Bat origin of a new human coronavirus: there and back again

Xiang Li[1,3†], Yuhe Song[4†], Gary Wong[1,2] & Jie Cui[1,2*]

[1]*CAS Key Laboratory of Molecular Virology & Immunology, Institut Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai 200031, China;*
[2]*Center for Biosafety Mega-Science, Chinese Academy of Sciences, Shanghai 200031, China;*
[3]*University of Chinese Academy of Sciences, Beijing 100049, China;*
[4]*School of Life Sciences, Shanghai University, Shanghai 200444, China*

A novel coronavirus reported to cause pneumonia in 41 people during December 2019 in Wuhan, the capital of Hubei province, China, was identified and designated as the 2019 novel coronavirus (2019-nCoV) (Zhu et al., 2020; Li et al., 2020). Infection with 2019-nCoV results in clusters of severe respiratory illness similar to that caused by severe acute respiratory syndrome coronavirus (SARS-CoV) and is associated with high mortality (Chen et al., 2020; Huang et al., 2020). This new coronavirus belongs to genus *Betacoronavirus*, family Coronaviridae, with a single-stranded, positive-sense RNA genome, but genetically distinct from SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV). Up to February 4th, 2020, at least 20,630 cases have been confirmed in China and another 23 countries, with 426 deaths (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/).

After epidemiological investigation, it was revealed that a local seafood market, where wild animals and live poultry were for sale, was linked to this undiagnosed pneumonia (Wu et al., 2020). Wu et al. collected bronchoalveolar lavage fluid from a patient who worked in the seafood market and performed full genome sequencing of 2019-nCoV. It was shown that 2019-nCoV had a high nucleotide sequence similarity to a bat SARS-related CoV (bat-SL-CoVZC45, accession No.

MG772933) and only 79.5% genome sequence similarity to SARS-CoV. Phylogenetic analysis showed that the virus was clustered with SARS-CoV and bat SARSr-CoVs in subgenus *Sarbecovirus* (Wu et al., 2020; Wei et al., 2020; Yu et al., 2019; Xu et al., 2020). Genomic analysis of 2019-nCoV and four typical coronaviruses (bat SARSr-CoV-Rp3, CoV-ZC45, CoV-ZXC21, and SARS-CoV-Tor2) showed a likelihood of recombination between 2019-nCoV and other coronaviruses among the subgenus (Wu et al., 2020).

So how did this novel coronavirus emerge? To investigate the origin of 2019-nCoV, Zhou et al. sequenced samples from seven patients and bats, and found that this new coronavirus shared 96.2% overall genome sequence identity with a bat coronavirus RaTG13 from horseshoe bats (*Rhinolophus*) (Table 1). This discovery provided further evidence that 2019-nCoV most likely originated from bats (Zhou et al., 2020). Previously, horseshoe bats were identified as natural hosts for SARS-related coronaviruses which were the direct progenitors for the origin of SARS-CoV (Hu et al., 2017). Although some amino acid residues were different in the receptor binding domain (RBD) of 2019-nCoV compared to SARS-CoV, it seems that the structure of the RBD still allowed for the use of ACE2 (angiotensin-converting enzyme 2) as an entry receptor. It was suggested that humans were infected with the virus directly from intermediate (none-bats) hosts via contact (Zhou et al., 2020). Certainly, more field work may be needed to find the inter-

Table 1 Sequence identities of 2019-nCoV compared with bat SARSr-CoV and SARS-CoV[a]

	RaTG13	BJ01
ORF1a	96.0/98.0	78.8/80.3
ORF1ab	96.5/96.2	82.3/74.8
S	92.9/97.3	78.4/45.2
ORF3a	96.3/97.8	75.8/72.5
E	99.6/100	93.9/94.8
M	95.5/99.6	85.9/90.4
ORF6	98.4/100	88.6/67.2
ORF7a	95.6/97.5	82.2/85.4
ORF8	97.0/95.1	44.0/50.0
ORF10	99.2/97.4	93.2/82.0
N	96.9/99.0	88.6/90.1
Whole genome	96.0/-	82.29/-

a) This table demonstrates identities of 2019-nCoV WIV04 (GISAID accession No. EPI_ISL_402124) compared with bat SARSr-CoV RaTG13 (GenBank accession No. MN996532.1) and SARS-CoV BJ01 (GenBank accession No. AY278488.2). Identities were calculated by comparing nucleotide sequences and amino acid sequences BJ01 (nt/aa%) respectively. The nucleotide sequence identities were calculated by using blastn and the amino acid identities were computed by blastp (https://blast.ncbi.nlm.nih.gov/Blast.cgi). S, spike glycoprotein; E, envelope small membrane protein; M, membrane protein; N, nucleoprotein.

mediate hosts.

Subsequent familial cluster patients implied that 2019-nCoV spreads viaperson-to-person transmission. Of six family members in Shenzhen who returned from Wuhan, five had been infected with 2019-nCoV (Chan et al., 2020). One family member, who did not travel to Wuhan but had contact with other family members, also became infected. Similar to the beginning of the SARS-CoV pandemic, 2019-nCoV showed mild infectivity. By comparing early submitted sequences, 17 nonsynonymous changes were observed in the open reading frame 1ab, spike gene, open reading frame 7a, and open reading frame 8 (Wei et al., 2020). The same mutations were observed in a familial cluster, suggesting viral mutation could have occurred during person-to-person transmission. Thus, it is urgent to tightly monitor the mutation and adaptation of the virus.

The 2019-nCoV has become a major public health concern as it continues to expand and infect more people. From the available data, given that 2019-nCoV sequences are clustered into one branch with no clear phylogenetic topology (Lu et al., 2020), it could be inferred that there was one single introduction of the virus into the human population, likely from market animals, the so-called intermediate hosts. Usually, adaptive evolution is constantly ongoing as the virus spreads among new hosts. Emerging viruses have certain chances to become endemic or epidemic once viral amino acid changes in key regions have been fixed in the populations. Because only limited sequences were released, there is general lack of understanding of the evolution and epidemiology of the new human-infecting virus, particularly with

respect to the population genetics. Thus, a timely publicity of more complete viral sequences from patients is needed (Wu and Poo, 2020), so that scientists across different disciplines could contribute to combating the threat posed by 2019-nCoV.

References

Chan, J.F.W., Yuan, S., Kok, K.H., To, K.K.W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C.Y., Poon, R.W.S., et al. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet, https://doi.org/10.1016/S0140-6736(20)30154-9.

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet, https://doi.org/10.1016/S0140-6736(20)30211-7.

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 Pathog 13, e1006698.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet, https://doi.org/10.1016/S0140-6736(20)30183-5.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med, https://doi.org/10.1056/NEJMoa2001316.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet, https://doi.org/10.1016/S0140-6736(20)30251-8.

Wei, X., Li, X., and Cui, J. (2020). Evolutionary perspectives on novel coronaviruses identified in pneumonia cases in China. Natl Sci Rev, https://doi.org/10.1093/nsr/nwaa009.

Wu, C.I., and Poo, M. (2020). Very fast evolution, not-so-fast publication—A proposed solution. Natl Sci Rev, https://doi.org/10.1093/nsr/nwaa010.

Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., et al. (2020). A new coronavirus associated with human respiratory disease in China. Nature, https://doi.org/10.1038/s41586-020-2008-3.

Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W., and Hao, P. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Sci China Life Sci, https://doi.org/10.1007/s11427-020-1637-5.

Yu, P., Hu, B., Shi, Z.L., and Cui, J. (2019). Geographical structure of bat SARS-related coronaviruses. Infect Genet Evol 69, 224–229.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med, https://doi.org/10.1056/NEJMoa2001017.

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, https://doi.org/10.1038/s41586-020-2012-7.