

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

Researchers 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.



Food and surfaces are sampled for traces of the virus at a wet market in China.

But Linfa Wang, a virologist at Duke–National University of Singapore Medical School, doubts whether closer relatives will be found, given that the cave has already been exhaustively sampled. “If you gave me a billion dollars, I would not sample in Mojiang cave. I would sample in southeast Asia,” adds Wang, who says that testing should extend to lesser-sampled regions such as Cambodia, where relatives of SARS-CoV-2 have been isolated.

To pin down when the virus first occurred in people, the “obvious low-hanging fruit” is to test archived blood samples at the Wuhan Blood Center, says Wang. The centre receives 200,000 donations a year and stores them for two years. Other researchers say that more details about the first known cases are needed, including genome sequences to reveal their diversity and what the early virus looked like.

What’s next in the origins search?

With many unanswered questions, a lot still needs to be done. The visit to Wuhan was the first phase in a WHO process that started in May 2020; it will be followed by longer-term studies, according to an agreement between China and the WHO.

Dominic Dwyer, a virologist at New South Wales Health Pathology in Sydney and a WHO team member, says that some of the work has already started, including a re-analysis of surveillance data on influenza-like illnesses in China before January 2020. But other work, including testing donated blood, will take longer to get under way. Other projects are identifying wildlife farms that supply markets in Wuhan and assessing how long the virus can persist in frozen foods that might also have been a source of transmission.

WHO director-general Tedros Adhanom Gebreyesus said in a statement that the agency

is committed to continuing the search. “It is clear that we need more research across a range of areas, which will entail further field visits.”

The focus of research into COVID-19’s origins should be to mitigate future animal–human spillover of viruses with pandemic potential, says David Heymann, an infectious disease epidemiologist at the London School of Hygiene & Tropical Medicine. “We need to change the paradigm from rapid detection and response, to prevention at the source,” he adds.

“There won’t be an undisputable origins answer. All we’ll have are likelihoods and probabilities.”

Groups not involved with the WHO study have already ramped up sampling and sequencing of archived samples from bats and possible intermediate animals across Asia, to look for potential ancestral viruses.

What’s next for the lab-leak theory?

The report concludes that the chances of the virus having originated in a lab accident are slim. But there is growing pressure, including some from researchers, for a more comprehensive inquiry into this possible route.

WHO team members did not have the required background to investigate a biosafety breach, says Filippa Lentzos, a biosecurity researcher at King’s College London.

Tedros agreed that further investigation was needed, and that he is willing to send additional missions. He said that the team’s assessment, based on lab visits and interviews with researchers, was not extensive enough, adding that “all hypotheses remain on the table”.

A key argument against the lab-leak hypothesis is that SARS-CoV-2 was unknown before the pandemic, with no trail in public databases and research articles. But some researchers say that it is not possible to know exactly which viruses had been sampled around the world.

Wang says it’s highly unlikely that a lab would keep such information secret, but he adds that there’s a small chance that someone inadvertently got infected by an unknown virus while collecting bat samples in a cave, and that this infection seeded the pandemic.

Immunologist Nikolai Petrovsky at Flinders University in Adelaide, Australia, says that given the lack of evidence, the team “would have been best to have been silent on the question because, scientifically, we simply don’t know”.

What do scientists think of the WHO’s handling of the mission?

The question of the pandemic’s origins has been politically fraught from the start. Many researchers say that the team did an excellent job of synthesizing the available evidence under difficult circumstances.

Given the constraints the investigators were under, the report is a “useful and thorough appraisal of what we know and what we need to do”, says Holmes. Constraints included the brief time frame, restrictions on access to material and the narrow scope of the work, he says.

But others are disappointed with how politics seems to have overshadowed the search for answers. The “WHO have handled this as a diplomatic mission and not as an independent scientific investigation into all possible sources”, says Petrovsky.

Following the report’s release, the United States and more than a dozen other countries issued a statement raising concerns about delays and the team’s limited access to raw data. However, others point out that for China to open up and allow the team in to conduct such an investigation was in itself unusual. “If you consider other highly industrialized countries, I am not sure they would,” says Heymann.

Will scientists ever find the origin?

Given the politics and the many unanswered questions, some scientists have wondered whether the origins of the pandemic might forever remain elusive. But those familiar with tracing origins say that it takes time, and luck.

It took years to understand the sources of many human viruses, says Robertson. “Viruses are tricky, as rare events can have massive implications.” However, with sufficient sampling, researchers should be able to identify where, and in which animals, the ancestors of SARS-CoV-2 were circulating, he says.

But Lentzos argues that the origins might forever be shrouded in uncertainty: “I seriously doubt we’ll find a smoking gun. There won’t be an undisputable origins answer. All we’ll have are likelihoods and probabilities.”