

## RESEARCH HIGHLIGHT



## Tracing the origins of SARS-CoV-2: lessons learned from the past

Qihui Wang<sup>1</sup>✉, Hua Chen<sup>2</sup>, Yi Shi<sup>1</sup>, Alice C. Hughes<sup>3</sup>, William J. Liu<sup>4</sup>, Jingkun Jiang<sup>5</sup>, George F. Gao<sup>1</sup>, Yongbiao Xue<sup>2</sup> and Yigang Tong<sup>6</sup>

© CEMCS, CAS 2021

*Cell Research* (2021) 31:1139–1141; <https://doi.org/10.1038/s41422-021-00575-w>

**The origins of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) remain 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 the lessons learned from the endless battle against pathogens and accumulated research data with regard to the origins and intermediate hosts, we present multiple potential locations as the natural reservoirs of SARS-CoV-2.**

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.<sup>1</sup> After the outbreaks of two zoonotic coronaviruses (CoVs), severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), researchers worldwide have reached a consensus that the occurrence of the next CoV spillover event is only a matter of time, as supported by research data and the natural laws of pathogen emergence.<sup>2</sup> In other words, the outbreak of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is actually a gray rhino event that was predicted by professionals.

Considering that the known coronaviruses are zoonotic viruses, to change such an upward trend and prevent future spillover events, it is crucial to identify the origins and intermediate hosts of known pathogens. For this purpose, lessons must be learned from the endless battle between humans and their pathogens.

First, determining the origins of a pathogen requires solid evidence. Specifically, highly similar sequence-related viruses must be identified from an animal that shares an ecological link with the virus' reservoir host or a known intermediate host. Here, we use the origin tracing of MERS-CoV as an example. Strong evidence indicates that the 2012 MERS-CoV outbreak was driven by a dromedary-to-human spillover event.<sup>3</sup> Bats are the suspected natural reservoir of MERS-CoV. However, no virus with a whole genome highly homologous to MERS-CoV, besides some similar ones, has been identified from any bat species to date,<sup>3</sup> which prevents drawing a conclusion that MERS-CoV originated from bats. In contrast, another CoV, swine acute diarrhea syndrome coronavirus (SADS-CoV), which causes the death of piglets, was

quickly determined as a bat-origin CoV after its outbreak because a highly similar virus (98.48% identity), bat CoV HKU2, was found in bats living in a cave near the infected pig farms.<sup>3</sup>

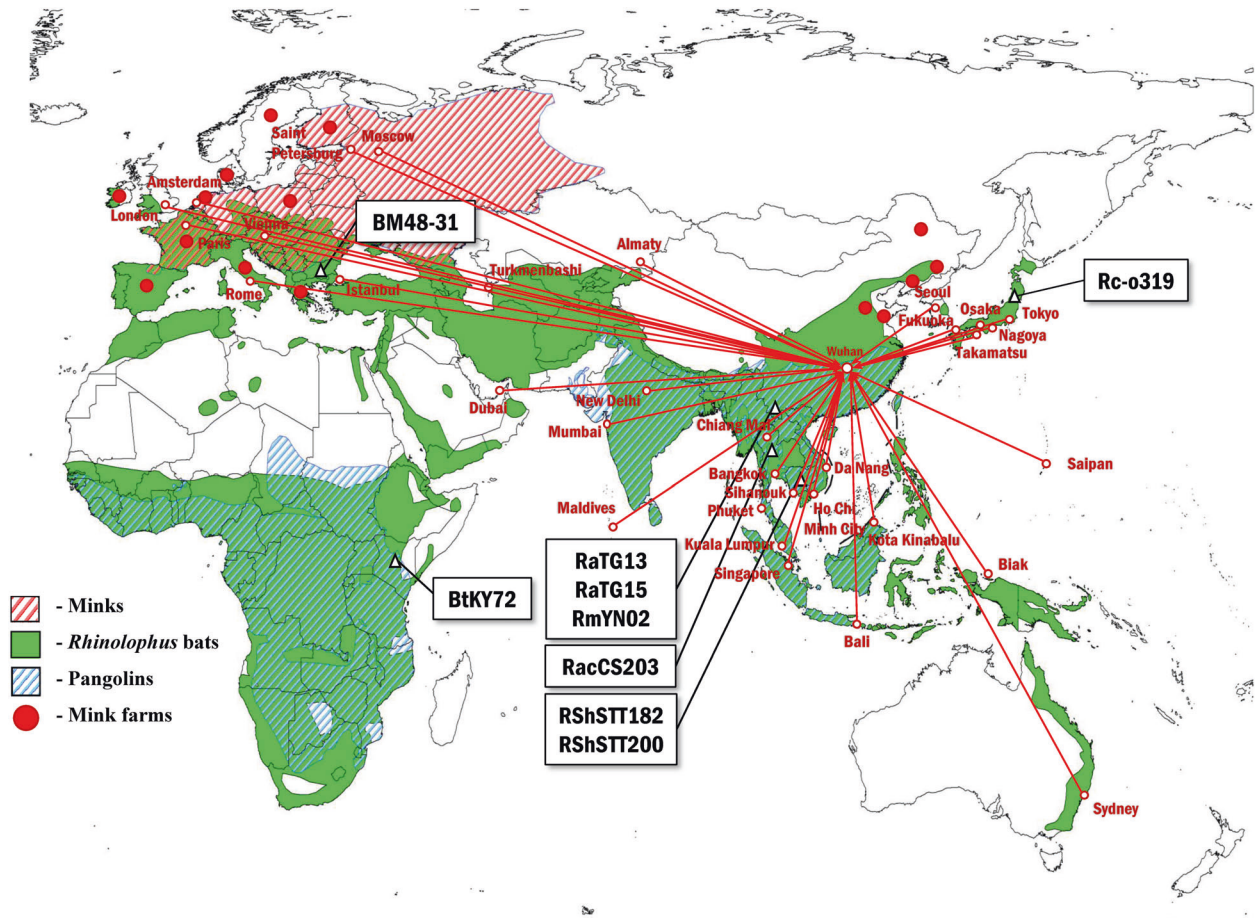
Second, tracing the origins of a virus could require decades of continuous research, but the accumulated data would lay the foundation of future origin-tracing capability. For example, it has long been known that the influenza A virus circulates in wild aquatic birds and can be transmitted to other avian and mammalian hosts.<sup>4</sup> In the past century, extensive surveillance of influenza A viruses in animals and humans has created an enormous amount of genome sequence data. Using the database that compiles these data, the origins of some newly emergent influenza A strains have been quickly traced, e.g., the H1N1 pandemic strain in 2009 and the H7N9 avian influenza strain in 2013.<sup>4,5</sup>

Third, the location of the first outbreak might be far from the place of origin. For example, human immunodeficiency virus (HIV) was believed to have originated in the United States when it was first identified in the 1980s. Since then, scientists and health workers have been increasingly aware of HIV and officially recognized AIDS as a new human infectious disease. However, subsequent studies discovered a blood sample with HIV taken in 1959 from a man living in Kinshasa in the Democratic Republic of the Congo, which confirmed the first verified case of HIV in Africa.<sup>6</sup> Thus, the place where a new infectious disease is reported may not be the original place of disease occurrence.

In spite of the widespread hypotheses/"theories" of laboratory leakage, we agree with the analysis on the genome of SARS-CoV-2 that it is unlikely a laboratory product.<sup>7</sup> Therefore, to trace the origins of SARS-CoV-2 as a zoonotic virus, it is crucial to learn from history. First, the progenitor of the virus, which has strong similarity to SARS-CoV-2, must be found from a geographically and ecologically relevant animal before drawing conclusions. Second, origin tracing must not rush to a conclusion before accumulating sufficient evidence. Third, the fact that the location of the first outbreak might not be the place of origin must be kept in mind.

To find the progenitor of SARS-CoV-2 in animals, a number of SARS-related CoVs (sarbecoviruses) from around the world have been investigated, including RaTG13/RaTG15/RmYN02 (southern China), RshSTT182/RshSTT200 (Cambodia), Rc-o139 (Japan),

<sup>1</sup>CAS Key Laboratory of Pathogen Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences (CAS), Beijing, China. <sup>2</sup>Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Centre for Bioinformation, Beijing, China. <sup>3</sup>Center of Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan, China. <sup>4</sup>National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention (China CDC), Beijing, China. <sup>5</sup>School of Environment, Tsinghua University, Beijing, China. <sup>6</sup>Beijing Advanced Innovation Center for Soft Matter Science and Engineering, College of Life Science and Technology, Beijing University of Chemical Technology, Beijing, China. ✉email: wangqihui@im.ac.cn



**Fig. 1** Distribution of *Rhinolophus*, pangolin and mink species, showing locations of bat sarbecoviruses discovered and the main distribution areas of mink farms.<sup>12</sup> Red lines indicate international flight routes to Wuhan. Animal distribution data are from the database of International Union for conservation of Nature (IUCN) Red list of Threatened Species (<https://www.iucnredlist.org/>). Air route information is from the website of Wuhan Tianhe Airport (<http://www.whairport.com/>).

RacCS203 (Thailand), BM48-31 (Bulgaria), and BtKY72 (Kenya).<sup>8</sup> Notably, the vast majority of the sarbecoviruses were discovered from bats of the *Rhinolophus* genus,<sup>8</sup> making *Rhinolophus* bats the potential reservoir hosts of SARS-CoV-2. However, as the closest known sarbecovirus related to SARS-CoV-2, RaTG13 still displays significant differences from SARS-CoV-2 with regard to its genome sequence, receptor-binding pattern, and potential host range,<sup>9</sup> whether bats represent the potential natural host of SARS-CoV-2 remains inconclusive. According to the World Health organization (WHO)-convened Global Study of Origins of SARS-CoV-2: China Part (hereafter referred to as the “WHO report”), direct zoonotic spillover is considered to be a possible-to-likely pathway.<sup>8</sup> Therefore, a global search for natural reservoirs with the potential to carry SARS-CoV-2-like viruses is urgently needed.

The WHO report also concluded that the introduction of SARS-CoV-2 through an intermediate host is considered to be a likely-to-very likely pathway.<sup>8</sup> Since the outbreak in Wuhan, a nationwide survey was quickly conducted to examine the presence of SARS-CoV-2 virus or antibody in livestock, poultry and wild animals, in order to identify potential intermediate hosts. Over 80,000 stocked or fresh samples were analyzed, but none was found to be positive.<sup>8</sup> To further search for potential intermediate hosts of SARS-CoV-2, a number of mammalian species were investigated more thoroughly, including domesticated animals (e.g., horses, pigs, and cows), companion animals (e.g., cats and dogs), and wild animals (e.g., bats, pangolins, minks, foxes, and civets). Research data show that the angiotensin-converting enzyme 2 (ACE2)

receptor from many of these species has a binding affinity to the SARS-CoV-2 receptor-binding domain (RBD) similar to human ACE2, suggesting potential cross-species transmission paths between these animals and humans.<sup>10</sup> Among the possible intermediate hosts of SARS-CoV-2, pangolins and minks have attracted more attention than others. Pangolins have been found to host at least two CoVs, GX/P2V/2017 and GD/1/2019, that are closely related to SARS-CoV-2.<sup>11</sup> Minks might also be an intermediate host because the only reported SARS-CoV-2 outbreak in animals occurred in the mink population in Europe. This indicates that SARS-CoV-2 is well adapted to minks, and minks might have played an important role in the evolution of SARS-CoV-2.<sup>12</sup> These possibilities must be taken into consideration to unravel the mystery of the intermediate host of SARS-CoV-2.

The cross-species transmission of SARS-CoV-2 from the reservoir host to the intermediate host requires that the two hosts live in proximity and share ecological links. Considering the potential reservoir hosts and intermediate hosts, the location of origin of SARS-CoV-2 could be in regions where the distribution of *Rhinolophus* bats overlaps with that of pangolins, minks, or other potential intermediate hosts. Mustelids (which includes mink) are distributed across the entire old world. Therefore, we mapped the overall distribution area of 98 *Rhinolophus* species, eight pangolin species, and the wild European mink (*Mustela lutreola*), together with the main distribution area of mink farms.<sup>12</sup> We then marked the locations where bat sarbecoviruses were discovered and international flight routes to Wuhan (Fig. 1). The distribution area

of *Rhinolophus* species covers the southern portion of the Eurasian continent, the islands of Southeast Asia, and most of sub-Saharan Africa, which overlaps with that of pangolins in southern China, Southeast Asia, India, and sub-Saharan Africa. The European mink is distributed across Europe, which overlaps with the *Rhinolophus* distribution area in southern Europe. However, the majority of minks in Eurasia are the millions of American minks (*Neovison vison*) kept in mink farms in various European countries and China,<sup>12</sup> whose distribution overlaps with the *Rhinolophus* distribution area in southern European countries such as Italy, Greece, Spain, and France, as well as some northern Chinese provinces.

These data suggest that sarbecovirus spillover from *Rhinolophus* to pangolins could occur in Southeast Asia, southern China, India, and sub-Saharan Africa, while cross-species transmission from *Rhinolophus* to minks could occur in southern Europe. Importantly, most of these regions show evidence of sarbecovirus circulation in bats, which could allow multiple SARS-CoV-2-like viruses to evolve independently. Therefore, a global search for sarbecoviruses needs to be conducted in *Rhinolophus* bats, pangolins, and minks before considering other potential intermediate hosts (such as carnivores) distributed across the old world, in order to trace the origins of SARS-CoV-2. The abovementioned places should receive a higher priority.

Aside from the distribution area of hosts, evolution analyses could also help locate the origins of SARS-CoV-2. Specifically, accurate inference of the time to the most recent common ancestor (TMRCA) and initial evolutionary trajectories of the early SARS-CoV-2 sequences would facilitate unraveling the origins of SARS-CoV-2. The TMRCA of the early SARS-CoV-2 sequences was inferred by more than 10 studies (summarized in Table 8 in "Molecular Epidemiology" section of the WHO report), and most of these point estimates are between mid-November 2019 and mid-December 2019, indicating that SARS-CoV-2 might have originated at an earlier time and from outside of the Wuhan Seafood Market.<sup>8,13</sup> Furthermore, by constructing a haplotype network of the early SARS-CoV-2 genomes, the viral sequences can primarily be divided into two lineage clades, among which, the samples isolated from the Huanan Seafood Market mainly cluster with the descendant lineages rather than the ancestral lineages. This also indicates that the source of the SARS-CoV-2 in the market could be imported from elsewhere.<sup>14</sup>

In addition, as a hub of international communication in central China, Wuhan received extensive international flights from cities around the world before the SARS-CoV-2 pandemic (Fig. 1). Notably, many of these flights to Wuhan departed from Southeast Asian countries that overlap with the *Rhinolophus* and pangolin distributions, as well as locations of multiple known sarbecoviruses. As mentioned in the WHO report, introduction through

cold/food chain products is considered as a possible pathway, which was supported by human infection by contaminated cold chain products in Qingdao.<sup>15</sup> Therefore, before the pandemic, Wuhan was already at a risk of importing SARS-CoV-2 through cold chain cargoes from other parts of the world.

Eighteen months have passed since the identification of SARS-CoV-2, with no progenitor virus identified. The origin-tracing progress has long been hindered by politicization, unfounded slander and widespread laboratory leakage hypothesis. It is high time to start the real global search for sarbecoviruses in the potential locations to identify the origins, intermediate hosts, and transmission paths of SARS-CoV-2. Tracing the origins of a virus is a difficult task. A solid conclusion is the result of an enormous amount of work, patience, global cooperation, some luck, and possibly decades of continuous research, as has been accomplished for the influenza virus. However, such work is indispensable for reducing the frequency of the inevitable pathogen emergences and the damage of outbreaks, as it is crucial to the common health of all mankind.

## REFERENCES

1. Gao, G. F. *Cell* **172**, 1157–1159 (2018).
2. Su, S. et al. *Trends Microbiol.* **24**, 490–502 (2016).
3. Cui, J., Li, F. & Shi, Z. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
4. Gao, R. et al. *N. Engl. J. Med.* **368**, 1888–1897 (2013).
5. Smith, G. J. et al. *Nature* **459**, 1122–1125 (2009).
6. Sharp, P. M. & Hahn, B. H. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **365**, 2487–2494 (2010).
7. Andersen, K. G. et al. *Nat. Med.* **26**, 450–452 (2020).
8. WHO-convened global study of origins of SARS-CoV-2: China Part. <https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part> (2021).
9. Liu, K. et al. *Cell* **184**, 3438–3451.e3410 (2021).
10. Wu, L. et al. *Cell Discov.* **6**, 68 (2020).
11. Zhou, P. & Shi, Z. *Science* **317**, 120–122 (2021).
12. Fenollar, F. et al. *Front. Microbiol.* **12**, 663815 (2021).
13. Liu, Q. et al. *Genom. Proteom. Bioinf.* **18**, 640–647 (2020).
14. Yu, W., Tang, G., Zhang, L. & Corlett, R. T. *Zool. Res.* **41**, 247–257 (2020).
15. Ma, H. et al. *China CDC Weekly* **3**, 637–644 (2021).

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to Qihui Wang.

**Reprints and permission information** is available at <http://www.nature.com/reprints>