

To stop the next pandemic, we need to unravel the origins of COVID-19

David A. Relman^{a,b,c,d,1} 

We find ourselves ten months into one of the most catastrophic global health events of our lifetime and, disturbingly, we still do not know how it began. What's even more troubling is that despite the critical importance of this question, efforts to investigate the origins of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus and of the associated disease, coronavirus disease 2019 (COVID-19), have become mired in politics, poorly supported assumptions and assertions, and incomplete information.

SARS-CoV-2 is a betacoronavirus whose apparent closest relatives, RaTG13 and RmYN02, are reported to have been collected from bats in 2013 and 2019,

respectively, in Yunnan Province, China (1). COVID-19 was first reported in December 2019 more than 1,000 miles away in Wuhan City, Hubei Province, China. Beyond these facts, the "origin story" is missing many key details, including a plausible and suitably detailed recent evolutionary history of the virus, the identity and provenance of its most recent ancestors, and surprisingly, the place, time, and mechanism of transmission of the first human infection. Even though a definitive answer may not be forthcoming, and even though an objective analysis requires addressing some uncomfortable possibilities, it is crucial that we



To avoid or mitigate the dire consequences of this and future pandemics (here, people in PPE bury a victim in Delhi, India in June), unraveling the origins of SARS-CoV-2 and COVID-19 will be essential—even though a definitive answer may be elusive, and an objective analysis means broaching some uncomfortable possibilities. Image credit: Shutterstock/PradeepGaur.

^aDepartment of Medicine, Stanford University School of Medicine, Stanford, CA 94305; ^bDepartment of Microbiology & Immunology, Stanford University School of Medicine, Stanford, CA 94305; ^cCenter for International Security and Cooperation, Stanford University, Stanford, CA 94305; and ^dInfectious Diseases Section, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA 94304

The author declares no competing interest.

Published under the [PNAS license](#).

Any opinions, findings, conclusions, or recommendations expressed in this work are those of the author and have not been endorsed by the National Academy of Sciences.

¹Email: relman@stanford.edu.

pursue this question. Preventing the next pandemic depends on understanding the origins of this one.

There are several potential origin scenarios. First, SARS-CoV-2 may have evolved in bats, which are known reservoirs of immense coronavirus diversity (2), and then spread directly, or indirectly via an intermediate host, to humans through natural mechanisms. The degree of anticipated but undiscovered natural diversity clearly lends support to this scenario, as well as support to other scenarios. Second, SARS-CoV-2 or a recent ancestor virus may have been collected by humans from a bat or other animal and then brought to a laboratory where it was stored knowingly or unknowingly, propagated and perhaps manipulated genetically to understand its biological properties, and then released accidentally.

A deliberative process for investigating the origins of this pandemic must be representative of all relevant disciplines, expertise, and stakeholders; must achieve political neutrality, scientific balance, and access to all relevant information and samples; and must operate with transparency and independent oversight. Without these features, it will not be credible, trustworthy, or effective.

Some have argued that a deliberate engineering scenario is unlikely because one would not have had the insight *a priori* to design the current pandemic virus (3). This argument fails to acknowledge the possibility that two or more as yet undisclosed ancestors (i.e., more proximal ancestors than RaTG13 and RmYN02) had already been discovered and were being studied in a laboratory—for example, one with the SARS-CoV-2 backbone and spike protein receptor-binding domain, and the other with the SARS-CoV-2 polybasic furin cleavage site. It would have been a logical next step to wonder about the properties of a recombinant virus and then create it in the laboratory. Alternatively, the complete SARS-CoV-2 sequence could have been recovered from a bat sample and viable virus resurrected from a synthetic genome to study it, before that virus accidentally escaped from the laboratory. The third scenario, seemingly much less likely, involves laboratory manipulation or release, with the clear intention of causing harm.

Even though strong opinions abound, none of these scenarios can be confidently ruled in or ruled out with currently available facts. Just because there are no public reports of more immediate, proximal ancestors in natural hosts, doesn't mean that these ancestors don't exist in natural hosts or that COVID-19 didn't begin as a spillover event. Nor does it mean that they have not been recovered and studied, or deliberately recombined in a laboratory.

Why do these distinctions matter? If we find more concrete evidence of a "spill-over" event with SARS-CoV-2 passing directly from bat to human, then efforts

to understand and manage the bat–human interface need to be significantly strengthened. But if SARS-CoV-2 escaped from a lab to cause the pandemic, it will become critical to understand the chain of events and prevent this from happening again. Rather than resorting to hunches or finger-pointing, each scenario must be systematically and objectively analyzed using the best available science-based approaches. There is a path to greater clarity. It requires scientific rigor, forensic approaches, deliberate methods, transparency, and cooperation.

In an effort to reveal the origins of the pandemic, researchers so far have focused on the SARS-CoV-2 genome sequence. However, the sequence of the pandemic virus tells us only so much. First, the closest known relatives, RaTG13 and RmYN02, are not that close (4). Second, there is probably more than one recent ancestral lineage that contributes to SARS-CoV-2 because its genome shows evidence of recombination between different parental viruses. In nature, recombination is common among coronaviruses. But it's also common in some research laboratories where recombinant engineering is used to study those viruses. The bottom line is simple: We need to identify the immediate parent(s) of SARS-CoV-2, and they're missing.

To find its parents and understand its recent history, we need 1) additional genome sequences of coronaviruses from relevant bats and other suspect hosts—some of these likely exist already in laboratories, given the efforts so far undertaken to survey bats in particular (2, 5); 2) measurements of SARS-CoV-2 evolution under a variety of defined conditions so that differences between viral genomes can be understood better as differences in time on an evolutionary clock; and 3) data from antibody surveys of humans at high risk of coronavirus exposure and from past cases of similar disease, so that previously unrecognized encounters can be revealed. In addition, we need to address whether there is information about host or environmental samples that contain recent ancestors of SARS-CoV-2, data perhaps not yet publicly available. More generally, are there relevant scientific data, including from coronavirus engineering work in laboratories, that have not been shared widely? Who knew what about relevant viruses and cases of disease before December 2019, and when? This information will go a long way toward clarifying the origins of this pandemic, even if certainty continues to elude us.

The means are just as important as the goals. An investigative process should be transparent, collaborative, international, and, to the extent possible, devoid of political interest. Recent, productive scientific collaborations between the United States and China, for example, provide hope that such a process can be achieved. But the kind of effort required will need to expand far beyond what's taken place so far, and nations other than the United States and China will need to be involved. Conflicts of interest by researchers, administrators, and policymakers on all sides must be revealed and addressed, and all relevant global constituencies must be included. Both the World

Health Organization and *The Lancet* COVID-19 Commission (6) have hinted that they have taken some first steps, but their efforts so far have been cloaked in secrecy (7, 8). A deliberative process for investigating the origins of this pandemic must be representative of all relevant disciplines, expertise, and stakeholders; must achieve political neutrality, scientific balance, and access to all relevant information and samples; and must operate with transparency and independent oversight. Without these features, it will not be credible, trustworthy, or effective.

A more complete understanding of the origins of COVID-19 clearly serves the interests of every person in every country on this planet. It will limit further re-criminations and diminish the likelihood of conflict; it will lead to more effective responses to this pandemic, as well as efforts to anticipate and prevent the next one. It will also advance our discussions about risky science. And it will do something else: Delineating COVID-19's origin story will help elucidate the nature of our very precarious coexistence within the biosphere.

-
- 1 Y.-Z. Zhang, E. C. Holmes, A genomic perspective on the origin and emergence of SARS-CoV-2. *Cell* **181**, 223–227 (2020).
 - 2 A. Latinne *et al.*, Origin and cross-species transmission of bat coronaviruses in China. *Nat. Commun.* **11**, 4235 (2020).
 - 3 K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).
 - 4 M. F. Boni *et al.*, Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat. Microbiol.* (2020).
 - 5 Z. Wu *et al.*, Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *ISME J.* **10**, 609–620 (2016).
 - 6 J. D. Sachs *et al.*; The Lancet COVID-19 Commission. *Lancet* **396**, 454–455 (2020).
 - 7 World Health Organization, WHO experts to travel to China. <https://www.who.int/news-room/detail/07-07-2020-who-experts-to-travel-to-china/>. Accessed 20 September 2020.
 - 8 P. Nuki, S. Newey, Scientists to examine possibility Covid leaked from lab as part of investigation into virus origins. *The Telegraph*, 15 September 2020. <https://www.telegraph.co.uk/global-health/science-and-disease/scientists-examine-possibility-covid-leaked-lab-part-investigation/>. Accessed 27 September 2020.