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.

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 SARSr-CoVs. The concatenated conserved replicate domains for coronavirus species demarcation by the International Committee on Taxonomy of Viruses showed ≥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-SARS-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-CoVs 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.
DISPATCHES

A similarity plot showed that SARS-CoV-2 is most closely related to SARSr-Ra-BatCoV RaTG13 (Pu’er), Pangolin-SARSr-CoVs (Guangzhou and Nanning), and SARSr-Rp-BatCoV ZC45 (Zhoushan). Time of sampling and percentage genotype 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 SARS-CoVs; green indicates SARSr-Rp-BatCoV 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; SARS-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/
Possible Bat Origin of SARS-CoV-2

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/s41556-020-0236-1
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.0507235102
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
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: [EID Podcast: Nipah Virus Transmission from Bats to Humans Associated with Drinking Traditional Liquor Made from Date Palm Sap, Bangladesh, 2011–2014](http://www2c.cdc.gov/podcasts/player.asp?f=8642667)
Possible Bat Origin of Severe Acute Respiratory Syndrome Coronavirus 2

Appendix

### Appendix Table
Severe acute respiratory syndrome coronavirus strains used in this study

| Virus | GenBank accession no. |
|-------|------------------------|
| SARS-Rp-BatCoV ZC45/R.pusillus/Zhejiang/2015 | MG772934.1 |
| SARS-Rp-BatCoV ZXC21/R.pusillus/Zhejiang/2015 | MG772933.1 |
| SARS-Rm-BatCoV Longquan-140/R.monoceros/Zhejiang/2012 | KF294457.1 |
| SARS-Rs-BatCoV HKU3-1/R.sinicus/Hong Kong/2005 | DQ022305.2 |
| SARS-Rs-BatCoV HKU3-12/R.sinicus/Hong Kong/2007 | GQ153547.1 |
| SARS-Rs-BatCoV HKU3-11/R.sinicus/Hong Kong/2007 | GQ153546.1 |
| SARS-Rs-BatCoV HKU3-9/R.sinicus/Hong Kong/2006 | GQ153544.1 |
| SARS-Rs-BatCoV HKU3-6/R.sinicus/Hong Kong/2005 | GQ153541.1 |
| SARS-Rs-BatCoV HKU3-5/R.sinicus/Hong Kong/2005 | GQ153540.1 |
| SARS-Rs-BatCoV HKU3-4/R.sinicus/Hong Kong/2005 | GQ153539.1 |
| SARS-Rs-BatCoV HKU3-3/R.sinicus/Hong Kong/2005 | DQ084200.1 |
| SARS-Rs-BatCoV HKU3-2/R.sinicus/Hong Kong/2005 | DQ084199.1 |
| SARS-Rs-BatCoV HKU3-13/R.sinicus/Hong Kong/2007 | GQ153548.1 |
| SARS-Rs-BatCoV HKU3-10/R.sinicus/Hong Kong/2006 | GQ153545.1 |
| SARS-Rs-BatCoV HKU3-8/R.sinicus/Guangdong/2006 | GQ153543.1 |
| SARS-Rs-BatCoV HKU3-7/R.sinicus/Guangdong/2006 | GQ153542.1 |
| SARS-Rs-BatCoV Rp3/R.sinicus/Guangxi/2004 | DQ071615.1 |
| SARS-Rs-BatCoV Rs4247/R.sinicus/Yunnan/2013 | KY171548.1 |
| SARS-Rs-BatCoV Rs4237/R.sinicus/Yunnan/2013 | KY171474.1 |
| SARS-Rs-BatCoV Rs4081/R.sinicus/Yunnan/2012 | KY171431.1 |
| SARS-Rs-BatCoV Rs4255/R.sinicus/Yunnan/2013 | KY171419.1 |
| SARS-As-BatCoV As65266/Aselliscus stoliczkanus/Yunnan/2014 | KU182964.1 |
| SARS-Rs-BatCoV JTMC15/R.ferrumequinum/Jilin/2013 | MG772934.1 |
| SARS-Rm-BatCoV 279/2005/R.macrotis/Hubei/2005 | DQ648857.1 |
| SARS-Ra-BatCoV YN2018D/R.spphinis/Yunnan/2016 | MK211378.1 |
| SARS-Ra-BatCoV YN2018C/R.spphinis/Yunnan/2016 | MK211377.1 |
| SARS-Rm-BatCoV Rm1/R.macrotis/Hubei/2004 | DG412043.1 |
| SARS-Rs-BatCoV Rs3367/R.sinicus/Yunnan/2012 | KC89106.2 |
| SARS-Rs-BatCoV RsSHC014/R.sinicus/Yunnan/2011 | KC89105.1 |
| SARS-Rs-BatCoV WIV1 | KF364697.1 |
| SARS-Rs-BatCoV Rs9401/R.sinicus/Yunnan/2015 | KY171521.1 |
| SARS-Rs-BatCoV Rs4874/R.sinicus/Yunnan/2013 | KY171501.1 |
| SARS-Ra-BatCoV YN2018A/R.spphinis/Yunnan/2016 | MK211375.1 |
| SARS-Rs-BatCoV Anlong-103/Guizhou/R.sinicus/2013 | KY177085.1 |
| SARS-Rs-BatCoV Anlong-112/Guizhou/R.sinicus/2013 | KY177085.1 |
| Human SARS-CoV TQR2/Toronto/Mar/2003 | NC_004718.1 |
| Human SARS-CoV GZ02/Guangdong/Feb/2003 | KY179056.1 |
| SARS-Rs-BatCoV Rs7327/R.sinicus/Yunnan/2014 | KY179151.1 |
| SARS-Rs-BatCoV Rs4231/R.sinicus/Yunnan/2013 | KY179146.1 |
| SARS-Rs-BatCoV Rf4092/R.ferrumequinum/Yunnan/2012 | KY179145.1 |
| UNVERIFIED: SARS-Rp-BatCoV F46/R.pusillus/Yunnan/2012 | KU973692.1 |
| SARS-Rs-BatCoV WIV16/R.sinicus/Yunnan/2013 | KT443826.1 |
| SARS-Rs-BatCoV YN2018B/R.spphinis/Yunnan/2016 | KU182964.1 |
| Human SARS-CoV BJ01/Beijing | KY179056.1 |
| Civet SARS-CoV SZ16/Shenzhen/2013 | KY179056.1 |
| Civet SARS-CoV SZ3/Shenzhen/2013 | KY179056.1 |
| Civet SARS-CoV GD69/Guangdong/May/2003 | KY179056.1 |
| SARS-Rs-BatCoV YNLF_34C/R.ferrumequinum/Yunnan/2013 | KP886809.1 |
| SARS-Rs-BatCoV YNLF_31C/R.ferrumequinum/Yunnan/2013 | KP886808.1 |
| SARS-Rs-BatCoV Rs4084/R.sinicus/Yunnan/2012 | KY179056.1 |
| Civet SARS-CoV PC4-227/Guangdong/2004 | KY179056.1 |
| Civet SARS-CoV PC4-136/Guangdong/2004 | KY179056.1 |
| Civet SARS-CoV PC4-13/Guangdong/2004 | KY179056.1 |
| Human SARS-CoV GZ0402 | KY179056.1 |
| Virus                      | GenBank accession no. |
|---------------------------|-----------------------|
| SARSr-Rf-BatCoV Rf1/R.ferumequinum/Hubei/2004 | DQ412042.1 |
| SARSr-Rf-BatCoV 273/2005/R.ferumequinum/Hubei/2005 | DQ648856.1 |
| SARSr-Ra-BatCoV LYRa11/R.affinis/Yunnan/2011 | KF569996.1 |
| SARSr-Rf-BatCoV Ri-SC2018/Rhinolophus sp./2016 | MK211374.1 |
| Civet SARS-CoV A001       | FJ959407.1 |
| Civet SARS-CoV civet020/Guangdong/2004 | AY572038.1 |
| Civet-Rp-BatCoV Rp/Shaanxi/2011/R.pusillus/Shaanxi/2011 | JX993987.1 |
| SARS-Rf-BatCoV Jiyuan-84/R.ferumequinum/ Henan/2012 | KY770860.1 |
| Human SARS-CoV GZ2041/Guangdong/Dec2003 | AY568539.1 |
| Civet SARS-CoV civet007/Guangdong/2004 | AY572034.1 |
| Civet SARS-CoV civet010/Guangdong/2004 | AY572035.1 |
| Civet SARS-CoV B039/Guangdong/2004 | AY668664.1 |
| SARS-Rs-BatCoV Cp/Yunnan2011/Chaerephon plicata/Yunnan/2011 | JX993988.1 |
| Civet SARS-CoV A022/Guangdong/2004 | AY668663.1 |
| SARS-Rf-BatCoV SX2013/R.ferumequinum/Shanxi/2013 | KJ473813.1 |
| SARS-Rf-BatCoV HeB2013/R.ferumequinum/Hebei/2013 | KJ473812.1 |
| SARS-Rs-BatCoV Re672/2006/R.sinicus/Guizhou/2006 | FJ588686.1 |
| SARS-Rs-BatCoV YN2013/R.sinicus/Yunnan/2013 | KJ473816.1 |
| SARS-Rs-BatCoV GX2013/R.sinicus/Guangxi/2013 | KJ473815.1 |
| SARS-Rf-BatCoV JL2012/R.ferumequinum/Jilin/2012 | KJ473811.1 |
| SARS-Rf-BatCoV 16BO133/R.ferumequinum/Korea/2016 | KY938558.1 |
| SARS-Rs-BatCoV BtKY72/Rhinolophus sp./Kenya/2007 | KY352407.1 |
| SARS-Rb-BatCoV BM48-31/BGR/2008/R.blasii/Bulgaria/2008 | GU190215.1 |
| SARS-Rs-BatCoV RaTG13/Yunnan/2013 | MN996532.1 |
| SARS-CoV-2 Wuhan-Hu-1/Hubei/2019 | MN908947.1 |
| SARS-CoV-2 WIV02/Hubei/2019 | MN996527.1 |
| SARS-CoV-2 HKU-SZ_002a/Guangdong/2020 | MN938384.1 |
| SARS-CoV-2 HKU-SZ_005b/Guangdong/2020 | MN975262.1 |
| Pangolin SARS-CoV Guangxi/P4L2017 | MT040333.1 |
| Pangolin SARS-CoV MP789 | MT084071.1 |
| SARS-CoV-2 HK20/Hong Kong/2020 | MT186683 |

*SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SARSr-CoV, severe acute respiratory syndrome–related coronavirus.*
Appendix Figure 1. Multiple alignment of amino acid sequences of the receptor-binding domain of the spike proteins of SARS-CoV-2; human, pangolin, and civet SARSr-CoVs; and corresponding sequences of SARSr-BatCoVs in different *Rhinolophus* species. Asterisks indicate positions that have fully conserved residues. Dashes indicate deletions. Amino acid deletions in some SARSr-BatCoVs are highlighted in orange; the 5 critical residues for receptor binding in human SARS-CoV at positions 442, 472, 479, 487, and 491 are highlighted in blue, and the polybasic cleavage site is highlighted in purple. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SARSr-CoV, severe acute respiratory syndrome–related coronavirus.
Appendix Figure 2. Simplot analysis using the genome sequence of SARS-CoV-2 strain HK20 as the query sequence. Simplot analysis was conducted with Simplot version 3.5.1 (F84 model; window size, 400 bp; step, 40 bp) on nucleotide alignment, generated with ClustalX. The teal line denotes SARSr-Ra-BatCoV RaTG13, the red line denotes pangolin-SARSr-CoV-GX_P4L, the green line denotes SARSr-Rp-BatCoV strain ZC45, the blue line denotes SARSr-Rs-BatCoV strain Rs3367, and the black line denotes SARSr-Rm-BatCoV strain Longquan-140. Phylogenetic trees were constructed by maximum-likelihood method using the generalized time reversible + gamma + invariant substitution model based on nucleotides sequences for the regions from the 5’end to position 11502, position 11502 to 21509, position 21509 to 25928 and position 25928 to 3’ end. Bootstrap values were calculated from 1,000 trees. Only bootstrap values of >70% are shown. E, envelope; Hel 1, helicase; M, matrix; N, nucleocapsid; RBD, receptor-binding protein; PL; papain-like protease; RdRp, RNA-dependent RNA polymerase; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SARSr, severe acute respiratory syndrome–related coronavirus.
Appendix Figure 3. Phylogenetic analysis of nonstructural protein 3 (nsp3), nsp5, nsp13, spike protein, open reading frame 8 (ORF8), and nucleocapsid protein of severe acute respiratory syndrome–related coronavirus causing the SARS epidemic during 2003. Viruses were isolated from bats, pangolins, civets, and humans. Trees were constructed by using the maximum-likelihood method and Jones-Taylor-Thornton + gamma (nsp3, nsp5, nsp13, and nucleocapsid); leaving group + gamma + invariant (spike); and Jones-Taylor-Thornton + invariant (ORF8) substitution models. Bootstrap values were calculated from 1,000 trees. Amino acid positions 1666, 306, 600, 1225, 108 and 414 in nsp3, nsp5, nsp13, spike, ORF8, and nucleocapsid, respectively, were included in the analysis. Scale bar indicates estimated number of amino acid substitutions per 20, 200, 200, 50, 5, and 50 positions in nsp3, nsp5, nsp13, spike, ORF8, and nucleocapsid, respectively.