# Phyloevolutionary analysis of SARS-CoV-2 in Nigeria

E. H. Awoyelu[1], E. K. Oladipo[2], B. O. Adetuyi[1], T. Y. Senbadejo[3], O. M. Oyawoye[2] and J. K. Oloke[1]

1) *Department of Natural Sciences, Precious Cornerstone University, Ibadan, Oyo State,* 2) *Department of Microbiology, Laboratory of Molecular Biology, Immunology and Bioinformatics, Adeleke University, Ede* and 3) *Biological Sciences Department, Fountain University, Osogbo, Osun State, Nigeria*

## Abstract

Phyloepidemiological approaches have provided specific insight into understanding the emergence and evolution of infection. Knowledge on the outbreak and spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Nigeria would assist in provision of preventive measures to reduce transmission among populations at risk. This study aimed to investigate the evolution of SARS-CoV-2 in Nigeria. A total of 39 complete genomes of SARS-CoV-2 were retrieved from the GISAID EpiFlu™ database on 29 March 2020 to investigate its evolution in Nigeria. Sequences were selected based on the travel history of the individual and the collection date. Other sequences were not selected because they were short, contained artefacts, were not from an original source or had insufficient information. Evolutionary history was inferred using the maximum likelihood method based on the general time reversible model. A phylogenetic tree was constructed to determine the common ancestor of each strain. The phylogenetic analysis showed that the strain in Nigeria clustered in a monophyletic clade with a Wuhan sublineage. Nucleotide alignment also showed a 100% similarity indicating a common origin of evolution. Comparative analysis showed 27 972 (93.6%) identical sites and 97.6% pairwise identity with the consensus. The study evidently showed the entire outbreak of SARS-CoV-2 infection in Nigeria stemmed from a single introduction sharing consensus similarity with the reference SARS-CoV-2 human genome from Wuhan. Preventive measures that can limit the spread of the infection among populations at risk should be implemented.
© 2020 Published by Elsevier Ltd.

**Corresponding author:** A. Elukunbi, Department of Natural Sciences, Precious Cornerstone University, Ibadan, Oyo State, Nigeria.
**E-mail:** elukunbi@yahoo.com

## Introduction

The first confirmed case of coronavirus disease 2019 (COVID-19) in Nigeria was announced on 27 February 2020, when an infected traveller from one of the WHO-identified high-risk countries—Italy—arrived by commercial aircraft into Lagos. Although, the traveller's movement was restricted, another positive case was reported in Ewekoro, Ogun State, a Nigerian citizen who had had contact with the Italian citizen. Despite lockdown in some states and several precautionary measures put in place to prevent and contain the spread of the disease, the Nigeria Centre for Disease Control has since reported over 400 cases of infected individuals with 17 deaths in about 20 States.

Incidences of emerging/re-emerging viral infections have significantly affected human health despite extraordinary progress in the area of biomedical knowledge [1]. The key to understanding this emergence and evolution of novel viruses is subject to knowledge of the intricate host—pathogen—environment relationship [2]. Understanding the modes of transmission of an emerging infectious disease continues to be a key factor in implementing effective public health measures [3].

Proper tracking of genome sequences has helped to ensure optimal virus diagnostic tests, tracking and tracing of the ongoing outbreak and proper identification of potential intervention options [4]. Reports have suggested the route of transmission to be airborne [5], or via direct contact, droplets

and transmission from mildly ill or asymptomatic individuals [6,7]. However, lack of evidence on transmission dynamics can lead to inconsistencies in the isolation guidelines.

Phyloepidemiological approaches have given specific insight into understanding the emergence and evolution of emerging and re-emerging viruses, particularly severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [8]. Knowledge on the outbreak and spread of SARS-CoV-2 in Nigeria would help in providing preventive measures and reducing transmission among populations at risk. Hence, this study aimed to investigate the evolution of SARS-CoV-2 in Nigeria.

## Methods

### Nucleotide sequence retrieval and alignment tool
A total of 39 complete genomes (Table 1) of novel SARS-CoV-2 were retrieved from the GISAID database (https://www.epicov.org) on 29 March 2020. Sequences retrieved include those from China, Italy, France, Nigeria, South Africa and Congo.

Sequences were selected based on the travel history of the patient and the collection date. The sequences were then aligned to obtain the conserved regions using multiple sequence alignment with the aid of CLUSTAL W on MEGA X [15–17].

### Evolution analysis
The sequences were subjected to evolutionary divergence analysis. The phylogenetic tree was constructed to determine the common ancestor of each strain using MEGA 5.2 [18–20].

Comparative analysis of strains within clades was performed on GENEIOUS PRIME (https://www.geneious.com/) based on statistical analysis to determine positions in the genomic sequences from Nigeria that significantly differ from other strains.

## Results

### Evolutionary history
The analysis involved 39 nucleotide sequences. The evolutionary history was inferred using a maximum likelihood method based on the general time reversible model. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed.

Branches corresponding to partitions reproduced in <50% of bootstrap replicates were collapsed. Initial tree(s) for the heuristic search were obtained automatically by applying neighbour-joining and BIONJ algorithms to a matrix of pairwise distances estimated using the maximum composite likelihood approach, and then selecting the topology with superior log likelihood value.

A discrete gamma distribution was used to model evolutionary rate differences among sites (five categories (+$G$, parameter = 200.0000)). The rate variation model allowed for some sites to be evolutionarily invariable ((+$I$), 0.0000% sites). Codon positions included were 1st+2nd+3rd + Noncoding. All positions containing gaps and missing data were eliminated. There was a total of 25 459 positions in the final data set.

The maximum likelihood tree is shown in Fig. 1.

Multiple sequence alignment using CLUSTALW (https://www.ebi.ac.uk/Tools/services/rest/muscle), showed that all the genomes that formed a clade with the strain from Nigeria generally had more than 70% similarity in the genetic sequence.

### Conserved variants
Fig. 2(a–g) showed consensus similarities and variants between three strains from Wuhan, China and Nigeria, including the reference human SARS-CoV-2 genome. Sequences in the alignment were compared to the consensus to identify polymorphisms. At each position, the consensus is the allele with frequency >50%. N is ambiguity if no allele exceeds 50%.

**TABLE 1. Characteristics of the 40 SARS-CoV-2 genomic sequences used in this study**

| Country | Accession number | Month/year of collection | Age (year) | Sex |
|---|---|---|---|---|
| Nigeria | EPI_ISL_413550 | 02/2020 | 44 | M |
| Congo | EPI_ISL_417950 | 03/2020 | 48 | M |
| | EPI_ISL_417442 | 03/2020 | 39 | F |
| | EPI_ISL_417436 | 03/2020 | 43 | M |
| South Africa | EPI_ISL_417186 | 03/2020 | 38 | M |
| France | EPI_ISL_406596 | 01/2020 | 31 | F |
| | EPI_ISL_408431 | 01/2020 | 53 | M |
| | EPI_ISL_411220 | 01/2020 | 30 | F |
| | EPI_ISL_410720 | 01/2020 | 31 | F |
| | EPI_ISL_410486 | 02/2020 | 56 | F |
| | EPI_ISL_414623 | 02/2020 | 36 | M |
| | EPI_ISL_414624 | 02/2020 | — | — |
| | EPI_ISL_414626 | 02/2020 | 88 | F |
| | EPI_ISL_416502 | 02/2020 | 70 | — |
| | EPI_ISL_414625 | 02/2020 | 59 | F |
| | EPI_ISL_416749 | 03/2020 | 44 | F |
| | EPI_ISL_416746 | 03/2020 | 61 | M |
| | EPI_ISL_416751 | 03/2020 | 16 | M |
| | EPI_ISL_416756 | 03/2020 | 88 | F |
| | EPI_ISL_416752 | 03/2020 | 49 | F |
| China | EPI_ISL_402119 | 12/2019 | 49 | F |
| | EPI_ISL_402123 | 12/2019 | 65 | M |
| | EPI_ISL_403929 | 12/2019 | 52 | F |
| | EPI_ISL_402124 | 12/2019 | 49 | F |
| | EPI_ISL_406716 | 01/2020 | — | — |
| | EPI_ISL_408978 | 02/2020 | 65 | F |
| | EPI_ISL_408514 | 02/2020 | | |
| | EPI_ISL_412982 | 02/2020 | — | M |
| Italy | EPI_ISL_413489 | 03/2020 | 38 | F |
| | EPI_ISL_417921 | 03/2020 | 32 | M |
| | EPI_ISL_417922 | 02/2020 | 41 | M |
| | EPI_ISL_417923 | 03/2020 | 53 | M |
| | EPI_ISL_412974 | 02/2020 | 57 | M |
| | EPI_ISL_412973 | 02/2020 | 38 | M |
| | EPI_ISL_417421 | 01/2020 | 45 | M |
| | EPI_ISL_410545 | 01/2020 | 66 | F |
| | EPI_ISL_412974 | 01/2020 | 57 | M |
| | EPI_ISL_417491 | 03/2020 | 75 | M |

hCoV-19/Italy/INMI1-isI/2020 EPI ISL 410545
hCoV-19/Wuhan/WH05/2020 EPI ISL 408978
hCoV-19/France/IDF0372-isI/2020 EPI ISL 410720
hCoV-19/France/IDF0386-isIP3/2020 EPI ISL 411220
hCoV-19/France/IDF0372/2020 EPI ISL 406596
hCoV-19/Italy/SPL1/2020 EPI ISL 412974
hCoV-19/France/GE1583/2020 EPI ISL 414623
hCoV-19/Congo/158/2020 EPI ISL 417950
hCoV-19/Wuhan/IVDC-HB-04/2020 EPI ISL 402120
91
hCoV-19/Wuhan/IPBCAMS-WH-01/2019 EPI ISL 402123
hCoV-19/France/IDF0626/2020 EPI ISL 408431
hCoV-19/Wuhan/WIV04/2019 EPI ISL 402124
hCoV-19/Wuhan/IVDC-HB-envF13-20/2020 EPI ISL 408514
hCoV-19/Wuhan/IPBCAMS-WH-04/2019 EPI ISL 403929
hCoV-19/China/WHU01/2020 EPI ISL 406716
hCoV-19/France/RA739/2020 EPI ISL 410486
hCoV-19/Wuhan/HBCDC-HB-06/2020 EPI ISL 412982
hCoV-19/Italy/INMI3/2020 EPI ISL 418166
● hCoV-19/Nigeria/Lagos01/2020 EPI ISL 413550
hCoV-19/France/Valence 425/2020 EPI ISL 416746
94
hCoV-19/Italy/INMI4/2020 EPI ISL 418167
hCoV-19/Italy/UniMI03/2020 EPI ISL 417447
hCoV-19/Italy/INMI5/2020 EPI ISL 418168
hCoV-19/South Africa/R03006-20/2020 EPI ISL 417186
hCoV-19/France/N1620/2020 EPI ISL 414624
hCoV-19/Italy/UnivPM1/2020 EPI ISL 417491
hCoV-19/Italy/INMI5/2020 EPI ISL 417923
hCoV-19/Congo/KN-0072/2020 EPI ISL 417442
hCoV-19/Congo/KN-0051/2020 EPI ISL 417436
hCoV-19/France/Valence 532/2020 EPI ISL 416749
hCoV-19/France/Clermont-Ferrand 650/2020 EPI ISL 416752
hCoV-19/France/Lyon 06531/2020 EPI ISL 416756
hCoV-19/France/Clermont-Ferrand 651/2020 EPI ISL 416751
hCoV-19/Italy/CDG1/2020 EPI ISL 412973
hCoV-19/France/PL1643/2020 EPI ISL 414625
hCoV-19/France/B2330/2020 EPI ISL 416502
hCoV-19/France/HF1684/2020 EPI ISL 414626
hCoV-19/Italy/UniSR1/2020 EPI ISL 413489
hCoV-19/Italy/FVG-ICGEB S8/2020 EPI ISL 417421
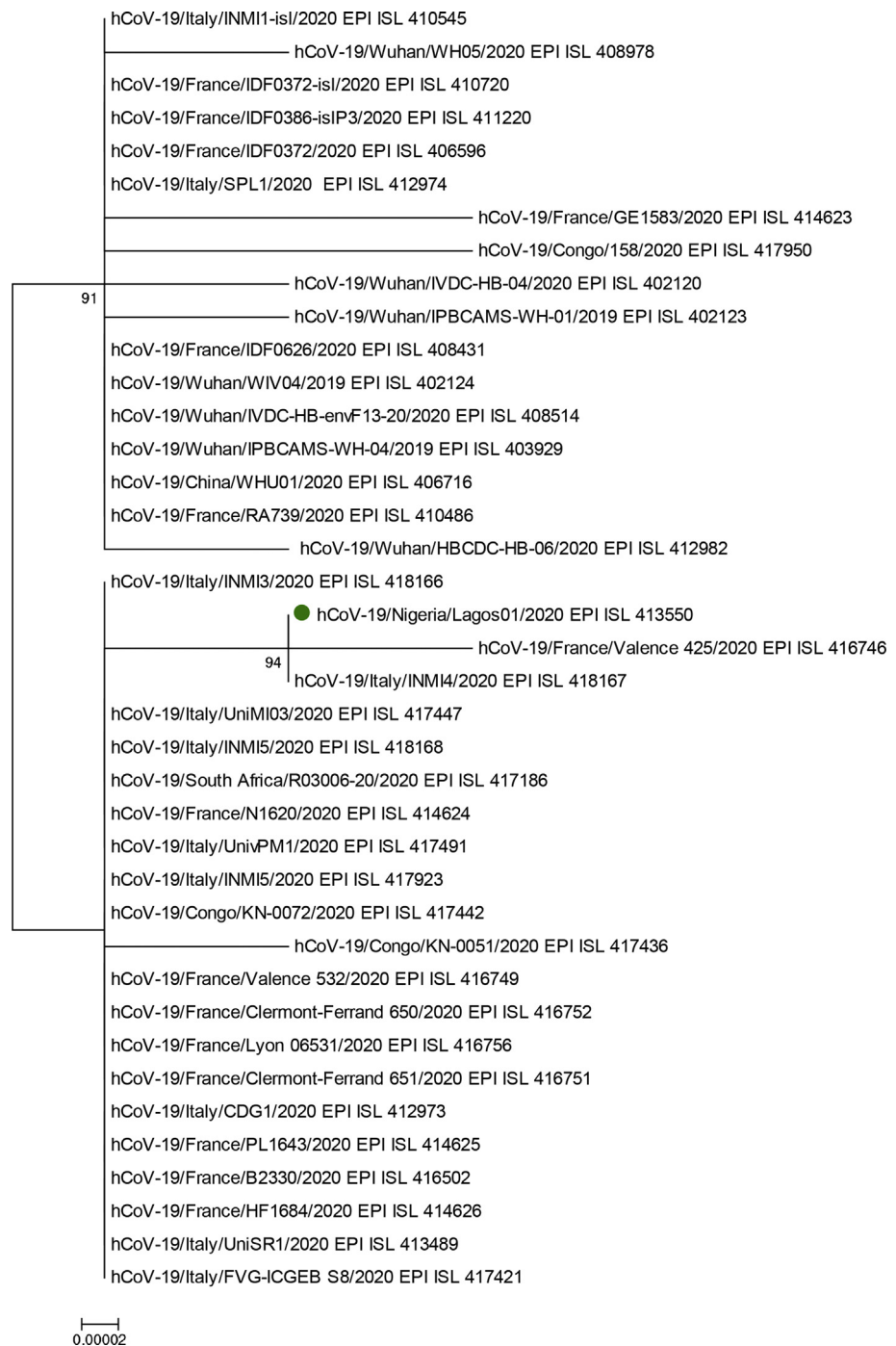
0.00002

**FIG. 1.** Maximum Likelihood Tree of SARS-CoV-2. Sequences were aligned using MAFFT and tree reconstruction using MEGA 5.0.

## Discussion

Knowledge on the transmission chain of an emerging or re-emerging virus combined with sequence data provide more insight into explaining occurrence of mutation and spread of infection [9]. In the Nigerian COVID-19 outbreak, viewed by itself, Nigerian sequence data submitted by Okwuraiwe et al. [10] have provided a clear picture that the entire outbreak stemmed from a single introduction into the country. This is an indication that the first case is an imported case that has turned into a serious health challenge in society.

Along with the somewhat close proximity of the sample collection dates to each other, it is possible to reconstruct a

**FIG. 2.** (a–f) Consensus variants between the four strains studied; (g) consensus variants showing gaps in the genomic sequence from Nigeria.

transmission chain and infer that the patient from Nigeria contracted the virus through a patient who became infected with the Wuhan strain, as seen from the phylogenetic tree. The two strains are in the same clade by sharing the genetic information. The reconstruction can be explained through the findings of Parvez and Parveen [1] that the emergence of infectious diseases in naive regions is dependent on the movement of pathogens via trade and travel, whereas local emergence is driven by a combination of environmental and social changes [11].

Multiple sequence alignment separated the strains studied into three distinct clades depending on their divergence from their common ancestor. The strains from Wuhan further

subdivided the clades into three subclades. The strain from Nigeria was found in the Wuhan subclade 3 together with some strains from Congo and France. The strains that formed a monophyletic clade with Wuhan subclade 3 resembled the genetic sequence with >70% similarity.

More importantly, the tree confirmed that the outbreak in Nigeria was the result of a single introduction from China/Wuhan through an imported case of an Italian. More specifically, the imported SARS-CoV-2 strain from Nigeria is a descendant of the China/Wuhan strain as also described by Zhu et al. [12].

Comparative analysis of the strain from Nigeria, two strains from Wuhan sharing the same clade and the reference human SARS-CoV-2 genome was performed. Results from GENEIOUS PRIME showed that all four sequences had 27 972 (93.6%) identical sites and 97.6% pairwise identity. The strain from Nigeria and Wuhan strain (WH05/2020) had more genome sequence similarity compared with strain WH01/2020. They shared consensus similarity with the reference SARS-CoV-2 human genome, showing a common descendant as observed from other studies [13,14].

The strain imported into Nigeria by the Italian shared <20% variant characteristics with Wuhan strain WH01/2020. Compared with the consensus, the strain from Nigeria had 49 gaps, 39 unknowns and seven point mutations. More than 80% of these differences were unique to Nigeria.

In summary, on the basis of the evolutionary analysis, it is evident that human-to-human transmission occurred, hence preventive measures should be adhered to so as to control the spread of the virus. The study showed that the entire outbreak of COVID-19 infection in Nigeria stemmed from a single introduction sharing consensus similarity with the reference SARS-CoV-2 human genome from Wuhan.

Establishment of the phyloevolutionary relationship of the Nigerian reference sequence for SARS-CoV-2 could benefit the biological study of this virus, and the diagnosis, clinical monitoring and intervention for SARS-CoV-2 infections in Nigeria.

## Conflict of interest

The authors' declare no conflict of interest.

## References

[1] Parvez MK, Parveen S. Evolution and emergence of pathogenic viruses: past, present and future. Intervirology 2017;60:1–7.

[2] Weiss SR, Leibowitz JL. Coronavirus pathogenesis. Adv Virus Res 2011;81:85–164.

[3] Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 2003. https://doi.org/10.1016/j.vaa.2019.11.008.

[4] Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 395 2020. pp. 470-473.

[5] Booth TF, Kournikakis B, Bastien N, Ho J, Kobasa D, Stadnyk L. Detection of airborne severe acute respiratory syndrome (SARS) coronavirus and environmental contamination in 20 SARS outbreak units. J Infect Dis 2005;191:1472–7.

[6] Omrani AS, Matin MA, Haddad Q, Al-Nakhli D, Memish ZA, Albarrak AM. A family cluster of Middle East respiratory syndrome coronavirus infections related to a likely unrecognized asymptomatic or mild case. Int J Infect Dis 2013;17:e668–72.

[7] Paules CI, Marston HI, Fauci AS. Coronavirus infections—more than just the common cold. JAMA 2020. epub ahead of print.

[8] Avise JC. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, Massachusetts, 2000. viii1447 pp. 49.95, (ISBN 0-674-66638-0). Available from, https://www.researchgate.net/publication/275147931_Phylogeography_The_History_and_Formation_of_Species [accessed Jun 17 2020].

[9] Folarin OA, Ehichioya D, Schaffner SF, Winnicki SM, Wohl S, Eromon P, et al. Ebola virus epidemiology and evolution in Nigeria. J Infect Dis 2016;214(Suppl. 3). Epub ahead of print.

[10] Okwuraiwe A, Onwuamah C, Amoo O, Rosemary A, Salako B, James A, et al. First African SARS-CoV-2 genome sequence from Nigerian COVID-19 case. 2020. https://doi.org/10.13140/RG.2.2.22914.63685.

[11] Lu R, Zhao Z, Li J, Niu P, Yang B, Wu H, et al. Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020;395:565–74.

[12] Zhu N, Zhang DY, Wang W, Li X, Yang B, Song J, et al. For the China novel coronavirus investigating and research team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020;382:727–33.

[13] Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. For the Washington state 2019-nCoV case investigation team*. First case of 2019 novel coronavirus in the United States. N Engl J Med 2020;382:929–36.

[14] Arima Y, Shimada T, Suzuki M, Suzuki T, Kobayashi Y, Tsuchihashi Y, et al. Severe acute respiratory syndrome coronavirus 2 infection among returnees to Japan from Wuhan, China, 2020. Emerg Infect Dis 2020;6:7.

[15] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395:497–506.

[16] Qin E, Zhu Q, Yu M, Fan B, Chang G, Si B, et al. A complete sequence and comparative analysis of a SARS-associated virus (isolate BJ01). Chin Sci Bull 2003;48:941.

[17] Yu WB, Tang GD, Zhang L, Corlett RT. Decoding the evolution and transmission of the novel pneumonia coronavirus (SATS-CoV-2) using whole genome data. 2020. Preprint.

[18] Nei M, Kumar S. Molecular evolution and phylogenetics. New York: Oxford University Press; 2000.

[19] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011;28:2731–9.

[20] Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 1985;39:783–91.