



Evolution and genetic diversity of SARS-CoV-2 in Africa using whole genome sequences



Babatunde Olarenwaju Motayo^{a,*}, Olukunle Oluwapamilerin Oluwasemowo^b, Babatunde Adebisi Olusola^c, Paul Akiniyi Akinduti^{d,*}, Olamide T. Arege^c, Yemisi Dorcas Obafemi^d, Adedayo Omotayo Faneye^c, Patrick Omoregie Isibor^d, Oluwadurotimi Samuel Aworunse^d, Solomon Uche Oranusi^d

^a Federal Medical Center, Abeokuta, Nigeria

^b Fifty-four Gene, Lagos, Nigeria

^c Department of Virology, College of Medicine, University of Ibadan, Nigeria

^d Department of Biological Science, Covenant University, Ota, Nigeria

ARTICLE INFO

Article history:

Received 29 July 2020

Received in revised form 20 November 2020

Accepted 26 November 2020

Keywords:

SARSCoV-2
Virus evolution
Phylogeny
Africa

ABSTRACT

Background: The ongoing SARS-CoV-2 pandemic was introduced into Africa on 14th February 2020 and has rapidly spread across the continent causing a severe public health crisis and mortality. We investigated the genetic diversity and evolution of this virus during the early outbreak months, between 14th February to 24th April 2020, using whole genome sequences.

Methods: We performed recombination analysis against closely related CoV strains, Bayesian time scaled phylogeny, and investigation of spike protein amino acid mutations.

Results: Recombination signals were observed between the Afr-SARS-CoV-2 sequences and reference sequences within the RdRPs and S genes. The evolutionary rate of the Afr-SARS-CoV-2 was 4.133×10^{-4} Highest Posterior Density (HPD 4.132×10^{-4} to 4.134×10^{-4}) substitutions/site/year. The time to most recent common ancestor (TMRCA) of the African strains was December 7th 2019, (95% HPD November 12th 2019–December 29th 2019). The Afr-SARCoV-2 sequences diversified into two lineages A and B, with B being more diverse with multiple sub-lineages confirmed by both maximum clade credibility (MCC) tree and PANGOLIN software. There was a high prevalence of the D614G spike protein amino acid mutation 59/69 (82.61%) among the African strains.

Conclusion: This study has revealed a rapidly diversifying viral population with the G614G spike protein variant dominating. We advocate for up scaling NGS sequencing platforms across Africa to enhance surveillance and aid control effort of SARS-CoV-2 in Africa.

© 2020 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Towards the end of December 2019, Chinese authorities, through the World Health Organization office in China, made the world known of a new pathogen responsible for a series of pneumonia associated infections in Wuhan, Hubei province (WHO 2020a). The pathogen was later identified to be a novel coronavirus closely related to the severe acute respiratory syndrome virus (SARS), with a possible bat origin (Zhou et al., 2020). The World

Health Organization named the disease COVID-19 (Chan et al., 2020), and later declared it a pandemic on 11th March 2020, prompting concerted efforts towards prevention and control worldwide (WHO 2020a). On February 11th 2020 the international committee on the taxonomy of viruses (ICTV) adopted the name SARS-CoV-2 following the report of their coronavirus working group CSG-1CTV, (The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2, 2020). The virus has been placed in the subgenera *Sarbecovirus*, genus *Betacoronavirus*, subfamily *Coronavirinae*, family *Coronaviridae* (de Groot et al., 2013; Gorbalenya et al., 2020).

Coronaviruses are enveloped viruses containing a single-stranded positive sense RNA genome with a size of between 26

* Corresponding authors.

E-mail addresses: babatundemotayo@yahoo.com (B.O. Motayo), niyiakinduti@gmail.com (P.A. Akinduti).

kb to 32 kb (Masters and Pearlman 2013). They are responsible for a host of human and animal infections. The *Betacoronaviruses* contain the most medically important species of human coronaviruses such as HuCoVOC43 and HuCoVHKU13. The severe acute respiratory syndrome coronavirus SARS-CoV and the Middle East respiratory syndrome coronavirus MERS are also members of this group, and have been reported as high consequence pathogens with zoonotic potential and ability to cause large-scale epidemics (Lau et al., 2005; Zaki et al., 2012). Genomic and structural analyses have revealed that SARS-CoV-2 encodes four structural proteins spike (S), membrane (M), envelope (E) and nucleocapsid (N) proteins as well as several non-structural proteins (Chen et al., 2020; Lu et al., 2020). The spike protein is the major antigenic protein responsible for initiating infection, via attachment of its receptor binding domain (RBD) to the SARS-CoV/SARS-CoV-2 receptor angiotensin converting enzyme 2 ACE 2 (Donnelly et al., 2004; Monteil et al., 2020).

Globally, there have been 5,226,268 confirmed SARS-CoV-2 cases globally, with 335,218 deaths as of 21st of May 2020 (ECDC 2020). The coronavirus pandemic began in Egypt, Africa on the 14th February 2020 with an Italian who returned into the country (WHO 2020b). As of 21st of May there have been 95,332 cases in Africa, with 2995 deaths and 35,519 recoveries covering fifty-four countries in Africa with South Africa having the highest number of cases at 18,003 (WHO 2020c). Several reports have traced the evolutionary origins of SARS-CoV-2 to SARSr-CoV from bats (Zhou et al., 2020) and Pangolins (Lam et al. 2020).

Phylogenetic analysis has shown that the virus has diversified through the duration of the pandemic into two major lineages, A and B, with several sub-lineage diversifications (Rambaut et al. 2020a, 2020b). The majority of the reports were generated using genome sequences of SARS-CoV-2 from North America, Europe and Asia (Rambaut et al. 2020a, 2020b). There has been a paucity of data on the genetic evolution of SARS-CoV-2 sequences from Africa, despite the increasing number of genome sequence submissions into the Global Initiative for Sharing of Avian Influenza Data (GISAID) database from Africa. There were 97 whole genome sequences available in the GISAID database as at 24th April 2020. The majority of the published information on SARS-CoV-2, particularly in Sub-Saharan Africa, has been on the socio-economic impact of the virus in the region (Akinduti et al., 2020; Olasehinde et al., 2020; Oleribe et al., 2020). This gap in knowledge prompted the conceptualization of this study. Describing the genetic diversity and evolutionary dynamics of SARS-CoV-2 will facilitate real time surveillance, and understanding of antigenic diversity and virus transmission patterns. This study was therefore designed to determine the genetic diversity and evolutionary history of SARS-CoV-2 genome sequences isolated in Africa.

Materials and methods

Data curation

Full genome sequences with high coverage were downloaded from the GISAID database. As of 24th April there were 97 full genome sequences from Africa available in the GISAID database. This included 69 high coverage genomes, defined as sequences with < 1% unidentified nucleotides, <0.05% mutations not found in another isolate, and no indel mutations not verified by the submitter. This selection was made through an option in the database that automatically filters only high coverage genomes. Another 151 high coverage full genome sequences were also downloaded from three continents North America (USA), Asia (China and South Korea) and Europe (England, Italy and Germany). Three different datasets were then generated from

these sequences. The first dataset consisted of high coverage full genome sequences from Africa, along with SARS-CoV-2 reference genome sequence from Wuhan, China consisting of Bat and Pangolin SARS related reference sequences and SARS-CoV reference sequences (n = 76). The first dataset was used for the evolutionary and Bayesian phylogenetic analysis of the African SARS-CoV-2 genome sequences. The second dataset consisted of complete genome sequences from Africa, North America, Asia and Europe (n = 220), the second dataset was used for the generation of Bayesian phylogenetic data, as well as lineage determination. The third dataset consisted of complete spike protein (S) gene sequences from Africa and Bat and Pangolin SARS-related reference S gene sequences (n = 69). This dataset was used exclusively for spike protein amino acid motif analysis and visualization to determine significant mutations of the African SARS-CoV-2 spike protein sequences. In addition to the sequence data retrieved from GISAID, clinico-demographic information was retrieved from 33 of the sequence submissions from Africa between February 14th and April 24th 2020. This was the total number of submissions that had demographic information relating to the infected patients. Table 1 shows a summary of the demographic distribution by country of the 33 patients.

Phylogenetic analysis

Whole genome sequences downloaded from the GISAID database were aligned using MAFFT v7.222 (FF-NS-2 algorithm) following default settings (Katoh et al., 2019). Maximum likelihood phylogenetic analysis was performed using the general time reversible nucleotide substitution model with gamma distributed rate variation GTR-Γ (Yang, 1994) with 1000 bootstrap replicates using IQ-TREE software (Nguyen et al., 2015). Lineage assignments for the SARS-CoV-2 sequences were conducted using the Phylogenetic Assignment of Named Global Outbreak LINEages tool (PANGOLIN), available at <http://github.com/hCoV-2019/pangolin>.

Recombination analysis

We analyzed potential recombination events using the Recombination Detection Program (RPD) software (Martin et al., 2015). The analysis was conducted on whole genome sequences of identified lineages among the 69 African isolates, using RDP, bootscan analysis, GENECOV, Chimera, SISCAN, 3SEQ, and maximum chi-square methods. A putative recombination event was passed only if three of the above mentioned methods gave a positive recombination signal (Liu et al., 2010).

Table 1
Distribution of demographic information of available from SARSCoV-2 sequences submitted to GISAID between February 12 and April 24th 2020.

Variable	Country				
	Senegal	Algeria	Nigeria	DRC	South Africa
Gender					
Male	16	2	1	1	1
Female	9	1	Nil	Nil	2
Age range					
> 15	2	Nil	Nil	Nil	Nil
16-45	12	2	1	Nil	3
>45	11	1	Nil	1	Nil
Clinic Outcome					
Diseased	1	Nil	Nil	Nil	Nil
Hospital	24	3	1	1	3
Mild	Nil	Nil	Nil	Nil	Nil

Evolutionary and timescaled phylogenetic analysis

Temporal clock signal was analyzed among the aligned sequences using TempEst version 1.5 (Rambaut et al., 2020). The root-to-tip divergence and sampling dates supported the use of molecular clock analysis in this study. Phylogenetic trees were generated by Bayesian inference through Markov chain Monte Carlo (MCMC), implemented in BEAST version 1.10.4 (Suchard et al., 2018). We partitioned the coding genes into first + second and third codon positions and applied a separate Hasegawa-Kishino-Yano (HKY + G) substitution model with gamma-distributed rate heterogeneity among sites to each partition (Hasegawa et al., 1985). The relaxed clock with Gaussian Markov Random Field Skyride plot (GMRF) coalescent prior was selected for the final analysis, after running different models and comparing them using Bayes factor with marginal likelihood estimated using the path sampling and stepping stone methods implemented in BEAST version 1.10.4 (Suchard et al., 2018). One hundred million MCMC chains were run with 1 0% burn in. Results were then visualized with Tracer version 1.8. (<http://tree.bio.ed.ac.uk/software/tracer/>). The effective sampling size values (ESS), were above 200 indicating sufficient sampling. Bayesian skyride analysis was carried out to visualize the epidemic evolutionary history using Tracer version 1.8.

Analysis of spike protein

Complete S protein gene sequence of Afr-SARS-CoV-2 was aligned along with RaTG13 BtCoV and Pangolin SARSr-CoV sequences using MAFFT (Katoh et al., 2019). The alignment was then edited and visualized using BioEdit software (Hall, 1999).

Results and discussion

The current global SARS-CoV-2 pandemic, otherwise known as COVID-19, began on the African continent from a European returnee in Egypt on February 17th 2020 (WHO 2020a). It has since spread to virtually all the countries within the African region. This study was based on sequences generated during the early phase of the pandemic in Africa between February 2020 and April 2020. Sixty-nine high coverage full genome sequences from six African countries, namely Algeria (3), Senegal (20), Democratic Republic of Congo (DRC) (35), Nigeria (1), Ghana (6) and South Africa (4) were analyzed. Basic demographic information was available for 33 of the isolates, and has been summarized in Table 1. From the patient demography males were more infected than females, and the majority of individuals infected were within the adult working age

range of 16–45. This result is an indication of the adverse economic effect of the pandemic on the economies of African countries as the number of work hours will be grossly affected by infected workers calling in sick or going into self-isolation owing to secondary contacts with infected SARS-CoV-2 patients. Phylogenetic analysis of the African sequences showed clustering within the *Sabecovirus* sub-genus forming a sub-cluster with SARSr-CoV and PCoV (Fig. 1a) as previously reported in several studies (Zhao et al., 2020; Lam et al., 2020; Zhang et al., 2020). The root-to-tip regression analysis showed a weaker signal with a correlation of coefficient of 0.995 and $R^2 = 0.991$, this is evidenced by the short distance between the points to the line of best fit (Supplementary Figure 1).

Results of recombination analysis of the African SARS-CoV-2 (Afr-SARS-CoV-2) sequences against the reference whole genome sequences of SARS-CoV, MERS-CoV and BtCoV RaTG13 were presented using the RDP program (Martin et al., 2015). Recombination events evidenced by recombination breakpoint around nucleotides 27,150 and 27270, were detected by 5 of the 7 detection methods incorporated within the RDP program. The events were observed between African SARS-CoV-2 sequences and reference sequence (Major recombinant hCoV-19 Pangolin/Guangu P4L/2017; Minor parent hCoV-19 B batYunan/RaTG13) within the RdRP and S gene regions (Fig. 1b). This result is consistent with a previous report from Saudi Arabia which investigated the recombination between SARS-CoV-2 and closely related viruses such as SARS-CoV and MERS (Nour et al., 2020). The evolutionary rate for the Afr-SARS-CoV-2 isolates during the period under study was 4.133×10^{-4} substitutions/site/year (high posterior density interval HPD, 4.132×10^{-4} to 4.134×10^{-4}). This is slightly higher than that of a report from early outbreak strains from China with a rate of 3.345×10^{-4} (Li et al., 2020), however, it is lower than the calculated global SARS-CoV-2 evolutionary rate estimated to be 8.0×10^{-4} as reported by Nexstrain (www.nextstrain.org/ncov/global). There seems to be a gradual increase in SARS-CoV-2 viral evolution as the virus is established within new populations, as evidenced by the slight increase in the rate in African strains compared to the early Chinese strains. However, the results from this study may not represent the true evolutionary trend of the virus as nucleotide substitution rates are still limited and the number of African sequences few. However, as the pandemic progresses and more African SARS-CoV-2 genomes are sequenced and submitted the evolutionary trend will be clearer. The MCC tree of the African SARS-CoV-2 sequences shows that they have evolved into two major lineages, A and B, with lineage B being more diverse. The majority of the African SARS-CoV-2 sequences clustered within lineage B, while three Ghanaian, three Congolese, and four

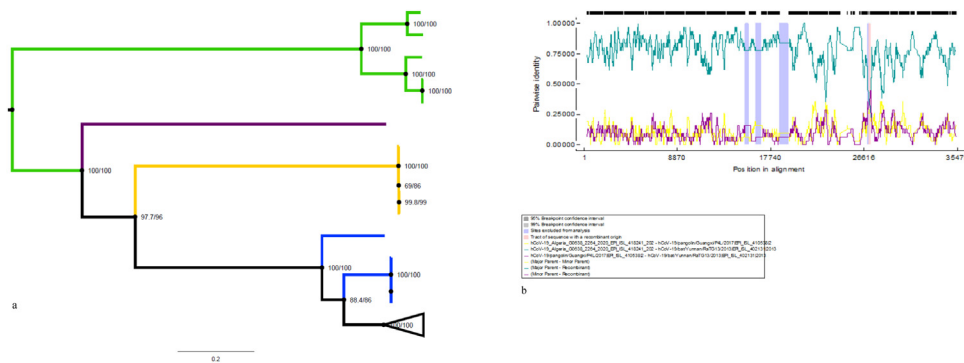


Fig. 1. (a) Maximum Likelihood phylogenetic tree of AfrSARSCoV2 full genome sequences with 100% bootstrap value, along with Pangolin SARS-like (Red), Bat SARS-like CoV (Blue), SARS-CoV (Blue), MERS-CoV (Yellow) and alpha CoV (Green) full genome sequences. The empty triangle represents Afr-SARS-CoV = 2 sequences; (b) Boot scan plot of complete genome sequences of Afr-SARS-CoV-2 sequences analyzed with the RDP recombination software. The legend shows the identity of the sequences scanned within the plot; the light blue bars indicate the portions of the genome with recombinant signals in reference to the major and minor recombinant parent sequences.

Senegalese strains clustered along with the reference Chinese and South Korean strains within lineage A (Fig. 2a). The MCC tree for the dataset containing global reference sequences also showed a similar topology with Fig. 2a. The tree (Fig. 2b) showing phylogeny of African SARS-CoV-2 strains along with other global reference sequences, was distributed into two major lineages, A and B, with lineage B further diversifying into about four sub-lineages, and lineage A diversifying into only two sub-lineages (Fig. 2b). The Afr-SARS-CoV-2 strains were intermixed with the global sequences within both lineages, lineage B consisted mainly of strains from Germany, England, Italy and the USA, intermixed with African strains; while lineage A consisted mainly of strains from South Korea and China with a few African strains from Senegal, Ghana and DRC. The result of the genotype analysis using the genotyping tool PANGOLIN was largely in conformity with observed phylogenetic analysis. Fig. 3a shows a summary of the lineage distribution of the isolates by country of origin using the PANGOLIN genotyping tool. The complete distribution of the strains according to lineage and country is shown in supplementary Table 2. From the analysis with PANGOLIN, lineage B.1 was the most commonly encountered and the most widely distributed, consisting of 93 sequences from seven countries, followed by lineage B.2 and genotype B. Lineage A had 15 positive sequences from six countries. The majority of the sequences recorded high bootstrap values with over 70% of the sequences recording a bootstrap value of above 80%. This shows that the PANGOLIN is a reliable tool with a broad scope of functions including a user friendly and interactive representation of phylogenetic clustering of the identified sub-lineages and lineages by means of graphical images of the trees generated using virtually all available SARS-CoV-2 sequences available on GISAID platform as a reference. The genotyping tool which was recently introduced, has been shown by previous studies to be very useful in accurately predicting SARS-CoV-2 lineage assignments (Xavier et al., 2020). The TMRCA of the African SARS-CoV-2 strains was December 7th 2019, 95% HPD interval (November 12th 2019 to December 29th 2019), while the TMRCA of all the sequences under analysis was 14th October 2019, 95% HPD (July 27th 2019 to December 17th 2019). Our TMRCA was lower than a similar study which reported a TMRCA of 14th October 2019 among global isolates including Chinese isolates (Li et al., 2020), but was slightly higher than another recent study investigating the evolutionary dynamics of the ongoing SARS-CoV-2 epidemic in Brazil which reported a TMRCA of 10th February 2020 (Xavier et al., 2020). These slight differences in time of origin observed from different studies can be

due to the differences in the number of sequences analyzed and different Bayesian models employed for analyses (although majority of reports utilize the coalescent relaxed models) (Li et al., 2020). The epidemic history of the ongoing outbreak was investigated using the Bayesian Skyline Plot BSP. The BSP showed a steady increase in viral population as the outbreak progressed during the study period (Fig. 3b). This observation was expected as the viral sequence population will increase as infection spreads. A major limitation was the small number of sequences analyzed and the short study duration; therefore, our results may not reflect the exact viral population dynamic of the outbreak in Africa.

The Afr-SARS-CoV-2 sequences were analyzed for the D614G mutation within the S1 subunit of the spike protein, which has been reported to contribute to increased transmissibility of SARS-CoV-2 (Korber et al., 2020a). Fig. 4 shows a representative amino acid alignment of selected Afr-SARS-CoV-2 sequences along with reference sequences of BtCoV RaTG13 and PCoV. Our results revealed high prevalence of D614G mutation among Afr-SARS-CoV-2 at 59/69 (82.61%). The mutation was recorded in isolates from all African countries analyzed in this study, Supplementary Figure 2. Prior to this report the D614G spike mutation was found predominantly in Europe accompanied by high number of cases and significant mortality rate (Pachetti et al., 2020; Korber et al., 2020b). The introduction of this strain in Africa is worrisome, considering the population densities of most African cities and the poor state of public health infrastructure to support medical intervention of symptomatic SARS-CoV-2 cases. Although more evidence is still required to determine the extent of the effect of the D614G mutation on the virulence factors of the virus, current evidence from in vitro studies seem to support the hypothesis of increased transmissibility of this variant of the virus (Korber et al., 2020a; Hu et al., 2020).

In conclusion, we have reported the genetic diversity and evolutionary history of SARS-CoV-2 isolated in Africa during the early outbreak period. Our findings have identified diverse sub-lineages of SARS-CoV-2 currently circulating among Africans. We identified a relatively high prevalence of the D614G spike protein variant of the virus capable of rapid transmission in all countries sampled. Major limitations to this study were the lack sufficient patient information from the originating samples which would have helped in further linking epidemiologic data to the sequence data, and the relatively low amount of sequence submission available in GISAID database from Africa as at the time of this study, compared with those of other regions such as Europe and Asia. We

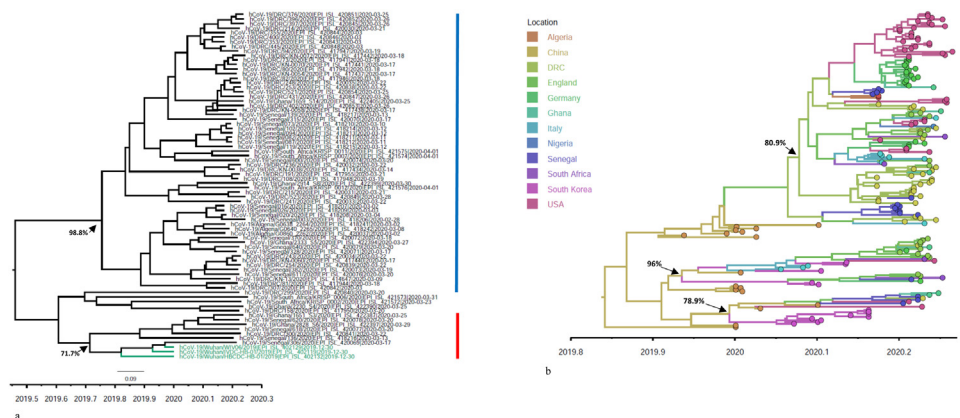


Fig. 2. (a) Time-scaled MCC tree of Afr-SARS-CoV-2 sequences, green labels represent reference SARS-CoV-2 isolates from Wuhan, China. The blue horizontal line represents lineage B isolates, while the red horizontal line represents lineage A isolates. Posterior probability values of the major lineages are indicated as percentages using arrows; (b) Time scaled MCC tree of complete genome sequences of Afr-SARS-CoV-2 isolates along with isolates from Asia, Europe and North America. The legend indicates the color code for each country of origin of the isolates. The red horizontal line represents lineage B.1 isolates, the green horizontal line represents lineage B.2, while the blue horizontal line represents lineage A isolates. Posterior probability values of the major lineages are indicated as percentages using arrows.

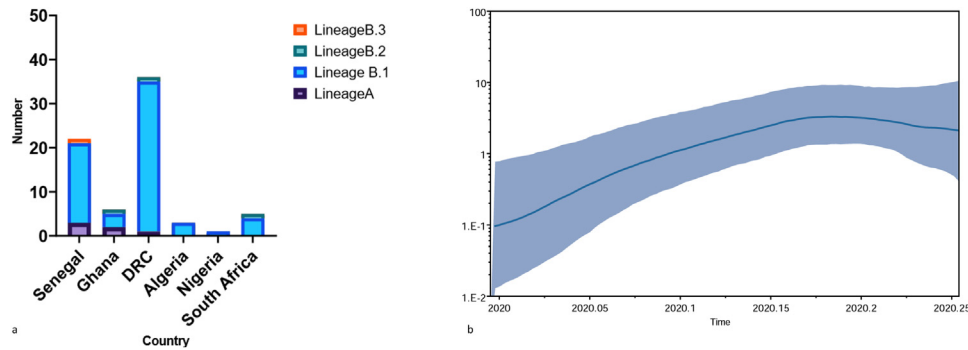


Fig. 3. (a) Bar chart showing lineage distribution of Afr-SARS-CoV-2 sequences according to country of origin; (b) Bayesian Skyline plot of Afr-SARS-CoV-2 sequences through the early period of the pandemic in Africa.

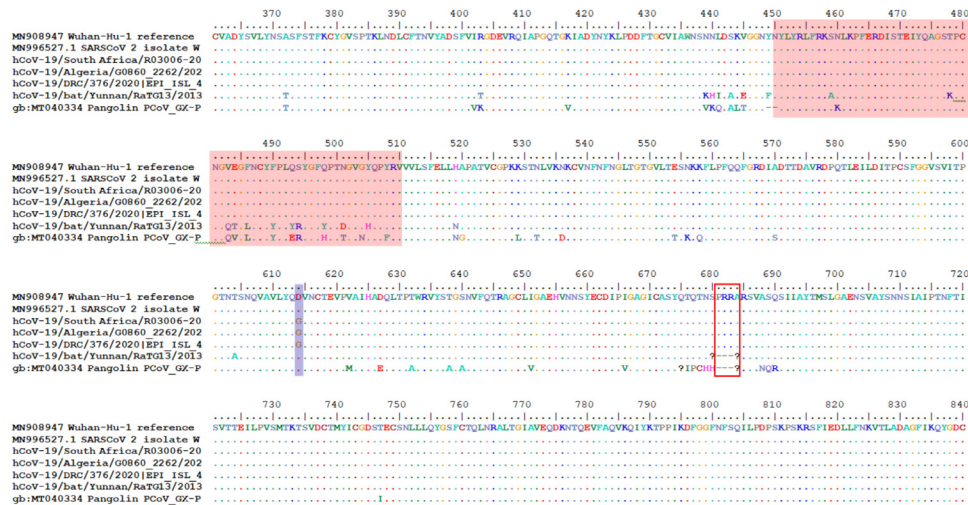


Fig. 4. Amino acid alignment of the partial S gene sequences covering amino acid positions 360 to 840, of selected Afr-SARS-CoV-2 isolates along with reference sequences of closely related PCoV and bat RaTG13. The red shaded region represents the receptor binding domain; the blue shaded box represents the D614G motif, while the empty red box represents the polybasic cleavage site bordering the S1/S2 sub-unit. The question mark represents missing information sites in amino acid alignment.

advocate for upscale of NGS capacity for whole genome sequencing of SARS-CoV-2 samples across the African continent to support surveillance and control effort in Africa.

Conflict of interest statement

The authors declare that there are no conflicts of interests in regards to the publication of this study.

Funding source

The authors did not receive any form of funding to conduct this research.

Ethical approval

The study did not require an ethical approval, so none was sought.

Acknowledgements

We are grateful to all the authors, originating and submitting laboratories from Global Initial on Sharing All Influenza Data (GISAID’s EpiCoV database. <http://www.gisaid.org>.) for making the sequences available for use in our study. We also acknowledge the

management of Covenant University Otta for their support to publish this work.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ijid.2020.11.190>.

References

Akinduti PA, Obafemi YD, Oranusi SU. Sero-epidemiological impact of SARSCoV-2 on the socio-demographic status of African populace. *Covenant J Life Sci* 2020;34:6–9.

Gorbalenya E, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. The species severe acute respiratory syndrome-related virus: classifying 2019-nCoV and naming it SARSCoV-2. *Nat Microbiol* 2020;5:536–44.

Chan JFW, Yuan S, Kok KH, To KK, Chu H, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 2020;395(10223):514–23.

de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, et al. Commentary: Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J Virol* 2013;15: 87(14):7790–2.

Donnelly CA, Fisher MC, Fraser C, Ghani AC, Riley S, Ferguson NM, et al. Epidemiological and genetic analysis of severe acute respiratory syndrome. *Lancet Infect Dis* 2004;4: 672e83.

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;5(4):536–44, doi: <http://dx.doi.org/10.1038/s41564-020-0695-z>.

- European Center for Disease Control ECDC. Covid-19 situation world wide as at 21th May. <https://www.ecdc.europa.eu/en/covid-19/situation-updates>.
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT". *Nucl Acids Symp Ser* 1999;41:95–8.
- Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 1985;22:160–74, doi:<http://dx.doi.org/10.1007/BF02101694>.
- Hu J, He C-L, Gao Q-Z, Zhang G-J, Cao X-X, Long Q-X, et al. The D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity and decreases neutralization sensitivity to individual convalescent sera. *bioRxiv* 2020;2020:06.20.161323.
- Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 2019;20(4):1160–6, doi:<http://dx.doi.org/10.1093/bib/bbx108> In this issue.
- Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 2020a;182(4):812–27.
- Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *bioRxiv* 2020b;. . preprint<https://doi.org/10.1101/2020.04.29.069054>.
- Lam TT, Jia N, Zhang W, Shum MH, Jiang J, Zhu H, et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* 2020;. . (2020) <https://doi.org/10.1038/s41586-020-2169-0>.
- Lau SK, Woo PCY, Li KSM, Huang Y, Tsoi H, Wong BHL, et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A* 2005;102:14040–5, doi:<http://dx.doi.org/10.1073/pnas.0506735102>.
- Li X, Zai J, Zhao Q, Nie Q, Li Y, Foley BT, et al. Evolutionary history, potential intermediate animal host, and cross species analysis of SARS-CoV-2. *J Med Virol* 2020;92(6):602–11.
- Liu X, Wu C, Chen AY. Codon usage bias and recombination events for neuraminidase and hemagglutinin genes in Chinese isolates of influenza A virus subtype H9N2. *Arch Virol* 2010;155(5):685–93.
- Lu R, Zaou X, 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; doi:10.1016/S0140-6736(20)30251-8 (2020).
- Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* 2015;1(1).
- Masters PS, Perlman S. Chapter 28, Coronaviridae. In: Knipe D, Howley P, editors. *Fields Virology* 6th edition. Lippincott Williams and Wilkins; 2013.
- Monteil V, Kwon H, Prado P, Hagelkrüys A, Wimmer RA, Stahl M, et al. Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. *Cell* 2020; In press. 16.
- Nguyen L-T, Schmidt HA, Haeseler Arndt von, Minh Bui Quang. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32(1):268–74, doi:<http://dx.doi.org/10.1093/molbev/msu300>. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4271533/>.
- Nour I, Alanazi IO, Hanif A, Kohl Eifan S. Insights into evolution and recombination of pandemic SARS-CoV-2 using Saudi Arabian sequences. *bioRxiv* 2020; preprint. doi: <https://doi.org/10.1101/2020.05.13.093971>.
- Olasehinde GI, Akinduti PA, Akinola OO, Ipadeola AF, Adebayo GP. COVID-19 Pandemic: Perception, Practices and Preparedness in Nigeria. *Pan Afr J Life Sci* 2020;4(1):30–4.
- Oleribe OO, Osita-Oleribe P, Salako BL, Ishola TA, Fertleman M, Tylor-Robinson SD. COVID-19 experience: Taking the right steps at the right time to prevent avoidable Morbidity and Mortality in Nigeria and other nation of the World. *Int J Gen Med* 2020;13:491–5.
- Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med* 2020;18:179.
- Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2020a;2: vev007.
- Rambaut A, Holmes EC, Hill V, O'Toole Á, McCrone J, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *bioRxiv* 2020b; doi:<https://doi.org/10.1101/2020.04.17.046086>.
- Suchard MA, Lemey P, Baele G, Ayers DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018;4: vey016. DOI:10.1093/ve/vey016.
- World Health Organization (WHO). Novel Coronavirus (2019-nCoV) Situation Report - 1, 21 January 2020. WHO Bull; 2020.
- World Health Organization (WHO). WHO Africa/Second case of nCoV confirmed in Africa. <https://www.afro.who.int/news/second-covid-19-case-confirmed-africa>.
- World Health Organization (WHO). Coronavirus disease 2019 (COVID-19). Situation Report – 51 11 March 2020. 2020 Available: https://www.who.int/docs/defaultsource/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10.
- Xavier J, Giovanetti M, Adelino T, Fonseca V, daCosta BA, Ribero AA, et al. The ongoing COVID-19 epidemic in Minas Gerais, Brazil: insights from epidemiological data and SARS-CoV-2 whole genome sequencing. *medRxiv* 2020; preprint doi: <https://doi.org/10.1101/2020.05.05.20091611>.
- Yang Ziheng. Maximum Likelihood Phylogenetic Estimation from DNA Sequences with Variable Rates over Sites: Approximate Methods. *J Mol Evol* 1994;39(3):306–14.
- Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 2012;367:1814–20 doi:<https://doi.org/10.1056/NEJMoa1211721> (2012).
- Zhang T, Wu Q, Zang Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr Biol* 2020;. doi:<http://dx.doi.org/10.1016/j.cub.2020.03.022>.
- Zhou P, Lou Yang X, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579:270–3, doi:<http://dx.doi.org/10.1038/s41586-020-2012-7>.