PNAS OPINION

The evidence remains clear: SARS-CoV-2 emerged via the wildlife trade

Robert F. Garry[a,b,c,1]

In the case of both the COVID-19 pandemic and the West African Ebola outbreak, a rush to judgment about virus origins obscured the facts and led researchers and policymakers to make errant claims that wasted time and resources. Image credit: Shutterstock/Sergey Uryadnikov.

During the large Ebola outbreak of 2013–2016, conspiracists used various media platforms to accuse our research group in Sierra Leone of unleashing Ebola virus (EBOV) on the people of West Africa (1). The main argument: EBOV must have leaked from a laboratory run by the Ministry of Health and Sanitation and our group in Kenema, located about 50 miles southwest of the village in Guinea in which EBOV emerged. The virus, so went their rationale, had previously only emerged in Middle African countries, such as the Democratic Republic of the Congo, which is more than 1,500 miles away from West Africa (Fig. 1). Some pundits charged that we were running a bioweapons laboratory funded by the "New World Order" as part of their clandestine efforts to establish a one-world government. Other theorists suggested that we had altered EBOV to make it more infectious—perhaps even airborne.

In reality, we did not have EBOV in our laboratory and therefore could not have released or engineered it. The NIH funded us to develop countermeasures for Lassa virus, a hemorrhagic fever virus unrelated to EBOV (2). The West African Ebola outbreak showed that viruses can move large distances either via human travelers, commerce, or—sometimes—on the wings of bats.

Fast forward to the last days of 2019 and the first reports of a severe pneumonia (coronavirus disease 2019 [COVID-19]) in Wuhan, China. Public health officials quickly determined that the illness was caused by a novel coronavirus, subsequently named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (3). Politicians in the United States, and others eager to assign blame, reacted by pointing out the presence of a laboratory that studies coronaviruses, the Wuhan Institute of Virology (WIV). Once again, the guilt-by-proximity argument came to the fore, serving to shift accountability for the pandemic to a geopolitical rival and distract attention from public health response deficiencies.

Author affiliations: [a]Department of Microbiology and Immunology, Tulane University Medical Center, New Orleans, LA 70112; [b]Zalgen Labs, LLC, Frederick, MD 21703; and [c]Global Viral Network, Baltimore, MD 21201

R.F.G. is co-founder of Zalgen labs, a biotechnology company developing countermeasures to emerging viruses, including SARS-CoV-2.

[1]To whom correspondence may be addressed. Email: rfgarry@tulane.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2214427119/-/DCSupplemental.

Published November 10, 2022.

In both the case of SARS-CoV-2 and EBOV, a rush to judgment obscured the facts and led researchers and policymakers to make errant claims that wasted time and resources. Discerning the origins of COVID-19 is an important mission. It must be a mission that's approached with care—false leads not only unfairly place blame on ethical research and researchers but also fail to advance public health so that we're better prepared for the next pandemic. The following careful telling of events early on in the pandemic elucidates the facts—and the distortions.

Leaks Unlikely

The faulty lab leak narrative leaves out some crucial details. Most lab leak proponents don't mention that most major Chinese cities have one or more active coronavirus laboratories. The Chinese government established these laboratories after multiple spillovers of the first SARS-CoV in 2002 through 2004, which caused approximately 8,000 cases of severe respiratory disease worldwide and at least 744 deaths. Those who suggest that the pandemic is the result of a lab leak often also note that the closest related bat coronaviruses to SARS-CoV-2 have been found only in southern China or in Laos, about 750 miles away from Wuhan (4) (Fig. 1). They argue that the virus could not have traversed such a distance without causing COVID-19 cases along the way. These comments, however, show an ignorance of some crucial points: the West African EBOV precedent, which showed that viruses can emerge or reemerge large distances from the site of their initial spillover, and the fact that SARS-CoV emerged multiple times in Chinese megacities similar distances from where its closest bat progenitors have been found.

On January 10, 2020, Edward Holmes of the University of Sydney, Australia—working on behalf of a consortium led by Yong-Zhen Zhang of Fudan University, Shanghai, China—became the first person to release the genomic sequence of the novel coronavirus (5). That sequence spurred a flurry of activity by virologists reminiscent of when the first genomic sequence of SARS-CoV was published in 2003 (6). My long-time collaborator William Gallaher and I quickly analyzed the amino acid sequence of the SARS-CoV spike protein. We correctly predicted that the coronavirus spike protein fits the general scaffold of HIV-1 and influenza virus glycoprotein structures (Fig. S1). Likewise, the site for cleavage of the SARS-CoV-2 spike protein into its two subunits, S1 and S2, was quickly identified by ourselves and others: It is a furin cleavage site (FCS). Acquisition of a FCS can render a low pathogenicity avian influenza virus increasingly transmissible and highly pathogenic. SARS-CoV accomplishes its high pathogenicity, a 9% case fatality rate, without a FCS at this location. Although the FCS confers increased transmissibility, the level of SARS-CoV-2 pathogenicity—although substantial—does not approach that of SARS-CoV.

In another instance, the newly released SAR-CoV-2 sequence was promptly used to support the possibility of an unnatural origin for SARS-CoV-2. On January 31, 2020, a preprint purported that the SARS-CoV-2 proteins contain unique inserts with an "uncanny similarity" to HIV-1 proteins (7). The clear insinuation was that SARS-CoV-2 had been engineered in a laboratory. The authors of this preprint suggested that some inserts are "related" to the HIV glycoprotein (Gp120) (Fig. 2A). They suggested that another insert from the HIV-1 Group antigen (Gag) created a portion of the FCS. The inserts are short or required insertions

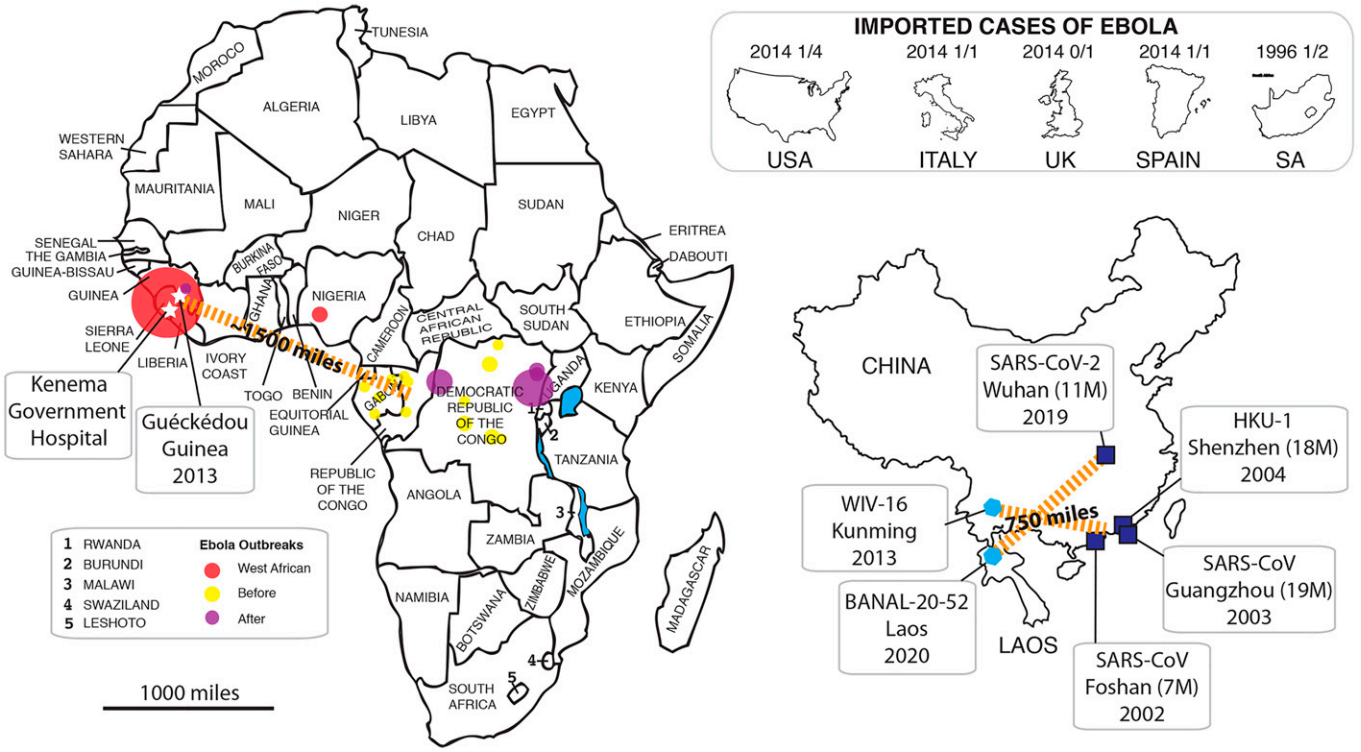

Fig. 1. The data clearly suggest that viruses can travel long distances in relatively short periods of time.

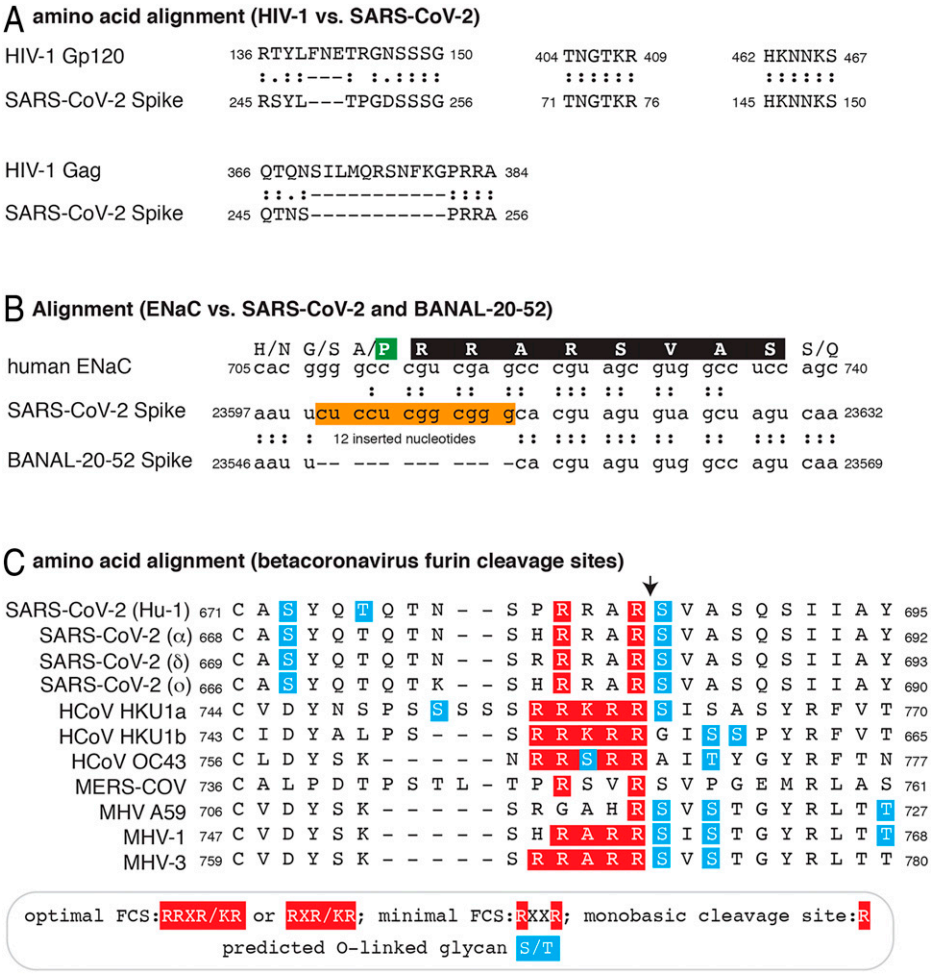

Fig. 2. SARS-CoV-2 Spike Alignments: (*A*) Alignment of the SARS-CoV-2 with HIV-1 proteins. (*B*) Amino acid alignment of human amiloride-sensitive epithelial sodium channel a subunit (ENaC) with Spike of SARS-CoV-2 showing eight common amino acids RRARSVAS. Modified from (20). (*C*) Amino acid alignment of the S1/S2 junction of SARS-CoV-2 Spike with spikes of other betacoronaviruses.

for partial alignment and therefore are likely to be present purely by chance. Although this preprint was retracted based on these fundamental flaws, the media firestorm fortified lab leak conspiracists worldwide and intensified suspicion of virologists at the WIV.

During a February 1, 2020, teleconference, organized by Wellcome Trust director Jeremy Farrar, virologists, evolutionary biologists, and NIH administrators assessed the likelihood that SARS-CoV-2 may have been engineered and leaked, by accident or intentionally, from the WIV or another laboratory. Strict confidentiality was necessary to allow for uninhibited discussion and ensure that isolated comments would not be misconstrued out of context. The conclusion of the teleconference participants was that detailed analyses and more data were needed. Some participants felt that it was important not to dismiss the remote possibility of a lab leak out of hand.

A few of us, however, continued to analyze SARS-CoV-2 features, including the spike's FCS and receptor binding domain (RBD), which on initial inspection were deemed significant to the origins investigation. Our subsequent peer-reviewed study (8) discussed the possibility of purposeful laboratory manipulation as insinuated in the "HIV-1 insert" preprint. However, we determined a natural origin of SARS-CoV-2 was more likely based on a comparative analysis of genomic data. NIH Administrators on the call did not influence or edit our work.

Paucity of Proof

To this day, no scientific data exist to support a lab leak of SARS-CoV-2. Multiple theories regarding the unnatural origins of SARS-CoV-2 have flourished, as was the case with disproven "lab origin" theories of other pandemic viruses, including HIV. All but one lab leak theory is predicated on laboratory manipulation of SARS-CoV-2 or a close (>99% similar) progenitor. Excepting this one, lab leak theories must involve a conspiracy and a cover-up. Lab leak proponents suggest that an undisclosed progenitor of SARS-CoV-2 may have been passaged on human cells or experimental animals or genetically engineered for enhanced virulence. Some theories posit that not only is SARS-CoV-2 an engineered virus but that related coronaviruses of bats and pangolin have also been faked or engineered to cover up the role of WIV in starting the COVID-19 pandemic. A role of the Chinese military at WIV has also been proposed. Lab leak theories are often bolstered by racist tropes that suggest that epidemiological, genetic, or other scientific data have been purposefully withheld or altered to obscure the origin of the virus (9). A small but vocal group of scientists have

taken up the SARS-CoV-2 lab leak cudgel as an opportunity to advance long-standing positions in opposition to virology research that they consider to be risky and criticize the NIH administrators who oversee virology funding (10, 11).

An often-cited SARS-CoV-2 origin theory suggests that a scientist could have been unknowingly infected while doing field work or after unsuccessfully attempting virus culture from a bat sample (4, 10). This hypothesis is said to show that not all lab leak theories involve a conspiracy or cover-up, because no person would be aware of the accidental infection. Although such a proposal may comfort lab leak proponents sensitive to the conspiracy theorist label, it is a very weak argument. In this hybrid scenario, SARS-CoV-2 is a natural virus, not a lab-created one. Moreover, a sample containing enough infectious virus to infect a human would likely also be detectable by nucleic acid sequencing or virus culture. Most animal encounters by coronavirus virologists in the field are with bats, not other wildlife species. Recent data show that SARS-CoV-2 did not come directly from a bat to a human; rather, it first evolved in an intermediate wildlife host (12, 13). It is extremely unlikely that a non-bat intermediate wildlife host transferred SARS-CoV-2 to a scientist directly or through a collected sample. Compared with the millions of worldwide encounters of humans with wildlife that could transmit an infectious agent, the number of high-risk exposures of scientists doing field or laboratory work is miniscule.

Early support for a lab leak has come from assertions that SARS-CoV-2 emerged full-blown as a human-adapted virus with unusual genetic stability (14), properties the authors suggested could be the result of laboratory manipulation. SARS-CoV-2 is clearly not "well adapted" just for humans. Rather, it is capable of effective spread not only in humans but also to a diverse group of mammals, including mink, otters, deer, and various canids and felines (15). The virus has also shown remarkable genetic plasticity with the ability to produce variants with improved FCS and RBD. Recent studies suggest that wild species, such as white-tailed deer, may continue to harbor SARS-CoV-2 variants even after they no longer circulate in humans (16). The animal reservoir of SARS-CoV-2 has not yet been identified, but that doesn't imply the possibility of a lab origin. Indeed, the reservoirs for SARS-CoV and many other emerging pathogens, notably EBOV, have not been identified.

With its high cross-species transmissibility, SARS-CoV-2 would be expected to quickly reach the human population via intermediate hosts. Extensive spread of SARS-CoV-2 in wildlife has not been observed and in hindsight is not expected. In contrast, SARS-CoV had spread widely from its as-yet-undetermined reservoir into intermediate hosts, such as civets, ferret badgers, and raccoon dogs, and genetically diversified before it made multiple jumps to humans. This possibility was not accounted for by the authors of the preprint (14) that suggested that SARS-CoV-2 may have been adapted in a laboratory to infect humans.

Another unsound theory suggesting an engineered SARS-CoV-2 was featured in a recent PNAS opinion piece (17). The authors opined that an eight-amino acid identity (RRARSVAS) of SARS-CoV-2 FCS with one of the FCS of human amiloride-sensitive epithelial sodium channel a subunit (ENaC) was strong evidence that WIV scientists had conspired with coronavirus virologists at the University of North Carolina to produce SARS-CoV-2 (Fig. 2B). One of the authors subsequently proclaimed, in other media reports, that he was convinced that American biotechnology was likely responsible for the COVID-19 pandemic (18, 19). As was the case with the retracted preprint discussed above (7), the short amino acid similarity is quite simply happenstance. Several other coronaviruses share five of the eight amino acids (RSVAS) with ENaC (20) (Fig. 2B). It is also certainly not unusual for a FCS to be present at the junction between the S1 and S2 subunits of a betacoronavirus spike protein (21) (Fig. 2C). The two betacoronaviruses that cause common colds, OC43 and HKU1, have a FCS in that location.

The flawed research, and flawed conclusions, continue apace. On October 20, while this Opinion was in press, Bruttel and colleagues published a preprint claiming that the pattern of restriction enzyme (RE) sites in the SARS-CoV-2 genome indicates that it has a synthetic origin (22). The authors suggest that short nucleotide sequences (sites) recognized by specific type II shifted (IIS) RE were added and removed from a bat coronavirus genome or combination of genomes to facilitate laboratory manipulations that produced SARS-CoV-2. The pattern of type IIS RE sites in the SARS-CoV-2 genome is not unusual and does not prove a laboratory origin. Although not all of these RE sites are present in each genome, all sites are represented in genomes of bat coronaviruses that are relatives of SARS-CoV-2. The genome of RpYN06, a virus isolated from a horseshoe bat (Rhinolophus pusillus) sampled in May 2020 (23) lacks the two RE sites the authors speculate were removed from the SARS-CoV-2 genome. A natural evolutionary pathway involving an RpYN06-like coronavirus accounts for the absence of these sites in SARS-CoV-2.

Wildlife Origins

The FCS of SARS-CoV-2 was not bioengineered—full stop. S1/S2 cleavage sites of coronaviruses are frequently modified by insertions or deletions (indels) during evolution. The SARS-CoV-2 FCS was generated by a 12-nucleotide out-of-frame insertion. There is no rational reason for a scientist to perform an out-of-frame insertion. In nature, however, insertions would be expected to occur without regard to reading frame. A natural12 base insertion was recently detected near the S1/S2 junction of an alphacoronavirus (24), and a (net) 6 base insertion occurred at another location in Omicron BA.1. The SARS-CoV-2 spike protein insertion adds the amino acid proline (P) before the three amino acids (RRA) that create the SARS-CoV-2 FCS. A proline is not present in the ENaC FCS. Its presence is inconsistent with the theory advanced in the recent PNAS Opinion (17) (Fig. 2B). This proline is in fact replaced in several SARS-CoV-2 variants (Alpha, Delta, Omicron), resulting in increased infectivity (Fig. 2C). Furthermore, the SARS-CoV-2 FCS contains a previously undescribed feature, O-linked glycans, that a laboratory researcher could not have known to include.

The data that have accumulated since our first origins study (8) was published provide clear insight into how SARS-CoV-2 emerged via the wildlife trade (Fig. S2). Epidemiological, phylogenetic, and serological evidence indicates

that SARS-CoV-2 did not circulate widely in humans before November 2019 (12). The Huanan Market was one of only four locations that sold live wildlife, but not bats, in Wuhan (25). The earliest known COVID-19 cases from December 2019, including those without reported direct links, lived close to the Huanan market (13, 26). SARS-CoV-2–positive environmental samples were spatially associated with vendors in the southwestern corner of the market, the area selling live mammals. An iron cage (13) and drainage from this area (27) were positive for SARS-CoV-2 nucleic acids. Linkage to the wildlife trade provides simple explanations for the early epidemiology of known cases and phylogenic analyses indicating that two different lineages of SARS-CoV-2 emerged at the Huanan market (12).

In the last two decades, four novel coronaviruses with the ability to cause widespread human infections have emerged via zoonoses. Middle East Respiratory Syndrome coronavirus, which frequently spills over from camels to humans, and SARS-CoV have higher pathogenicity than SARS-CoV-2 and the common cold coronavirus HKU-1 but are not highly transmissible from human-to-human. Because other novel coronaviruses with high pathogenicity or transmissibility could emerge in the future, it is essential that we develop a greater understanding of the diversity of this virus family in wild animals. Such studies will require extensive international cooperation and careful attention to biosecurity and biosafety concerns for handling potential pandemic pathogens in the field.

All this means we need to develop, in advance, effective medical countermeasures to potential pandemic coronaviruses, including diagnostics, vaccines, and therapeutics. Prevention efforts for the next coronavirus must also focus on increased surveillance at the animal–human interface—as well as stringent oversight of the wildlife and fur trade. And all of these approaches must be predicated on sound science and evidence.

ACKNOWLEDGMENTS. Essential input and discussion came from colleagues too numerous to list. Alex Crits-Christoph and Jonathan Pekar provided analysis of SARS-CoV-2 restriction enzyme sites. Work on emerging viruses in the Garry Laboratory is supported by the National Institutes of Health, the Coalition for Epidemic Preparedness Innovations, the Burroughs Wellcome Fund, the Wellcome Trust, the Center for Disease Prevention and Control, and the European & Developing Countries Clinical Trials Partnership.

1. K. Sanders, Bloggers: Bill and Melinda Gates, George Soros funded Sierra Leone lab that started Ebola outbreak (2014). https://www.politifact.com/factchecks/2014/aug/11/blog-posting/bloggers-bill-and-melinda-gates-george-soros-funde/. Accessed 11 August 2022.
2. A. Goba *et al.*; Viral Hemorrhagic Fever Consortium, An outbreak of Ebola virus disease in the Lassa Fever Zone. *J. Infect. Dis.* **214** (suppl. 3), S110–S121 (2016).
3. N. Zhu *et al.*; China Novel Coronavirus Investigating and Research Team, A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
4. A. Chan, M. Ridley, *Viral: The Search for the Origin of COVID-19* (News Corp: Harper-Collins, New York, 2021).
5. E. Holmes, Novel 2019 coronavirus genome (2020). https://virological.org/t/novel-2019-coronavirus-genome/319. Accessed 11 August 2022.
6. P. A. Rota *et al.*, Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394–1399 (2003).
7. P. Pradhan *et al.*, (2020) Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag. *bioRxiv*. https://doi.org/10.1101/2020.01.30.927871 (2 February 2020).
8. 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).
9. J. Cho, (2021) Lab-Leak, gain-of-function and the media myths swirling around the Wuhan Institute of Virology. https://www.mintpressnews.com/lab-leak-gain-function-media-myths-swirling-around-wuhan-institute/278555/. Accessed 11 August 2022.
10. Relman DA (2020) Opinion: To stop the next pandemic, we need to unravel the origins of COVID-19. *Proc. Natl. Acad. Sci. U S A* **117**, 29246–29248 (2020).
11. R. Ebright, (2022) For the hearing Revisiting Gain of Function Research. https://www.hsgac.senate.gov/imo/media/doc/Ebright. Accessed 11 August 2022.
12. J. E. Pekar *et al.*, The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. *Science* **377**, 960–966 (2022).
13. M. Worobey *et al.*, The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science* **377**, 951–959 (2022).
14. S. Zhan, B. Deverman, Y. Chan, (2020) SARS-CoV-2 is well adapted for humans. What does this mean for re-emergence? doi.org/10.1101/2020.05.01.073262 (2 May 2020).
15. M. Koopmans, SARS-CoV-2 and the human-animal interface: outbreaks on mink farms. *Lancet Infect. Dis.* **21**, 18–19 (2021).
16. L. C. Caserta *et al.*, (2022) White-tailed deer (Odocoileus virginianus) may serve as a wildlife reservoir for nearly extinct SARS-CoV-2 variants of concern. https://www.biorxiv.org/content/10.1101/2022.09.02.506368v1 (2 September 2020).
17. N. L. Harrison, J. D. Sachs, A call for an independent inquiry into the origin of the SARS-CoV-2 virus. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2202769119 (2022).
18. C. Boyd, (2022) Covid leaked from an AMERICAN lab and not the notorious Chinese facility at centre of pandemic cover-up, claims top US professor described as 'Xi propagandist.' https://www.dailymail.co.uk/health/article-10980715/Covid-leaked-AMERICAN-lab-claims-professor-Jeffrey-Sachs.html. Accessed 11 August 2022.
19. R. F. Kennedy, Jr, J. D. Sachs, (2022) Origins of the Virus with Jeffrey Sachs https://pca.st/f4qhwghp. Accessed 11 August 2022.
20. R. F. Garry, SARS-CoV-2 furin cleavage site was not engineered. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2211107119 (2022).
21. Y. Wu, S. Zhao, Furin cleavage sites naturally occur in coronaviruses. *Stem Cell Res. (Amst.)* **50**, 102115 (2020).
22. V. Bruttel, A. Washburne, A. VanDongen, Endonuclease fingerprint indicates a synthetic origin of SARS-CoV-2. https://doi.org/10.1101/2022.10.18.512756 (20 October 2022).
23. H. Zhou *et al.*, Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell* **184**, 4380–4391.e14 (2021).
24. S. K. P. Lau *et al.*, Fatal pneumonia associated with a novel genotype of human coronavirus OC43. *Front. Microbiol.* **12**, 795449 (2022).
25. X. Xiao, C. Newman, C. D. Buesching, D. W. Macdonald, Z. M. Zhou, Animal sales from Wuhan wet markets immediately prior to the COVID-19 pandemic. *Sci. Rep.* **11**, 11898 (2021).
26. M. Worobey, Dissecting the early COVID-19 cases in Wuhan. *Science* **374**, 1202–1204 (2021).
27. G. Gao *et al.*, (2022) Surveillance of SARS-CoV-2 in the environment and animal samples of the Huanan Seafood Market. https://doi.org/10.21203/rs.3.rs-1370392/v1 (25 February 2022).