



The Wuhan Institute of Virology carries out research on coronaviruses because these pathogens are endemic to the region where it's located.

THE COVID LAB-LEAK HYPOTHESIS: WHAT SCIENTISTS DO AND DON'T KNOW

Nature examines arguments that the coronavirus SARS-CoV-2 escaped from a lab in China, and the science behind them.

By Amy Maxmen & Smriti Mallapaty

Debate over the idea that the coronavirus SARS-CoV-2 emerged from a laboratory has escalated over the past few weeks, coinciding with the annual World Health Assembly, at which the World Health Organization (WHO) and officials from nearly 200 countries discussed the COVID-19 pandemic. After last year's assembly, the WHO agreed to sponsor the first phase of an investigation into the pandemic's origins; this probe took place in China in early 2021.

Most scientists say SARS-CoV-2 probably has a natural origin, and was transmitted from an animal to humans. However, a lab leak has not been ruled out, and many are calling for a deeper investigation into the hypothesis that the virus emerged from the Wuhan Institute

of Virology (WIV), located in the Chinese city where the first COVID-19 cases were reported. On 26 May, US President Joe Biden tasked the US Intelligence Community to join efforts to find SARS-CoV-2's origins, whatever they might be, and report back in 90 days.

Australia, the European Union and Japan have also called for a robust investigation into SARS-CoV-2's origins. The WHO has yet to reveal the next phase of its investigation. But China has asked that the probe examine other countries. Such reticence, and the fact that China has withheld information in the past, has fuelled suspicions about a 'lab leak'. For instance, Chinese government officials suppressed crucial public-health data at the start of the COVID-19 pandemic, and during the 2002–04 severe acute respiratory syndrome (SARS) epidemic, according to high-level reports^{1,2}.

Nature looks at the key arguments that support a lab leak, and the extent to which research has answers.

There's not yet any substantial evidence for a lab leak. Why are scientists still considering it?

Scientists don't have enough evidence about the origins of SARS-CoV-2 to rule out the lab-leak hypothesis, or to prove the alternative – that the virus has a natural origin. Many infectious-disease researchers agree that the most probable scenario is that the virus evolved naturally and spread from a bat either directly to a person or through an intermediate animal. Most emerging infectious diseases begin with a spillover from nature, as was seen with HIV, influenza epidemics, Ebola outbreaks and the coronaviruses that caused the SARS

News in focus

epidemic beginning in 2002 and the Middle East respiratory syndrome (MERS) outbreak beginning in 2012.

Researchers have some leads that support a natural origin. Bats are known carriers of coronaviruses, and scientists have determined³ that the genome of SARS-CoV-2 is most similar to that of RATG13, a coronavirus that was first found in a horseshoe bat (*Rhinolophus affinis*) in the southern Chinese province of Yunnan in 2013. But RATG13's genome is only 96% identical to SARS-CoV-2's, suggesting that a closer relative of the virus – the one passed to humans – remains unknown.

Still, the possibility remains that SARS-CoV-2 escaped from a lab. Although lab leaks have never caused an epidemic, they have resulted in small outbreaks involving well-documented viruses. A relevant example happened in 2004, when two researchers were independently infected by the virus that causes SARS at a virology lab in Beijing that studied the disease. They spread the infection to seven others before the outbreak was contained.

What are the key arguments for a lab leak?

In theory, COVID-19 could have come from a lab in a few ways. Researchers might have collected SARS-CoV-2 from an animal and maintained it in their lab to study, or they might have created it by engineering coronavirus genomes. In these scenarios, a person in the lab might have then been accidentally or deliberately infected with the virus, and then spread it to others – sparking the pandemic. There is currently no clear evidence to back these scenarios, but they aren't impossible.

People have made a number of arguments for a lab origin for SARS-CoV-2 that are

currently conjecture.

One holds that it's suspicious that, almost a year and a half into the pandemic, SARS-CoV-2's closest relative still hasn't been found in an animal. Another suggests it is no coincidence that COVID-19 was first detected in Wuhan, the site of a top lab studying coronaviruses – the WIV.

Some lab-leak proponents contend that the virus contains unusual features and genetic sequences signalling that it was engineered by humans. And some say that SARS-CoV-2 spreads among people so readily that it must have been created with that intention. Another argument suggests that SARS-CoV-2 might be derived from coronaviruses found in an unused mine where WIV researchers collected samples from bats between 2012 and 2015.

So what do infectious-disease researchers and evolutionary biologists say about these arguments?

Is it suspicious that no animal has been identified as transmitting the virus to humans?

Outbreak-origin investigations often take years, and some culprits remain unknown. It took 14 years to nail down the origin of the SARS epidemic, which began with a virus in bats that spread to humans, most likely through civets. To date, a complete Ebola virus has never been isolated from an animal in the region where the world's largest outbreak occurred between 2013 and 2016.

Origin investigations are complicated because outbreaks among animals that aren't the main hosts of a particular virus, such as civets in the case of SARS, are often sporadic. Researchers must find the right animal before it dies or clears the infection. And, even if the

animal tests positive, viruses found in saliva, faeces or blood are often degraded, making it difficult to sequence the pathogen's whole genome.

Scientists have made some progress since the pandemic began, however. A report posted to the preprint server bioRxiv on 27 May suggests that RmYN02, a coronavirus in bats in southern China, might be more closely related to SARS-CoV-2 than RATG13 is⁴.

As for finding an intermediate host animal, researchers in China have tested more than 80,000 wild and domesticated animals; none has been positive for SARS-CoV-2. But this number is a tiny fraction of the animals in the country. To narrow the search down, researchers say, more strategic testing is needed to isolate, for example, animals that come in close contact with people.

Is it suspicious that the WIV is in Wuhan?

Virology labs tend to specialize in the viruses around them, says Vincent Munster, a virologist at the Rocky Mountain Laboratories, a division of the US National Institutes of Health, in Hamilton, Montana. The WIV specializes in coronaviruses because many have been found in and around China. Munster names other labs that focus on endemic viral diseases: haemorrhagic fever labs in Africa and dengue-fever labs in Latin America, for example. "Nine out of ten times, when there's a new outbreak, you'll find a lab that will be working on these kinds of viruses nearby," says Munster.

Researchers note that a coronavirus outbreak in Wuhan isn't surprising, because it's a city of 11 million people in a broader region where coronaviruses have been found. It contains an airport, train stations and markets selling goods and wildlife transported there from around the region⁵ – meaning a virus could enter the city and spread rapidly.

Does the virus have features that suggest it was created in a lab?

Several researchers have looked into whether features of SARS-CoV-2 signal that it was bio-engineered. One of the first teams to do so, led by Kristian Andersen, a virologist at Scripps Research in La Jolla, California, determined that this was "improbable" for a few reasons, including a lack of signatures of genetic manipulation⁶. Since then, others have asked whether the virus's furin cleavage site is evidence of engineering, because SARS-CoV-2 has these sites but its closest relatives don't. The furin cleavage site is important because it's in the virus's spike protein, and cleavage of the protein at that site is necessary for the virus to infect cells.

But there are furin cleavage sites in many other coronaviruses, such as those that cause colds⁷. Viruses containing the site are scattered across the coronavirus family tree, rather than



Scientists found SARS-CoV-2's closest known relative, RATG13, in a horseshoe bat.



Researchers work inside a lab at the WIV.

confined to a group of closely related viruses. Stephen Goldstein, a virologist at the University of Utah in Salt Lake City, says the site probably evolved multiple times because it provides an evolutionary advantage.

Another feature of SARS-CoV-2 that has drawn attention is a combination of nucleotides that underlie a segment of the furin cleavage site: CGG (these encode the amino acid arginine). A *Medium* article that speculates on a lab origin for SARS-CoV-2 (see go.nature.com/3xko) quotes David Baltimore, a Nobel laureate and professor emeritus of biology at the California Institute of Technology in Pasadena, as saying that viruses don't usually have that particular code for arginine, but humans often do – a “smoking gun”, hinting that researchers might have tampered with SARS-CoV-2's genome.

Andersen says that Baltimore was incorrect about that detail, however. In SARS-CoV-2, about 3% of the nucleotides encoding arginine are CGG, he says. And he points out that around 5% of those encoding arginine in the virus that caused the original SARS epidemic are CGG, too. In an e-mail to *Nature*, Baltimore says Andersen could be correct that evolution produced SARS-CoV-2, but adds that “there are other possibilities and they need careful consideration, which is all I meant to be saying”.

Is it true that SARS-CoV-2 must have been engineered, because it's perfect for causing a pandemic?

Many scientists say no. Just because the virus spreads among humans doesn't mean it was designed to do so. It also flourishes among mink and infects a host of carnivorous mammals. And it wasn't optimally transmissible among humans for the better part of last year. Rather, new, more efficient variants

have evolved around the world. To name one example, the highly transmissible variant of SARS-CoV-2 first reported in India (B.1.617.2, or Delta) has mutations in the nucleotides encoding its furin cleavage site that seem to make the virus better at infecting cells⁸.

Did researchers collect SARS-CoV-2 from a mine?

Researchers from the WIV collected hundreds of samples from bats roosting in a mine between 2012 and 2015, after several miners working there had fallen ill with an unknown respiratory disease. Back at the lab, WIV researchers detected nearly 300 coronaviruses in the bat samples⁹, but they were able

“We want an answer. But we may have to keep piecing bits of evidence together.”

to get whole or partial genomic sequences from fewer than a dozen, and none of those that were reported was SARS-CoV-2. During the WHO-led origins probe earlier this year, WIV researchers told investigators that they cultured only three coronaviruses at the lab, and none was closely related to SARS-CoV-2.

Although the investigators didn't sift through freezers at the WIV to confirm this information, the low number of genomes and cultures doesn't surprise virologists. Munster says it's exceedingly difficult to extract intact coronaviruses from bat samples. Virus levels tend to be low in the animals, and viruses are often degraded in faeces, saliva and droplets of blood. Additionally, when researchers want to study or genetically alter viruses, they need

to keep them (or synthetic mimics of them) alive, by finding the appropriate live animal cells for the viruses to inhabit in the lab, which can be a challenge.

So, for SARS-CoV-2 to have come from this mine in China, WIV researchers would have had to overcome some serious technical challenges – and they would have kept the information secret and misled investigators on the WHO-led mission, scientists point out. There's no evidence of this, but it can't be ruled out.

What's next for lab-leak investigations?

Biden asked the US Intelligence Community to report back to him in 90 days. Perhaps this investigation will shed light on undisclosed US intel reported by *The Wall Street Journal* (see go.nature.com/3zsk) suggesting that three staff members at the WIV were sick in November 2019, before the first cases of COVID-19 were reported in China. The article claims that US officials have different opinions on the quality of that intel. And researchers at the WIV have maintained that staff at the institute tested negative for antibodies that would indicate SARS-CoV-2 infection before January 2020.

This month, Anthony Fauci, Biden's chief medical adviser, asked Chinese officials to release the hospital records of WIV staff members. Others have asked for blood samples from WIV staff members, and access to WIV virus samples, lab notebooks and hard drives. But it's unclear what such requests will yield, because China has not conceded to demands for a full lab investigation. A spokesperson for the Ministry of Foreign Affairs of the People's Republic of China, Zhao Lijian, said that US labs should instead be investigated, and that some people in the United States “don't care about facts or truth and have zero interest in a serious science-based study of origins”.

As Biden's investigation commences and the WHO considers the next phase in its origin study, pandemic experts are bracing themselves for a long road ahead. “We want an answer,” says Jason Kindrachuk, a virologist at the University of Manitoba in Winnipeg, Canada. “But we may have to keep piecing bits of evidence together as weeks and months and years move forward.”

1. The Independent Panel for Pandemic Preparedness & Response. ‘COVID-19: Make it the last pandemic’ (Independent Panel, 2021).
2. Huang, Y. in *Learning from SARS: Preparing for the Next Disease Outbreak: Workshop Summary* (eds Knobler S. et al.) 116–136 (National Academies Press, 2004).
3. Zhou, P. et al. *Nature* **579**, 270–273 (2020).
4. Lytras, S. et al. Preprint at bioRxiv <https://doi.org/10.1101/2021.01.22.427830> (2021).
5. Xiao, X. et al. *Sci. Rep.* **11**, 11898 (2021).
6. Andersen, K. G. et al. *Nature Med.* **26**, 450–452 (2020).
7. Wu, Y. & Zhao, S. *Stem Cell Res.* **50**, 102115 (2020).
8. Peacock, T. P. et al. Preprint at bioRxiv <https://doi.org/10.1101/2021.05.28.446163> (2021).
9. Guo, H. et al. Preprint at bioRxiv <https://doi.org/10.1101/2021.05.21.445091> (2021).