Could the Biogeographic Background Affect the Emergence of SARS-CoV-2 Variants?

Tetsuya Akaishi, MD, PhD

Department of Education and Support for Regional Medicine, Tohoku University, Seiryo-machi 1-1, Aoba-ku, Sendai, Miyagi, Japan.

J Gen Intern Med 37(10):2562–3
DOI: 10.1007/s11606-022-07632-z

By the end of 2021, there were 41 potentially important SARS-CoV-2 variant strains.[1] Among these, 12 lineages (29.3%) were thought to emerge from two countries, 8 (19.5%) from the UK and 4 (9.8%) from South Africa. Based on epidemiologic data released from the Johns Hopkins University,[2] COVID-19 cases in these two countries totalled approximately 13 million, accounting for 5.0% (3.9% for the UK and 1.1% for South Africa) of cases worldwide. This may be partially explained by the disproportionate testing and shared genome sequences from these countries. The numbers of the total reported COVID-19 cases and shared SARS-CoV-2 genome sequences, based on the data released from GISAID,[3] are shown in Fig. 1, together with the numbers of detected potentially consequential variants in each of the countries. The UK was among the leaders for testing and

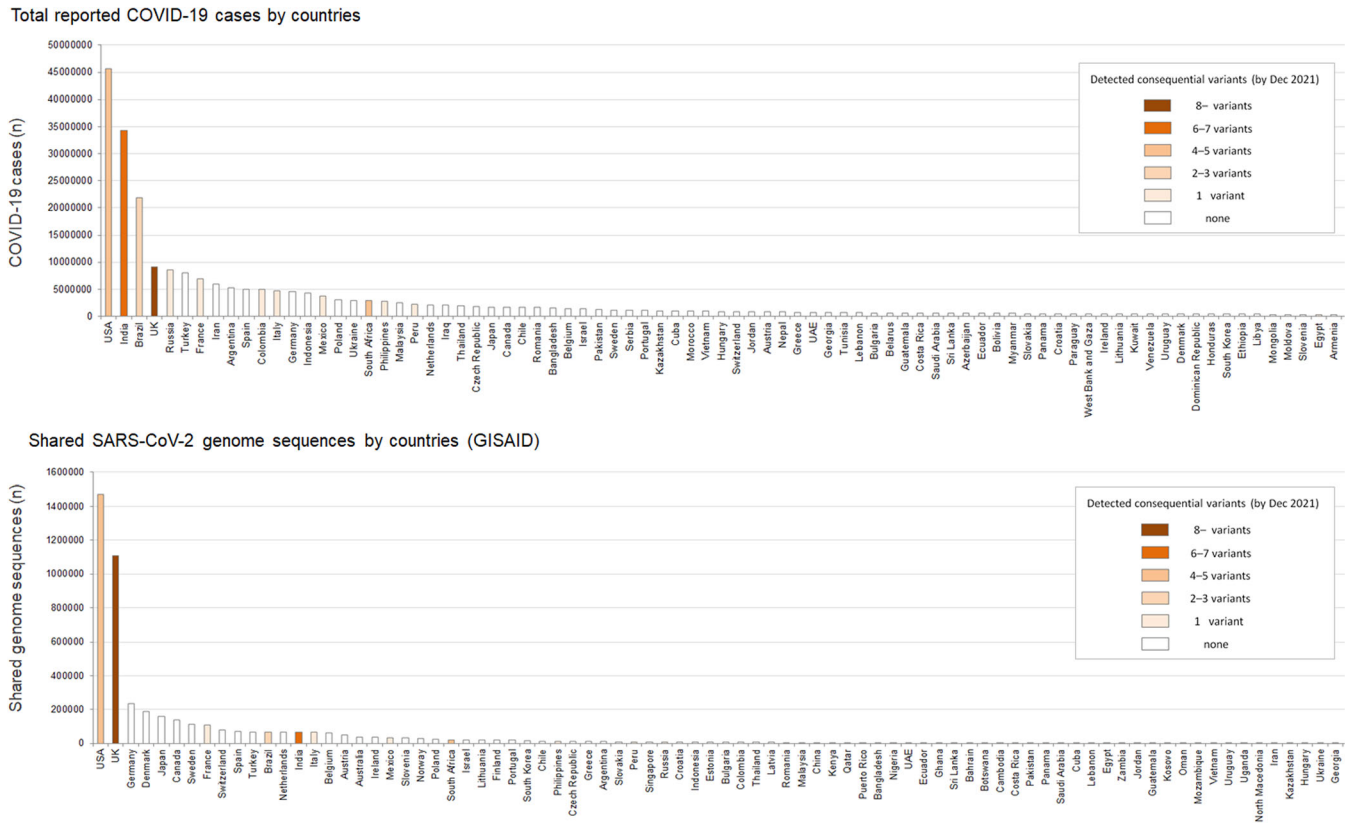

Fig. 1 Numbers of total reported COVID-19 cases and shared SARS-CoV-2 genome sequences by countries. The shown graphs are the numbers of total reported COVID-19 cases and shared SARS-CoV-2 genome sequences, based on the data released from GISAID, in each of the countries worldwide by December 2021. The colors of the bars represent the total numbers of detected potentially consequential SARS-CoV-2 variant lineages in each of the countries

Received December 15, 2021
Accepted April 19, 2022
Published online April 27, 2022

sharing the virus genome sequences throughout the pandemic accounting for 23.6% of the total shared genome sequences worldwide. While shared genome sequences from South Africa accounted for less than 1%, 4 potentially consequential variants were detected there. While the USA accounted for 31.3% of shared genome sequences and 18.8% of COVID-19 cases worldwide, the USA accounted for only 9.8% of the 41 variant lineages. India, a country with diverse biogeography, had the second largest number (13.9%) of total COVID-19 cases, and 14.6% of the consequential variant lineages, but only accounted for 1.4% of the shared genome sequences worldwide.

If we suppose that the occurrence of potentially consequential variant strains emerges randomly, the numbers of first detected variant strains in each country should be proportional to the number of accumulated cases occurring in each country. That this did not happen suggests that the occurrence of potentially consequential SARS-CoV-2 variant lineages may not be completely random among the total reported COVID-19 cases or among cases with shared SARS-CoV-2 genome sequences. Other unknown biogeographical factors may exist that may influence the development of consequential variants. Further studies that explore potential biogeographical or other unknown factors may be warranted.

Corresponding Author: *Tetsuya Akaishi, MD, PhD; Department of Education and Support for Regional Medicine, Tohoku University, Seiryo-machi 1-1, Aoba-ku, Sendai, Miyagi 980-8574, Japan (e-mail: t-akaishi@med.tohoku.ac.jp).*

Declarations:

Conflict of Interest: *The author declares that he does not have a conflict.*

REFERENCES

1. **World Health Organization.** Tracking SARS-CoV-2 variants (URL: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/).
2. **Johns Hopkins University, Corona virus resource center**. COVID-19 Dashboard (URL: https://coronavirus.jhu.edu/map.html).
3. **GISAID**. Sequences shared via GISAID since 10 January 2020. (URL: https://www.gisaid.org/index.php?id=208).