Prevalence of COVID-19 genomic variation in Africa: a living systematic review protocol

George Adjei[1,2] • Yeetey A. Enuameh[2,3] • Nicholas E. Thomford[4,5]

[1]Department of Community Medicine, School of Medical Sciences, University of Cape Coast, Cape Coast, Ghana, [2]The Kintampo Health Research Centre: A JBI Centre of Excellence, Kintampo, Ghana, [3]Department of Epidemiology and Biostatistics, School of Public Health, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, [4]Department of Medical Biochemistry, School of Medical Sciences, University of Cape Coast, Cape Coast, Ghana, and [5]Division of Human Genetics, Department of Pathology, University of Cape Town, Cape Town, South Africa

ABSTRACT

Objective: The objective of this living systematic review is to synthesize the available evidence on the prevalence of types of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genomic variations in Africa.

Introduction: The burden of the coronavirus disease 2019 (COVID-19) pandemic on the health, well-being, and global economy (especially the fragile economies of African countries) is significant. Profiling the genomic and geographical variations of SARS-CoV-2, a causative agent of COVID-19, may be important for future decision-making, policy guidelines, and development of drugs and vaccines. However, little is known about the up-to-date prevalence of genomic and geographical variations of SARS-CoV-2 virus on the African continent.

Inclusion criteria: This living systematic review will include studies on the prevalence of SARS-CoV-2 genetic strains and mutations obtained from sequencing data of samples from individuals of all ages and sexes using the next generation sequencing approaches in studies conducted in Africa.

Methods: The search strategy will be developed to retrieve both published and unpublished data. Published data will be obtained from electronic databases. Unpublished data will be obtained from conference proceedings, preprints, theses/dissertations, electronic search engines, and COVID-19–dedicated websites. Relevant published or unpublished data in the English language from January 2020 will be considered. Studies will be selected based on the inclusion criteria of the review. The selected studies will be critically appraised for methodological quality by two independent reviewers and data extracted from eligible studies. Finally, meta-analysis will be done, if feasible, to pool prevalence estimates after heterogeneity of the data has been analyzed.

Systematic review registration number: PROSPERO CRD42020211451

Keywords: Africa; COVID-19; genome; SARS-CoV-2 and mutations

JBI Evid Synth 2022; 20(1):158–163.

Introduction

Coronavirus disease 2019 (COVID-19), which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was initially reported in Wuhan, China, in mid-December 2019.[1] COVID-19 has a high transmission rate and was able to spread to over 210 countries in less than four months from the point of discovery.[2] Due to its high rate of transmission, the World Health Organization declared it a global pandemic on March 11, 2020.[3] By September 22, 2020, there were 31,174,627 confirmed cases of COVID-19 worldwide, with 962,613 deaths.[3] Although there is a considerable case-fatality rate of the disease worldwide, it varies across geographical regions.[2]

SARS-CoV-2 is a single-stranded ribonucleic acid (RNA) virus with four structural proteins (spike [S], envelope [E], membrane [M], and nucleocapsid [N]), 16 non-structural proteins (nsp1–16), and five accessory proteins (ORF3a, ORF6, ORF7a, ORF7b, and ORF8).[4,5] Current available evidence has shown that there are geographic genomic diversities and recurrent mutations in the viral genome across

Correspondence: George Adjei, geoz30@yahoo.com

The authors declare no conflict of interest.

DOI: 10.11124/JBIES-20-00516

geographic regions, reiterating its global diversity.[2] Mutation of the virus occurs at a low rate and the S protein has been reported to have high affinity for angiotensin converting enzyme-2, thus mediating binding and entry into host cells.[6] Because of this, most of the 60-plus approaches of SARS-CoV-2 candidate vaccines under development have targeted the S protein.[7] Currently, few vaccines have been proven to be effective against the novel SARS-CoV-2.[8-10] In addition, there is no systematic, reliable, and robust method of estimating prevalence of types of variations in genes to ascertain dominant genes and their associated mutations in geographical regions of Africa.

Currently, 57 countries in Africa are affected by COVID-19, with total confirmed cases of 1,152,822 and confirmed deaths of 24,882 as of September 22, 2020.[3] The five highest-ranked countries contributing to the confirmed cases are South Africa (57.4%), Ethiopia (6.0%), Nigeria (5.0%), Algeria (4.3%), and Ghana (4.0%).[3] Severe diseases, including pneumonia, are reported in 15% of COVID-19 cases, and 5% of cases also experience respiratory- and organ-related failures.[11] However, Africa's poor health systems, coupled with its fragile economies, call for strategies that efficiently make use of scarce resources in managing any unexpected exponential increase in outbreak of the disease.[12,13] Until April 2020, the total number of working ventilators in public hospitals across 41 countries in Africa was fewer than 2000, compared to more than 120,000 in the United States.[14] This requires strategies to make efficient use of scarce resources to manage any unexpected exponential increase in outbreak of the disease. One of these strategies may be to document the prevalent genes and mutations in the geographical settings of Africa in order to provide targeted interventions, such as effective drugs and vaccines development.

SARS-CoV-2 has variations in the viral genome.[15] Thus, as more complete sequencing information becomes available over time, the need to identify geographic virus variants may be important for future decision-making in resource allocation and provision of appropriate interventions. Therefore, the need to profile the viral genes and mutations in a reliable, robust, and systematic manner may be important.

The traditional method of producing systematic reviews could be employed to do the profiling by estimating the prevalence; however, high volumes of peer-reviewed articles and gray literature on the virus are being produced rapidly,[16] reflecting the urgency to deploy the most up-to-date information as quickly as possible. Hence, there is a need to conduct a living systematic review (LSR), which is a review characterized by continual surveillance of evidence, incorporating new evidence as it becomes available.[17] This would produce synthesized evidence based on high-quality studies using systematic, rigorous, and robust methods to ensure the evidence is as salient and recent as possible.

Living systematic reviews are characterized by regular updates of evidence that may provide health workers and policy makers with reliable and up-to-date evidence to inform their decision-making on practices and policy guidelines. The reliable and regularly updated evidence could also inform African governments' strategies to combat the COVID-19 pandemic. Currently, many approaches are being explored to create effective vaccines and drugs[7]; however, structured proteins that are considered to play important roles in the viral entry, fusion, and survival in the host genetic cells are considered a major target for drug and vaccine development.[18] Therefore, it is imperative to profile mutations (variants) of structural proteins; for instance, the prevalence of specific variants, such as D614G identified in the S protein, in two or more studies conducted in different countries within Africa could be profiled by pooling the existing variants. Any time a new prevalent of the variant becomes available, it can be incorporated in the pooling of results.[2]

There are continual worldwide clinical trials on drugs and candidate vaccines for COVID-19, of which a slim proportion (2.6%) is taking place in Africa.[19] Providing up-to-date prevalence of genomic variation may inform decisions in more studies carried out in the future on drug and vaccine development on the continent. In addition, no LSRs on the prevalence of specific structural protein mutations of SARS-CoV-2 in Africa were found after searching PROSPERO, Cochrane Database of Systematic Reviews, PubMed, and *JBI Evidence Synthesis*. Hence, there is a gap in using up-to-date and systematic methods of documenting prevalence of specific structural mutations (variants) in Africa to identify the predominant structural protein variants in specific geographical areas.

The objective of this LSR is to synthesize the available evidence on the prevalence of types of

SARS-CoV-2 genomic variations in Africa. More specifically, the aim is to estimate the prevalence of specific mutations (variants) associated with structural proteins in North Africa and sub-Saharan Africa.

Review questions

i) What are the specific variants of structural proteins in North Africa?
ii) What are the specific variants of structural proteins in sub-Saharan Africa?

Inclusion criteria

Participants

The LSR will consider studies that included individuals of all ages and sexes.

Condition

This LSR will include studies on the prevalence of SARS-CoV-2 genetic strains and mutations obtained from sequencing data of individual samples using next generation sequencing approaches.

Context

The LSR will consider studies on the prevalence of SARS-CoV-2 in any geographical setting in Africa. Specifically, published or unpublished data on the prevalence of SARS-CoV-2 genomic variants conducted in any country in Africa will be considered. The review will consider the pooling in the context of North Africa and sub-Saharan Africa, because they differ in terms of sociocultural characteristics, resources, and climatic conditions.[20]

Types of studies

Analytical observational studies, including analytical cross-sectional studies, will be considered for inclusion. This review will also consider descriptive observational studies including descriptive cross-sectional studies for inclusion.

Methods

The baseline LSR will be conducted in accordance with JBI methodology for prevalence reviews using the JBI System for the Unified Management, Assessment and Review of Information (JBI SUMARI; JBI, Adelaide, Australia).[21] In addition, useful features of the LSR protocol that are missing in the JBI guidelines will be incorporated using the Cochrane guidelines for LSR to augment the review protocol.[17]

Search strategy

The search strategy aims to find both published and unpublished studies. A three-step search strategy will be utilized in this review. An initial limited search of MEDLINE (PubMed) and CINAHL (EBSCO) was undertaken followed by analysis of the text words contained in the title and abstract, and of the index terms used to describe article. A second search using all identified keywords and index terms will be undertaken across all included databases. Thirdly, the reference lists of all identified reports and articles that meet the inclusion criteria will be searched for additional studies. Studies published in the English language will be considered for inclusion in this review. The African continent is predominantly made up of Francophone and Anglophone countries. Searching in French may give a true picture of the evidence, but the review team lack the capacity to effectively translate French into English. In addition, the team lacks sufficient funds to engage the services of a professional translator. Studies published from January 2020 will be considered for inclusion in this review. The start date for the search was chosen on the basis that no confirmed COVID-19 case in Africa was reported before that period.

The databases to be searched include MEDLINE (PubMed), LitCovid, Embase (Ovid), Scopus (Elsevier), CINAHL (EBSCO), ScienceDirect (Elsevier), and SciELO (Web of Science).

The search for unpublished studies will include the World Health Organization COVID-19 website, the Centers for Disease Control and Prevention COVID-19 website, Google Scholar, ProQuest Dissertations and Theses, conference proceedings, bioRxiv (preprints), medRxiv (preprints), and SSRN (preprints). The following initial keywords will be used: "COVID-19," "SARS-CoV-2," "mutations," "sequencing," "prevalence," "percentage," "genome," "strains," "genetic," "variants," "Africa."

An initial search strategy has been developed for MEDLINE (PubMed; see Appendix I).

Study selection

Following the search, all citations (published or unpublished data) that meet the inclusion criteria of the review will be collated and uploaded into Mendeley V.1.19.8 (Mendeley Ltd., Elsevier, Netherlands) to de-duplicate citations. Following a pilot test, titles and abstracts will then be screened by two independent reviewers for assessment against the inclusion criteria for the review. All citation details will be exported to MS Excel (Redmond, Washington, USA), and each full-text article that potentially meets the inclusion criteria of the review will be assessed against the inclusion criteria by two independent reviewers. Full-text articles that do not meet the inclusion criteria will be excluded, and the reason for exclusion will be provided in an appendix of the final baseline review. Any disagreements that arise between reviewers at any stage of the selection process will be resolved through discussion. The results of the search will be reported in full in the review and reported in a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram.[22]

Assessment of methodological quality

All citations that meet the inclusion criteria of the review will be critically appraised for methodological quality by two independent reviewers using the standardized critical appraisal instrument for prevalence studies developed by JBI.[21] Authors of articles will be contacted to request missing or additional information where required. Any disagreements between reviewers that arise during the critical appraisal process will be resolved through discussion. The results of the critical appraisal will be reported in narrative and tabular format in the review. Following critical appraisal, studies that do not meet a certain quality threshold will be excluded. This decision will be based on a total score greater than 5, which is more than 60% on the nine-item JBI critical appraisal tool.

Data extraction

Data will be extracted from citations included in the review by two independent reviewers using the JBI standardized data extraction instrument for prevalence studies.[21] The data extracted includes specific details about the condition, context, population, and study methods of significance to the review question and specific objectives. Any disagreements between the reviewers will be resolved through discussion

Data synthesis

Data for prevalent genomic variations extracted from included studies (published or unpublished) will, where possible, be pooled in statistical meta-analysis using JBI SUMARI.[21] Prior to pooling, data will be logarithmically transformed because all ratio data (such as prevalence data) tend to be skewed to the right.[23] Selected prevalence estimates will be combined statistically to obtain a pooled estimate of prevalence and corresponding 95% confidence interval. A funnel plot will be used to visually assess publication bias, and the Egger test will be used to statistically assess publication bias if there are 10 or more studies.[24] Heterogeneity will be assessed statistically using the standard χ^2 (Cochran's Q) and I^2 tests; any I^2 tests above 75% will be indicative of significant heterogeneity.[25] In addition, the heterogeneity among studies will be checked manually in terms of study population, study settings, conditions, and methods. Random-effects models will be used for pooling the prevalence data.[26] Subgroup analysis will be carried out based on covariates, such as study design, context, condition, year of study, and population, where they have the potential to have impact on summary estimates. Where statistical pooling is not possible, the findings will be presented in narrative form including tables and figures to aid in data presentation, where appropriate. If pooling is possible, the antilog for the pooled estimates of prevalence and the corresponding 95% confidence intervals will be found after pooling.

Living systematic review considerations

Electronic databases will be searched monthly and auto-alerts set up (where possible) to deliver a monthly search yield by email. Search methods and strategies will be reviewed yearly to ensure they reflect any terminology changes in the topic area or in the database. All gray literature will be searched monthly, except conference proceedings, which will be searched yearly.

Whenever any new evidence (published or unpublished data) relevant to the review is identified, data will be extracted and the risk of bias appropriately assessed. We will wait until the accumulating evidence has impact on the conclusion of the review before incorporating it and re-publishing the review. The impact will be measured by the increase in precision of the pooled estimate (ie, the reduction in the length of 95% confidence interval).

Each year, reviewers will consider the necessity for the review to continue to be living by assessing ongoing relevance of the question to decision-makers. The review will be transitioned out of living mode if it lacks relevance to decision-makers.

The LSR status will be updated every month by stating the last search date; the number of studies identified, together with the DOI or hyperlinks of studies; and a statement indicating whether the new evidence made an impact on the conclusion of the study. JBI SUMARI files (eg, PRISMA flow diagram, forest plots) will be updated accordingly.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, *et al*. A novel coronavirus from patients 218 with pneumonia in China. N Engl J Med 2019;382(8):727–33.
2. Mercatelli D, Giorgi FM. Geographic and genomic distribution SARS-CoV-2 mutations. Front Microbiol 2020;11:1800.
3. World Health Organization. Corona virus dashboard [internet]. WHO; n.d. [cited 2021 Mar 10]. Available from: https://covid19.who.int/?gclid=EAIaIQobChMIp8jooub6wIV1ODICh2k_QKfEAAYASAAEgJ5ivD_BwE.
4. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, *et al*. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell Host Microbe 2020; 27(3):325–8.
5. Malik YA. Properties of coronavirus and SARS-CoV-2. Malays J Pathol 2020;42(1):3–11.
6. Hirano T, Murakami M. COVID-19: a new virus, but a familiar receptor and cytokine release syndrome. Immunity 2020;52(5):731–3.
7. 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 (preprint) [internet]. 2020 [cited 2020 Nov 13]. Available from: https://www.biorxiv.org/content/10.1101/2020.04.29.069054v2.
8. Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Alev PK, *et al*. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of randomised controlled trials in Brazil, South Africa, and the UK. Lancet 2021;397(10269):99–111.
9. Baden LR, Sahly EL, Essink B, Kotloff K, Frey S, Novak R, *et al*. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. N Engl J Med 2021;384(5):403–16.
10. Logunov DY, Dolzhikova IV, Shcheblyakov DV, Tukhvatulin AI, Zubkova OV, Dzharullaevu DV, *et al*. Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: an interim analysis of a randomised controlled phase 3 trial in Russia. Lancet 2021;397:671–8.
11. Harries AD, Martinez L, Chakaya JM. Monitoring the COVID-19 pandemic in sub-Saharan Africa: focusing on health facility admissions and deaths. Int J Tuberc 2020;24(5):550–2.
12. Margolin E, Burgers WA, Sturrock ED, Mendelson M, Chapman R, Douglass N, *et al*. Prospects for SARS-CoV-2 diagnostics, therapeutics and vaccines in Africa. Nat Rev 2020;18:690–704.
13. African Development Bank Group. African economic outlook 2021. From debt resolution to growth: the road ahead for Africa [internet]. African Development Group. 2021 [cited 2021 Mar 15]. Available from: https://www.afdb.org/en/documents/african-economic-outlook-2021.
14. Maclean R, Marks S. 10 African countries have no ventilators. That's only part of the problem [internet]. The New York Times. 2020 Apr 18 [cited 2020 Sep 23]. Available from: https://www.nytimes.com/2020/04/18/world/africa/africa-coronavirus-ventilators.html?searchResultPosition=1.
15. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, *et al*. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020;395:565–74.
16. Teixeria da Silva JA, Tsigaris P, Erfanmanesh M. Publishing volumes in major databases related to Covid-19. Scientometrics 2021;126:831–42.
17. Brooker J, Synnot A, McDonald S, Elliot J, Turner T, Hodder R, *et al*. Guidance for the production and publication of Cochrane living systematic reviews: Cochrane reviews in living mode [internet]. 2019 [cited 2021 Mar 14]. Available from: https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/201912_LSR_Revised_Guidance.pdf.
18. Naqvi AAT, Fatima K, Mohammed T, Fatima U, Singh IK, Singh A, *et al*. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach. Biochim Biophys Acta Mol Basis Dis 2020; 1866(10):165878.
19. Nkeck JN, Ndoadoumgue AL, Temgoua MN. COVID 19 pandemic, status of clinical trials in Africa on May 2020: need to reinforce. Pan Afr Med J 2020;35(Suppl 2):87.
20. Henerson JV, Turner MA. Urbanization in developing world: too early or too slow? J Econ Prespect 2020;34:150–73.
21. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Chapter 5: systematic reviews of prevalence and incidence. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis [internet]. Adelaide: JBI; 2020 [cited 2021 Mar 10]. Available from: https://synthesismanual.jbi.global.
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021; 372:380.
23. Snell KIE, Ensor J, Debray TPA, Moons KGM, Riley RD. Meta-analysis of prediction model performance across multiple studies: which scale helps ensure between-study normality for the C-statistic and calibration measures? Stat Methods Med Res 2018;27(11):3505–22.
24. Jin ZC, Zhou XH, He J. Statistical methods for dealing with publication bias in meta-analysis. Stat Med 2015;34(2):343–60.
25. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–60.
26. Nikolakopoulou A, Mavridis D, Salanti G. Demystifying fixed and random effects meta-analysis. Evid Based Mental Health 2014;17(2):53–7.

Appendix I: Search strategy

MEDLINE (PubMed)

Search conducted on July 20, 2021.

Search	Query	Records retrieved
1	SARS-CoV-2 [Mesh] OR Coronavirus [Mesh] OR COVID-19 [Mesh] OR "Coronavirus Envelope" [tw] OR "Coronavirus M Proteins" [Mesh] OR "Coronavirus Nucleocapsid" [tw] OR "Spike Glycoprotein" [Mesh] OR NSP4 [tw] OR NSP7 [tw] OR "nucleocapsid phosphoprotein" [tw] OR ORF1ab [tw] OR "Spike protein" [tw] OR "COVID-19 virus" [tw] OR "SARS Coronavirus 2" [tw] OR 2019-nCov [tw] OR "Severe Acute Respiratory Syndrome Coronavirus 2" [tw] OR "Wuhan Coronavirus" [tw] OR Genome [tw] OR Betacoronavirus [tw] OR "structural proteins" [tw] OR "Angiotensin-Converting Enzyme 2" [Mesh] OR Polyprotein [Mesh] OR protein [tw] OR "Viral Protein" [Mesh] OR "Viral Genome Packaging" [Mesh] OR "Viral Packaging Sequence" [Mesh] OR "Viroporin Proteins" [Mesh]	2,294,056
2	Mutation [tw] OR sequencing [tw] OR variant [tiab] OR "genomic variation" [tiab] OR "Next Generation Sequencing" [tw] OR "genomes sequencing" [tw] OR strain [tw] OR isolated [tw] OR Spectrum [tw] OR Analysis [tw] OR Mutation [tw] OR sequencing [tw] OR variant [tiab] OR "genomic variation" [tiab] OR "Next Generation Sequencing" [tw] OR "genomes sequencing" [tw] OR strain [tw] OR isolated [tw] OR Spectrum [tw] OR Analysis [tw]	4,732,291
3	Prevalence [tw] OR Percentage [tw] OR Frequency [tw] OR Frequencies [tw] OR Proportion [tw]	1,629,244
4	Algeria [tw] OR Angola [tw] OR Benin [tw] OR Botswana [tw] OR "Burkina Faso" [tw] OR Burundi [tw] OR "Cape Verde" [tw] OR Cameroon [tw] OR "Central African Republic" [tw] OR Chad [tw] OR Comoros [tw] OR "Democratic Republic of Congo" [tw] OR Congo [tw] OR "Ivory Coast" [tw] OR Djibouti [tw] OR Egypt [tw] OR "Equatorial Guinea" [tw] OR Eritrea [tw] OR Ethiopia [tw] OR Gabon [tw] OR Gambia [tw] OR Ghana [tw] OR Guinea [tw] OR Guinea-Bissau [tw] OR Kenya [tw] OR Lesotho [tw] OR Liberia [tw] OR Libya [tw] OR Madagascar [tw] OR Malawi [tw] OR Mali [tw] OR Mauritania [tw] OR Mauritius [tw] OR Morocco [tw] OR Mozambique [tw] OR Namibia [tw] OR Niger [tw] OR Nigeria [tw] OR Rwanda [tw] OR "Sao Tome and Principe" [tw] OR Senegal [tw] OR Seychelles [tw] OR "Sierra Leone" [tw] OR Somalia [tw] OR "South Africa" [tw] OR "South Sudan" [tw] OR Sudan [tw] OR Swaziland [tw] OR Tanzania [tw] OR Togo [tw] OR Tunisia [tw] OR Uganda [tw] OR Zambia [tw] OR Zimbabwe [tw] OR "North Africa" [tw] OR "sub-Saharan Africa" [tw] OR Africa [tw]	309,176
5	Filters: Publication date from 2020/01/01 to 2021/12/31; Language: English; Species: Humans	
	Combination of results 1, 2, 3, 4, and 5 linked by the Boolean term "AND"	349