The lag in SARS-CoV-2 genome submissions to GISAID

To the Editor — Genomic surveillance of the evolving SARS-CoV-2 strains is an important tool for helping control the pandemic[1]. For efficient surveillance, the first major requirement for analysis of how the virus is evolving and spreading is the availability of all sequenced genomes on an open-access platform that is accessible to researchers worldwide. Therefore, soon after researchers became aware of COVID-19, toward the end of 2019, the Global Initiative on Sharing All Influenza Data (GISAID), an existing platform for sharing influenza virus sequences, began receiving deposits of SARS-CoV-2 genome sequences. Here we report an analysis of the median collection to submission time (CST) lag for SARS-CoV-2 sequences to GISAID on a country-by-country basis. Our results suggest that researchers in the United Kingdom are the fastest, logging sequences in a median time of 16 days, which is not only more than 5 times as fast as the upload times of sequences originating from industrial countries such as Japan or Canada, but also 18 times as fast as that of Qatar, among the countries that have sequenced over 1,000 genomes.

As of now, GISAID is the largest open-access portal, hosting the genome sequences and related epidemiological and clinical data of more than 1.7 million SARS-CoV-2 strains. Thanks to ongoing genomic surveillance using GISAID data, several new SARS-CoV-2 variants, such as B.1.1.7 (Alpha; first identified in the United Kingdom), B.1.351 (Beta; first identified in South Africa), B.1.1.28 (Gamma; P.1, first identified in Brazil), B.1.617.2 (Delta; first identified in India), B.1.617.1 (Kappa; first identified in India), P.3 (Theta; first identified in the Philippines), and B.1.427 and B.1.429 (Epsilon; first identified in the United States), have been identified[2–5]. This information has been used to update public health policies for the control of COVID-19 infections[6,7].

Considering the benefits of genomic surveillance[6,8], scientists have pressured countries to increase their sequencing capacity, and this has led to several initiatives, such as COG-UK (United Kingdom; https://www.cogconsortium.uk/), INSA-COG (India; https://pib.gov.in/PressReleseDetailm.aspx?PRID=1684782), NGS-SA (South Africa; http://www.krisp.org.za/ngs-sa/ngs-sa_network_for_genomic_surveillance_south_africa/) and SPHERES (United States; https://www.cdc.gov/coronavirus/2019-ncov/covid-data/spheres.html). Even so, although an increasing fraction of COVID-19 samples are being sequenced[9], an equally important issue is how soon the sequences are being submitted to GISAID. Rapid submission is important as it enables the international community to analyze the variants emerging around the world quickly and provide actionable information to governments.

Our statistical analysis (Fig. 1, Supplementary Fig. 1 and Supplementary Tables 1 and 2) for the 1,718,035 SARS-CoV-2 strains submitted to GISAID (as of 27 May 2021) has determined that CST lag per strain ranges from 1 day to over 1 year. We have also calculated the median CST lag for each country. Examining the median CST lag values for countries that have sequenced over 1,000 SARS-CoV-2 genomes, we note that the CST lag from the United Kingdom is the shortest (namely, 16 days for over 417,000 genomes). This is almost a week faster than the CST lags of 25 and 26 days for over 590,000 and 498,000 genomes in the rest of Europe and the United States, respectively. For Canada, the CST lag is over five times as long: 88 days for over 44,000 genomes. Among the countries of Oceania, the CST lag for New Zealand is 40 days for over 1,000 genomes, whereas for Australia it is 51 days for over 17,000 genomes. In Asia, the median CST lag is 72 days for over 89,000 genomes, with Singapore having the shortest lag, 26 days for 2,405 genomes, and Qatar the longest, 289 days for 2,298 genomes. India's median CST lag is 57 days for 15,614 genomes whereas Japan, which has sequenced the most genomes in Asia, has taken 79 days for over 37,000 genomes. For South American countries, the median lag is 61 days for over 18,000 genomes, whereas countries in Africa have taken 50 days for over 7,000 genomes (Supplementary Table 2).

Coming to the rate of sequencing, top-performing countries Iceland, Australia, New Zealand and Denmark have sequenced ~77%, 59%, 39% and 35% of their positive samples, respectively (Supplementary Table 1). The United States and United Kingdom have sequenced over 400,000 genomes each, which is 1.5% and 9.3% of their respective positive samples. India, being the second-largest country on the basis of both total population and known COVID-19 cases, has sequenced a mere 0.05% of the collected samples. On average, African, Asian and South American countries have sequenced a mere 0.36%, 0.21% and 0.07% of their total COVID-19 samples, respectively, whereas this number is 1.9%, 1.4% and 37% for European, North American and Oceania countries. Population-wise, most of the European countries, the United States, Israel (Asia) and the island of Réunion have sequenced samples from over 1,000 people per million population (ppmp). Among countries with over 100 million population, including Brazil (50 ppmp), India (11 ppmp), Indonesia (6 ppmp), Nigeria (4 ppmp) and Pakistan (1 ppmp), only the United States (1,497 ppmp) and Japan (297 ppmp) have sequenced over 100 ppmp. Cumulatively, African, Asian and South American countries have sequenced only 14, 21 and 49 ppmp whereas this number is 1,198, 948 and 607 ppmp for European, North American and Oceania countries (Supplementary Table 2).

Several reasons may explain the delay in sequence submissions to GISAID. Submission times are based on (i) the time taken from sample collection from a patient to RNA isolation in the lab and its dispatch to the sequencing center and (ii) the time from RNA sample arrival at the sequencing center to uploading of the sequence to GISAID. Countries like the United Kingdom and Denmark with a short median CST lag have strong public health systems, allowing efficient sample and metadata collection and smoother coordination between the sample collection center, the RNA isolation lab and the sequencing lab. Countries without such a strong system are at a disadvantage and may face additional logistical problems in sample and metadata collection and shipping because of lockdown-related restrictions. Several countries might have a shortage of labs that can handle COVID-19 samples or might have an overly centralized system wherein only a few labs are authorized to handle such samples, causing a delay in sequencing and submission. A paucity of funds or restrictions on importing reagents and
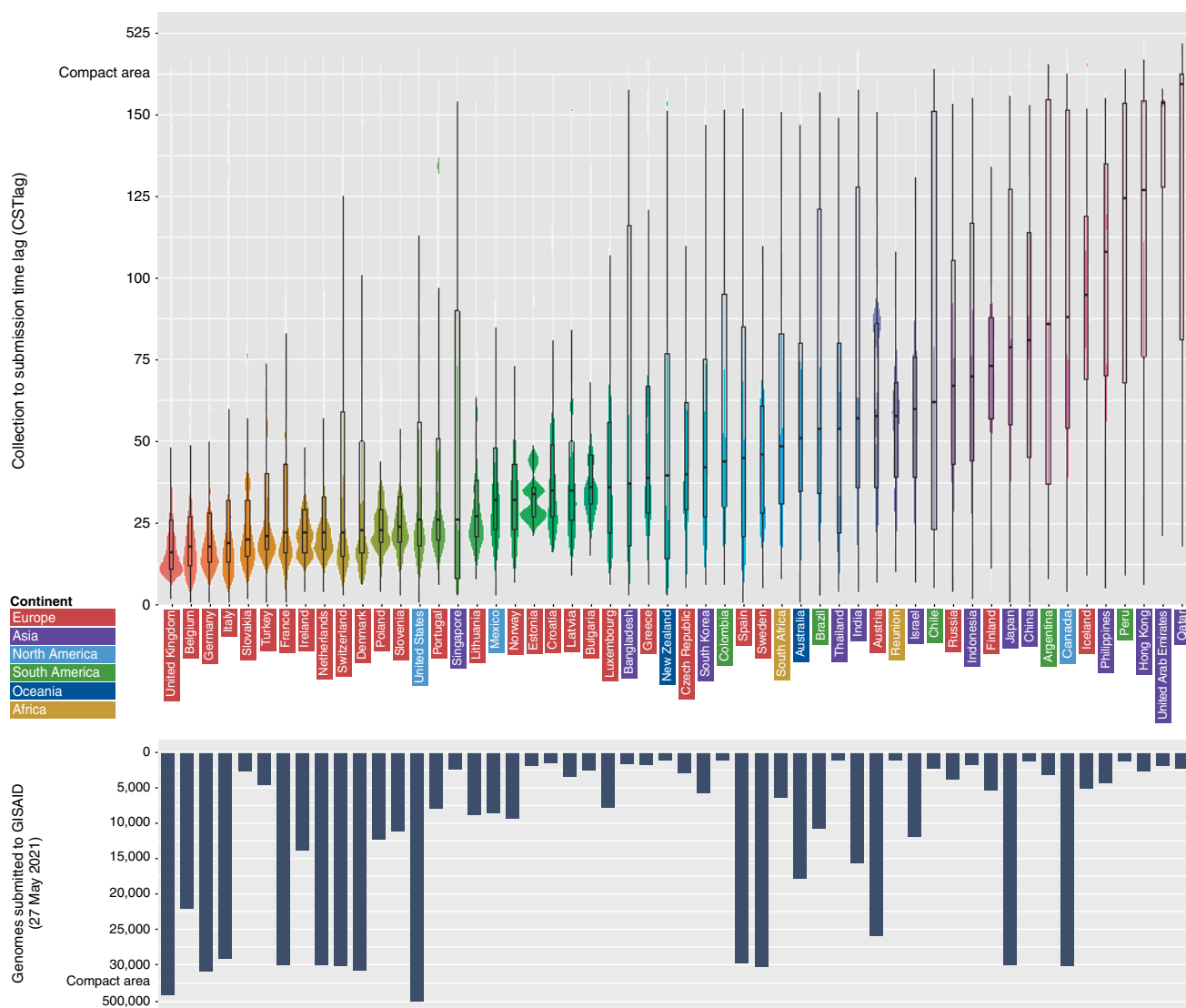
Fig. 1 | Violin plot illustrating the CST lag values for the 54 countries that have sequenced over 1,000 genomes. The box plot inside the violin plot depicts the median CST lag per country. Outlier CST lag entries are not shown. Country name is color-coded according to continent. We have also graphed the relative distribution of the number of genome sequences submitted by each country as a bar plot.

equipment would also add to the delay. The use of older sequencing technologies that are low throughput and more expensive per sample would complicate matters further.

Most of the countries with a short CST lag are industrialized nations that are likely to have strong linkages between the clinical and scientific establishments, although Japan and Canada are outliers. This is not always so for other countries. Some of the countries with a longer CST lag have a less developed public health system. They might also have had to establish new collaborations and institutional arrangements to help deal with the pandemic. All of this would have taken time, which would have affected work on the ground. Some of the possible causes for delay

listed above are known to have been true in India, for instance, and are being resolved[10,11].

Sometimes, even after rapid sequencing, genomes may not be promptly uploaded to GISAID, and there may be several reasons for that. First, the importance of genomic surveillance may not have been well understood, especially in the early months of the pandemic. Second, there may be a wish to withhold information, to publish or patent first. Although there is a general understanding that scientists do not publish work based on others' data unless the latter are acknowledged or have already published on them, there have been breaches of this professional norm, leading to some hesitation in sharing unpublished data[12]. Third, several governments may be

particularly sensitive to the issue of virulent strains being named after their countries. The WHO initiative of renaming variants with Greek letters may help resolve this issue[5]. Finally, in many countries, significant bureaucratization or political interference at various steps, from sample collection to uploading sequences to GISAID, can add to the delay. In several countries—including the United States—where most of the testing and sampling is carried out by private diagnostic labs, there is no financial incentive to share data, and the labs there may prefer to discard samples rather than pay for their storage or shipping. Although one does not know the extent of various problems in each country, clearly far more samples have been tested than have

been sequenced; and far more samples have been sequenced than are represented in GISAID.

In countries with a longer CST lag, the new variants may have enough time to establish themselves across a region[13] if quick tracking, tracing and actions to stop transmission are not undertaken. Therefore, this issue must receive urgent attention and bottlenecks that prevent a lower CST lag must be addressed.

Overall, an effective genomic surveillance system requires not only sequencing a major fraction of SARS-CoV-2 strains from COVID-19 patients, but also rapid genome submission to open access platforms like GISAID. This will enable researchers across the globe to track the evolved variants and their mutations, epidemiology and biological consequences, which will provide crucial inputs for appropriate and effective public health policies ☐

Kishan Kalia, Gayatri Saberwal [ID] and Gaurav Sharma [ID] ✉

Institute of Bioinformatics and Applied Biotechnology (IBAB), Bengaluru, India.
✉e-mail: gauravsharma@ibab.ac.in

References

1. Lo, S. W. & Jamrozy, D. *Nat. Rev. Microbiol.* **18**, 478 (2020).
2. Anonymous. *Lancet* **397**, 445 (2021).
3. Burki, T. *Lancet* **397**, 462 (2021).
4. Abdool Karim, S. S. & de Oliveira, T. *N. Engl. J. Med.* **384**, 1866–1868 (2021).
5. WHO. *WHO Activities* https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (2021).
6. Cyranoski, D. *Nature* **589**, 337–338 (2021).
7. Grubaugh, N. D., Hodcroft, E. B., Fauver, J. R., Phelan, A. L. & Cevik, M. *Cell* **184**, 1127–1132 (2021).
8. Anonymous. *Nat. Biotechnol.* **39**, 527 (2021).
9. Geoghegan, J. L. et al. *Emerg. Infect. Dis.* **27**, 1317–1322 (2021).
10. Office of the Principal Scientific Adviser to the Government of India. Data access and enhancing research collaborations of urgent importance. https://static.psa.gov.in/psa-prod/psa_custom_files/Data%20access_PSA%20website.pdf (2021).
11. Agrawal, A. *Nature* **594**, 9 (2021).
12. Van Noorden, R. *Nature* **590**, 195–196 (2021).
13. Ascoli, C. A. *Nat. Biotechnol.* **39**, 274–275 (2021).

ⓡ Check for updates

The need for new test verification and regulatory support for innovative diagnostics

To the Editor — The pressing need to address the COVID-19 pandemic led to the creation of novel large-scale cooperative programs among the US government, including the US National Institutes of Health (NIH) and the US Food and Drug Administration (FDA); academia; and private industry. A case in point is the NIH's Rapid Acceleration of Diagnostics (RADx), the diagnostic testing arm of Operation Warp Speed — a public–private partnership to fast-track SARS-CoV-2 vaccines, diagnostics and therapeutics. As the name suggests, the goal of RADx is to accelerate development, verification, validation, FDA Emergency Use Authorization (EUA) and deployment of diagnostic tests to detect infection with SARS-CoV-2. Availability of fast, accurate and inexpensive testing is a key component of efforts stemming from the pandemic. RADx was created in response to our nation's declared state of emergency that caused a sharp increased demand for testing in traditional clinical laboratories, resulting in a nationwide shortage. This cascade of events motivated swift innovation of new technologies using alternative test materials and methods, and even alternative biospecimen sample types. As investigators

in the RADx initiative, we gained a unique perspective on the frenetic pace of test development, performance testing and regulatory assessments in response to the COVID-19 pandemic. Consistent with an editorial recently published here in *Nature Biotechnology*[1], our experience suggests that RADx provides a high-reward approach for a relatively small investment. Furthermore, we believe the RADx model could be broadly established and applied beyond the current pandemic as a foundational resource to improve the developmental pathway for novel laboratory and point-of-care tests for all varieties of diseases.

For its RADx infrastructure, the NIH leveraged its Point-of-Care Technology Research Network (POCTRN), of which our center, the Atlanta Center for Microsystems Engineering Point-of-Care Technologies (ACME POCT), serves as the national RADx test verification hub. In this role, and in collaboration with the FDA, we provided independent and impartial verification data on the performance of COVID-19 diagnostic tests developed by private companies or academic inventors. RADx and ACME POCT also created a close working partnership with FDA

leadership and regulatory experts to convey clear regulatory guidelines and processes to developers, who often had little experience in this area. RADx also provided test developers access to a network of experienced entrepreneurs to help them avoid common mistakes that could hinder innovation, evaluation and deployment of COVID diagnostics.

The ACME POCT has evaluated over 80 different diagnostic technologies for detection of SARS-CoV-2 infection. These products, which typically detect either SARS-CoV-2 antigens or SARS-CoV-2 RNA, originated from a spectrum of applicants ranging from well-established companies to small teams from academic laboratories. For test verification at the ACME POCT, each diagnostic technology was put through a battery of different evaluations to determine preliminary analytical and clinical performance (for example, limits of detection, sensitivity, specificity and cross-reactivity), repeatability and usability. Whenever possible, tests were cross-compared with one another to provide the NIH and RADx leadership the most objective performance data on which to base their decisions to fund test validation, further manufacturing,