Spring Festival travel rush, during which hundreds of millions of people will travel through China.

As a typical RNA virus, the average evolutionary rate for coronaviruses is roughly 10^{-4} nucleotide substitutions per site per year,[1] with mutations arising during every replication cycle. It is, therefore, striking that the sequences of 2019-nCoV from different patients described here were almost identical, with greater than 99·9% sequence identity. This finding suggests that 2019-nCoV originated from one source within a very short period and was detected relatively rapidly. However, as the virus transmits to more individuals, constant surveillance of mutations arising is needed.

Phylogenetic analysis showed that bat-derived coronaviruses fell within all five subgenera of the genus Betacoronavirus. Moreover, bat-derived coronaviruses fell in basal positions in the subgenus Sarbecovirus, with 2019-nCoV most closely related to bat-SL-CoVZC45 and bat-SL-CoVZXC21, which were also sampled from bats.[23] These data are consistent with a bat reservoir for coronaviruses in general and for 2019-nCoV in particular. However, despite the importance of bats, several facts suggest that another animal is acting as an intermediate host between bats and humans. First, the outbreak was first reported in late December, 2019, when most bat species in Wuhan are hibernating. Second, no bats were sold or found at the Huanan seafood market, whereas various non-aquatic animals (including mammals) were available for purchase. Third, the sequence identity between 2019-nCoV and its close relatives bat-SL-CoVZC45 and bat-SL-CoVZXC21 was less than 90%, which is reflected in the relatively long branch between them. Hence, bat-SL-CoVZC45 and bat-SL-CoVZXC21 are not direct ancestors of 2019-nCoV. Fourth, in both SARS-CoV and MERS-CoV, bats acted as the natural reservoir, with another animal (masked palm civet for SARS-CoV[35] and dromedary camels for MERS-CoV)[36] acting as an intermediate host, with humans as terminal hosts. Therefore, on the basis of current data, it seems likely that the 2019-nCoV causing the Wuhan outbreak might also be initially hosted by bats, and might have been transmitted to humans via currently unknown wild animal(s) sold at the Huanan seafood market.

Previous studies have uncovered several receptors that different coronaviruses bind to, such as ACE2 for SARS-CoV[29] and CD26 for MERS-CoV.[30] Our molecular modelling showed structural similarity between the receptor-binding domains of SARS-CoV and 2019-nCoV. Therefore, we suggest that 2019-nCoV might use ACE2 as the receptor, despite the presence of amino acid mutations in the 2019-nCoV receptor-binding domain. Although a previous study using HeLa cells expressing ACE2 proteins showed that 2019-nCoV could employ the ACE2 receptor,[17] whether these mutations affect ACE2 binding or change receptor tropism requires further study.

Recombination has been seen frequently in coronaviruses.[1] As expected, we detected recombination in the Sarbecoviruses analysed here. Our results suggest that recombination events are complex and are more likely occurring in bat coronaviruses than in 2019-nCoV. Hence, despite its occurrence, recombination is probably not the reason for emergence of this virus, although this inference might change if more closely related animal viruses are identified.

In conclusion, we have described the genomic structure of a seventh human coronavirus that can cause severe pneumonia and have shed light on its origin and

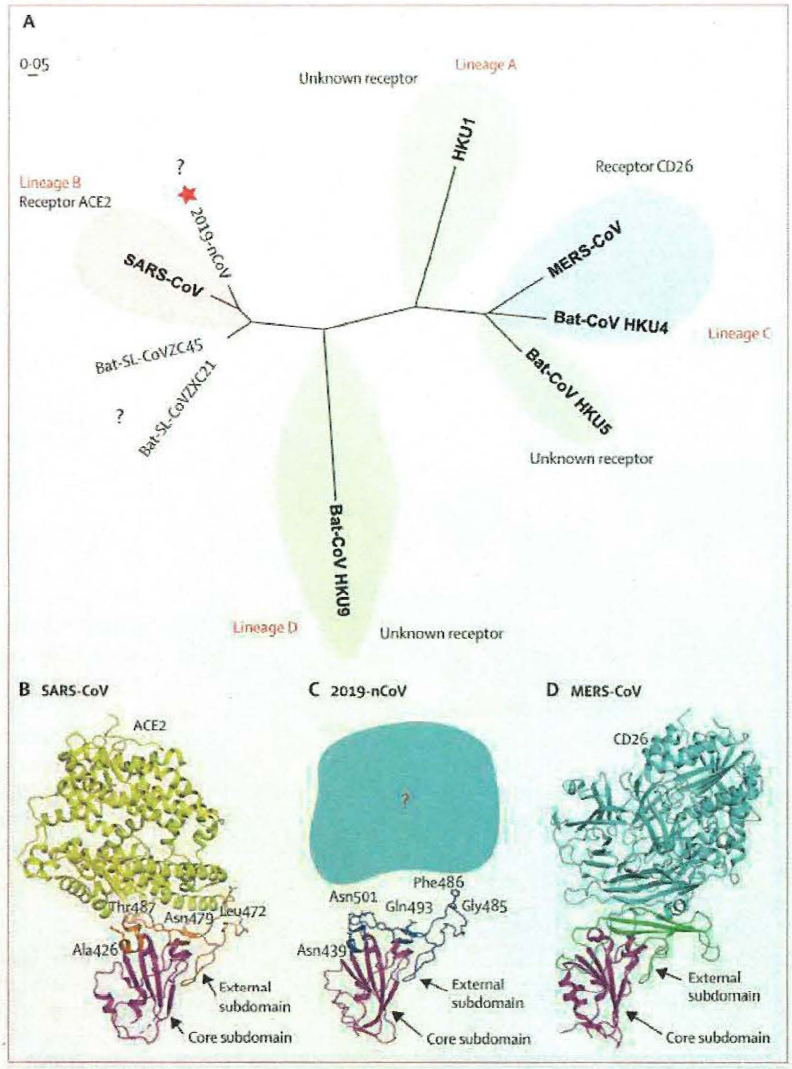

Figure 5: **Phylogenetic analysis and homology modelling of the receptor-binding domain of the 2019-nCoV, SARS-CoV, and MERS-CoV**

(A) Phylogenetic analysis of the receptor-binding domain from various betacoronaviruses. The star highlights 2019-nCoV and the question marks means that the receptor used by the viruses remains unknown. Structural comparison of the receptor-binding domain of SARS-CoV (B), 2019-nCoV (C), and MERS-CoV (D) binding to their own receptors. Core subdomains are magenta, and the external subdomains of SARS-CoV, 2019-nCoV, and MERS CoV are orange, dark blue, and green, respectively. Variable residues between SARS-CoV and 2019-nCoV in the receptor-binding site are highlighted as sticks. CoV=coronavirus. 2019-nCoV=2019 novel coronavirus. SARS-CoV=severe acute respiratory syndrome coronavirus. MERS=Middle East respiratory syndrome coronavirus.

receptor-binding properties. More generally, the disease outbreak linked to 2019-nCoV again highlights the hidden virus reservoir in wild animals and their potential to occasionally spill over into human populations.

Contributors

GFG, WT, WS, WC, WX, and GW designed the study. RL, XZ, PN, HW, WW, BH, NZ, XM, WZ, LZ, JC, YM, JW, YL, JY, ZX, JM, WJL, and DW did the experiments. BY, FZ, and ZH provided samples. WS, WC, WT, JL, HS, YB, LW, TH, and HZ analysed data. WS, WT, and JL wrote the report. ECH and GFG revised the report.

Declaration of interests

We declare no competing interests.

Data sharing

Data are available on various websites and have been made publicly available (more information can be found in the first paragraph of the Results section).

Acknowledgments

This work was supported by the National Key Research and Development Programme of China (2016YFD0500301, 2020YFC0840800, 2020YFC0840900), the National Major Project for Control and Prevention of Infectious Disease in China (2017ZX10104001, 2018ZX10101002, 2018ZX10101004, and 2018ZX10732-401), the State Key Research Development Program of China (2019YFC1200501), the Strategic Priority Research Programme of the Chinese Academy of Sciences (XDB29010102), and the Academic Promotion Programme of Shandong First Medical University (2019QL006 and 2019PT008). WS was supported by the Taishan Scholars Programme of Shandong Province (ts201511056).

References

1 Su S, Wong G, Shi W, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol* 2016; **24:** 490–502.

2 Cavanagh D. Coronavirus avian infectious bronchitis virus. *Vet Res* 2007; **38:** 281–97.

3 Ismail MM, Tang AY, Saif YM. Pathogenicity of turkey coronavirus in turkeys and chickens. *Avian Dis* 2003; **47:** 515–22.

4 Zhou P, Fan H, Lan T, et al. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* 2018; **556:** 255–58.

5 Peiris JS, Guan Y, Yuen KY. Severe acute respiratory syndrome. *Nat Med* 2004; **10** (suppl 12): S88–97.

6 Chan-Yeung M, Xu RH. SARS: epidemiology. *Respirology* 2003; **8** (suppl): S9–14.

7 Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 2012; **367:** 1814–20.

8 Lee J, Chowell G, Jung E. A dynamic compartmental model for the Middle East respiratory syndrome outbreak in the Republic of Korea: a retrospective analysis on control interventions and superspreading events. *J Theor Biol* 2016; **408:** 118–26.

9 Lee JY, Kim YJ, Chung EH, et al. The clinical and virological features of the first imported case causing MERS-CoV outbreak in South Korea, 2015. *BMC Infect Dis* 2017; **17:** 498.

10 Tan W, Zhao X, Ma X, et al. A novel coronavirus genome identified in a cluster of pneumonia cases—Wuhan, China 2019–2020. *China CDC Weekly* 2020; **2:** 61–62.

11 Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020; published online Jan 24. DOI:10.1056/NEJMoa2001017.

12 Chan JFW, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 2020; published online Jan 24. https://doi.org/10.1016/S0140-6736(20)30154-9.

13 Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; published online Jan 24. https://doi.org/10.1016/S0140-6736(20)30183-5.

14 Niu P, Shen J, Zhu N, Lu R, Tan W. Two-tube multiplex real-time reverse transcription PCR to detect six human coronaviruses. *Virol Sin* 2016; **31:** 85–88.

15 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25:** 1754–60.

16 Zhao Y, Tang H, Ye Y. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* 2012; **28:** 125–26.

17 Nurk S, Bankevich A, Antipov D, et al. Assembling genomes and mini-metagenomes from highly chimeric reads. In: Deng M, Jiang R, Sun F, Zhang X, eds. Research in computational molecular biology (RECOMB 2013): lecture notes in computer science, vol 7821. Berlin: Springer, 2013: 158–70.

18 Pan M, Gao R, Lv Q, et al. Human infection with a novel, highly pathogenic avian influenza A (H5N6) virus: virological and clinical findings. *J Infect* 2016; **72:** 52–59.

19 Marchler-Bauer A, Bo Y, Han L, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 2017; **45:** D200–03.

20 Lole KS, Bollinger RC, Paranjape RS, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 1999; **73:** 152–60.

21 Nakamura T, Yamada KD, Tomii K, Katoh K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* 2018; **34:** 2490–92.

22 Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; **30:** 1312–13.

23 Hu D, Zhu C, Ai L, et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg Microbes Infect* 2018; **7:** 154.

24 Li F. Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol* 2016; **3:** 237–61.

25 Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol* 2015; **23:** 468–78.

26 Wang Q, Wong G, Lu G, Yan J, Gao GF. MERS-CoV spike protein: targets for vaccines and therapeutics. *Antiviral Res* 2016; **133:** 165–77.

27 He Y, Zhou Y, Liu S, et al. Receptor-binding domain of SARS-CoV spike protein induces highly potent neutralizing antibodies: implication for developing subunit vaccine. *Biochem Biophys Res Commun* 2004; **324:** 773–81.

28 Li F. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *J Virol* 2012; **86:** 2856–58.

29 Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 2005; **309:** 1864–68.

30 Lu G, Hu Y, Wang Q, et al. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* 2013; **500:** 227–31.

31 Wang N, Shi X, Jiang L, et al. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res* 2013; **23:** 986–93.

32 Wang Q, Qi J, Yuan Y, et al. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* 2014; **16:** 328–37.

33 Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018; **46:** W296–303.

34 Prabakaran P, Gan J, Feng Y, et al. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *J Biol Chem* 2006; **281:** 15829–36.

35 Guan Y, Zheng BJ, He YQ, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 2003; **302:** 276–78.

36 Alagaili AN, Briese T, Mishra N, et al. Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. *mBio* 2014; **5:** e00884-14.

37 Zhou P, Yang X-L, Wang X-G, et al. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *bioRxiv* 2020; published online Jan 23. DOI:10.1101/2020.01.22.914952.