Tracing the Origin of SARS-CoV-2: Lessons Learned from the Past

Authors: Qihui Wang, Hua Chen, Yi Shi, Alice C. Hughes, William J. Liu, Jingkun Jiang, George F. Gao, Yongbiao Xue, Yigang Tong, Qihui Wang

Date: 2021-07-28T00:00:00+00:00

Abstract

The origin of SARS-CoV-2 remains unclear. Understanding how, when, and where SARS-CoV-2 transmitted from its natural host to humans is crucial for preventing future coronavirus outbreaks. Drawing on lessons from the continuous battle against pathogens and integrating currently available research data on the origin of SARS-CoV-2 and its intermediate hosts, we propose that multiple locations worldwide are possible origins of SARS-CoV-2.

Full Text

Tracing the Origin of SARS-CoV-2: Lessons Learned from the Past

Qihui Wang[1], Hua Chen[2], Yi Shi[1], Alice C. Hughes[3], William J. Liu[4], Jingkun Jiang[5], George F. Gao[1], Yongbiao Xue[2], Yigang Tong[6]

[1]CAS Key Laboratory of Pathogen Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China
[2]Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Centre for Bioinformation, Beijing, 100101, China
[3]Landscape Ecology Group, Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, 666303, China
[4]National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, 102206, China
[6]Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, P. R. China

Correspondence: Qihui Wang (wangqihui@im.ac.cn)

Abstract

The origin of SARS-CoV-2 remains elusive. Understanding how, when, and where SARS-CoV-2 was transmitted from its natural reservoir to human beings is crucial for preventing future coronavirus outbreaks. With lessons learned from the endless battle with pathogens and accumulated research data regarding origins and intermediate hosts, we present multiple potential locations as natural reservoirs of SARS-CoV-2.

Introduction

Emerging and re-emerging infectious diseases pose a significant threat to human health, economy, and security worldwide. In recent years, we have witnessed the emergence of novel pathogens at an accelerating rate[1], most of which are zoonotic, including Nipah virus, influenza virus, and especially coronaviruses (CoVs). After the outbreaks of severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), researchers worldwide have reached a consensus that the next CoV spillover event was only a matter of time, as supported by research data and the natural laws of pathogen emergence[3]. In other words, the SARS-CoV-2 outbreak is actually a gray rhino event that was predicted by professionals.

To change this upward trend and prevent future spillover events, it is crucial to identify the origins and intermediate hosts of known pathogens. Important lessons must be learned from the endless battle between humans and their pathogens.

Lessons from Past Outbreaks

First, determining the origins of a pathogen requires solid evidence. Specifically, a highly similar virus must be identified from an animal that shares an ecological link with the virus' s reservoir host or a known intermediate host. The origin tracing of MERS-CoV provides a clear example. Strong evidence indicates that the 2012 MERS-CoV outbreak was driven by a dromedary-to-human spillover event[4], but the animal that transmitted MERS-CoV to dromedaries remains unclear. Bats are hypothesized to be the natural reservoir for MERS-CoV because bat CoV HKU4 displays sequence homology and similar receptor binding patterns with MERS-CoV, suggesting that MERS-CoV may be an HKU4-related virus that originated from bats. However, no virus with a genome highly homologous to MERS-CoV has been identified from any bat species[4], which prevents definitive conclusions regarding MERS-CoV' s origin. In contrast, swine acute diarrhea syndrome coronavirus (SADS-CoV), which causes piglet mortality, was quickly determined to be bat-origin after its outbreak because a highly similar virus (98.48% identity), bat CoV HKU2, was found in bats living in a cave near infected pig farms[4].

Second, tracing viral origins can require decades of continuous research, but accumulated data form the foundation for future origin tracing capabilities. For example, influenza A virus has long been known to circulate in wild aquatic birds and can be transmitted to other avian and mammalian hosts through mutation and reassortment[5]. Over the past century, extensive surveillance of influenza A viruses in animals and humans has created an enormous genome sequence database. Using this database, the origins of newly emergent influenza A strains have been quickly traced, such as the H1N1 pandemic strain in 2009 and the H7N9 avian influenza strain in 2013[6],[7].

Third, the location of the first outbreak may be far from the place of origin. Human immunodeficiency virus (HIV) provides a compelling example. HIV was believed to have originated in the United States when first identified in the 1980s. Since then, scientists and health workers have become increasingly aware of HIV and officially recognized AIDS as a new human infectious disease. However, subsequent studies discovered a 1959 blood sample with HIV from a man living in Kinshasa, Democratic Republic of the Congo, confirming the first verified HIV case in Africa[8]. Thus, the location where a new infectious disease is reported may not be where it originally emerged.

Applying Lessons to SARS-CoV-2 Origin Tracing

Tracing SARS-CoV-2’ s origins requires applying these historical lessons. First, a virus progenitor with strong similarity to SARS-CoV-2 must be found in a geographically and ecologically relevant animal before drawing conclusions. Second, origin tracing must not rush to judgment before accumulating sufficient evidence. Third, we must remember that the first outbreak location may not be the place of origin.

To find SARS-CoV-2’s progenitor in animals, numerous SARS-related CoVs (sarbecoviruses) worldwide have been investigated, including RaTG13/RaTG15/RmYN02 (southern China), RshSTT182/RshSTT200 (Cambodia), Rc-o139 (Japan), RacCS203 (Thailand), BM48-31 (Bulgaria), and BtKY72 (Kenya)[9]. Notably, all these sarbecoviruses were discovered in bats of the *Rhinolophus* genus[9], making *Rhinolophus* bats potential reservoir hosts. However, RaTG13, the closest known sarbecovirus related to SARS-CoV-2, still displays significant differences in genome sequence, receptor binding pattern, and host range[10], suggesting that bats as natural hosts remains inconclusive. According to the World Health Organization (WHO)-convened Global Study of Origins of SARS-CoV-2: China Part (hereafter “WHO report”), direct zoonotic spillover is considered a possible-to-likely pathway[9]. Therefore, a global search for natural reservoirs carrying SARS-CoV-2-like viruses is urgently needed.

The WHO report also concluded that SARS-CoV-2 introduction through an intermediate host is likely to very likely[9]. To identify potential intermediate hosts, numerous mammalian species have been investigated, including domesticated animals (e.g., horses, pigs, cows), companion animals (e.g., cats, dogs), and wild animals (e.g., bats, pangolins, minks, foxes, civets). Research shows that the angiotensin-converting enzyme 2 (ACE2) receptor from many species binds the SARS-CoV-2 receptor-binding domain (RBD) with affinity similar to human ACE2, suggesting potential cross-species transmission paths[11]. Among possible intermediate hosts, pangolins and minks have attracted particular attention. Pangolins host at least two CoVs, GX/P2V/2017 and GD/1/2019, closely related to SARS-CoV-2[12]. Minks may also be intermediate hosts because the only reported SARS-CoV-2 animal outbreak occurred in European mink populations, indicating SARS-CoV-2 is well-adapted to minks and that minks may have played an important role in SARS-CoV-2 evolution[13]. All possibilities must be considered to unravel the intermediate host mystery.

Geographic Distribution of Potential Hosts

Cross-species transmission from reservoir to intermediate host requires host proximity and ecological links. Therefore, SARS-CoV-2' s origin location could be where *Rhinolophus* bat distribution overlaps with pangolins, minks, or other potential intermediate hosts. Mustelids (including mink) are distributed across the entire Old World. We mapped the distribution of 98 *Rhinolophus* species, eight pangolin species, and wild European mink (*Mustela lutreola*), along with major mink farm distribution areas[13], and marked bat sarbecovirus discovery locations and international flight routes to Wuhan (Figure 1 [Figure 1: see original paper]).

Rhinolophus species distribution covers southern Eurasia, Southeast Asian islands, and most of sub-Saharan Africa, overlapping with pangolins in southern China, Southeast Asia, India, and sub-Saharan Africa. European mink is distributed across Europe, overlapping with *Rhinolophus* in southern Europe. However, most Eurasian minks are millions of American minks (*Neovison vison*) in farms across various European countries and China[13], overlapping with *Rhinolophus* in southern European countries (Italy, Greece, Spain, France) and some northern Chinese provinces.

Figure 1. Distribution of *Rhinolophus*, pangolin, and mink species, showing bat sarbecovirus discovery locations and main mink farm distribution areas[13]. Red lines indicate international flight routes to Wuhan. Animal distribution data are from the International Union for Conservation of Nature (IUCN) Red List of Threatened Species database (https://www.iucnredlist.org/). Air route information is from Wuhan Tianhe Airport website (http://www.whairport.com/).

These data suggest multiple worldwide locations where SARS-CoV-2 could transmit from natural reservoirs to intermediate hosts, even before considering other potential hosts and intermediate hosts (such as other native carnivores) distributed across the Old World. Specifically, sarbecovirus spillover from *Rhinolophus* to pangolins could occur in Southeast Asia, southern China, India, and sub-Saharan Africa, while *Rhinolophus*-to-mink transmission could occur in southern Europe. Both routes could lead to viral adaptation and potential human infection. Importantly, most regions show evidence of bat sarbecovirus circulation, allowing multiple SARS-CoV-2-like viruses to evolve independently. Therefore, sarbecovirus surveillance in *Rhinolophus* bats, pangolins, and minks from these regions is needed before determining SARS-CoV-2’ s origin.

Temporal and Spatial Considerations

Beyond host distribution, evolutionary analyses can help locate SARS-CoV-2’ s origin. Accurate inference of the time to most recent common ancestor (TMRCA) and initial evolutionary trajectories of early SARS-CoV-2 sequences would facilitate origin determination.

The TMRCA of early SARS-CoV-2 sequences was inferred to be November 28, 2019, with a 95% CI of [Oct 20, 2019, Dec 9, 2019], indicating COVID-19 may have originated earlier and outside Wuhan Seafood Market[14]. Furthermore, constructing a haplotype network of early SARS-CoV-2 genomes reveals two primary lineage clades; samples from Huanan Seafood Market cluster with descendant rather than ancestral lineages, suggesting the market’ s CoV source was likely imported[15].

Additionally, as a central China international communication hub, Wuhan received extensive global flights before the pandemic (Figure 1). Notably, many flights to Wuhan departed from Southeast Asian countries overlapping with *Rhinolophus* and pangolin distributions and known sarbecoviruses. As mentioned in the WHO report, cold/food chain product introduction is considered a possible pathway. Therefore, Wuhan was already at high risk of importing SARS-CoV-2 via cold chain cargo before the pandemic.

Conclusion

In conclusion, as stated in the WHO report, direct zoonotic spillover is possible-to-likely, while introduction through an intermediate host is likely-to-very-likely. SARS-CoV-2 introduction via cold/food chain products is possible, while a laboratory incident causing the outbreak is extremely unlikely. More evidence must be collected to identify SARS-CoV-2’ s origins, intermediate hosts, and transmission pathways[9]. Tracing viral origins and intermediate hosts is difficult; solid conclusions result from enormous work, patience, global cooperation, some luck, and possibly decades of continuous research, as accomplished for influenza virus. Understanding species ecology, host interactions, and landscape management impacts on future spillover risks are also important considerations. Such work is indispensable for reducing the frequency of inevitable pathogen emergences and outbreak damage, as it is crucial to the common health of all mankind.

References

1. Gao GF. From “A” IV to “Z” IKV: attacks from emerging and re-emerging pathogens. Cell 2018; 172:1157-1159.

2. Plowright RK, Parrish CR, McCallum H et al. Pathways to zoonotic spillover. Nat Rev Microbiol 2017; 15:502-510.
3. Su S, Wong G, Shi W et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends Microbiol 2016; 24:490-502.
4. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 2019; 17:181-192.
5. Su S, Bi Y, Wong G, Gray GC, Gao GF, Li S. Epidemiology, evolution, and recent outbreaks of avian influenza virus in China. J Virol 2015; 89:8671-8676.
6. Smith GJ, Vijaykrishna D, Bahl J et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature 2009; 459:1122-1125.
7. Gao R, Cao B, Hu Y et al. Human infection with a novel avian-origin influenza A (H7N9) virus. N Engl J Med 2013; 368:1888-1897.
8. Sharp PM, Hahn BH. The evolution of HIV-1 and the origin of AIDS. Philos Trans R Soc Lond B Biol Sci 2010; 365:2487-2494.
9. WHO-convened global study of origins of SARS-CoV-2: China Part. Available from: https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part.
10. Liu K, Pan X, Li L et al. Binding and molecular basis of the bat coronavirus RaTG13 virus to ACE2 in humans and other species. Cell 2021; 184:3438-3451 e3410.
11. Wu L, Chen Q, Liu K et al. Broad host range of SARS-CoV-2 and the molecular basis for SARS-CoV-2 binding to cat ACE2. Cell Discov 2020; 6:68.
12. Lam TT, Jia N, Zhang YW et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature 2020; 583:282-285.
13. Fenollar F, Mediannikov O, Maurin M et al. Mink, SARS-CoV-2, and the human-animal interface. Front Microbiol 2021; 12:663815.
14. Liu Q, Zhao S, Shi CM et al. Population genetics of SARS-CoV-2: disentangling effects of sampling bias and infection clusters. Genomics Proteomics Bioinformatics 2020.
15. Yu WB, Tang GD, Zhang L, Corlett RT. Decoding the evolution and transmissions of the novel pneumonia coronavirus (SARS-CoV-2/HCoV-19) using whole genomic data. Zool Res 2020; 41:247-257.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.