# Are pangolins the intermediate host of the 2019 novel coronavirus (2019-nCoV)?

Ping Liu<sup>1\*</sup>, Jing-Zhe Jiang<sup>2\*</sup>, Xiu-Feng Wan<sup>3,4,5,6,7</sup>, Yan Hua<sup>8</sup>, Xiaohu Wang<sup>9</sup>, Fanghui Hou<sup>10</sup>, Jing Chen<sup>9</sup>, Jiejian Zou<sup>10</sup>, Jinping Chen<sup>1</sup>†

<sup>1</sup>Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Guangdong Institute of Applied Biological Resources, Guangzhou, Guangdong Province 510260, China. <sup>2</sup>Key Laboratory of South China Sea Fishery Resources Exploitation & Utilization, Ministry of Agriculture, South China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou, Guangdong Province 510300, China. <sup>3</sup>Department of Molecular Microbiology and Immunology, School of Medicine, University of Missouri, Columbia, MO, 65211 USA. <sup>4</sup>Department of Electrical Engineering & Computer Science, College of Engineering, University of Missouri, Columbia, MO, 65211 USA. 5 Missouri University Center for Research on Influenza Systems Biology (CRISB), University of Missouri, Columbia, MO, 65211 USA. <sup>6</sup>Bond Life Sciences Center, University of Missouri, Columbia, MO, 65211 USA. <sup>7</sup>MU Informatics Institute, University of Missouri, Columbia, MO, 65211 USA. 8Guangdong Provincial Key Laboratory of Silviculture, Protection and Utilization, Guangdong Academy of Forestry, Guangzhou, Guangdong Province 510520, China. <sup>9</sup>Institute of Animal Health, Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong Province 510640, China. <sup>10</sup>Guangdong Provincial

Wildlife Rescue Center, Guangzhou, Guangdong Province 510520, China.

\*These authors contributed equally to this work.

†Corresponding author. Email: <a href="mailto:chenjp@giabr.gd.cn">chenjp@giabr.gd.cn</a>

## Abstract

The outbreak of 2019-nCoV pneumonia (COVID-19) in the city of Wuhan, China has resulted in more than 70,000 laboratory confirmed cases, and recent studies showed that 2019-nCoV (SARS-CoV-2) could be of bat origin but involve other potential intermediate hosts. In this study, we assembled the genomes of coronaviruses identified in sick pangolins. The molecular and phylogenetic analyses showed that pangolin Coronaviruses (pangolin-CoV) are genetically related to both the 2019-nCoV and bat Coronaviruses but do not support the 2019-nCoV arose directly from the pangolin-CoV. Our study also suggested that pangolin be natural host of *Betacoronavirus*, with a potential to infect humans. Large surveillance of coronaviruses in pangolins could improve our understanding of the spectrum of coronaviruses in pangolins. Conservation of wildlife and limits of the exposures of humans to wildlife will be important to minimize the spillover risks of coronaviruses from wild animals to humans.

#### Introduction

In December 2019, there was an outbreak of pneumonia with an unknown cause in Wuhan, Hubei province in China, with an epidemiological link to the Huanan Seafood Wholesale Market, which is a live animal and seafood market. Clinical presentations of this disease greatly resembled viral pneumonia. Through deep sequencing on the lower respiratory tract samples of patients, a novel coronavirus named the 2019 novel coronavirus (2019-nCoV) was identified [1]. Within less than 2 months, the viruses have spread to all provinces across China and 23 additional countries. As of February 19, 2020, the epidemic has resulted in 72,532 laboratory confirmed cases, 1,872 of which were fatal. With nearly three weeks of locking down Wuhan (and followed by many other cities across China), the toll of new cases and deaths are still rising.

To effectively control the diseases and prevent new spillovers, it is critical to identify the animal origin of this newly emerging coronavirus. In this wet market of Wuhan, high viral loads were reported in the environmental samples. However, variety of animals, including some wildlife, were sold on this market, and the number and species were very dynamics. It remains unclear which animal initiated the first infections.

Coronaviruses cause respiratory and gastrointestinal tract infections and are genetically classified into four major genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus*. The former two genera primarily infect mammals, whereas the latter two predominantly infect birds [2]. In addition to the 2019-nCoV, *Betacoronavirus* caused the 2003 SARS (severe acute respiratory

syndrome) outbreaks and the 2012 MERS (Middle East respiratory syndrome) outbreaks in humans [3, 4]. Both SARS-CoV and MERS-CoV are of bat origin, but palm civets were shown to be an intermediate host for SARS-CoV [5] and dromedary camels for MERS-CoV [6].

Approximate 30-thousand-base genome of coronavirus codes up to 11 proteins, and the surface glycoprotein S protein binds to receptors on the host cell, initiating virus infection. Different coronaviruses can use distinct host receptors due to structural variations in the receptor binding domains of the virus S protein.

SARS-CoV uses angiotensin-converting enzyme 2 (ACE2) as one of the main receptors [7] with CD209L as an alternative receptor [8], whereas MERS-CoV uses dipeptidyl peptidase 4 (DPP4, also known as CD26) as the primary receptor.

Computational modeling analyses suggested that, similar to SARS-CoV, the 2019-nCoV uses ACE2 as the receptor [9].

Not soon after the release of the 2019-nCoV genome, a scientist released a full genome of a coronavirus, Bat-CoV-RaTG13, from bat (*Rhinolophus sinicus*), which is colonized in Yunan province, nearly 2,000 km away from Wuhan. Bat-CoV-RaTG13 was 96% identical at the whole genome level to the 2019-nCoV, suggesting the 2019-nCoV, could be of bat origin [10]. However, with rare direct contacts between such bats and humans, similar to SARS-CoV and MERS-CoV, it seems to be more likely that the spillover of 2019-nCoV to humans from another intermediate host rather than directly from bats.

The goal of this study is to determine the genetic relationship between a

coronavirus from two groups of sick pangolins and the 2019-nCoV and to assess whether pangolins could be a potential intermediate host for the 2019-nCoV.

## Results

In March of 2019, we detected *Betacoronavirus* in three animals from two sets of smuggling *Malayan* pangolins (*Manis javanica*) (n=26) intercepts by Guangdong customs [11]. All three animals suffered from serious respiratory diseases and failed to be rescued by the Guangdong Wildlife Rescue Center [11] (Table S2). Through metagenomic sequencing and *de novo* assembling, we recovered 38 contigs ranging from 380 to 3,377 nucleotides, and the nucleotide sequence identities among the contigs from these three samples were 99.54%. Thus, we pooled sequences from three samples and assembled the draft genome of this pangolin origin coronavirus, so called pangolin-CoV-2020 (Accession No.: GWHABKW00000000), which was approximately 29,380 nucleotides, with approximately 84% coverage of the virus genome (Figure 1a).

Strikingly, genomic analyses suggested the pangolin-CoV-2020 has a high identity with both 2019-nCoV and Bat-CoV-RaTG13, the proposed origin of the 2019-nCoV [10] (Figure 1b; Figure 1c). The nucleotide sequence identity between pangolin-CoV-2020 and 2019-nCoV was 90.23% whereas the protein sequence identities for individual proteins can be up to 100% (Table S3; Table S4). The nucleotide sequence between pangolin-CoV-2020 and Bat-CoV-RaTG13 was 90.15% whereas that for the corresponding regions between 2019-nCoV and

Bat-CoV-RaTG13 was 96.12% (Table 1).

The nucleotide sequence identities of the surface glycoprotein Spike (S) protein genes between pangolin-CoV-2020 and 2019-nCoV was 82.21%, and the Bat-CoV-RaTG13 and 2019-nCoV shared the highest sequence identity of 92.59% (Table 1). There was a low similarity of 72.63% between the S genes of pangolin-CoV-2020 and SARS-CoV. Nucleotide sequence analyses suggested the S gene was relatively more genetic diverse in the S1 region than the S2 region (Figure 2a). Furthermore, the S proteins of pangolin-CoV-2020 and 2019-nCoV had a sequence identity of 89.78% (Table 2), sharing a very conserved receptor binding motif (RBM) (Figure S1), which is more conserved than in Bat-CoV-RaTG13. These results support that pangolin-CoV-2020 and 2019-nCoV, and SARS-CoV could all share the same receptor ACE2. The presence of highly identical RBMs in pangolin-CoV-2020 and 2019-nCoV means that this motif was likely already present in the virus before jumping to humans. However, it is interesting that both pangolin-CoV-2020 and Bat-CoV-RaTG13 lack a S1/S2 cleavage site (~680-690 aa) whereas 2019-nCoV possess (Figure S1).

Phylogenetic analyses suggested that the S genes of pangolin-CoV-2020, 2019-nCoV and three bat origin coronaviruses (Bat-CoV-RaTG13, Bat-CoV-CVZXC21, and Bat-Cov-CVZC45) were genetically more similar to each other than other viruses in the same family (Figure 2b). The S gene of Bat-CoV-RaTG13 was genetically closer to each other than pangolin-CoV-2020, Bat-CoV-CVZXC21, and Bat-Cov-CVZC45. Similar tree topologies were observed

for the RdRp gene and other genes (Figure 3a-d; Figure S2a-h). At the whole genomic level, the 2019-nCoV is also genetically closer to Bat-CoV-RaTG13 than pangolin-CoV-2020 (Figure 1c).

## **Discussion**

In this study, we assembled the genomes of coronaviruses identified in sick pangolins and our results showed that a pangolin coronavirus (pangolin-CoV-2020) is genetically associated with both 2019-nCoV and a group of bat coronaviruses. There is a high sequence similarity between pangolin-CoV-2020 and 2019-nCoV. However, phylogenetic analyses did not support the 2019-nCoV arose directly from the pangolin-CoV-2020.

It is of interest that the genomic sequences for coronaviruses detected from two batches of pangolins intercepted by two different customs at different dates were all be associated with bat coronaviruses. The reads from the third pangolin acquired in July of 2019 were relatively less abundant than the two from the first two pangolin samples acquired in March of 2019. Although we are unclear whether these two batches of smuggling and exotic pangolins were from the same origin, our results indicated that pangolin be a natural host for *Betacoronaviruses*, which could be enzootic in pangolins. All three exotic pangolins detected with *Betacoronaviruses* in this study were very sick with serious respiratory diseases and failed to be rescued. However, these pangolins were very stressful in the transportation freight when being intercepted by the customs. It is unclear whether this coronavirus a common virus

flora in the respiratory tracts of pangolins. Nevertheless, the pathogenesis of this coronavirus to pangolin remains to be studied.

Compared to the genomic sequence of pangolin-CoV-2020 we assembled in this study, phylogenetic trees suggested that a bat origin coronavirus (i.e., Bat-CoV-RaTG13) was more genetically close to the 2019-nCoV at both individual gene and genomic sequence level. Interestingly, the cleavage site between S1 and S2 at the 2019-nCoV had multiple insertions (i.e. PRRA), compared to that of Bat-CoV-RaTG13 and pangolin-CoV-2020, which were similar. Thus, although it is clear that 2019-nCoV is of bat origin, it is likely another intermediate host could be involved in emergence of the 2019-nCoV.

The S protein of coronaviruses bind to host receptors via receptor-binding domains (RBDs), and plays an essential role in initiating virus infection and determines host tropism [2]. A prior study suggested that the 2019-nCoV, SARS-CoV, and Bat-CoV-RaTG13 had similar RBDs, suggested all of them use the same receptor ACE2 [9]. Our analyses showed that pangolin-CoV-2020 had a very conserved RBD to these three viruses rather than MERS-CoV, suggesting that pangolin-CoV is very likely to use ACE2 as its receptor. On the other hand, ACE2 receptor is present in pangolins with a high sequence conservation with those in the gene homolog in humans. However, the zoonosis of this pangolin-CoV-2020 remains unclear.

The host range of animal origin coronaviruses was promiscuous [12]. It is critical to determine the natural reservoir and the host range of coronaviruses, especially their potential of causing zoonosis. In the last two decades, besides the 2019-nCoV, SARS

and MERS caused serious outbreaks in humans, lead to thousands of deaths [3, 4, 13, 14]. Although all of three zoonotic coronaviruses were shown to be of bat origin, they seemed to use different intermediate hosts. For example, farmed palm civets were suggested to be an intermediate host for SARS to be spilled over to humans although the details on how to link bat and farmed palm civets are unclear [15, 16, 17]. Most recently, dromedary camels in Saudi Arabia were shown to harbor three different coronavirus species, including a dominant MERS-CoV lineage that was responsible for the outbreaks in the Middle East and South Korea during 2015 [18]. Although this present study does not support pangolins would be an intermediate host for the emergence of the 2019-nCoV, our results do not prevent the possibility that other CoVs could be circulating in pangolins. Thus, large surveillance of coronaviruses in the pangolins could improve our understanding the spectrum of coronaviruses in the pangolins. Conservation of wildlife and limits of the exposures of humans to wildlife will be important to minimize the spillover risks coronaviruses from wild animals to humans.

In summary, this study suggested pangolins be a natural host of *Betacoronavirus*, with an unknown potential to infect humans. However, our data do not support the 2019-nCoV evolved directly from the pangolin-CoV.

#### **Materials and Methods**

**Data selection**. During our routine wildlife rescue efforts, one of the goals was to

identify pathogens causing wildlife diseases. In 2019, we were involved in two events of pangolin rescues, one involved with 21 smuggling pangolins in March and the second with 6 smuggling pangolins in July. From those pangolins failed to be rescued, we collected samples from different tissues and subjected for metagenomic analyses. Through viral metagenomics analyses of lower respiratory tract samples from these pangolins, we detected coronavirus in three individual animals [11]. Two of these animals were from the first batch of Malayan pangolins intercepted by Meizhou, Yangjiang, and Jiangmen customs, and the third one was from the second batch in a freight being transported from Qingyuan to Heyuan. The RNA samples from these three individuals were subjected to deep sequencing. To determine the read abundance of coronaviruses in each sample, we mapped clean reads without ribosomes and host sequences to an in-house virus reference data separated from the GenBank non-redundant nucleotide database.

Genomic assembly and sequence analyses. After examining the high similarity among the samples from three animals, to maximize the coverage of the virus genome, clean reads from three animals were pooled together and *de novo* assembled using MEGAHIT v1.2.9 [19]. The assembled contigs were used as references for mapping those the rest unmapped reads using Salmon v0.14.1 [20], and multiple rounds were implemented to maximize the mapping (Table S2).

A total of 38 contigs were identified to be highly similar to the 2019-nCoV genome (accession MN908947.3) using BLASTn and tBLASTx. GapFiller v1.10 and

SSPACE v3.0 were used to fill gaps and draft pangolin-CoV-2020 genome was constructed with ABACAS v1.3.1 (http://abacas.sourceforge.net/) [21, 22, 23].

Multiple sequence alignments were conducted using CLUSTAL Ov1.2.4 [24]. Simplot analyses were conducted with SimPlot v3.5.1 to determine the sequence similarity among 2019-nCoV (MN908947.3), pangolin-CoV-2020, Bat-CoV-RaTG13 (MN996532.1), and SARS-CoV (AY395003.1) at both the genomic sequence level and at individual gene level [25]. Sequence identity was calculated utilizing p-diatance in MEGA v10.1.7 [26].

Phylogenetic analyses. We downloaded 44 full-length genome sequences of coronaviruses isolated from different hosts from the public database (Table S1). Phylogenetic analyses were performed based on their whole genome sequences, encoding ORFs of RNA-dependent RNA polymerase (RdRp gene), the receptor binding protein spike protein (S gene), small envelope protein (E gene), as well as all other gene sequences were conducted utilizing Mrbayes [27] with 50,000,000 generations and the 25% of the generations as burnin. Best models were determined by jModeltest v2.1.7 [28]. Then, all the trees were visualized and exported as vector diagrams with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

**Acknowledgements:** We thank the De-Chun Lin and Tao Jin from Magigene Biotech. and Hanghui Kong from South China Botanical Garden support for bioinformatics analysis. This project was supported by wildlife disease monitoring and early warning

system maintenance project from National Forestry and Grassland Administration (2019072), GDAS Special Project of Science and Technology Development (grant number 2020GDASYL-20200103090, 2018GDASCX-0107), Guangzhou Science Technology and Innovation Commission (grant number 201804020080), Natural Science Foundation of China (grant number 31972847), Guangzhou science and technology project (grant number 2019001), and 2019-nCoV wildlife origin project from Guangdong department of science and technology.

## References

- 1. Song Z, Xu Y, Bao L, Zhang L, Yu P, Qu Y, et al. From SARS to MERS, thrusting coronaviruses into the spotlight. Viruses. 2019; 11(1):59.
- Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host & Microbe. 2020.
- 3. Drosten C, Günther S, Preiser W, Van Der Werf S, Brodt HR, Becker S, .et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. New England journal of medicine. 2003; 348(20):1967-1976.
- Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA.
   Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia.
   New England Journal of Medicine. 2012; 367(19):1814-1820.
- Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science. 2003; 302(5643):276-8.
- Azhar EI, El-Kafrawy SA, Farraj SA, Hassan AM, Al-Saeed MS, Hashem AM, et al. Evidence for camel-to-human transmission of MERS coronavirus. New England Journal of Medicine. 2014; 370(26):2499-2505.
- 7. 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(7477):535-8.
- 8. Jeffers SA, Tusell SM, Gillim-Ross L, Hemmila EM, Achenbach JE, Babcock GJ,

- et al. CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. Proceedings of the National Academy of Sciences. 2004; 101(44):15748-53.
- 9. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. Journal of Virology. (2020).
- 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;
   1-4.
- Liu P, Chen W, Chen JP. Viral Metagenomics Revealed Sendai Virus and Coronavirus Infection of Malayan Pangolins (*Manis javanica*). Viruses. 2019; 11(11), 979.
- 12. MacLachlan NJ, Dubovi EJ. Fenner's Veterinary Virology Fifth Ed., Chapter 29, Flaviviridae, West Nile Virus. 2017.
- World Health Organization. (2003). Consensus document on the epidemiology of severe acute respiratory syndrome (SARS) (No. WHO/CDS/CSR/GAR/2003.11).
   World Health Organization.
- 14. Cunha CB, Opal SM. Middle East respiratory syndrome (MERS) A new zoonotic viral pneumonia. Virulence. 2014; 5(6):650-654.
- 15. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005; 310(5748):676-679.
- 16. Wang LF, Shi Z, Zhang S, Field H, Daszak P, Eaton BT. Review of bats and

- SARS. Emerging infectious diseases. 2006; 12(12):1834.
- 17. Wang LF, Eaton BT. Bats, civets and the emergence of SARS. In Wildlife and emerging zoonotic diseases: the biology, circumstances and consequences of cross-species transmission (pp. 325-344). Springer, Berlin, Heidelberg. 2007.
- 18. Sabir JS, Lam TTY, Ahmed MM, Li L, Shen Y, Abo-Aba SE, et al. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. Science. 2016; 351(6268):81-84.
- 19. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. 2016; 102:3-11.
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nature methods. 2017; 14(4):417.
- 21. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics. 2009; 25(15):1968-1969.
- 22. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011; 27(4):578-579.
- 23. Nadalin F, Vezzi F, Policriti A. GapFiller: a de novo assembly approach to fill the gap within paired reads. BMC bioinformatics. 2012; 13(S14):S8.
- 24. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Science. 2018; 27(1):135-45.

- 25. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. Journal of virology. 1999; 73(1):152-160.
- 26. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution. 2018; 35(6):1547-1549.
- 27. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001; 17(8);754-755.
- 28. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nature methods. 2012; 9(8):772-772.

Table 1. Nucleotide sequence identities among the genes of pangolin-CoV-2020 and other representative coronavirus against 2019 nCoV.

	Genes (%)											
	S	E	M	N	ORF1a	ORF1b	RDRP	ORF3a	ORF6	ORF7a	ORF7b	ORF10
Pangolin-CoV-2020	83.05	99.11	93.38	95.58	89.33	89.66	91.28	92.36	95.53	93.39	91.47	99.15
Bat-CoV-RaTG13	93.11	99.55	95.93	96.90	96.04	97.31	97.80	96.24	98.36	95.59	99.22	99.15
Bat-CoV-CVZXC21	76.79	98.67	93.39	91.17	90.93	86.17	86.99	88.85	95.08	89.62	95.35	100.00
Bat-Cov-CVZC45	77.19	98.67	93.39	91.09	91.03	86.07	86.70	87.76	95.08	89.31	94.57	99.15
SARS	74.54	94.44	85.37	88.78	76.08	86.26	88.65	75.55	76.50	84.11	86.18	93.16

**Table 2.** Protein sequence identities among the genes of pangolin-CoV-2020 and other representative coronavirus against 2019 nCoV.

	Amino acid (%)											
	S	E	M	N	ORF1a	ORF1b	RDRP	ORF3a	ORF6	ORF7a	ORF7b	ORF10
Pangolin-CoV-2020	89.78	100.00	98.63	97.06	95.84	99.20	99.34	97.03	96.49	97.49	95.24	97.30
Bat-CoV-RaTG13	97.69	100.00	99.55	99.04	98.07	99.37	99.57	97.79	100.00	97.49	97.65	97.30
Bat-CoV-CVZXC21	82.08	100.00	98.64	94.10	95.72	95.51	95.69	91.66	93.22	90.09	92.77	100.00
Bat-Cov-CVZC45	82.66	100.00	98.64	94.10	95.74	95.81	96.03	90.47	93.22	89.04	92.77	97.30
SARS	78.30	95.74	90.02	90.76	80.90	95.63	96.48	96.80	62.68	88.11	84.18	82.31

Figure legends

Figure 1. Genomic comparison of pangolin-CoV-2020, 2019-nCoV, and other coronaviruses. a) genomic alignment of pangolin-CoV-2020 and 2019-nCoV, white indicates missing sequence; b) Similarity plot based on the full-length genome sequence of 2019-nCoV. Full-length genome sequences of Bat-CoV-RaTg13, Bat-CoV-SL-CoVZXC21, SARS, and pangolin-CoV-2020 draft genome were used as subject sequences; c) Phylogenetic tree based on nucleotide sequences of complete genomes of coronaviruses.

**Figure 2. Genetic analyses of the spike surface glycoprotein of pangolin-CoV-2020, 2019-nCoV, and other coronaviruses.** a) similarity plot based on the spike surface glycoprotein amino acid and nucleotide sequence of 2019-nCoV. Bat-CoV-RaTg13, and pangolin-CoV-2020 were used as subject sequences; b) phylogenetic tree of S genes.

Figure 3. Phylogenetic analyses of a) small envelope gene, b) RNA-dependent RNA polymerase (RdRp) gene, c) matrix protein, and d) nucleocapsid protein sequences of coronaviruses from different hosts.

# **Supplementary Information**

# **Supplementary Tables**

**Table S1.** Accession numbers and strain IDs of coronaviruses strains isolated from different hosts.

**Table S2.** Number of sequencing reads assigned to different viruses in each pangolin sample. We only focus on Coronaviruses in this study.

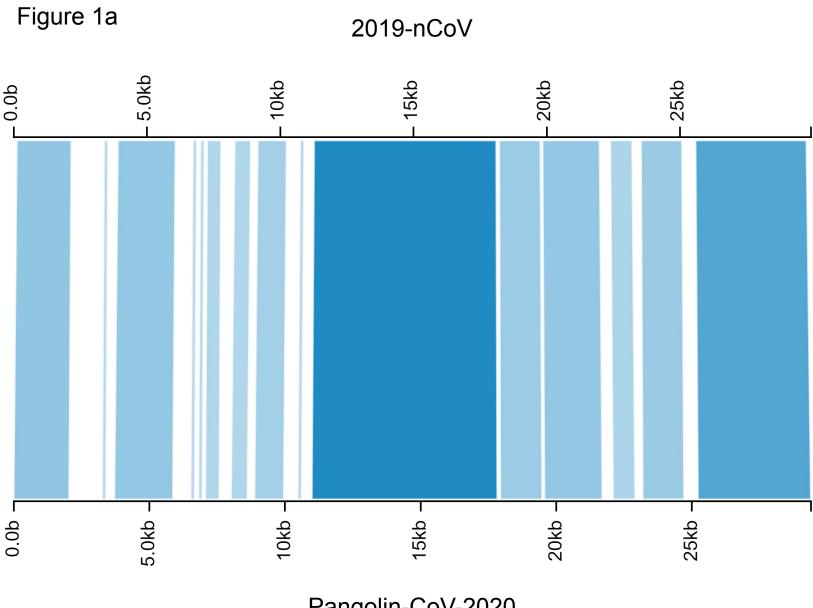
**Table S3.** The blast results for the assembled nucleotide contigs of pangolin-CoV-2020 and 2019-nCoV.

**Table S4.** The blast results for the translated proteins from assembled nucleotide contigs of pangolin-CoV-2020 and 2019-nCoV.

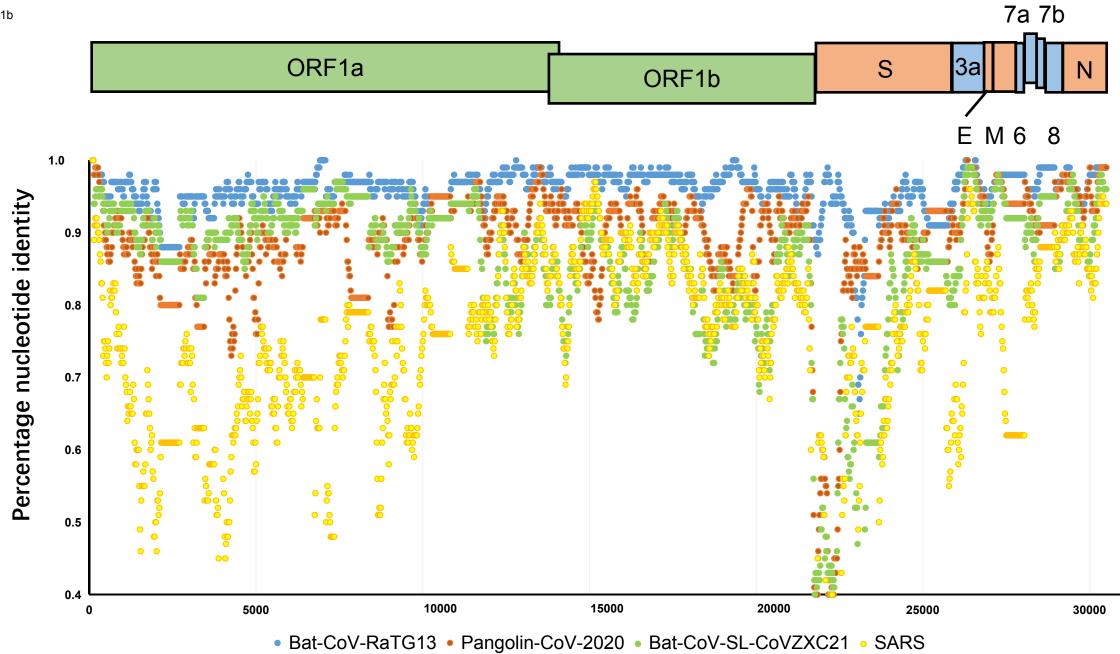
# **Supplementary Figures**

**Figure S1.** Amino acid sequence alignment of the spike surface glycoprotein of the Pangolin-CoV-2020 with 2019-nCoV and Bat-CoV-RaTG13.

**Figure S2. Phylogenetic analyses of** a) ORF1a, b) ORF1b, c) ORF3a, d) ORF6, e) ORF7a, f) ORF7b, g) ORF8, and h) ORF10 gene sequences of coronaviruses from different hosts.



Pangolin-CoV-2020



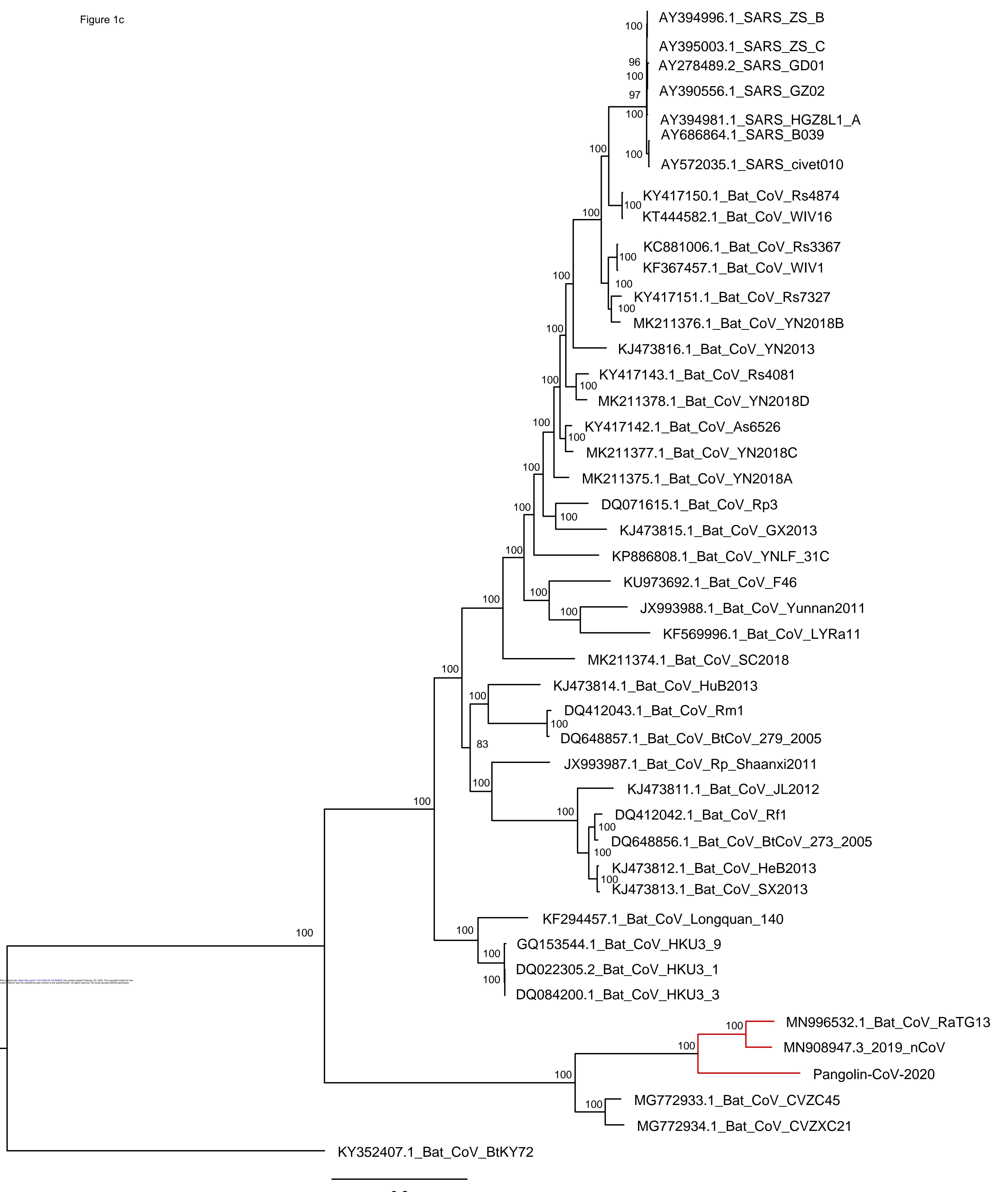


Figure 2a

