

NBER WORKING PAPER SERIES

A BAYESIAN ASSESSMENT OF THE ORIGINS OF COVID-19
USING SPATIOTEMPORAL AND ZOOLOGICAL DATA

Andrew T. Levin

Working Paper 33428
<http://www.nber.org/papers/w33428>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2025

The author has no financial interests nor any other conflicts of interest related to this study. No funding was received for conducting this study. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2025 by Andrew T. Levin. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

A Bayesian Assessment of the Origins of COVID-19 using Spatiotemporal and Zoonotic Data
Andrew T. Levin
NBER Working Paper No. 33428
January 2025
JEL No. I10, I18

ABSTRACT

This paper uses Bayesian methods in conjunction with spatiotemporal and zoonotic data to evaluate the odds ratio for two hypotheses regarding the origin of the COVID-19 pandemic, namely, an accidental laboratory leak of a chimera virus or the transmission of a natural virus from an infected wildlife mammal. The overall Bayes factor is decomposed into 4 components: (1) the odds that the outbreak would occur in the People's Republic of China (PRC); (2) the odds that the outbreak would occur in Wuhan, conditional on its location in PRC; (3) the odds of observing the spatiotemporal pattern of confirmed COVID-19 cases with no known link to the specific wholesale market where wildlife mammals were being sold, conditional on the outbreak taking place in Wuhan; and (4) the odds of observing the spatiotemporal pattern of confirmed vendor cases at that market, conditional on the outbreak taking place in Wuhan. These four conditional Bayes factors are estimated as 2.3:1, 20:1, 27:1, and 12:1, respectively, and hence the overall odds ratio is 14,900:1, indicating overwhelming evidence in favor of the hypothesis that the pandemic resulted from an accidental lab leak. This conclusion is robust to alternative specifications of the detailed statistical analysis.

Andrew T. Levin
Department of Economics
Dartmouth College
6106 Rockefeller Hall
Hanover, NH 03755
and NBER
andrew.t.levin@dartmouth.edu

All supplementary materials, data, and software code have been posted at the following github repository:
<https://andrewtlewin.github.io/bayesian-analysis-of-covid-origins>

Introduction

A half-decade has passed since the onset of the COVID-19 pandemic. The World Health Organization (WHO) reports more than 7 million confirmed COVID-19 fatalities, while estimates based on excess mortality indicate an even higher global toll of about 15 to 35 million deaths.¹⁻⁴ Systematic reviews have documented extensive disparities in these outcomes, with very high incidence of SARS-CoV-2 infections in socioeconomically disadvantaged communities and in longer-term care facilities, while elevated rates of mortality have been directly linked to higher age and impaired healthcare access.⁵⁻⁹ Moreover, the pandemic has been associated with sharply lower economic growth in many low-income countries, with significant adverse consequences for their public health and wellbeing.^{10, 11}

At the early stages of the pandemic, a broad consensus of scientists inferred that the SARS-CoV-2 virus had zoonotic origins, similar to the previous coronavirus outbreaks of SARS and MERS that were traced to palm civets and dromedary camels, respectively.¹²⁻¹⁶ In reaching that conclusion, it was noted that the earliest COVID-19 cases occurred in the city of Wuhan in the People's Republic of China (PRC), and a high proportion of those cases were linked to the Huanan Wholesale Seafood Market.¹⁷⁻¹⁹ In early 2021 WHO published the findings of a joint investigation by Chinese and international experts, including detailed information on 172 early COVID-19 cases for which onset of symptoms occurred before the end of 2019.²⁰ Subsequently, Worobey et al. (2022) analyzed spatial data from the WHO-2021 report and concluded that the Huanan Market was the epicenter of the COVID-19 pandemic and that the initial outbreak was linked to a specific cluster of market shops where live wild mammals had been sold.²¹ However, no existing study has produced any statistical analysis of the spatiotemporal characteristics of the initial case data.

Among the species of wildlife mammals for sale at the Huanan Market, the raccoon dog (*Nyctereutes procyonoides*) has been highlighted as the most likely source of zoonotic transmission of a virus lineage that – like the SARS and MERS viruses -- presumably originated in horseshoe bats.^{19, 21, 22} In fact, epidemiological investigations of the SARS outbreak found that virus not only in palm civets but in raccoon dogs that were being sold for human consumption in Guangdong Province.²³⁻²⁶ In late 2019, wild raccoon dogs were observed for sale at the Huanan Market, and swab samples collected in early 2020 contained traces of raccoon dog DNA as well as viral particles of SARS-CoV-2.^{21, 27, 28} However, there has been no systematic analysis of the extent to which the zoonotic transmission of SARS-CoV-2 is consistent with the geographical distribution of wild or farmed raccoon dogs.

Beyond those specific gaps in the scientific literature, one basic question has been left unanswered: Why did the COVID-19 outbreak occur in Wuhan?²⁹ More specifically, is its location within PRC associated with an elevated risk of a zoonotic outbreak?³⁰ Conversely, could the virus have escaped from a research laboratory in Wuhan? An important factor underlying these questions is that during the decade of the 2010s, laboratories in Wuhan were at the forefront of global research on bat-related coronaviruses.^{31, 32} Moreover, much of this research was conducted at relatively limited precautions corresponding to BioSafety Level 2 (BSL-2), as distinct from the greater precautions of BSL-3 and BSL-4.^{33, 34} Numerous studies have considered these questions by analyzing the genetic sequence of SARS-CoV-2 and comparing its key features to those of other sarbecoviruses (that is, SARS-related betacoronaviruses).³⁵⁻⁴⁰ Nonetheless, there have been persistent concerns about whether the COVID-19 pandemic may have been triggered by an accidental leak of a chimera virus developed in a research laboratory rather than by zoonotic transmission from an infected animal at the Huanan market.⁴¹⁻⁴⁶

The WHO 2021 report contained a brief discussion of these scenarios and concluded that zoonotic transmission was “very likely” and that an accidental leak was “extremely unlikely.”²⁰ However, other researchers have arrived at markedly different conclusions by combining probabilistic assessments encompassing a wide range of evidence.⁴⁷⁻⁵¹ Some national government agencies have published their own independent assessments of accidental vs. zoonotic transmission, but their conclusions partly reflect non-public information obtained from confidential sources.⁵²⁻⁵⁵ Meanwhile, there has been a dearth of formal statistical analysis of publicly-available spatiotemporal and zoonotic data using standard Bayesian methods.⁵⁶⁻⁵⁸

Bayesian procedures can be used to weigh the relative probabilities of competing hypotheses, as distinct from classical statistical methods that focus on testing a single null hypothesis.^{59, 60} In particular, the *Bayes factor* indicates the odds that the observed outcome would occur under each of two competing hypotheses; the *prior odds ratio* incorporates subjective probability assessments along with other sources of information beyond the formal statistical framework; and the *posterior odds ratio* is obtained as the product of the Bayes factor and the prior odds ratio. In many contexts, the Bayes factor can be decomposed into several distinct conditional factors, thereby elucidating the analysis more clearly than would be feasible using a single “black-box” statistical model aimed at jointly encapsulating all dimensions of the observed data.⁶¹

This paper uses Bayesian methods in conjunction with spatiotemporal and zoonotic data to evaluate the odds ratio for the following two hypotheses:

A: The SARS-CoV-2 outbreak was caused by an accidental laboratory leak of a chimera virus.

Z: The SARS-CoV-2 outbreak was caused by zoonotic transmission from an infected animal.

To facilitate the transparency of this assessment, the overall Bayes factor is decomposed into four components: (1) the likelihood that PRC would be the epicenter of the pandemic; (2) the likelihood that the epicenter would be located in Wuhan, conditional on its occurrence in PRC; (3) the likelihood of observing the spatiotemporal pattern of confirmed COVID-19 cases with no known link to the Huanan Market, conditional on the outbreak taking place in Wuhan; and (4) the likelihood of observing the spatiotemporal pattern of confirmed vendor cases at the Huanan Market. Thus, the posterior odds ratio can be expressed as follows:

$$(1) \quad \text{POSTERIOR_ODDS}_{AZ} = \text{PRIOR_ODDS}_{AZ} * BF_1 * BF_2 * BF_3 * BF_4$$

Each conditional Bayes factor $BF_i = L_{iA} / L_{iZ}$, where L_{ij} denotes the conditional marginal likelihood of observing the specified data for component $i = (1, \dots, 4)$ and hypothesis $j = (A, Z)$. For each component, the conditional Bayes factor is computed using a baseline specification of the underlying data-generating process, and then robustness analysis is used to assess its sensitivity to alternative specifications, similar to approaches for addressing model uncertainty that are common in other fields such as meteorology and economic forecasting.⁶²⁻⁶⁴

The remainder of this paper is organized as follows. Section 1 briefly discusses the prior odds of the two hypotheses. Section 2 uses data on the global distribution of bats and the global distribution of research on bat-related viruses to assess the likelihood of a coronavirus outbreak in PRC under each hypothesis. Section 3 assesses the conditional likelihood of a coronavirus outbreak in Wuhan, focusing

on the geographical distribution of wildlife mammals that were being sold for human consumption in PRC. Sections 4 and 5 conduct spatiotemporal analysis of initial case data in the vicinity of Wuhan and within the Huanan market, respectively. Section 6 discusses the results, identifies limitations of the analysis, and highlights directions for further research. Methodological details are provided in the Supplementary Materials. All software and data used in this project have been posted on Github.

1. Prior Odds

Since the prior odds are specified apart from any formal statistical procedure, a common approach is to place equal prior probabilities on the two hypotheses – commonly known as *diffuse priors* or *flat priors* -- so that the posterior odds are determined solely by analyzing the observed data, i.e., “letting the data speak.” However, such an approach may well exclude background information that should be reflected in the quantitative assessment of odds. In the present context, the consideration of such information is important not only for specifying the prior odds but for illuminating the scientific consensus at the onset of the COVID-19 pandemic and providing helpful context for the formal statistical analysis below.

1.1 Key Features of Pre-COVID Pandemics

Table 1 summarizes key features of pre-COVID pandemics in the 20th and 21st centuries, that is, infectious disease outbreaks resulting in substantial fatalities across multiple countries.⁶⁵⁻⁶⁷

Pandemic	Date of Outbreak	Disease	Zoonotic Transmission	Estimated Fatalities	Sources
Spanish Flu	1918	Influenza A/H1N1	aquatic birds	15-100 million	68-70
Asian Flu	1957	Influenza A/H2N2	aquatic birds	0.7-1.5 million	71
Hong Kong Flu	1968	Influenza A/H3N2	aquatic birds	1 million	72
Ebola	1976	Ebola virus	bats/primates	13 thousand	73
AIDS	1981	HIV	primates	40 million	74, 75
SARS	2002	Betacoronavirus	bats/civets	783	76
Swine Flu	2008	Influenza A/H1N1	swine	150-450 thousand	77
MERS	2012	Betacoronavirus	bats/camels	888	78

It is notable that each of these pandemics was associated with a virus.^{79, 80} Over prior centuries and millenia, pandemics were frequently associated with pathogenic bacteria (e.g., cholera, typhoid, and typhus), but the impact of those diseases has been contained or even eliminated by the use of antibiotics and hygienic measures.⁸¹ Likewise, improvements in sanitation and other public health measures have greatly diminished the burden of vector-borne diseases (e.g., bubonic plague) that are primarily spread by insects or other animals, although mosquito-borne diseases such as malaria and yellow fever continue to afflict many populations in tropical regions.^{82, 83} By contrast, the HIV and Ebola viruses are generally spread by contact with human bodily fluids, while the deadly strains of influenza and coronavirus are respiratory illnesses that mainly spread via airborne particles.^{84, 85}

Furthermore, each of the diseases listed in Table 1 emerged from a zoonotic source, often involving distinct species that act as natural reservoirs or intermediate hosts for multiple strains of the evolving virus.^{86,87} For example, aquatic birds have been identified as a natural reservoir for numerous strains of influenza virus, which infects those species without any significant pathological effects, while swine have been identified as an intermediate host in which mutations produce new variants that are particularly virulent in humans.⁸⁸ Similarly, bats have been identified as a natural reservoir for the group of filament-shaped viruses (*Filoviridae*) that includes the Ebola and Marburg viruses and for the group of crown-shaped viruses (*Coronaviridae*) that includes SARS and MERS as well as SARS-CoV-2.⁸⁹⁻⁹¹

1.2 Lessons from the SARS Pandemic

A number of key aspects of the SARS pandemic provide important background and context in considering prior odds regarding the origins of SARS-CoV-2:

- *Pattern of Early SARS Cases.* The earliest identified cases of SARS, with symptom onset between November 2002 and January 2003, occurred independently in at least five different well-separated municipalities in southern Guangdong province, suggesting multiple introductions of a virus or similar viruses from a common source.^{76, 92} A subsequent study identified asymptomatic SARS-like infections in 1.8% of serological samples collected from healthy Hong Kong adults in May 2001, consistent with sequential instances of interspecies transmission of virus strains that had not yet adapted efficiently to a human host.⁹³
- *Link to Wildlife Animals.* Restaurant chefs and others handling wildlife animals for human consumption accounted for 39% of confirmed SARS cases during the initial phase of the epidemic, whereas no cases were identified in farmers handling livestock or poultry.⁹² A serology study of food market vendors in Shenzhen found anti-SARS antibodies in 40% of wildlife animal traders and only 5% of vegetable vendors.²³
- *Identification of Zoonotic Source.* In a study of serology samples from vendors at three live animal markets in Guangzhou, the incidence of SARS-CoV antibodies was highest (72%) among those selling masked palm civets (*Paguma larvata*) and lowest (9%) among those selling snakes.⁹⁴ In early 2004, researchers found live virus in 100% of oral and rectal swabs collected from a random sample of civets at a wildlife market.⁹⁵ Genetic analysis indicated a sequence identity of 99.8% between the animal viruses and the SARS-CoVs isolated from human patients, differing by a single segment whose deletion evidently occurred during adaptation of the animal-derived virus to its new human host.²³
- *Mode of Zoonosis.* Among the earliest SARS cases were the cook and chief waiter at a Guangzhou restaurant that served palm civets and a customer of that restaurant who was seated near a group of wire cages that typically held six live palm civets.⁹⁶ More broadly, large numbers of masked palm civets were sold for human consumption in southern China. A survey conducted in 1993-96 found that masked palm civets were the most popular species of exotic mammals sold at restaurants in Guanxi, a province adjacent to Guangdong.⁹⁷ Biweekly surveys in 2001-02 indicated an average of about 190 civets on sale at the Chatou wholesale food market in Guangzhou.⁹⁸ In fact, sales at that particular market occasionally exceeded 500 civets; see the Supplementary Materials for further details.

- *Role as Amplifying Intermediary.* Serology studies found no SARS-CoV infections in wild or farmed civets, except for a single transit center where civets were purchased and then exported to southeast Asia.^{95, 99, 100} Moreover, genetic analysis indicated that strains of SARS-CoV were evolving rapidly in palm civets as well as humans.¹⁰¹ These findings indicated that the virus had recently jumped from another species and had spread to civets at wildlife markets.⁹⁵ In effect, civets were an amplifying intermediary but not the natural host of the virus.^{24, 76}
- *Identification of Host Reservoir.* In the wake of the SARS outbreak, closely related viruses were identified in horseshoe bats (*Rhinolophidae*) and other bat families.^{99, 102, 103} However, SARS-CoV does not infect any of the bat species that inhabit southern China and replicates very inefficiently in other bat species.^{24, 104, 105} Evidently, the virus was transmitted from bats to the intermediary host and then evolved over a substantial period prior to its zoonotic spillover to humans.^{106, 107}
- *Mode of Transmission from Bats to Civets.* The most closely related bat viral strains (WIV1 and WIV16) exhibit similarity of about 96% to the SARS-CoV genome; those strains were cultured from samples collected at an abandoned mine in Yunnan province, about 1500km from Guangzhou.¹⁰⁸⁻¹¹⁰ The transmission from bats to civets presumably occurred via fecal matter rather than direct contact.^{102, 103} because stacked wire cages were commonly used for the transportation and sale of civets at wildlife markets.⁹⁵ Thus, the ancestor of SARS likely infected civets in warehouses or other facilities where bats roosted overhead and defecated on the closely-packed cages below. The virus then began spreading among the tens of thousands of civets that were being transported and consumed each year, evolving into a strain that was infectious to humans as well as various other types of mammals, including domestic cats, ferrets, foxes, and raccoon dogs.^{24, 26, 76, 102, 103, 111}

1.3 Lessons from the MERS Pandemic

The lessons from SARS were highly relevant when the MERS outbreak transpired in 2012. Genetic analysis of the MERS virus revealed similarities to other betacoronaviruses associated with various species of bats around the globe, triggering an intense search for zoonotic sources.^{78, 112-118} Serological analysis detected anti-MERS antibodies in 100% of a sample of 50 dromedary camels from Oman; MERS viral strains were later found in dromedaries in Egypt and Kenya.¹¹⁹⁻¹²¹ In contrast, no such antibodies were detected in any other species of livestock.^{119, 122} Moreover, a detailed case study documented the direct transmission of MERS from sick camels to a human patient in Saudi Arabia.¹²³

However, a notable difference between SARS and MERS is in the timing of the viral jump from bats to an intermediate host species. In the case of SARS, that cross-species transmission appears to have been roughly contemporaneous with the first wave of human infections. In contrast, retrospective analysis of archived camel sera indicate that the MERS virus had been circulating in Middle Eastern dromedaries since the early 1980s with practically no genetic modifications.¹²⁴ In effect, camels had become a host reservoir for the virus long before the onset of the MERS pandemic.¹²⁵

1.4 Zoonotic Implications for SARS-Cov-2

Soon after the onset of the COVID-19 pandemic, numerous parallels with SARS and MERS were identified. Genetic analysis indicated that this novel coronavirus was closely related to SARS-CoV,

thereby leading to its designation as SARS-CoV-2.^{126, 127} Each of these viruses (SARS, MERS, and SARS-CoV-2) infect cells by binding to a specific receptor, i.e., angiotensin-converting enzyme 2 (ACE2).^{128, 129} The nucleotide sequence of the receptor binding domain (RBD) in the SARS-CoV-2 genome is very similar to two SARS-related bat viruses (Rs3367 and RaTG13) that had previously been studied by researchers at the Wuhan Institute of Virology (WIV), leading to the inference that SARS-CoV-2 also originated in bats.^{108, 127, 130-132} Nonetheless, the SARS-CoV-2 virus does not infect bats, and its genome differs from that of all known bat viruses by 1000+ nucleotides, and hence this virus must have evolved in intermediate or transient hosts prior to its zoonotic transmission to humans.¹³³

Moreover, as noted above, a substantial fraction of the earliest COVID-19 cases were directly linked to the Huanan Wholesale Seafood Market.^{17, 18} Media reports highlighted an array of exotic animals that were on sale at the Huanan market, including the species associated with the SARS pandemic (palm civets and raccoon dogs) as well as meat from camels (the zoonotic host for MERS).¹³⁴

Qualitative Assessment	Odds Ratio	Qualitative Assessment	Odds Ratio
hypotheses A & B equally likely	1:1	hypotheses A & B equally likely	1:1
anecdotal evidence for A	3:1	anecdotal evidence for B	1:3
strong evidence for A	10:1	strong evidence for B	1:10
very strong evidence for A	30:1	very strong evidence for B	1:30
extremely strong evidence for A	100:1	extremely strong evidence for B	1:100

Source: Lee and Wagenmakers (2013), table 7.1.

Consequently, when the WHO convened its international review in early 2021, the members of its panel concluded that the virus had most likely originated in bats and then spread to humans through an intermediate host, whereas a scenario involving a laboratory accident was judged as “extremely unlikely.” As shown in Table 2, in a Bayesian context that assessment would correspond to odds of about 100:1 or more in favor of hypothesis Z relative to hypothesis A. In early 2023, a group of 156 virologists stated that “currently the zoonosis hypothesis has the strongest supporting evidence.”¹³⁵ Likewise, the editors of three virology journals concluded that hypotheses involving an accidental lab leak were “unlikely.”¹³⁶ In spring 2024, an article in the *Annual Review of Virology* concluded that “The available data clearly point to a natural zoonotic emergence within, or closely linked to, the Huanan Seafood Wholesale Market in Wuhan.”¹³⁷

1.5 Accidental Laboratory Leaks

The prospect of an accidental pathogenic leak has been a longstanding matter of concern, leading to the standardization of biosafety practices at research laboratories. For example, in the USA and the European Union (EU), a lab operating at biosafety level 4 (BSL-4) must be used for investigating highly infectious and potentially lethal microbes for which no vaccine or treatment is currently available, whereas biosafety level 2 (BSL-2) is appropriate for working with moderately hazardous agents associated with mild disease and no aerosol transmission.¹³⁸ Such standards may be quite different than the regulations applicable to facilities engaged in manufacturing vaccines and other biochemical substances, where susceptibility to accidental leaks arises from other factors.¹³⁹

Nonetheless, strict safety standards cannot completely eliminate laboratory accidents due to human error or equipment failure.¹⁴⁰⁻¹⁴³ A report by the U.S. Center for Disease Control & Prevention (CDC) identified multiple instances between 2003 and 2010 in which personnel at U.S. BLS-3 laboratories contracted lab-acquired infections without providing any notification to authorities.¹⁴⁴ A subsequent study estimated a probability of 20% that an accident at a U.S. research lab would result in an infectious disease outbreak over a 10-year period.¹⁴⁵ Accidental lab leaks have also been identified in PRC, including a leak of the SARS virus from the China CDC research lab in 2004 and a *brucella* leak from the Lanzhou Veterinary Research Institute in early December 2019.s.¹⁴⁶⁻¹⁴⁸

Concerns about the risk of a catastrophic leak were magnified in conjunction with the rapid development and dissemination of tools for manipulating the genome of pathogenic viruses.¹⁴⁹⁻¹⁵¹ For example, the method of serial passage can facilitate the rapid evolution and optimization of the virus by repeated cycles of inoculation, reproduction and harvesting in sequential cohorts of infected animals or cell cultures.¹⁵² Alternatively, researchers can create a chimera virus by directly modifying its genetic sequence (via insertion, deletion or substitution), replicate the virus using polymerase chain reaction (PCR), and evaluate its infectiousness and virulence using cell cultures.¹⁵³

Genetic Sequence	# Nucleotides	Relative Scale	Length
Homo sapiens	6 000 000 000	Earth to Moon	400,000 km
E. coli	4 600 000	New York City to Washington DC	300 km
Sars-Cov-2	29 900	Eiffel Tower to Louvre Museum	2 km
S gene (spike)	3 822	Tower Bridge to Tower of London	240 m
Furin cleavage site	12	Small Desk	0.8 m

As shown in Table 3, the genetic sequence of a coronavirus is far smaller than that of a common bacteria and miniscule compared to the human genome, and the insertion of a few codons can dramatically affect its properties. For example, some contemporary avian influenza viruses differ by only a few amino acids from the virus that caused the 1918 pandemic.¹⁵⁴

Thus, experiments with chimera viruses need not involve an entire laboratory but can be conducted by one or more graduate students or postdoctoral researchers working on a limited budget. Such considerations were apparent in a 2012 commentary published by the director of the U.S. National Institute for Allergy and Infectious Disease:

*...consider this hypothetical scenario: An important gain-of-function experiment involving a virus with serious pandemic potential is performed in a well-regulated, world class laboratory by experienced investigators, but the information from the experiment is then used by another scientist who does not have the same training and facilities and is not subject to the same regulations. In an unlikely but conceivable turn of events, what if that scientist becomes infected with the virus, which leads to an outbreak and ultimately triggers a pandemic?*¹⁵⁵

Likewise, in a review article published in March 2019, China's CDC director stated:

The release of biological agents, whether due to natural, accidental or deliberate causes, is among the most serious challenges to humanity... Advances in biomedical technologies, such as genome

*editing and synthetic biotechnology, have the potential to provide new avenues for biological intervention in human diseases...However, the proliferation of such technologies means they will also be available to the ambitious, careless, inept, and outright malcontents, who may misuse them in ways that endanger our safety.*¹⁵⁶

In many legal jurisdictions such as the USA and the EU, research on highly infectious and pathogenic respiratory viruses may only be conducted in BSL-3 and BSL-4 labs to mitigate the risk of accidental release.^{157, 158} By contrast, Chinese laboratories were routinely conducting such research in BSL-2 and BSL-3 labs with deficiencies in biosafety training and equipment.¹⁵⁹⁻¹⁶¹ In a September 2019 review article, WIV's deputy director stated that "*some BSL-3 laboratories run on extremely minimal operational costs, or in some cases, none at all.*"¹⁶² In light of such deficiencies, one prominent virologist indicated that "*lab-acquired infections occur much more frequently at BSL-2...if you study hundreds of different bat viruses at BSL-2, your luck may eventually run out.*"¹⁶³ Another virologist described the risks associated with such research as "*unacceptable.*"¹⁶⁴

Genetic analysis has flagged characteristics of SARS-CoV-2 that may point to a non-natural origin:

- *Adaptation to Humans.* The genetic sequence of SARS-CoV-2 remained extraordinarily stable from late 2019 through summer 2020, undergoing practically no adaptive mutations even as it spread worldwide, indicating that the ancestral strain of SARS-CoV-2 was remarkably well adapted to humans.¹⁶⁵ One specific mutation (involving a single nucleotide in the spike segment of the genome) arose in February 2020 and became universally dominant thereafter; the timing of that mutation showed that the ancestral virus could not have been fully optimized for human respiratory infection using serial passage methods.¹⁶⁶⁻¹⁶⁹ Two other single-nucleotide mutations were identified during late spring and summer, but the Alpha variant of concern did not emerge until September 2020.^{167, 170, 171} By comparison, the SARS and MERS viruses each underwent much more extensive adaptive mutations during the early phases of those pandemics.^{35, 172, 173}
- *Binding Efficiency to Cell Receptors.* Every coronavirus has spike proteins that include a receptor-binding domain (RBD), which has a specific pattern of amino acids that enables the virus to bind to a specific receptor protein protruding from the surface of the host cell.¹⁷⁴ In SARS-like viruses, the RBD binds to ACE-2, a protein that occurs in nearly all vertebrates but with wide differences in relative frequency and surface characteristics.^{131, 175} Notably, the RBD of COVID-19 has peak efficiency in binding to the ACE-2 protein of humans and apes; lower efficiency in binding to the ACE-2 proteins of other mammals; and no capacity to bind effectively to the ACE-2 proteins of fish, amphibians, reptiles, and birds.^{36, 176-178} One group of researchers concluded that "*SARS-CoV-2 may not be especially adapted to the ACE-2 of any of its putative intermediate hosts.*"¹⁷⁹
- *Cleavage of Spike Protein.* The coronavirus spike protein has two components: the S1 unit contains the RBD that binds to a specific receptor on the surface of the host cell, and then the S2 unit facilitates entry into the host cell.¹⁷⁴ In SARS and other related viruses, the cleavage of these two components is initiated at the S1/S2 junction during cell infection after the S1 unit has bound to the host receptor.^{180, 181} SARS-CoV-2 is unique among SARS-related viruses in having a polybasic sequence of amino acids at the S1/S2 junction that facilitates cleavage by furin protease during the process of viral maturation within the infected host cell, thereby "pre-activating" the virus for efficient entry into other cells.^{128, 182-186} Researchers have shown

that this polybasic furin cleavage site is essential in enabling SARS-CoV-2 to infect human respiratory cells.^{187, 188}

Enigmatic Evolutionary Origins. Polybasic furin cleavage sites have been identified in other pathogenic human viruses (including HIV, Ebola, and Marburg) and in various avian, rodent, and bat-related viruses.¹⁸⁹⁻¹⁹⁴ The polybasic furin cleavage site generally has the pattern *RR-RR*, *R-RR*, or *R-KR* involving the two highly basic acids arginine (*R*) and lysine (*K*). By contrast, the SARS-CoV-2 virus has a distinctive polybasic sequence (*RR-R*) that has not been found in any other SARS-related coronavirus.^{195, 196} By aligning the genomes of SARS and SARS-CoV-2, it is evident that this cleavage site involves four amino acids at the S1/S2 junction of the viral genome.^{14, 37, 38, 40, 197-202} As shown in Table 4, all the SARS-related bat viruses with high overall genetic similarity to SARS-CoV-2 have a monobasic amino acid at the S1/S2 junction; none of those viruses have a polybasic sequence that would facilitate more efficient furin cleavage.²⁰³ Nonetheless, gene editing has been used to insert furin cleavage sites into chimera viruses, and in 2018 WIV was preparing to investigate potential furin cleavage sites in SARS-related viruses and to culture those viral chimera in cultures of human respiratory cells.^{180, 204-211}

Table 4: The S1/S2 Cleavage Site in SARS-CoV-2 and Other SARS-Related Bat Viruses

Description	Location	Bat Species	Identifier	Sequence of Amino Acids											
				S	P	R	R	A	R	S	V	A	S	Q	
SARS-CoV-2	Wuhan, PRC	----	Wuhan-Hu-1	S	P	R	R	A	R	S	V	A	S	Q	
	Yunnan, PRC	<i>R. affinis</i>	RaTG13	•	◦	◦	◦	◦	R	•	•	•	•	•	
	Laos	<i>R. malayanus</i>	BANAL-20-52	•	◦	◦	◦	◦	R	•	•	•	•	•	
	Laos	<i>R. pusillus</i>	BANAL-20-103	•	◦	◦	◦	◦	R	•	•	•	•	•	
	Laos	<i>R. marshalli</i>	BANAL-20-236	•	◦	◦	◦	◦	R	•	•	•	•	•	
Bat Viruses	Cambodia	<i>R. shameli</i>	RShSTT-182	•	◦	◦	◦	◦	R	•	•	•	•	•	
	Yunnan, PRC	<i>R. malayanus</i>	RmYN02	•	•	A	◦	A	R	◦	•	G	T	N	
	Laos	<i>R. malayanus</i>	BANAL-20-116	•	•	A	◦	A	R	◦	•	G	T	N	
	Thailand	<i>R. acuminatus</i>	RacCS203	•	•	A	◦	A	R	◦	•	G	T	N	

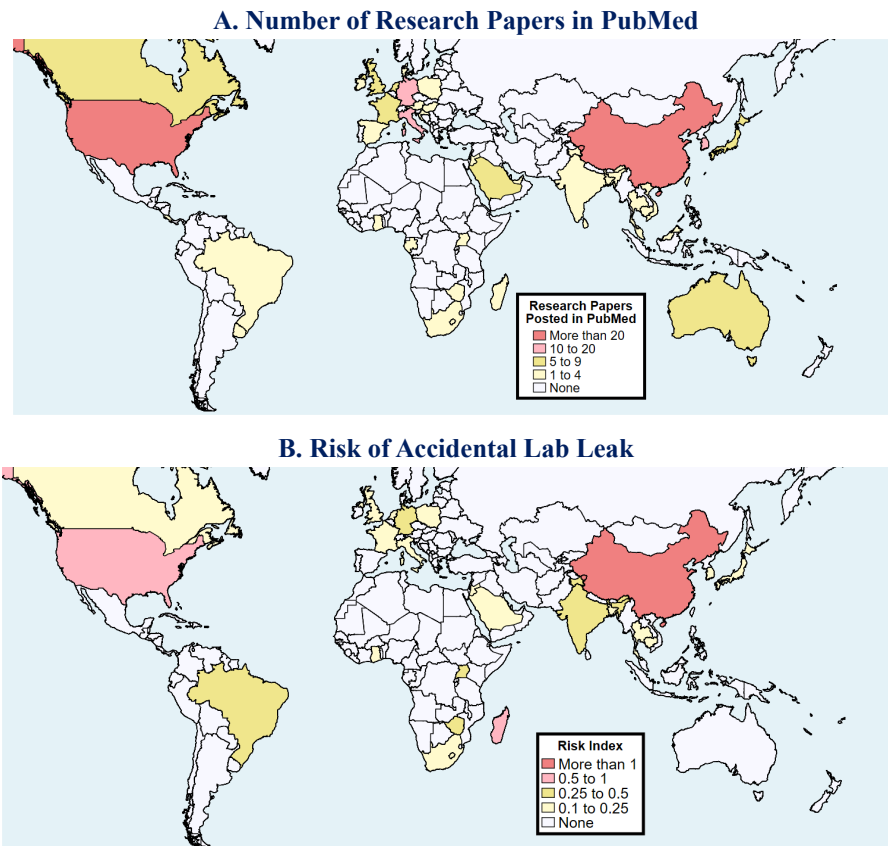
Note: This table compares the genetic sequence of SARS-CoV-2 with eight other bat viruses at the S1/S2 junction of the spike protein, where each bat viral sequence has been aligned to highlight the similarities with SARS-CoV-2. The amino acid arginine (R) is basic and positively charged, and hence this amino acid is indicated in bold green, whereas the other amino acids in these sequences are neutral and uncharged: alanine (A), asparagine (N), glycine (G), glutamine (Q), proline (P), serine (S), threonine (T), and valine (V). For each aligned position in each bat viral strain, a solid bullet (•) indicates that the amino acid is identical to that of SARS-CoV-2, whereas a hollow bullet (◦) indicates the absence of any amino acid at that aligned position. The blue vertical bar indicates the cleavage site.

1.6 Assessment of Prior Odds

The Lancet COVID-19 Commission, a global panel of 28 experts, carefully considered the likelihood that SARS-CoV-2 arose from a zoonotic spillover or a laboratory accident; its final report, issued in October 2022, indicated that panelists “held diverse views about the relative probabilities of the two explanations, and both possibilities require further scientific investigation.”²¹² Likewise, WHO’s Scientific Advisory Group for the Origins of Novel Pathogens (SAGO) stated in June 2022 that it was “not able to identify any conclusive findings” regarding the origins of SARS-CoV-2.²¹³ In April 2023, the COVID Crisis Group’s final report indicated that: “Both theories remain plausible....At present, we just do not think there is enough evidence available, yet, to come down hard either way.”²¹⁴

From a Bayesian perspective, these judgments would naturally suggest the use of flat priors for the present analysis, i.e., prior odds of 1:1 for the two hypotheses A and Z . In representing other perspectives, it will also be helpful to assess robustness using alternative priors of 1:100 or 100:1; such priors might reflect alternate interpretations of the genetic data and any other evidence not incorporated into the formal statistical analysis of this study.

Figure 1: Global Risks of Bat-Related Coronavirus Research



Note: The upper panel shows the country-specific volume of research on bat-related viruses, and the lower panel shows an index of the risk of an accidental lab leak of a bat virus, using the methodology described in Table 5.

2. Why Did the Pandemic Start in PRC?

This section focuses on assessing the conditional Bayes factor BF_1 , i.e., the relative odds that the global pandemic outbreak from a bat-related coronavirus would be located in PRC under each of the two competing hypotheses A (accidental lab leak) and Z (zoological spillover). This conditional Bayes factor can be expressed as follows:

$$(2) \quad BF_1 = \frac{\text{Marginal Likelihood}(\text{PRC} \mid \text{hypothesis } A)}{\text{Marginal Likelihood}(\text{PRC} \mid \text{hypothesis } Z)}$$

The numerator indicates the marginal likelihood that the pandemic outbreak would be observed in PRC under hypothesis A , and the denominator is the marginal likelihood of that outcome under hypothesis Z . Thus, each of these marginal likelihoods can be evaluated using geographical information. In particular, data on the global distribution of bat-related viral research during 2010-19 can be used to assess the probability of a catastrophic lab leak occurring in PRC rather than some other location, and data on the global distribution of bat species can be used to assess the probability of a bat-originated zoonotic spillover occurring in PRC rather than in some other location.

2.1 The Global Distribution of Bat-Related Viral Research

To assess the country-specific risk of a catastrophic lab leak, it would be ideal if one could refer to a multinational tabulation of risk characteristics of the projects being conducted within each individual lab in each jurisdiction, comparable to the macroprudential information and stress tests that are routinely reported for systemically important financial institutions.^{215, 216} Unfortunately, no such tabulation of biological research laboratories currently exists.²¹²

Thus, in the absence of such detailed information, country-specific risks can be assessed using tabulations of articles and working papers on bat-related coronavirus research.⁴⁹ In particular, the PubMed database provides global coverage of citation data for nearly 40 million biomedical and life sciences research articles, books, and working papers, and each database entry indicates the country where that research item was published.²¹⁷

The upper panel of Figure 1 shows each country's share of new research papers on bat-related coronaviruses that were posted in PubMed between 2015 and 2019, based on the national affiliation of each paper's lead author. This tabulation was performed using the all-fields search string (*("bat" AND "coronavirus") OR "alphacoronavirus" OR "betacoronavirus"*); the composition is similar to that obtained by Phelps et al. (2019), whose tabulation encompassed all bat-viral research since 1948.²¹⁸

Of course, lab-based research generally involves a substantial capital investment to establish the lab and purchase the requisite equipment, which is then used to produce a stream of research output over an extended period thereafter. Thus, the risk analysis performed in this paper focuses on countries that have produced at least five publications on bat-related coronavirus research.

In the absence of detailed information on regulation and training, a reasonable proxy is to utilize each country's level of GDP per capita, as computed by the World Bank based on purchasing-power parity (PPP) adjustments.²¹⁹ Such an approach is broadly consistent with the judgmental assessments of national biosafety standards used in the Global Health Security Index (GHSI).²²⁰ Specifically, the GHSI incorporates qualitative questions for two components of biosafety: (1) *Has the country established laws and regulations to prevent accidental exposure to harmful biological agents, and is a specific government agency responsible for enforcing those biosafety standards?* (2) *Does the country have standardized requirements for biosafety training?* As shown in the Supplementary Materials, six advanced and emerging-market countries (Australia, France, Germany, Jordan, South Africa, and USA) have a 100% GHSI rating on national biosafety standards, whereas six other developing countries (Egypt, Ghana, Kenya, Phillipines, PRC, and Thailand) have zero ratings on that criterion.

In assessing the risk of a catastrophic lab accident, it is also important to account for systematic differences in national biosafety regulations and enforcement mechanisms as well as differences in education and training of laboratory personnel (including technicians and support staff). In a review article published in September 2019 the China CDC's biosafety director acknowledged that *"a comprehensive system of legal and regulatory standards is lacking for BSL-2 laboratories in China...we do not have enough well-trained and experienced laboratory biosafety (LB) specialists. ...Compared to developed countries, China is still in the beginning stages of LB development... The design and reliability of our LB system also lacks acute evaluation criteria and schemes."*²²¹

Table 5: Country-Specific Risks of Bat-Related Coronavirus Research

Country	Country-Specific Indicators			Share of Global Total (%)		
	Papers in PubMed	GDP Per Capita (\$)	Risk Index	Papers in PubMed	Risk Index	Alternative Risk Index
PRC	82	18,465	4.97	30.1	43.6	56.0
USA	62	69,459	1.00	22.8	8.8	11.3
Madagascar	1	1,757	0.64	0.4	5.6	--
Uganda	1	2,693	0.42	0.4	3.6	--
Zimbabwe	1	3,413	0.33	0.4	2.9	--
Germany	16	62,507	0.29	5.9	2.5	3.2
India	2	7,964	0.28	0.7	2.5	3.2
Brazil	4	17,650	0.25	1.5	2.2	2.9
Korea	10	46,904	0.24	3.7	2.1	2.7
Cambodia	1	4,786	0.23	0.4	2.0	--

Note: This table provides quantitative information for the 10 countries assessed as having the highest risk of bat-related coronavirus research; corresponding information for all other countries is provided in the Supplementary Materials. For each country, the first data column shows the total number of research papers on bat-related coronaviruses posted in the PubMed database during 2015-19, based on the affiliation of each paper's lead author. The second data column shows the World Bank's estimate of GDP per capita in 2019 using PPP-adjusted constant international dollars.²¹⁹ The third data column reports the assessment of bat-related coronavirus research risk, computed as the number of research papers divided by the level of GDP per capita and indexed to the US value of this ratio. The fourth and fifth columns show the country's share of global research papers and its share of global risk of an accidental lab leak. The final column shows an alternative risk metric that excludes countries for which only a single bat viral research paper was posted in PubMed during 2015-19.

Contemporaneously, WIV's biosafety director stated: "Currently, most laboratories lack specialized biosafety managers and engineers. In such facilities, some of the skilled staff is composed by part-time researchers. This makes it difficult to identify and mitigate potential safety hazards."¹⁶²

Thus, this paper computes the country-specific risk index by taking the tally of bat-related coronavirus research papers for each country and dividing by GDP per capita in 2019. Table 5 reports this risk index and the indicators used in computing it for each of the ten countries identified as having the highest risk of an accidental lab leak; the Supplementary Materials provide corresponding information for all other countries included in this analysis.

These results provide a reasonable estimate for $Marginal_Likelihood[PRC | hypothesis A]$, that is, the numerator of the first Bayes Factor. In particular, conditional on the hypothesis that an accidental lab leak of a bat-related coronavirus occurred somewhere in the world, this risk assessment indicates a probability of 43.6% that the lab accident occurred in PRC. Nonetheless, it is apparent that this estimate is subject to substantial uncertainty, roughly consistent with the interval of about 30 to 56% implied by the alternative approaches that are also shown in Table 5.

2.2 The Global Distribution of Bat-Linked Zoonosis Risk

Bats were identified a century ago as the likely host reservoir for the rabies virus, and more recently as a host for the viruses that cause other human diseases, including Ebola, Hendra, Marburg, MERS, Nipah, and SARS.²²²⁻²²⁵ Bats have a number of extraordinary characteristics that facilitate their role as a viral host.²²⁶⁻²²⁹ As the sole class of flying mammal, bats have distinct respiratory and immune systems that minimize the pathogenic effects of persistent viral infections.^{225, 230, 231} A high proportion of bats reside in roosts or colonies that are often shared by more than one species, thereby enabling respiratory viruses to spread rapidly both within and across bat species.^{232, 233} The capacity for flight has enabled bats to spread across the globe, including locations such as the Hawaiian islands that are nearly 4000 km from the nearest landfall.²³⁴ The species diversity of bats is greater than that of any other mammal, reflecting bats' adaptation to a wide array of ecological conditions including tropical forests and alpine terrain.²³⁵ Onsite sampling and echolocation studies have found bats in all locations worldwide except for Antarctica, Greenland, and Iceland.²³⁶

During the decade of the 2010s, a number of studies reached somewhat disparate conclusions in assessing hotspots for bat-related zoonotic outbreaks, but none of those studies identified PRC as the most likely location. For example, a worldwide review of all known bat viruses found that *“there is no geographical bias based on viral genus, i.e. alpha and beta coronaviruses are equally likely to be found in all regions.”*²³⁷ A global analysis of all mammalian hosts identified the Amazon region of South America as the most likely source of bat-related zoonotic outbreaks, based on the diversity of species of bats and coronaviruses as well as the close proximity of bats to human populations.²³⁸

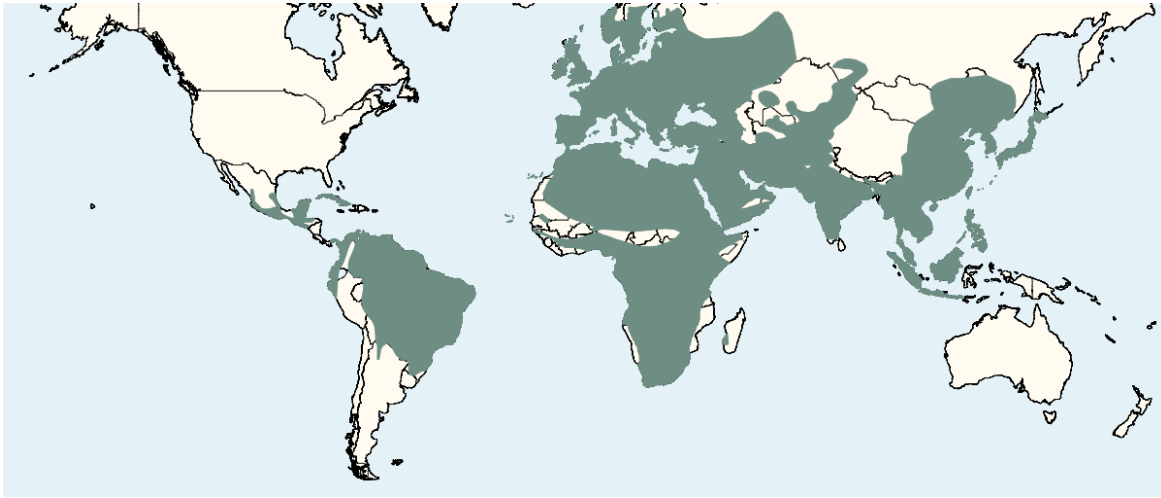
Another global study of bat viruses incorporated further indicators, including the incidence of bushmeat hunting and livestock production, and identified *“hotspots of virus sharing primarily in sub-Saharan Africa, as well as in South and East Asia, Southern Europe, and Central America.”*²³⁹ Such analysis may have spurred a CNN report on *“Virus Hunters Searching for the Next Disease Outbreak”*, which profiled research on cave-dwelling bats near Johannesburg, South Africa.²⁴⁰

Figure 2 shows the overall ranges of all known bat species in which MERS-related viruses and SARS-related viruses have been identified using pharyngeal/anal swabs or excreta samples. These tabulations of bat species have been compiled from the global surveys of Frutos et al. (2021) and Wu et al. (2023).^{233, 241} The geographical range of each bat species is shown using the geospatial coordinates published by the International Union of Conservation of Nature (IUCN), an organization that is widely recognized as the global authority on the status of biological species.²⁴²

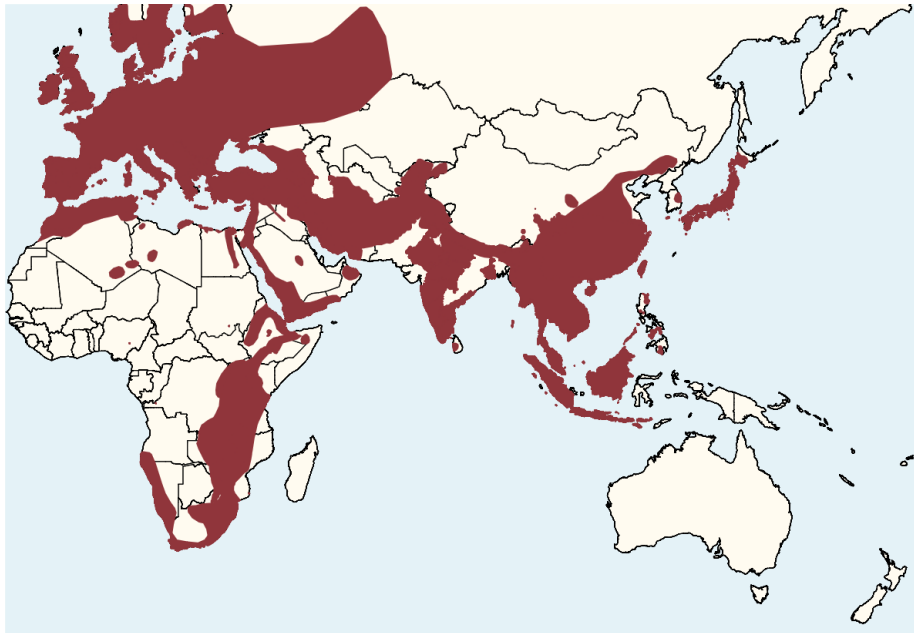
As shown in the upper panel of the figure, bat species carrying MERS-related viruses are present throughout most of the inhabited world. Exceptions include some parts of South America (e.g., west of the Andes), North America (north of the Sierra Madre del Sur), northeastern Asia (Siberia, Mongolia, and northwestern China), and Oceania (Australia, New Zealand, New Guinea, and Polynesia). As shown in the lower panel, SARS-related bat viruses have been identified in a similar swath of the eastern hemisphere (except sub-Saharan Africa) but not in the western hemisphere.

Figure 2: The Global Distribution of Bat-Hosted BetaCoronaviruses

A. MERS-Related Viruses



B. SARS-Related Viruses



Note: The upper panel shows the overall range of all known bat species in which MERS-related viruses have been identified (using pharyngeal/anal swabs or excreta samples), and the bottom panel shows the overall range of all known bat species in which SARS-related viruses have been identified. Each group of bat species has been tabulated using the global surveys of Frutos et al. (2021) and Wu et al. (2023). The range of each bat species is shown using the geospatial coordinates published by the International Union of Conservation of Nature (IUCN), an organization that is widely recognized as the global authority on the status of biological species. See the Supplementary Materials for additional details.

Geographic Range of Bat Host Species	Population in 2019 (millions)		Ratio (%)
	PRC	Total	
All Coronaviruses	1,415.9	7,742.3	18.3
SARS-Related Viruses	1,009.8	5,207.9	19.4
MERS-Related Viruses	1,415.9	7,134.3	19.8

Note: This table provides quantitative information about human populations living within the ranges of bat species identified as hosting coronaviruses (first row), SARS-related viruses (second row), and MERS-related viruses, using the tabulations described in Table 4 and the geospatial data shown in Figure 2. The first two data columns shows the PRC population and total population (in millions) living in the specified bat range, and the third data column shows the ratio (in %). Details of these calculations are provided in the Supplementary Materials.

These results provide a reasonable assessment for $Marginal_Likelihood[PRC | hypothesis Z]$, that is, the denominator of the first Bayes Factor. In particular, conditional on the hypothesis that the COVID-19 pandemic was triggered by a zoonotic spillover of a bat-related coronavirus, the middle row of Table 6 indicates a probability of 19.4% that this spillover occurred in PRC. Moreover, the alternative estimates are well aligned, implying a plausible interval of about 18 to 20% for this marginal likelihood.

2.3 Bayesian Implications

In light of the results given in Tables 4 and 5, the value of BF_1 can now be computed as the ratio of 43.6% to 19.4%, implying a conditional odds ratio of 2.25 to 1 for hypothesis A vs. hypothesis Z . Referring back to the qualitative interpretations in Table 2, this conditional odds ratio is equivalent to “anecdotal evidence” in favor of hypothesis A . In effect, observing that the COVID-19 pandemic started in PRC provides some useful evidence but by itself would not be sufficient to offset strong priors in favor of hypothesis Z .

3. Why Did the Pandemic Start in Wuhan?

This section focuses on assessing the conditional Bayes factor BF_2 . This factor is conditional on the pandemic outbreak having taken place in PRC, and hence it indicates the conditional odds that the pandemic outbreak would be located in Wuhan (rather than somewhere else in PRC) under each of the two competing hypotheses A (accidental lab leak) and Z (zoological spillover). This conditional Bayes factor can be expressed as follows:

$$(3) \quad BF_2 = \frac{\text{Marginal Likelihood}(\text{Wuhan} \mid \text{PRC} \ \& \ \text{hypothesis } A)}{\text{Marginal Likelihood}(\text{Wuhan} \mid \text{PRC} \ \& \ \text{hypothesis } Z)}$$

Section 3.1 reviews the relevant background information and then uses the qualitative assessments in Table 2 to specify the marginal likelihood of a lab leak in Wuhan vs. elsewhere in PRC. Section 3.2 uses zoonotic and geospatial information to assess the marginal likelihood that a zoonotic spillover would have occurred in Wuhan vs. elsewhere in PRC. Section 3.3 synthesizes these results.

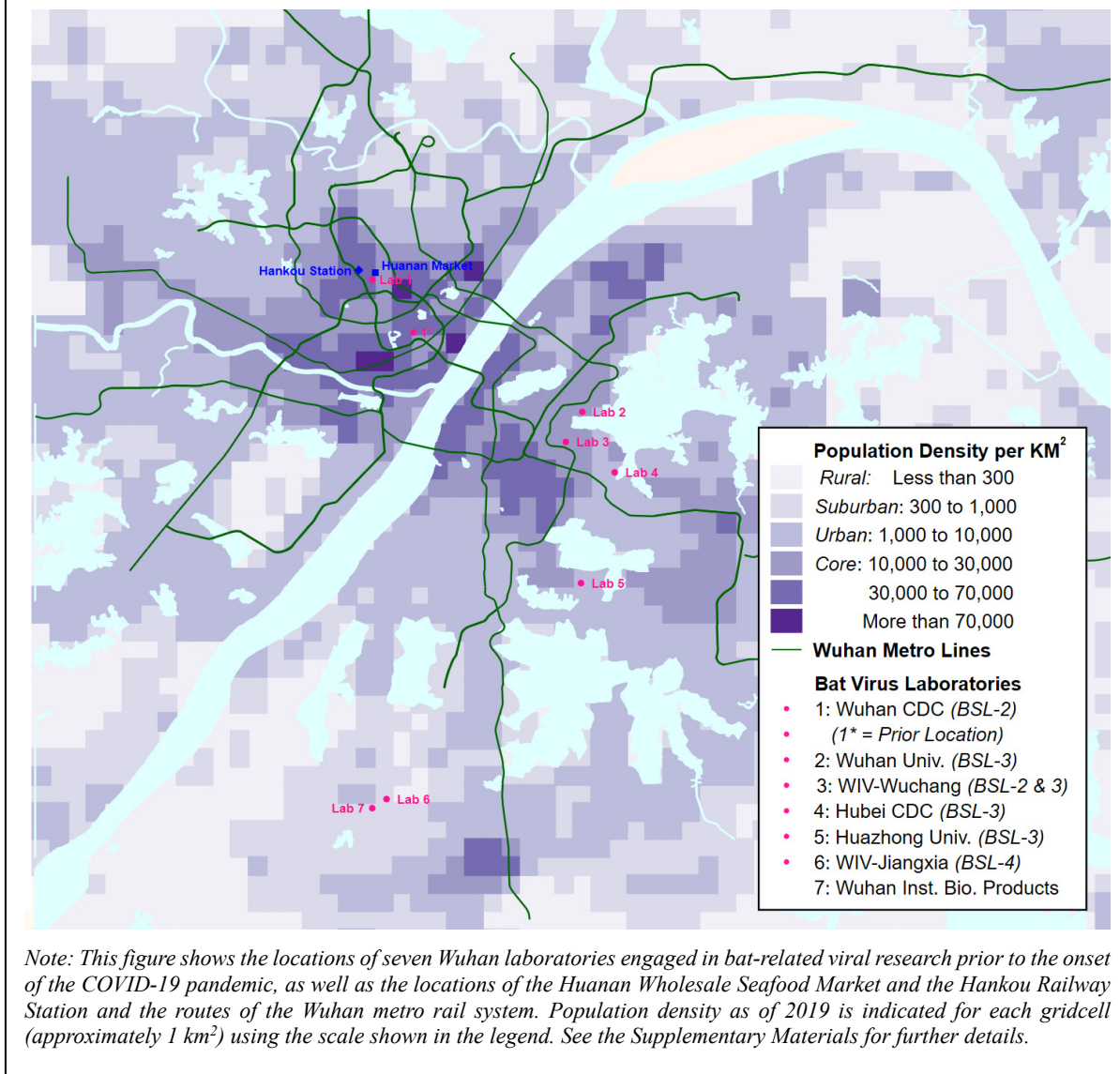
3.1 The Likelihood of a Bat-Related Viral Lab Accident in Wuhan

Wuhan is a city of 11 million people, located at the confluence of the Yangtze and Han rivers and serving as a focal point of commerce and transportation in central China. BBC News characterized Wuhan as “home to the world’s leading coronavirus research facility” and referred to WIV’s director of bat viral research as “China’s Batwoman.”²⁴³⁻²⁴⁶ Researchers at WIV played a leading role in collecting bat specimens from remote locations and performing pathbreaking genetic analysis of SARS-related viruses:

- During 2011-2015 WIV led expeditions to an abandoned mineshaft in Yunnan province (just a short distance from China’s southern border with Laos and Vietnam), where they collected fecal swabs from six species of bats and identified several novel SARS-like viruses.^{108, 130, 247-249}
- WIV subsequently conducted molecular analysis which revealed that these viruses were highly similar to the SARS-CoV virus, with nucleotide sequence identity of 95 to 96%.^{109, 250}
- WIV researchers used cutting-edge methods to sequence and modify viral genomes, produce chimeras and clones, and compare *in vitro* efficiency of viral variants using cell cultures.^{104, 109, 157, 159, 161, 206, 251-253}
- By 2018, WIV had upgraded its laboratory to “next-generation sequencing technology” that was used to perform more comprehensive sequencing of bat viruses.¹³²
- As of 2019, WIV’s internal database encompassed about 16,000 bat virus samples and genetic sequences, most of which were not accessible to researchers elsewhere.^{244, 246, 254, 255}

WIV’s work was generally performed in BSL-2 labs at its Wuchang site, although one study involving international coauthors was conducted using a BSL-3 lab.^{157, 159, 161, 243} WIV was chosen as the host for China’s first BSL-4 laboratory, which was launched in January 2018 at a new site in southern Wuhan, but that lab was used for analyzing hemorrhagic fever viruses and not for bat-related coronaviruses.¹⁶²

Figure 3: Bat-Related Coronavirus Research in Wuhan PRC



As shown in Figure 3, WIV's research activities appear to have generated agglomeration effects that facilitated a boom in such research at other nearby laboratories, including Wuhan University (a few blocks from WIV's Wuchang campus) and Huazhong Agricultural University (which is about 10km south).^{33, 205, 256-261} Experiments with primates and other live animals were conducted to assess the virulence of various viruses; such investigations are also subject to accidental risks.²⁶²⁻²⁶⁴

Research on bat viruses was also conducted at the Wuhan CDC, just across the river from WIV's Wuchang campus. In fall 2019, the Wuhan CDC moved about 3km to a location near Hankou Station, just a short walk (350m) from the Huanan Market. Coincidentally, a Wuhan CDC researcher whose work focused on identifying bat viruses was profiled in a CCTV video in early December 2019.²⁶⁵⁻²⁶⁸

Location	Number		Share of PRC Total (%)	
	Lead Authors	All PRC Authors	Lead Authors	All PRC Authors
Wuhan Institutions	18	163	22.0	19.6
<i>Wuhan Institute of Virology</i>	12	106	14.6	12.8
<i>Huazhong Agric. University</i>	6	53	7.3	6.4
<i>Wuhan University</i>	0	4	0	0.5
Other PRC Locations	64	667	78.0	81.4
Total	82	830	100.0	100.0

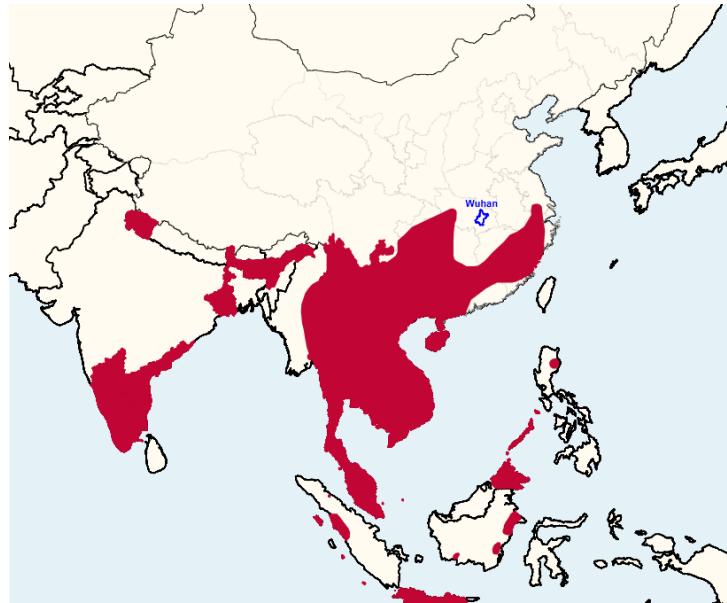
Note: This table reports on the affiliations of PRC authors of 82 bat-related coronavirus research papers that were posted in the PubMed database during 2015-19 and whose lead author was affiliated with a PRC institution. The first two columns indicate the number of lead authors and total authors from institutions located in Wuhan and from institutions elsewhere in PRC, and the final two columns show the corresponding shares (in %).

Although Wuhan served as a hub for the collection and analysis of bat viruses, such investigations were also being conducted at laboratories in Beijing, Hong Kong, Shanghai, and other PRC locations, using essentially the same biosafety standards as in Wuhan labs.^{102, 110, 269-277} As noted above in Section 2.1, PRC officials publicly expressed concerns about lab accidents in 2019, prior to the onset of the COVID-19 pandemic, and such concerns were not specifically directed towards any specific location.

Table 7 provides a tabulation of bat coronavirus research papers produced at PRC labs that were posted in PubMed during 2015-19, using the same search string as described above (*“bat” AND “coronavirus”*) OR *“alphacoronavirus”* OR *“betacoronavirus”*). This search identified 82 papers whose lead author was affiliated with a PRC institution, and those papers were coauthored by a total of 830 researchers with PRC affiliations. The lead authors of 18 papers (22%) were affiliated with research institutions in Wuhan. Among all PRC authors of those 82 papers, a total of 106 researchers (19.6%) had Wuhan affiliations. See the Supplementary Materials for further details.

Thus, under the premise of hypothesis *A* that the pandemic was triggered by an accident leak from a laboratory located in PRC, it is reasonable to assess a marginal likelihood of about 20% to the specific scenario in which the leak occurred in Wuhan.

Figure 4: The Distribution of Bats with SARS-CoV-2-Related Viruses



Note: This map shows the geographical range of bat species identified as hosts of viruses that are closely related to the SARS-CoV-2 virus; see Frutos et al. (2021) and Wu et al. (2023). The range of each bat species is shown using geospatial coordinates published by IUCN. See the Supplementary Materials for further details.

3.2 The Likelihood of a Zoonotic Spillover in Wuhan

Figure 4 shows the geographical distribution of bat species carrying viruses related to SARS-CoV-2. Systematic surveys of bats in central and southeast Asia have not found SARS-CoV-2 in any samples of bat excretions. Closely-related bat viruses have been identified in Laos and in two locations in Yunnan province, within 100km of the Laotian border and more than 1500km from Wuhan.^{233, 278}

At an early stage of the pandemic, bats were considered as a potential source of the virus, because prior studies had identified an elevated risk of zoonotic spillover in locations with high bat-human interactions.²⁵⁰ However, a direct zoonotic pathway from bats to humans has subsequently been ruled out, because molecular analysis and cell culture studies have found that the RBD protein of the SARS-CoV-2 virus has negligible efficiency in binding to the ACE-2 receptors of bat orders that are commonly found in PRC.^{178, 279-282} Phylogenetic analysis has also concluded that bats could not have been the direct zoonotic source of SARS-CoV-2.²⁸³

Thus, the zoonotic spillover of SARS-CoV-2 must have involved one or more intermediate species that were infected by an ancestor of the virus (presumably through exposure to live virus in bat excretions) and then served as host reservoirs in which it evolved to become highly transmissible to humans. However, given the specificity of the binding mechanism of SARS-CoV-2, a wide array of animals can be ruled out as intermediate hosts in its zoonotic transmission.

The RBD protein of SARS-CoV-2 cannot bind to the ACE-2 receptors of fish, amphibians, reptiles, or birds.¹⁷⁸ Moreover, common domesticated livestock – including cattle, goats, sheep, rabbits, and swine – can be infected with SARS-CoV-2 by intranasal inoculations, but none of those animals shed any infectious virus via nasal, oral, or faecal excretions.^{280, 284-286} A subsequent study of rabbits confirmed

those findings but found some nasal viral shedding at very high doses of intranasal inoculation.²⁸⁷ Some common household pets (notably, cats, ferrets, and hamsters) can be infected and shed live virus.^{280, 288-296} However, domestic pets have minimal exposure to bat excretions and hence very unlikely to have been intermediate hosts prior to its transmission to humans.

Evidently, some species of wildlife mammals—either free-roaming or farmed—must have served as the intermediate host(s) for the evolving virus. Moreover, the zoonotic spillover of SARS-CoV-2 to humans almost surely involved interactions with live mammals, because the virus generally spreads through airborne transmission, not via direct contact with fresh or frozen meat or other packaged foods.²⁹⁷

In considering the zoonotic hypothesis, it should be noted that exotic animals were not a common feature of Wuhan's cuisine, even prior to the pandemic.²⁹⁸ A systematic survey of wildlife sales at Wuhan markets in 2017-19 found that sales of live wildlife mammals for human consumption averaged about 150 animals per week in a megacity of more than 10 million residents.²⁷ Evidently, the number of wildlife mammals being sold in Wuhan prior to the pandemic was very far below the incidence of sales observed in the southern provinces of PRC prior to the outbreak of the SARS virus.

In assessing the likelihood that the zoonotic spillover occurred in Wuhan, it is reasonable to focus on the Huanan Wholesale Seafood Market. The systematic survey of Wuhan markets identified seven Huanan Market shops selling wildlife mammals as well as six such shops at two other wholesale markets (Qiyimen Zhengshian and Baishazhou, both located on the other bank of the Yangtze River).²⁷ A substantial fraction of early confirmed COVID-19 cases were linked to the Huanan Market, whereas the Wuhan CDC found no systematic links to any other Wuhan markets.²⁰

At the onset of the pandemic, pangolins—an order of scaly anteaters—were suspected as a potential source of zoonotic transmission.^{299, 300} However, pangolins are an endangered species in PRC, and no pangolins were observed for sale at Wuhan markets.^{27, 301} Palm civets, which were the primary zoonotic link for the SARS virus, were sold occasionally at Wuhan markets (with average sales of about two civets per week) but were not observed at the Huanan Market in October or November 2019.^{21, 27} Farmed mink have been found to be highly susceptible to SARS-CoV-2 infection, with viral shedding and airborne transmission to other mink as well as humans.^{302, 303} However, no mink were observed at the Huanan Market in fall 2019, and typical sales elsewhere in Wuhan were merely 2 to 3 mink per week.^{21, 27} Likewise, nutria, weasels, and wild boar were sold occasionally at other Wuhan markets (with total weekly sales of one nutria, two weasels, and one wild boar), but those three genera were not observed at the Huanan Market during fall 2019.^{21, 27} Various types of squirrels were sold at Wuhan markets, but only as pets rather than for human consumption.²⁷

Table 8: Wildlife Mammals Sold at Wuhan Markets in November 2019

Name	Genus	Susceptibility to SARS-CoV-2 Infection	Average Daily Sales	Significant DNA Readings at Huanan Market
Amur Hedgehog	Erinaceus	1 (very low)	11.1	no
Badger	Meles	1 (very low)	0..4	no
Reeves Muntjac	Muntiacus	1 (very low)	0.3	no
Porcupine	Hystrix	2 (low)	0.3	no
Red Fox	Vulpes	2 (low)	1.0	no
Chinese Hare	Lepus	3 (medium)	5.6	no
Bamboo Rat	Rhizomys	3 (medium)	1.4	yes
Himalayan Marmot	Marmota	3 (medium)	0.5	yes
Raccoon Dog	Nyctereutes	N/A	1.3	yes

Note: This table provides information about nine genera of wildlife mammals that were observed for sale at the Huanan Market in November 2019.^{21, 27} The first and second columns indicate the common name and genus. The third column reports the findings of Damas et al. (2021) regarding the binding efficiency of SARS-CoV-2 RBD to the ACE-2 receptor proteins of the specified mammal; that analysis did not include raccoon dogs. The fourth column indicates average daily sales at all Wuhan markets, computed using data from the survey of Xiao et al. (2021). The final column characterizes the genetic analysis of swab samples collected at the Huanan Market after its closing and indicates whether mitochondrial DNA for that genus exceeded the threshold of 1200 matching reads; see Table 13 and the Supplementary Materials for details.

Table 8 provides a synopsis of the nine genera of wildlife mammals that were sold at Wuhan markets in late 2019: badgers, bamboo rats, foxes, hares, hedgehogs, marmots, muntjacs, porcupines, and raccoon dogs. This tabulation reflects the findings of the systematic survey of wildlife sales at Wuhan markets as well as results from genetic analysis of swab samples collected in early 2020, shortly after the closure of the Huanan Market.^{21, 27, 28}

In contrast to palm civets, none of these mammal genera would have any natural interactions with bats or exposure to bat faeces. Masked palm civets are arboreal, living in family groups of 2-10 animals and sleeping during the day in tree dens.³⁰⁴ By contrast, all of the mammals listed in this table are ground-dwelling and solitary, and nearly all (except the Reeves muntjac) live in small burrows.³⁰⁵ For example, bamboo rats live alone in underground burrows except during breeding seasons when each adult male relocates to the burrow of its female mate.³⁰⁶

Six of these mammals have been found to have little or no susceptibility to SARS-CoV-2 infection, because the virus has negligible ability to bind to the ACE-2 receptor in those genera.¹⁷⁸ Three other mammal are moderately susceptible to SARS-CoV-2 infection but were unlikely to have been the intermediate host for the SARS-CoV-2 virus:

- *Chinese hares* were sold in small numbers at Wuhan markets (about 6-7 per day), but genetic analysis of swab samples from the Huanan Market indicates that very few if any live hares were sold there. These data are consistent with standard practices at hare and rabbit farms in PRC, which generally produced fresh or frozen meat rather than selling live animals.³⁰⁷

Figure 5: Characteristics of Farmed vs. Wild Raccoon Dogs

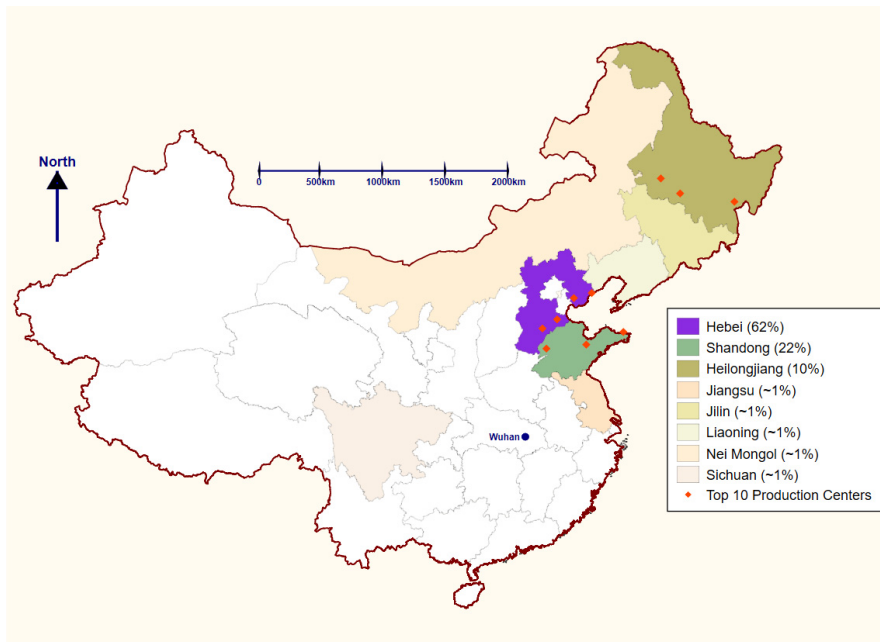


Note: The top row shows a raccoon dog fur farm in northeastern China, a farmed animal, and pelts prepared for sale. The bottom row shows a wild raccoon dog, a pair of mates foraging for food, and a young cub; these photos were taken by Midori Saeki in Ibaraki, Japan. Sources: Human Society International, Kauhala (2009).

- *Himalayan marmot*. Only a few marmots were observed for sale at Wuhan markets.²⁷ There is no commercial farming of this species, and wild populations are only found in Tibet, about 1000km from Wuhan.^{308, 309}
- *Bamboo rats*. An average of 10 bamboo rats per week were sold at Wuhan markets prior to the onset of the COVID-19 pandemic.²⁷ By comparison, PRC authorities reported that about 25 million bamboo rats were being farmed for human consumption, of which 18 million were in Guangxi and the remainder concentrated in Guangdong and other southern provinces.³¹⁰ Meanwhile, farms in rural areas of Hubei province were raising hundreds of thousands of wildlife animals, including bamboo rats.³¹¹ Evidently, nearly all of those farmed animals were being shipped to southern China. Thus, if bamboo rats had been the source of zoonotic spillover, then the earliest cases of SARS-CoV-2 would have been most likely to be observed in Guangxi or Guangdong, just as with the initial outbreaks of the SARS-CoV virus.

Thus, among the genera listed in Table 8, the raccoon dog has been a focal point for analyzing the zoonotic transmission of the SARS-CoV-2 virus.^{21, 312} Raccoon dogs are in the family of canids (which includes domestic dogs, coyotes, foxes, and wolves), with facial features similar to a raccoon and soft, thick fur similar to that of a mink.³¹³⁻³¹⁶ This species has been found to be susceptible to SARS-CoV-2 infection, although recent analysis indicates that the RBD of the virus does not bind very efficiently to the ACE-2 receptor of raccoon dogs.³¹⁷ In assessing scenarios in which this species was an intermediate or transitory host for SARS-CoV-2, it is essential to distinguish the relative likelihood of viral transmission in farmed vs. wild raccoon dogs.

Figure 6: Raccoon Dog Fur Farms in PRC as of 2019



Note: This figure shows each province's production of raccoon dog fur pelts in 2019 as a share of total PRC production, using data published by the China Leather Industry Association. As of 2019, 5 of the top 10 raccoon dog fur production centers were located in Hebei (Cangzhou, Hengshui, Qinhuangdao, Tangshan, Weifang), with the remainder in Shandong (Liaocheng, Weihai) and Heilongjiang (Daqing, Harbin, Jixi). See the Supplementary Materials for further details.

Farmed Raccoon Dogs. As of 2019, fur farms in PRC were raising about 14 million raccoon dogs for domestic sales and global exports, roughly similar to their output in 2010.³¹⁸⁻³²⁰ However, the likelihood that farmed raccoon dogs were the intermediate reservoir for SARS-CoV-2 is negligible due to a number of distinct factors:

- As shown in Figure 6, farming of raccoon dogs is concentrated in the northeastern provinces of PRC, with cooler temperatures suitable for animals with dense fur. Statistics published by the China Leather Industry Association indicate that about 90% of raccoon dog fur is produced on farms in Hebei and Shandong that are more than 1000km from Wuhan.^{318, 321}
- No bats with COVID-related viruses have been identified in the northern regions of PRC where fur farms are located, as shown above in Figure 4.^{233, 241, 242}
- A virological survey of non-livestock mammals in PRC collected swabs and tissue samples during late 2020 and 2021, including 95 farmed raccoon dogs from Hebei and Shandong, but found no betacoronaviruses in those samples.³²² This survey did find a number of other common viruses, including Alphacoronavirus 1—a virus that causes mild digestive symptoms in canine, feline, and porcine animals.^{323, 324}
- Another virome survey collected samples in 2023 from a variety of fur animals in PRC, including 75 farmed raccoon dogs from six provinces (Hebei, Heilongjiang, Henan, Jiangsu, Liaoning, Shandong) and 5 wild raccoon dogs from Inner Mongolia.³²⁵ This survey identified one sample containing canine respiratory virus—a betacoronavirus that is very common among dogs in kennels and other large facilities.^{326, 327} The survey also identified a number of samples with

Alphacoronavirus I as well as other common mammalian viruses associated with gastroenteritis, and the virus that causes Hepatitis E was identified in two samples (one of which was from a wild raccoon dog).³²⁵

- In contrast, farmed raccoon dogs are slaughtered and skinned onsite at each farm, which then ships the fur pelts to commercial centers, thereby eliminating the risk of disease or injury that would be associated with the transit of live animals.^{328, 329}
- The survey of live animal sales in Wuhan found that the raccoon dogs were wild animals that had been trapped, based on vendor interviews and direct observations of leg wounds.²⁷ Even if a sick raccoon dog had escaped from a fur farm somewhere in Hebei Province, it seems highly improbable that such an animal would roam far enough to be caught by a trapper in Hubei Province and then sold to a wildlife vendor at Huanan Market.

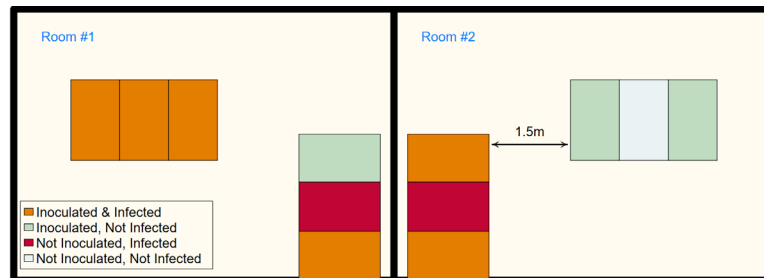
Wild Raccoon Dogs. The ecology of wild raccoon dogs is notably similar to that of badgers and true raccoons.³³⁰ Specifically, raccoon dogs are nocturnal, omnivorous, and solitary, mostly dwelling in forests with thick undercover and using fallen trees or rock crevices as dens, but may also be found scavenging for food in urban and suburban areas.³³¹⁻³³⁴ Wild raccoon dogs are strictly monogamous, and mated pairs often forage for food together, as shown in Figure 5.^{335, 336} Cubs are born once a year, and the male stays inside the den with the cubs while the female goes outside to forage, thereby facilitating the production of milk for the litter.^{313, 337}

The population density ranges from 0.8 to 1.5 km² for a pair of adult raccoon dogs, while juveniles often wander 10 to 20 km to find a new home.³³⁷⁻³³⁹ Raccoon dogs are unique among canids in hibernating during the winter season in regions with temperatures below freezing.³⁴⁰ The range of wild raccoon dogs in mainland China and Hong Kong has been carefully monitored over the past century, dating back to the pioneering work of Sowerby (1923) and Allen (1938).^{314, 341-345} See the Supplementary Materials for further details.

As solitary and monogamous animals dwelling in burrows, wild raccoon dogs do not have any interactions with bats, and hence it is very unlikely that raccoon dogs could have been the intermediate host for the evolving SARS-CoV-2 virus. Even if a wild raccoon dog did become infected with a bat virus (perhaps through exposure to bat feces), that infection would spread to its mate and young offspring but would be unlikely to infect any other raccoon dogs. By contrast, the masked palm civet—which was identified as the principal intermediary for the SARS virus—lives in trees and interacts with numerous potential mates during each semiannual breeding season.³⁴⁶

Moreover, wild raccoon dogs could not have been an intermediate host that magnified the virulence of SARS-CoV-2 prior to its spread to the human population. A laboratory study assessed the susceptibility of raccoon dogs to SARS-CoV-2 and found that the virus was transmitted via direct contact but not via airborne particles.^{347, 348} By contrast, a similar experimental design was used to study ferrets and found airborne transmission of SARS-CoV-2 over distances exceeding 1m.²⁹³

Figure 7: SARS-CoV-2 Susceptibility and Transmission in Raccoon Dogs



Note: This figure shows the experimental design of Freuling et al. (2021), who conducted a controlled experiment to assess raccoon dogs' susceptibility to SARS-CoV-2 infection. Each lab room (size: 12 x 7 m) contained two sets of three adjacent cages. Nine raccoon dogs were inoculated nasally with live virus, and three other raccoon dogs served as "sentinels" to assess the extent of transmission. Six of the inoculated animals and two of the sentinels developed viral infections in nasal passages, whereas three of the inoculated animals and one sentinel did not develop infections. See the Supplementary Materials for further details.

The design of this experiment is shown in Figure 7. Nine raccoon dogs were given nasal inoculations of live virus and placed in individual cages in two lab rooms. To assess transmission by direct contact, three "sentinel" raccoon dogs were placed in adjacent cages on the following day. Six of the nine inoculated animals developed SARS-CoV-2 infections, along with two of the sentinels. The distance between the third sentinel and the infected animals was only 1.5 to 3m, but that sentinel did not become infected. Viral antigen was detected within nasal passages but not in respiratory or rectal samples from the infected animals. Thus, the researchers concluded that "*raccoon dogs are susceptible to SARS-CoV-2 infection and can transmit the virus to direct in-contact animals.*"³⁴⁷

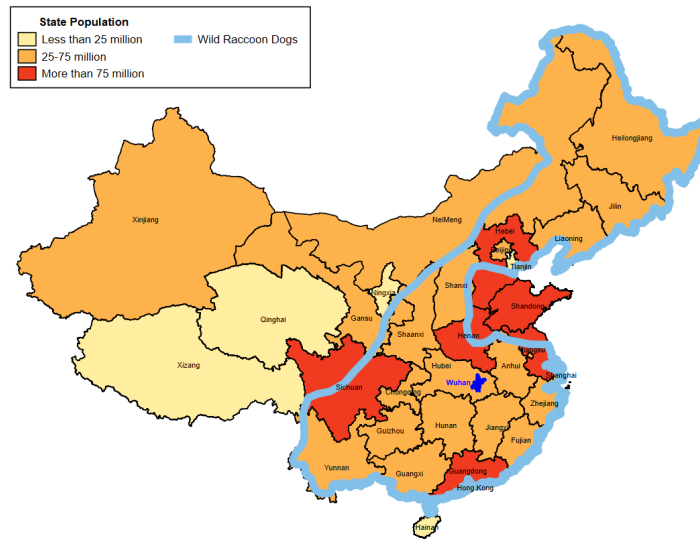
As shown in Figure 8, wild raccoon dogs are present in nearly all of the densely-inhabited regions of PRC with the exception of Shandong and a few other areas lacking sufficient forest cover.³⁴⁹ Thus, even if a wild raccoon dog residing in PRC became infected with the SARS-CoV-2 virus, there is no particular reason to expect that Wuhan would be the location of the zoonotic spillover to humans. In fact, the population of Wuhan comprises only 1.1% of the total PRC population residing in areas with wild raccoon dogs; see the Supplementary Materials for further details.

During the second and third weeks of January 2020, the China CDC initiated surveillance of wild mammals in the vicinity of Wuhan, including 15 wild raccoon dogs captured by traders who were suppliers to Huanan Market shops.³⁵⁰ Strains of alphacoronavirus were identified in 8 raccoon dogs, but none of the raccoon dogs were infected with SARS-CoV-2 or any other betacoronavirus. Evidently, SARS-CoV-2 was not endemic among wild raccoon dogs in the Wuhan area.

Moreover, it seems improbable that a raccoon dog from a different locale could have been infected with SARS-CoV-2 and then roamed to Wuhan prior to its capture. Bats carrying viruses related to SARS-CoV-2 have only been identified at a single location in PRC, near the southwestern edge of Yunnan province. Even if a wild raccoon dog in that area were infected with the virus, there would be no mode of transmission to wild raccoon dogs near Wuhan—a distance exceeding 1600 kilometers.

Consequently, it seems very unlikely that wild raccoon dogs were the intermediate host for the SARS-CoV-2 virus. The European CDC reached a very similar conclusion in its assessment of the risk of zoonotic spillover in western Europe from wild mustelids such as badgers and wild mink: "...due to

Figure 8: Range of Wild Raccoon Dogs in PRC



Note: This map shows the geographical range of wild raccoon dogs in PRC, using geographical coordinates produced by Kauhala and Saeki (2008) as published in the IUCN RedList. The population of each PRC province as of 2020 is taken from CIESIN. See the Supplementary Materials for further details.

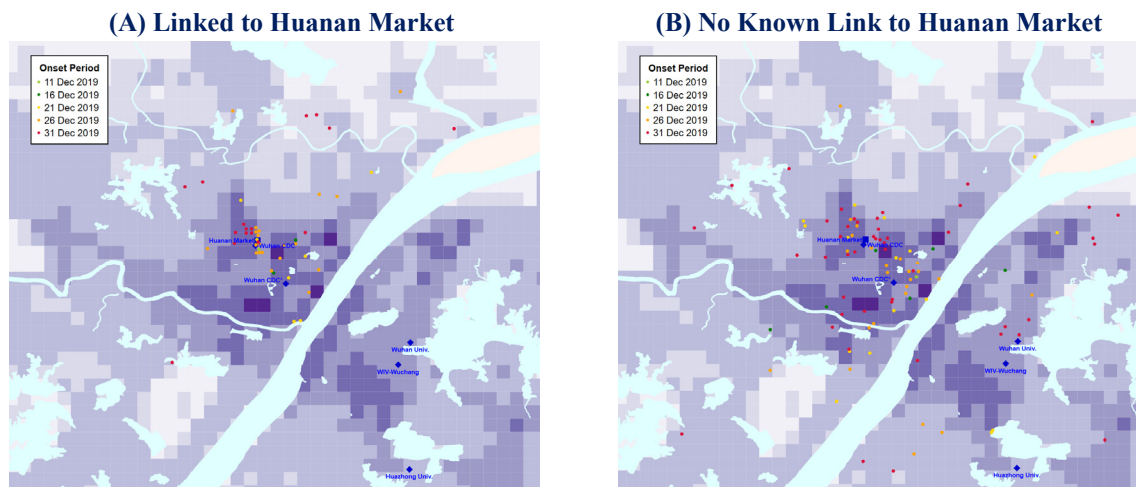
their elusive and solitary behaviour...there is a very low chance of contact with humans and/or other SARS-CoV-2 susceptible domestic animals. Therefore, the risk of wild mustelids becoming a reservoir for SARS-CoV-2...is very low."³⁵¹

A more plausible scenario is that wild raccoon dogs were merely a transient host of SARS-CoV-2. In that scenario, the virus evolved in the process of infecting a variety of wild and farmed mammals, with cross-species transmission occurring over an extended period prior to the zoonotic spillover to the human population in late 2019. However, there is no apparent reason why this scenario would have occurred in Wuhan rather than some other location in PRC. Consequently, this marginal likelihood may reasonably be judged at around 1%, that is, the ratio of Wuhan's population relative to the total population of PRC.

3.3 Bayesian Implications

In light of these findings, the value of BF_2 can now be computed using the marginal likelihood of 20% under hypothesis A and the marginal likelihood of less than 1% under hypothesis Z . Thus, the conditional odds ratio for hypothesis A vs. hypothesis Z is at least 20 or more. Referring back to the qualitative interpretations in Table 2, this conditional odds ratio is equivalent to "extremely strong" evidence in favor of hypothesis A . In effect, addressing the question of why the COVID-19 pandemic started in Wuhan provides very significant evidence regarding the origins of SARS-CoV-2. However, no conclusion should be reached without considering the spatiotemporal pattern of early cases.

Figure 9: COVID-19 Cases in Wuhan's Urban Core during December 2019



Note: This figure shows the residence locations of 145 confirmed COVID-19 cases in the urban core of Wuhan with symptom onset in December 2019. Panel A shows 47 cases that were identified as having links to the Huanan Market, and panel B shows the other 98 cases with no known link to this market. The 5-day onset period of each case is denoted using the symbols shown in the legend. Population density in 2019 for each gridcell (approximately 1km²) is shown using the same scale as in Figure 3. See the Supplementary Materials for further details.

4. Spatiotemporal Pattern of Early Cases in Wuhan

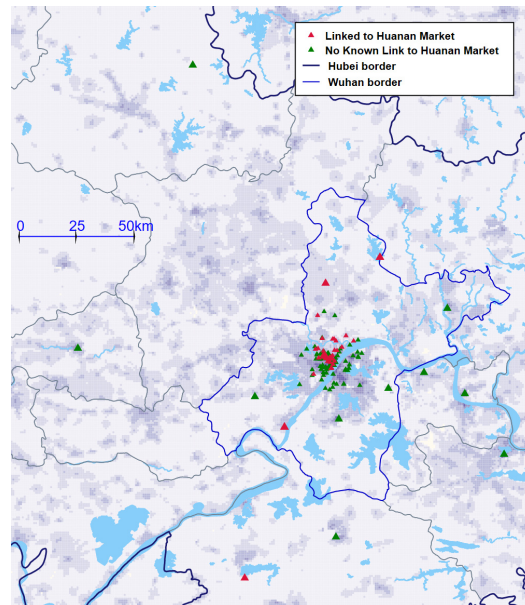
This section focuses on assessing the conditional Bayes factor BF_4 , which incorporates information from the timing and place of residence of confirmed COVID-19 cases in Wuhan with symptom onset by 31 December 2019. This Bayes factor is conditional on the pandemic outbreak having taken place in Wuhan, and hence it indicates the conditional odds of observing the spatiotemporal pattern of these early COVID-19 cases under hypothesis A (accidental lab leak) compared with hypothesis Z (zoological spillover). Specifically, this conditional Bayes factor can be expressed as follows:

$$(4) \quad BF_4 = \frac{\text{Marginal Likelihood}(\text{Pattern of Early Cases} | \text{hypothesis } A; \text{outbreak in Wuhan})}{\text{Marginal Likelihood}(\text{Pattern of Early Cases} | \text{hypothesis } Z; \text{outbreak in Wuhan})}$$

Figure 9 shows the residence locations of the confirmed cases that were identified as having links to the Huanan Market (left panel) and the cases that had no known link to this market (right panel). These data are taken from the Annex of the WHO 2021 Report, which showed the spatial distribution of cases using 5-day intervals for timing of symptom onset.²⁰ This figure also shows the population density of each 1km² gridcell using the same scale as in Figure 3. In effect, this figure encapsulates practically all of the information needed to conduct spatiotemporal analysis of early cases in Wuhan.

It should be noted that previous studies, starting with the pathbreaking work of Worobey et al. (2022), have focused on gauging the centroid and radial dispersion of the cumulative spatial distribution of these cases.^{21, 352-356} Such an approach can be effective in investigating a persistent source of toxicity at a fixed location, as demonstrated by the pioneering work of John Snow (1866), who mapped the spatial pattern of infections during a cholera outbreak in London and succeeded in identifying the site of the contaminated water pump.^{357, 358} Nonetheless, in analyzing the outbreak of an infectious disease, such an approach may neglect potentially significant information about the spatiotemporal characteristics of each wave of new cases.³⁵⁹

Figure 10: COVID-19 Cases in Hubei Province during December 2019



Note: This figure shows the residence location of 165 confirmed COVID-19 cases in Hubei PRC with symptom onset in December 2019 and information about exposure to Huanan Market. Population density in 2019 is shown using the same scale as in Figure 3. See the Supplementary Materials for further details.

Moreover, the statistical analysis of an infectious disease outbreak is intrinsically probabilistic, reflecting various sources of uncertainty about the location and timing of specific cases.^{360, 361} These considerations warrant the use of a spatial grid that facilitates the use of cell-specific indicators, such as population density, as well as spatiotemporal data such as the distance of each gridcell from the nearest cases that occurred in preceding periods. Consequently, both positive and negative gridcells may provide significant information about the characteristics of the outbreak, analogous to the classic case of “the dog that didn’t bark in the night.”³⁶²

A very striking pattern in Figure 9 is that all of the cases linked to Huanan Market were residents on the northwest bank of the Yangtze River (where that market is located), whereas a substantial number of cases with no known link to Huanan Market were residents on the southeast bank, in the general vicinity of laboratories engaged in bat viral research. Of course, formal statistical analysis of the spatiotemporal data is required to determine whether that pattern is significant or merely coincidental.

Thus, the remainder of this section is organized as follows: Section 4.1 examines the information content and limitations of the data on early COVID-19 cases. Section 4.2 describes the methodology for analyzing the spatiotemporal pattern of cases to assess the relatively likelihood of the two hypotheses A and Z . Section 4.3 presents the results of this analysis and the range of estimates for the Bayes factor, BF_4 .

4.1 Characteristics of the COVID-19 Case Data

Identification of Cases. The WHO 2021 Report provides a comprehensive review of the procedures used to identify early COVID-19 cases in Wuhan. The primary source of case data was the PRC’s National Notifiable Disease Reporting System (NNDRS), established in 2004 following the SARS pandemic. The Wuhan CDC searched the NNDRS database and identified 174 COVID-19 cases with symptom onset in 2019, including 100 laboratory-confirmed cases and 74 clinically diagnosed cases. Wuhan hospitals and outpatient clinics conducted a separate search to identify records of patients with COVID-like symptoms, and those records were reviewed by a panel of clinical experts who did not find any additional COVID-19 cases apart from the NNDRS cases identified by the Wuhan CDC.

Underascertainment of Infections. A large proportion of COVID-19 infections are asymptomatic or cause only mild illness.^{363, 364} However, all of the confirmed cases with onset in 2019 were hospitalized patients: Clinical confirmation mainly relied on lung x-rays showing “ground-glass” opacity (a characteristic of severe cases), and laboratory confirmation generally relied on blood samples because RT-PCR tests for SARS-CoV-2 did not become available until January 2020.³⁶⁵ Consequently, in characterizing those confirmed cases, the WHO 2021 Report noted that “*it is likely that others were infected at the time of recognition of early cases.*”²⁰

The likely rate of underascertainment can be gauged using the infection-hospitalization ratio (IHR), that is, the proportion of individuals infected with the original strain of SARS-CoV-2 whose severity required hospitalization.^{366, 367} Serological surveys of SARS-CoV-2 antibodies and hospitalization records indicate that the IHR is strongly linked to age, with a rate of 0.5% for adults ages 20-29 and rising by a factor of 2x with each additional 10 years of age.³⁶⁸ For example, the IHR is about 4.5% for an individual of age 56 (the median for confirmed cases with onset in December 2019), and hence the ratio of infections to hospitalizations at that age would be about 20:1.³⁶⁵ If infections were uniformly distributed across age groups, then about 7500 Wuhan residents were infected with SARS-CoV-2 by the end of December 2019, and the average underascertainment rate was about 44:1; see the Supplementary Materials for further details.

Links to Huanan Market. The Wuhan CDC conducted an in-depth epidemiological review of all confirmed COVID-19 cases, including age, sex, place of residence, occupation, travel history, contact with animals, social contacts, visits to Wuhan markets, and other exposure history. Among all 174 confirmed cases in Wuhan and its vicinity in Hubei Province, the review identified 55 cases with links to the Huanan Market, including 30 vendors, 12 wholesale purchasers from local hotels and restaurants, 3 retail purchasers, 2 couriers, 2 individuals who did not visit this market but had direct contact with someone who had exposure there, and 6 individuals who passed through this market without making any purchases.²⁰ Epidemiological studies have indicated that transmission risks for the original strain of SARS-CoV-2 were highest in instances of sustained contact (more than 15 minutes) with an infected person in an indoor space.³⁶⁹⁻³⁷⁶ It should also be noted that the information in the WHO Report was inconsistent about whether one specific individual with symptom onset on 20 December 2019 was linked to the Huanan Market (as in one figure in the main report) or unlinked (as in two figures in the annex of the report); see the Supplementary Materials for further details.

Cluster Cases. The Wuhan CDC’s epidemiological review identified seven “clusters” of cases where two or three COVID-19 patients had been in close contact with each other at home or work. As reported in the Annex to the WHO Report, three clusters were linked to Huanan Market: three

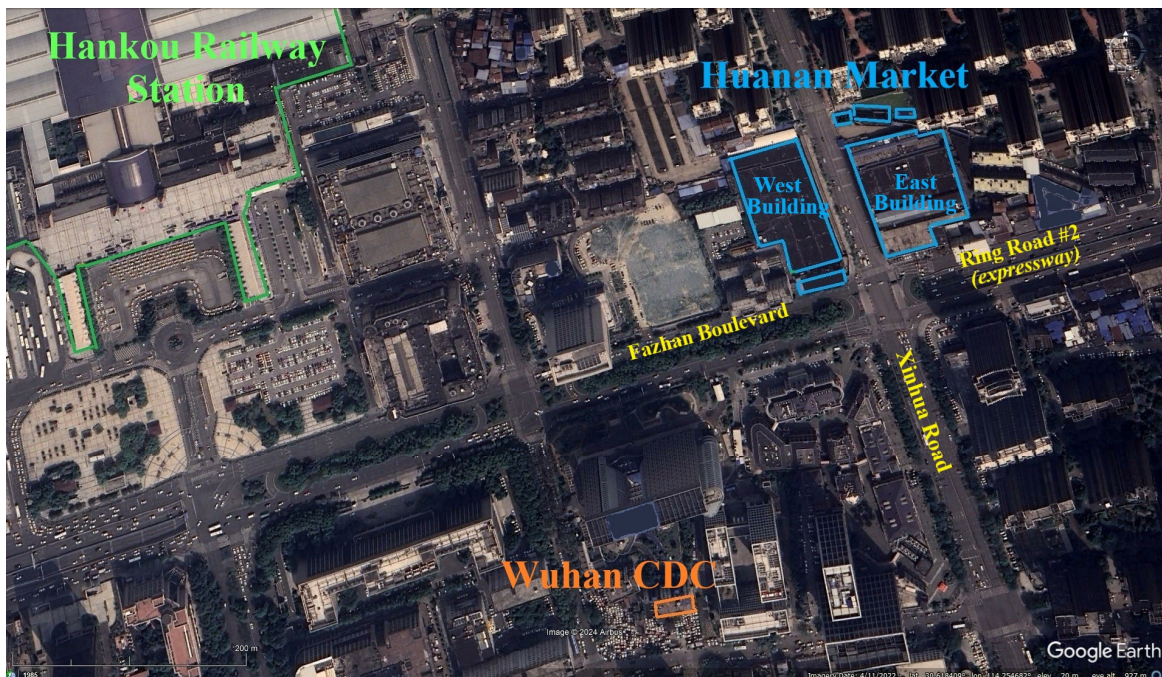
individuals who worked together at a frozen food shop, two who worked together at a seafood shop, and one wholesale distributor who frequently stocked up at Huanan Market and whose spouse also became ill. One of those individuals had symptom onset during the third week of December 2019, and the other six developed symptoms during the fourth week of that month.²⁰

Residential Locations of Cases. The WHO Report Annex included maps showing the residence locations for 165 confirmed cases with market exposure information, including the 55 cases linked to Huanan Market and 110 other cases with no known link to Huanan Market.²⁰ The geocoordinates of those home addresses were extracted and analyzed by Worobey et al. (2022).²¹ An independent digital extraction has been performed for the present study; those locations are shown in Figure 9 for the core Wuhan area and in Figure 10 for the relevant portion of Hubei Province. This digital extraction confirms that these geocoordinates can be consistently determined within a range of about 150- 200m, that is, about 0.001 degrees or 3 arc-seconds at the latitude of Wuhan; further details are provided in the Supplementary Materials. Thus, the residential locations of these cases can be compared to the population density of the corresponding gridcells with edges of 30 arc seconds (approximately 1km² at the equator and at the latitude of PRC); that density data was produced by WorldPop/CIESIN using United Nations population estimates for 2019.³⁷⁷

Timing of Symptom Onset. The Annex to the WHO 2021 Report provides data on the daily incidence of new cases based on self-reported onset of COVID-19 symptoms.^{20, 378} Consequently, the precise timing of individual cases is subject to several distinct sources of uncertainty. First, the time between infection and symptom onset can be highly variable, with a mean lag of 5 days and a 95% confidence interval of 2 to 14 days.^{365, 379} Second, the symptoms of COVID-19 infection may vary markedly across individuals; for example, 78% develop a fever and 58% have a dry cough.³⁸⁰ However, those symptoms are also associated with other seasonal respiratory illnesses that would also be circulating in Wuhan in late 2019. Third, the timing of symptom onset was self-reported and hence could be influenced by the typical interval of about 6 days between symptom onset and initial medical consultation for early COVID-19 cases.^{17, 365} These factors likely accounted for some differences in the timing of cases as published in initial reports compared to conclusions reached later by the Wuhan CDC's epidemiological review.^{17-21, 381, 382} The WHO 2021 Report also downplayed differences in case timing of one or two days as "*within a common incubation period.*"²⁰ The Supplementary Materials provide further details.

Spatiotemporal Pattern of Cases. The WHO Report Annex included a set of maps showing the residential locations of confirmed cases based on timing of symptom onset over 5-day intervals from 11 December to 31 December 2019.²⁰ The information from those maps has been digitally extracted so that the geocoordinates of each confirmed case can be matched precisely to a specific 5-day period of symptom onset; see the Supplementary Materials for further details.

Figure 11: The Location of the Huanan Market



Note: This figure is a satellite photo of the central core of Wuhan PRC, showing the location of the Huanan Wholesale Seafood Market, the Hankou Railway Station, and the Wuhan CDC, which moved to that location during fall 2019. Credit: © 2024 Airbus.

Potential Sampling Bias. Residents of neighborhoods near Huanan Market comprised a substantial fraction of the cases with no known link to that market. That pattern has raised concerns that sampling bias in the case data could have arisen from greater scrutiny of respiratory patients at hospitals near Huanan Market compared with hospitals elsewhere in Wuhan.^{46, 353, 382} However, several specific factors are relevant in mitigating such concerns:

- As shown in Figure 11, the Huanan Market is located in the central area of Wuhan’s urban core, about 400m from Hankou Railway Station (Wuhan’s hub for intercity train service) and about 800m from Jinjiadun Bus Terminal (the hub for intercity bus service).³⁸³ The Wuhan Museum and the 1911 Memorial Park are located just a few city blocks away. As of fall 2019, the Wuhan CDC was located within a few hundred meters of Huanan Market.
- It is clear that some Huanan Market vendors lived nearby the market.³⁸⁴ A vendor with an asymptomatic or mild infection may have unknowingly transmitted the disease to other neighborhood residents, and such cases could not be linked to Huanan Market because comprehensive contact tracing had not yet been initiated in December 2019.
- Similarly, Huanan’s wholesale products were mostly sold to restaurants near the market.³⁸⁴ Thus, a wholesale purchaser could have unknowingly transmitted the virus to the restaurant’s workers or customers, some of whom might well be residents of that neighborhood.
- Figures 9 and 10 confirm that the geographical scope of the Wuhan CDC’s epidemiological survey encompassed a wide area rather than being limited to the vicinity of the Huanan Market.

As noted above, a substantial fraction of unlinked cases were identified on the opposite bank of the Yangtze River. Among the seven cases located in outlying areas of Wuhan, three were linked to the Huanan Market and four had no known link to the market. Another seven cases were identified in other locations outside of Wuhan, but only one was linked to Huanan Market.

Characteristics of Earliest Cases. All six of the earliest confirmed cases (symptom onset during the week of 8-14 December 2019) were laboratory confirmed. The Wuhan CDC's review identified three cases with links to the Huanan Market (one shrimp vendor, one wholesale seafood buyer, and one visitor who passed through the market); there was no known link to Huanan Market for the other three cases (an office clerk, a commercial service worker, and a retiree). Given the range of ages of these cases (nearly all of whom were ages 40+), it is very likely that the true number of SARS-CoV-2 infections was substantially higher at that time.^{20, 378, 385}

Dating the Start of the Pandemic. Epidemiology studies have estimated that the number of infections was doubling in size every 3 to 7 days during the early stages of the pandemic, with estimates of an effective reproduction rate of 6 or more prior to any containment measures.^{365, 381, 386-388} Thus, the virus was almost surely spreading in Wuhan by late November and early December, prior to the onset dates for any of the confirmed cases, and the first SARS-CoV-2 infection might well have occurred in mid-to-late November 2019.^{385, 389} Moreover, the WHO panel conducted a comprehensive review of disease and mortality data, including cases of severe acute respiratory illnesses, and found no indications that SARS-CoV-2 had been spreading widely in Wuhan or other parts of Hubei province prior to December 2019.^{20, 390}

Similar findings were obtained by a serology study that analyzed serum samples from Wuhan blood donors collected between 01 September and 31 December 2019; that study found that about 0.5% of samples were reactive to SARS-CoV-2 antigens, as one would expect from an assay with imperfect specificity; however, there were zero positive results from a microneutralization assay that provides a gold standard for detecting SARS-CoV-2 antibodies.³⁹¹ Thus, the authors concluded that there was “no evidence of transmission of COVID-19 before December 2019 in Wuhan, China.”³⁹¹ Likewise, some serology studies conducted in western Europe found that a small proportion of pre-2020 samples were reactive to SARS-CoV-2 antigens, but followup studies found no positive results using microneutralization assays, and genetic analysis linked the first two infections in Italy to the arrival of tourists from PRC.³⁹²⁻³⁹⁵

4.2 Distinguishing the Two Hypotheses

The spatiotemporal characteristics of these case data can be analyzed using a Bayesian statistical framework, where the probability of a new case occurring within the boundaries of a given gridcell at a given time period is associated with the characteristics of that location and its distance from prior cases. Using this approach, the two competing hypotheses can be distinguished as follows:

Hypothesis Z. Under this hypothesis, the Huanan Market was “the early epicenter” of the pandemic, reflecting “the sustained presence of a potential source of virus transmission into the human population in late 2019, plausibly from infected live mammals sold at the Huanan market.”²¹ Specifically, the first SARS-CoV-2 infection(s) would have been caused by exposure to an infected wildlife mammal but may have been asymptomatic or mildly symptomatic and hence never hospitalized. Over subsequent weeks, the sustained presence of infected animals at the market would have led to further infections (including

severe cases involving hospitalization) and began spreading to community residents living near the market. In fact, spatial analysis of the cumulative case data found that those individuals with no known link to Huanan Market lived closer to the market than many of the individuals with direct links to the market, some of whom resided in outlying areas of Wuhan.²¹

Thus, this hypothesis indicates that all of the early cases in Wuhan were directly or indirectly linked to Huanan Market. In particular, the cases with no known link would have been most likely to occur in the neighborhood surrounding the market or in the vicinity of the residential location of linked cases living further away from the market. As in standard models of infectious disease transmission, the probability of new cases would be highest in densely populated locations at relatively short distances from the locations of prior cases. Moreover, the invisible spread from linked to unlinked cases would predominantly occur during mid-December 2019 and then wane towards the end of the month as the incidence of market infections subsided and the growing number of unlinked cases outside the market became the primary source of new infections.

Hypothesis A. Under this hypothesis, the first individuals to be infected with SARS-CoV-2 would have been lab researchers (perhaps asymptomatic or mildly symptomatic and hence not hospitalized) who then spread the virus to others, most likely through interactions involving sustained contact in indoor spaces. In fact, the restaurants adjacent to Huanan Market are very close to Hankou Railway Station and to the Wuhan CDC, and hence this hypothesis implies that the Huanan Market was merely “*a victim of the pandemic*” (as described by the director of the China CDC in May 2020) and that the SARS-CoV-2 virus “*came into the marketplace before it came out of the marketplace.*”^{389, 396} In effect, most of the early cases in Wuhan had no connection at all to the Huanan Market, either directly or indirectly.

Thus, unlinked cases would be relatively likely in the general vicinity of the BSL-2 labs where bat viruses were being studied, with SARS-CoV-2 spreading invisibly from a few infected researchers to other residents in those areas. On the northwest bank of the Yangtze River, such a pattern could be difficult to distinguish from that implied by hypothesis Z, due to the short distance between the Wuhan CDC and the Huanan Market. Consequently, the implications of hypothesis A would be most visible on the southeast bank where several BSL-2 labs were located, especially since the population density in those neighborhoods was much lower than in the dense urban core around the Hankou Railway Station.

4.3 Methodology

The method of probit analysis has been used in over half a million empirical studies in the natural and social sciences over many decades, dating back to the seminal contributions of Bliss (1934) and Finney (1947).³⁹⁷⁻³⁹⁹ In a spatial context, this method can be used to examine the incidence of positive and negative cases on a geospatial grid and assess the extent to which these outcomes are affected by various grid-specific factors such as population density. In a spatiotemporal context, the probability of a positive case may also be related to explanatory variables at previous points in time, such as the prior prevalence of cases in the neighborhood of each gridcell.

Following this approach, the pattern of early COVID-19 cases in the core Wuhan area can be analyzed using a geospatial grid of 37 x 43 cells, each of which spans about 1 km².³⁷⁷ This grid excludes waterways (lakes and rivers) as well as areas that are practically unpopulated (fewer than 300 inhabitants per km²). Thus, as shown in Figure 9, the grid has a total of 1,291 non-empty cells.

A much larger grid of about 250 x 250 cells could be used to encompass the entire Wuhan area, but such an approach would necessitate a more complex statistical framework to account for the extremely low probability of cases in nearly all of those cells.

Alternatively, the core Wuhan area could be analyzed using a much coarser grid of 11 x 10 cells (with an area of about 9 km² per cell), but such an approach would obscure the details of the spatiotemporal data and hence provide less information to distinguish the competing hypotheses. Moreover, a coarser grid would have numerous cells with multiple new cases in each time period, thereby warranting the use of alternative statistical procedures such as ordered probit (which distinguishes the number of new cases in each gridcell at each time period), poisson regression, negative binomial regression, and logistic regression. The Supplementary Materials provide further information about those alternative approaches.

Date of Symptom Onset	New Cases Linked to the Huanan Market		New Cases with No Known Link to the Huanan Market	
	Number of Gridcells	Percent of Gridcells	Number of Gridcells	Percent of Gridcells
8-11 December	1	0.08	1	0.08
12-16 December	2	0.54	7	0.54
17-21 December	8	1.08	14	1.08
22-26 December	16	1.70	22	1.70
27-31 December	13	3.10	40	3.10

Note: This table reports the pattern of confirmed cases in the core Wuhan area for each 5-day interval in December 2019, using a grid of 1,291 cells (each having area ~ 1 km²). For the period of 27-31 December, one gridcell has 3 confirmed cases with no known link to Huanan Market; there are 0, 1, or 2 unlinked cases for all other gridcells and 5-day intervals. See the Supplementary Materials for further details.

Table 9 tabulates the spatiotemporal pattern of 140 cases in the core Wuhan area, including 46 cases that were linked to the Huanan Market and 94 cases with no known link to that market. As noted above, the Wuhan CDC identified seven clusters of cases among household members or employees working at a specific Huanan Market shop; five of those clusters occurred within the core Wuhan area. The statistical framework used here is not well suited for studying patterns of viral spread within households or market shops, and hence this tabulation only includes the earliest case in each cluster.

The analysis is focused on the spatiotemporal pattern of confirmed cases with no known link to the Huanan Market, for whom the risk of infection could be systematically related to their place of residence; henceforth these cases are referred to as “*unlinked cases*.” By contrast, the confirmed cases that were linked to the Huanan Market presumably resulted from viral transmission within the market, and hence their risk of infection would have no direct connection to their place of residence.

In compartmental models of infectious disease, the rate of growth at the onset of an outbreak reflects the frequency of interactions between infected and susceptible individuals.^{360, 400} Thus, the logarithm of population density is commonly included as an explanatory variable in spatiotemporal studies,

reflecting the extent to which a disease may spread most rapidly in densely populated locations where such interactions occur most frequently.^{401, 402}

Prior studies of infectious diseases have also found that the probability of infection follows a power law with respect to geospatial distance from previous cases.^{403, 404} Thus, a key explanatory variable in this model is expressed in terms of the logarithm of the distance (measured in km) of each gridcell from the nearest case whose symptom onset occurred in the preceding period. Since the logarithm function diverges to $-\infty$ as its input becomes very small, this metric is truncated at a lower bound of zero, and an auxiliary binary variable (taking values of 0 or 1) can be included to indicate whether any prior cases occurred within a given cell or an adjacent cell.⁴⁰²

Thus, Hypothesis *Z* can be expressed as follows:

$$(5) \quad Prob[y_{it}] \sim f_Z \left(\begin{array}{l} \log(dist_linked_{i,t-1}), \text{ nearby_linked}_{i,t-1}, \\ \log(popdensity_i) \end{array} \right)$$

where the dependent variable y_{it} is a binary indicator that equals 1 if any unlinked case was a resident of gridcell i and had symptom onset during period t ; this indicator equals zero if there were no new unlinked cases in that gridcell during that time period. The first explanatory variable is the natural logarithm of the distance from the nearest linked case whose symptom onset occurred in the preceding 5-day period; the second explanatory variable is a binary indicator that equals 1 if the center of that gridcell is within 1km of the nearest linked case in the preceding time period, and 0 otherwise; and the third explanatory variable is the natural logarithm of the population density of that gridcell.

Using the same notation, Hypothesis *A* can be expressed as follows:

$$(6) \quad Prob[y_{it}] \sim f_A \left(\begin{array}{l} \log(dist_linked_{i,t-1}), \text{ southeast_bank}_i, \\ \text{northwest_bank} * \log(popdensity_i) \end{array} \right)$$

Under this hypothesis, the first explanatory variable denotes the natural logarithm of the distance from the nearest linked case in the preceding period; the second explanatory variable is a binary indicator that equals unity if the gridcell is on the southeast bank of the Yangtze River and zero otherwise; and the third explanatory variable is the interaction between population density and a binary indicator that equals unity if the gridcell is on the northwest bank of the Yangtze River, i.e., this variable equals the natural logarithm of population density if the gridcell is on the northwest bank and zero otherwise.

Under each hypothesis, the probit model also includes a binary indicator that equals unity for the final time interval (27-31 Dec) and zero for other time periods. This indicator encapsulates the extent to which the incidence of new unlinked cases may have accelerated in late December due to factors that were distinct and unrelated to the explanatory variables indicated in equations 5 and 6.

Each model is estimated using 5,028 observations, that is, 1,257 gridcells x 4 time periods; the first time period (11-15 Dec) serves as the initial condition for the lagged indicator(s) in each model. Estimation is performed using the Metropolis-Hastings algorithm, as implemented by the *statamh* command in Stata version 18.⁴⁰⁵⁻⁴⁰⁸ The priors for the model coefficients are specified as flat or diffuse. The posterior is estimated using 10 Markov chains, each of which is initialized with a burn-in of 10,000 iterations followed by sampling of 50,000 iterations. Markov Chain Monte Carlo (MCMC) convergence is assessed using the Gelman-Rubin criterion.^{409, 410} See the Supplementary Materials for further details.

Table 10: Probit Analysis of Unlinked Cases in the Core Wuhan Area during December 2019

Model	Model Coefficients			MCMC Convergence Criterion	Deviance Information Criterion	log Marginal Likelihood	Bayes Factor
	Indicator	Estimate	95% CI				
Hypothesis A (accidental lab leak)	southeast bank of Yangtze River	2.69	1.83 to 3.61	1.001	669.47	-338.27	27.1
	northwest bank x population density	0.31	0.21 to 0.41				
	distance to nearest prior linked case	-0.43	-0.58 to -0.28				
Hypothesis Z (zoonotic spillover)	population density	0.23	0.15 to 0.31	1.002	676.38	-341.57	
	distance to nearest prior linked case	-0.41	-0.55 to -0.27				
	prior linked case in same or adjacent cell	0.31	-0.09 to 0.71				
Variant of Hypothesis Z (excluding population density)	distance to nearest prior linked case	-0.55	-0.68 to -0.42	1.000	712.46	-358.41	560 million
	prior linked case in same or adjacent cell	0.33	-0.06 to 0.72				

Note: This table reports on spatiotemporal analysis of confirmed COVID-19 cases in the core Wuhan area during December 2019 for which there was no known link to the Huanan Market. For each hypothesis, the probit model includes the set of explanatory variables indicated in the second column as well as an intercept and a binary indicator for the final time period (27-31 Dec). Each model includes one or more lagged explanatory variables, and hence the data for the initial period (11-15 Dec) are used as initial conditions; thus, each model is estimated using 5,028 observations (1,257 gridcells x 4 time periods). Under each hypothesis, the coefficient estimate and 95% coverage interval for each explanatory variable is given in columns 3 and 4, and the maximum value of the Gelman-Rubin MCMC convergence criterion is shown in column 5. Column 6 indicates the Deviance Information Criterion (which is minimized for the preferred model); column 7 shows the natural logarithm of the average marginal likelihood (which is maximized for the preferred model); and column 8 shows the Bayes Factor, that is, the odds for Hypothesis A relative to Model Z or the variant of Model Z that excludes population density. See the Supplementary Materials for further details.

4.4 Results

As reported in Table 10, this probit analysis provides very strong support for hypothesis *A* relative to hypothesis *Z*, because the spatiotemporal pattern of unlinked cases was markedly different on the southeast bank (where the WIV and several other research labs were located) compared to the spatiotemporal pattern on the northwest bank (where the Huanan Market was located).

The results for hypothesis *A* are shown in the top panel of the table. As in prior studies of infectious disease, the probability of a new unlinked case within a specific 5-day interval is inversely related to that gridcell's distance from prior linked and unlinked cases. The probability of a new unlinked case is directly related to population density for gridcells on the northwest bank but not on the southeast bank. In effect, during December 2019 the virus was not only spreading across the dense urban core of Wuhan on the northwest bank but also spreading among residents on the southeast bank—even in lower-density neighborhoods near Wuhan University and other academic and research institutions. These results are consistent with the premise that the virus began spreading due to unknowing exposure to an infected lab researcher and that it “*came into the market before it came out of the market.*”

The results for hypothesis *Z* are shown in the middle panel of the table. For this model, the probability of a new unlinked case at a given time and location is strongly related to the gridcell's population density and its distance from the nearest linked case in the previous time period. In effect, the Huanan Market was the epicenter from which the virus began spreading radially outwards, and the unlinked cases in December 2019 were individuals who were unknowingly infected by exposure to market vendors or customers.

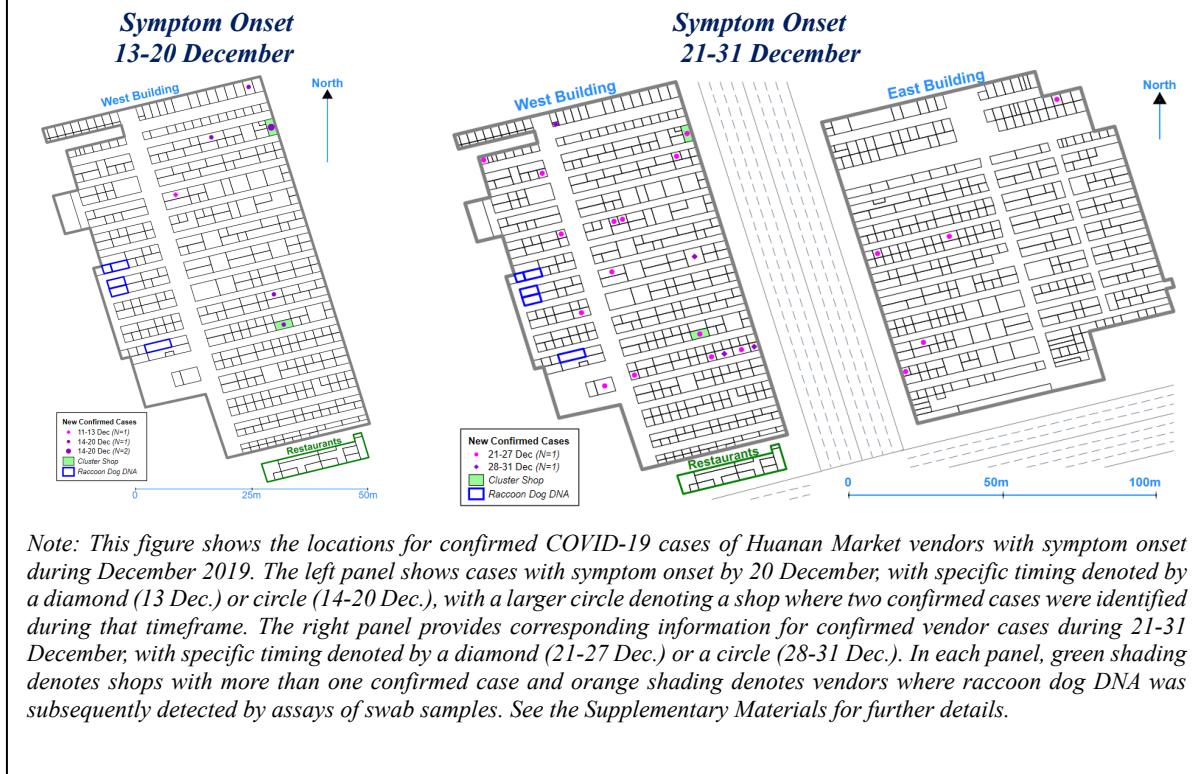
As shown in the table, this probit analysis produces a Bayes factor of 27.1, i.e., the conditional odds in favor of hypothesis *A* relative to hypothesis *Z* are about 27:1. Referring back to Table 2, these odds can be interpreted as “*strong evidence*” for an accidental lab leak.

Although not shown in the table, these results are highly robust to alternative estimation procedures. For example, when each model is estimated by maximum likelihood, the model for hypothesis *A* is strongly favored by Akaike's Information Criterion and the Bayesian Information Criterion. Similarly, the results are not sensitive to alternative specifications of the explanatory variables. For example, augmenting model *A* with the lagged distance to unlinked cases reduces the deviance information criterion (indicating that it has some additional explanatory power), but the log marginal likelihood is essentially identical to that shown in Table 10. See the Supplementary Materials for further details.

These findings differ starkly from those of Worobey et al. (2022), henceforth referred to as WOR22. Several differences in methodology are noteworthy: (i) WOR22 focused on the spatial pattern of cumulative cases as of 31 December 2019 but did not analyze the spatiotemporal evolution of cases over shorter time intervals. (ii) WOR22 assessed the centroid and radial pattern of cases but did not utilize a probabilistic framework for assessing the determinants of cases using a geospatial grid. (iii) WOR22 identified the centroids of linked vs. unlinked cases but did not compare other potential differences in the spatial distribution of such cases.

Finally, this spatiotemporal analysis clearly indicates the role of population density in determining the early spread of the SARS-CoV-2 virus. This result stands in stark contrast to the analysis of WOR22, who concluded that “*COVID-19 cases in December 2019 were associated with the Huanan market in a manner unrelated to Wuhan population density or demographic patterns.*” The spatiotemporal analysis indicates that such a conclusion is completely untenable, with odds of about 560 million to 1 against that particular version of hypothesis *Z*.

Figure 12: COVID-19 Cases of Huanan Market Vendors in December 2019



5. Spatiotemporal Pattern of Huanan Market Vendor Cases

This section focuses on assessing the conditional Bayes factor BF_5 , which incorporates information about the specific timing and location of confirmed COVID-19 cases of vendors at the Huanan Market with symptom onset by 31 December 2019. This factor is conditional on the pandemic outbreak having occurred in Wuhan and on the timing and residential locations of all confirmed early cases. Specifically, this conditional Bayes factor can be expressed as follows:

$$(7) \quad BF_5 = \frac{\text{Marg. Likelihood}(\text{Pattern of Vendor Stalls} | \text{hypothesis } A; \text{pattern of non-vendor cases})}{\text{Marg. Likelihood}(\text{Pattern of Vendor Stalls} | \text{hypothesis } Z; \text{pattern of non-vendor cases})}$$

In effect, this Bayes factor gauges the evidence that the SARS-CoV-2 virus “*came into the market before it came out of the market*” (according to hypothesis A) or was triggered by a zoonotic spillover from wildlife mammals being sold at the market (according to hypothesis Z).

Figure 12 shows the locations of market vendors with onset of symptoms during the periods of 13-20 December (left panel) and 21-31 December (right panel), with symbols denoting more specific timing of cases within each period. These data are taken from the Annex of the WHO 2021 Report, which depicted the spatial distribution of vendor cases at each of those dates; see the Supplementary Materials for details on the digitization of that information.²⁰ This figure also shows shops where wildlife mammals were sold, based on analysis of mitochondrial DNA (mtDNA) in the swab samples that were collected by the Wuhan CDC after the market was closed in January 2020.

Figure 13: Vendor Operations at the Huanan Market



Note: This figure illustrates the operations of the Huanan Market prior to the onset of the COVID-19 pandemic. The five photographs of market shops and restaurants were taken on 02 November 2019. The lower-right photo shows a live raccoon dog observed for sale at Huanan Market in 2014. Credits: Soho.com (main entrance), Arend Kuester (vendors & restaurants), Edward Holmes (raccoon dog).

None of the vendors whose symptom onset occurred by 20 December 2019 were working at wildlife shops; four of those seven vendors were employed at shops at the other end of the west wing. These spatiotemporal data are relevant for assessing hypotheses *A* and *Z* but have not been analyzed by prior studies, which focused on gauging the spatial centroid of vendor cases regardless of timing.^{21, 354}

5.1 Salient Characteristics of the Huanan Market

Following the onset of the COVID-19 pandemic, the Huanan Wholesale Seafood Market was frequently labeled as a “wet market” by scientific researchers as well as media commentators, with echoes of the burgeoning wildlife trade associated with SARS.^{27, 92, 134, 301, 365, 411-414} However, the label “wet market” does not fully convey its salient characteristics.⁴¹⁵

- *Location.* The Huanan Market was a relatively placid haven in the dense urban core of Wuhan, comprised of two main buildings and some adjacent structures, including a separate building with restaurants that served market vendors and clients as well as others living or working in the vicinity of the market.³⁵⁴ The east and west wings of the market were separated by Xinhua Road, a major thoroughfare about 40m wide.
- *Internal Layout.* In contrast to many congested retail centers, the Huanan Market was relatively spacious. The internal layout of the market is shown in Figure 12, which was produced by digitizing the maps in the WHO 2021 Report that indicated the actual sizes and shapes of the shops at the Huanan Market, as distinct from stylized diagrams with square cells that have been used in several epidemiological studies.^{20, 21, 28} These spatial data indicate dimensions of 3m × 3m for the median shop at the Huanan Market. Access to market shops was facilitated by common space and by evenly spaced corridors (commonly referred to as the “streets” of the market) with widths of 2.5m or greater. As shown in Figure 13, market shops generally had

an internal height of about 3m, with higher ceilings in common areas and corridors; see the Supplementary Materials for further details.^{20, 21, 416-418}

- *Wholesale Operations.* As evident from its name, the Huanan Wholesale Seafood Market was a wholesale outlet focused on bulk sales to purchasers from local restaurants and hotels. As of December 2019, there were 678 shops with 1162 workers serving about 10,000 clients per day.²⁰ In effect, a typical shop employed two vendors and served roughly 15 clients per day, that is, about one or two clients per hour. By contrast, as noted in the WHO 2021 Report, the Baishazhou Market on the other side of the river was a far larger enterprise that served individual customers as well as wholesale clients.²⁰
- *Food Products.* A majority of shops at Huanan Market sold aquatic products, including fresh and frozen seafood as well as live shellfish.²⁰ In addition, many shops sold poultry and livestock meat (both fresh and frozen) as well as vegetables and miscellaneous items (grains, oils, seasonings, dried fruits, and prepackaged foods).
- *Wildlife Mammal Sales.* The WHO 2021 Report indicated that 7 out of 678 shops at the market were engaged in sales of live wildlife mammals: 6 shops in the West Building that were selling bamboo rats, badgers, hedgehogs, porcupines, and rabbits, and one shop in the East Building that sold Sika deer.²⁰ (Three other market shops sold live reptiles, which are not infected by SARS-Cov-2.) The WHO Report included a statement from a witness who had observed a raccoon dog at the market in 2014 but did not indicate any sales of live raccoon dogs in late 2019, presumably because such sales had been occurring surreptitiously on the fringes of the market in relatively small numbers.^{20, 27} Across all Wuhan markets, a monthly survey observed daily average sales of 11 hedgehogs, 5 hares, 1 red fox, 1 bamboo rat, and 1 raccoon dog, and sales at the Huanan Market comprised a subset of that total.²⁷ Evidently, this context was markedly different from the wildlife markets in southern China at the onset of SARS, where hundreds of wildlife mammals were sold on a daily basis.^{97, 98}

Product Type	Vendors	Confirmed Cases	Incidence (95% CI)
Freshwater Aquatic	559	15	2.7% (1.5–4.2%)
Seafood	484	14	2.9% (1.6–4.5%)
Livestock Meat	318	9	2.9% (1.4–5.0%)
Miscellaneous	266	3	1.4% (0.4–2.9%)
Poultry Meat	230	8	3.5% (1.6–6.1%)
Vegetables	108	5	4.5% (1.6–8.6%)
Live Animals	15	0	5.0% (0.6–14%)
Total	1162	30	2.6% (1.8–3.5%)

Note: This table reports Bayesian mean estimates and 95% coverage intervals for the incidence of confirmed COVID-19 cases of vendors at Huanan Market during December 2019. Some shops sold multiple types of products, so the individual rows cannot be added to derive the totals shown. See the Supplementary Materials for further details.

5.2 Timing and Distribution of Cases

Among the six confirmed COVID-19 cases with symptom onset by 14 December 2019, only one (a shrimp vendor) was an employee at the Huanan Market.²⁰ Among the confirmed cases with symptom onset during the second half of that month, there were 6 vendors who developed symptoms by 20 December; 17 vendors with symptom onset during 21-27 December; and 6 vendors with onset during the final days of December 2019 before the market was closed. There were two clusters of COVID-19 cases at the market: 3 vendors at a frozen food shop, and 2 vendors at a seafood shop.

As shown in Table 11, the incidence of vendor cases was roughly proportional to the total number of vendors for each type of product. The mean and 95% coverage interval for each product type was estimated using a Bayesian procedure with a relatively flat prior. The overall incidence was 2.6% (95% CI: 1.8 to 3.5%), and the estimate for each specific product category was consistent with that overall estimate. Notably, there were no confirmed cases among the 15 vendors engaged in selling live animals, but that outcome is not particularly surprising from a Bayesian perspective given that there were only a few such vendors at Huanan Market. See the Supplementary Materials for details.

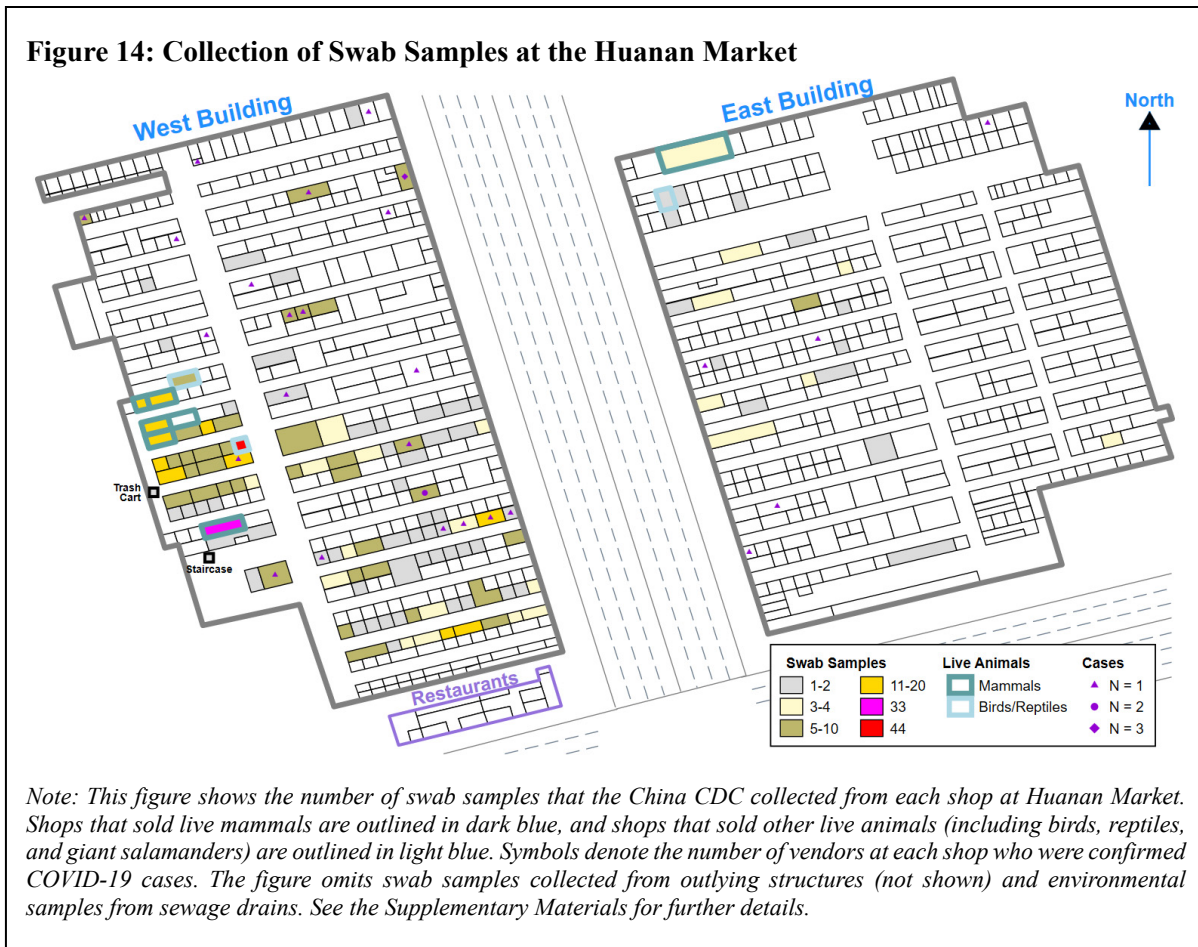
5.3 Modes of Transmission

The structural features of the Huanan Market would not be expected to facilitate airborne viral transmission within the market. As noted above, the original strain of SARS-CoV-2 was most likely to be transmitted in instances involving prolonged close contact with an infected individual (i.e., at least 15 minutes) in a confined space with limited ventilation.³⁶⁹⁻³⁷⁶ At the early stages of the pandemic, concerns were also raised about infectious droplets and viral particles on hard surfaces, but subsequent research found that such fomites posed a negligible risk of infection.⁴¹⁹⁻⁴²³

Another potential mode of transmission was flagged by a shrimp vendor (the earliest confirmed vendor case), who worried about having gotten infected while using the public toilets near the southwest corner of the west building.^{384, 424-428} Such concerns were not unreasonable given the notable role of sewage in the spread of the SARS virus.^{384, 424-428} However, epidemiological studies have concluded that SARS-CoV-2 is associated with a miniscule risk of faecal transmission.⁴²⁹⁻⁴³²

Local restaurants were identified as having a significant role in the spread of the SARS virus but were only briefly mentioned in the WHO 2021 Report, which noted that at Huanan Market, “*employees ate in the canteen of the market*” while “*vendors usually brought food from home or ate in neighboring canteens.*”²⁰ As shown in Figures 12 and 13, the restaurants at the Huanan Market were located in an adjacent structure on Fazhan Boulevard, at the southern end of the west wing, and hence were presumably oriented towards serving market workers and clients as well as community residents and pedestrians heading to/from Hankou Station.³⁵⁴ Under hypothesis *A*, those restaurants may have facilitated the transmission of the virus from local residents and passersby to market vendors (i.e., how it “*came into the market before it came out of the market*”), whereas the direction of transmission would be reversed under hypothesis *Z* (i.e., the virus originated at the market).

Likewise, neighborhood clinics, which provided services to Huanan workers as well as local residents, may have played a role in transmitting the virus. One such clinic provided antibiotic injections to vendors with symptoms of seasonal flu but later were confirmed with COVID-19.³⁸⁴ The director of that clinic, who lived in the neighborhood, was also infected and succumbed to the disease.^{356, 433}



As with Huanan restaurants, it is difficult to ascertain whether these clinics may have contributed to the spread of SARS-CoV-2, because that question was not addressed in the WHO 2021 Report.

5.4 Detection of SARS-CoV-2 in Swab Samples

In early 2020 the China CDC collected 651 swab samples from surfaces and objects in 135 shops at the Huanan Market. The China CDC's surveillance began when the market was closed and initially focused on using an RT-qPCR assay to clarify the mode of transmission of the SARS-CoV-2 virus; a subset of samples were then subjected to genomic analysis to identify matches to a SARS-CoV-2 reference sequence.^{28, 434}

Figure 14 shows the spatial characteristics of the sampling procedures. As the China CDC noted in its peer-reviewed study, "the selection of shops for sampling was biased because shops selling wildlife and shops linked to early cases were prioritized for sampling."²⁸ In particular, samples were gathered from 135 shops but not from the other 543 shops at the market.²⁸ One or two samples were taken from most shops, usually by applying swabs to solid surfaces such as floors, walls, shelves, or rolling shutters. At some shops, additional swabs were applied to other items such as furniture, appliances, containers, clothing, and kitchen utensils. The most intensive sampling occurred at specific shops engaged in the sale of wildlife mammals; in fact, 77 samples (more than 10% of the total) were obtained at two such shops.²⁸ See the Supplementary Materials for further details.

A. RT-qPCR Analysis						
Samples per Shop	Swab Samples			Market Shops		
	Samples Collected	Positive SARS-CoV-2	Percent	Shops Surveyed	Positive SARS-CoV-2	Percent
1-4	184	6	3.3	86	6	6.9
5-9	216	15	5.9	31	8	25.8
10-20	209	5	2.6	16	4	25.0
21 or more	77	9	11.7	2	2	100.0
Total	686	35	5.1	135	20	14.7
B. Genomic Analysis						
Samples per Shop	Swab Samples			Market Shops		
	Samples Analyzed	Positive Match to SARS-CoV-2	Percent	Shops Analyzed	Positive Match to SARS-CoV-2	Percent
Negative	94	3	3.1	10	0	0.0
Positive	33	30	90.9	19	19	100.0
Total	130	33	25.4	29	19	65.6
<i>Note: This table reports on RT-qPCR and genomic analysis of 687 swab samples that the China CDC collected from 136 shops at the Huanan Market in early 2020. (No samples were collected from the other 542 shops at the market.) The China CDC identified 130 of those swab samples (collected at 29 shops) as suitable for high-throughput processing and genomic analysis. See the Supplementary Materials for further details.</i>						

The upper panel of Table 12 gives a synopsis of the RT-qPCR results, which were positive for 35 samples taken from 20 shops. Even at shops where SARS-CoV-2 was identified, the RT-qPCR results were positive in a relatively small proportion of samples. For example, at two shops that had been selling wildlife, there were 9 positive results from 77 swab samples. Even at market shops where vendors became infected with SARS-CoV-2, viral particles evidently did not contaminate every hard surface or physical object within the shop.

Viral RNA has less ambient stability than DNA and tends to degrade fairly rapidly when exposed to air at typical room temperatures.^{420, 421, 435} Thus, the RT-qPCR detection rate using samples collected in early January 2020 might well understate the true incidence of infections of market vendors over preceding weeks in December 2019 (including asymptomatic and mild cases). Since the China CDC's swab sampling was most intensively conducted at or near the market shops selling wildlife mammals, such undetected SARS-CoV-2 infections would tend to be located at larger distances from those shops. This sampling bias could skew estimates of the spatial distribution of positive RT-qPCR readings and hence induce bias towards the hypothesis of zoonotic spillover.

Sampling occurred at 15 shops with confirmed COVID-19 vendor cases but not at 12 other shops where vendor cases had not yet been identified or confirmed.^{20, 28} For those 15 shops with early COVID-19 cases, the swab samples from 7 shops had at least one positive RT-qPCR reading, whereas the other 8 shops had uniformly negative readings. In effect, the sensitivity of the test procedure was less than 50%, well below that associated with intranasal swabs from human patients.^{28, 436, 437}

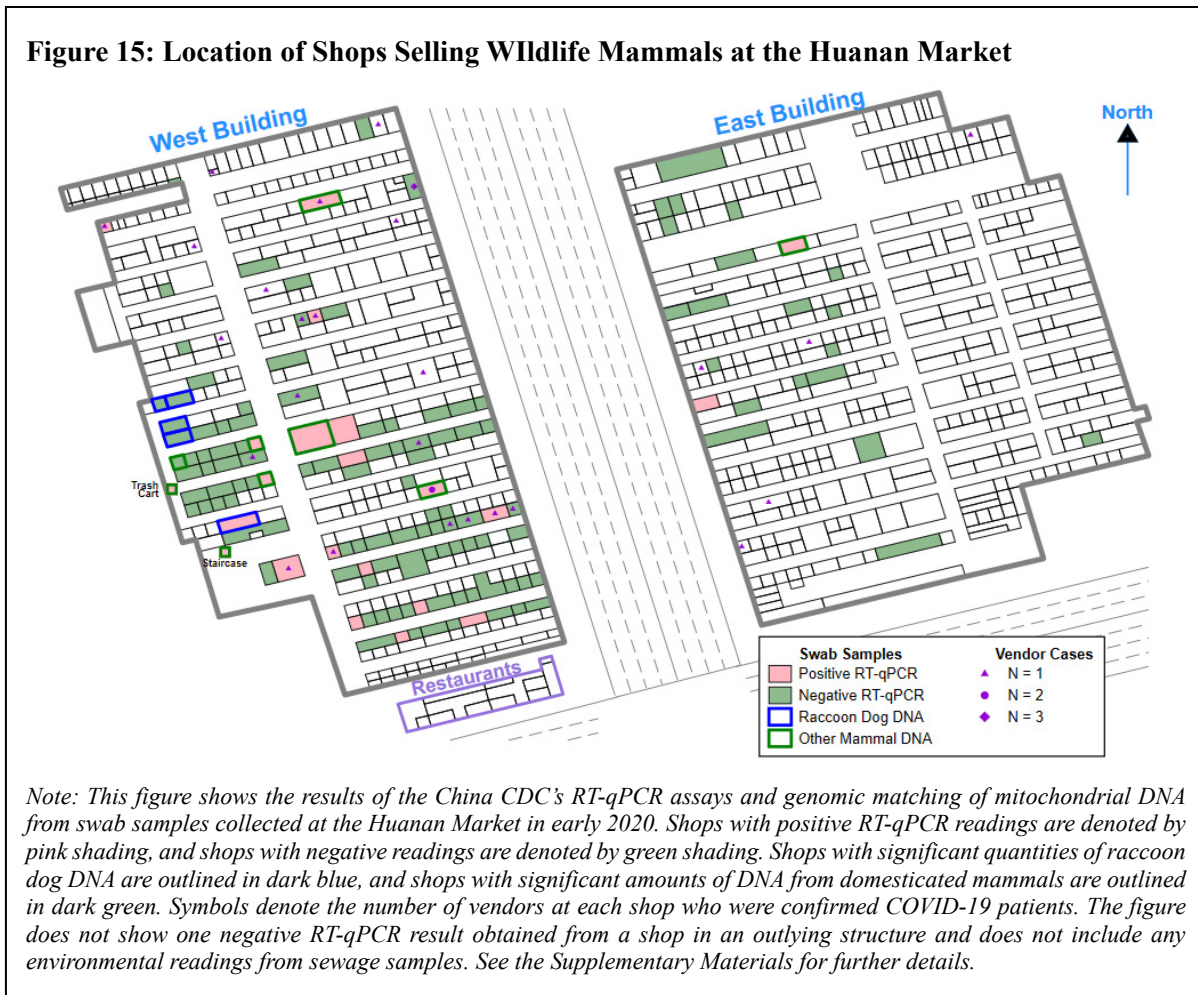
Nonetheless, these results do not necessarily reflect intrinsic limitations of the RT-qPCR assay used by the China CDC; rather, the attenuated sensitivity may reflect the fact that the swab samples were gathered from a wide variety of surfaces and objects and might also reflect variations in the extent of degradation of viral particles at the time those samples were collected. Nearly all of the shops with positive RT-qPCR readings were identified during the first two weeks of surveillance, but one additional shop was identified using swab samples collected during February 2020.

The lower panel of Table 12 gives a synopsis of the genomic analysis that was performed for a subset of 130 swab samples (collected at 29 shops) that were suitable for high-throughput processing.²⁸ Genomic matches to a SARS-CoV-2 reference sequence were obtained for 33 samples from 19 shops, all of which had at least one positive RT-qPCR reading; by contrast, there were no significant matches for any samples from the other 10 shops, all of which had uniformly negative RT-qPCR readings.²⁸ These results provide further support for the validity of the RT-qPCR assay in identifying the market shops that had been contaminated by the SARS-CoV-2 virus.

It is also noteworthy that positive matches were obtained for about two-thirds of the shops with genomic data ($19/29 = 66\%$), far higher than the positivity rate for RT-qPCR samples ($20/35$ shops = 15%) or the incidence of confirmed COVID-19 vendor cases ($27/678$ shops = 4%). That divergence reflects the extent of bias in the selection of shops for sampling as well as the number of samples obtained from each shop.

These sampling characteristics have substantive implications for determining the statistical methodology for analyzing the RT-qPCR data:

- *The RT-qPCR readings should be analyzed using each market shop as the unit of observation.* Previous studies have analyzed the positivity of individual swabs but have not considered the wide variation in the number of samples collected at each shop or in the types of surfaces and objects that were sampled.^{21, 312, 354} Ideally, such variations would be incorporated directly into the statistical analysis, but such an approach remains impractical given the absence of detailed information about the layout and furnishing of each shop and the precise position of each sampled surface or item within the shop.
- *The RT-qPCR data should be analyzed in terms of a binary indicator that incorporates negative as well as positive readings.* Previous studies have focused solely on shops with positive tests, thereby discarding relevant information and exacerbating the bias associated with the China CDC's sampling procedures.^{21, 312, 354} For the reasons discussed above, such sampling bias would tend to favor the hypothesis of zoonotic spillover.
- *The RT-qPCR data should be analyzed using a probabilistic framework* that allows for intrinsic randomness of viral RNA degradation as well as the various idiosyncratic factors that affected the China CDC's sampling procedures. In contrast, previous studies have followed a deterministic approach in focusing on the centroid of positive RT-qPCR results.^{21, 312, 354}



5.5 Location of Shops Selling Wildlife Mammals

The WHO 2021 Report listed six shops in the west wing that had been selling live exotic mammals (excluding livestock), but none of those shops disclosed any sales of raccoon dogs.²⁰ Survey data published in mid-2021 reported on citywide sales of raccoon dogs in Wuhan but did not specify whether such sales had occurred at Huanan Market.²⁷

However, DNA analysis of swab samples has confirmed the presence of raccoon dogs and has clarified their location within the market, which had previously remained obscure.²⁸ These samples contained a mean of 124 million genetic fragments (95% CI: 52 to 181 million) with an average length of 115 base pairs (95% CI: 43 to 200); see the Supplementary Materials. The DNA sequences of these fragments was then matched to the entries in two reference databases: the mitochondrial DNA of mammalian genera in the Barcode of Life Database (BOLD), and the whole-genome sequences of vertebrate species in the Nucleotide database maintained by the U.S. National Center for Biotechnology Information (NCBI).^{28, 438-442}

Table 13: Analysis of Mitochondrial DNA for Selected Mammals at the Huanan Market

Genus	Common Name	Description	Barcode of Life Database (BOLD)			
			Maximum # of matching mtDNA reads per sample	# Swab Samples		# Shops with more than 1200 matching mtDNA reads
				More than 20 matching mtDNA reads	More than 1200 matching mtDNA reads	
Pongo	Orangutan	Endangered	24	1	0	0
Hylobates	Gibbon	Endangered	35	1	0	0
+ Meles	Badger	Wildlife	41	1	0	0
Loxodonta	African Elephant	Endangered	42	3	0	0
Symphalangus	Siamang	Endangered	42	1	0	0
Mammuthus	Mammoth	Extinct	56	5	0	0
- Mustela	Weasel	Wildlife	71	6	0	0
Homotherium	Sabre-Tooth Cat	Extinct	87	2	0	0
Dicerorhinus	Sumatran Rhino	Endangered	106	1	0	0
+ Hystrix	Porcupine	Wildlife	126	2	0	0
Semnopithecus	Langur Monkey	Vulnerable	146	7	0	0
Panthera	Leopard	Vulnerable	149	16	0	0
+ Vulpes	Fox	Wildlife	174	14	0	0
- Melogale	Ferret	Wildlife	188	2	0	0
Ailurus	Red Panda	Endangered	261	11	0	0
- Paguma	Palm Civet	Wildlife	274	1	0	0
+ Lepus	Hare	Wildlife	379	30	0	0
+ Muntiacus	Muntjac	Wildlife	381	18	0	0
- Arctonyx	Badger	Wildlife	519	15	0	0
+ Erinaceus	Hedgehog	Wildlife	811	21	0	0
Elephas	Asian Elephant	Endangered	1,027	18	0	0
Genera with more than 1200 matching mtDNA reads						
Rattus	Common Rat	Vermin	2,352	38	3	2
+ Rhizomys	Bamboo Rat	Wildlife	2,691	37	1	1
Homo	Human	Human	3,157	62	5	5
Oryctolagus	Rabbit	Livestock	3,843	35	2	1
Ovis	Sheep	Livestock	3,917	31	4	4
Canis	Dog	Pets	4,535	48	5	3
Sus	Pig	Livestock	4,604	41	3	3
+ Marmota	Marmot	Wildlife	5,256	4	1	1
Lariscus	Ground Squirrel	Pets	8,188	67	7	3
+ Nyctereutes	Raccoon Dog	Wildlife	9,619	41	6	5
Bos	Cattle	Livestock	26,922	34	3	3

Note: This table summarizes the results of the China CDC's high-throughput analysis of genetic material from swab samples collected at the Huanan Market in early 2020, as reported in Supplementary Table S8 of Liu et al. (2023). Each row refers to a specific mammal genus for which at least 20 matching reads of mitochondrial DNA (mtDNA) were identified. The plus symbol (+) indicates that the specified mammal was observed at the Huanan Market in late 2019, and the minus symbol (-) indicates that the specified mammal was observed for sale at other Wuhan markets but not at the Huanan Market. Columns 2 and 3 indicate the common name and current status of the specified genus. Column 4 indicates the maximum number of matching mtDNA reads obtained from any single swab sample. Columns 5 and 6 indicate the number of samples with at least 20 and 1200 matching mtDNA reads, respectively. The lower panel of the table lists the genera for which samples with more than 1200 matching mtDNA reads were identified, and column 7 indicates the number of Huanan Market shops at which such samples were obtained. See the Supplementary Materials for further details.

In identifying the genera of mammals at the Huanan Market, the China CDC used a provisional threshold of at least 20 matching mtDNA reads.²⁸ As shown in Table 13, however, that threshold is not stringent enough to exclude two extinct genera (mammoth and sabre-tooth cats) nor several genera of rare and endangered mammals that were certainly not on sale at the Huanan Market. A higher threshold of 1200 matching mtDNA reads excludes all of those false positives.

The more stringent threshold (more than 1200 matching reads) identifies DNA of humans and various domesticated mammals that were present at the market (dogs, and livestock) as well as several other genera of mammals (including bamboo rats, marmots, rabbits, and squirrels) that were observed for sale at Wuhan markets and that were listed in the WHO 2021 Report's tabulation of wildlife mammal sales at the Huanan Market.^{20, 27}

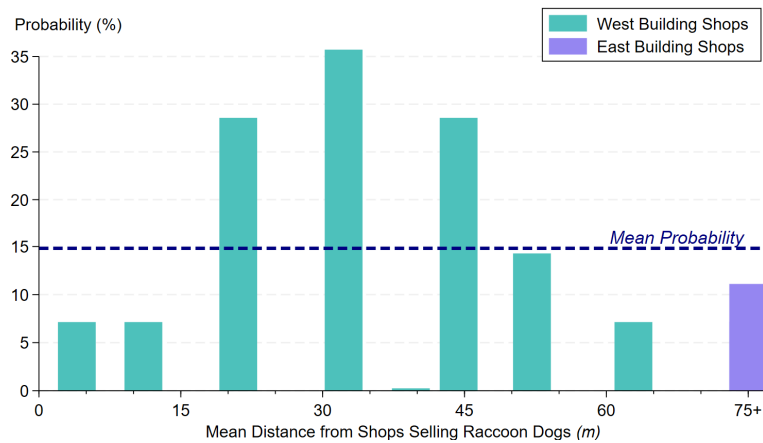
This analysis of mitochondrial DNA has demonstrated that raccoon dogs were indeed being sold by five shops at the Huanan Market.⁴⁴³ Figure 15 shows the location of those five shops, all of which were identified in the WHO 2021 Report as having sold wildlife mammals.²⁰ The same shops are also identified by analysis of the China CDC's tabulation of DNA matches with whole-genome sequences of vertebrate species using the NCBI Nucleotide database; for those data, rare and endangered mammals are excluded using a threshold of 1000 matching reads instead of the provisional threshold of 100 matches that was used in the China CDC's analysis.²⁸ See the Supplementary Materials for further details.

Nonetheless, the identification of these five shops may further diminish the likelihood that raccoon dogs were the primary intermediate host for SARS-CoV-2. The observational survey reported mean sales of 38 raccoon dogs per month at all Wuhan markets.²⁷ Even if all of those monthly sales were made by the shops at Huanan Market, each shop would typically be selling about 2 wild raccoon dogs per week—perhaps sometimes a male and female trapped together. Laboratory experiments have shown that raccoon dogs transmit SARS-CoV-2 via direct nasal contact.³⁴⁷ Thus, there would be negligible likelihood of the virus spreading among raccoon dogs sold at different shops, and a low likelihood that caged raccoon dogs transmitted the virus to any other wildlife animals, because raccoon dogs only shed SARS-CoV-2 through direct nasal contact, not faeces or airborne particles.³⁴⁷

A recent study has reanalyzed the China CDC's genomic data with the specific aim of identifying mitochondrial DNA matches with the wildlife mammals that were observed at Wuhan markets during fall 2019.³¹² The matching algorithm for that analysis was validated using simulations of randomly-generated synthetic data, but such an approach would not necessarily exclude spurious matches. In fact, the algorithm identified DNA matches to the red panda, an endangered species that was certainly not being sold at the Huanan Market, as well as several other mammal species that were not observed there in fall 2019: arctic fox, masked palm civet, nutria, and wild boar.^{27, 242, 312, 444} Notwithstanding its limitations, several aspects of that study's findings are relevant for the analysis conducted here:

- *Raccoon dogs were being sold at Huanan Market shops in late 2019.* This fact was not yet evident at the time of the WHO 2021 Report but has subsequently been confirmed by direct observation and genetic analysis of swab samples collected at the market.^{20, 21, 27, 312}
- *The raccoon dogs being sold at Huanan Market were trapped from the wild, not farmed for fur.* This characterization was evident from the market surveys that were conducted prior to the pandemic and is consistent with the practices of fur farms, which slaughter animals onsite

Figure 16: RT-qPCR Test Results and Distance from Wildlife Mammal Shops



Note: This figure uses the RT-qPCR test results from the 135 shops at the Huanan Market where swab samples were collected by the China CDC in early 2020, as well as the distance of each shop from the nearest seller of wildlife mammals (as shown in Figure 16). The 112 shops in the West Building are grouped into 8 bins (with 14 shops per bin), and the 23 shops in the East Building are grouped into a separate bin. Each bin's horizontal position indicates the mean distance of those shops from the nearest wildlife mammal sellers, and the vertical axis indicates the proportion of those shops with at least one positive RT-qPCR reading. See the Supplementary Materials for further details.

and then ship the fur pelts to wholesale centers.^{328, 329} A recent study also reached this conclusion using the China CDC's genomic data to identify the subspecies of raccoon dogs that were being sold at the Huanan Market; however, that study did not discuss the intrinsic difficulties of using mitochondrial DNA for intraspecies identification.^{312, 440, 445-449}

- *The SARS-CoV-2 virus contaminated at least one shop that was selling raccoon dogs.* In light of the idiosyncracies of the China CDC's sampling procedures and the degradation of viral RNA, any single swab sample might have a large amount of SARS-COV-2 residue or a large amount of raccoon dog DNA but would not necessarily have significant amounts of both types of genetic material.^{443, 450-453}
- Analysis of the genomic data collected by the China CDC cannot resolve the question of whether wildlife mammals at the Huanan Market were the source of the zoonotic spillover. In particular, these data cannot be used to determine whether any wildlife mammals were infected with SARS-CoV-2 and hence leave open the question of whether or not the virus "came into the market before it came out of the market."^{443, 454}
- Swab sample collection was intensified in the vicinity of the shops selling wildlife mammals, but the proportion of positive vs. negative RT-qPCR readings among those shops was no higher than in other areas of the Huanan Market. As shown in Figure 16, the proportion of positive readings was relatively low at the sampled shops that were closest to wildlife mammal sellers and relatively high in other areas of the market. The significance of this spatial pattern can be gauged using a probabilistic statistical framework.

5.6 Ancestral Strains of SARS-CoV-2

Complete genetic sequences of SARS-CoV-2 were obtained from three swab samples collected at the Huanan Market and from 14 patients with symptom onset prior to 31 December 2019, of whom 12 patients were vendors or purchasers at the Huanan Market.^{20, 21, 126} Those sequences have been classified into two viral strains A and B that differ by two nucleotides and encode proteins that differ by a single amino acid.⁴⁵⁵

For each of the nucleotides where these two strains differ, strain A exactly matches the genetic sequence of the bat virus RaTG13, whose genome has 96.8% similarity to SARS-CoV-2.^{127, 132} Thus, under the maintained assumption of zoonotic spillover, phylogenetic analysis has identified strain A as the ancestral strain of SARS-CoV-2, with strain B having subsequently arisen from a pair of neutral or moderately advantageous mutations to strain A.⁴⁵⁶ However, the temporal pattern of observed viral strains in December 2019 appears to be inconsistent with a single zoonotic spillover of strain A:

- *Strain A* was matched to genetic sequences from two patients with symptom onset on 26 and 27 December 2019, respectively. One of those individuals lived in the dense urban core of Wuhan, near the Hankou Station as well as the Huanan Market, and the other was staying at a hotel in that vicinity.^{20, 21, 126} In addition, Strain A was matched to one swab sample from the Huanan Market on 01 January 2020, namely, a pair of gloves at a shop on Street 7 in the west wing of the market.²⁸ There is no evidence that this viral strain was circulating in Wuhan prior to the last week of December 2019.
- *Strain B* was matched to all 12 genetic sequences from patients linked to the Huanan Market, eight of whom had symptom onset by 20 December 2019, including the earliest linked case of the shrimp vendor who had confirmed COVID-19 symptoms by 13 December.²⁰ Strain B was also matched to two of the three swab samples collected from that market.²⁸ These data would reasonably be viewed as supporting the premise that this viral strain was already circulating widely by mid-December 2019.

To alleviate that inconsistency, some analysts have conjectured that zoonotic spillover occurred on two distinct occasions, i.e., strains A and B were both circulating in an intermediate host, but the transmission of strain B to humans occurred a few weeks before the transmission of strain A. Using a complex statistical framework, Pekar et al. (2022) assessed the number of spillovers and reported a Bayes factor of 62 that conveyed very strong evidence for the dual spillover scenario.⁴⁵⁶ However, a subsequent erratum revised that Bayes factor to 4.3 (moderate evidence).⁴⁵⁷ And the correction of further modelling errors produces a Bayes factor below unity, indicating that the observed data on strains A and B are not sufficient to determine the number of zoonotic spillover events.^{458, 459}

Moreover, the scenario of dual zoonotic spillovers requires “*the sustained presence of a potential source of virus transmission into the human population in late 2019.*”²¹ But such an outcome seems inconsistent with the fact that raccoon dogs are solitary wild animals and were only trapped and sold sporadically by a few shops at the Huanan Market, with mean sales of about one raccoon dog per day. It is unclear how such an intermediate host could generate the spillover of two distinct strains of SARS-CoV-2 within just a few weeks. The hypothesis of an accidental lab leak does not impose any such constraints and hence seems favored by the principle of Occam’s razor.⁴⁶⁰

Table 14: Probit Analysis of RT-qPCR Samples from the Huanan Market

Model	Gelman-Rubin Convergence Criterion (R_c)	Distance to Nearest Raccoon Dog Shop		Deviance Information Criterion	log Marginal Likelihood	Bayes Factor
		Estimate	95% CI			
Hypothesis A (accidental lab leak)	1.0	---	---	115.24	-59.28	3.19
Hypothesis Z (zoonotic spillover)	1.0	-4.67	-12.1 to 2.30	116.35	-60.43	

Note: This table reports on probit analysis of RT-qPCR samples that the China CDC collected at the Huanan Market in early 2020. Under Hypothesis A, the probit model only includes an intercept, i.e., all shops have the same probability of a positive assay reading. Under Hypothesis Z, the probit model also includes each shop's distance to the nearest seller of raccoon dogs. For each hypothesis, the second column indicates the maximum value of the Gelman-Rubin convergence criterion (R_c) across the parameters of that model; the third and fourth columns report the coefficient estimate and 95% coverage interval for the distance of each shop from the nearest raccoon dog seller; the fifth column indicates the Deviance Information Criterion (DIC), which is minimized for the preferred model; the sixth column indicates the log marginal likelihood of each model; and the final column indicates the Bayes factor, i.e., the odds for Model A relative to Model Z. See the Supplementary Materials for further details.

5.7 Spatial Analysis of RT-qPCR Results

The two alternative hypotheses have divergent implications for the spatial pattern of RT-qPCR readings at the 135 Huanan Market shops where swab samples were collected. Under the zoonotic spillover hypothesis, the shops selling wildlife mammals were the epicenter of the pandemic, and hence the probability of a positive assay should be highest in the vicinity of those shops. In contrast, the accidental lab leak hypothesis indicates that the virus came from outside the market and hence the probability of a positive assay should be unrelated to the distance from wildlife mammal vendors.

Thus, the probabilistic framework can be expressed as follows:

$$(8) \quad \text{Prob}[PCR_j = 1] \sim f(RD_j)$$

where the indicator $PCR_j = 1$ for a positive assay and $PCR_j = 0$ for a negative assay; the explanatory variable RD_j denotes the distance of stall j from the nearest shop selling raccoon dogs, and the index $j = 1, \dots, 135$.

Under hypothesis A , the probit model only includes an intercept, i.e., all shops have an equal probability of a positive RT-qPCR reading. Under hypothesis Z , the probit model also includes each shop's distance to the nearest raccoon dog vendor (as shown in Figure 16). The priors for the model parameters are specified as diffuse. Each model is estimated using a Bayesian procedure with 10 MCMC chains, with 20,000 iterations per chain and burn-in of 10,000 observations. Convergence of the MCMC algorithm is verified when the Gelman-Rubin criterion (R_c) is less than 1.10.

As reported in Table 14, the probit analysis indicates that the spatial distribution of the RT-qPCR data is more consistent with hypothesis A than with hypothesis Z . The probability of a positive assay reading is not significantly related to the distance from raccoon dog vendors: The 95% coverage interval for that coefficient spans both positive and negative values. The Bayes factor indicates odds of about 3:1 in favor of hypothesis A relative to hypothesis Z .

Similar results were obtained using several alternative formulations of the probabilistic framework: (i) measuring each shop's distance from a specific vendor (located on Street 6, stalls 29 to 33) where the China CDC found the highest levels of raccoon dog mtDNA as well as positive RT-qPCR results; (ii) the inclusion of a binary indicator to distinguish the effects at shops in the East vs. West buildings; (iii) alternative specifications of the prior density; (iv) the use of logit rather than probit; and (v) estimation of the model using maximum likelihood rather than Bayesian MCMC. Detailed results are provided in the Supplementary Materials.

These results are markedly different from the conclusions of Worobey et al. (2022), which identified the wildlife mammal vendors at the Huanan Market as the epicenter of the COVID-19 pandemic.²¹ However, that study only considered the frequency of positive RT-qPCR readings and did not incorporate any information from negative readings. Consequently, its conclusions essentially reflected the extent to which the China CDC's sampling concentrated on the area of the market where wildlife shops were located. That sampling bias is mitigated by assessing the relative proportion of positive vs. negative assay readings. In fact, as shown in Figure 16, the spatial distribution of the relative incidence of positive RT-qPCR readings was *not* centered on the market shops where raccoon dogs and other wildlife mammals were being sold.

Nonetheless, a Bayes factor of about 3.2 cannot be viewed as decisive in weighing the two competing hypotheses. As indicated in Table 2 above, a Bayes factor of this magnitude should merely be viewed as equivalent to “anecdotal” evidence in favor of hypothesis *A*, reflecting the fact that this analysis involves a relatively small set of datapoints, namely, the 135 assay readings obtained after the Huanan Market was closed. In effect, the spatial analysis of the RT-qPCR data draws on far less information than the spatiotemporal analysis of confirmed vendor cases.

5.8 Spatiotemporal Analysis of Vendor Cases

The spatiotemporal analysis of vendor case data is conducted using a Bayesian framework in which the probability of a new case occurring at a specific shop at a particular time period is associated with the spatial and temporal characteristics of that location, e.g., its distance from prior vendor cases or wildlife mammal vendors. A total of 30 vendors were confirmed COVID-19 cases, including three vendors at a shop near the center of the west wing and two vendors at another shop in the northeast corner of that building.²⁰ The analysis reported here only includes the earliest case for each of those clusters, because the statistical framework is not well suited for modelling viral spread among vendors within a single market shop.

This analysis is conducted using digitized data from the maps in the Annex of the WHO 2021 Report, which showed the shops at which confirmed vendor cases had occurred as of four specific dates: 13 December, 20 December, 27 December, and 31 December 2019.²⁰ Thus, the temporal dimension of the analysis comprises those four time periods, and the spatial dimension consists of 735 spatially distinct shops at which vendors were employed. The market had a total of 678 operators, each of whom managed a primary shop that often spanned multiple contiguous stalls, and some operators also managed an annex shop that was not contiguous to their primary shop; in this analysis, each of those annexes is treated as a distinct shop.²⁰ The analysis excludes 49 vacant stalls, nearly all of which were in the east wing of the market.²¹ See the Supplementary Materials for further details.

As with the data on early COVID-19 cases in Wuhan, both positive and negative observations may provide significant information, and hence the probit model is well-suited for analyzing these data. In particular, the two competing hypotheses can be distinguished as follows:

- *Hypothesis A.* Under this hypothesis, the SARS-CoV-2 virus “*came into the market before it came out of the market.*” Consequently, the early vendor cases would have been infected elsewhere, such as a restaurant on the fringes of the Huanan Market. Subsequent vendor cases would reflect contagion from other local residents (given that many vendors lived in that neighborhood) or transmission among vendors within the market itself. As in standard models of infectious disease transmission, the likelihood of such infections would be highest for vendors in shops at shorter distances from those of prior vendor cases.
- *Hypothesis Z.* Under this hypothesis, the “*sustained presence*” of infected wildlife mammals at the Huanan Market was the source from which the SARS-CoV-2 virus was transmitted into the human population.²¹ Consequently, the spread of the virus within the market would be initially concentrated in the vicinity of those shops and then spread outwards to vendors in other parts of the market. This spatiotemporal pattern would be most apparent for the earliest cohorts of vendor cases and then wane as the incidence of market infections subsided in late December.

In particular, the implications of Hypothesis *A* can be expressed as follows:

$$(9) \quad \text{Prob}[y_{it} = 1] \sim f_{A,t}(\text{dist_nearest_case}_{i,t-1})$$

where the discrete dependent variable $y_{it} = 1$ if a vendor at shop i had onset of COVID-19 symptoms during time period t and $y_{it} = 0$ otherwise. The principal explanatory variable is the distance of shop i from the nearest shop at which a vendor had onset of COVID-19 symptoms in the prior period $t-1$. The probit function $f_{A,t}$ has a time subscript that denotes the inclusion of time-specific intercepts and slope coefficients. The time-specific intercepts allow for temporal variation in the market-wide probability of new cases, which would presumably reflect the rising incidence of cases in the local community to whom market vendors might be exposed. The time-specific slopes allow for temporal variation in the transmission of the virus from vendors working at nearby shops.

Conversely, the implications of Hypothesis *Z* can be expressed as follows:

$$(10) \quad \text{Prob}[y_{it} = 1] \sim f_{Z,t}(\text{dist_nearest_raccoon_dog_seller}_i)$$

For this hypothesis, the principal explanatory variable is the distance of shop i from the nearest shop where raccoon dogs were sold. This probit function $f_{Z,t}$ also includes time-specific intercepts and slope coefficients. The time-specific intercepts allow for temporal variation in the market-wide probability of new cases, which might simply reflect the relative proportions of market vendors who were susceptible to infection as distinct from those who had already been infected by the virus. The time-specific slopes allow for temporal variation in the ongoing transmission of the virus owing to the sustained presence of infected animals at the market.

Table 15: Probit Analysis of Confirmed Vendor Cases at the Huanan Market

Model	Gelman-Rubin Convergence Criterion (R_c)	Distance Coefficient			Deviance Information Criterion	log Marginal Likelihood	Bayes Factor
		Indicator	Estimate	95% CI			
Hypothesis A (accidental lab leak)	1.003	distance to nearest prior case	-12.1	-26.6 to -1.95	273.5	-125.45	11.95
Hypothesis Z (zoonotic spillover)	1.002	distance to nearest raccoon dog seller	-5.41	-13.8 to 1.39	274.6	-127.92	

Note: This table reports on probit analysis of confirmed COVID-19 vendor cases at the Huanan Market in Dec. 2019. Under Hypothesis A, the probit model includes each shop's distance to the nearest case with symptom onset occurred during the preceding time period. Under Hypothesis Z, the probit model includes each shop's distance to the nearest shop where raccoon dogs were sold. Each model is estimated using 2,105 observations (735 shops x 3 time periods). For each hypothesis, the second column indicates the maximum value of the Gelman-Rubin convergence criterion (R_c) across the parameters of that model; the third column indicates the distance indicator; the fourth and fifth columns report the coefficient estimate and 95% coverage interval for that distance indicator; the sixth column indicates the Deviance Information Criterion (DIC), which is minimized for the preferred model; the seventh column indicates the log marginal likelihood of each model; and the final column indicates the Bayes factor; i.e., the odds for Model A relative to Model Z. Both models incorporate time-specific effects and time-specific slopes. See the Supplementary Materials for further details.

As reported in Table 15, the probit analysis indicates that the spatiotemporal distribution of confirmed vendor cases is more consistent with hypothesis A than with hypothesis Z. The probability of a confirmed case at any particular market shop is inversely related to that shop's distance from the nearest case in the preceding time period; the 95% coverage interval for that coefficient is relatively wide but excludes zero. By contrast, the probability of a confirmed case at any particular shop is *not* significantly related to the distance from raccoon dog vendors, i.e., the 95% coverage interval for that distance coefficient spans both positive and negative values. The Bayes factor indicates odds of about 12:1 for hypothesis A relative to hypothesis Z; as indicated in Table 2, this result is interpretable as “strong evidence” for hypothesis A.

Similar results were obtained using several alternative formulations of the statistical framework: (i) measuring each shop's distance from the specific shop at the market where the China CDC found the highest levels of raccoon dog mtDNA; (ii) alternative specifications of the prior density; (iii) the use of logit rather than probit; and (iv) estimation using maximum likelihood rather than Bayesian MCMC. See the Supplementary Materials for details.

These results are starkly different from the conclusions of Worobey et al. (2022), which identified the wildlife shops at the Huanan Market as the epicenter of the COVID-19 pandemic.²¹ That study followed a deterministic approach in estimating the centroid of cumulative vendor cases at the end of December 2019 but did not incorporate any information about the temporal evolution of cases. Evidently, the cumulative spatial pattern was strongly influenced by the cases in late December and did not account for the fact that none of the earlier cases (i.e., symptom onset by 20 December) were in the vicinity of shops selling live wildlife mammals.

Figure 17: Spatiotemporal Analysis of Vendor Cases by Time Interval of Symptom Onset

Hypothesis A: Accidental Lab Leak

Hypothesis Z: Zoonotic Spillover

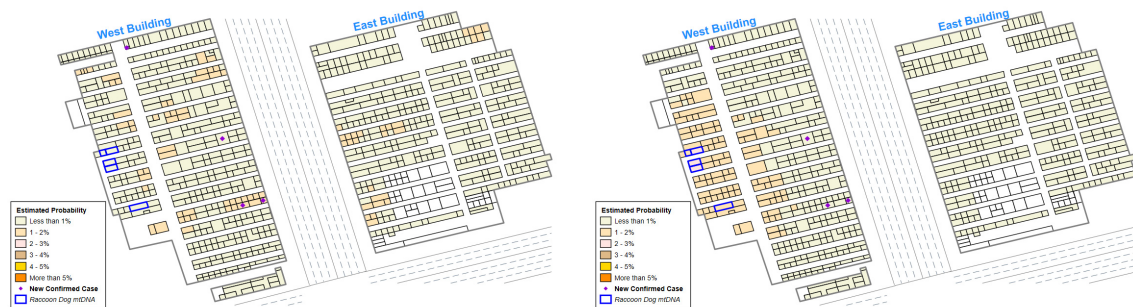
14-20 December 2019



21-27 December 2019



28-31 December 2019



Note: This figure compares the implications of Hypothesis A (left column) and Hypothesis Z (right column) for the spatiotemporal distribution of confirmed vendor cases at the Huanan Market, using the estimation results from the statistical models described in Table 13. In each panel, the shading of each market shop indicates that model's estimated probability of a new confirmed vendor case occurring at that shop during that time interval, and each diamond indicates that a new confirmed case actually occurred at that shop during that time interval. Shops with significant quantities of raccoon dog mtDNA are outlined in dark blue.

Table 16: Bayesian Assessment of the Origins of Two Pandemics

	SARS Pandemic	COVID-19 Pandemic
Prior Odds of Accidental Lab Leak vs. Zoonotic Spillover	1 in 100	1 to 1
Conditional Odds of Pandemic Outbreak in PRC	1 in 100	2.3 to 1
Conditional Odds of Epicenter in Specified Location	1 in 100 (Guangdong)	20 to 1 (Wuhan)
Conditional Odds of Observed Spatiotemporal Pattern of Cases	1 in 100	324 to 1
Posterior Odds of Accidental Lab Leak vs. Zoonotic Spillover	1 in 100 million	14,900 to 1

Note: This table provides a synopsis of Bayesian assessments of the origins of two pandemics, specified in terms of the odds of accidental lab leak vs. zoonotic spillover. The origins of the SARS pandemic are discussed in the text of section 6. For the COVID-19 pandemic, the prior odds are characterized in section 1.6; the odds of an outbreak in PRC are characterized in section 2.3; the conditional odds of the outbreak occurring in Wuhan are characterized in section 3.3; and the conditional odds for the observed spatiotemporal pattern of confirmed cases is given by the product of the results given in Tables 10 and 15, i.e., the spatiotemporal pattern of cases in the core Wuhan area with no known link to the Huanan Market, and the spatiotemporal pattern of vendor cases at Huanan Market shops.

6. Synthesis and Discussion

Since the onset of the pandemic, there have been recurring calls for scientific investigation of the origins of the SARS-CoV-2 virus.^{135, 198, 213, 461-465} In early 2021, the director-general of the WHO underscored the ethical imperative of identifying the origins of COVID-19, for the sake of those stricken by the disease as well as mitigating the risk of a recurrence.⁴⁶⁶ However, subsequent progress towards that goal has been hampered by the absence of a coherent statistical framework for addressing this intrinsically complex question.

Thus, a fundamental objective of this study has been the formulation of a Bayesian framework that can be used to assess the relative weight of the evidence for the two key hypotheses, that is, accidental lab leak vs. zoonotic spillover. This approach facilitates a direct comparison of these two hypotheses, as distinct from classical procedures that focus on testing a particular null hypothesis without specifying any particular alternative. Moreover, the various dimensions of the data can be represented in terms of several conditional Bayes factors, thereby facilitating the transparency of the statistical evaluation. The overall weight of the evidence is expressed by the posterior odds ratio, which is the product of the conditional Bayes factors multiplied by the prior assessment of odds.

Of course, Bayesian analysis does not necessarily provide a definitive conclusion; the results will be inconclusive if the available evidence is not sufficient to distinguish between competing hypotheses. For example, some data might provide support for one hypothesis whereas the rest of the data points to a different hypothesis. In that case, some conditional Bayes factors would be greater than 1 while the remaining factors would be less than 1, and hence the combined product of these factors would be close

to unity, i.e., the evidence was not very informative, and hence the posterior odds would be roughly the same as the prior odds—an outcome very different from the findings of this paper.

The statistical framework used in this study has decomposed the overall Bayes factor into four conditional components: (1) the likelihood that PRC would be the epicenter of the pandemic; (2) the likelihood that the epicenter would be located in Wuhan, conditional on its occurrence in PRC; (3) the likelihood of observing the spatiotemporal pattern of confirmed COVID-19 cases with no known link to the Huanan Market, conditional on the outbreak taking place in Wuhan; and (4) the likelihood of observing the spatiotemporal pattern of confirmed vendor cases within the Huanan market itself.

The results of this analysis are shown in Table 16. The posterior odds of about 14,900:1 indicate an overwhelming degree of support for the hypothesis that the COVID-19 pandemic originated from an accidental lab leak rather than zoonotic spillover. Those odds are computed using flat priors (1:1), but the results would decisively support the accidental lab leak hypothesis even if the prior odds were strongly tilted towards zoonotic spillover.

Moreover, this conclusion is not sensitive to the specific details of the statistical analysis, because all four conditional Bayes factors favor the hypothesis of an accidental lab leak. The product of the first two factors is 46:1, i.e., the fact that Wuhan was the epicenter of the pandemic is interpretable as “*very strong*” evidence for the accidental lab leak hypothesis. The product of the last two factors is 324:1, indicating that the spatiotemporal pattern of unlinked and vendor cases is interpretable as “*extremely strong*” evidence for this hypothesis. In effect, the coherence of all four conditional Bayes factors may be viewed as providing further confirmation for the robustness of this conclusion, analogous to the adage that “*the whole may be greater than the sum of its parts.*”

6.1 Comparison of the SARS vs. COVID-19 Pandemics

As shown in Table 16, this Bayesian framework can also be used to assess the origins of the SARS pandemic. Such an assessment clearly shows that this framework is not intrinsically skewed towards either hypothesis and hence serves as a useful cross-validation of the analysis as well as highlighting some key differences between the SARS outbreak and the COVID-19 pandemic.

Prior Odds. Prior to the outbreak of SARS in 2002, only two strains of human coronavirus had been identified (HCoV-229E and HCoV-OC43), both of which are generally associated with the mild symptoms of a common cold.⁴⁶⁷ A complete genetic sequence had only been produced for one of those coronaviruses (HCoV-229E), which was later classified as an alphacoronavirus whose genome differs markedly from that of betacoronaviruses such as SARS and whose lineage likely diverged about 5000 years ago.²⁷⁴ Moreover, genetic analysis of viral RNA was still costly and highly labor-intensive, requiring reverse transcription into complementary DNA, insertion into vector plasmids, and replication in bacterial cultures.⁴⁶⁸⁻⁴⁷⁰ Thus, at around the turn of the century, the risk of a catastrophic lab leak was largely associated with research on existing diseases or toxic substances, not the production of a novel viral chimera.

By contrast, during the decade of the 2010s genetic analysis was revolutionized by the CRISPR-Cas9 gene editing system, whose creators received the Nobel Prize in Chemistry in 2020.⁴⁷¹ This system provides a simple, precise, and cost-effective method of inserting or deleting nucleotide sequences at specific positions within a genome and is commonly referred to as “genetic scissors.”^{472, 473} CRISPR-Cas9 can be used directly with DNA viruses such as HIV and hepatitis C and can be applied

to complementary DNA of RNA viruses.⁴⁷⁴⁻⁴⁷⁷ Alternative methods such as CRISPR-Cas13 were developed to facilitate the direct editing of RNA sequences.⁴⁷⁸ By 2019, researchers had produced chimera from human and animal strains of coronavirus, including HCov-229E and MERS.^{261, 479-481} Thus, the marginal likelihood of catastrophic lab leak would plausible be far higher at that time than at the turn of the century, whereas the likelihood of zoonotic spillover might not have changed markedly over that timeframe. Consequently, the prior odds would strongly tilted towards a zoonotic source for the SARS pandemic, whereas flat priors may be reasonable as a baseline specification for analyzing the COVID-19 pandemic.

Location in PRC. At the time of the SARS outbreak, the PRC had not yet emerged as a leader in research on bat viruses. The PubMed database indicates that the term “coronavirus” was used in 1,975 papers published between January 1990 and December 2002, but only two of those papers were coauthored by individuals affiliated with PRC institutions; see the Supplementary Materials for further details.²¹⁷ Thus, even if a catastrophic lab leak had occurred at around that time, the marginal likelihood of its occurrence in PRC would have been miniscule.

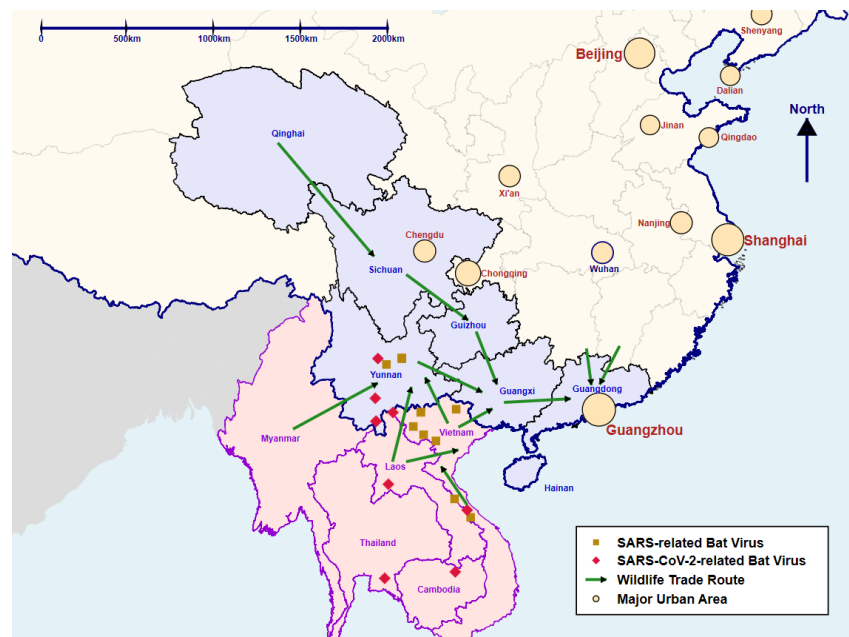
In contrast, scientific research on coronaviruses burgeoned worldwide and in the PRC in the wake of SARS. The PubMed database indicates that the term “coronavirus” was used in 9,533 papers published between January 2003 and December 2019, of which 2,020 papers (about 22%) were coauthored by researchers with PRC affiliations.²¹⁷ In early 2018 the WHO’s blueprint for R&D prioritization identified a novel “Disease X”, characterized as a virus most likely to emerge from a zoonotic source.⁴⁸²⁻⁴⁸⁶ During 2019 PRC authorities urged research labs to intensify their efforts along these lines, even though such research was generally being conducted under limited biosafety standards (BSL-2). Thus, the marginal likelihood of an accidental lab leak occurring in PRC was far higher than at the time of the SARS outbreak, whereas there was no notable change in the marginal likelihood of a zoonotic outbreak in PRC. As indicated in Table 16, the conditional odds of an accidental lab leak shifted from 1:100 for SARS (“extremely unlikely”) to about 2.3:1 for COVID-19.

Specific Location of Outbreak. As of 2002, Guangdong Province had no major biological research institutions; its state key laboratories in genomics and respiratory diseases were not established until 2007.⁴⁸⁷ Thus, at the time of the SARS outbreak, the marginal likelihood of a catastrophic lab leak in Guangdong was miniscule relative to other PRC locations with more advanced research facilities.

By contrast, in 2019 Wuhan was a paragon of bat viral research, but major labs elsewhere in PRC were also engaged in such research.⁴⁸⁸ Experiments with chimera viruses were typically conducted at modest biosafety standards (BSL-2) that were comparable to those of a dental office.^{157, 159, 161, 243} In 2019 WIV officials specifically prioritized research on a “*transmissible infectious disease caused by a novel coronavirus originated from bats*” and indicated that the WHO blueprint for Disease X had “*prompted us to accelerate our efforts...against divergent human coronaviruses and SARS-related coronaviruses.*”⁴⁸⁹ Thus, under the hypothesis of an catastrophic lab leak in PRC during late 2019, the foregoing analysis indicates a marginal likelihood of 20% that it would be located in Wuhan.

As for the marginal likelihood of zoonotic spillover, the trade and consumption of wildlife mammals in Guangdong was characterized as “*booming*” over the period preceding the SARS pandemic.⁹⁸ Masked palm civets were the most popular species of exotic mammal sold at restaurants in southern China; about 500 civets were observed for sale at a single market in Guangzhou in late 2001.⁹⁸ The provincial authorities had recently begun tightening the enforcement of wildlife regulations,

Figure 18: Locations of SARS-Related Bat Viruses and Wildlife Trade Routes



Note: This map shows wildlife trade routes, locations of SARS-related bat viruses, and major urban areas in PRC. Specifically, the squares and diamonds denote specific locations in PRC and southeast Asia where researchers have collected bat specimens and identified viruses with very high genetic similarity to SARS-CoV and SARS-CoV-2, respectively. The green arrows denote the principal domestic and cross-border routes used for transporting live wildlife animals for human consumption. The circles denote urban areas with population ≥ 5 million in 2020, using U.N. Population Division estimates, and the radius of each circle denotes its relative population size. See the Supplementary Materials for further details.

but conditions at such markets were still characterized as “appalling”, with animals packed into tiny spaces with no basic care for open wounds or infections.⁹⁸

Figure 18 shows the domestic and cross-border trade routes that were used for transporting live wildlife animals to southern China from the early 1990s through 2019; all of those wildlife trade routes culminated in Guangdong.^{97, 98, 310, 490-494} The figure also indicates the specific sites in Yunnan and in northern Vietnam where researchers have identified bat viruses with high genetic similarity to the SARS-CoV virus.^{108, 110, 247, 495} Notably, all of those sites are very close to major wildlife trade routes, and the Vietnamese border is only about 500km from the Guangdong municipalities where the earliest SARS cases were identified.

Consequently, the marginal likelihood of a zoonotic spillover in southern China was non-negligible in the early 2000s and surely far greater than the marginal likelihood of an accidental lab leak. Thus, Table 16 characterizes this conditional odds ratio for SARS as 1:100, i.e., the scenario of a lab leak in southern China was “extremely unlikely” relative to the hypothesis of a zoonotic spillover.

By contrast, as of fall 2019 only small numbers of wildlife mammals were being sold at Wuhan markets, with total citywide sales per day averaging about 11 hedgehogs, 5 hares, 1 bamboo rat, 1 red fox, and 1 raccoon dog, with even less frequent sales of other exotic mammal species.²⁷ Moreover,

as shown in Figure 18, Wuhan did not serve as a hub or waypoint for any of the major wildlife trade routes that prevailed from the 1990s until the onset of the COVID-19 pandemic.

Indeed, there is no apparent reason that the likelihood of a zoonotic spillover in Wuhan in late 2019 would have been any higher than in any other location in PRC, and hence this likelihood may reasonably be judged at around 1%, i.e., the ratio of Wuhan's population relative to the national total. This marginal likelihood is markedly smaller than that of a catastrophic lab leak in Wuhan, and hence this conditional odds ratio is assessed at about 20:1 for the COVID-19 pandemic—nearly the inverse of the corresponding odds ratio for SARS.

This assessment is underscored by the geospatial data shown in Figure 18. The bat viruses with high genetic similarity to SARS-CoV-2 have been identified from specimens collected in Yunnan province and from other southeast Asian locations.^{127, 132, 203, 495-497} Those studies indicate that a bat virus from the Vientiane province of Laos is the closest known relative to SARS-CoV-2.⁴⁹⁸⁻⁵⁰⁰ However, but the distance from that location to Wuhan is nearly 2000km. If the zoonotic spillover occurred there, then why were the earliest confirmed cases of COVID-19 identified in Wuhan rather than in Guangdong or some other major urban center of PRC?

Spatiotemporal Distribution of Cases. During the SARS pandemic, the pattern of confirmed cases pointed clearly to a zoonotic spillover. For example, the earliest cases included a cook, waiter, and customer at a restaurant where palm civets were held in wire cages prior to being consumed.⁹⁶ Restaurant chefs and others handling wildlife animals for human consumption accounted for 39% of confirmed SARS cases.⁹² A serology study found anti-SARS antibodies in 40% of wildlife animal traders and only 5% of vegetable vendors.²³ Live virus was identified in 100% of samples collected from palm civets and raccoon dogs at a particular food market, with a 99.8% genetic match to samples of SARS from human patients. Independent outbreaks occurred at multiple locations in late 2002 and early 2003, consistent with repeated instances of zoonotic spillover.⁹² Subsequent analysis identified a cluster of asymptomatic SARS-like infections in spring 2001, indicating that viral strains had undergone substantial evolution prior to the onset of the pandemic.⁹³

By contrast, none of those patterns were evident at the onset of the COVID-19 pandemic. The Wuhan CDC identified 176 early cases (symptom onset by 31 December 2019), including 5 farmers and 30 vendors at the Huanan Market, none of whom were engaged in the sale of wildlife mammals. During the period of 7-18 January 2020, the China CDC collected tissue samples from raccoon dogs and other mammals (including bats) that were captured in the rural areas of Wuhan by local traders who had previously been trapping them for sale at the Huanan Market, but SARS-CoV-2 was not found in any of those samples.³⁵⁰ Subsequent surveillance did not find SARS-CoV-2 in any farmed or wild animals anywhere in PRC.^{322, 325, 501} Spatiotemporal analysis of the case data indicates a conditional odds ratio of 369:1, which can be interpreted as “extremely strong” evidence in favor of the accidental lab leak hypothesis relative to the hypothesis of zoonotic spillover.

Table 17: Comparing the Receptor Binding Domain in SARS-CoV-2 vs. Two Bat Viruses

	<i>Position of Codon in Spike Gene</i>											
	448	453	458	463	468	473	478	483	488	493	498	503
SARS-CoV-2	NYNYL	YRLFR	KSNLK	PFERD	ISTEI	YQAGS	TPCNG	VEGFN	CYFPL	QSYGF	QPTNG	VGYPQ
BANAL-52 (Vientiane, Laos)	•••••	•••••	•••••	•••••	•••••	•••••	•••••	•••••	•••••	•••••	•••••	H•••••
RaTG13 (Yunnan, PRC)	•F•••	•••••	•A•••	•••••	•••••	•••••	•••••	K••••	QT•L•	•Y••	YR•••	Y••D••H••

Note: This table indicates the sequence of amino acids for the core region of the receptor binding domain (RBD) of the ancestral strain of SARS-CoV-2 (sequence Wuhan-Hu-1) compared to two bat viruses. RaTG13 was collected in Yunnan Province in 2013 (Genbank ID: MN996532), and BANAL-52 was collected in the Vientiane Province of Laos in 2020 (Genbank ID: MZ937000). Each column refers to a specific position in the spike gene of SARS-CoV-2, and each letter denotes the codon for a specific amino acid (N = asparagine, Y = tyrosine, etc.). For SARS-CoV-2, the green bold letters denote the amino acids in the RBD that have been identified as critical for binding to the human ACE-2 receptor. For each bat virus, a bullet (•) indicates that the amino acid at that position is identical to that of SARS-CoV-2, whereas a letter indicates a different amino acid at that position, and bold purple denotes differences at positions that are critical for the functioning of the RBD in SARS-CoV-2.

6.2 Limitations of the Analysis

The analysis in this paper is subject to a number of substantive limitations, each of which warrants further research that will provide additional clarity about the origins of the COVID-19 pandemic.

Other Types of Viruses. Both of the specific origin hypotheses considered in this paper are framed in terms of a bat-related coronavirus, reflecting the fact that the genome of SARS-CoV-2 is highly similar to other strains of betacoronavirus that originated in bats. Thus, the statistical analysis in Sections 2 and 3 has been focused on assessing the geospatial distribution of bat families and locations of bat viral research laboratories. Nonetheless, this focus should not be interpreted as downplaying the risk that the next pandemic could involve a different type of virus such as influenza. Indeed, as noted above, the Spanish flu was the most catastrophic pandemic of the 20th century.

Genetic Analysis. The Bayesian framework of this paper has focused on assessing conditional odds using spatiotemporal and zoonotic data, whereas the implications of genetic analysis have been confined to the specification of the prior odds. However, Bayesian analysis can also be applied to genetic data to elucidate the odds of the extraordinary characteristics of SARS-CoV-2 arising under each of the two origin hypotheses, i.e., the accidental leak of a lab-produced chimera vs. zoonotic transmission from a host animal. Several distinct sources of data are relevant for such analysis:

- *Comparisons to other Bat Viruses.* In early 2020, WIV reported that SARS-CoV-2 exhibited similarity of 96.1% to a bat virus (RaTG13) that had been collected in 2013 at a location in Yunnan province.¹²⁷ From July 2020 to January 2021, an international team collected and sequenced bat viruses from several locations in Laos and identified one virus (BANAL-52) with somewhat higher similarity of 96.8% to SARS-CoV-2.⁴⁹⁶ Moreover, as shown in Table 17, the core region of the RBD of BANAL-52 virus is nearly identical to that of SARS-CoV-2, whereas the corresponding region of RaTG13 diverges at five of the six critical positions.^{498, 502}

Comprehensive surveys of bat viruses have not identified any viruses anywhere in PRC with greater similarity to SARS-CoV-2.^{233, 278}

- *Pre-Pandemic Evolution.* Researchers have compared the genetic sequence of SARS-CoV-2 to a wide range of bat viruses with the aim of identifying point mutations and recombination events and estimating the timing of its evolutionary divergence from other bat viruses.^{283, 503-507} One such study used Bayesian methods and concluded that the SARS-CoV-2 virus diverged from RaTG13 several decades ago.¹⁰⁷ However, such analysis has generally been conducted under the assumption of a zoonotic origin, but such approaches should be used to assess and compare the odds of both origin hypotheses.^{239, 495, 508-510}
- *SARS-CoV-2 Evolution Since 2020.* The SARS-CoV-2 virus spread worldwide during 2020-21 and infected more than 1 billion people by the time that the Omicron variant was identified in November 2021; the virus likely underwent about 50 billion replication cycles over that time interval.⁵¹¹⁻⁵¹⁴ Genetic analysis of omicron variant BA.2.76 found similarity of 99.7% to the ancestral strain of SARS-CoV-2 (Wuhan-Hu-1), with nearly all of the mutations involved a single nucleotide substitution or deletion, and none involving more than two consecutive nucleotides.⁵¹⁵ This finding underscores the efficacy of the coronaviral proofreading system in minimizing replication errors but may also reflect the extent to which the ancestral strain of SARS-CoV-2 was already very well adapted to humans.⁵¹⁶⁻⁵¹⁹
- *Bat Viral Genomic Data.* Prior to the pandemic, WIV maintained a password-protected database of about 16,000 samples of bat viruses, most of which had been collected and sequenced by WIV researchers for work that had not yet been published, and hence those genetic sequences had not yet been uploaded to any public databases such as Genbank or GISAID; this WIV database was taken offline in early 2020 due to concerns about potential security breaches.^{186, 244, 246, 254}

The China CDC has released detailed information about its environmental sampling at the Huanan Market, including all of the raw genomic data from its high-throughput processing and analysis. Likewise, comprehensive surveys of bat species and bat-related viruses in PRC have been published in international journals. Along similar lines, it would now be ideal for PRC authorities to disseminate all of the pre-pandemic genomic data that was stored or produced at Wuhan research labs, which could facilitate further systematic assessment of both origin hypotheses.^{520, 521}

Nonetheless, genomic data may not provide definitive conclusions about the origins of SARS-CoV-2. It is evident that the virulence of this virus was not optimized using complex numerical algorithms or extensive serial passaging in cell cultures.¹² If SARS-CoV-2 was indeed created as an artificial chimera, then it may have been produced in lab experiments conducted by a single graduate student or postdoctoral researcher.^{522, 523} Such research could be conducted using routine methods and off-the-shelf lab materials in conjunction with the genomic data on bat viruses (such as RaTG13) that were stored in internal WIV databases. Such experiments could have been conducted with little or no supervision and practically no electronic records apart from a single workstation.

Moreover, if SARS-CoV-2 originated in a research laboratory, then the initial infections would almost certainly have been invisible, because severe SARS-CoV-2 infections are extremely rare for healthy younger adults (ages 20-39).^{5, 366, 368} Once the virus began spreading more widely, all evidence of its

origins would have been promptly deleted so that WIV officials and other investigators could find no indication of its existence in any internal WIV databases.²⁴³⁻²⁴⁶

COVID-19 Case Data. The statistical analysis in Sections 4 and 5 uses data on confirmed COVID-19 cases that were identified by the Wuhan CDC. One notable limitation of those data is that all of the confirmed cases with symptom onset in December 2019 were hospitalized patients, because RT-PCR tests for the live SARS-CoV-2 virus were not yet available.³⁶⁵ Consequently, the probit analysis reported here only elucidates the spatiotemporal characteristics of severe COVID-19 cases that required hospitalization. However, a high proportion of SARS-CoV-2 infections are mild or asymptomatic, and hence it is likely that several thousand Wuhan residents were infected with the virus by the end of 2019.^{20, 363-365, 378, 385} Thus, it would be ideal to extend this analytical framework to provide further insights about the spatiotemporal pattern of all early SARS-CoV-2 infections, not just the most severe cases. Moreover, the PRC authorities could facilitate such research by releasing de-identified data for all confirmed cases with symptom onset in 2019, including each individual's age, sex, and location of residence and employment.

The COVID-19 case data may also exhibit underascertainment of severe COVID-19 cases, notwithstanding the Wuhan CDC's intensive review of hospital records and other information. As in many locations worldwide, such underascertainment could reflect instances in which an individual succumbed to COVID-19 without having been hospitalized and with no post-mortem analysis.⁷ Bayesian analysis of serum samples collected in spring 2020 indicates that Wuhan's infection fatality ratio (IFR) was 0.23% (CI: 0.19-0.26%), and epidemiology studies indicate that about two-thirds of COVID-19 fatalities occurred within 14 days of symptom onset during the first wave of the pandemic.^{5, 7, 524, 525} Based on those data, there may well have been 50+ COVID-19 fatalities in Wuhan during December 2019, only a few of whom were hospitalized and hence considered by the Wuhan CDC's epidemiological investigation.

Nonetheless, these limitations of the COVID-19 case data may not have material implications for the key findings of this paper. The spatiotemporal analysis in Section 4 indicates that the incidence of early cases on the southeast bank of Yangtze River – the area where bat viral research laboratories were located – was markedly higher than what would be predicted solely from population density and distance from the dense urban core of Wuhan, and that result might well be strengthened by the inclusion of additional cases that were not considered during the Wuhan CDC's investigation. Likewise, the findings of Section 5 would be unlikely to be overturned by further information about vendor cases at the Huanan Market, because those cases were scrutinized by the Wuhan CDC and in the China CDC's analysis.²⁰

Zoonotic Hosts. The genetic data from the China CDC's environmental surveillance is invaluable for identifying the location of shops at the Huanan Market where wildlife mammals were sold.²⁸ However, such data do not indicate the number of animals of each species that were being sold by each individual shop during the final months of 2019. Ironically, such information was collected for a monthly survey of wildlife sales at Wuhan markets, but that study only reported citywide averages over the full survey period (May 2017 to November 2019), and its supplementary materials did not include any market-specific or chronological data.²⁷ With the passing of nearly four years after the publication of that article, it would be ideal for its authors to now post all of the relevant underlying data. It would also be helpful

to conduct a systematic study of viral transmission in each of those species using the ancestral strain of the virus (Wuhan-Hu-1).⁵²⁶

Alternative Locations of Zoonotic Spillover. The analysis in Sections 4 and 5 above has focused on a zoonotic spillover from wildlife mammals at the Huanan Market. However, further research could consider alternative scenarios in which SARS-CoV-2 originated elsewhere and was promptly transmitted to residents in Wuhan, which would then have emerged as the location of the earliest confirmed cases even though Wuhan was not actually the site of zoonotic origin. Several such locations could be considered as plausible alternatives:

- *Yunnan Province.* In 2015, WIV researchers collected serum samples from 218 individuals living near the Yunnan caves where bat specimens were being collected, and a bat-related viral strain was identified in six of those samples (2.7%) but not in any samples from Wuhan residents (who served as the control group).^{104, 108, 109, 250} Unlike SARS or SARS-CoV-2, that particular viral strain does not bind to the human ACE2 receptor, but the findings confirmed that individuals living near bat caves have elevated levels of exposure to bat-related viruses. However, experimental studies and phylogenetic analysis have uniformly concluded that bats could not have been the direct zoonotic source of SARS-CoV-2.^{178, 279-283} Consequently, this scenario raises a host of difficult questions. If a Yunnan resident had in fact been “patient zero”, what was the intermediate animal host? Why was the virus transmitted to Wuhan instead of some other location in PRC? And how was it transmitted to Wuhan (a distance exceeding 1500km) without leaving any traces of its origin? An infected individual from rural Yunnan would presumably have traveled to Wuhan on an intercity bus or train originating in Kunming (the province’s capital and primary transportation hub) but there were no confirmed COVID-19 cases in Kunming or anywhere else in Yunnan prior to 21 January 2020.⁵²⁷ Moreover, such modes of transportation would be associated with a high rate of viral transmission to other passengers and hence widespread dispersion of SARS-CoV-2 across southern and central PRC, rather than the observed pattern in which all of the early cases were located in Wuhan itself.
- *Guangdong Province.* During 2009-2013 researchers conducted surveillance in 12 Guangdong districts where large numbers of wildlife animals were being sold at markets and restaurants.⁵²⁸ About two-thirds of the study participants reported having eaten wild animals, and nearly half were employed in occupations involving meat preparation (hunters, butchers, and restaurant workers); however, the incidence of SARS antibodies in each of those subgroups was not statistically distinguishable from that of the general population, while antibodies to rodent-related viruses (hanta and bunya) were somewhat more common in butchers than in other occupational categories.⁵²⁸ Thus, these results do not indicate an elevated risk of bat-related viruses in Guangdong during the timeframe that the surveillance was conducted. If the SARS-CoV-2 virus did originate in Guangdong, why was its dissemination markedly different from that of SARS-CoV? Given that Guangdong is highly urbanized and densely populated, it would seem likely that the earliest cases of COVID-19 would have been observed there rather than in Wuhan. Consequently, the marginal likelihood of a zoonotic spillover in Guangdong can reasonably be assessed as even lower than in Yunnan.

- *Southeast Asia.* In many tropical areas of southeast Asia, consumption of wildlife animals (“bushmeat”) is relatively common, and serology studies have found elevated incidence of SARS-related viruses among bushmeat hunters as well as individuals working in bat guano extraction and forestry-related occupations.^{529, 530} Thus, prior to the COVID-19 pandemic, the incidence of bushmeat hunting was identified as a key factor in assessing the global distribution of zoonotic bat viruses.²³⁹ As noted above, the virus (BANAL-52) whose genome is closest to that of SARS-CoV-2 was extracted from bat specimens collected in Vientiane, Laos. However, a serology study conducted in August-September 2020, encompassing Vientiane and four other Laotian provinces, found “no evidence for significant SARS-CoV-2 circulation” as of that date. Moreover, there is no apparent link between Vientiane and Wuhan (apart from bat viral research), and hence it seems very unlikely that SARS-CoV-2 could have originated in Laos and been promptly transmitted to residents in Wuhan without leaving any traces of its origin.

Scientific Epistemology. This study has not provided any systematic characterization of the evolving views of public health officials or researchers regarding the origins of SARS-CoV-2, which would necessarily encompass a wide array of public documents, articles, working papers, interviews, and social media interchanges. Such a systematic review would be a substantive contribution to the field of scientific epistemology and might well have practical implications for pandemic preparedness and crisis management.⁵³¹⁻⁵³⁴

At the onset of the COVID-19 pandemic, China CDC officials proceeded on the premise that the novel virus had been transmitted from wildlife mammals, especially once it was confirmed that the novel virus was closely related to SARS-CoV.^{16, 535} However, those officials subsequently realized that this premise was not consistent with the incoming evidence: (a) SARS-CoV-2 was far more contagious than what would be expected for a virus that had not yet circulated widely among humans; (b) no trace of the virus was found in any animal samples from the Huanan Market; (c) no COVID-19 cases were found in wildlife traders or vendors; and (d) the earliest confirmed cases included a large number of individuals with no known link to the Huanan Market.¹⁶

Thus, in May 2020 the now-retired China CDC director explained: “*At first, we assumed the seafood market might have [produced] the virus, but now the market is more like a victim.*”³⁹⁶ In a subsequent BBC interview, that same official was asked about the scenario of an accidental lab leak and replied: “*Don’t rule out anything.*”⁵³⁶

Although the WHO 2021 Report characterized the hypothesis of an accidental lab leak as “*extremely unlikely*”, the director-general of WHO contradicted that judgment at a press conference in July 2021, stating that “*there was a premature push to especially reduce one of the options like the lab theory...I was a lab technician myself, an immunologist, and I have worked in the lab, and lab accidents happen. It’s common; I have seen it happening and I have myself had errors, so it can happen.*”^{466, 537}

Nonetheless, prominent researchers have downplayed or dismissed the hypothesis of an accidental lab leak in journal articles, editorials, and media interviews.^{14, 135-137, 538-545} For example, in fall 2024 the *Journal of Virology* published an editor-endorsed commentary coauthored by 41 virologists and entitled “*The harms of promoting the lab leak hypothesis for SARS-CoV-2 origins without evidence*” which warned that “*the resulting anti-science movement puts the research community, scientific research, and pandemic preparedness at risk.*”⁵⁴⁶ It remains to be seen whether those views may be reconsidered in light of the findings of the present study.

Public Policy Implications. The foregoing analysis has focused on analyzing the likely origin of the SARS-CoV-2 virus but has not considered any potential public policy implications for the regulation of virology research or wildlife trade. In light of that analysis, however, it is evident that some key factors associated with the risk of zoonotic spillover have been receding notably in recent years, whereas other zoonotic risk factors remain elevated:

- *Culinary Use of Wildlife Mammals.* In February 2020 the PRC adopted a permanent and comprehensive ban on the farming, sale, and consumption of all terrestrial wildlife animals.⁵⁴⁷
- *Fur Farming.* PRC fur farms produced 50 million pelts in 2018, about half of the global total, but their production fell sharply to about 10 million pelts in 2023, roughly similar to the drop in global volume of furs.⁵⁴⁸⁻⁵⁵⁰ The price of a mink pelt (in inflation-adjusted U.S. dollars) fell by about 75% from 2011 to 2023.⁵⁵¹ The European Commission is now actively considering a complete ban on fur farming.⁵⁵² These developments reflect public concerns about humane treatment of animals as well as innovations in producing high-quality “faux fur” substitutes.⁵⁵³
- *Bat Guano.* Small-scale farms use bat guano as an organic fertilizer in some regions of southeast Asia and sub-Saharan Africa.⁵⁵⁴ Apart from potential zoonotic risks, the effectiveness of bat guano is limited by unpredictable composition and rapid decomposition and hence is rarely used by larger and more productive farms.⁵⁵⁵
- *Bushmeat.* No comprehensive global data has quantified the incidence of bushmeat activities (i.e., the hunting and sale of wild animals for human consumption), but systematic reviews indicate that bushmeat is a substantial component of nutritional intake in tropical low-income rural areas in Brazil, sub-Saharan Africa, and southeast Asia.^{556, 557} Policies that foster sustained economic development will be crucial for reducing dependence on bushmeat in those areas.⁵⁵⁸

By contrast, the revolution in genetic editing has continued at an extraordinary pace in recent years. For example, the following innovations were described in peer-reviewed journals during 2024:

- *Click Editing.*⁵⁵⁹ This method enables precise and programmable genome engineering from simple DNA templates. The accompanying *Nature Biotechnology* briefing described it as “well validated” and “akin to clicking and editing the text of a word processing document.”
- *Seamless Insertions.*⁵⁶⁰ The CRISPR-Cas9 method is combined with the single-stranded annealing repair pathway to facilitate the insertion of a genetic sequence at a specific location in the genome without leaving any unwanted markers at the integration site.
- *Customized Coronavirus Receptors.*⁵⁶¹ This article demonstrates that functional viral receptors can be engineered to match the receptor-binding domain (RBD) of a specific coronavirus. Evidently, a similar approach could be used to produce a customized viral RBD that binds to any specified receptor in a host cell.
- *CRISPRkit.*⁵⁶² This kit provides an inexpensive, safe, and user-friendly approach for introducing CRISPR technology to high school students. CRISPR methods are routinely used by undergraduate biology students, some of whom collaborate with graduate students and postdoctoral researchers who are overseen by a more senior principal investigator.

Further enhancements to the precision, ease, and cost-effectiveness of genetic editing will surely continue over the next 5-10 years. Such technical enhancements may even be boosted by more complex algorithms developed with the assistance of artificial intelligence and implemented with quantum computing.

The analysis of this paper indicates extraordinarily high odds that the COVID-19 pandemic resulted from an accidental leak of a viral chimera. But even if there were certainty that SARS-CoV-2 had a zoonotic origin, the risks posed by chimera viruses will surely intensify over coming years. In benign scenarios, such a leak could involve a mildly virulent chimera produced by a biology student without adequate supervision or safeguards. However, far more adverse scenarios—perhaps involving hostile actors working with low-cost equipment and supplies—cannot be dismissed. Moreover, unlike SARS-CoV-2, such a viral chimera would not necessarily be most severe for older adults. The worst pandemic in modern history—the Spanish flu—mainly caused fatalities in younger adults ages 18-39 years.⁶⁸⁻⁷⁰

Conclusion

It is extraordinarily difficult for anyone to be fully objective in considering a catastrophic event that has caused millions of fatalities. Nonetheless, assessing the origins of the SARS-CoV-2 virus is intrinsically a matter of analyzing and weighing observable facts, without influence from ideology or geopolitics or conflicts of interest. However, progress towards this goal has been hampered by the absence of a coherent statistical framework.

Thus, a key contribution of this study is the formulation of a Bayesian framework that can be used to assess the relative weight of the evidence. This framework indicates an overwhelming degree of support for the hypothesis that the COVID-19 pandemic originated from an accidental lab leak rather than zoonotic spillover. This conclusion is not sensitive to the specific details of the analysis, because each of the conditional Bayes factors points in the same direction. Nonetheless, such an important issue cannot and should not be settled by any single academic study; further research is clearly warranted.

References

1. World Health Organization. *WHO COVID-19 Dashboard*. 2024 [cited 2024 17 June 2024]; Available from: <https://data.who.int/dashboards/covid19/deaths?n=c>.
2. The Economist. *The pandemic's true death toll*. 2022 [cited 2024 17 June 2024]; Available from: <https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker>.
3. Msemburi, W., et al., *The WHO estimates of excess mortality associated with the COVID-19 pandemic*. *Nature*, 2023. **613**(7942): p. 130-137.
4. Karlinsky, A. and D. Kobak, *Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset*. *eLife*, 2021. **10**: p. e69336.
5. Levin, A.T., et al., *Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications*. *Eur J Epidemiol*, 2020. **35**(12): p. 1123-1138.
6. Levin, A.T., et al., *COVID-19 prevalence and mortality in longer-term care facilities*. *Eur J Epidemiol*, 2022. **37**(3): p. 227-234.
7. Levin, A.T., et al., *Assessing the burden of COVID-19 in developing countries: systematic review, meta-analysis and public policy implications*. *BMJ Glob Health*, 2022. **7**(5).
8. Mena, G.E., et al., *Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile*. *Science*, 2021. **372**(6545): p. eabg5298.
9. Pugh, S., et al., *A Hierarchical Bayesian Model for Estimating Age-Specific COVID-19 Infection Fatality Rates in Developing Countries*. *Statistics in Medicine*. **n/a**(n/a).
10. Chrimes, T., et al., *The Great Reversal: Prospects, Risks, and Policies in International Development Association (IDA) Countries*. 2024: The World Bank.
11. The World Bank, *Global Economic Prospects (June 2024)*. Available from: <https://www.worldbank.org/en/publication/global-economic-prospects>.
12. Andersen, K.G., et al., *The proximal origin of SARS-CoV-2*. *Nat Med*, 2020. **26**(4): p. 450-452.
13. Liu, Y.-C., R.-L. Kuo, and S.-R. Shih, *COVID-19: The first documented coronavirus pandemic in history*. *Biomedical Journal*, 2020. **43**(4): p. 328-333.
14. Holmes, E.C., et al., *The origins of SARS-CoV-2: A critical review*. *Cell*, 2021. **184**(19): p. 4848-4856.
15. da Silva, S.J.R., et al., *Two Years into the COVID-19 Pandemic: Lessons Learned*. *ACS Infect Dis*, 2022. **8**(9): p. 1758-1814.
16. Yang, D.L., *Wuhan: How the COVID-19 Outbreak in China Spiraled Out of Control*. 2024: Oxford University Press.
17. Huang, C., et al., *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China*. *Lancet*, 2020. **395**(10223): p. 497-506.
18. Ren, L.L., et al., *Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study*. *Chin Med J (Engl)*, 2020. **133**(9): p. 1015-1024.
19. Worobey, M., *Dissecting the early COVID-19 cases in Wuhan*. *Science*, 2021. **374**(6572): p. 1202-1204.

20. World Health Organization, *WHO-convened Global Study of Origins of SARS-CoV-2: China*. 2021.
21. Worobey, M., et al., *The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic*. *Science*, 2022. **377**(6609): p. 951-959.
22. Spinney, L., *Germany's Covid-19 expert: 'For many, I'm the evil guy crippling the economy'*, in *The Guardian*. 2020: London, UK. Available from: <https://www.theguardian.com/world/2020/apr/26/virologist-christian-drosten-germany-coronavirus-expert-interview>.
23. Guan, Y., et al., *Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China*. *Science*, 2003. **302**(5643): p. 276-278.
24. Shi, Z. and Z. Hu, *A review of studies on animal reservoirs of the SARS coronavirus*. *Virus Res*, 2008. **133**(1): p. 74-87.
25. Graham, R.L. and R.S. Baric, *Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission*. *J Virol*, 2010. **84**(7): p. 3134-46.
26. Xu, L., et al., *Angiotensin-converting enzyme 2 (ACE2) from raccoon dog can serve as an efficient receptor for the spike protein of severe acute respiratory syndrome coronavirus*. *J Gen Virol*, 2009. **90**(Pt 11): p. 2695-2703.
27. Xiao, X., et al., *Animal sales from Wuhan wet markets immediately prior to the COVID-19 pandemic*. *Sci Rep*, 2021. **11**(1): p. 11898.
28. Liu, W.J., et al., *Surveillance of SARS-CoV-2 at the Huanan Seafood Market*. *Nature*, 2023.
29. Zapatero Gaviria, A. and R. Barba Martin, *What do we know about the origin of COVID-19 three years later?* *Rev Clin Esp (Barc)*, 2023. **223**(4): p. 240-243.
30. Domingo, J.L., *An updated review of the scientific literature on the origin of SARS-CoV-2*. *Environ Res*, 2022. **215**(Pt 1): p. 114131.
31. Chan, A. and M. Ridley, *Viral: The Search for the Origin of COVID-19*. 2021: Harper Press.
32. Young, A., *Pandora's Gamble: Lab Leaks, Pandemics, and a World at Risk*. 2023: Center Street Publishers.
33. Bostickson, B. and Y. Ghannam, *Wuhan laboratories, bat research, and biosafety*. 2021.
34. Demaneuf, G., *BSL Laboratories in Wuhan and their roles in coronavirus research*. 2022. Available from: <https://medium.com/@gillesdemaneuf/overview-of-biological-laboratories-in-wuhan-with-their-roles-in-coronavirus-research-bca6c1cd1f74>.
35. Zhan, S.H., B.E. Deverman, and Y.A. Chan, *SARS-CoV-2 is well adapted for humans. What does this mean for re-emergence?* *bioRxiv*, 2020: p. 2020.05.01.073262.
36. Piplani, S., et al., *In silico comparison of SARS-CoV-2 spike protein-ACE2 binding affinities across species and implications for virus origin*. *Sci Rep*, 2021. **11**(1): p. 13063.
37. Seyran, M., et al., *Questions concerning the proximal origin of SARS-CoV-2*. *J Med Virol*, 2021. **93**(3): p. 1204-1206.
38. Whittaker, G.R., *SARS-CoV-2 spike and its adaptable furin cleavage site*. *Lancet Microbe*, 2021. **2**(10): p. e488-e489.
39. Goh, G.K., et al., *Computational, Experimental, and Clinical Evidence of a Specific but Peculiar Evolutionary Nature of (COVID-19) SARS-CoV-2*. *J Proteome Res*, 2022. **21**(4): p. 874-890.

40. Lubinski, B. and G.R. Whittaker, *The SARS-CoV-2 furin cleavage site: natural selection or smoking gun?* *The Lancet Microbe*, 2023. **4**(8): p. e570.
41. Xiao, B., L. Xiao, and M. Elvey, *The possible origins of 2019-nCoV coronavirus*. 2020.
42. Hudson, J. and N. Jones, *State Department releases cable that launched claims that coronavirus escaped from Chinese lab*, in *Washington Post*. 2020.
43. Dou, E. and L. Kuo, *A scientist adventurer and China's 'Bat Woman' are under scrutiny as coronavirus lab-leak theory gets another look*, in *Washington Post*. 2021.
44. Wade, N., *The origin of COVID: Did people or nature open Pandora's box at Wuhan?*, in *Bulletin of the Atomic Scientists*. 2021. Available from: <https://thebulletin.org/2021/05/the-origin-of-covid-did-people-or-nature-open-pandoras-box-at-wuhan/>.
45. Markson, S., *What Really Happened In Wuhan: A Virus Like No Other, Countless Infections, Millions of Deaths*. 2021: HarperCollins.
46. Bahry, D., *Rational Discourse on Virology and Pandemics*. *mBio*, 2023. **14**(3): p. e00313-23.
47. Demaneuf, G. and R. de Maistre, *Outlines of a probabilistic evaluation of possible SARS-CoV-2 origins*. 2020.
48. Quay, S., *A Bayesian analysis concludes beyond a reasonable doubt that SARS-CoV-2 is not a natural zoonosis but instead is laboratory derived*. 2021.
49. Coccia, M., *Meta-analysis to explain unknown causes of the origins of SARS-COV-2*. *Environ Res*, 2022. **211**: p. 113062.
50. Weissman, M.B., *An Inconvenient Probability: Bayesian analysis of the probable origins of Covid*. 2024. Available from: <https://michaelweissman.substack.com/p/an-inconvenient-probability-v57>.
51. Chen, X., et al., *Use of a risk assessment tool to determine the origin of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)*. *Risk Anal*, 2024.
52. Gordon, M. and W. Strobel, *FBI Director Says Covid Pandemic Likely Caused by Chinese Lab Leak*, in *Wall Street Journal*. 2023. Available from: <https://www.wsj.com/articles/fbi-director-says-covid-pandemic-likely-caused-by-chinese-lab-leak-13a5e69b>.
53. Gordon, M. and W. Strobel, *Lab Leak Most Likely Origin of Covid-19 Pandemic, Energy Department Now Says*, in *Wall Street Journal*. 2023. Available from: <https://www.wsj.com/articles/covid-origin-china-lab-leak-807b7b0a>.
54. US Director of National Intelligence, *Report on Potential Links Between the Wuhan Institute of Virology and the Origins of COVID-19*, O.o.t.d.o.N. Intelligence, Editor. 2023: Washington, DC.
55. Redfield, R., *Written Testimony*. 2023, U.S House of Representatives Select Committee on the Coronavirus Crisis: Washington, DC.
56. Gelman, A., *About that claim that "SARS-CoV-2 is not a natural zoonosis but instead is laboratory derived"*. 2021. Available from: <https://statmodeling.stat.columbia.edu/2021/02/01/about-that-claim-that-sars-cov-2-is-not-a-natural-zoonosis-but-instead-is-laboratory-derived/>.
57. Monath, T.P., L.H. Kahn, and B. Kaplan, *Introduction: one health perspective*. *Ilar j*, 2010. **51**(3): p. 193-8.

58. Zinsstag, J., et al., *From “one medicine” to “one health” and systemic approaches to health and well-being*. Preventive Veterinary Medicine, 2011. **101**(3): p. 148-156.
59. Gelman, A., et al., *Bayesian Data Analysis*. 2021: Chapman & Hall/CRC.
60. Lee, M.D. and E.-J. Wagenmakers, *Bayesian Cognitive Modeling: A Practical Course*. 2013: Cambridge University Press.
61. Mahajan, S., *The Art of Insight in Science and Engineering: Mastering Complexity*. 2014: MIT Press.
62. Levin, A.T., *What Central Bankers Can Learn from Weather Forecasters*. International Banker, 2016.
63. Levin, A.T. and J.C. Williams, *Robust Monetary Policy with Competing Reference Models*. Journal of Monetary Economics, 2003. **50**: p. 945-975.
64. Taylor, J.B., *Introduction*, in *Monetary Policy Rules*, J.B. Taylor, Editor. 1999, University of Chicago Press.
65. Hubálek, Z., *Emerging Human Infectious Diseases: Anthroponoses, Zoonoses, and Sapronoses*. Emerging Infectious Disease journal, 2003. **9**(3): p. 403.
66. Jones, K.E., et al., *Global trends in emerging infectious diseases*. Nature, 2008. **451**(7181): p. 990-3.
67. Woolhouse, M. and E. Gaunt, *Ecological Origins of Novel Human Pathogens*. Critical Reviews in Microbiology, 2007. **33**(4): p. 231-242.
68. Johnson, N.P. and J. Mueller, *Updating the accounts: global mortality of the 1918-1920 “Spanish” influenza pandemic*. Bull Hist Med, 2002. **76**(1): p. 105-15.
69. Spreuwenberg, P., M. Kroneman, and J. Paget, *Reassessing the Global Mortality Burden of the 1918 Influenza Pandemic*. American Journal of Epidemiology, 2018. **187**(12): p. 2561-2567.
70. Taubenberger, J.K., *The origin and virulence of the 1918 “Spanish” influenza virus*. Proc Am Philos Soc, 2006. **150**(1): p. 86-112.
71. Viboud, C., et al., *Global Mortality Impact of the 1957–1959 Influenza Pandemic*. The Journal of Infectious Diseases, 2016. **213**(5): p. 738-745.
72. Kilbourne, E.D., *Influenza pandemics of the 20th century*. Emerg Infect Dis, 2006. **12**(1): p. 9-14.
73. Pigott, D.M., et al., *Mapping the zoonotic niche of Ebola virus disease in Africa*. eLife, 2014. **3**: p. e04395.
74. Sharp, P.M. and B.H. Hahn, *Origins of HIV and the AIDS Pandemic*. Cold Spring Harbor Perspectives in Medicine, 2011. **1**(1).
75. Zhu, T., et al., *An African HIV-1 sequence from 1959 and implications for the origin of the epidemic*. Nature, 1998. **391**(6667): p. 594-597.
76. Wang, L.F. and B.T. Eaton, *Bats, civets and the emergence of SARS*. Curr Top Microbiol Immunol, 2007. **315**: p. 325-44.
77. Dawood, F.S., et al., *Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study*. The Lancet Infectious Diseases, 2012. **12**(9): p. 687-695.

78. Han, H.J., H. Yu, and X.J. Yu, *Evidence for zoonotic origins of Middle East respiratory syndrome coronavirus*. J Gen Virol, 2016. **97**(2): p. 274-280.
79. Woolhouse, M.E., et al., *Assessing the Epidemic Potential of RNA and DNA Viruses*. Emerg Infect Dis, 2016. **22**(12): p. 2037-2044.
80. Woolhouse, M.E.J. and L. Brierley, *Epidemiological characteristics of human-infective RNA viruses*. Sci Data, 2018. **5**: p. 180017.
81. Harper, K., *Plagues upon the Earth*. 2021: Harper.
82. Gubler, D.J., *The Global Threat of Emergent/Reemergent Vector-Borne Diseases*, in *Vector-Borne Diseases: Understanding the Environmental, Human Health, and Ecological Connections*, I.o.M.U.F.o.M. Threats., Editor. 2008, National Academies Press: Washington DC. p. 43-64.
83. Keeling, M.J. and C.A. Gilligan, *Bubonic plague: a metapopulation model of a zoonosis*. Proc Biol Sci, 2000. **267**(1458): p. 2219-30.
84. Su, S., et al., *Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses*. Trends Microbiol, 2016. **24**(6): p. 490-502.
85. World Health Organization, *Consensus document on the epidemiology of severe acute respiratory syndrome (SARS)*, D.o.C.D.S.a. Response, Editor. 2003. Available from: https://iris.who.int/bitstream/handle/10665/70863/WHO_CDS_CSR_GAR_2003.11_eng.pdf.
86. Holmes, E.C. and A.J. Drummond, *The evolutionary genetics of viral emergence*. Curr Top Microbiol Immunol, 2007. **315**: p. 51-66.
87. Kane, Y., G. Wong, and G.F. Gao, *Animal Models, Zoonotic Reservoirs, and Cross-Species Transmission of Emerging Human-Infecting Coronaviruses*. Annu Rev Anim Biosci, 2023. **11**: p. 1-31.
88. Glezen, W.P., *Emerging Infections: Pandemic Influenza*. Epidemiologic Reviews, 1996. **18**(1): p. 64-76.
89. Vijaykrishna, D., et al., *Evolutionary insights into the ecology of coronaviruses*. J Virol, 2007. **81**(8): p. 4012-20.
90. Ye, Z.W., et al., *Zoonotic origins of human coronaviruses*. Int J Biol Sci, 2020. **16**(10): p. 1686-1697.
91. Markotter, W., et al., *Bat-borne viruses in Africa: a critical review*. Journal of Zoology, 2020. **311**(2): p. 77-98.
92. Xu, R.-H., et al., *Epidemiologic Clues to SARS Origin in China*. Emerging Infectious Disease journal, 2004. **10**(6): p. 1031.
93. Zheng, B.J., et al., *SARS-related virus predating SARS outbreak, Hong Kong*. Emerg Infect Dis, 2004. **10**(2): p. 176-8.
94. Yu, D., et al., *Prevalence of IgG Antibody to SARS-Associated Coronavirus in Animal Traders --- Guangdong Province, China, 2003*. MMWR Weekly, 2003. **52**(41): p. 986-987.
95. Kan, B., et al., *Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms*. J Virol, 2005. **79**(18): p. 11892-900.
96. Wang, M., et al., *SARS-CoV Infection in a Restaurant from Palm Civet*. Emerging Infectious Disease journal, 2005. **11**(12): p. 1860.

97. Yiming, L. and D. Li, *The dynamics of trade in live wildlife across the Guangxi border between China and Vietnam during 1993–1996 and its control strategies*. Biodiversity & Conservation, 1998. **7**: p. 895-914.
98. Lee, K.S., M.W.N. Lau, and B.P.L. Chan, *Wild Animal Trade Monitoring at Selected Markets in Guangzhou and Shenzhen, South China, 2000-2003*, in *Technical Report #2*. 2004, Kadoorie Farm & Botanic Garden.
99. Poon, L.L.M., et al., *Identification of a Novel Coronavirus in Bats*. Journal of Virology, 2005. **79**(4): p. 2001-2009.
100. Tu, C., et al., *Antibodies to SARS coronavirus in civets*. Emerg Infect Dis, 2004. **10**(12): p. 2244-8.
101. Song, H.-D., et al., *Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human*. Proceedings of the National Academy of Sciences, 2005. **102**(7): p. 2430-2435.
102. Lau, S.K.P., et al., *Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats*. Proceedings of the National Academy of Sciences, 2005. **102**(39): p. 14040-14045.
103. Li, W., et al., *Bats Are Natural Reservoirs of SARS-Like Coronaviruses*. Science, 2005. **310**(5748): p. 676-679.
104. Ren, W., et al., *Difference in Receptor Usage between Severe Acute Respiratory Syndrome (SARS) Coronavirus and SARS-Like Coronavirus of Bat Origin*. Journal of Virology, 2008. **82**(4): p. 1899-1907.
105. Hou, Y., et al., *Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry*. Archives of Virology, 2010. **155**(10): p. 1563-1569.
106. Hon, C.-C., et al., *Evidence of the Recombinant Origin of a Bat Severe Acute Respiratory Syndrome (SARS)-Like Coronavirus and Its Implications on the Direct Ancestor of SARS Coronavirus*. Journal of Virology, 2008. **82**(4): p. 1819-1826.
107. Boni, M.F., et al., *Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic*. Nature Microbiology, 2020. **5**(11): p. 1408-1417.
108. Ge, X.Y., et al., *Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor*. Nature, 2013. **503**(7477): p. 535-8.
109. Yang, X.L., et al., *Isolation and Characterization of a Novel Bat Coronavirus Closely Related to the Direct Progenitor of Severe Acute Respiratory Syndrome Coronavirus*. J Virol, 2015. **90**(6): p. 3253-6.
110. Lau, S.K., et al., *Severe Acute Respiratory Syndrome (SARS) Coronavirus ORF8 Protein Is Acquired from SARS-Related Coronavirus from Greater Horseshoe Bats through Recombination*. J Virol, 2015. **89**(20): p. 10532-47.
111. Martina, B.E.E., et al., *SARS virus infection of cats and ferrets*. Nature, 2003. **425**(6961): p. 915-915.
112. Chan, J.F., et al., *Is the discovery of the novel human betacoronavirus 2c EMC/2012 (HCoV-EMC) the beginning of another SARS-like pandemic?* J Infect, 2012. **65**(6): p. 477-89.
113. Lu, G. and D. Liu, *SARS-like virus in the Middle East: a truly bat-related coronavirus causing human diseases*. Protein Cell, 2012. **3**(11): p. 803-5.

114. Ithete, N.L., et al., *Close Relative of Human Middle East Respiratory Syndrome Coronavirus in Bat, South Africa*. Emerging Infectious Disease journal, 2013. **19**(10): p. 1697.
115. Corman, V.M., et al., *Rooting the Phylogenetic Tree of Middle East Respiratory Syndrome Coronavirus by Characterization of a Conspecific Virus from an African Bat*. Journal of Virology, 2014. **88**(19): p. 11297-11303.
116. Chan, J.F., et al., *Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease*. Clin Microbiol Rev, 2015. **28**(2): p. 465-522.
117. Munster, V.J., et al., *Replication and shedding of MERS-CoV in Jamaican fruit bats (*Artibeus jamaicensis*)*. Sci Rep, 2016. **6**: p. 21878.
118. Tan, C.C.S., et al., *Genomic screening of 16 UK native bat species through conservationist networks uncovers coronaviruses with zoonotic potential*. Nature Communications, 2023. **14**(1): p. 3322.
119. Reusken, C.B.E.M., et al., *Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study*. The Lancet Infectious Diseases, 2013. **13**(10): p. 859-866.
120. Chu, D.K., et al., *MERS coronaviruses in dromedary camels, Egypt*. Emerg Infect Dis, 2014. **20**(6): p. 1049-53.
121. Deem, S.L., et al., *Serological Evidence of MERS-CoV Antibodies in Dromedary Camels (*Camelus dromedaries*) in Laikipia County, Kenya*. PLOS ONE, 2015. **10**(10): p. e0140125.
122. Adney, D.R., et al., *Inoculation of Goats, Sheep, and Horses with MERS-CoV Does Not Result in Productive Viral Shedding*. Viruses, 2016. **8**(8): p. 230.
123. Azhar, E.I., et al., *Evidence for Camel-to-Human Transmission of MERS Coronavirus*. New England Journal of Medicine, 2014. **370**(26): p. 2499-2505.
124. Müller, M., et al., *MERS Coronavirus Neutralizing Antibodies in Camels, Eastern Africa, 1983–1997*. Emerging Infectious Disease journal, 2014. **20**(12): p. 2093.
125. de Wit, E., et al., *SARS and MERS: recent insights into emerging coronaviruses*. Nat Rev Microbiol, 2016. **14**(8): p. 523-34.
126. Lu, R., et al., *Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding*. The Lancet, 2020. **395**(10224): p. 565-574.
127. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin*. Nature, 2020. **579**(7798): p. 270-273.
128. Shang, J., et al., *Cell entry mechanisms of SARS-CoV-2*. Proceedings of the National Academy of Sciences, 2020. **117**(21): p. 11727-11734.
129. Xiong, Q., et al., *Close relatives of MERS-CoV in bats use ACE2 as their functional receptors*. Nature, 2022. **612**(7941): p. 748-757.
130. Ge, X.-Y., et al., *Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft*. Virologica Sinica, 2016. **31**(1): p. 31-40.
131. Shang, J., et al., *Structural basis of receptor recognition by SARS-CoV-2*. Nature, 2020. **581**(7807): p. 221-224.
132. Zhou, P., et al., *Addendum: A pneumonia outbreak associated with a new coronavirus of probable bat origin*. Nature, 2020. **588**(7836): p. E6.

133. Liu, S.-L., et al., *No credible evidence supporting claims of the laboratory engineering of SARS-CoV-2*. *Emerging Microbes & Infections*, 2020. **9**(1): p. 505-507.
134. Yee, J., *Bizarre Wuhan Wet Market Menu Shows Over 100 Wild Animals Sold As Food; Link With Virus Unclear*, in *MustShare News*. 2020. Available from: <https://mustsharenews.com/wuhan-market-animals-menu/>
135. Goodrum, F., et al., *Virology under the Microscope—a Call for Rational Discourse*. *mBio*, 2023. **14**(1): p. e00188-23.
136. Alwine, J.C., et al., *A Critical Analysis of the Evidence for the SARS-CoV-2 Origin Hypotheses*. *Journal of Virology*, 2023. **97**(4): p. e00365-23.
137. Holmes, E.C., *The Emergence and Evolution of SARS-CoV-2*. *Annual Review of Virology*, 2024. **11**(Volume 11, 2024): p. 21-42.
138. U.S. Center for Disease Control & Prevention, *Biosafety in Microbiological and Biomedical Laboratories*. 6th ed. 2024.
139. Pappas, G., *The Lanzhou Brucella Leak: The Largest Laboratory Accident in the History of Infectious Diseases?* *Clin Infect Dis*, 2022. **75**(10): p. 1845-1847.
140. Emond, R.T., et al., *A case of Ebola virus infection*. *British Medical Journal*, 1977. **2**(6086): p. 541-544.
141. Beer, B., R. Kurth, and A. Bukreyev, *Characteristics of Filoviridae: Marburg and Ebola Viruses*. *Naturwissenschaften*, 1999. **86**(1): p. 8-17.
142. Geddes, A.M., *The history of smallpox*. *Clinics in Dermatology*, 2006. **24**(3): p. 152-157.
143. Ristanović, E.S., et al., *A forgotten episode of Marburg virus disease: Belgrade, Yugoslavia, 1967*. *Microbiology and Molecular Biology Reviews*, 2020. **84**(2): p. 10.1128/mmr.00095-19.
144. Henkel, R.D., T. Miller, and R.S. Weyant, *Monitoring Select Agent Theft, Loss and Release Reports in the United States—2004–2010*. *Applied Biosafety*, 2012. **17**(4): p. 171-180.
145. Klotz, L.C. and E.J. Sylvester, *The consequences of a lab escape of a potential pandemic pathogen*. *Front Public Health*, 2014. **2**: p. 116.
146. Parry, J., *Breaches of safety regulations are probable cause of recent SARS outbreak, WHO says*. *BMJ*, 2004. **328**(7450): p. 1222.
147. Normile, D., *Mounting Lab Miscues Raise SARS Fears -- Chinese outbreak redoubles concerns about lab safety*. *Science*, 2004.
148. Cyranoski, D., *Chinese institutes investigate pathogen outbreaks in lab workers*. *Nature*, 2019.
149. Lipkin, W.I., *Biocontainment in Gain-of-Function Infectious Disease Research*. *mBio*, 2012. **3**(5): p. 10.1128/mbio.00290-12.
150. Lipsitch, M., *Why Do Exceptionally Dangerous Gain-of-Function Experiments in Influenza?*, in *Influenza Virus: Methods and Protocols*, Y. Yamauchi, Editor. 2018, Springer New York: New York, NY. p. 589-608.
151. Banerjee, A., et al., *Tools to study pathogen-host interactions in bats*. *Virus Research*, 2018. **248**: p. 5-12.
152. Herfst, S., et al., *Airborne Transmission of Influenza A/H5N1 Virus Between Ferrets*. *Science*, 2012. **336**(6088): p. 1534-1541.

153. Imai, M., et al., *Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets*. *Nature*, 2012. **486**(7403): p. 420-428.
154. Watanabe, T., et al., *Circulating avian influenza viruses closely related to the 1918 virus have pandemic potential*. *Cell Host Microbe*, 2014. **15**(6): p. 692-705.
155. Fauci, A.S., *Research on Highly Pathogenic H5N1 Influenza Virus: The Way Forward*. *mBio*, 2012. **3**(5): p. 10.1128/mbio.00359-12.
156. Gao, G.F., *For a better world: Biosafety strategies to protect global health*. *Biosafety and Health*, 2019. **1**(1): p. 1-3.
157. Menachery, V.D., et al., *A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence*. *Nature Medicine*, 2015. **21**(12): p. 1508-1513.
158. Menachery, V.D., et al., *Trypsin Treatment Unlocks Barrier for Zoonotic Bat Coronavirus Infection*. *J Virol*, 2020. **94**(5).
159. Zeng, L.P., et al., *Bat Severe Acute Respiratory Syndrome-Like Coronavirus WIV1 Encodes an Extra Accessory Protein, ORFX, Involved in Modulation of the Host Immune Response*. *J Virol*, 2016. **90**(14): p. 6573-6582.
160. Chinese Society of Laboratory Animal Science, *Seminar on key aspects of BSL-3 laboratory operation and maintenance*. 2018. Available from: <https://web.archive.org/web/20210105120613/https://www.lascn.net/Print.aspx?id=68780>.
161. Guo, H., et al., *Evolutionary Arms Race between Virus and Host Drives Genetic Diversity in Bat Severe Acute Respiratory Syndrome-Related Coronavirus Spike Genes*. *Journal of Virology*, 2020. **94**(20): p. 10.1128/jvi.00902-20.
162. Zhiming, Y., *Current status and future challenges of high-level biosafety laboratories in China*. *Journal of Biosafety and Biosecurity*, 2019. **1**(2): p. 123-127.
163. Jacobsen, R., "We never created a supervirus." *Ralph Baric explains gain-of-function research.*, in *MIT Technology Review*. 2021. Available from: <https://www.technologyreview.com/2021/07/26/1030043/gain-of-function-research-coronavirus-ralph-baric-vaccines/>.
164. Lipkin, W.I., *Interview Transcript*. 2023, U.S. House of Representatives - Select Committee on the Coronavirus Pandemic: Washington, DC. Available from: <https://oversight.house.gov/wp-content/uploads/2023/07/2023.04.06-Lipkin-Transcript.pdf>.
165. Berche, P., *Gain-of-function and origin of Covid19*. *Presse Med*, 2023. **52**(1): p. 104167.
166. Groves, D.C., S.L. Rowland-Jones, and A. Angyal, *The D614G mutations in the SARS-CoV-2 spike protein: Implications for viral infectivity, disease severity and vaccine design*. *Biochem Biophys Res Commun*, 2021. **538**: p. 104-107.
167. Korber, B., et al., *Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus*. *Cell*, 2020. **182**(4): p. 812-827.e19.
168. Yurkovetskiy, L., et al., *Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant*. *Cell*, 2020. **183**(3): p. 739-751.e8.
169. Guo, C., et al., *The D614G mutation redirects SARS-CoV-2 spike to lysosomes and suppresses deleterious traits of the furin cleavage site insertion mutation*. *Science Advances*, 2022. **8**(51): p. eade5085.

170. Fauver, J.R., et al., *Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States*. Cell, 2020. **181**(5): p. 990-996.e5.
171. Thomson, E.C., et al., *Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity*. Cell, 2021. **184**(5): p. 1171-1187.e20.
172. Forni, D., et al., *Molecular Evolution of Human Coronavirus Genomes*. Trends in Microbiology, 2017. **25**(1): p. 35-48.
173. Hodcroft, E.B., et al., *Spread of a SARS-CoV-2 variant through Europe in the summer of 2020*. Nature, 2021. **595**(7869): p. 707-712.
174. Decaro, N., *Betacoronavirus*, in *The Springer Index of Viruses*, C. Tidona and G. Darai, Editors. 2011, Springer New York: New York, NY. p. 385-401.
175. Wan, Y., et al., *Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus*. Journal of Virology, 2020. **94**(7): p. 10.1128/jvi.00127-20.
176. Zhao, X., et al., *Broad and Differential Animal Angiotensin-Converting Enzyme 2 Receptor Usage by SARS-CoV-2*. J Virol, 2020. **94**(18).
177. Conceicao, C., et al., *The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins*. PLoS Biol, 2020. **18**(12): p. e3001016.
178. Damas, J., et al., *Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates*. Proc Natl Acad Sci U S A, 2020. **117**(36): p. 22311-22322.
179. Zhai, X., et al., *Comparison of Severe Acute Respiratory Syndrome Coronavirus 2 Spike Protein Binding to ACE2 Receptors from Human, Pets, Farm Animals, and Putative Intermediate Hosts*. J Virol, 2020. **94**(15).
180. Belouzard, S., V.C. Chu, and G.R. Whittaker, *Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites*. Proceedings of the National Academy of Sciences, 2009. **106**(14): p. 5871-5876.
181. Li, F., *Structure, Function, and Evolution of Coronavirus Spike Proteins*. Annu Rev Virol, 2016. **3**(1): p. 237-261.
182. Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein*. Cell, 2020. **181**(2): p. 281-292.e6.
183. Coutard, B., et al., *The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade*. Antiviral Research, 2020. **176**: p. 104742.
184. Takeda, M., *Proteolytic activation of SARS-CoV-2 spike protein*. Microbiology and Immunology, 2022. **66**(1): p. 15-23.
185. Jackson, C.B., et al., *Mechanisms of SARS-CoV-2 entry into cells*. Nat Rev Mol Cell Biol, 2022. **23**(1): p. 3-20.
186. Segreto, R. and Y. Deigin, *The genetic structure of SARS-CoV-2 does not rule out a laboratory origin: SARS-COV-2 chimeric structure and furin cleavage site might be the result of genetic manipulation*. Bioessays, 2021. **43**(3): p. e2000240.
187. Hoffmann, M., H. Kleine-Weber, and S. Pöhlmann, *A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells*. Molecular Cell, 2020. **78**(4): p. 779-784.e5.

188. Johnson, B.A., et al., *Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis*. Nature, 2021. **591**(7849): p. 293-299.
189. Hallenberger, S., et al., *Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160*. Nature, 1992. **360**(6402): p. 358-361.
190. de Haan, C.A., et al., *Cleavage inhibition of the murine coronavirus spike protein by a furin-like enzyme affects cell-cell but not virus-cell fusion*. J Virol, 2004. **78**(11): p. 6048-54.
191. Pyrc, K., B. Berkhout, and L. van der Hoek, *The Novel Human Coronaviruses NL63 and HKU1*. Journal of Virology, 2007. **81**(7): p. 3051-3057.
192. Yamada, Y. and D.X. Liu, *Proteolytic Activation of the Spike Protein at a Novel RRRR/S Motif Is Implicated in Furin-Dependent Entry, Syncytium Formation, and Infectivity of Coronavirus Infectious Bronchitis Virus in Cultured Cells*. Journal of Virology, 2009. **83**(17): p. 8744-8758.
193. Braun, E. and D. Sauter, *Furin-mediated protein processing in infectious diseases and cancer*. Clinical & Translational Immunology, 2019. **8**(8): p. e1073.
194. Wu, Y. and S. Zhao, *Furin cleavage sites naturally occur in coronaviruses*. Stem Cell Res, 2020. **50**: p. 102115.
195. Tang, T., et al., *Proteolytic Activation of SARS-CoV-2 Spike at the S1/S2 Boundary: Potential Role of Proteases beyond Furin*. ACS Infect Dis, 2021. **7**(2): p. 264-272.
196. Liu, B., et al., *A comprehensive dataset of animal-associated sarbecoviruses*. Scientific Data, 2023. **10**(1): p. 681.
197. Baltimore, D., *The Debate over Origins of SARS-CoV-2*. CalTech Weekly, 2021.
198. Harrison, N.L. and 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, 2022. **119**(21): p. e2202769119.
199. Garry, R.F., *SARS-CoV-2 furin cleavage site was not engineered*. Proceedings of the National Academy of Sciences, 2022. **119**(40): p. e2211107119.
200. Harrison, N.L. and J.D. Sachs, *Reply to Garry: The origin of SARS-CoV-2 remains unresolved*. Proceedings of the National Academy of Sciences, 2022. **119**(45): p. e2215826119.
201. Romeu, A.R., *Probable human origin of the SARS-CoV-2 polybasic furin cleavage motif*. BMC Genomic Data, 2023. **24**(1): p. 71.
202. Chretien, J.-P. and R.G. Cutlip, *Critical analysis of Andersen et al.: The proximal origin of SARS-CoV-2*. 2020, U.S. National Center for Medical Intelligence: Washington, DC USA. Available from: <https://www.congress.gov/118/chrg/CHRG-118hhrg53002/CHRG-118hhrg53002.pdf>.
203. Zhou, H., et al., *A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein*. Current Biology, 2020. **30**(11): p. 2196-2203.e3.
204. Follis, K.E., J. York, and J.H. Nunberg, *Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry*. Virology, 2006. **350**(2): p. 358-69.
205. Li, W., et al., *A Single Point Mutation Creating a Furin Cleavage Site in the Spike Protein Renders Porcine Epidemic Diarrhea Coronavirus Trypsin Independent for Cell Entry and Fusion*. Journal of Virology, 2015. **89**(15): p. 8077-8081.

206. Yang, Y., et al., *Two Mutations Were Critical for Bat-to-Human Transmission of Middle East Respiratory Syndrome Coronavirus*. J Virol, 2015. **89**(17): p. 9119-23.
207. EcoHealth Alliance, *Project DEFUSE: Defusing the Threat of Bat-borne Coronaviruses*. 2018. Available from: <https://drasticresearch.org/wp-content/uploads/2021/09/main-document-preempt-volume-1-no-ess-hr00118s0017-ecohealth-alliance.pdf>.
208. Deigin, Y., *SARS-CoV-2 is precisely the virus that WIV was hunting in 2019*. Medium.com, 2024.
209. Anand, P., et al., *SARS-CoV-2 strategically mimics proteolytic activation of human ENaC*. eLife, 2020. **9**: p. e58603.
210. García-Caballero , A.n., et al., *ENaC Proteolytic Regulation by Channel-activating Protease 2*. Journal of General Physiology, 2008. **132**(5): p. 521-535.
211. Kota, P., et al., *The N terminus of α -ENaC mediates ENaC cleavage and activation by furin*. Journal of General Physiology, 2018. **150**(8): p. 1179-1187.
212. Sachs, J.D., et al., *The Lancet Commission on lessons for the future from the COVID-19 pandemic*. The Lancet, 2022. **400**(10359): p. 1224-1280.
213. Scientific Advisory Group for the Origins of Novel Pathogens (SAGO), *Preliminary Report*. 2022, World Health Organization,. Available from: https://cdn.who.int/media/docs/default-source/scientific-advisory-group-on-the-origins-of-novel-pathogens/sago-report-09062022.pdf?sfvrsn=42b55bbc_1&download=true.
214. The COVID Crisis Group, *Lessons from the COVID War: An Investigative Report*. 2023, New York, NY: Public Affairs.
215. Maddaloni, A. and J.-L. Peydro, *Monetary Policy, Macropprudential Policy, and Banking Stability: Evidence from the Euro Area*. International Journal of Central Banking, 2013. **9**(1): p. 121-169.
216. Resti, A., *Review of: Banking's Final Exam: Stress Testing and Bank-Capital Reform*. Journal of Economic Literature, 2018. **56**(1): p. 271-273.
217. U.S. National Center for Biotechnology Information. *PubMed Database*. 2024; Available from: <https://pubmed.ncbi.nlm.nih.gov/>.
218. Phelps, K.L., et al., *Bat Research Networks and Viral Surveillance: Gaps and Opportunities in Western Asia*. Viruses, 2019. **11**(3).
219. World Bank, *World Development Indicators Database - GDP Per Capita PPP (constant 2021 international \$)*, I.C. Program, Editor. 2024: Washington DC. Available from: <https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD>.
220. Johns Hopkins Center for Health Security & Nuclear Threat Initiative, *Global Health Security Index: Advancing Collective Action and Accountability Amid Global Crisis*. 2021.
221. Wu, G., *Laboratory biosafety in China: past, present, and future*. Biosafety and Health, 2019. **1**(2): p. 56-58.
222. Carini, A., *A Large Rabies Outbreak (Sur Une Grande Epizootie de Rage)*. Annal Inst Pasteur 1911. **25**: p. 843–846.
223. Cheetham, S. and W. Markotter, *Bats as Reservoirs of Viral Zoonoses*, in *50 Years of Bat Research: Foundations and New Frontiers*, B.K. Lim, et al., Editors. 2021, Springer International Publishing: Cham. p. 313-330.

224. Geng, R. and P. Zhou, *Severe acute respiratory syndrome (SARS) related coronavirus in bats*. *Animal Diseases*, 2021. **1**(1): p. 4.
225. Letko, M., et al., *Bat-borne virus diversity, spillover and emergence*. *Nat Rev Microbiol*, 2020. **18**(8): p. 461-471.
226. Williams, E.P., et al., *Common Themes in Zoonotic Spillover and Disease Emergence: Lessons Learned from Bat- and Rodent-Borne RNA Viruses*. *Viruses*, 2021. **13**(8).
227. Goh, G., et al., *Complementary regulation of caspase-1 and IL-1 β reveals additional mechanisms of dampened inflammation in bats*. *Proceedings of the National Academy of Sciences*, 2020. **117**(46): p. 28939-28949.
228. Luis, A.D., et al., *A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special?* *Proc Biol Sci*, 2013. **280**(1756): p. 20122753.
229. Brook, C.E. and A.P. Dobson, *Bats as 'special' reservoirs for emerging zoonotic pathogens*. *Trends Microbiol*, 2015. **23**(3): p. 172-80.
230. Irving, A.T., et al., *Lessons from the host defences of bats, a unique viral reservoir*. *Nature*, 2021. **589**(7842): p. 363-370.
231. Wu, Z., et al., *Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases*. *Microbiome*, 2018. **6**(1): p. 178.
232. Irving, A.T. and L.-F. Wang, *Are Bats Special?*, in *Bats and Viruses: Current Research and Future Trends*, E. Corrales-Aguilar and M. Schwemmler, Editors. 2020, Caister Academic Press: Norfolk, UK. p. 214-221.
233. Wu, Z., et al., *A comprehensive survey of bat sarbecoviruses across China in relation to the origins of SARS-CoV and SARS-CoV-2*. *Natl Sci Rev*, 2023. **10**(6): p. nwac213.
234. Bonaccorso, F.J. and L.P. McGuire, *Modeling the Colonization of Hawaii by Hoary Bats (*Lasiurus cinereus*)*, in *Bat Evolution, Ecology, and Conservation*, R.A. Adams and S.C. Pedersen, Editors. 2013, Springer New York: New York, NY. p. 187-205.
235. Alves, D.M.C.C., et al., *Geographic variation in the relationship between large-scale environmental determinants and bat species richness*. *Basic and Applied Ecology*, 2018. **27**: p. 1-8.
236. Walters, C.L., et al., *Challenges of Using Bioacoustics to Globally Monitor Bats*, in *Bat Evolution, Ecology, and Conservation*, R.A. Adams and S.C. Pedersen, Editors. 2013, Springer New York: New York, NY. p. 479-499.
237. Anthony, S.J., et al., *Global patterns in coronavirus diversity*. *Virus Evol*, 2017. **3**(1): p. vex012.
238. Olival, K.J., et al., *Host and viral traits predict zoonotic spillover from mammals*. *Nature*, 2017. **546**(7660): p. 646-650.
239. Brierley, L., et al., *Quantifying Global Drivers of Zoonotic Bat Viruses: A Process-Based Perspective*. *Am Nat*, 2016. **187**(2): p. E53-64.
240. Bresnahan, S. *Virus hunters searching for next outbreak*. 2017.
241. Frutos, R., et al., *Emergence of Bat-Related Betacoronaviruses: Hazard and Risks*. *Frontiers in Microbiology*, 2021. **12**.

242. International Union for Conservation of Nature (IUCN), *The IUCN Red List of Threatened Species*. 2018. Available from: <https://www.iucnredlist.org>.
243. Cohen, J., *Wuhan coronavirus hunter Shi Zhengli speaks out*. *Science*, 2020. **369**(6503): p. 487-488.
244. Qiu, J., *Meet the scientist at the center of the covid lab leak controversy*, in *MIT Technology Review*. 2022. Available from: <https://www.technologyreview.com/2022/02/09/1044985/shi-zhengli-covid-lab-leak-wuhan/>.
245. Qiu, J., *How China's 'Bat Woman' Hunted Down Viruses from SARS to the New Coronavirus*. *Scientific American*, 2020.
246. Sudworth, J., *Covid: Wuhan scientist would 'welcome' visit probing lab leak theory*, in *BBC News*. 2020. Available from: <https://www.bbc.com/news/world-asia-china-55364445>.
247. Hu, B., et al., *Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus*. *PLoS Pathog*, 2017. **13**(11): p. e1006698.
248. Yang, X.-L., et al., *Genetically Diverse Filoviruses in Rousettus and Eonycteris Bats, China, 2009 and 2015*. *Emerging Infectious Disease journal*, 2017. **23**(3): p. 482.
249. Luo, Y., et al., *Longitudinal Surveillance of Betacoronaviruses in Fruit Bats in Yunnan Province, China During 2009–2016*. *Virologica Sinica*, 2018. **33**(1): p. 87-95.
250. Wang, N., et al., *Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China*. *Virol Sin*, 2018. **33**(1): p. 104-107.
251. Zhang, Q., et al., *IFNAR2-dependent gene expression profile induced by IFN- α in Pteropus alecto bat cells and impact of IFNAR2 knockout on virus infection*. *PLOS ONE*, 2017. **12**(8): p. e0182866.
252. Wang, Y., et al., *Coronavirus nsp10/nsp16 Methyltransferase Can Be Targeted by nsp10-Derived Peptide *In Vitro* and *In Vivo* To Reduce Replication and Pathogenesis*. *Journal of Virology*, 2015. **89**(16): p. 8416-8427.
253. Zhou, P., et al., *Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin*. *Nature*, 2018. **556**(7700): p. 255-258.
254. Bostickson, B. and G. Demaneuf, *DRASTIC Investigation into Wuhan Institute of Virology Pathogen Databases*. Zenodo, 2021.
255. Suryanarayanan, S., *Wuhan Institute of Virology has many unreported bat virus samples, collaborating virologist says*. 2021, U.S. Right to Know. Available from: <https://usrtk.org/covid-19-origins/wuhan-institute-of-virology-has-many-unreported-bat-virus-samples-collaborating-virologist-says/>.
256. Chen, F., et al., *Decline of transmissible gastroenteritis virus and its complex evolutionary relationship with porcine respiratory coronavirus in the United States*. *Sci Rep*, 2019. **9**(1): p. 3953.
257. Zeng, Z., et al., *Coexistence of multiple genotypes of porcine epidemic diarrhea virus with novel mutant S genes in the Hubei Province of China in 2016*. *Virologica Sinica*, 2017. **32**(4): p. 298-306.
258. Yuan, Y., et al., *Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains*. *Nature Communications*, 2017. **8**(1): p. 15092.

259. Han, H.-J., et al., *Novel coronaviruses, astroviruses, adenoviruses and circoviruses in insectivorous bats from northern China*. *Zoonoses and Public Health*, 2017. **64**(8): p. 636-646.
260. OpenStreetMap (© OpenStreetMap contributors), *OpenStreetMap Database*,. 2024, OpenStreetMap Foundation,: Cambridge, UK. Available from: <https://www.openstreetmap.org/>.
261. Shen, Z., et al., *A conserved region of nonstructural protein 1 from alphacoronaviruses inhibits host gene expression and is critical for viral virulence*. *Journal of Biological Chemistry*, 2019. **294**(37): p. 13606-13618.
262. Zhang, X.-J., *The Institute of Model Animals of Wuhan University, China*. *European Heart Journal*, 2016. **37**(43): p. 3257-3259.
263. Luo, F., et al., *Evaluation of Antibody-Dependent Enhancement of SARS-CoV Infection in Rhesus Macaques Immunized with an Inactivated SARS-CoV Vaccine*. *Virol Sin*, 2018. **33**(2): p. 201-204.
264. Guo, M., et al., *Biosafety and data quality considerations for animal experiments with highly infectious agents at ABSL-3 facilities*. *Journal of Biosafety and Biosecurity*, 2019. **1**(1): p. 50-55.
265. Guo, W.P., et al., *Phylogeny and origins of hantaviruses harbored by bats, insectivores, and rodents*. *PLoS Pathog*, 2013. **9**(2): p. e1003159.
266. Shi, M., et al., *Redefining the invertebrate RNA virosphere*. *Nature*, 2016. **540**(7634): p. 539-543.
267. China Science Communication, *Youth in the Wild: Invisible Defender*. 2019, CCTV,.
268. Wang, Z., *Bat Hunter Tian Junhua*, in *Health Times*. 2020, People's Daily Online. Available from: http://paper.people.com.cn/jksb/html/2020-03/13/content_1979666.htm.
269. Lau, S.K.P., et al., *Isolation and Characterization of a Novel Betacoronavirus Subgroup A Coronavirus, Rabbit Coronavirus HKU14, from Domestic Rabbits*. *Journal of Virology*, 2012. **86**(10): p. 5481-5496.
270. Wong, A.C.P., et al., *Global Epidemiology of Bat Coronaviruses*. *Viruses*, 2019. **11**(2): p. 174.
271. Woo, P.C., et al., *Molecular diversity of coronaviruses in bats*. *Virology*, 2006. **351**(1): p. 180-7.
272. Woo, P.C., et al., *Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features*. *J Virol*, 2007. **81**(4): p. 1574-85.
273. Woo, P.C.Y., et al., *Coronavirus Diversity, Phylogeny and Interspecies Jumping*. *Experimental Biology and Medicine*, 2009. **234**(10): p. 1117-1127.
274. Woo, P.C.Y., et al., *Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavirus*. *Journal of Virology*, 2012. **86**(7): p. 3995-4008.
275. Wu, Z., et al., *Virome Analysis for Identification of Novel Mammalian Viruses in Bat Species from Chinese Provinces*. *Journal of Virology*, 2012. **86**(20): p. 10999-11012.

276. Wu, Z., et al., *Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases*. The ISME Journal, 2016. **10**(3): p. 609-620.
277. Jiang, S., et al., *Roadmap to developing a recombinant coronavirus S protein receptor-binding domain vaccine for severe acute respiratory syndrome*. Expert Review of Vaccines, 2012. **11**(12): p. 1405-1413.
278. Han, Y., et al., *Panoramic analysis of coronaviruses carried by representative bat species in Southern China to better understand the coronavirus sphere*. Nature Communications, 2023. **14**(1): p. 5537.
279. Lau, S.K.P., et al., *Differential Tropism of SARS-CoV and SARS-CoV-2 in Bat Cells*. Emerging Infectious Disease journal, 2020. **26**(12): p. 2961.
280. Schlottau, K., et al., *SARS-CoV-2 in fruit bats, ferrets, pigs, and chickens: an experimental transmission study*. Lancet Microbe, 2020. **1**(5): p. e218-e225.
281. Liu, K., et al., *Cross-species recognition of SARS-CoV-2 to bat ACE2*. Proc Natl Acad Sci U S A, 2021. **118**(1).
282. Yan, H., et al., *ACE2 receptor usage reveals variation in susceptibility to SARS-CoV and SARS-CoV-2 infection among bat species*. Nature Ecology & Evolution, 2021. **5**(5): p. 600-608.
283. Pekar, J.E., et al., *The recency and geographical origins of the bat viruses ancestral to SARS-CoV and SARS-CoV-2*. bioRxiv, 2023: p. 2023.07.12.548617.
284. Meekins, D.A., et al., *Susceptibility of swine cells and domestic pigs to SARS-CoV-2*. Emerging Microbes and Infections, 2020. **9**(1): p. 2278-2288.
285. Bosco-Lauth, A.M., et al., *Susceptibility of livestock to SARS-CoV-2 infection*. Emerging Microbes and Infections, 2021. **10**(1): p. 2199-2201.
286. Meekins, D.A., N.N. Gaudreault, and J.A. Richt, *Natural and Experimental SARS-CoV-2 Infection in Domestic and Wild Animals*. Viruses, 2021. **13**(10).
287. Mykytyn, A.Z., et al., *Susceptibility of rabbits to SARS-CoV-2*. Emerging Microbes & Infections, 2021. **10**(1): p. 1-7.
288. Bosco-Lauth, A.M., et al., *Experimental infection of domestic dogs and cats with SARS-CoV-2: Pathogenesis, transmission, and response to reexposure in cats*. Proc Natl Acad Sci U S A, 2020. **117**(42): p. 26382-26388.
289. Kim, Y.I., et al., *Infection and Rapid Transmission of SARS-CoV-2 in Ferrets*. Cell Host Microbe, 2020. **27**(5): p. 704-709 e2.
290. Shi, J., et al., *Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2*. Science, 2020. **368**(6494): p. 1016-1020.
291. Sit, T.H.C., et al., *Infection of dogs with SARS-CoV-2*. Nature, 2020. **586**(7831): p. 776-778.
292. Wu, L., et al., *Broad host range of SARS-CoV-2 and the molecular basis for SARS-CoV-2 binding to cat ACE2*. Cell Discov, 2020. **6**: p. 68.
293. Kutter, J.S., et al., *SARS-CoV and SARS-CoV-2 are transmitted through the air between ferrets over more than one meter distance*. Nature Communications, 2021. **12**(1): p. 1653.
294. Haagmans, B.L. and M.P.G. Koopmans, *Spreading of SARS-CoV-2 from hamsters to humans*. The Lancet, 2022. **399**(10329): p. 1027-1028.

295. Yen, H.-L., et al., *Transmission of SARS-CoV-2 delta variant (AY.127) from pet hamsters to humans, leading to onward human-to-human transmission: a case study*. The Lancet, 2022. **399**(10329): p. 1070-1078.
296. Li, M., et al., *Comparative susceptibility of SARS-CoV-2, SARS-CoV, and MERS-CoV across mammals*. The ISME Journal, 2023. **17**(4): p. 549-560.
297. Sills, J., et al., *Airborne transmission of SARS-CoV-2*. Science, 2020. **370**(6514): p. 303-304.
298. Hessler, P., *Nine Days in Wuhan, the Ground Zero of the Coronavirus Pandemic*, in *New Yorker*. 2020. Available from: <https://www.newyorker.com/magazine/2020/10/12/nine-days-in-wuhan-the-ground-zero-of-the-coronavirus-pandemic>.
299. Lam, T.T., et al., *Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins*. Nature, 2020. **583**(7815): p. 282-285.
300. Liu, P., et al., *Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)?* PLOS Pathogens, 2020. **16**(5): p. e1008421.
301. Macdonald, D.W., *The wet market sources of Covid-19: bats and pangolins have an alibi*, in *Oxford News Blog*. 2021. Available from: <https://www.ox.ac.uk/news/science-blog/wet-market-sources-covid-19-bats-and-pangolins-have-alibi>.
302. Boklund, A., et al., *SARS-CoV-2 in Danish Mink Farms: Course of the Epidemic and a Descriptive Analysis of the Outbreaks in 2020*. Animals (Basel), 2021. **11**(1).
303. Oude Munnink, B.B., et al., *Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans*. Science, 2021. **371**(6525): p. 172-177.
304. Wozencraft, W.C., *Genus Paguma - Masked Palm Civet*, in *The Mammals of China*, A.T. Smith and Y. Xie, Editors. 2013, Princeton University Press.
305. Smith, A.T. and Y. Xie, *The Mammals of China*. 2013: Princeton University Press.
306. Smith, A.T., *Genus Rhizomys - Bamboo Rat*, in *The Mammals of China*, A.T. Smith and Y. Xie, Editors. 2013, Princeton University Press.
307. Dalle Zotte, A., *Rabbit farming for meat purposes*. Animal Frontiers, 2014. **4**(4): p. 62-67.
308. Shrestha, T., *Himalayan Marmot (Marmota himalayana)*. IUCN Red List of Threatened Species, 2016.
309. Li, Y.-M., et al., *Illegal wildlife trade in the Himalayan region of China*. Biodiversity and Conservation, 2000. **9**: p. 901-918.
310. Thomson, B., *China culls 1.6 tonnes of rats bred for their meat as it cracks down on the wild animal trade due to COVID-19*, in *Daily Mail*. 2020: London, U.K. Available from: <https://www.dailymail.co.uk/news/article-8447367/China-culls-1-6-tonnes-bamboo-rats-bred-meat.html>.
311. Standaert, M. and E. Dou, *In search for coronavirus origins, Hubei caves and wildlife farms draw new scrutiny*, in *Washington Post*. 2020: Washington, DC. Available from: https://www.washingtonpost.com/world/asia_pacific/china-covid-bats-caves-hubei/2021/10/10/082eb8b6-1c32-11ec-bea8-308ea134594f_story.html.
312. Crits-Christoph, A., et al., *Genetic tracing of market wildlife and viruses at the epicenter of the COVID-19 pandemic*. Cell, 2024. **187**(19): p. 5468-5482.e11.
313. Kauhala, K., *The Raccoon Dog: A Successful Canid*. Canid News, 1994: p. 37-40.

314. Kauhala, K. and M. Saeki, *The IUCN Red List of Threatened Species: Nyctereutes procyonoides*. 2008.
315. Wozencraft, W.C., *Carnivora*, in *The Mammals of China*, A.T. Smith and Y. Xie, Editors. 2013, Princeton University Press.
316. Lan, T., et al., *The chromosome-scale genome of the raccoon dog: Insights into its evolutionary characteristics*. iScience, 2022. **25**(10): p. 105117.
317. Luo, C., et al., *Receptor binding and structural basis of raccoon dog ACE2 binding to SARS-CoV-2 prototype and its variants*. PLOS Pathogens, 2024. **20**(12): p. e1012713.
318. China Leather Industry Association, *Statistical Report on the Production of Skins of Mink, Fox and Raccoon in China*. 2021. Available from: <https://www.chinaleather.org/front/article/115225/366>.
319. Petry, M. and B. Liting, *PRC - Fur Animals and Products*, U.D.o.A.-F.A. Service, Editor. 2010: Washington DC. Available from: https://apps.fas.usda.gov/newgainapi/api/report/downloadreportbyfilename?filename=Fur%20Animals%20and%20Products_Beijing_China%20-%20Peoples%20Republic%20of_5-25-2010.pdf.
320. Humane Society International, *Fur farm investigation exposes immense suffering and disease risk in China*. 2024. Available from: <https://www.humanesociety.org/blog/undercover-investigation-exposes-suffering-china-fur-farm>.
321. Center for International Earth Science Information Network - CIESIN - Columbia University, *Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11*. 2018, NASA Socioeconomic Data and Applications Center (SEDAC): Palisades, New York. Available from: <https://doi.org/10.7927/H4JW8BX5>.
322. He, W.-T., et al., *Virome characterization of game animals in China reveals a spectrum of emerging pathogens*. Cell, 2022. **185**(7): p. 1117-1129.e8.
323. Lefkowitz, E., et al., *Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV)*. Nucleic Acids Research, 2017.
324. Walker, P., et al., *Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses* Archives of Virology, 2022.
325. Zhao, J., et al., *Farmed fur animals harbour viruses with zoonotic spillover potential*. Nature, 2024.
326. Buonavoglia, C., et al., *Canine Coronavirus Highly Pathogenic for Dogs*. Emerging Infectious Disease journal, 2006. **12**(3): p. 492.
327. Timurkan, M.O., et al., *Molecular characterization of canine coronaviruses: an enteric and pantropic approach*. Archives of Virology, 2021. **166**(1): p. 35-42.
328. Hsieh, Y., et al., *Fun Fur? A report on the Chinese fur industry*. 2006, Swiss Animal Protection SAP.
329. Lung, Y.-C. and S. Lin, *China's Fur Trade and Its Position in the Global Fur Industry*. 2019, ActAsia. Available from: <https://www.actasia.org/wp-content/uploads/2019/10/China-Fur-Report-7.5.pdf>.
330. Kauhala, K. and R. Kowalczyk, *The Raccoon Dog (Nyctereutes Procyonoides) in the Community of Medium-Sized Carnivores in Europe: Its Adaptations, Impact on Native Fauna, and Management of the Population*, in *CABI Compendium*. 2004.

331. Kauhala, K., M. Kaunisto, and E. Helle, *Diet of the Raccoon dog, Nyctereutes procyonoides*. Z. Säugetierkund, 1993. **58**: p. 129-136.
332. Zoller, H. and F. Drygala, *Activity patterns of the invasive raccoon dog (Nyctereutes procyonoides) in North East Germany*. Folia Zoologica, 2013. **62**(4): p. 290-296.
333. Xin, W., *Shanghai learning to live with its raccoon dogs*, in *China Daily*. 2023.
334. Townshend, T., *Guide to the Mammals of Beijing*, B. Beijing, Editor. 2018.
335. Kauhala, K., *Nyctereutes Procyonoides (Raccoon Dog)*. 2009, Centre for Agriculture and Bioscience International.
336. Drygala, F., et al., *Spatial organisation and intra-specific relationship of the raccoon dog Nyctereutes procyonoides in Central Europe*. Wildlife Biology, 2008. **14**(4): p. 457-466.
337. Kauhala, K. and R. Kowalczyk, *Invasion of the raccoon dog Nyctereutes procyonoides in Europe: History of colonization, features behind its success, and threats to native fauna*. Current Zoology, 2011. **57**(5): p. 584-598.
338. Jeong, W., et al., *Home range differences by habitat type of raccoon dogs Nyctereutes procyonoides (Carnivora: Canidae)*. Journal of Asia-Pacific Biodiversity, 2017. **10**(3): p. 349-354.
339. Suld, K., U. Saarma, and H. Valdmann, *Home ranges of raccoon dogs in managed and natural areas*. PLoS One, 2017. **12**(3): p. e0171805.
340. Carr, K., *Nyctereutes Procyonoides, raccoon dog*. 2004, Animal Diversity Web.
341. Sowerby, A.C., *Wolves, Wild Dogs and Foxes*, in *A Naturalist in Manchuria*. 1923, Tientsin Press.
342. Allen, G.M., *Nyctereutes procyonoides procyonoides (Gray), Raccoon Dog*, in *The Mammals of China and Mongolia*. 1938, American Museum of Natural History. p. 346-349.
343. Wozencraft, W.C., *Genus Nyctereutes - Raccoon Dog*, in *The Mammals of China*, A.T. Smith and Y. Xie, Editors. 2013, Princeton University Press.
344. Kauhala, K. and M. Saeki, *Nyctereutes procyonoides, raccoon dog*. 2016, IUCN.
345. Marshall, P., *Hong Kong Mammals*. Journal of the Hong Kong Branch of the Royal Asiatic Society, 1967: p. 11-20.
346. Duckworth, J.W., et al., *Paguma larvata (masked palm civet)*. IUCN Red List of Threatened Species, 2016.
347. Freuling, C.M., et al., *Susceptibility of Raccoon Dogs for Experimental SARS-CoV-2 Infection*. Emerg Infect Dis, 2020. **26**(12): p. 2982-2985.
348. Latham, J., *The Great Raccoon Dog Mystery*, in *Independent Science News*. 2023. Available from: <https://jonathanlatham.net/the-great-raccoon-dog-mystery/>.
349. Li, C., et al., *A Circa 2010 Thirty Meter Resolution Forest Map for China*. Remote Sensing, 2014. **6**(6): p. 5325-5343.
350. Wang, W., et al., *Coronaviruses in wild animals sampled in and around Wuhan at the beginning of COVID-19 emergence*. Virus Evol, 2022. **8**(1): p. veac046.
351. Boklund, A., et al., *Monitoring of SARS-CoV-2 infection in mustelids*. EFSA Journal, 2021. **19**(3).

352. Stoyan, D. and S.N. Chiu, *Statistics did not prove that the Huanan Seafood Wholesale Market was the early epicentre of the COVID-19 pandemic*. Journal of the Royal Statistical Society Series A: Statistics in Society, 2024.
353. Weissman, M.B., *Proximity ascertainment bias in early COVID case locations*. Journal of the Royal Statistical Society Series A: Statistics in Society, 2024.
354. Zhang , D., et al., *Zoonosis at the Huanan Seafood Market: A Critique*, in *Chemistry & Physics Faculty Proceedings*, N.S. University, Editor. 2022.
355. Débarre, F. and M. Worobey, *Confirmation of the centrality of the Huanan market among early COVID-19 cases: Reply to Stoyan and Chiu (2024)*. Arxiv.org, 2024.
356. Débarre, F. and M. Worobey, *No evidence of systematic proximity ascertainment bias in early COVID-19 cases in Wuhan: Reply to Weissman (2024)*. Arxiv.org, 2024.
357. Begum, F., *Mapping Disease: John Snow and Cholera*. 2016, Royal College of Surgeons of England,. Available from: <https://www.rcseng.ac.uk/library-and-publications/library/blog/mapping-disease-john-snow-and-cholera/>.
358. Paneth, N., et al., *A rivalry of foulness: official and unofficial investigations of the London cholera epidemic of 1854*. American Journal of Public Health, 1998. **88**(10): p. 1545-1553.
359. Keeling, M.J. and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*. 2008, Princeton: Princeton University Press.
360. Allen, L.J.S., *A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis*. Infectious Disease Modelling, 2017. **2**(2): p. 128-142.
361. Held, L., S. Meyer, and J. Bracher, *Probabilistic forecasting in infectious disease epidemiology: the 13th Armitage lecture*. Statistics in Medicine, 2017. **36**(22): p. 3443-3460.
362. Doyle, A.C., *The Adventure of Silver Blaze*, in *Memoirs of Sherlock Holmes*. 1893, G. Newnes Ltd.: London, UK.
363. Davies, N.G., et al., *Age-dependent effects in the transmission and control of COVID-19 epidemics*. Nat Med, 2020. **26**(8): p. 1205-1211.
364. Li, R., et al., *Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)*. Science, 2020. **368**(6490): p. 489-493.
365. Li, Q., et al., *Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia*. N Engl J Med, 2020. **382**(13): p. 1199-1207.
366. Verity, R., et al., *Estimates of the severity of coronavirus disease 2019: a model-based analysis*. The Lancet Infectious Diseases, 2020. **20**(6): p. 669-677.
367. Imai, N., et al., *Estimating the potential total number of novel coronavirus cases in Wuhan City, China*. 2020, Imperial College London.
368. Herrera-Esposito, D. and G. de Los Campos, *Age-specific rate of severe and critical SARS-CoV-2 infections estimated with multi-country seroprevalence studies*. BMC Infect Dis, 2022. **22**(1): p. 311.
369. Bulfone, T.C., et al., *Outdoor Transmission of SARS-CoV-2 and Other Respiratory Viruses: A Systematic Review*. The Journal of Infectious Diseases, 2020. **223**(4): p. 550-561.
370. European Centre for Disease & Prevention, *Contact tracing: public health management of persons, including healthcare workers, who have had contact with COVID-19 cases in the European Union – third update*. 2020. Available from:

- <https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-contact-tracing-public-health-management-third-update.pdf>.
371. U.S. Center for Disease Control & Prevention, *Coronavirus 2019 - Definition of Close Contact*. 2020. Available from: <https://web.archive.org/web/20201101000502/https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/appendix.html>.
 372. Fouda, B., et al., *Identifying SARS-CoV2 transmission cluster category: An analysis of country government database*. Journal of Infection and Public Health, 2021. **14**(4): p. 461-467.
 373. Qian, H., et al., *Indoor transmission of SARS-CoV-2*. Indoor Air, 2021. **31**(3): p. 639-645.
 374. Greenhalgh, T., et al., *Ten scientific reasons in support of airborne transmission of SARS-CoV-2*. The Lancet, 2021. **397**(10285): p. 1603-1605.
 375. Duval, D., et al., *Long distance airborne transmission of SARS-CoV-2: rapid systematic review*. BMJ, 2022. **377**: p. e068743.
 376. Ferretti, L., et al., *Digital measurement of SARS-CoV-2 transmission risk from 7 million contacts*. Nature, 2024. **626**(7997): p. 145-150.
 377. WorldPop and Columbia University Center for International Earth Science Information Network (CIESIN), *Global High Resolution Population Denominators Project*. 2024.
 378. Wang, D., et al., *Epidemiological characteristics and the entire evolution of coronavirus disease 2019 in Wuhan, China*. Respiratory Research, 2020. **21**(1): p. 257.
 379. Linton, N.M., et al., *Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data*. J Clin Med, 2020. **9**(2).
 380. Grant, M.C., et al., *The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries*. PLoS One, 2020. **15**(6): p. e0234765.
 381. Shi, F., et al., *The comparison of epidemiological characteristics between confirmed and clinically diagnosed cases with COVID-19 during the early epidemic in Wuhan, China*. Global Health Research and Policy, 2021. **6**(1): p. 18.
 382. Demaneuf, G., *Arrested Development: The number of Wuhan cases of COVID-19 with onset in 2019*. 2022, ResearchGate. Available from: https://www.researchgate.net/publication/364644929_Arrested_Development_The_number_of_Wuhan_cases_of_COVID-19_with_onset_in_2019.
 383. GoogleEarth, *Satellite Images of Wuhan PRC*, c. Airbus, Editor. 2024. Available from: <https://www.google.com/permissions/geoguidelines>.
 384. Peng, Y., *Searching for the first infected person in the Huanan Seafood Market*. 2020, The Paper. Available from: https://www.thepaper.cn/newsDetail_forward_6681256.
 385. Jijón, S., et al., *Using early detection data to estimate the date of emergence of an epidemic outbreak*. PLOS Computational Biology, 2024. **20**(3): p. e1011934.
 386. Rocklöv, J., H. Sjödin, and A. Wilder-Smith, *COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures*. J Travel Med, 2020. **27**(3).

387. Wu, J.T., K. Leung, and G.M. Leung, *Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study*. The Lancet, 2020. **395**(10225): p. 689-697.
388. Liu, W., et al., *Spatiotemporal analysis of COVID-19 outbreaks in Wuhan, China*. Scientific Reports, 2021. **11**(1): p. 13648.
389. Cohen, J., *Wuhan seafood market may not be source of novel virus spreading globally*. Science, 2020.
390. Kong, W.-H., et al., *SARS-CoV-2 detection in patients with influenza-like illness*. Nature Microbiology, 2020. **5**(5): p. 675-678.
391. Chang, L., et al., *Serosurvey for SARS-CoV-2 among blood donors in Wuhan, China from September to December 2019*. Protein & Cell, 2022. **14**(1): p. 28-36.
392. Carrat, F., et al., *Evidence of early circulation of SARS-CoV-2 in France: findings from the population-based "CONSTANCES" cohort*. Eur J Epidemiol, 2021. **36**(2): p. 219-222.
393. La Rosa, G., et al., *SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring*. Sci Total Environ, 2021. **750**: p. 141711.
394. Montomoli, E., et al., *Timeline of SARS-CoV-2 Spread in Italy: Results from an Independent Serological Retesting*. Viruses, 2022. **14**(1): p. 61.
395. Giovanetti, M., et al., *The first two cases of 2019-nCoV in Italy: Where they come from?* Journal of Medical Virology, 2020. **92**(5): p. 518-521.
396. Global Times, *Wuhan's Huanan seafood market a victim of COVID-19: CDC director*, in *People's Daily*. 2020: Beijing PRC. Available from: <https://www.globaltimes.cn/content/1189506.shtml>.
397. Bliss, C.I., *The Method of Probits*. Science, 1934. **79**(2037): p. 38-39.
398. Finney, D.J., *Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve*. 1947: Cambridge University Press.
399. Google Scholar Database, *Articles with Keyword "Probit" in Title*. 2024. Available from: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C30&q=probit&btnG=.
400. Kermack, W.O., A.G. McKendrick, and G.T. Walker, *A contribution to the mathematical theory of epidemics*. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 1927. **115**(772): p. 700-721.
401. Held, L. and M. Paul, *Modeling seasonality in space-time infectious disease surveillance data*. Biometrical Journal, 2012. **54**(6): p. 824-843.
402. Meyer, S. and L. Held, *Power-law models for infectious disease spread*. The Annals of Applied Statistics, 2014. **8**(3): p. 1612-1639, 28.
403. Brockmann, D., L. Hufnagel, and T. Geisel, *The scaling laws of human travel*. Nature, 2006. **439**(7075): p. 462-465.
404. Deardon, R., et al., *Inference for Individual-Level Models of Infectious Diseases in Large Populations*. Stat Sin, 2010. **20**(1): p. 239-261.
405. Metropolis, N., et al., *Equation of State Calculations by Fast Computing Machines*. The Journal of Chemical Physics, 1953. **21**(6): p. 1087-1092.
406. Hastings, W.K., *Monte Carlo sampling methods using Markov chains and their applications*. Biometrika, 1970. **57**(1): p. 97-109.

407. Gelfand, A.E. and A.F.M. Smith, *Sampling-Based Approaches to Calculating Marginal Densities*. Journal of the American Statistical Association, 1990. **85**(410): p. 398-409.
408. StataCorp LLC, *Bayesian Analysis in Stata*. 2024. Available from: <https://www.stata.com/features/overview/bayesian-analysis/>.
409. Brooks, S.P. and A. Gelman, *General Methods for Monitoring Convergence of Iterative Simulations*. Journal of Computational and Graphical Statistics, 1998. **7**(4): p. 434-455.
410. Gelman, A. and D.B. Rubin, *Inference from Iterative Simulation Using Multiple Sequences*. Statistical Science, 1992. **7**(4): p. 457-472.
411. Haider, N., et al., *COVID-19-Zoonosis or Emerging Infectious Disease?* Front Public Health, 2020. **8**: p. 596944.
412. Riou, J. and C.L. Althaus, *Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020*. Euro Surveill, 2020. **25**(4).
413. Bloomberg News, *Wuhan Is Returning to Life. So Are Its Disputed Wet Markets*. 2020. Available from: <https://www.bloomberg.com/news/articles/2020-04-08/wuhan-is-returning-to-life-so-are-its-disputed-wet-markets>.
414. Hobbs, E.C. and T.J. Reid, *Animals and SARS-CoV-2: Species susceptibility and viral transmission in experimental and natural conditions, and the potential implications for community transmission*. Transbound Emerg Dis, 2021. **68**(4): p. 1850-1867.
415. Westcott, B. and S. Want, *China's wet markets are not what some people think they are*, in *CNN News*. 2020. Available from: <https://www.cnn.com/2020/04/14/asia/china-wet-market-coronavirus-intl-hnk/index.html>.
416. First Life, *Crab Season: How to buy crabs in Wuhan and catch crabs in the lake*. 2018, Sohu.com. Available from: https://m.sohu.com/a/256151004_739135.
417. Kuester, A., *Photos of Huanan Market*. 2019, Flickr.com. Available from: <https://www.flickr.com/photos/arendpictures/49333662807/in/album-72157712524297396/>.
418. Daszak, P., *Slide Presentation about WHO Report on COVID Origins*. 2021. Available from: <https://downloads.vanityfair.com/ecohealth-alliance/peter-daszak-powerpoint.pdf>.
419. Derqui, N., et al., *Risk factors and vectors for SARS-CoV-2 household transmission: a prospective, longitudinal cohort study*. The Lancet Microbe, 2023. **4**(6): p. e397-e408.
420. Harvey, A.P., et al., *Longitudinal Monitoring of SARS-CoV-2 RNA on High-Touch Surfaces in a Community Setting*. Environmental Science & Technology Letters, 2021. **8**(2): p. 168-175.
421. Liu, H., et al., *Investigating SARS-CoV-2 persistent contamination in different indoor environments*. Environmental Research, 2021. **202**: p. 111763.
422. Pastorino, B., et al., *Prolonged Infectivity of SARS-CoV-2 in Fomites*. Emerg Infect Dis, 2020. **26**(9): p. 2256-7.
423. Short, K.R. and B.J. Cowling, *Assessing the potential for fomite transmission of SARS-CoV-2*. The Lancet Microbe, 2023. **4**(6): p. e380-e381.
424. Courtier-Orgogozo, V. and F.A. de Ribera, *SARS-CoV-2 infection at the Huanan seafood market*. Environ Res, 2022. **214**(Pt 1): p. 113702.
425. Page, J., W. Fan, and N. Khan, *How It All Started: China's Early Coronavirus Missteps*, in *Wall Street Journal*. 2020. Available from: <https://www.wsj.com/articles/how-it-all-started-chinas-early-coronavirus-missteps-11583508932>.

426. Cheng, P.K.C., et al., *Viral shedding patterns of coronavirus in patients with probable severe acute respiratory syndrome*. *The Lancet*, 2004. **363**(9422): p. 1699-1700.
427. Leung, W.K., et al., *Enteric involvement of severe acute respiratory syndrome-associated coronavirus infection I* The authors thank Man-ye Yung, Sara Fung, Dr. Bonnie Kwan, and Dr. Thomas Li for their help in retrieving patient information. *Gastroenterology*, 2003. **125**(4): p. 1011-1017.
428. McKinney, K.R., Y.Y. Gong, and T.G. Lewis, *Environmental transmission of SARS at Amoy Gardens*. *J Environ Health*, 2006. **68**(9): p. 26-30; quiz 51-2.
429. Albert, S., et al., *Lack of evidence for infectious SARS-CoV-2 in feces and sewage*. *Eur J Clin Microbiol Infect Dis*, 2021. **40**(12): p. 2665-2667.
430. Cerrada-Romero, C., et al., *Excretion and viability of SARS-CoV-2 in feces and its association with the clinical outcome of COVID-19*. *Sci Rep*, 2022. **12**(1): p. 7397.
431. Guo, M., et al., *Reply to: Rectally shed SARS-CoV-2 lacks infectivity: time to rethink faecal–oral transmission?* *Nature Reviews Gastroenterology & Hepatology*, 2021. **18**(9): p. 669-670.
432. Pedersen, R.M., et al., *Rectally shed SARS-CoV-2 lacks infectivity: time to rethink faecal–oral transmission?* *Nature Reviews Gastroenterology & Hepatology*, 2021. **18**(9): p. 669-669.
433. Caixin News Desk, *Private clinic doctors fighting against COVID-19*, in *Caixin*. 2020. Available from: <https://china.caixin.com/2020-04-07/101539697>.
434. Corman, V.M., et al., *Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR*. *Euro Surveill*, 2020. **25**(3).
435. Chin, A.W.H., et al., *Stability of SARS-CoV-2 in different environmental conditions*. *The Lancet Microbe*, 2020. **1**(1): p. e10.
436. Dong, L., et al., *Highly accurate and sensitive diagnostic detection of SARS-CoV-2 by digital PCR*. *Talanta*, 2021. **224**: p. 121726.
437. Yu, F., et al., *Assessment of the Diagnostic Ability of Four Detection Methods Using Three Sample Types of COVID-19 Patients*. *Frontiers in Cellular and Infection Microbiology*, 2021. **11**.
438. Johns, G.C. and J.C. Avise, *A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene*. *Molecular Biology and Evolution*, 1998. **15**(11): p. 1481-1490.
439. Valentini, A., F. Pompanon, and P. Taberlet, *DNA barcoding for ecologists*. *Trends Ecol Evol*, 2009. **24**(2): p. 110-7.
440. Ratnasingham, S. and P.D. Hebert, *A DNA-based registry for all animal species: the barcode index number (BIN) system*. *PLoS One*, 2013. **8**(7): p. e66213.
441. U.S. National Center for Biotechnology Information, *Nucleotide (nt) Database*. 2024. Available from: <https://www.ncbi.nlm.nih.gov/nucleotide/>.
442. Wood, D.E., J. Lu, and B. Langmead, *Improved metagenomic analysis with Kraken 2*. *Genome Biology*, 2019. **20**(1): p. 257.
443. Debarre, F., *What we can and cannot learn from SARS-CoV-2 and animals in metagenomic samples from the Huanan market*. *Virus Evol*, 2024. **10**(1): p. vead077.
444. Ruan, T., et al., *Research on the Changes in Distribution and Habitat Suitability of the Chinese Red Panda Population*. *Animals*, 2024. **14**(3): p. 424.

445. Luo, A., et al., *Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals*. BMC Genomics, 2011. **12**(1): p. 84.
446. Kim, S.-I., et al., *Phylogeography of Korean raccoon dogs: implications of peripheral isolation of a forest mammal in East Asia*. Journal of Zoology, 2013. **290**(3): p. 225-235.
447. Morón-López, J., et al., *Intraspecies variation of the mitochondrial genome: An evaluation for phylogenetic approaches based on the conventional choices of genes and segments on mitogenome*. PLOS ONE, 2022. **17**(8): p. e0273330.
448. Elyasigorji, Z., et al., *Mitochondrial genes as strong molecular markers for species identification*. The Nucleus, 2023. **66**(1): p. 81-93.
449. Li, J., et al., *Applying DNA barcoding to conservation practice: a case study of endangered birds and large mammals in China*. Biodiversity and Conservation, 2017. **26**(3): p. 653-668.
450. Coenen Ashley, R. and S. Weitz Joshua, *Limitations of Correlation-Based Inference in Complex Virus-Microbe Communities*. mSystems, 2018. **3**(4): p. 10.1128/msystems.00084-18.
451. Carr, A., et al., *Use and abuse of correlation analyses in microbial ecology*. The ISME Journal, 2019. **13**(11): p. 2647-2655.
452. Bloom, J.D., *Association between SARS-CoV-2 and metagenomic content of samples from the Huanan Seafood Market*. Virus Evolution, 2023. **9**(2).
453. Bloom, J.D., *Importance of quantifying the number of viral reads in metagenomic sequencing of environmental samples from the Huanan Seafood Market*. Virus Evol, 2024. **10**(1): p. vead089.
454. Woolhouse, M.E., *Did COVID come from an animal market? Here's what the new evidence really tells us*, in *The Conversation*. 2024. Available from: <https://theconversation.com/did-covid-come-from-an-animal-market-heres-what-the-new-evidence-really-tells-us-239533>.
455. Rambaut, A., et al., *A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology*. Nature Microbiology, 2020. **5**(11): p. 1403-1407.
456. Pekar, J.E., et al., *The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2*. Science, 2022. **377**(6609): p. 960-966.
457. Pekar, J.E., et al., *Erratum for "The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2"*. Science, 2023. **382**(6667): p. ead10585.
458. Demaneuf, G., *Backstage Story: The October 2023 correction to Pekar et al. (2022)*. 2024. Available from: <https://gillesdemaneuf.medium.com/backstage-story-the-oct-2023-correction-to-pekar-et-al-e167080e957d>.
459. PubPeer Forum, *Comments on "The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2"*. 2024. Available from: <https://pubpeer.com/publications/3FB983CC74C0A93394568A373167CE>.
460. Baker, A., *Simplicity*, in *Stanford Encyclopedia of Philosophy*, E.N. Zalta, Editor. 2022.
461. Bloom, J.D., et al., *Investigate the origins of COVID-19*. Science, 2021. **372**(6543): p. 694-694.
462. Hanage, W., *Amid a muddled quest for covid's source, a crucial message*, in *Washington Post*. 2021.
463. van Helden, J., et al., *An appeal for an objective, open, and transparent scientific debate about the origin of SARS-CoV-2*. Lancet, 2021. **398**(10309): p. 1402-1404.

464. Gostin, L.O. and G.K. Gronvall, *The Origins of Covid-19 — Why It Matters (and Why It Doesn't)*. New England Journal of Medicine, 2023. **388**(25): p. 2305-2308.
465. Young, A., *How did the Covid pandemic begin? We need to investigate all credible hypotheses*, in *The Guardian (U.K.)*. 2023. Available from: <https://www.theguardian.com/commentisfree/2023/mar/02/covid-19-pandemic-origin-answers-political-feud>.
466. World Health Organization, *COVID-19 Virtual Press Conference Transcript*. 2021. Available from: <https://www.who.int/publications/m/item/covid-19-virtual-press-conference-transcript--15-july-2021>.
467. Liu, D.X., J.Q. Liang, and T.S. Fung, *Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (Coronaviridae)*. Encyclopedia of Virology, 2021: p. 428-440.
468. Agostini, H.J., J.D. Carroll, and K.W. Minton, *Identification and characterization of uvrA, a DNA repair gene of Deinococcus radiodurans*. Journal of Bacteriology, 1996. **178**(23): p. 6759-6765.
469. Makarova, K.S., et al., *Genome of the Extremely Radiation-Resistant Bacterium <i>Deinococcus radiodurans</i> Viewed from the Perspective of Comparative Genomics*. Microbiology and Molecular Biology Reviews, 2001. **65**(1): p. 44-79.
470. Walker, S.C., J.M. Avis, and G.L. Conn, *General plasmids for producing RNA in vitro transcripts with homogeneous ends*. Nucleic Acids Research, 2003. **31**(15): p. e82-e82.
471. Jinek, M., et al., *A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity*. Science, 2012. **337**(6096): p. 816-821.
472. Adli, M., *The CRISPR tool kit for genome editing and beyond*. Nature Communications, 2018. **9**(1): p. 1911.
473. Ledford, H. and E. Callaway, *Pioneers of revolutionary CRISPR gene editing win chemistry Nobel*, in *Nature*. 2020. Available from: <https://www.nature.com/articles/d41586-020-02765-9>.
474. Ferlin, J., et al., *Investigation of the role of GBF1 in the replication of positive-sense single-stranded RNA viruses*. Journal of General Virology, 2018. **99**(8): p. 1086-1096.
475. Lee, C., *CRISPR/Cas9-Based Antiviral Strategy: Current Status and the Potential Challenge*. Molecules, 2019. **24**(7): p. 1349.
476. van Diemen, F.R., et al., *CRISPR/Cas9-Mediated Genome Editing of Herpesviruses Limits Productive and Latent Infections*. PLOS Pathogens, 2016. **12**(6): p. e1005701.
477. Tang, N., et al., *Application of CRISPR-Cas9 Editing for Virus Engineering and the Development of Recombinant Viral Vaccines*. The CRISPR Journal, 2021. **4**(4): p. 477-490.
478. Freije, C.A., et al., *Programmable Inhibition and Detection of RNA Viruses Using Cas13*. Molecular Cell, 2019. **76**(5): p. 826-837.e11.
479. Cockrell, A.S., et al., *A mouse model for MERS coronavirus-induced acute respiratory distress syndrome*. Nature Microbiology, 2016. **2**(2): p. 16226.
480. de Wilde, A.H., et al., *Coronaviruses and arteriviruses display striking differences in their cyclophilin A-dependence during replication in cell culture*. Virology, 2018. **517**: p. 148-156.
481. Peng, Q., et al., *Rapid manipulation of the porcine epidemic diarrhea virus genome by CRISPR/Cas9 technology*. Journal of Virological Methods, 2020. **276**: p. 113772.

482. Barnes, T., *World Health Organisation fears new 'Disease X' could cause a global pandemic*, in *The Independent*. 2018: London, U.K. Available from: <https://www.independent.co.uk/news/science/disease-x-what-is-infection-virus-world-health-organisation-warning-ebola-zika-sars-a8250766.html>.
483. R.J., *What is Disease X?*, in *The Economist*. 2018. Available from: <https://www.economist.com/the-economist-explains/2018/03/23/what-is-disease-x>.
484. World Health Organization, *2018 Annual Review of Diseases Prioritized under the Research & Development Blueprint*. 2018. Available from: <https://www.who.int/docs/default-source/blue-print/2018-annual-review-of-diseases-prioritized-under-the-research-and-development-blueprint.pdf>.
485. World Health Organization, *List of Blueprint Priority Diseases*. 2018. Available from: <https://web.archive.org/web/20200301083134/http://origin.who.int/blueprint/priority-diseases/en/>.
486. Daszak, P., *We Knew Disease X Was Coming. It's Here Now.*, in *New York Times*. 2020. Available from: <https://www.nytimes.com/2020/02/27/opinion/coronavirus-pandemics.html>.
487. Weinstein, E.S., et al., *China's State Key Laboratory System: A View into China's Innovation System*. 2022, Georgetown University Center for Security and Emerging Technology. Available from: <https://cset.georgetown.edu/publication/map-of-chinas-state-key-laboratory-system>.
488. Cui, J., F. Li, and Z.-L. Shi, *Origin and evolution of pathogenic coronaviruses*. *Nature Reviews Microbiology*, 2019. **17**(3): p. 181-192.
489. Jiang, S. and Z.-L. Shi, *The First Disease X is Caused by a Highly Transmissible Acute Respiratory Syndrome Coronavirus*. *Virologica Sinica*, 2020. **35**(3): p. 263-265.
490. Li, W. and H. Wang, *Wildlife Trade in Yunnan Province, China, at the Border with Vietnam*. *Traffic Bulletin*, 1999. **18**(1): p. 21-30.
491. Zhang, L., N. Hua, and S. Sun, *Wildlife trade, consumption and conservation awareness in southwest China*. *Biodiversity and Conservation*, 2008. **17**(6): p. 1493-1516.
492. Bell, D., S. Robertson, and P.R. Hunter, *Animal origins of SARS coronavirus: possible links with the international trade in small carnivores*. *Philos Trans R Soc Lond B Biol Sci*, 2004. **359**(1447): p. 1107-14.
493. Lau, M.W.-N., J.R. Fellowes, and B.P.L. Chan, *Carnivores (Mammalia: Carnivora) in South China: a status review with notes on the commercial trade*. *Mammal Review*, 2010. **40**(4): p. 247-292.
494. Chow, A.T., S. Cheung, and P.K. Yip, *Wildlife markets in South China*. *Human-Wildlife Interactions*, 2014. **8**(1): p. 108-112.
495. Hassanin, A., V. Tu, and T. Görföl, *Phylogeographic evolution of horseshoe bat sarbecoviruses in Vietnam and implications for the origins of SARS-CoV and SARS-CoV-2*. *ResearchSquare*, 2023.
496. Temmam, S., et al., *Bat coronaviruses related to SARS-CoV-2 and infectious for human cells*. *Nature*, 2022. **604**(7905): p. 330-336.
497. Wang, J., et al., *Individual bat virome analysis reveals co-infection and spillover among bats and virus zoonotic potential*. *Nature Communications*, 2023. **14**(1): p. 4079.

498. Chen, S., et al., *Emergence of crucial evidence catalyzing the origin tracing of SARS-CoV-2*. PLOS ONE, 2024. **19**(8): p. e0309557.
499. Delaye, L. and L. Román-Padilla, *Untangling the Evolution of the Receptor-Binding Motif of SARS-CoV-2*. Journal of Molecular Evolution, 2024. **92**(3): p. 329-337.
500. Latinne, A., et al., *Origin and cross-species transmission of bat coronaviruses in China*. Nature Communications, 2024. **15**(1): p. 10705.
501. Wang, X., et al., *Identification of coronaviruses in farmed wild animals reveals their evolutionary origins in Guangdong, southern China*. Virus Evolution, 2022. **8**(1).
502. Ou, X., et al., *Host susceptibility and structural and immunological insight of S proteins of two SARS-CoV-2 closely related bat coronaviruses*. Cell Discov, 2023. **9**(1): p. 78.
503. LaTourrette, K., et al., *Genome-Wide Variation in Betacoronaviruses*. Journal of Virology, 2021. **95**(15): p. 10.1128/jvi.00496-21.
504. Patiño-Galindo, J., et al., *Recombination across distant coronavirid species and genera is a rare event with distinct genomic features*. Journal of Virology, 2024. **98**(12): p. e01100-24.
505. Tang, X., et al., *On the origin and continuing evolution of SARS-CoV-2*. National Science Review, 2020. **7**(6): p. 1012-1023.
506. Zhou, H., et al., *Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses*. Cell, 2021. **184**(17): p. 4380-4391.e14.
507. Tan, C.C.S., L. van Dorp, and F. Balloux, *The evolutionary drivers and correlates of viral host jumps*. Nature Ecology & Evolution, 2024. **8**(5): p. 960-971.
508. Latinne, A., et al., *Origin and cross-species transmission of bat coronaviruses in China*. Nat Commun, 2020. **11**(1): p. 4235.
509. Hassanin, A., et al., *Inferring the ecological niche of bat viruses closely related to SARS-CoV-2 using phylogeographic analyses of Rhinolophus species*. Sci Rep, 2021. **11**(1): p. 14276.
510. Sánchez, C.A., et al., *A strategy to assess spillover risk of bat SARS-related coronaviruses in Southeast Asia*. Nature Communications, 2022. **13**(1): p. 4380.
511. Barber, R.M., et al., *Estimating global, regional, and national daily and cumulative infections with SARS-CoV-2 through Nov 14, 2021: a statistical analysis*. The Lancet, 2022. **399**(10344): p. 2351-2380.
512. Bergeri, I., et al., *Global SARS-CoV-2 seroprevalence from January 2020 to April 2022: A systematic review and meta-analysis of standardized population-based studies*. PLOS Medicine, 2022. **19**(11): p. e1004107.
513. Mautner, L., et al., *Replication kinetics and infectivity of SARS-CoV-2 variants of concern in common cell culture models*. Virology Journal, 2022. **19**(1): p. 76.
514. Li, W., et al., *Virus shedding dynamics in asymptomatic and mildly symptomatic patients infected with SARS-CoV-2*. Clinical Microbiology and Infection, 2020. **26**(11): p. 1556.e1-1556.e6.
515. Yin, Q., et al., *Comprehensive genomic analysis of the SARS-CoV-2 Omicron variant BA.2.76 in Jining City, China, 2022*. BMC Genomics, 2024. **25**(1): p. 378.
516. Robson, F., et al., *Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting*. Molecular Cell, 2020. **79**(5): p. 710-727.

517. Grellet, E., et al., *Replication of the coronavirus genome: A paradox among positive-strand RNA viruses*. Journal of Biological Chemistry, 2022. **298**(5).
518. Wang, H., et al., *Recombination in Positive-Strand RNA Viruses*. Frontiers in Microbiology, 2022. **13**.
519. Bonavita, C.M., H.L. Wells, and S.J. Anthony, *Cellular dynamics shape recombination frequency in coronaviruses*. PLOS Pathogens, 2024. **20**(9): p. e1012596.
520. Van Kerkhove, M.D., *Share all SARS-CoV-2 data immediately*. Science, 2023. **380**(6640): p. 11-11.
521. World Health Organization, *SAGO statement on newly released SARS-CoV-2 metagenomics data from China CDC on GISAIID*. 2023. Available from: <https://www.who.int/news/item/18-03-2023-sago-statement-on-newly-released-sars-cov-2-metagenomics-data-from-china-cdc-on-gisaid>.
522. Garry, R.F., *Slack comments about proximal origins of COVID-19 2020*. Available from: https://usrtk.org/wp-content/uploads/2023/10/Proximal_Origin_Slack_OCRd.pdf.
523. Wilf, S., *COVID-19 Origin Debate Results*. 2024. Available from: <https://cards.rootclaim.com/analysis/65892f07d97bdb00199d4366>.
524. Li, Z., et al., *Antibody seroprevalence in the epicenter Wuhan, Hubei, and six selected provinces after containment of the first epidemic wave of COVID-19 in China*. The Lancet Regional Health – Western Pacific, 2021. **8**.
525. He, Z., et al., *Seroprevalence and humoral immune durability of anti-SARS-CoV-2 antibodies in Wuhan, China: a longitudinal, population-level, cross-sectional study*. The Lancet, 2021. **397**(10279): p. 1075-1084.
526. Latham, J., *The Great Raccoon Dog Mystery*. Independent Science News, 2023.
527. World Health Organization, *Novel Coronavirus (2019-nCoV) Situation Report #2*. 2020.
528. Monagin, C., et al., *Serologic and behavioral risk survey of workers with wildlife contact in China*. PLOS ONE, 2018. **13**(4): p. e0194647.
529. Evans, T.S., et al., *Exposure to diverse sarbecoviruses indicates frequent zoonotic spillover in human communities interacting with wildlife*. International Journal of Infectious Diseases, 2023. **131**: p. 57-64.
530. Pruvot, M., et al., *Toward a quantification of risks at the nexus of conservation and health: The case of bushmeat markets in Lao PDR*. Science of the Total Environment, 2019. **676**: p. 732-745.
531. Kuhn, T.S., *The Structure of Scientific Revolutions*. 1962: Univ. Chicago Press.
532. Wray, K.B., et al., *Kuhn's The Structure of Scientific Revolutions at 60*. 2024: Cambridge University Press.
533. Dang, H., *Paradigm Shifts and Group Belief Change*, in *Kuhn's The Structure of Scientific Revolutions at 60*, K.B. Wray, Editor. 2024, Cambridge University Press: Cambridge. p. 147-164.
534. Kindi, V., *Kuhn and the History of Science*, in *Kuhn's The Structure of Scientific Revolutions at 60*, K.B. Wray, Editor. 2024, Cambridge University Press: Cambridge. p. 254-274.

535. Wang, F. and S. Yang, *SARS Experience Guides China's Effort to Contain New Virus*, in *Wall Street Journal*. 2020. Available from: <https://www.wsj.com/articles/sars-experience-guides-chinas-effort-to-contain-new-virus-11578681205>.
536. Sudworth, J. and S. Maybin, *COVID: Top Chinese scientist says don't rule out lab leak*, in *BBC News*. 2020. Available from: <https://www.bbc.com/news/world-asia-65708746>.
537. Cerdas, S., et al., *Mission report of delegation to the 75th World Health Assembly*. 2021, European Parliament Committee on the Environment, Public Health & Food Safety. Available from: https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/ENVI/CR/2022/10-24/Missionreport-75thWHA-Geneva_EN.pdf.
538. Lytras, S., et al., *The animal origin of SARS-CoV-2*. *Science*, 2021. **373**(6558): p. 968-970.
539. Holmes, E.C., *The COVID lab leak theory is dead. Here's how we know the virus came from a Wuhan market*, in *The Conversation*. 2022. Available from: <https://theconversation.com/the-covid-lab-leak-theory-is-dead-heres-how-we-know-the-virus-came-from-a-wuhan-market-188163>.
540. Garry, R.F., *The evidence remains clear: SARS-CoV-2 emerged via the wildlife trade*. *Proceedings of the National Academy of Sciences*, 2022. **119**(47): p. e2214427119.
541. Jiang, X. and R. Wang, *Wildlife trade is likely the source of SARS-CoV-2*. *Science*, 2022. **377**(6609): p. 925-926.
542. Maxmen, A., *Wuhan market was epicentre of pandemic's start, studies suggest*. *Nature*, 2022.
543. Gallagher, J., *Genetic ghosts suggest Covid's market origins*. 2024, BBC. Available from: <https://www.bbc.com/news/articles/cy8095xjg4po>.
544. Mallapaty, S., *Sick animals suggest COVID pandemic started in Wuhan market*, in *Nature*. 2024. Available from: <https://www.nature.com/articles/d41586-024-03968-0>.
545. Woodford, J., *Evidence points to Wuhan market as Source of COVID-19 Outbreak*, in *New Scientist*. 2024. Available from: <https://www.newscientist.com/article/2448671-evidence-points-to-wuhan-market-as-source-of-covid-19-outbreak/>.
546. Alwine, J., et al., *The harms of promoting the lab leak hypothesis for SARS-CoV-2 origins without evidence*. *Journal of Virology*, 2024. **98**(9): p. e01240-24.
547. PRC National People's Congress, *Complete Ban of Illegal Wildlife Trade and the Elimination of the Unhealthy Habit of Indiscriminate Wild Animal Meat Consumption For the Protection of Human Life and Health*. 2020. Available from: http://www.npc.gov.cn/englishnpc/lawsofthepc/202003/t20200302_384868.html.
548. China Leather Industry Association, *China Fur Commission*. 2013.
549. China Leather Industry Association, *Statistical Report on the Skins of Mink, Foxes and Raccoon Dogs in China (2023)*. 2024. Available from: <https://www.chinaleather.org/front/article/131902/>.
550. Humane Society International. *The Fur Trade*. 2024 [30 December 2024]; Available from: <https://www.hsi.org/news-resources/fur-trade/>.
551. U.S. National Agricultural Statistics Service, *Mink Methodology and Quality Measures 2024*. Available from: https://data.nass.usda.gov/Publications/Methodology_and_Data_Quality/Mink/08_2024/minkqm24.pdf.

552. European Commission Directorate-General for Health & Food Safety, *Request for a scientific opinion and technical assistance concerning the welfare of fur animals*. 2023. Available from: <https://open.efsa.europa.eu/questions/EFSA-Q-2023-00869>.
553. FurFree Alliance, *The Fur Industry is in Economic Decline*. 2023. Available from: https://www.furfreealliance.com/wp-content/uploads/2023/11/Fur-Free_Factsheets_A4_Fur-industry-is-economically-unviable_13.pdf.
554. Dimande, P., M. Arrobas, and M.Â. Rodrigues, *Under a Tropical Climate and in Sandy Soils, Bat Guano Mineralises Very Quickly, Behaving More like a Mineral Fertiliser than a Conventional Farmyard Manure*. *Agronomy*, 2023. **13**(5): p. 1367.
555. Marwa, E., T. Andrew, and A. Hatibu, *Challenges Facing Effective use of Bat Guano as Organic Fertilizer in Crop Production: A Review*. *International Journal of Engineering and Applied Sciences (IJEAS)*, 2021. **8**.
556. Aragão Silva, J.A., et al., *Use of wild vertebrates for consumption and bushmeat trade in Brazil: a review*. *Journal of Ethnobiology and Ethnomedicine*, 2023. **19**(1): p. 64.
557. Jagadesh, S., et al., *Mapping Global Bushmeat Activities to Improve Zoonotic Spillover Surveillance by Using Geospatial Modeling*. *Emerging Infectious Disease journal*, 2023. **29**(4): p. 742.
558. Luiselli, L., et al., *Understanding the influence of non-wealth factors in determining bushmeat consumption: Results from four West African countries*. *Acta Oecologica*, 2019. **94**: p. 47-56.
559. Ferreira da Silva, J., et al., *Click editing enables programmable genome writing using DNA polymerases and HUH endonucleases*. *Nature Biotechnology*, 2024.
560. Aguilar, G., et al., *Seamless knockins in Drosophila via CRISPR-triggered single-strand annealing*. *Developmental Cell*, 2024.
561. Liu, P., et al., *Design of customized coronavirus receptors*. *Nature*, 2024. **635**(8040): p. 978-986.
562. Collins, M., et al., *A frugal CRISPR kit for equitable and accessible education in gene editing and synthetic biology*. *Nature Communications*, 2024. **15**(1): p. 6563.