CellPress
OPEN ACCESS

Cell
Leading Edge

Review

The origins of SARS-CoV-2: A critical review

Edward C. Holmes,[1,*] Stephen A. Goldstein,[2] Angela L. Rasmussen,[3] David L. Robertson,[4] Alexander Crits-Christoph,[5] Joel O. Wertheim,[6] Simon J. Anthony,[7] Wendy S. Barclay,[8] Maciej F. Boni,[9] Peter C. Doherty,[10] Jeremy Farrar,[11] Jemma L. Geoghegan,[12,13] Xiaowei Jiang,[14] Julian L. Leibowitz,[15] Stuart J.D. Neil,[16] Tim Skern,[17] Susan R. Weiss,[18] Michael Worobey,[19] Kristian G. Andersen,[20] Robert F. Garry,[21,22] and Andrew Rambaut[23,*]

[1]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia
[2]Department of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA
[3]Vaccine and Infectious Disease Organization, University of Saskatchewan, Saskatoon, SK S7N 5E3, Canada
[4]MRC-University of Glasgow Centre for Virus Research, Glasgow G61 1QH, UK
[5]Department of Plant and Microbial Biology, University of California-Berkeley, Berkeley, CA 94704, USA
[6]Department of Medicine, University of California-San Diego, La Jolla, CA 92093, USA
[7]Department of Pathology, Microbiology, and Immunology, University of California-Davis School of Veterinary Medicine, Davis, CA 95616, USA
[8]Department of Infectious Disease, Imperial College, London W2 1PG, UK
[9]Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA
[10]Department of Microbiology and Immunology, The University of Melbourne at the Doherty Institute, 792 Elizabeth Street, Melbourne, VIC 3000, Australia
[11]The Wellcome Trust, London NW1 2BE, UK
[12]Department of Microbiology and Immunology, University of Otago, Dunedin 9010, New Zealand
[13]Institute of Environmental Science and Research, Wellington 5022, New Zealand
[14]Department of Biological Sciences, Xi'an Jiaotong-Liverpool University (XJTLU), Suzhou, China
[15]Department of Microbial Pathogenesis and Immunology, Texas A&M University, College Station, TX 77807, USA
[16]Department of Infectious Diseases, King's College London, Guy's Hospital, London SE1 9RT, UK
[17]Max Perutz Labs, Medical University of Vienna, Vienna Biocenter, Dr. Bohr-Gasse 9/3, 1030 Vienna, Austria
[18]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[19]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA
[20]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037, USA
[21]Department of Microbiology and Immunology, Tulane University School of Medicine, New Orleans, LA 70112, USA
[22]Zalgen Labs, Germantown, MD 20876, USA
[23]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, UK
*Correspondence: edward.holmes@sydney.edu.au (E.C.H.), a.rambaut@ed.ac.uk (A.R.)
https://doi.org/10.1016/j.cell.2021.08.017

SUMMARY

Since the first reports of a novel severe acute respiratory syndrome (SARS)-like coronavirus in December 2019 in Wuhan, China, there has been intense interest in understanding how severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in the human population. Recent debate has coalesced around two competing ideas: a "laboratory escape" scenario and zoonotic emergence. Here, we critically review the current scientific evidence that may help clarify the origin of SARS-CoV-2.

EVIDENCE SUPPORTING A ZOONOTIC ORIGIN OF SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS 2

Coronaviruses have long been known to present a high pandemic risk. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the ninth documented coronavirus that infects humans and the seventh identified in the last 20 years (Lednicky et al., 2021; Vlasova et al., 2021). All previous human coronaviruses have zoonotic origins, as have the vast majority of human viruses. The emergence of SARS-CoV-2 bears several signatures of these prior zoonotic events. It displays clear similarities to SARS-CoV that spilled over into humans in Foshan, Guangdong province, China in November 2002, and again in Guangzhou, Guangdong province in 2003 (Xu et al., 2004). Both these SARS-CoV emergence events were associated with markets selling live animals and involved species, particularly civets and raccoon dogs (Guan et al., 2003), that were also sold live in Wuhan markets in 2019 (Xiao et al., 2021) and are known to be susceptible to SARS-CoV-2 infection (Freuling et al., 2020). Animal traders working in 2003, without a SARS diagnosis, were documented to have high levels of immunoglobulin G (IgG) to SARS-CoV (13% overall and >50% for traders specializing in civets) (Centers for Disease Control and Prevention, 2003). Subsequent serological surveys found ~3% positivity rates to SARS-related coronaviruses (SARSr-CoV) in residents of Yunnan province

Organization, 2021), this is expected given high rates of asymptomatic transmission and undocumented secondary transmission events and was similarly observed in early SARS-CoV cases in Foshan (Xu et al., 2004).

During 2019, markets in Wuhan—including the Huanan market—traded many thousands of live wild animals including high-risk species such as civets and raccoon dogs (Xiao et al., 2021). Following its closure, SARS-CoV-2 was detected in environmental samples at the Huanan market, primarily in the western section that traded in wildlife and domestic animal products, as well as in associated drainage areas (World Health Organization, 2021). Although animal carcasses retrospectively tested negative for SARS-CoV-2, these were unrepresentative of the live animal species sold and specifically did not include raccoon dogs and other animals known to be susceptible to SARS-CoV-2 (Xiao et al., 2021).

The earliest split in the SARS-CoV-2 phylogeny defines two lineages—denoted A and B (Rambaut et al., 2020)—that likely circulated contemporaneously (Figure 1A). Lineage B, which became dominant globally, was observed in early cases linked to the Huanan market and environmental samples taken there, whereas lineage A contains a case with exposure to other markets (Figures 1A and 1B) as well as with later cases in Wuhan and other parts of China (World Health Organization, 2021). This phylogenetic pattern is consistent with the emergence of SARS-CoV-2 involving one or more contacts with infected animals and/or traders, including multiple spill-over events, as potentially infected or susceptible animals were moved into or between Wuhan markets via shared supply chains and sold for human consumption (Xiao et al., 2021). The potential emergence of SARS-CoV-2 across multiple markets again mirrors SARS-CoV in which high levels of infection, seroprevalence, and genetic diversity in animals were documented at both the Dongmen market in Shenzhen (Yaqing, 2004; Guan et al., 2003) and the Xinyuan market in Guangzhou (Tu et al., 2004; Wang et al., 2005).

Viruses closely related to SARS-CoV-2 have been documented in bats and pangolins in multiple localities in South-East Asia, including in China, Thailand, Cambodia, and Japan (Lytras et al., 2021; Zhou et al., 2021), with serological evidence for viral infection in pangolins for more than a decade (Wacharapluesadee et al., 2021). However, a significant evolutionary gap exists between SARS-CoV-2 and the closest related animal viruses: for example, the bat virus RaTG13 collected by the WIV has a genetic distance of ~4% (~1,150 mutations) to the Wuhan-Hu-1 reference sequence of SARS-CoV-2, reflecting decades of evolutionary divergence (Boni et al., 2020). Widespread genomic recombination also complicates the assignment of which viruses are closest to SARS-CoV-2. Although RaTG13, sampled from a *Rhinolophus affinis* bat in Yunnan (Zhou et al., 2020b), has the highest average genetic similarity to SARS-CoV-2, a history of recombination means that three other bat viruses—RmYN02, RpYN06, and PrC31—are closer in most of the virus genome (particularly ORF1ab) and thus share a more recent common ancestor with SARS-CoV-2 (Li et al., 2021; Lytras et al., 2021; Zhou et al., 2021). None of these three closer viruses were collected by the WIV and all were sequenced after the pandemic had begun (Li et al., 2021; Zhou et al., 2020a, 2021). Collectively, these data demonstrate beyond reasonable doubt that RaTG13 is not the progenitor of SARS-CoV-2, with or without laboratory manipulation or experimental mutagenesis.

No bat reservoir or intermediate animal host for SARS-CoV-2 has been identified to date. This is presumably because the right animal species and/or populations have not yet been sampled and/or any progenitor virus may be at low prevalence. Initial cross-species transmission events are also very likely to go undetected. Most SARS-CoV-2 index case infections will not have resulted in sustained onward transmission (Pekar et al., 2021), and only a very small fraction of spillovers from animals to humans result in major outbreaks. Indeed, the animal origins of many well-known human pathogens, including Ebola virus, hepatitis C virus, poliovirus, and the coronaviruses HCoV-HKU1 and HCoV-NL63, are yet to be identified, while it took over a decade to discover bat viruses with >95% similarity to SARS-CoV and able to use hACE-2 as a receptor (Hu et al., 2017).

COULD SARS-CoV-2 HAVE ESCAPED FROM A LABORATORY?

There are precedents for laboratory incidents leading to isolated infections and transient transmission chains, including SARS-CoV (Parry, 2004). However, with the exception of Marburg virus (Ristanović et al., 2020), all documented laboratory escapes have been of readily identifiable viruses capable of human infection and associated with sustained work in high titer cultures (Geddes, 2006; Lim et al., 2004; Senior, 2003). The 1977 A/H1N1 influenza pandemic, that most likely originated from a large-scale vaccine challenge trial (Rozo and Gronvall, 2015), is the only documented example of a human epidemic or pandemic resulting from research activity. No epidemic has been caused by the escape of a novel virus, and there is no data to suggest that the WIV—or any other laboratory—was working on SARS-CoV-2, or any virus close enough to be the progenitor, prior to the COVID-19 pandemic. Viral genomic sequencing without cell culture, which was routinely performed at the WIV, represents a negligible risk because viruses are inactivated during RNA extraction (Blow et al., 2004). No case of laboratory escape has been documented following the sequencing of viral samples.

Known laboratory outbreaks have been traced to both workplace and family contacts of index cases and to the laboratory of origin (Geddes, 2006; Lim et al., 2004; Ristanović et al., 2020; Senior, 2003). Despite extensive contact tracing of early cases during the COVID-19 pandemic, there have been no reported cases related to any laboratory staff at the WIV, and all staff in the laboratory of Dr. Shi Zhengli were said to be seronegative for SARS-CoV-2 when tested in March 2020 (World Health Organization, 2021), with the laboratory reportedly following the appropriate biosafety protocols during their coronavirus work (Cohen, 2020). During a period of high influenza transmission and other respiratory virus circulation (Liu et al., 2020a), reports of illnesses would need to be confirmed as caused by SARS-CoV-2 to be relevant. Epidemiological modeling suggests that the number of hypothetical cases needed to result in multiple hospitalized COVID-19 patients prior to December 2019 is incompatible with observed clinical, genomic, and epidemiological data (Pekar et al., 2021).