

Possible Bat Origin of Severe Acute Respiratory Syndrome Coronavirus 2

Susanna K.P. Lau,¹ Hayes K.H. Luk,¹ Antonio C.P. Wong,¹ Kenneth S.M. Li, Longchao Zhu, Zirong He, Joshua Fung, Tony T.Y. Chan, Kitty S.C. Fung, Patrick C.Y. Woo

We showed that severe acute respiratory syndrome coronavirus 2 is probably a novel recombinant virus. Its genome is closest to that of severe acute respiratory syndrome–related coronaviruses from horseshoe bats, and its receptor-binding domain is closest to that of pangolin viruses. Its origin and direct ancestral viruses have not been identified.

Seventeen years after the severe acute respiratory syndrome (SARS) epidemic, an outbreak of pneumonia, now called coronavirus disease (COVID-19), was reported in Wuhan, China. Some of the early case-patients had a history of visiting the Huanan Seafood Wholesale Market, where wildlife mammals are sold, suggesting a zoonotic origin. The causative agent was rapidly isolated from patients and identified to be a coronavirus, now designated as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses (1). SARS-CoV-2 has spread rapidly to other places; 113,702 cases and 4,012 deaths had been reported in 110 countries/areas as of March 10, 2020 (2). In Hong Kong, 130 cases and 3 deaths had been reported.

SARS-CoV-2 is a member of subgenus *Sarbecovirus* (previously lineage b) in the family *Coronaviridae*, genus *Betacoronavirus*, and is closely related to SARS-CoV, which caused the SARS epidemic during 2003, and to SARS-related-CoVs (SARSr-CoVs) in horseshoe bats discovered in Hong Kong and mainland China (3–5). Whereas SARS-CoV and Middle East respiratory syndrome coronavirus were rapidly traced to their immediate animal sources (civet and dromedaries, respectively), the origin of SARS-CoV-2 remains obscure.

Author affiliations: The University of Hong Kong, Hong Kong, China (S.K.P. Lau, H.K.H. Luk, A.C.P. Wong, K.S.M. Li, L. Zhu, Z. He, J. Fung, T.T.Y. Chan, P.C.Y. Woo); United Christian Hospital, Hong Kong (K.S.C. Fung)

SARS-CoV-2 showed high genome sequence identities (87.6%–87.8%) to SARSr-Rp-BatCoV-ZXC21/ZC45, detected in *Rhinolophus pusillus* bats from Zhoushan, China, during 2015 (6). A closer-related strain, SARSr-Ra-BatCoV-RaTG13 (96.1% genome identity with SARS-CoV-2), was recently reported in *Rhinolophus affinis* bats captured in Pu'er, China, during 2013 (7). Subsequently, Pangolin-SARSr-CoV/P4L/Guangxi/2017 (85.3% genome identity to SARS-CoV-2) and related viruses were also detected in smuggled pangolins captured in Nanning, China, during 2017 (8) and Guangzhou, China, during 2019 (9). To elucidate the evolutionary origin and pathway of SARS-CoV-2, we performed an in-depth genomic, phylogenetic, and recombination analysis in relation to SARSr-CoVs from humans, civets, bats, and pangolins (10).

The Study

We downloaded 4 SARS-CoV-2, 16 human/civet-SARSr-CoV, 63 bat-SARSr-CoV and 2 pangolin-SARSr-CoV genomes from GenBank and GISAID (<https://www.gisaid.org>). We also sequenced the complete genome of SARS-CoV-2 strain HK20 (GenBank accession no. MT186683) from a patient with COVID-19 in Hong Kong. We performed genome, phylogenetic, and recombination analysis as described (11).

The 5 SARS-CoV-2 genomes had overall 99.8%–100% nt identities with each other. These genomes showed 96.1% genome identities with SARSr-Ra-BatCoV-RaTG13, 87.8% with SARSr-Rp-BatCoV-ZC45, 87.6% with SARSr-Rp-BatCoV-ZXC21, 85.3% with pangolin-SARSr-CoV/P4L/Guangxi/2017, and 73.8%–78.6% with other SARSr-CoVs, including human/civet-SARSr-CoVs (Table 1, <https://wwwnc.cdc.gov/EID/article/26/7/20-0092-T1.htm>).

Most predicted proteins of SARS-CoV-2 showed high amino acid sequence identities with that of SARSr-Ra-BatCoV RaTG13, except the receptor-binding

domain (RBD) region. SARS-CoV-2 possessed an intact open reading frame 8 without the 29-nt deletion found in most human SARS-CoVs. The concatenated conserved replicase domains for coronavirus species demarcation by the International Committee on Taxonomy of Viruses showed $\geq 92.9\%$ aa identities (threshold $>90\%$ for same species) between SARS-CoV-2 and other SARSr-CoVs, supporting their classification under the same coronavirus species (Table 2) (1).

Unlike other members of the subgenus *Sarbecovirus*, SARS-CoV-2 has a spike protein that contains a unique insertion that results in a potential cleavage site at the S1/S2 junction, which might enable proteolytic processing that enhances cell-cell fusion. SARS-CoV-2 was demonstrated to use the same receptor, human angiotensin-converting enzyme 2 (hACE2), as does SARS-CoV (7). The predicted RBD region of SARS-CoV-2 spike protein, corresponding to aa residues 318–513 of SARS-CoV (12), showed the highest (97% aa) identities with pangolin-SARSr-CoV/MP789/Guangdong and 74.1%–77.7% identities with human/civet/bat-SARSr-CoVs known to use hACE2 (Table 1). Moreover, similar to the human/civet/bat-SARSr-CoV hACE2-using viruses, the 2 deletions (5 aa and 12

aa) found in all other SARSr-BatCoVs (10) were absent in SARS-CoV-2 RBD (Appendix Figure 1, <https://wwwnc.cdc.gov/EID/article/26/7/20-0092-App1.pdf>). Of the 5 critical residues needed for RBD-hACE2 interaction in SARSr-CoVs (13), 3 (F472, N487, and Y491) were present in SARS-CoV-2 RBD and pangolin SARSr-CoV/MP789/2019-RBD.

Phylogenetic analysis showed that the RNA-dependent RNA polymerase gene of SARS-CoV-2 is most closely related to that of SARSr-Ra-BatCoV RaTG13, whereas its predicted RBD is closest to that of pangolin-SARSr-CoVs (Figure 1). This finding suggests a distinct evolutionary origin for SARS-CoV-2 RBD, possibly as a result of recombination. Moreover, the SARS-CoV-2 RBD was also closely related to SARSr-Ra-BatCoV RaTG13 and the hACE2-using cluster containing human/civet-SARSr-CoVs and Yunnan SARSr-BatCoVs previously successfully cultured in VeroE6 cells (4,5).

To identify putative recombination events, we performed sliding window analysis using SARS-CoV-2-HK20 as query and SARSr-Ra-BatCoV RaTG13, pangolin-SARSr-CoV/P4L/Guangxi/2017, SARSr-Rp-BatCoV ZC45, SARSr-Rs-BatCoV Rs3367, and SARSr-Rs-BatCoV Longquan-140 as potential parents

Table 2. Percentage amino acid identity between 7 conserved domains of the replicase polyprotein for species demarcation in SARS-CoV-2 and selected members of the subgenus *Sarbecovirus**

Virus	% Amino acid identity compared with that for SARS-CoV-2							Seven concatenated domains
	ADRP	nsp5	nsp12	nsp13	nsp14	nsp15	nsp16	
Human SARS-CoV TOR2/Toronto/Mar2003	79.3	96.1	96.4	99.8	95.1	88.7	93.3	95.0
Civet SARS-CoV SZ3/Shenzhen/2013	79.3	96.1	96.4	99.7	95.1	88.7	93.6	95.0
Civet SARS-CoV PC4–136/Guangdong/2004	78.4	96.1	96.4	99.3	94.7	88.7	93.3	94.8
Human SARS-CoV GZ0402	78.4	95.8	96.2	99.5	94.9	51.7	93.3	94.8
SARSr-Rs-BatCoV HKU3–1/R.sinicus/Hong Kong/2005	79.3	95.4	95.6	98.8	94.7	88.2	93.3	94.5
SARSr-Rp-BatCoV Rp/Shaanxi2011/R.pusillus/Shaanxi/2011	78.4	96.1	96.1	99.3	96.0	88.4	93.6	95.0
SARSr-Rs-BatCoV Rs672/2006/R.sinicus/Guizhou/2006	80.2	95.4	96.5	99.0	95.6	88.2	94.3	95.0
SARSr-Rs-BatCoV WIV1	80.2	95.8	96.2	99.5	95.4	89.0	93.0	95.0
SARSr-Rf-BatCoV YNLF_31C/R.ferrumequinum/Yunnan/2013	80.2	95.8	96.2	99.3	95.6	88.7	93.0	95.0
SARSr-Rf-BatCoV 16BO133/R.ferrumequinum/Korea/2016	79.3	95.1	96.1	98.7	94.9	88.2	92.6	94.6
SARSr-Rf-BatCoV JTMC15/R.ferrumequinum/Jilin/2013	79.3	94.8	96.2	98.5	94.9	88.2	92.6	94.5
SARSr-Rs-BatCoV BtKY72/Rhinolophus sp./Kenya/2007	74.8	95.4	95.0	96.8	92.8	87.3	90.6	92.9
SARSr-Rm-BatCoV Longquan-140/R.monoceros/Zhejiang/2012	79.3	95.8	95.7	99.2	94.9	89.0	94.3	94.8
SARSr-Rb-BatCoV BM48–31/BGR/2008/R.blasii/Bulgaria/2008	78.4	94.1	95.2	97.8	93.5	89.9	88.6	93.5
SARSr-Rp-BatCoV ZXC21/R.pusillus/Zhejiang/2015	88.3	99.0	95.6	98.8	94.7	88.2	98.0	95.4
SARSr-Rp-BatCoV ZC45/R.pusillus/Zhejiang/2017	92.8	99.0	95.9	99.3	94.5	89.0	98.0	95.8
Pangolin-SARSr-CoV Guangxi/P4L/2017	85.5	97.1	97.9	98.2	97.0	94.5	97.7	96.9
SARSr-Ra-BatCoV RaTG13/R.affinis/Yunnan/2013	96.7	99.3	99.6	99.7	99.2	97.7	100	99.2

*ADRP, ADP-ribose-1" phosphatase; nsp, nonstructural protein; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SARSr-CoV, severe acute respiratory syndrome-related coronavirus.

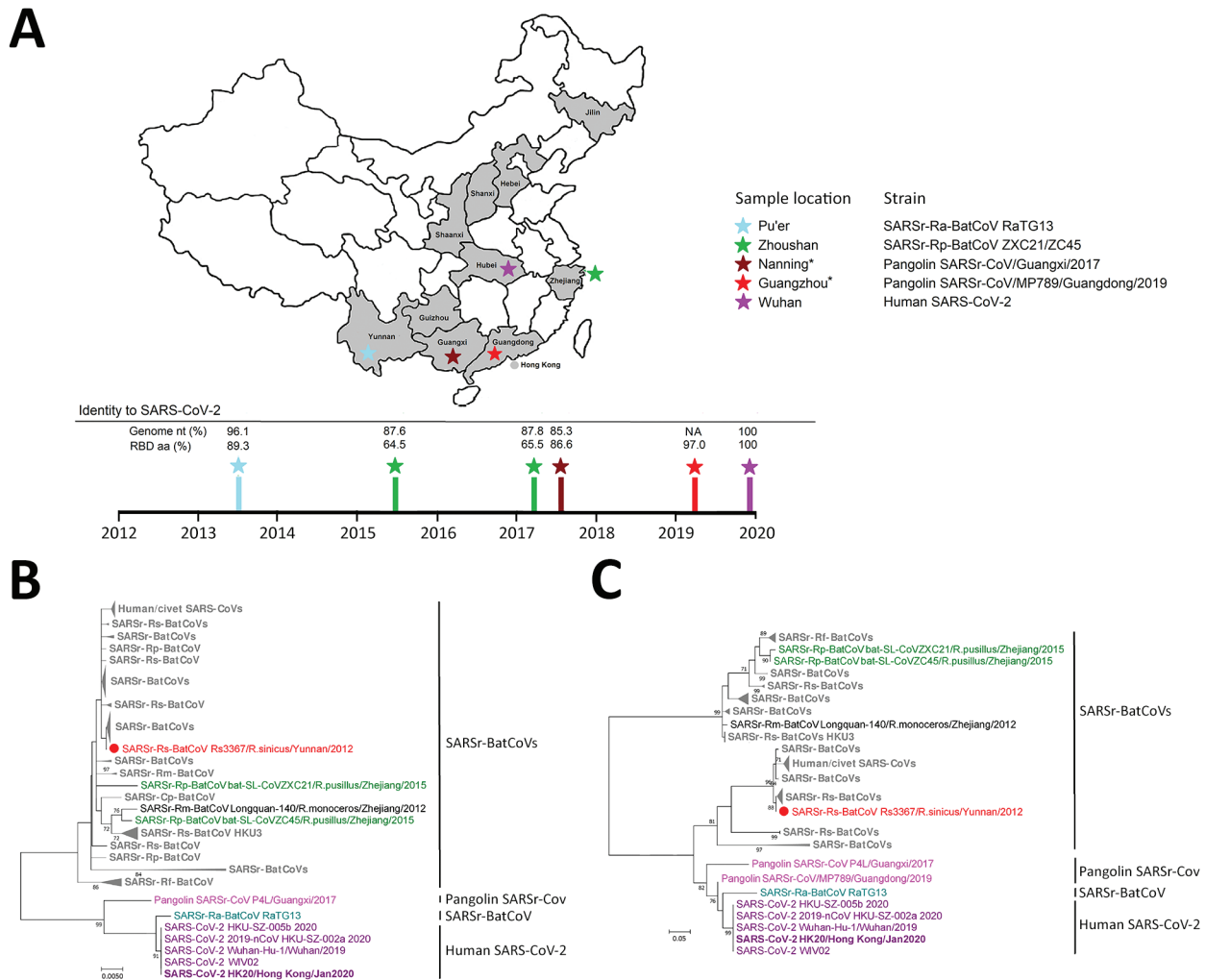


Figure 1. Geographic and phylogenetic comparisons of SARS-CoV-2 isolates with closely related viruses. A) Locations in China where SARS-CoV-2 first emerged (Wuhan), and where closely related viruses were found, including SARSr-Ra-BatCoV RaTG13 (Pu'er), Pangolin-SARSr-CoVs (Guangzhou and Nanning), and SARSr-Rp-BatCoV ZC45 (Zhoushan). Time of sampling and percentage genome identities to SARS-CoV-2 are shown. *Guangzhou and Nanning. The geographic origin of smuggled pangolins remains unknown. B, C) Phylogenetic analyses of RdRp (B) and RBD (C) domains of SARSr-CoVs. Trees were constructed by using maximum-likelihood methods with Jones-Taylor-Thornton plus gamma plus invariant sites (RdRp) and Whelan and Goldman plus gamma (RBD) substitution models. A total of 745 aa residues for RdRp and 177 aa residues for RBD were included in the analyses. Numbers at nodes represent bootstrap values, which were calculated from 1,000 trees. Only bootstrap values >70% are shown. Purple indicates SARS-CoV-2 (strain HK20 in bold); teal indicates SARSr-Ra-BatCoV RaTG13; pink indicates pangolin SARSr-CoVs; green indicates SARSr-Rp-BatCoVs ZXC21 and ZC45; red indicates SARSr-Rs-BatCoV Rs3367; black indicates SARSr-Rs-BatCoV Longquan-140; gray indicates remaining SARSr-BatCoVs. Dots indicate SARSr-BatCoVs reported to use angiotensin-converting enzyme 2 as receptor. Scale bars indicate estimated number of amino acid substitutions per 200 aa residues for RdRp and per 20 aa residues for RBD. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SARSr-CoV, severe acute respiratory syndrome-related coronavirus; NA, not available; RBD, receptor-binding domain; RdRp, RNA-dependent RNA polymerase.

(Figure 2; Appendix Figure 2). A similarity plot showed that SARS-CoV-2 is most closely related to SARSr-Ra-BatCoV RaTG13 in the entire genome, except for its RBD, which is closest to pangolin-SARSr-CoV/MP789/Guangdong, and shows potential recombination breakpoints. Moreover, different regions of SARS-CoV-2 genome showed different similarities to pangolin-SARSr-CoV/

P4L/Guangxi/2017, SARSr-Rp-BatCoV ZC45, SARSr-Rs-BatCoV Rs3367, and SARSr-Rs-BatCoV Longquan-140, as supported by phylogenetic analysis (Appendix Figures 2, 3).

Sequence alignment around the RBD supported potential recombination between SARSr-Ra-BatCoV RaTG13 and pangolin-SARSr-CoV/MP789/

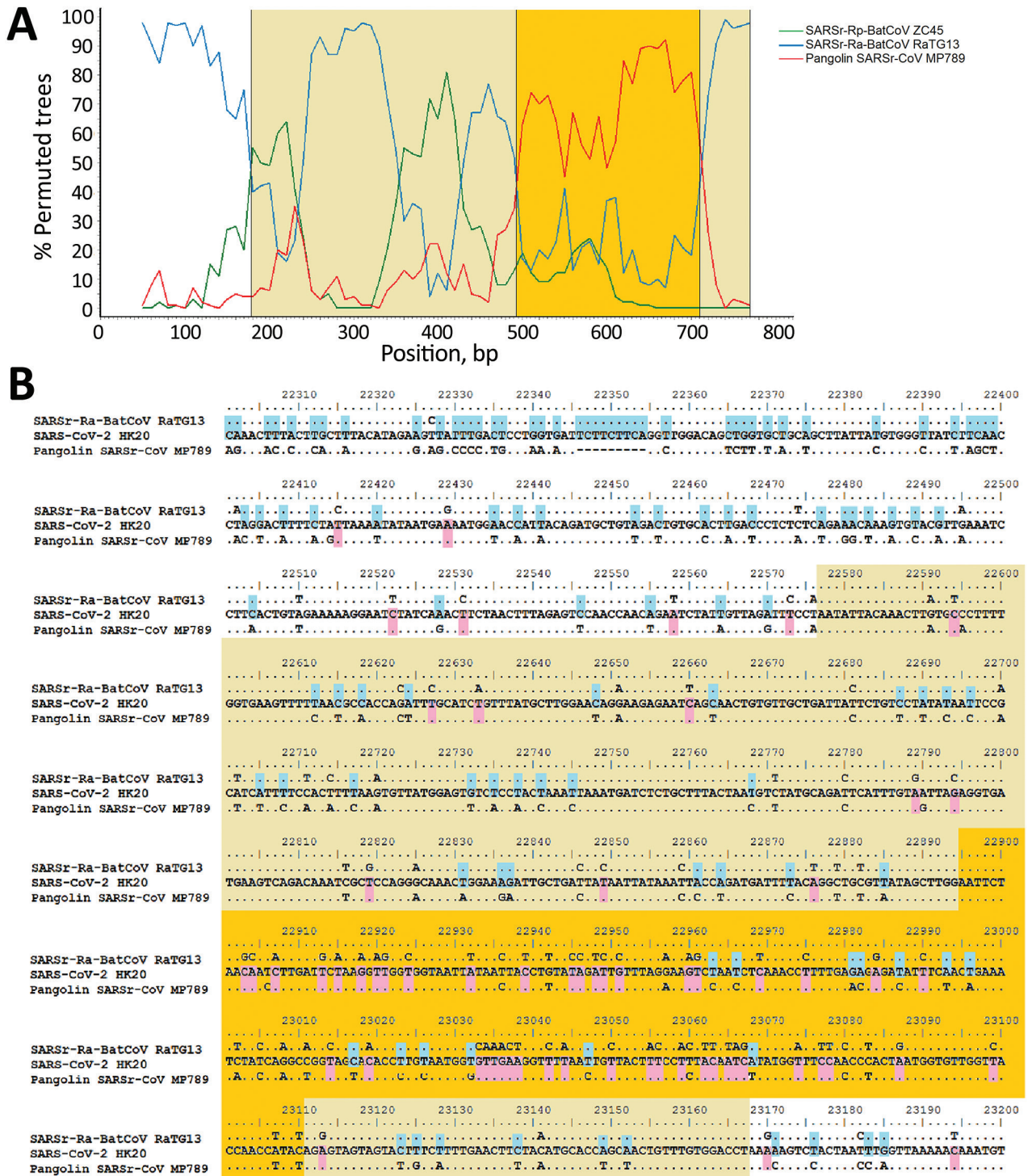


Figure 2. Bootscan analysis and nucleotide sequence alignment for SARS-CoV-2 isolates and closely related viruses. A) Bootscan analysis using the partial spike gene (positions 22397–23167) of SARS-CoV-2 strain HK20 as query sequence. Bootscanning was conducted with SimPlot version 3.5.1 (<https://sray.med.som>) (F84 model; window size, 100 bp; step, 10 bp) on nucleotide alignment, generated with ClustalX (<http://www.clustal.org>). B) Multiple alignment of nucleotide sequences from genome positions 22300 to 23700. Yellow indicates receptor binding domain; orange indicates receptor binding motif; pink indicates bases conserved between SARS-CoV-2 HK20 and Pangolin-SARSr-CoV/MP789/Guangdong/2019; and blue indicates bases conserved between SARS-CoV-2 HK20 and SARSr-Ra-BatCoVs RaTG13.

Guangdong/2019 and the receptor-binding motif region showing exceptionally high sequence similarity to that of pangolin-SARSr-CoV/MP789/Guangdong/2019. This finding suggested that SARS-CoV-2 might be a recombinant virus between viruses closely related to SARSr-Ra-BatCoV RaTG13 and pangolin-SARSr-CoV/MP789/Guangdong/2019.

Conclusions

Despite the close relatedness of SARS-CoV-2 to bat and pangolin viruses, none of the existing SARSr-CoVs represents its immediate ancestor. Most of the genome region of SARS-CoV-2 is closest to SARSr-Ra-BatCoV-RaTG13 from an intermediate horseshoe bat in Yunnan, whereas its RBD is closest to that of pangolin-SARSr-CoV/MP789/Guangdong/2019 from smuggled pangolins in Guangzhou. Potential recombination sites were identified around the RBD region, suggesting that SARS-CoV-2 might be a recombinant virus, with its genome backbone evolved from Yunnan bat virus-like SARSr-CoVs and its RBD region acquired from pangolin virus-like SARSr-CoVs.

Because bats are the major reservoir of SARSr-CoVs and the pangolins harboring SARSr-CoVs were captured from the smuggling center, it is possible that pangolin SARSr-CoVs originated from bat viruses as a result of animal mixing, and there might be an unidentified bat virus containing an RBD nearly identical to that of SARS-CoV-2 and pangolin SARSr-CoV. Similar to SARS-CoV, SARS-CoV-2 is most likely a recombinant virus originated from bats.

The ability of SARS-CoV-2 to emerge and infect humans is likely explained by its hACE2-using RBD region, which is genetically similar to that of culturable Yunnan SARSr-BatCoVs and human/civet-SARSr-CoVs. Most SARSr-BatCoVs have not been successfully cultured *in vitro*, except for some Yunnan strains that had human/civet SARS-like RBDs and were shown to use hACE2 (4,5). For example, SARSr-Rp-BatCoV ZC45, which has an RBD that is more divergent from that of human/civet-SARSr-CoVs, did not propagate in VeroE6 cells (6). Factors that determine hACE2 use among SARSr-CoVs remain to be elucidated.

Although the Wuhan market was initially suspected to be the epicenter of the epidemic, the immediate source remains elusive. The close relatedness among SARS-CoV-2 strains suggested that the Wuhan outbreak probably originated from a point source with subsequent human-to-human transmission, in contrast to the polyphyletic origin of Middle East respiratory syndrome coronavirus (14). If the Wuhan

market was the source, a possibility is that bats carrying the parental SARSr-BatCoVs were mixed in the market, enabling virus recombination. However, no animal samples from the market were reported to be positive. Moreover, the first identified case-patient and other early case-patients had not visited the market (15), suggesting the possibility of an alternative source.

Because the RBD is considered a hot spot for construction of recombinant CoVs for receptor and viral replication studies, the evolutionarily distinct SARS-CoV-2 RBD and the unique insertion of S1/S2 cleavage site among *Sarbecovirus* species have raised the suspicion of an artificial recombinant virus. However, there is currently no evidence showing that SARS-CoV-2 is an artificial recombinant, which theoretically might not carry signature sequences. Further surveillance studies in bats are needed to identify the possible source and evolutionary path of SARS-CoV-2.

This study was partly supported by the theme-based research scheme (project no. T11-707/15-R) of the University Grant Committee; Health and Medical Research Fund of the Food and Health Bureau of HKSAR; Consultancy Service for Enhancing Laboratory Surveillance of Emerging Infectious Disease for the HKSAR Department of Health and the University Development Fund of the University of Hong Kong.

About the Author

Dr. Lau is a professor and head of the Department of Microbiology at The University of Hong Kong, Hong Kong, China. Her primary research interest is using microbial genomics for studying emerging infectious diseases, including coronaviruses.

References

1. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol.* 2020;5:536-44. <https://doi.org/10.1038/s41564-020-0695-z>
2. World Health Organization. Coronavirus disease 2019 (COVID-19) situation report 50, March 10, 2020 [cited 2020 Apr 11]. <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200310-sitrep-50-covid-19.pdf>
3. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A.* 2005;102:14040-5. <https://doi.org/10.1073/pnas.0506735102>
4. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature.* 2013;503:535-8. <https://doi.org/10.1038/nature12711>
5. Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, et al. Discovery of a rich gene pool of bat SARS-related

- coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog.* 2017;13:e1006698. <https://doi.org/10.1371/journal.ppat.1006698>
6. Hu D, Zhu C, Ai L, He T, Wang Y, Ye F, et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg Microbes Infect.* 2018;7:154. <https://doi.org/10.1038/s41426-018-0155-5>
 7. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020;579:270–3. <https://doi.org/10.1038/s41586-020-2012-7>
 8. Liu P, Chen W, Chen JP. Viral metagenomics revealed sendai virus and coronavirus infection of Malayan pangolins (*Manis javanica*). *Viruses.* 2019;11:E979. <https://doi.org/10.3390/v11110979>
 9. Lam TT, Shum MH, Zhu HC, Tong YG, Ni XB, Liao YS, et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature.* 2020;March 26: Epub ahead of print. <https://doi.org/10.1038/s41586-020-2169-0>
 10. Luk HK, Li X, Fung J, Lau SK, Woo PC. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. *Infect Genet Evol.* 2019;71:21–30. <https://doi.org/10.1016/j.meegid.2019.03.001>
 11. Lau SKP, Li KS, Huang Y, Shek CT, Tse H, Wang M, et al. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related *Rhinolophus* bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. *J Virol.* 2010;84:2808–19. <https://doi.org/10.1128/JVI.02219-09>
 12. Wong SK, Li W, Moore MJ, Choe H, Farzan M. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem.* 2004;279:3197–201. <https://doi.org/10.1074/jbc.C300520200>
 13. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* 2005;24:1634–43. <https://doi.org/10.1038/sj.emboj.7600640>
 14. Lau SK, Wernery R, Wong EY, Joseph S, Tsang AK, Patteril NA, et al. Polyphyletic origin of MERS coronaviruses and isolation of a novel clade A strain from dromedary camels in the United Arab Emirates. *Emerg Microbes Infect.* 2016;5:e128. <https://doi.org/10.1038/emi.2016.129>
 15. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020;395:497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)

Address for correspondence: Susanna K.P. Lau or Patrick C.Y. Woo, Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Rm 26, 19/F, Block T, Queen Mary Hospital, 102 Pokfulam Rd, Hong Kong, China; email: skplau@hku.hk or pcywoo@hku.hk

EID Podcast: Nipah Virus Transmission from Bats to Humans Associated with Drinking Traditional Liquor Made from Date Palm Sap, Bangladesh, 2011–2014

Nipah virus (NiV) is a paramyxovirus, and *Pteropus* spp. bats are the natural reservoir. From December 2010 through March 2014, hospital-based encephalitis surveillance in Bangladesh identified 18 clusters of NiV infection. A team of epidemiologists and anthropologists investigated and found that among the 14 case-patients, 8 drank fermented date palm sap (*tari*) regularly before their illness, and 6 provided care to a person infected with NiV. The process of preparing date palm trees for *tari* production was similar to the process of collecting date palm sap for fresh consumption. Bat excreta was reportedly found inside pots used to make *tari*. These findings suggest that drinking *tari* is a potential pathway of NiV transmission.



Visit our website to listen: **EMERGING INFECTIOUS DISEASES**
<http://www2c.cdc.gov/podcasts/player.asp?f=8642667>