

Where did SARS-CoV-2 come from?

Thomas Leitner^{1,*} and Sudhir Kumar^{2,3}

¹Theoretical Biology and Biophysics group, Los Alamos National Laboratory, Los Alamos, NM 87545

²Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA 19122

³Center for Excellence in Genome Medicine and Research, King Abdulaziz University, Jeddah, Saudi Arabia

*Corresponding author. tkl@lanl.gov

Keywords: origins, evolution, pandemic, SARS-CoV-2

ABSTRACT

Identifying the origin of SARS-CoV-2, the etiological agent of the current COVID-19 pandemic, may help us to avoid future epidemics of coronavirus and other zoonoses. Several theories about the zoonotic origin of SARS-CoV-2 have recently been proposed. While *Betacoronavirus* found in *Rhinolophus* bats from China have been broadly implicated, their genetic dissimilarity to SARS-CoV-2 is so high that they are highly unlikely to be its direct ancestors. Thus, an intermediary host is suspected to link bat to human coronaviruses. Based on genomic CpG dinucleotide patterns in different coronaviruses from different hosts, it was suggested that SARS-CoV-2 might have evolved in a canid gastro-intestinal tract prior to transmission to humans. However, similar CpG patterns are now reported in coronaviruses from other hosts, including bats themselves and pangolins. Therefore, reduced genomic CpG alone is not a highly predictive biomarker, suggesting a need for additional biomarkers to reveal intermediate hosts or tissues. The hunt for the zoonotic origin of SARS-CoV-2 continues.

TEXT

About seven months after a new coronavirus started to spread among humans in Wuhan, China, almost 10 million confirmed cases and half a million deaths have occurred worldwide. The new coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is the second zoonosis since a similar Chinese bat coronavirus, SARS-CoV-1, caused an epidemic of severe human respiratory disease 17 years ago (Drosten et al. 2003). Several other coronaviruses infect humans, including the Middle East respiratory syndrome (MERS)-CoV that came from camels and also caused a severe human disease (Dudas et al. 2018), and four coronaviruses that typically cause only mild disease (Corman et al. 2018). Also, many SARS-related coronaviruses have been identified in bats (Li et al. 2005; Yang et al. 2013; Hu et al. 2017), with the potential to infect humans (Ge et al. 2013). Understanding the zoonotic origins of the coronavirus and other viruses is critical because such knowledge can be used to prevent future zoonotic outbreaks.

While it is possible that a bat coronavirus jumped directly to a human, the closest known bat virus, RaTG13 found in a *Rhinolophus affinis* bat (Zhou et al. 2020), shows 96% genomic similarity to SARS-CoV-2. Moreover, RaTG13 differs substantially in the receptor-binding domain (RBD) of the spike protein, suggesting that it may not bind efficiently to the human receptor ACE2 (Andersen et al. 2020). Other bat coronaviruses found in *Rhinolophus pusillus* are less similar, at about 88% genomic similarity (Li et al. 2005). Hence, it seems likely that either there are other, closer, coronaviruses in bats as yet unsampled, or another host species has acted as an intermediary between bats and humans. In either case, because SARS-CoV-2 spreads so acquired the necessary mutations in the RBD to make it transmissible between humans before its zoonotic transfer.

Several candidates for the intermediary host have been proposed. Early circumstantial evidence pointed to snakes sold at the Wuhan market where the SARS-CoV-2 outbreak started, as the codon usage of SARS-CoV-2 was similar to that observed in snakes. However, no coronavirus has been found in snakes. Turtles were subsequently proposed based on predicted spike RBD and ACE2 interactions. Both snakes and turtles were later rejected as candidate intermediate hosts as stronger spike RBD-ACE2 interactions were predicted in ruminants and rodents (Luan et al. 2020). Pangolins were implicated based on the identification of several SARS-CoV-2-related viruses, including one with a similar RBD to SARS-CoV-2 (Lam et al. 2020), and 90-100% amino acid identity to different SARS-CoV-2 proteins (Xiao et al. 2020).

Recently, feral dogs were proposed to be the intermediary host of SARS-CoV-2 (Xia 2020). Following the hypothesis that genomic methylation may be influenced by defenses specific to a host environment, Xia compared CpG deficiencies of many alphacoronavirus and betacoronavirus genomes from multiple host species to that of SARS-CoV-2. The closest match occurred in canine coronaviruses, which are known to infect canine digestive tracts (Licitra et al. 2014). The observation that many SARS-CoV-2 infected humans report digestive symptoms and have high viral loads in stool samples (Wölfel et al. 2020) provided tangential support to the idea that this virus may also infect digestive tissues. Based on CpG patterns, Xia (2020)

suggested that a bat virus had entered the canine digestive tract where it experienced high RNA methylation activity (giving it the observed CpG signature) before transmission to humans.

Pollock et al. (2020) have now tested Xia (2020)'s hypothesis using an extended dataset, and report multiple bat and pangolin betacoronaviruses with CpG patterns similar to those found in dogs. Canine coronaviruses were not the only viruses with CpG patterns similar to those observed in SARS-CoV-2. Therefore, reduced genomic CpG content alone cannot predict the zoonotic origin of SARS-CoV-2, even though Xia (2020) reported reduced genomic CpG signatures in mammalian intermediate hosts of SARS and MERS outbreaks. Additional reliable biomarkers are necessary to conclusively determine intermediate hosts and host tissues. Broader sampling of wildlife species for coronaviruses (Pollock et al. 2020; Xia 2020) will also be vital to discovering intermediate hosts and tissues in which SARS-CoV-2 gained the ability to cause a pandemic.

Identifying the zoonotic source of an emerging pathogen may facilitate efficient containment in the early stages of an outbreak. Furthermore, virus sequence data from the actual source-host provides an essential outgroup for accurate assessment of early spread, as using a virus too distantly-related or divergent may mislead molecular epidemiological reconstructions. This effect was recently shown for SARS-CoV-2 (Mavian et al. 2020). Crucially, identification of the source host may prevent future viral introductions through higher awareness, systematic screening, and development of testing protocols. Identification of the zoonotic origin of SARS-CoV-2 may be particularly challenging, as coronaviruses frequently recombine and are found in many different host species in the wild (Graham and Baric 2010). The hunt for the source is far from over, and the origin of the pandemic will likely only be revealed through more extensive sampling and careful phylogenetic analyses.

REFERENCES

Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. 2020. The proximal origin of SARS-CoV-2. *Nat Med* 26:450-452.

Corman VM, Muth D, Niemeyer D, Drosten C. 2018. Chapter Eight - Hosts and Sources of Endemic Human Coronaviruses. In: Kielian M, Mettenleiter TC, Roossinck MJ, editors. *Adv Virus Res*: Academic Press. p. 163-188.

Drosten C, Günther S, Preiser W, van der Werf S, Brodt H-R, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RAM, et al. 2003. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *New Engl J Med* 348:1967-1976.

Dudas G, Carvalho LM, Rambaut A, Bedford T. 2018. MERS-CoV spillover at the camel-human interface. *eLife* 7:e31257.

Ge X-Y, Li J-L, Yang X-L, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, et al. 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 503:535-538.

Graham RL, Baric RS. 2010. Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission. *J Virol* 84:3134-3146.

Hu B, Zeng L-P, Yang X-L, Ge X-Y, Zhang W, Li B, Xie J-Z, Shen X-R, Zhang Y-Z, Wang N, et al. 2017. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Path* 13:e1006698.

Lam TT-Y, Jia N, Zhang Y-W, Shum MH-H, Jiang J-F, Zhu H-C, Tong Y-G, Shi Y-X, Ni X-B, Liao Y-S, et al. 2020. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* forthcoming.

Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, et al. 2005. Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science* 310:676-679.

Licitra BN, Duhamel GE, Whittaker GR. 2014. Canine Enteric Coronaviruses: Emerging Viral Pathogens with Distinct Recombinant Spike Proteins. *Viruses* 6:3363-3376.

Luan J, Jin X, Lu Y, Zhang L. 2020. SARS-CoV-2 spike protein favors ACE2 from Bovidae and Cricetidae. *J Med Virol*:1-8.

Mavian C, Pond SK, Marini S, Magalis BR, Vandamme A-M, Dellicour S, Scarpino SV, Houldcroft C, Villabona-Arenas J, Paisie TK, et al. 2020. Sampling bias and incorrect rooting make phylogenetic network tracing of SARS-CoV-2 infections unreliable. *Proc Natl Acad Sci USA* 117:12522-12523.

Pollock D, Castoe T, Perry B, Lytras S, Wade K, Robertson DL, Holmes EC, Boni MF, Kosakovsky Pond SL, Parry R, et al. 2020. Viral CpG deficiency provides no evidence that dogs were intermediate hosts for SARS-CoV-2. *Mol Biol Evol* forthcoming.

Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, Niemeyer D, Jones TC, Vollmar P, Rothe C, et al. 2020. Virological assessment of hospitalized patients with COVID-2019. *Nature* 581:465-469.

Xia X. 2020. Extreme Genomic CpG Deficiency in SARS-CoV-2 and Evasion of Host Antiviral Defense. *Mol Biol Evol* msaa094.

Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou J-J, Li N, Guo Y, Li X, Shen X, et al. 2020. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* forthcoming.

Yang L, Wu Z, Ren X, Yang F, He G, Zhang J, Dong J, Sun L, Zhu Y, Du J, et al. 2013. Novel SARS-like betacoronaviruses in bats, China, 2011. *Emerging Infect Dis* 19:989-991.

Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, et al. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270-273.