54. Tobian AAR, Cohn CS, Shaz BH. COVID-19 convalescent plasma. Blood. 2022;140(3):196–07.
55. Taboada B, Zárate S, Iša P, Boukadida C, Vazquez-Perez JA, Muñoz-Medina JE, et al. Genetic analysis of SARS-CoV-2 variants in Mexico during the first year of the COVID-19 pandemic. Viruses. 2021;13(11):2161. https://doi.org/10.3390/v13112161.
56. Yadav PD, Nyayanit DA, Majumdar T, Patil S, Kaur H, Gupta N, et al. An epidemiological analysis of SARS-CoV-2 genomic sequences from different regions of India. Viruses. 2021;13(5):925. https://doi.org/10.3390/v13050925.
57. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, et al. Escape from neutralizing antibodies by SARS-CoV-2 Spike protein variants. eLife. 2020;9:e61312. https://doi.org/10.7554/eLife.61312.
58. Wahid M, Jawed A, Mandal RK, Dailah HG, Janahi EM, Dhama K, et al. Variants of SARS-CoV-2, their effects on infection, transmission and neutralization by vaccine-induced antibodies. Eur Rev Med Pharmacol Sci. 2021;25(18):5857–64.

Publisher's note