Improving estimates of SARS-CoV-2 Omicron prevalence and growth rates through increased data sharing

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

1. Department of Molecular Medicine and Haematology, University of the Witwatersrand, 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. Botswana Harvard AIDS Institute Partnership and Botswana Harvard HIV Reference Laboratory, Gaborone, Botswana
4. Centre for Epidemic Response and Innovation, Stellenbosch University, Stellenbosch, South Africa
5. KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), University of KwaZulu-Natal, Durban, South Africa
6. Laboratory of Clinical and Epidemiological Virology, KU Leuven, Leuven, Belgium
7. Department of Computer Science, University of Oxford, Oxford, United Kingdom
8. Department of Zoology, University of Oxford, Oxford, United Kingdom
9. Department of Mathematics, Imperial College London, London, UK
10. Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark
11. MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, United Kingdom
12. The Abdul Latif Jameel Institute for Disease and Emergency Analytics (J-IDEA), School of Public Health, Imperial College London, London, United Kingdom
13. Instituto de Medicina Tropical, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil

* Correspondence to T.O.: tulio@sun.ac.za, S.B.: samir.bhatt@sund.ku.dk, S. F.: seth.flaxman@cs.ox.ac.uk, N. R. F.: nfaria@ic.ac.uk

Against a background of global dominance by SARS-CoV-2 variant of concern B.1.617.2 (Delta), a newly emerged variant, Omicron (also named B.1.1.529, GR/484A or 21K) was designated a variant of concern on 26th November 2021[1]. Given its very recent detection, little is known about the epidemiological or clinical properties of Omicron. Since the identification and exemplary early data sharing from Southern African researchers on 24th November 2021, global sequencing efforts are now centered on detecting Omicron in order to characterize its properties. The worldwide spread of Omicron and the critically important practice of rapid data sharing means that the number of Omicron sequences deposited in GISAID will most likely soon increase markedly. While increases in data will be useful for improving diagnostics and therapeutics as well as epidemiological investigations of virus transmission and spread (e.g. detection of local vs imported cases[2]), naive estimates of prevalence and growth rates based on either the observed fraction or count of Omicron sequences over time could suffer from extreme biases. These biases may arise for instance from selective sequencing of clinical/environmental samples or isolates that, as we discuss in detail below, have a diagnostic PCR result consistent with an S gene dropout, often referred to as spike gene target failure (SGTF).

The widely used Thermo Fisher TaqPath COVID-19 PCR assay was instrumental in tracking the spread of the B.1.1.7 (Alpha) variant of concern (VOC) in England[3] because a deletion at Alpha's S amino acid positions 69 and 70 (spike Δ69–70) in the spike (S) protein meant that only the N and ORF1a gene PCR targets were positive. This S-gene PCR negative signature is also denoted S-gene target failure (SGTF). By December 2020, SGTF was a highly accurate and rapid proxy for estimating growth of the Alpha VOC, with many SGTF cases later confirmed to be Alpha by whole viral genome sequencing. In spring 2021 in England, the Alpha variant became the dominant circulating lineage. However, the early spread of the Delta VOC, which lacks the spike Δ69–70 deletion and is therefore S-gene positive on TaqPath PCR tests, was tracked through a decreasing fraction of SGTF results[4]. As the Omicron VOC shares the spike Δ69–70 deletion and Alpha VOC cases have now dropped to negligible levels worldwide, the frequency of SGTF can once again be used as a rapid proxy for the frequency of Omicron cases. But herein lies the challenge: preferentially sequencing samples with a SGTF result,

even if all viral sequences (not just Omicron) are subsequently deposited in GISAID, will lead to virus genomic datasets which are unlikely to be representative of the true underlying prevalence of Omicron in any given location at any point in time.

We present two simple and concrete recommendations: the first is that depositors to GISAID make use of the newly introduced non-mandatory “sampling strategy” field (see first 115 Omicron submissions to GISAID until 27 November 2021 plotted, stratified by sampling strategy, on our GitHub[5]), noting selection, sampling, and reporting strategies (including whether samples were selected for sequencing based on PCR results indicative of SGTF). Virus genomic datasets can then be collated from samples representing a random selection of cases in a given location and analysed to generate accurate estimates of Omicron relative epidemic growth. Standard sampling strategies typically include random community sampling (preferred sampling strategy for estimating lineage growth[2,6]), outbreak investigation (including selection of SGTF dropout samples), vaccine breakthroughs and sequencing of cases in returning travellers.

Rapidly tracking SARS-CoV-2 lineages, including Omicron, through GISAID[7], Pango Lineages[8] and NextStrain[9] has provided valuable understanding of their spread in near-real time. However, genomic sequencing intensities and turnaround times vary significantly across the world, with most countries requiring >21 days of turnaround times between sample collection and data deposition on GISAID[6]. Moreover, sampling strategies used to select samples for sequencing are heterogeneous across geographic regions mostly due to intermittent funding resources and investment in research[6].

Delta currently accounts for the large majority of cases worldwide[6] and this VOC is S-gene positive (S+) on TaqPath PCR. Omicron cases are characterised by a contrasting S-gene negative (S-) profile on PCR assays and cases are expected to increase over the next weeks. Thus, our second recommendation, subject to privacy concerns, is that countries start prioritising the release of daily counts of cases and death data disaggregated by S+, S-, and unknown (as reported for instance, by the Office for National Statistics Infection survey[10]).

In the current context of heterogeneous genomics capacities, SGTF data could serve as a proxy for rapid and more accurate estimates of epidemiological quantities associated with the new VOC[11]. Specifically, (i) SGTF data could serve as a second data source to validate the fraction

of infections which are caused by Omicron versus Delta as Omicron spreads worldwide. This can further help to contextualize potential biases associated with non-representative global genomic sampling efforts. Also, (ii) SGTF data can allow the stratification of mortality and hospitalisation data to facilitate the statistical estimation of population changes in severity from observational data. Moreover, (iii) in lower resource settings where genomic sampling is absent, poor or characterised by long turnaround times[6], SGTF data can help inform on a global scale on the risk Omicron poses to pandemic control. The benefits of the use of SGTF data for Omicron monitoring in low resource settings could further be made cost effective through the use of VOC PCR alternatives[12]. Finally, (iv) through data synthesis with serological data[13], SGTF data shared in real-time could further help to evaluate the degree of immune protection conferred by natural- and vaccine-elicited immunity to the Omicron VOC.

The many unknowns regarding Omicron's relative transmissibility and impact on immune escape means that the risk profile this new variant poses to the global pandemic cannot be easily determined. To evaluate this risk and guide policy there is an urgent need to incentivize the quick sharing of both well-annotated genomic and SGTF stratified surveillance data globally. By acting with speed, transparency, and consistency we can establish norms to support better global responses to novel variants.

References:

1 Classification of Omicron (B.1.1.529): SARS-CoV-2 variant of Concern. https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern (accessed Nov 27, 2021).

2 WHO. Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health. 2021; published online Jan 8. https://www.who.int/publications/i/item/9789240018440 (accessed Nov 27, 2021).

3 Volz E, Mishra S, Chand M, *et al.* Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* 2021; **593**: 266–9.

4 Mishra S, Mindermann S, Sharma M, *et al.* Changing composition of SARS-CoV-2 lineages and rise of Delta variant in England. *EClinicalMedicine* 2021; **39**: 101064.

5 COVID-19 Omicron Repository. GitHub https://github.com/CADDE-CENTRE/covid19omicron (accessed Nov 30, 2021).

6 Brito AF, Semenova E, Dudas G, *et al.* Global disparities in SARS-CoV-2 genomic surveillance. *medRxiv* 2021; published online Aug 26. DOI:10.1101/2021.08.21.21262393.

7 Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance. 2017; **22**. DOI:10.2807/1560-7917.es.2017.22.13.30494.

8 Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2. 2021; published online Feb 4. https://virological.org/t/tracking-the-international-spread-of-sars-cov-2-lineages-b-1-1-7-and-b-1-351-501y-v2/592 (accessed Feb 10, 2021).

9 Hadfield J, Megill C, Bell SM, *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018; **34**: 4121–3.

10 Coronavirus (COVID-19) Infection Survey, UK Statistical bulletins. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/previousReleases (accessed Nov 27, 2021).

11 Kucharski AJ, Hodcroft EB, Kraemer MUG. Sharing, synthesis and sustainability of data analysis for epidemic preparedness in Europe. *Lancet Reg Health Eur* 2021; **9**: 100215.

12 Vogels CBF, Breban MI, Ott IM, *et al.* Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. *PLoS Biol* 2021; **19**: e3001236.

13 Arora RK, Joseph A, Van Wyk J, *et al.* SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. *Lancet Infect Dis* 2021; **21**: e75–6.

Acknowledgments

We gratefully thank originating and submitting laboratories who contributed genomic data and corresponding metadata on Omicron cases to GISAID analysed in Figure 1: CERI, Centre for Epidemic Response and Innvoation, Stellenbosch University and KRISP, KZN Research Innovation and Sequencing Platform, UKZN, South Africa; Hong Kong Department of Health, China; KU Leuven, Rega Institute, Clinical and Epidemiological Virology, Belgium; Botswana Harvard HIV Reference Laboratory, Botswana; National Institute for Communicable Diseases of the National Health Laboratory Service, South Africa; and the Shamir Medical Center (Asaf Harofe), Israel.