Cite as: L. Scott *et al.*, *Science* 10.1126/science.abn4543 (2021).

Track Omicron's spread with molecular data

Lesley Scott[1], Nei-yuan Hsiao[2,3], Sikhuline Moyo[4], Lavanya Singh[5,6], Houriiyah Tegally[5,6], Graeme Dor[1], Piet Maes[7], Oliver G. Pybus[8,9], Moritz U. G. Kraemer[8], Elizaveta Semenova[10], Samir Bhatt[11,12,13], Seth Flaxman[14], Nuno R. Faria[8,12,13,15], Tulio de Oliveira[5,6]*

[1]Department of Molecular Medicine and Haematology, University of the Witwatersrand, Johannesburg, South Africa. [2]Division of Medical Virology, Wellcome Centre for Infectious Diseases in Africa, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa. [3]National Health Laboratory Service, Cape Town, South Africa. [4]Botswana Harvard AIDS Institute Partnership and Botswana Harvard HIV Reference Laboratory, Gaborone, Botswana. [5]Centre for Epidemic Response and Innovation, Stellenbosch University, Stellenbosch, South Africa. [6]KwaZulu-Natal Research Innovation and Sequencing Platform, University of KwaZulu-Natal, Durban, South Africa. [7]Laboratory of Clinical and Epidemiological Virology, Katholieke Universiteit Leuven, Leuven, Belgium. [8]Department of Zoology, University of Oxford, Oxford, UK. [9]Royal Veterinary College, London, UK. [10]Department of Mathematics, Imperial College London, London, UK. [11]Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark. [12]MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, UK. [13]The Abdul Latif Jameel Institute for Disease and Emergency Analytics, School of Public Health, Imperial College London, London, UK. [14]Department of Computer Science, University of Oxford, Oxford, UK. [15]Instituto de Medicina Tropical, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil.

*Corresponding author. Email: tulio@sun.ac.za

On 26 November, the newly emerged variant Omicron was designated a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variant of concern (VOC) (*1*). Rapid polymerase chain reaction (PCR) test results could improve estimates of the prevalence of Omicron around the world. The widely used Thermo Fisher TaqPath COVID-19 PCR assay was valuable in tracking the spread of the Alpha (B.1.1.7) VOC (*2*) because a deletion of amino acids 69 and 70 in Alpha's spike gene (Δ69–70) yields a distinct absent S-gene (S^-) despite positive test results. The Delta VOC lacks this deletion and is therefore S-gene positive (S^+) on TaqPath PCR tests (*3*). The Omicron VOC shares the spike Δ69–70 deletion with Alpha, which has dropped to negligible levels worldwide. Therefore, the frequency of S^- results can be used as a rapid proxy for the frequency of Omicron cases, provided initial detection of local circulation had been confirmed by sequencing.

To put this data to use, countries should prioritize the release of daily counts of cases, hospitalizations, and deaths disaggregated by S^+, S^-, and unknown [e.g., (*4*)] as much as possible while taking logistical and privacy concerns into account. S-gene data could serve as a proxy for estimates of Omicron VOC prevalence (*5*) and help to understand the fraction of infections caused by Omicron (versus Delta) and the severity of Omicron cases, as measured by mortality and hospitalization. In low-resource settings where genomic sampling is absent, infrequent, or characterized by long turnaround times (*6*), S-gene data will help reveal the risk Omicron poses to pandemic control. Finally, through synthesis with serological data (*7*), S-gene data—shared in real time—could help to evaluate the degree of immune protection conferred by natural- and vaccine-elicited immunity in Omicron cases.

Although S-gene data will be informative, preferential sequencing of samples with an S^- result will lead to virus genomic datasets that are unrepresentative of the true underlying spatiotemporal prevalence of Omicron. To provide adequate context for genome sequences, depositors to the Global Initiative on Sharing All Influenza Data (GISAID) database should use the newly introduced nonmandatory "sampling strategy" field to note how cases are selected and sampled for virus genome sequencing, including whether samples were specifically targeted for sequencing based on S^- PCR results. [We have used this field to plot the first 115 Omicron submissions to GISAID, stratified by sampling strategy (*8*).] Virus genomic datasets then can be compiled from cases known to have been sampled randomly from a given population and analyzed to generate more accurate estimates of Omicron's growth relative to other variants. Standard sampling strategies include random community sampling [the preferred sampling strategy for estimating lineage growth (*6*, *9*)], targeted surveillance of defined subpopulations (e.g., vaccine breakthrough cases or international travelers), and enhanced sampling to investigate specific outbreaks or clusters.

Tracking SARS-CoV-2 lineages and variants, including Omicron, through GISAID (*10*), Pango lineages (*11*), and NextStrain (*12*) has provided valuable information about their spread in close to real time. However, genome sequencing intensities and turnaround times vary substantially across the world; in most countries, it takes more than 21 days after sample collection to deposit data in GISAID (*6*). Moreover, sampling strategies used to select samples for sequencing are heterogeneous across geographic regions (*6*) and often not reported in virus genome metadata. To evaluate risk and guide policy, there is an urgent need to incentivize the quick sharing of well-annotated genomic and S-gene–stratified surveillance data globally. By acting with speed, transparency, and consistency, we can establish norms to support better global responses to newly emerging variants.

REFERENCES AND NOTES

1. World Health Organization (WHO), "Classification of Omicron (B.1.1.529): SARS-CoV-2 variant of concern" (2021); www.who.int/news/item/26-11-2021-

classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern.
2. E. Volz *et al.*, *Nature* **593**, 266 (2021).
3. S. Mishra *et al.*, *EClinicalMedicine* **39**, 101064 (2021).
4. Office for National Statistics, "Coronavirus (COVID-19) Infection Survey, UK Statistical bulletins" (2021); www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/previousReleases.
5. A. J. Kucharski, E. B. Hodcroft, M. U. G. Kraemer, *Lancet Reg. Health Eur.* **9**, 100215 (2021).
6. A. F. Brito *et al.*, *medRxiv* 10.1101/2021.08.21.21262393 (2021).
7. R. K. Arora *et al.*, *Lancet Infect. Dis.* **21**, e75 (2021).
8. "COVID-19 Omicron Repository," GitHub (2021); https://github.com/CADDF-CENTRE/covid19omicron.
9. WHO, "Genomic sequencing of SARS-CoV-2: A guide to implementation for maximum impact on public health" (2021); www.who.int/publications/i/item/9789240018440.
10. Y. Shu, J. McCauley, Euro Surveill. **22**, 30494 (2017).
11. A. O'Toole, "Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2, Virological.com (2021); https://virological.org/t/tracking-the-international-spread-of-sars-cov-2-lineages-b-1-1-7-and-b-1-351-501y-v2/592.
12. J. Hadfield *et al.*, *Bioinformatics* **34**, 4121 (2018).

Published online 9 December 2021
10.1126/science.abn4543

Track Omicron's spread with molecular data

Lesley ScottNei-yuan HsiaoSikhuline MoyoLavanya SinghHouriiyah TegallyGraeme DorPiet MaesOliver G. PybusMoritz U. G. KraemerElizaveta SemenovaSamir BhattSeth FlaxmanNuno R. FariaTulio de Oliveira

Science Commentary, **Ahead of Print**