



University researchers unpack SARS-CoV-2 samples for genomic sequencing.

# WHY US CORONAVIRUS TRACKING CAN'T KEEP UP WITH VARIANTS

Despite its huge virus-sequencing capacity, the country faces funding and coordination roadblocks.

By Amy Maxmen

**A**s COVID-19 cases surge again in the United States, coronavirus variants are on the rise. But researchers fear that the country is ramping up surveillance of the coronavirus SARS-CoV-2 too slowly, allowing these variants – which evidence shows<sup>1,2</sup> could make vaccines less effective – to spread undetected in one of the countries hit hardest by the disease.

Laboratories supported by the US government have doubled the rate at which they are sequencing SARS-CoV-2 genomes over the past two months. Despite this, the number of genomes that the country shared in the online genome repository GISAID in March represented only 1.6% of its positive COVID-19 cases that month. And the United States lags behind at least 30 countries in terms of the sequencing it has done throughout the pandemic, according to GISAID data.

This frustrates researchers because the United States has the equipment and expertise to be doing much more. “We have enough sequencers to sequence SARS-CoV-2 from every case, 100 times over,” says Kristian Andersen, an immunologist at the Scripps Research Institute in La Jolla, California.

A dozen academic researchers at some of the leading virus-sequencing labs in the United States tell *Nature* that a series of problems is holding the country back. Last year, university labs were doing the majority of coronavirus sequencing in the country. They are still responsible for approximately 40% of the sequences on GISAID, with companies and government labs now adding to the effort.

US President Joe Biden listed variant surveillance as a priority for his administration

**“We have enough sequencers to sequence SARS-CoV-2 from every case, 100 times over.”**

immediately after taking office in January, and it was specified as a response measure that will receive some of the US\$19 billion in COVID-19 relief funds announced that month. But researchers say federal money isn't flowing fast enough and systemic problems in sharing samples and data are preventing them from ramping up sequencing.

It's not for lack of trying. The US Centers for Disease Control and Prevention (CDC) is

additionally investing \$200 million to expand surveillance at university, government and company labs, and it has launched initiatives to connect researchers at these labs and track data. The agency is continuing to fund several universities as part of its long-standing Emerging Infections Program, in which academic labs partner with state health departments. But genomic surveillance at this scale has never been done anywhere before – and the fragmented US health system makes coordination a gargantuan task, researchers agree. Hospitals, diagnostic testing labs, local health departments and sequencing centres have rarely worked in unison with one another.

“The biggest challenge is that we don't have a single health system,” says Art Reingold, an epidemiologist at the University of California, Berkeley. “It's a nightmare.”

## Spotty surveillance

To test for SARS-CoV-2, researchers extract RNA from a specimen and search for genetic fragments indicative of the virus. Some university labs that became testing centres during the pandemic sequenced the entire genomes of viruses when time and money allowed, putting them at the forefront of surveillance efforts last year. “Our sequencers have been humming since the start of the pandemic,” says Pavitra Roychoudhury, a computational biologist in the clinical virology department at the University of Washington (UW) in Seattle. But a lack of money for supplies and for paying researchers has frequently prevented Roychoudhury's lab from sequencing its total capacity of about 1,000 genomes per week.

The CDC tried to help with surveillance last spring by launching a programme, called Sequencing for Public Health Emergency Response, Epidemiology and Surveillance (SPHERES), to connect researchers at universities, companies and government labs. Although it was successful at creating a flow of information between these entities, SPHERES didn't dole out funding and therefore didn't significantly increase sequencing rates.

With the CDC now awarding funds to labs nationwide, the country's sequencing will soon accelerate, says Duncan MacCannell, the chief scientific officer at the CDC's Office of Advanced Molecular Detection. In March, nearly 29,000 coronavirus sequences from the United States were uploaded onto GISAID, which has the most SARS-CoV-2 sequences in the world, even though it doesn't contain all of the genomes sequenced. But at the current pace of COVID-19 infections in the United States, the country must reach about 23,000 sequences per week to sequence 5% of all cases, a benchmark considered sufficient to detect emerging variants. (This figure comes from a modelling study<sup>3</sup> published on the medRxiv preprint server without peer review and funded by the biotechnology company

Illumina, based in San Diego, California.)

But researchers at several labs say a lack of samples is as big of a problem as a lack of funding. “We could easily run 1,500 samples each week, but we’re running about 380,” says Lea Starita, a genomicist at the UW Northwest Genomics Center in Seattle. “Someone needs to be willing to fork samples over.”

The problem is that most COVID-19 tests are conducted in diagnostic labs at companies that don’t regularly do genomic sequencing. These labs frequently discard samples after testing, because saving them requires extra labour and storage.

But if a health department wants a deeper investigation into an individual case, officials might ask researchers at a nearby university to sequence the sample. “So we have to scramble to go back to the [testing] lab, and say, ‘Do you still have the specimen for Mr Jones? Save it! Save it!’ And that’s a huge challenge,” explains William Schaffner, an infectious-disease specialist at Vanderbilt University in Nashville, Tennessee, who works with the Tennessee Department of Health through the CDC’s Emerging Infections Program.

The CDC is all too aware of the problems. “We have a very distributed testing system, and private testing labs that aren’t incentivized to hang onto samples,” says MacCannell. He and his colleagues are helping diagnostic labs to either ramp up their own sequencing or connect with labs that can. The agency has also provided guidance on how public-health labs can partner with academic institutions for coronavirus surveillance. “One of our long-standing goals,” MacCannell says, “is to figure out better ways to engage with academics throughout the public-health system.”

### Data flow

Certain university labs, such as the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, don’t have a problem getting samples, because they’ve served as major testing centres throughout the pandemic. Health departments and hospitals in their states were already shipping specimens to these labs. But every researcher interviewed by *Nature* – including MacCannell at the CDC – complained about a lack of information connected to samples.

Such data are needed to uncover where variants are spreading, which variants make the coronavirus more contagious and whether variants help the coronavirus to evade vaccines or natural immunity from a prior infection. This information is scattered like crumbs along the path that a sample travels, but hospitals, health departments and labs are often reluctant to release data because of privacy or proprietary reasons. Stacia Wyman, a computational genomicist at the Innovative Genomics Institute at the University of California, Berkeley, says, “It’s tough to know what’s allowed, and public-health departments don’t

have a huge bandwidth for this.”

MacCannell says siloed data have been a problem for the CDC for many years. “Historically, disease surveillance has been very difficult because many states are uncomfortable with details being provided in public databases.” But he hopes that the need to keep tabs on coronavirus variants will help researchers at unconnected institutions to find ways to share information that could save lives.

In that vein, a platform to share de-identified data on individual COVID-19 cases launched last month. And a philanthropic organization funding the platform, the Rockefeller Foundation, has announced plans to build an even larger version, with the goal of including data

from genomic sequencing and analyses presented in ways to help inform policies.

However, MacCannell and other researchers argue that a government agency, such as the CDC, is best positioned to cut through the red tape that prevents samples from moving to sequencing labs or data from flowing. “I’m convinced that we can do this, and that we can be nimble,” says MacCannell. “But, you know, it is challenging in a pandemic.”

1. McCallum, M. et al. Preprint at bioRxiv <https://doi.org/10.1101/2021.03.31.437925> (2021).
2. Garcia-Beltran, W. F. et al. *Cell* <https://doi.org/10.1016/j.cell.2021.03.013> (2021).
3. Vavrek, D. et al. Preprint at medRxiv <https://doi.org/10.1101/2021.01.12.21249613> (2021).

## WHAT’S NEXT IN THE SEARCH FOR COVID’S ORIGINS

World Health Organization report makes a reasonable start, scientists say, but many questions remain.

By Smriti Mallapaty

**R**esearchers say that a World Health Organization (WHO) report on the pandemic’s origins offers an in-depth summary of available data, including previously unseen granular details. But much remains to be done to establish the provenance of the virus – something that will help to prevent future pandemics.

The report was the result of a joint investigation between Chinese and international researchers that included a four-week trip earlier this year to Wuhan, China, where COVID-19 was first detected.

The details in the report were helpful, but didn’t include much new information, says David Robertson, a virologist at the University of Glasgow, UK. “The extensive data presented confirmed a lot of what was already known, particularly on the timing of events and early cases in Wuhan.”

*Nature* spoke to scientists about what needs to come next.

### What’s new in the report?

The report described the results of many lines of investigation, including when the virus SARS-CoV-2 was first detected in people and which animals might have harboured it. It places the start of the outbreak in the months before mid-December 2019, suggesting that the virus was spreading undetected. It was perhaps introduced to the community through an

unknown animal intermediary between bats, which carried an ancestral virus, and people.

The team didn’t find that intermediary species, even though researchers in China tested tens of thousands of samples from wildlife and livestock, but team members point to wild-animal markets for future leads.

### What do scientists want to know?

Many key questions remain, including what the intermediate animal carrier was, and where and when the spillover occurred. The report includes recommendations for further studies, including following the trail of farmers and suppliers trading in animals and animal products at markets across Wuhan.

The priority should be to “follow the animals”, starting at the Huanan market, where many early cases were identified, says Eddie Holmes, a virologist at the University of Sydney in Australia. Given the large number of species that SARS-CoV-2 can infect, sampling should be expansive, say researchers.

And it should definitely include bats. The most similar virus known to SARS-CoV-2 is a coronavirus called RaTG13, isolated from a bat in a mine in Mojiang, southern China. But RaTG13 shares just 96% of its genome with SARS-CoV-2, meaning that it is only distantly related. Virginie Courtier, an evolutionary geneticist at the Jacques Monod Institute in Paris, says that more bats should be sampled from the mine, and researchers should share the sequences of coronaviruses isolated there.