

## VIEWPOINTS

## Global variation in sequencing impedes SARS-CoV-2 surveillance

Dana C. Crawford<sup>1</sup>\*, Scott M. Williams<sup>1</sup>

Departments of Population and Quantitative Health Sciences and Genetics and Genome Sciences, Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, Ohio, United States of America

\* [dana.crawford@case.edu](mailto:dana.crawford@case.edu)

## OPEN ACCESS

**Citation:** Crawford DC, Williams SM (2021) Global variation in sequencing impedes SARS-CoV-2 surveillance. *PLoS Genet* 17(7): e1009620. <https://doi.org/10.1371/journal.pgen.1009620>

**Editor:** Gregory S. Barsh, HudsonAlpha Institute for Biotechnology, UNITED STATES

**Published:** July 15, 2021

**Copyright:** © 2021 Crawford, Williams. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This publication was made possible by the Clinical and Translational Science Collaborative of Cleveland (to DCC), UL1TR002548 from the National Center for Advancing Translational Sciences (NCATS) component of the National Institutes of Health and NIH Roadmap for Medical Research (to DCC). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: SMW is a section editor of PGEN.

## Viewpoint

Surveillance is essential to successful, rapid response of infectious disease outbreaks. While public health surveillance has historically focused on monitoring clinical cases and consequences of infection (e.g., case reports and hospitalizations) [1], technological advances in genomic sequencing rooted in the Human Genome Project and other large-scale investments in human genetics and genomic research and technologies now allow the unprecedented opportunity for pathogen surveillance down to base pair patterns of variation. Despite the availability and ubiquity of sequencing in several countries, the adoption of genomics as a strategy for pathogen surveillance has been slow, difficult, and inconsistent.

An extreme example of the slow adoption is that, as of early April 2021, the United States ranked 33th in the world in Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) sequencing for variant surveillance, up from 36th a few weeks earlier [2–4]. For a country that led the sequencing of the human genome more than 2 decades ago, the lack of accrual of sequencing data in the midst of this pandemic is unbelievable and problematic. Capacity and expertise exist across several key academic centers and industry partners, yet these resources remained relatively dormant until recently with fewer than 7,000 new sequences the week of March 6 from more than 415,000 new cases in the same period [5–7]. At that point in time, several commentaries and interviews [2,3,8] provided factors associated with this genomic bottleneck, e.g., lack of funding, lack of robust sample tracking and data pipelines, and strict regulations governing biospecimen and data sharing. These variables, which are not unique to the US, have been successfully addressed by other countries such as Denmark and the United Kingdom (UK) [9]. Given the role of the US in developing and utilizing sequencing technologies in the study of human diseases, initial failure to extend this into the study of SARS-CoV-2 to track viral mutations and improve public health represents a troubling, self-inflicted barrier to battling Coronavirus Disease 2019 (COVID-19), especially in light of the high prevalence of SARS-Cov-2 infection in the US over the past year.

Historically, genomic studies in the US have been financially and scientifically well supported by several agencies including the US Department of Energy, National Institutes of Health (NIH), and private industry. Today's mega-biobanks with electronic health records (EHRs) and other health-related data linked to DNA samples collected for genome-wide genotyping and whole genome sequencing are efforts supported by the Veterans Administration [10], medical centers, NIH, and private industry. Nearly absent from the US genomic-centric efforts is explicit collaboration with and investment by the primary public health agency, the Centers for Disease Control and Prevention (CDC). Maintaining zip code is more relevant to health than genetic code [11], the CDC did not initially prioritize genomic research with

respect to public health. With the real time evolution of SARS-CoV-2 and the resulting impact on disease transmission and disease severity, this bias has created a gaping hole in our understanding of the trajectory of COVID-19. Recently, this has been at least partially addressed by new initiatives and funding through programs such as National SARS-CoV-2 Strain Surveillance (NS3) system. Initiated in November 2020, this program is partnering the CDC as of early 2021 with state health departments to process and sequence 750 samples per week and with commercial diagnostic labs to sequence 6,000 samples per week [12]. CDC is also working with 7 universities to conduct genomic surveillance research. In addition, the CDC-led SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology, and Surveillance (SPHERES) consortium has been developed to coordinate among more than 160 institutions active in SARS CoV-2 sequencing.

In stark contrast to the relatively uncoordinated and antiquated approach to SARS-CoV-2 genomic surveillance in America, successful public health sequencing surveillance programs outside the US embraced genomic technology early in the pandemic. China initiated this early by producing the first SARS-CoV-2 sequence [13], whose public widespread dissemination enabled near real-time worldwide sequence comparisons and the unprecedented rapid development of successful vaccines. The UK, ranked ninth in SARS-CoV-2 sequencing as of late January 2021, formed the COVID-19 Genomics UK (COG-UK) Consortium in April 2020 and have since sequenced more than 200,000 viral genomes, representing approximately 6% of reported COVID-19 cases in the UK [2,14]. As of April 2021, the UK climbed to fifth ranking with more than 8% of cases having been sequenced [4]. The COG-UK Consortium, financially supported by the Department of Health and Social Care, UK Research and Innovation, Wellcome, and Wellcome Sanger Institute, and the Consortium, includes several public health groups, universities, and others as their scientific partners [14]. Many of the same groups are core supporters of the UK Biobank, a cohort of 500,000 participants with genome-wide and health data available as a major worldwide research resource for health outcomes of interest, now including COVID-19 [15]. Meanwhile, the US had sequenced less than 0.5% of confirmed cases [2] with plans to ramp up sequencing [6,16], but this effort has coalesced only a year after the first US case of COVID-19 was confirmed in Washington State [17].

Although finances and limited supplies represented key impediments early in the US, this is not the case at present [4]. The CDC has recently committed more than \$200 million to enhance the sequencing. Rather now the major issue may be that the US does not have an organized, ongoing population-based research cohort that can be leveraged for COVID-19 studies, genetic or otherwise, forcing investigators to scramble to form ad hoc consortia for the collection of data from electronic health records [18] or to augment existing public [19] and private [20] genomic collections with COVID-19 data. Data access is siloed and samples are held (or discarded [4]) by a plethora of disconnected labs, both public and private. This balkanization of the public health and testing efforts has not only slowed the process; it has substantially increased expenses. The White House recently announced a \$1 billion dollar influx to increase sequencing capacity [21]. In comparison, the UK has had 2 major influxes of money into SARS-CoV-2 sequencing efforts totaling 20 million pounds in March 2020 producing more than 200,000 sequences [22] and an additional 12 million to produce sequence data from at least 20,000 cases per week. The results are clear; as of March 2021, the UK has generated approximately 40% of the SARS-CoV-2 sequences toward the global surveillance effort [23] for a fraction of the investment expected in the US.

Direct comparisons between non-US successful SARS-CoV-2 sequencing surveillance efforts and the US efforts are difficult and somewhat unfair given that the federal response to the pandemic was initiated under an administration that has since been replaced. Also, the American healthcare system and associated governmental agencies are mostly patchwork and

disparate. The CDC, part of the US Department of Health and Human Services (HHS), typically leads disease surveillance and works in conjunction with other HHS agencies, such as the Indian Health Services, as well as public health agencies organized at the state level. The latter rely primarily on healthcare organizations for data on reportable diseases. Financial and technical resources at the state and local level can vary substantially, explaining in part why Washington State has sequenced 4.84% of their confirmed cases compared with just 0.45% in Ohio. The difference in sequencing observed between Washington State and practically every other US state may also be due to both the history of SARS-CoV-2 in the US and the existing public health genomics research activities [17,24–27].

Given that SARS-CoV-2 is a novel zoonotic disease with no prior human infections, sequencing and analysis inform both the trajectory of the outbreak as well as its evolution [28–30]. The opening of a new niche for the evolution of the virus makes tracking human borne mutations critical to our surveillance and control, as many of these mutations may not have been beneficial to the virus in other hosts and hence would not have survived earlier. This is of particular importance in areas with high incidence. For example, even though as of this writing, the rate of infection is waning in the US, due in part to vaccinations, it is raging in other parts of the world, such as India, with little to no access to vaccines. The current crisis in India and the past year's tragedy in America has created two of among the largest viral populations in the world that can mutate into more transmissible [31] and more severe [32,33] versions of the original virus [34]. The emergence of B.1.1.7, B.1.351 [35], P.1 [36], among others, is a reminder that investments in SARS-CoV-2 genomics need to continue and be expanded as other variants are probably not be far behind given the worldwide variability in vaccination rates and adherence to COVID-19 precautions. Even though the US has set into motion funding and efforts to correct for its initial dearth of sequencing, the pipeline both in the US and globally will require additional and sustained support as the pandemic moves from locale to locale.

Apart from increased capacity for sequencing and analysis, provisions are also sorely needed to link genetic data to clinical and epidemiological data sources for public health research. These critical data linkages remain problematic in the US and resource-limited countries, but they are essential [37,38]. For countries with the adequate resources, increased sequencing capacity and the development of informatics and bioinformatics pipelines and workflows need to be adapted and adopted via international efforts. The pandemic is an evolving phenomenon, requiring worldwide genomic expertise and technology as part of effective SARS-CoV-2 surveillance. When linked to clinical and epidemiological data, the same expertise will help in understanding the factors relevant in variable host susceptibility and response to infection pre- or postvaccination, independent of and interacting with the genetic code of the evolving virus that knows no zip code or international boundaries.

## References

1. Choi BCK. The Past, Present, and Future of Public Health Surveillance. *Scientifica*. 2012; 2012:875253. <https://doi.org/10.6064/2012/875253> PMC3820481 PMID: 24278752
2. Wadman M. United States rushes to fill void in viral sequencing. *Science*. 2021; 371(6530):657–8. <https://doi.org/10.1126/science.371.6530.657> PMID: 33574189
3. Wan W, Guarino B. Why American is 'Flying Blind' to Mutations. *Washington Post* January. 2021; 29.
4. Maxmen A. Why US coronavirus tracking can't keep up with concerning variants. *Nature*. 2021; 592(7854):336–7. <https://doi.org/10.1038/d41586-021-00908-0> PMID: 33828280
5. The COVID Tracking Project: The Atlantic; 2021 [updated March 7, 2021]. Available from: [Covidtracking.com](https://covidtracking.com).
6. Centers for Disease Control and Prevention (CDC). Genomic Surveillance for SARS-CoV-2 Variants. [cited 2021 Feb 21]. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance.html>.

7. Centers for Disease Control and Prevention (CDC). COVID-19 2021 [cited 2021 Mar 15]. Available from: [www.cdc.gov/coronavirus/2019-ncov/index.html](http://www.cdc.gov/coronavirus/2019-ncov/index.html).
8. Hodcroft EB, De Maio N, Lanfear R, MacCannell DR, Minh BQ, Schmidt HA, et al. Want to track pandemic variants faster? Fix the bioinformatics bottleneck. *Nature*. 2021; 591(7848):30–3. <https://doi.org/10.1038/d41586-021-00525-x> PMID: 33649511
9. Kupferschmidt K. Danish scientists see tough times ahead as variant rises. *Science*. 2021; 371(6529):549–50. <https://doi.org/10.1126/science.371.6529.549> PMID: 33542115
10. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016; 70:214–23. <https://doi.org/10.1016/j.jclinepi.2015.09.016> PMID: 26441289
11. Potentially Preventable Deaths from the Five Leading Causes of Death. In: Skinner T, editor. CDC Telet briefing: Centers for Disease Control and Prevention; 2014. Available from: <https://www.cdc.gov/media/releases/2014/t0501-preventable-deaths.html>.
12. Abbasi J. How the US Failed to Prioritize SARS-CoV-2 Variant Surveillance. *JAMA*. 2021; 325(14):1380–2. <https://doi.org/10.1001/jama.2021.3368> PMID: 33760030
13. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020; 579(7798):265–9. <https://doi.org/10.1038/s41586-020-2008-3> PMID: 32015508
14. COVID-19 Genomics UK (COG-UK) Consortium 2020. [cited 2021 Feb 17]. Available from: <https://www.cogconsortium.uk/>.
15. Armstrong J, Rudkin JK, Allen N, Crook DW, Wilson DJ, Wyllie DH, et al. Dynamic linkage of COVID-19 test results between Public Health England's Second Generation Surveillance System and UK Biobank. *Microb Genom*. 2020; 6(7). <https://doi.org/10.1099/mgen.0.000397> PMID: 32553051
16. Callaway E. Multitude of coronavirus variants found in the US—but the threat is unclear. *Nature*. 2021; 591(7849):190. <https://doi.org/10.1038/d41586-021-00564-4> PMID: 33674807
17. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med*. 2020; 382(10):929–36. <https://doi.org/10.1056/NEJMoa2001191> PMC7092802 PMID: 32004427
18. Bennett TD, Moffitt RA, Hajagos JG, Amor B, Anand A, Bissell MM, et al. The National COVID Cohort Collaborative: Clinical Characterization and Early Severity Prediction. *medRxiv*. 2021:2021.01.12.21249511. <https://doi.org/10.1101/2021.01.12.21249511> PMID: 33469592
19. National Institutes of Health (NIH). All of Us Research Program launches COVID-19 research initiatives. Available from: <https://www.nih.gov/news-events/news-releases/all-us-research-program-launches-covid-19-research-initiatives>.
20. Shelton JF, Shastri AJ, Ye C, Weldon CH, Filshtein-Sonmez T, Coker D, et al. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat Genet*. 2021. <https://doi.org/10.1038/s41588-021-00854-7> PMID: 33888907
21. Fact Sheet: Biden Administration Announces \$1.7 Billion Investment to Fight COVID-19 Variants. The White House; April 16, 2021. Available from: <https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/16/fact-sheet-biden-administration-announces-1-7-billion-investment-to-fight-covid-19-variants/>
22. Nelson MI. Tracking the UK SARS-CoV-2 outbreak. *Science*. 2021; 371(6530):680–1. <https://doi.org/10.1126/science.abg2297> PMID: 33574202
23. Kirka D. UK variant hunters lead global race to stay ahead of COVID. *Associated Press* March. 2021; 28.
24. Kim AE, Brandstetter E, Graham C, Heimonen J, Osterbind A, McCulloch DJ, et al. Evaluating Specimen Quality and Results from a Community-wide, Home-Based Respiratory Surveillance Study. *J Clin Microbiol*. Forthcoming [2021]. <https://doi.org/10.1128/JCM.02934-20> PMID: 33563599
25. Fink S, Baker M. It's Just Everywhere Already': How Delays in Testing Set Back the U. S Coronavirus Response. *The New York Times*. March 10, 2020.
26. Chu HY, Englund JA, Starita LM, Famulare M, Brandstetter E, Nickerson DA, et al. Early Detection of Covid-19 through a Citywide Pandemic Surveillance Platform. *N Engl J Med*. 2020; 383:185–7. Epub 2020/05/02. <https://doi.org/10.1056/NEJM2008646> PMID: 32356944; PMC7206929.
27. Seattle Coronavirus Assessment Network (SCAN). The SCAN Dashboard 2021 [cited 2021 Mar 15]. Available from: <https://scanpublichealth.org/dashboard>.
28. Bedford T, Greninger AL, Roychoudhury P, Starita LM, Famulare M, Huang M-L, et al. Cryptic transmission of SARS-CoV-2 in Washington state. *Science*. 2020; 370(6516):571. <https://doi.org/10.1126/science.abc0523> PMID: 32913002

29. Müller NF, Wagner C, Frazar CD, Roychoudhury P, Lee J, Moncla LH, et al. Viral genomes reveal patterns of the SARS-CoV-2 outbreak in Washington State. *Sci Transl Med*. 2021:eabf0202. <https://doi.org/10.1126/scitranslmed.abf0202> PMID: 33941621
30. du Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science*. 2021; 371(6530):708–12. <https://doi.org/10.1126/science.abf2946> PMID: 33419936
31. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*. 2021; 593(7858):266–9. <https://doi.org/10.1038/s41586-021-03470-x> PMID: 33767447
32. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, et al. Genetic mechanisms of critical illness in COVID-19. *Nature*. 2021; 591(7848):92–8. Epub 2020/12/12. <https://doi.org/10.1038/s41586-020-03065-y> PMID: 33307546.
33. Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K, Danon L. Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ*. 2021; 372:n579. <https://doi.org/10.1136/bmj.n579> PMID: 33687922
34. Mascola JR, Graham BS, Fauci AS. SARS-CoV-2 Viral Variants—Tackling a Moving Target. *JAMA*. 2021; 325(13):1261–2. <https://doi.org/10.1001/jama.2021.2088> PMID: 33571363
35. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*. 2021; 592(7854):438–43. <https://doi.org/10.1038/s41586-021-03402-9> PMID: 33690265
36. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DdS, Mishra S, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*. 2021:eabh2644. <https://doi.org/10.1126/science.abh2644> PMID: 33853970
37. Walensky RP, Walke HT, Fauci AS. SARS-CoV-2 Variants of Concern in the United States—Challenges and Opportunities. *JAMA*. 2021; 325(11):1037–8. <https://doi.org/10.1001/jama.2021.2294> PMID: 33595644
38. Tracking COVID-19 Variants Act, House of Representatives (2021). Available from: <https://www.congress.gov/bill/117th-congress/senate-bill/236>.