phrp
OSONG PUBLIC HEALTH AND RESEARCH PERSPECTIVES

pISSN 2210-9099 eISSN 2233-6052
https://doi.org/10.24171/j.phrp.2021.0329
Osong Public Health Res Perspect 2022;13(2):101-113

Original Article

Phylogenetic and genome-wide mutational analysis of SARS-CoV-2 strains circulating in Nigeria: no implications for attenuated COVID-19 outcomes

Daniel B. Kolawole, Malachy I. Okeke

Department of Natural and Environmental Sciences, Biomedical Science Concentration, School of Arts and Sciences, American University of Nigeria, Yola, Nigeria

ABSTRACT

Objectives: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of coronavirus disease 2019 (COVID-19). The COVID-19 incidence and mortality rates are low in Nigeria compared to global trends. This research mapped the evolution of SARS-CoV-2 circulating in Nigeria and globally to determine whether the Nigerian isolates are genetically distinct from strains circulating in regions of the world with a high disease burden.

Methods: Bayesian phylogenetics using BEAST 2.0, genetic similarity analyses, and genome-wide mutational analyses were used to characterize the strains of SARS-CoV-2 isolated in Nigeria.

Results: SARS-CoV-2 strains isolated in Nigeria showed multiple lineages and possible introductions from Europe and Asia. Phylogenetic clustering and sequence similarity analyses demonstrated that Nigerian isolates were not genetically distinct from strains isolated in other parts of the globe. Mutational analysis demonstrated that the D614G mutation in the spike protein, the P323L mutation in open reading frame 1b (and more specifically in NSP12), and the R203K/ G204R mutation pair in the nucleocapsid protein were most prevalent in the Nigerian isolates.

Conclusion: The SARS-CoV-2 strains in Nigeria were neither phylogenetically nor genetically distinct from virus strains circulating in other countries of the world. Thus, differences in SARS-CoV-2 genomes are not a plausible explanation for the attenuated COVID-19 outcomes in Nigeria.

Keywords: Africa; COVID-19; Mutation; Nigeria; Phylogenetic analysis; SARS-CoV-2

Received: December 1, 2021
Revised: January 27, 2022
Accepted: March 28, 2022

Corresponding author:
Malachy I. Okeke
Department of Natural and Environmental Sciences, Biomedical Science Concentration, School of Arts and Sciences, American University of Nigeria, 98 Lamido Zubairu Way, PMB 2250 Yola, Adamawa State, Nigeria
E-mail: malachy.okeke@aun.edu.ng

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiological agent of coronavirus disease 2019 (COVID-19). This disease was first documented in Wuhan, Hubei

Province, China in December 2019 and was later recognized as a pandemic by the World Health Organization (WHO) in March of the following year [1]. The first documentation of COVID-19 in Nigeria was on February 27, 2020, after an Italian citizen who worked in Lagos, Nigeria, returned from Milan, Italy on February 25 [2].

The COVID-19 outbreak in Sub-Saharan Africa and Nigeria has not mirrored that of the rest of the world despite realities such as poor health care systems for testing, reporting, tracing, and treating cases, lack of access to clean water, poor living conditions that make social distancing difficult, and the high burden of comorbidities and infectious diseases such as HIV [3–7]. Among African countries, Nigeria and a few others have been recognized as particular hotspots for the importation of COVID-19 from China [8,9]. Even the WHO had major concerns for Africa, estimating a death toll of up to 190,000 people [10]. However, this projection has not yet materialized, and Nigeria and most Sub-Saharan African countries have had relatively few cases of infection, morbidity, and mortality from COVID-19 according to statistical data provided by the WHO [11]. The low level of COVID-19 testing in Nigeria may partly account for this disparity. However, this study did not seek to test this hypothesis; instead, it considered another factor that may have contributed to the attenuated disease outcomes— namely, mutations. Mutations are random in nature, and disadvantageous mutations are more likely than advantageous mutations. It is worth considering whether the strains that entered Nigeria were naturally less transmissible and/or fatal than those that have circulated elsewhere in the world. A first step in examining this hypothesis is to compare the genomes of SARS-CoV-2 circulating in Nigeria with virus genomes from other parts of the world. Therefore, this study sought to characterize the identity of SARS-CoV-2 strains isolated within Nigeria by comparative phylogenetic and mutational analysis with strains obtained elsewhere in the world.

Materials and Methods

Data Acquisition and Sequence Alignment

SARS-CoV-2 genomes were obtained from the Global Initiative to Share All Influenza Data (GISAID) website (http://gisaid.com). A total of 230 sequences were collected; 100 of them were Nigerian sequences; which were randomly picked (from a total of 408 complete and high-coverage Nigerian SARS-CoV-2 genomes as of May 29, 2021), while the other 130 sequences, which included the reference Wuhan sequence (EPI_ISL_402125), were randomly picked from the rest of the world with a major focus on the 31 countries analysed in Table 1 along with Nigeria. The sequence IDs (GISAID, Pango, and Nextclade) of the 230 genomes, as well as the acknowledgement list of the sequence submitters of all 230 isolates, are shown in the supplementary data (Tables S1 and S2). Although the sequences were randomly chosen, only complete and high-coverage genomes were selected for analyses. Two sequence groups were created; the first contained only the Nigerian isolates and the reference sequence, and the other contained all isolates. The sequence groups were then aligned using MAFFT (for multiple alignment using fast Fourier transformation) [12]. After alignment, the Nigerian sequence group had a length of 29,903 nucleotides, while the second sequence group containing all local and global isolates had a length of 30,009 nucleotides.

COVID-19 statistical data, such as total cases, total deaths, total cases and deaths per 1 million population, and tests per 1 million cases for all countries and regions of the world were obtained from Worldometer's website using the Wayback Machine to access the COVID-19 statistics from November 30, 2020 at 22:52:43 WAT (West Africa Time).

Bayesian Phylogenetics

The sequence alignments were entered into Bayesian Evolutionary Analysis Utility (BEAUti) to set up the parameters of the Bayesian analysis [13]. The tips of all taxa were dated with their corresponding collection date, the substitution model was set to general time reversible, and the gamma category count was set to 4. The strict clock model and coalescent exponential population tree prior were selected, and a monophyletic group was specified as a tree prior, excluding the Wuhan reference sequence (this was done in order to specify the Wuhan reference sequence as an outgroup). Finally, the analysis was set to run for 10 million generations. The above steps were repeated for the second sequence group (all 230 isolates); the only difference this time around was that a relaxed clock model was selected instead of the strict clock model. XML-formatted files were generated and inputted into the BEAST 2.0 program (BEAST Developers; https://www.beast2.org/), which ran Markov-chain Monte Carlo analyses using the specified parameters.

Maximum clade credibility trees were generated with the use of the package TreeAnnotator from the "trees" files that were output by the BEAST analyses. The maximum clade credibility trees were visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). A timescale was generated by selecting the "Reverse Axis" option in the Scale Axis panel and setting the "Offset by" option in the Time Scale panel to the collection time (in decimal dates) of the most recently collected sequence sample, which corresponded to

https://doi.org/10.24171/j.phrp.2021.0329

Table 1. COVID-19 cases, deaths, and tests in 32 random countries as of November 30, 2020

Country	Total case	Total death	Total case/1 M population	Death/1 M population	Test/1 M population
Algeria	82,221	2,410	1,861	55	-
Argentina	1,418,807	38,473	31,274	848	85,375
Australia	27,902	908	1,089	35	390,289
Canada	374,051	12,076	9,875	319	301,291
China	86,530	4,634	60	3	111,163
Denmark	80,481	837	13,874	144	1,280,256
Egypt	115,541	6,636	1,120	64	9,697
Finland	24,912	399	4,493	72	348,911
France	2,222,488	52,731	34,018	807	315,590
Germany	1,068,980	16,830	12,742	201	332,072
Greece	105,271	2,406	10,121	231	228,189
Iceland	5,392	26	15,759	76	1,140,016
India	9,462,739	137,649	6,829	99	101,313
Ireland	72,544	2,053	14,624	414	396,069
Israel	336,160	2,864	36,549	311	609,406
Italy	1,601,554	55,576	26,505	920	363,181
Japan	146,760	2,119	1,162	17	27,729
Kazakhstan	131,659	1,990	6,977	105	222,942
New Zealand	2,056	25	411	5	254,998
Nigeria	67,412	1,173	324	6	3,632
Norway	35,971	332	6,614	61	408,924
Pakistan	398,024	8,025	1,788	36	24,742
Peru	962,530	35,923	29,026	1,083	152,566
Senegal	16,089	333	951	20	13,642
Singapore	58,218	29	9,919	5	757,851
South Africa	790,004	21,535	13,251	361	90,295
Spain	1,664,945	45,069	35,604	964	491,694
Sweden	243,129	6,681	24,012	660	313,550
Taiwan	675	7	28	0	4,629
United Arab Emirates	168,860	572	16,989	58	1,682,936
United Kingdom	1,629,657	58,448	23,954	859	639,079
United States of America	13,874,550	273,963	41,815	826	583,206

M, million; -, data unavailable.

2021.36 (that is, May 13, 2021).

Sequence Similarity Matrices

A sequence similarity matrix for the amino acid transcripts of the Nigerian sequence group was generated using the Base-by-Base bioinformatics program [14]. The coding regions for all isolates in this group were extracted, translated, concatenated, and then realigned before their similarity was assessed. The extraction of the nucleotide-coding regions was made possible by the annotation provided by the National Center for Biotechnology Information.

Genome-Wide Mutational Analysis

After obtaining the amino acid coding sequences of all open reading frames (ORFs) within the SARS-CoV-2 genome for the Nigerian sequence group, each ORF was manually examined to note amino acid differences from the Wuhan reference sequence.

Results

The COVID-19 Epidemic Was Mild in Nigeria Compared to the Rest of the World

Table 1 shows COVID-19 cases, deaths, and test statistics in 32 countries (as of November 30, 2020). Nigeria's cases and deaths per 1 million population were among the lowest of the 32 countries presented in Table 1. In fact, Nigeria is among the 5 countries that have the lowest cases and deaths per 1 million population alongside other countries, such as China, New Zealand, Senegal, Singapore, and Taiwan (Table 1). However, the very low COVID-19 case count in Nigeria needs to be interpreted in the context that Nigeria's rate of COVID-19 tests per 1 million population is among the lowest in the world (Table 1).

SARS-Cov-2 Isolated from Nigeria Resolved into Multiple Clades and Lineages

The Bayesian phylogenetic analysis of the Nigerian sequence group produced a phylogenetic tree showing that the SARS-CoV-2 strains circulating in Nigeria clustered into multiple clades (Figure 1). At the time of data collection (May 29, 2021), there were 7 clades of SARS-CoV-2 circulating in Nigeria, including G (B.1), GH (B.1), GR (B.1.1), GR (L.3), GRY (B.1.1.7), V (B), and S (A), with lineage frequencies of 7%, 10%, 32%, 16%, 24%, 4%, and 7%, respectively. The Nigerian isolates belonging to the S (A) and V (B) clades are shown as ancestral as they were closer to the reference Wuhan strain, whereas the GRY (B.1.1.7) clade was the most distant from the reference strain (Figure 1).

Time-Scaled Maximum Clade Credibility Trees of SARS-CoV-2

One aim of the Bayesian phylogenetic analyses was to identify the strains of SARS-CoV-2 circulating in Nigeria and possible regions from which they were imported. The clustering patterns of the Nigerian strains within the broader global cluster are depicted in Figures 2 and 3. The strict and relaxed clock phylogenies produced slightly different versions of the evolutionary history of SARS-CoV-2 worldwide. Although the clustering of isolates in the clock phylogenies was mostly similar, the overall branching patterns of some clades showed variance.

According to the strict clock Bayesian phylogenetic analyses, the Nigerian isolates of the GRY(B.1.1.7) clade clustered most closely with isolates from Israel (EPI_ISL_2182149 [B.1.17]) and India (EPI_ISL_2017759 [B.1.1]) (Figure 2). The isolates of the GR (B.1.1) clade in Nigeria were very diverse, and Figures 1–3 depict the diversity of this clade (highlighted in pink and appropriately labeled). The Bayesian phylogenetic analyses demonstrated that the GR (B.1.1) clade can be divided into at least 3 distinct sub-clades (Figure 1). As a result of this clade's diversity, it clustered with a host of isolates from different regions all over the globe (Figures 2 and 3), including isolates from Peru (EPI_ISL_1111139 [B.1.1.348], EPI_ISL_1111445 [B.1.1]), Ecuador (EPI_ISL_660069 [B.1.1.203]), Argentina (EPI_ISL_2105573 [P.2], EPI_ISL_2135164 [P.2]), South Africa (EPI_ISL_498093 [B.1.1], EPI_ISL_2285331 [B.1.1.448]), Iceland (EPI_ISL_829280 [B.1.1.170], EPI_ISL_1586502 [B.1.1]), and Germany (EPI_ISL_572397 [B.1.1], EPI_ISL_574259 [B.1.1]) (Figure 2). GISAID analyses have previously depicted another distinct GR sub-clade other than GR (B.1.1) that has been designated with a different Pango lineage (L.3). The Nigerian isolates of this clade clustered most closely with other GR (B.1.1) isolates, probably because there were no foreign representatives of this clade. Notable examples were from Germany (EPI_ISL_732563 [B.1.1]) and Greece (EPI_ISL_451310 [B.1.1]). The Nigerian isolates belonging to the GH (B.1) clade (colored yellow) clustered notably with isolates from France (EPI_ISL_560644 [B.1], EPI_ISL_2293487 [B.1.160]), Colombia (EPI_ISL_445219 [B.1]), Norway (EPI_ISL_2226159 [B.1]), Israel (EPI_ISL_447432 [B.1]), and Egypt (EPI_ISL_2232405 [B.1.170]) (Figure 2). The Nigerian isolates of the G (B.1) clade (colored red) most closely clustered with other B.1 isolates, with noteworthy isolates from New Zealand (EPI_ISL_637094 [B.1]), India (EPI_ISL_1544153 [B.1]), and other African nations such as Senegal (EPI_ISL_1827949 [B.1], EPI_ISL_1167163 [B.1]), and South Africa (EPI_ISL_464129 [B.1.8]) (Figure 2). It is also important to note that, unlike what is clearly shown in Figure 1, the G clade more closely clustered with the GR (B.1.1) and GR (L.3) set of clades instead of the GH (B.1) clade. The Nigerian isolates within the V (B) clade clustered most closely with the other foreign isolates belonging to this clade. The V (B) clade did not have many representative isolates. Finally, the Nigerian isolates of the S (A) clade most closely clustered with the foreign representatives of this clade (Figure 2). This clade also did not have many representative isolates. There were no Nigerian isolate representatives belonging to the L (B), GR (C, N, P, and R), GV (B.1.177), GK (AY and B.1.617.2), and O (B) clades.

Another phylogenetic tree was constructed, this time using the relaxed clock model. Both trees were highly similar, except for a few differences. The GRY (B.1.1.7) clade in the relaxed-clock tree was basically identical to its counterpart in the strict-clock tree; clustering most closely with the exact same 2 sequences from India and Israel (Figures 2 and 3). As in the strict clock phylogenetic tree, the GR (B.1.1) clade was very diverse and clustered with a host of isolates from widely different global regions (Figures 2 and 3). Its common clustering patterns were also evident in the relaxed-clock tree. One major difference between the 2 clock trees is apparent in the GR (L.3) clade. While in the strict-clock tree, this clade most closely clustered with isolates EPI_ISL_732563 (B.1.1) and EPI_ISL_451310 (B.1) from Germany and Greece, respectively, the relaxed-clock tree did not show support for such close clustering patterns. Instead, the isolates of the L.3 lineage clustered most closely with 2 sequences from Italy and the United States; EPI_ISL_528947 (B.1.1) and EPI_ISL_1823584 (B.1.1.263) (Figure 3). The clustering patterns of the GH (B.1) clade were very similar for both clock trees; the same group of isolates that clustered with this clade in the strict-clock tree also closely clustered in the relaxed-clock tree (Figures 2 and 3). The G (B.1) clade here was also very similar to its counterpart in the strict-clock tree, also exhibiting a closer relationship

https://doi.org/10.24171/j.phrp.2021.0329

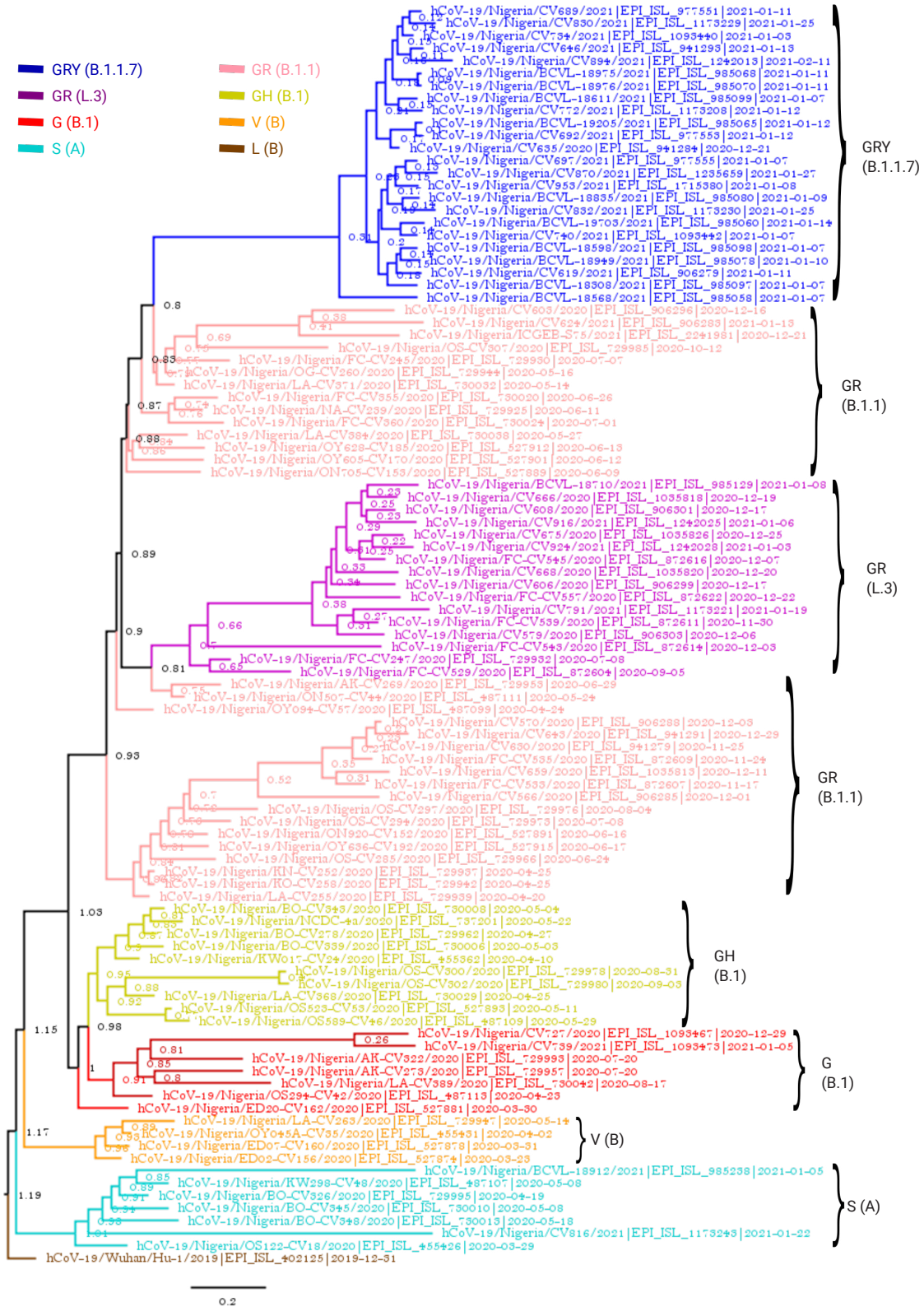

Figure 1. Strict clock Bayesian phylogeny of 100 full-genome severe acute respiratory syndrome coronavirus 2 Nigerian isolates plus the reference Wuhan isolate (hCoV-19/Wuhan/Hu-1/2019|EPI_ISL_402125|2019-12-31). The parameters included a general time reversible substitution model, the gamma category count set to 4, a coalescent exponential tree prior, and a custom tree prior to specify the Wuhan isolate "Wuhan/Hu-1/2019" as an outgroup. The analysis was run for 10 million generations.

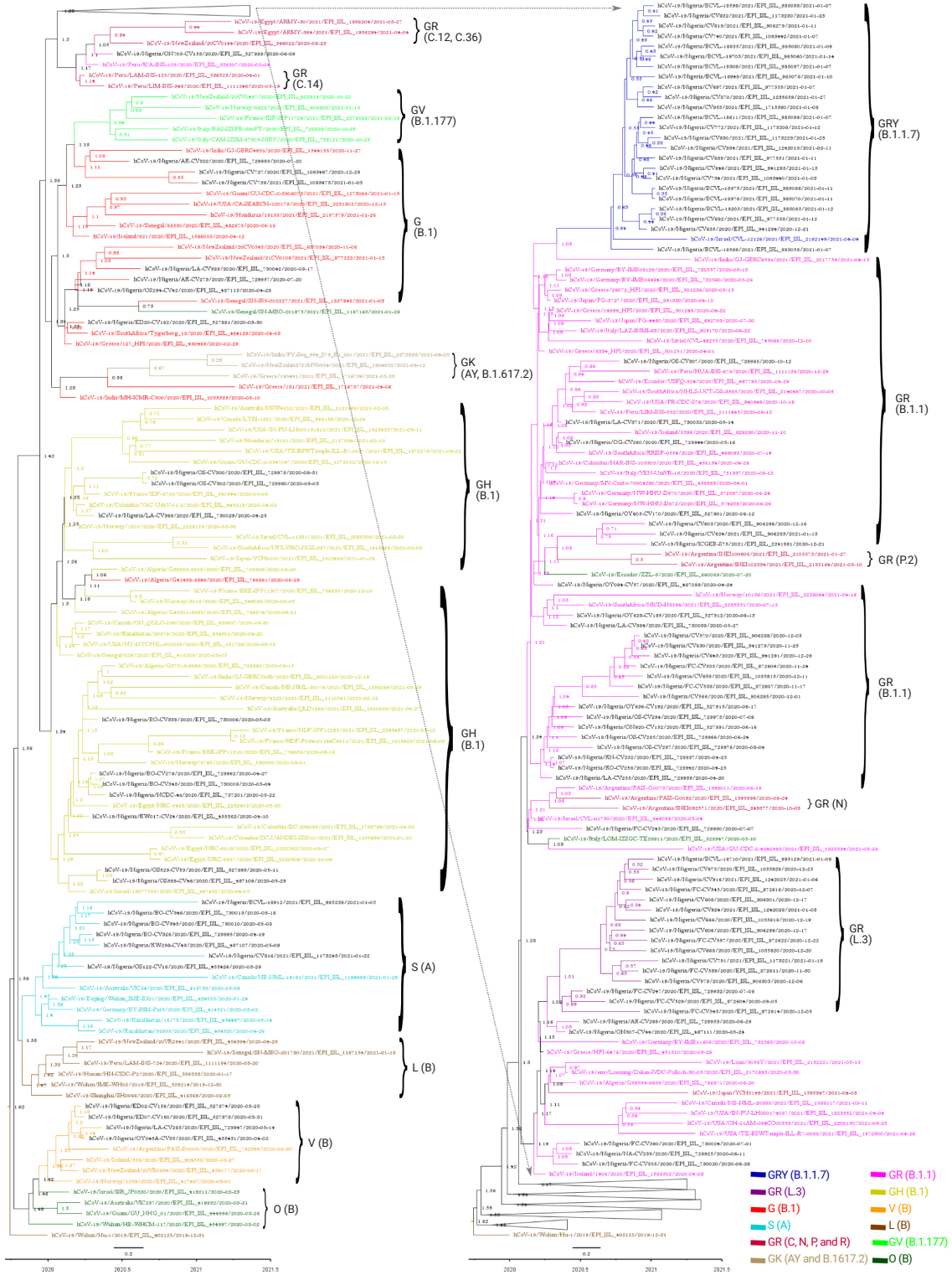

Figure 2. Strict clock time-scaled Bayesian phylogeny of 230 full-genome severe acute respiratory syndrome coronavirus 2 global isolates, including the reference Wuhan isolate (hCoV-19/Wuhan/Hu-1/2019|EPI_ISL_402125|2019-12-31). The parameters included a general time reversible substitution model, the gamma category count set to 4, a coalescent exponential tree prior, and a custom tree prior to specify the Wuhan isolate "Wuhan/Hu-1/2019" as an outgroup. The branches of all isolates are colored according to the Global Initiative to Share All Influenza Data clade and Pango lineage they belong to. However, the taxon labels of the Nigerian isolates are left colored in the default black, while the other isolates are colored depending on their clade.

https://doi.org/10.24171/j.phrp.2021.0329

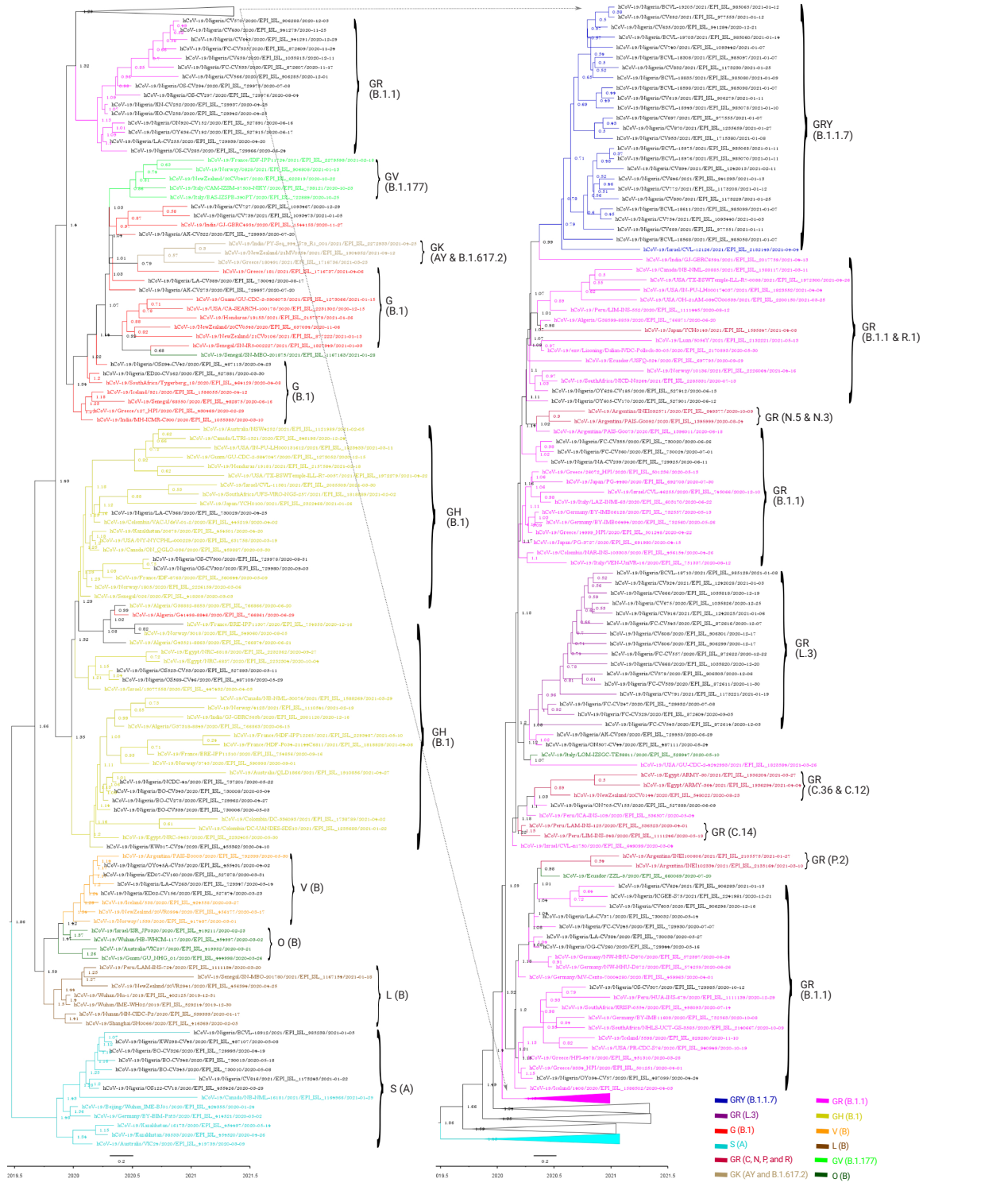

Figure 3. Relaxed clock time-scaled Bayesian phylogeny of 230 full-genome severe acute respiratory syndrome coronavirus 2 global isolates, including the reference Wuhan isolate (hCoV-19/Wuhan/Hu-1/2019|EPI_ISL_402125|2019-12-31). The parameters included a general time reversible substitution model, the gamma category count set to 4, a coalescent exponential tree prior, and a custom tree prior to specify the Wuhan isolate "Wuhan/Hu-1/2019" as an outgroup. The Nigerian isolates are highlighted in red. The branches of all isolates are colored according to the Global Initiative to Share All Influenza Data clade and Pango lineage they belong to. However, the taxon labels of the Nigerian isolates are left colored in the default black, while the other isolates are colored depending on their clade.

with the GR (B.1.1 and L. 3) set of clades than with the GH (B.1) clade, as suggested in Figure 1. The V (B) and S(A) clades in the relaxed-clock tree showcased a similar clustering pattern as in its strict clock counterpart (Figures 2 and 3). Finally, another major difference between the 2 clock trees was the positioning of the V (B), O (B), L (B), and S (A) clades on the trees. In the strict-clock tree, the V (B) and O (B) pair is shown to be more ancestral to the L (B) and S (A) pair (barring the Wuhan reference sequence), while in the relaxed-clock tree, the S (A) clade was depicted as being more ancestral to the other 3 clades, which formed a cluster (Figures 2 and 3).

Sequence Similarity Matrix of Isolates Correlated with Phylogenetic Grouping

As a means of evaluating the similarity of Nigerian isolates to one another and to the reference, a sequence similarity matrix is shown in Table 2. Isolates are represented by their GISAID accession identification/GISAID clade/Pango lineage. All isolates were highly similar, showing a similarity score of no less than 99.5% (Table 2). On average, isolates within the same clade were more similar than those of different clades. For example, the average sequence similarity value of isolates within a clade was 99.93%, while the average sequence similarity value of all isolates across different clades was only 99.84%. However, this was not always the case. As shown in Table 2, the S (A) clade was an exception to this rule. The intra-clade similarity score was only 99.76%, while the average inter-clade similarity score between the S clade and all other clades was 99.78% (Table 2).

D614G Was a Predominant Mutation in the S Gene of Nigerian Isolates, and Other SNPs and Deletions Occurred in Other Parts of the Genome

The D614G mutation in the spike (S) protein has received intense attention because of its association with increased infectivity. To check for the prevalence of this mutation in the available Nigerian isolates, position 614 of the S protein in respect to the reference Wuhan sequence was examined. This examination revealed that the D614G mutation existed in numerous Nigerian isolates (Table 3). A majority of the Nigerian isolates were G variants, only 10 isolates were D variants, and the isolate EPI_ISL_1173243 had an "N" amino acid at locus 614 of the S gene.

Tables 3–5 depict all single-nucleotide polymorphisms (SNPs) and deletions in all genes of the Nigerian isolates in comparison to the Wuhan reference sequence. Each isolate is represented by its GISAID accession identification number (Tables 3–5). Apart from the nucleotide substitutions observed at the 614th position of the spike gene, many other mutations occurred throughout the SARS-CoV-2 genome, most of which were SNPs. A few indels and run-on mutations were also noted. Twenty-nine amino acid mutations were present in over 10% of all Nigerian isolates. Out of these 29 mutations, 4 were present in over 70% of Nigerian isolates. These 4 were P323L in NSP12 (87.0%) (RNA-dependent RNA polymerase) of ORF1ab, D614G (88.0%) in the spike protein, and R203K (72.0%) and G204R (72.0%) in the nucleocapsid protein (ORFN).

Discussion

The observation of a mild COVID-19 epidemic in Nigeria from the onset of the epidemic through November 30, 2020 is indeed unexpected, as the Nigerian numbers best those of some countries that are believed to have handled the pandemic better [15–17]. Several hypotheses have been suggested to suggest this, primarily centering on the early action taken by Africa against the infection, likely fueled by African countries' familiarity with emerging infectious diseases, a warmer climate, a reduced genetic predisposition to the disease as a result of the reduced expression of ACE2 (the receptor implicated SARS-CoV-2 infectivity) in Africans, and a relatively young population [3,4,9,18–22]. Nonetheless, it should also be noted that only 3,632 tests were recorded per 1 million Nigerians, the lowest among the 32 countries shown in Table 1. As a result, the supposed mild COVID-19 epidemic in Nigeria is clouded by the fact that COVID-19 testing in that time period was very underwhelming. The proportion of COVID-19 tests per population should be considered as an important factor when interpreting the response of the country to the pandemic.

One aim of the Bayesian phylogenetic analyses in the present study was to identify the strains of SARS-CoV-2 circulating in Nigeria and possible regions from which they were imported. Bayesian trees with both strict and relaxed clock phylogenies showed that the individual SARS-CoV-2 clades in Nigeria clustered with and likely originated from various countries around the world. Some clades were likely imported from Asia, while others were likely imported from various countries in Europe and South America. The importation of SARS-CoV-2 from China into Nigeria is a very feasible occurrence considering air traffic flow patterns and bilateral trade agreements between the 2 nations [23–27]. The importation of COVID-19 from Europe is also as likely; in fact, Nigeria's index case was a man from Italy who traveled to Nigeria [2]. Other evidence supporting the importation of SARS-CoV-2 from Europe has been highlighted by 2 independent studies [28,29]. Therefore,

https://doi.org/10.24171/j.phrp.2021.0329

https://doi.org/10.24171/j.phrp.2021.0329

Table 2. Sequence similarity matrix of Nigerian isolates: 2 clade representatives are selected for each clade

Variable	Reference	985098/ GRY/ B.1.1.7	985070/ GRY/ B.1.1.7	730024/ GR/ B.1.1.484	527915/ GR/B.1.1	1242028/ GR/L.3	729932/ GR/L.3	729962/ GH/ B.1.462	730029/ GH/B.1	729993/ G/B.1	527881/ G/B.1	527874/ V/B	729947/ V/B	985238/ S/A.27	455426/ S/A
Reference		99.81	99.79	99.93	99.95	99.87	99.93	99.97	99.94	99.93	99.98	99.96	99.95	99.77	99.95
985098/GRY/B.1.1.7	99.81		99.98	99.82	99.85	99.76	99.82	99.82	99.79	99.78	99.83	99.77	99.76	99.61	99.76
985070/GRY/B.1.1.7	99.79	99.98		99.80	99.82	99.74	99.80	99.80	99.77	99.76	99.81	99.75	99.74	99.59	99.74
730024/GR/B.1.1.484	99.93	99.82	99.80		99.96	99.88	99.94	99.94	99.91	99.90	99.95	99.89	99.88	99.70	99.88
527915/GR/B.1.1	99.95	99.85	99.82	99.96		99.90	99.96	99.96	99.93	99.92	99.97	99.91	99.90	99.72	99.90
1242028/GR/L.3	99.87	99.76	99.74	99.88	99.90		99.90	99.88	99.85	99.84	99.89	99.82	99.81	99.66	99.81
729932/GR/L.3	99.93	99.82	99.80	99.94	99.96	99.90		99.94	99.91	99.90	99.95	99.89	99.88	99.70	99.88
729962/GH/B.1.462	99.97	99.82	99.80	99.94	99.96	99.88	99.94		99.97	99.94	99.99	99.93	99.92	99.74	99.92
730029/GH/B.1	99.94	99.79	99.77	99.91	99.93	99.85	99.91	99.97		99.91	99.96	99.90	99.89	99.71	99.89
729993/G/B.1	99.93	99.78	99.76	99.90	99.92	99.84	99.90	99.94	99.91		99.95	99.89	99.88	99.70	99.88
527881/G/B.1	99.98	99.83	99.81	99.95	99.97	99.89	99.95	99.99	99.96	99.95		99.94	99.93	99.75	99.93
527874/V/B	99.96	99.77	99.75	99.89	99.91	99.82	99.89	99.93	99.90	99.89	99.94		99.99	99.73	99.91
729947/V/B	99.95	99.76	99.74	99.88	99.90	99.81	99.88	99.92	99.89	99.88	99.93	99.99		99.72	99.90
985238/S/A.27	99.77	99.61	99.59	99.70	99.72	99.66	99.70	99.74	99.71	99.70	99.75	99.73	99.72		99.76
455426/S/A	99.95	99.76	99.74	99.88	99.90	99.81	99.88	99.92	99.89	99.88	99.93	99.91	99.90	99.76	

Table 3. Mutational profile of ORFS, ORF3a, ORF4, and ORF5 showing single-nucleotide polymorphisms and deletions

Variable	Isolate	Mutation	Isolate	Mutation	Isolate	Mutation	Isolate	Mutation	Isolate	Mutation
ORFS (surface glycoprotein)	730020	L5F	985238	L18F	1035813 +6	T22I	1093467 +1	Q52R	1093467 +1	A67V
	1093440 +25	Del 69-70	872616	G75R	985068 +1	D80Y	487111 +1	Del 141-144	1093440 +27	Del 144
	487109	Del 241-243	124025	G261R	527901	A262S	1035813 +5	T385I	2241981 +2	N439K
	1035813 +20	L452R	1035820	S477I	1093467 +2	E484K	1093440 +24	N501Y	872614	A520S
	487109	A522V	1093440 +23	A570D	1035813 +88	D614G	1173243	D614N	906283	V622F
	985238	A653V	906283	H655Y	527901	Q675R	729953	Q675H	872614	Q677H
	1093440 +27	P681H	730006	Q690L	1093440 +23	T716I	872604 +1	E780Q	985238	D796Y
	1093467 +1	F888L	1093440 +23	S982A	1035818	T1117I	1093440 +23	D1118H	730038	D1118Y
	729925	G1167V	985238	G1219V	985097	C1250F	729978 +1	L54F		
ORF3a protein	487099	D27Y	729966	I35T	985238	V50A	527901	L52F	487109 +9	Q57H
	872611	V77F	729966	L85F	872622	L101P	487109 +1	A110V	729993	T151I
	1173243	S171L	1173243	G172C	487111	Q185R	2241981 +2	S195F	872614	T223I
	729957	G224C	906296	D238Y	730029	P240S	455431 +3	G251V	729976	T269M
	1173243	257-271 point mutations	985238	258-271 point mutations						
ORFE (envelope protein)	1093467 +1	L21F								
ORFM (membrane glycoprotein)	1235659	W75L	1093467 +1	I82T	527878	C86F				

ORF, open reading frame.

Table 4. Mutational profile of ORF1ab and all its constituent non-structural proteins showing single-nucleotide polymorphisms and deletions

ORF1ab polyproteins	Isolate	Mutation	Isolate	Mutation	Isolate	Mutation	Isolate	Mutation	Isolate	Mutation
NSP1	729932	P6H								
	872614	Q66R								
	872609 +2	Del 141-143								
NSP2	985099	G29C	985238	P106L	729976	H208Y	730038	Q431H		
	906283	K82Q	1173243	K110N	729976	T256I	729930	T528I		
	730029 +2	T85I	941293	D155G	1035818 +12	L266F	455426	D582G		
	729993	K102E	730042	G165D	730013	A336V	455431 +3	P585S		
			906303	N168D	729976	Q376K	729995	T601A		
NSP3	729978 +1	P108L	2241981	G250R	527889	T428I	729930	S1038F	527901	H1307Y
	729978 +1	P108L	729930	G307C	527912	G476D	906285	T1063I	1093440 +23	I1412T
	906283	S176G	872616	V325F	872604	A480V	730042	P1103L	1173243	I1426T
	1093440 +22	T183I	1173243	S370L	729957	P822S	1093440	Y1185C	730042	P1469H
	872614	Del 206-207	2241981	P395L	729993	P874S	1093467 +1	T1189I	729976	A1527V
	906296	A231V	729966	D411Y	1093440 +23	A890D	730013	I1239T	527889	Q1756H
					487099	P968H	527901	L1244F	1035818 +12	P1921L
NSP4	729925	V30F								
	985238	D217G								
	1035818 +12	Y300S								
NSP5	527889	G15S								
	2241981	T21I								
	1235659	P241L								
	1173243	A255V								
	527901 +1	A260V								
NSP6	906285	W31C	729966	A136V						
	1035813 +17	L37F	906296	L146F						
	487109 +1	A46V	906285 +1	V149F						
	985060	A51V	730038	V182F						
	1093440 +25	Del 106-108	729993	C197F						
			1035818 +12	P198L						
NSP7	872622	L71F								
	1242025 +1	A80V								
NSP8	729930	R51C								
	1235659 +1	T145I								
	872616	T148I								
	729957	A152V								
NSP12 (RNA-dependent RNA polymerase)	1173243	T141I	1093467 +1	P323F						
	1093440 +22	P227L	730013	A529V						
	527889	G228D	730006	G774C						
	1093442 +1	V299F	455426	T806I						
	1035813 +87	P323L	729985	Q822H						
NSP13 (Helicase)	729985	S38L	985078	V98F						
	1173243 +1	S74L	729966	D105G						
	729978 +1	S74P	1093467	G118C						
	985238	P77L	487109 +1	V169F						
			1035818	A296T						
NSP14	730020 +3	A119S	527889	F326L						
	1093467	D222Y	487113	S418I						
	1173243	T250I	941284 +2	P451S						
NSP15	1715380	K12N	729957	R206M						
	985129	S147I								
	455426	S154F								
NSP16	729995	A83S								
	1093467	A116S								
	730020 +2	K182N								

ORF, open reading frame.

https://doi.org/10.24171/j.phrp.2021.0329

Table 5. Mutational profile of ORF6, ORF7a, ORF8, ORF9, and ORF10 showing single-nucleotide polymorphisms and deletions

Variable	Isolate	Mutation	Isolate	Mutation	Isolate	Mutation	Isolate	Mutation	Isolate	Mutation
ORF6 protein	1093473	Del 2	2241981	N39Y						
ORF7a protein	1242028	T14I	1173230	C35R	729957	P99L	527912	T120I		
ORF8 protein	729978 +1	S24L	527889	T26I	1093440 +23	Del 27	985097	W45C	1093440 +23	R52I
	730020	Del 59	729953	Del 65	729976	A65V	906299 +1	Del 66	906299 +1	S67A
	729953	K68E	1093440 +22	Del 68	1093440 +22	Y73C	1242013	S82A	455426 +6	L84S
	730013	Del 110	985238	D119I	2241981 +2	F120V	985238	Del 120-121		
ORFN (nucleocapsid phosphoprotein)	1093467 +1	Del 2	1093440 +23	D3L	1093467+1	D3Y	1093467 +1	A12G	730042	A35V
	872604	D128G	941291	D128Y	527912	T141I	1173243	D144H	729957	T148A
	729947	S193I	1173208	S194L	1035813 +72	R203K	1035813 +72	G204R	1093467 +3	T205I
	730032	A208G	730032	Del 209	1035813 +6	G212C	1093440 +23	S235F	2241981 +1	G238C
	906296	G238R								
ORF10 protein	487111 +1	I4T								

ORF, open reading frame.

with the observation that Nigerian isolates of SARS-CoV-2 were likely imported from Asia and Europe, the assumption that the strains circulating within Nigeria are different from strains circulating globally does not hold. This, in turn, does not explain the low transmissibility and fatality of the strains within Nigeria, especially considering the fact that most of the regions they are believed to have been imported from did not have attenuated disease outcomes, as highlighted in Table 1.

The sequence similarity analyses suggested that there was not much divergence between Nigerian isolates from the reference sequence (and from all SARS-CoV-2 isolates in general); therefore, the virus genome alone may not be a sufficient variable to explain the milder outcomes of COVID-19 in Nigeria. Other factors such as early action taken against COVID-19, a warmer climate, reduced genetic predisposition to the disease as a result of a reduced expression of ACE2, and a relatively young population may singly or in combination account for the attenuated disease outcomes in Nigeria. The sequence similarity matrix of Nigerian isolates corresponded to the phylogenetic groupings of clades; isolates within the same phylogenetic clade tended to have more similar sequences than isolates from different clades. This observation is significant because it further supports the result obtained via Bayesian phylogenetic inference.

The mutational analysis demonstrated that there were only very few non-synonymous substitutions in the genomes of Nigerian isolates compared to the reference Wuhan strain. Most non-synonymous mutations were present in 1 or a few isolates, although a few existed in numerous isolates. The D614G, P323L, R203K, and G204R mutations occurred in numerous Nigerian isolates. This same set of mutations was also discovered to be predominant in a SARS-CoV-2 epidemiological study conducted in Morocco [30].

There was an 88% prevalence of the D614G mutation in Nigerian isolates, and most of the D614G variants belonged to the S (A) clade of SARS-CoV-2, most likely because this clade is believed to be ancestral. Although this mutation is believed to be a relatively new mutation absent in the ancestral lineage, it has gained traction and is currently a dominant mutation in the population of SARS-CoV-2 worldwide, including in Nigeria and Africa [28,31,32]. Although there is a possibility that the G614 variant is evolutionarily favored over the D variant, there is no evidence of impact on disease severity and on therapeutic development [31,33–35]. The G variant, however, has been associated with increased transmission fitness, viral loads, and younger patient age [36,37].

The P323L mutation was also another predominant mutation in the Nigerian isolates studied, with a prevalence of 87%.

This mutation has been highlighted in the literature, and it was touted as an important variant in a large-scale study [38]. The P323L mutation may cause structural changes in NSP12, altering its interaction with NSP8 and affecting viral replication in host cells [39,40].

The R203K and G204R mutations were both present in ORFN, the nucleocapsid protein. These mutations have previously been identified in other studies [32,39,41]. Research on this mutation pair has shown that viruses possessing these mutations gain a replication advantage over the R203 and G204 variants [42]. This mutation pair has shown increased infectivity in the lung cells of humans and in hamsters [42]. In a study that modeled and analysed mutant protein structures, R203K and G204R were identified as mutations that caused significant changes to protein structure; moreover, this mutation pair also affected the affinity of intra-viral protein interactions [43].

Overall, we have shown that the SARS-CoV-2 strains circulating in Nigeria as of May 29, 2021 clustered into 7 different clades and were introduced into the country through multiple and unrelated introductions from Asia, Europe, South America, and Africa. The Nigerian SARS-CoV-2 isolates were also not genetically and phylogenetically distinct from strains circulating in other parts of the world. Future work will aim to associate the identified mutations with phenotypic characteristics through functional analysis.

Supplementary Material

Table S1. Sequence IDs, GISAID clades, Pango and Nextclade lineages of SARS-CoV-2 isolates used in the study; Table S2. Acknowledgement list of sequence submitters. Supplementary data are available at https://doi.org/10.24171/j.phrp.2021.0329.

Notes

Ethics Approval
Not applicable.

Conflicts of Interest
The authors have no conflicts of interest to declare.

Funding
None.

Availability of Data
Data are available on request to the corresponding author.

References

1. World Health Organization (WHO). WHO director-general's opening remarks at the media briefing on COVID-COVID-19: 11 March 2020 [Internet]. Geneva: WHO; 2020 [cited 2020 Oct 8]. Available from: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.
2. Nigeria Centre for Disease Control. First case of corona virus disease confirmed in Nigeria [Internet]. Abuja: Nigeria Centre for Disease Control; 2020 [cited 2020 Oct 8]. Available from: https://ncdc.gov.ng/news/227/first-case-of-corona-virus-disease-.
3. Torti C, Mazzitelli M, Trecarichi EM, et al. Potential implications of SARS-CoV-2 epidemic in Africa: where are we going from now? BMC Infect Dis 2020;20:412.
4. Lalaoui R, Bakour S, Raoult D, et al. What could explain the late emergence of COVID-19 in Africa? New Microbes New Infect 2020;38:100760.
5. Otu A, Ebenso B, Labonte R, et al. Tackling COVID-19: can the African continent play the long game? J Glob Health 2020;10:010339.
6. Armah FA, Ekumah B, Yawson DO, et al. Access to improved water and sanitation in sub-Saharan Africa in a quarter century. Heliyon 2018;4:e00931.
7. Racelma K. Towards African cities without slums [Internet]. New York: Africa Renewal; 2012 [cited 2020 Oct 13]. Available from: https://www.un.org/africarenewal/magazine/april-2012/towards-african-cities-without-slums.
8. Gilbert M, Pullano G, Pinotti F, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. Lancet 2020;395:871–7.
9. Kapata N, Ihekweazu C, Ntoumi F, et al. Is Africa prepared for tackling the COVID-19 (SARS-CoV-2) epidemic: lessons from past outbreaks, ongoing pan-African public health efforts, and implications for the future. Int J Infect Dis 2020;93:233–6.
10. World Health Organization (WHO). New WHO estimates: up to 190 000 people could die of COVID-19 in Africa if not controlled [Internet]. Geneva: WHO; 2020 [cited 2020 Oct 13]. Available from: https://www.afro.who.int/news/new-who-estimates-190-000-people-could-die-covid-19-africa-if-not-controlled.
11. World Health Organization (WHO). WHO coronavirus (COVID-19) dashboard [Internet]. Geneva: WHO; 2021 [cited 2021 Aug 31]. Available from: https://covid19.who.int.
12. Katoh K, Misawa K, Kuma K, et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 2002;30:3059–66.
13. Bouckaert R, Vaughan TG, Barido-Sottani J, et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol 2019;15:e1006650.
14. Hillary W, Lin SH, Upton C. Base-By-Base version 2: single nucleotide-level analysis of whole viral genome alignments. Microb Inform Exp 2011;1:2.
15. Wittenberg-Cox A. What do countries with the best coronavirus responses have in common? Women leaders. Forbes [Internet]. 2020 Apr 13 [cited 2020 Oct 17]. Available from: https://www.forbes.com/sites/avivahwittenbergcox/2020/04/13/what-do-countries-

with-the-best-coronavirus-reponses-have-in-common-women-leaders/.

16. Henley J. Female-led countries handled coronavirus better, study suggests. The Guardian [Internet]. 2020 Aug 18 [cited 2020 Oct 17]. Available from: https://www.theguardian.com/world/2020/aug/18/female-led-countries-handled-coronavirus-better-study-jacinda-ardern-angela-merkel.
17. Partridge-Hicks S. 5 Countries that are getting COVID-19 responses right. Global Citizen [Internet]. 2020 Sep 12 [cited 2020 Oct 17]. Available from: https://www.globalcitizen.org/en/content/countries-with-best-covid-responses/.
18. Zhao Y, Zhao Z, Wang Y, et al. Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2 [Preprint]. Posted 2020 Apr 9. bioRxiv 2020.01.26.919985. https://doi.org/10.1101/2020.01.26.919985.
19. Cao Y, Li L, Feng Z, et al. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. Cell Discov 2020;6:11.
20. Huang Z, Huang J, Gu Q, et al. Optimal temperature zone for the dispersal of COVID-19. Sci Total Environ 2020;736:139487.
21. Triplett M. Evidence that higher temperatures are associated with lower incidence of COVID-19 in pandemic state, cumulative cases reported up to March 27, 2020 [Preprint]. Posted 2020 Apr 12. medRxiv 2020.04.02.20051524. https://doi.org/10.1101/2020.04.02.20051524.
22. The Lancet. COVID-19 in Africa: no room for complacency. Lancet 2020;395:1669.
23. AnnaAero. The dragon awakes: over 2.5 million flew China-Africa last year [Internet]. AnnaAero; 2019 [cited 2020 Dec 7]. Available from: https://www.anna.aero/2019/09/13/the-dragon-awakes-over-2-5-million-flew-china-africa-last-year/.
24. Zhou Y. Air traffic between China and Africa has jumped 630% in the last decade. Quartz Africa [Internet]. 2019 Jul 28 [cited 2020 Dec 7]. Available from: https://qz.com/africa/1675287/china-to-africa-flights-jumped-630-in-the-past-nine-years/.
25. Pirie G. China has taken a different route to involvement in African aviation. The Conversation [Internet]. 2019 Sep 1 [cited 2020 Dec 7]. Available from: http://theconversation.com/china-has-taken-a-different-route-to-involvement-in-african-aviation-122522.
26. The World Bank. Air transport, passengers carried: Nigeria [Internet]. The World Bank Group; 2020 [cited 2020 Dec 7]. Available from: https://data.worldbank.org/indicator/IS.AIR.PSGR?locations = NG.
27. Jeremiah. Nigeria-China relations and 20 years of FOCAC. Leadership [Internet]. 2020 [cited 2020 Dec 7]. Available from: https://leadership.ng/nigeria-china-relations-and-20-years-of-focac/.
28. Motayo BO, Oluwasemowo OO, Olusola BA, et al. Evolution and genetic diversity of SARS-CoV-2 in Africa using whole genome sequences. Int J Infect Dis 2021;103:282–7.
29. Tegally H, Wilkinson E, Lessells RJ, et al. Sixteen novel lineages of SARS-CoV-2 in South Africa. Nat Med 2021;27:440–6.
30. Badaoui B, Sadki K, Talbi C, et al. Genetic diversity and genomic epidemiology of SARS-CoV-2 in Morocco. Biosaf Health 2021;3:124–7.
31. 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.e19.
32. Shaibu JO, Onwuamah CK, James AB, et al. Full length genomic sanger sequencing and phylogenetic analysis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Nigeria. PLoS One 2021;16:e0243271.
33. Grubaugh ND, Hanage WP, Rasmussen AL. Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. Cell 2020;182:794–5.
34. Vaidyanathan G. Vaccine makers in Asia rush to test jabs against fast-spreading COVID variant. Nature 2021 Jan 12 [Epub]. https://doi.org/10.1038/d41586-021-00041-y.
35. Garcia-Beltran WF, Lam EC, St Denis K, et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. Cell 2021;184:2372–83.e9.
36. Volz E, Hill V, McCrone JT, et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. Cell 2021;184:64–75.e11.
37. Plante JA, Liu Y, Liu J, et al. Spike mutation D614G alters SARS-CoV-2 fitness. Nature 2021;592:116–21.
38. Miao M, Clercq E, Li G. Genetic diversity of SARS-CoV-2 over a one-year period of the COVID-19 pandemic: a global perspective. Biomedicines 2021;9:412.
39. Garvin MR, T Prates E, Pavicic M, et al. Potentially adaptive SARS-CoV-2 mutations discovered with novel spatiotemporal and explainable AI models. Genome Biol 2020;21:304.
40. Chand GB, Banerjee A, Azad GK. Identification of novel mutations in RNA-dependent RNA polymerases of SARS-CoV-2 and their implications on its protein structure. PeerJ 2020;8:e9492.
41. Xu Y, Kang L, Shen Z, et al. Dynamics of severe acute respiratory syndrome coronavirus 2 genome variants in the feces during convalescence. J Genet Genomics 2020;47:610–7.
42. Wu H, Xing N, Meng K, et al. Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2. Cell Host Microbe 2021;29:1788–1801.
43. Wu S, Tian C, Liu P, et al. Effects of SARS-CoV-2 mutations on protein structures and intraviral protein-protein interactions. J Med Virol 2021;93:2132–40.