SCIENCE CHINA
Life Sciences

Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission

Xintian Xu, Ping Chen, Jingfang Wang, Jiannan Feng, Hui Zhou, Xuan Li, Wu Zhong and Pei Hao

Citation: SCIENCE CHINA Life Sciences; doi: 10.1007/s11427-020-1637-5

View online: http://engine.scichina.com/doi/10.1007/s11427-020-1637-5

Published by the Science China Press

SCIENCE CHINA
Life Sciences

•LETTER TO THE EDITOR•

https://doi.org/10.1007/s11427-020-1637-5

Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission

Xintian Xu[1†], Ping Chen[2,5†], Jingfang Wang[3†], Jiannan Feng[4], Hui Zhou[2], Xuan Li[2*], Wu Zhong[4*] & Pei Hao[1,5*]

[1]*Key Laboratory of Molecular Virology and Immunology, Institut Pasteur of Shanghai, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Shanghai 200031, China;*
[2]*Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China;*
[3]*Key Laboratory of Systems Biomedicine, Ministry of Education, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China;*
[4]*National Engineering Research Center for the Emergence Drugs, Beijing Institute of Pharmacology and Toxicology, Beijing 100850, China;*
[5]*The Joint Program in Infection and Immunity: a. Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 510623, China; b. Institute Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai 200031, China*

Received January 16, 2020; accepted January 20, 2020; published online January 21, 2020

Citation: 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 63, https://doi.org/10.1007/s11427-020-1637-5

Dear Editor,

The occurrence of concentrated pneumonia cases in Wuhan city, Hubei province of China was first reported on December 30, 2019 by the Wuhan Municipal Health Commission (WHO, 2020). The pneumonia cases were found to be linked to a large seafood and animal market in Wuhan, and measures for sanitation and disinfection were taken swiftly by the local government agency. The Centers for Disease Control and Prevention (CDC) and Chinese health authorities later determined and announced that a novel coronavirus (CoV), denoted as Wuhan CoV, had caused the pneumonia outbreak in Wuhan city (CDC, 2020). Scientists from multiple groups had obtained the virus samples from hospitalized patients (Normile, 2020). The isolated viruses were morphologically identical when observed under electron microscopy.

One genome sequence (WH-Human_1) of the Wuhan CoV was first released on Jan 10, 2020, and subsequently five additional Wuhan CoV genome sequences were released (Zhang, 2020; Shu and McCauley, 2017) (Table S1 in Supporting Information). The current public health emergency partially resembles the emergence of the SARS outbreak in southern China in 2002. Both happened in winter with initial cases linked to exposure to live animals sold at animal markets, and both were caused by previously unknown coronaviruses. As of January 15, 2020, there were more than 40 laboratory-confirmed cases of the novel Wuhan CoV infection with one reported death. Although no obvious evidence of human-to-human transmission was reported, there were exported cases in Hong Kong China, Japan, and Thailand.

Under the current public health emergency, it is imperative to understand the origin and native host(s) of the Wuhan CoV, and to evaluate the public health risk of this novel

†Contributed equally to this work
*Corresponding authors (Pei Hao, email: phao@ips.ac.cn; Wu Zhong, email: zhongwu@bmi.ac.cn; Xuan Li, email: lixuan@sippe.ac.cn)

coronavirus for transmission cross species or between humans. To address these important issues related to this causative agent responsible for the outbreak in Wuhan, we initially compared the genome sequences of the Wuhan CoV to those known to infect humans, namely the SARS-CoV and Middle East Respiratory Syndrome (MERS)-CoV (Cotten et al., 2013). The sequences of the six Wuhan CoV genomes were found to be almost identical (Figure S1A in Supporting Information). When compared to the genomes of SARS-CoV and MERS-CoV, the WH-human_1 genome that was used as representative of the Wuhan CoV, shared a better sequence homology toward the genomes of SARS-CoV than that of MERS-CoV (Figure S1B in Supporting Information). High sequence diversity between Wuhan-human_1 and SARS-CoV_Tor2 was found mainly in ORF1a and spike (S-protein) gene, whereas sequence homology was generally poor between Wuhan-human_1 and MERS-CoV.

To understand the origin of the Wuhan CoV and its genetic relationship with other coronaviruses, we performed phylogenetic analysis on the collection of coronavirus sequences from various sources. The results showed the Wuhan CoVs were clustered together in the phylogenetic tree, which belong to the Betacoronavirus genera (Figure 1A). Betacoronavirus is enveloped, single-stranded RNA virus that infects wild animals, herds and humans, resulting in occasional outbreaks and more often infections without apparent symptoms. The Wuhan CoV cluster is situated with the groups of SARS/SARS-like coronaviruses, with bat coronavirus HKU9-1 as the immediate outgroup. Its inner joint neighbors are SARS or SARS-like coronaviruses, including the human-infecting ones (Figure 1A, marked with red star). Most of the inner joint neighbors and the outgroups were found in various bats as natural hosts, e.g., bat coronaviruses HKU9-1 and HKU3-1 in Rousettus bats and bat coronavirus HKU5-1 in Pipistrellus bats. Thus, bats being the native host of the Wuhan CoV would be the logical and convenient reasoning, though it remains likely there was intermediate host(s) in the transmission cascade from bats to humans. Based on the unique phylogenetic position of the Wuhan CoVs, it is likely that they share with the SARS/SARS-like coronaviruses, a common ancestor that resembles the bat coronavirus HKU9-1. However, frequent recombination events during their evolution may blur their path, evidenced by patches of high-homologous sequences between their genomes (Figure S1B in Supporting Information).

Overall, there is considerable genetics distance between the Wuhan CoV and the human-infecting SARS-CoV, and even greater distance from MERS-CoV. This observation raised an important question whether the Wuhan CoV adopted the same mechanisms that SARS-CoV or MERS-CoV used for transmission cross species/humans, or involved a new, different mechanism for transmission.

The S-protein of coronavirus is divided into two functional units, S1 and S2. S1 facilitates virus infection by binding to host receptors. It comprises two domains, the N-terminal domain and the C-terminal RBD domain that directly interacts with host receptors (Li, 2012). To investigate the Wuhan CoV and its host interaction, we looked into the RBD domain of its S-protein. The S-protein was known to usually have the most variable amino acid sequences compared to those of ORF1a and ORF1b from coronavirus (Hu et al., 2017). However, despite the overall low homology of the Wuhan CoV S-protein to that of SARS-CoV (Figure S1 in Supporting Information), the Wuhan CoV S-protein had several patches of sequences in the RBD domain having a high homology to that of SARS-CoV_Tor2 and HP03-GZ01 (Figure 1B). The residues at positions 442, 472, 479, 487, and 491 in SARS-CoV S-protein were reported to be at receptor complex interface and considered critical for cross-species and human-to-human transmission of SARS-CoV (Li et al., 2005). Despite the patches of highly conserved regions in the RBD domain of the Wuhan CoV S-protein, four of the five critical residues are not preserved except Tyr491 (Figure 1B). Although the polarity and hydrophobicity of the replacing amino acids are similar, they raised serious questions about whether the Wuhan CoV would infect humans via binding of S-protein to ACE2, and how strong the interaction is for risk of human transmission. Note MERS-CoV S-protein displayed very little homology toward that of SARS-CoV in the RBD domain, due to the different binding target for its S-protein, the human dipeptidyl peptidase 4 (DPP4) (Raj et al., 2013).

To answer the serious questions and assess the risk of human transmission of the Wuhan CoV, we performed structural modeling of its S-protein and evaluated its ability to interact with human ACE2 molecules. Based on the computer-guided homology modeling method, the structural model of the Wuhan CoV S-protein was constructed by Swiss-model using the crystal structure of SARS coronavirus S-protein (PDB accession: 6ACD) as a template (Schwede et al., 2003). Note the amino acid sequence identity between the Wuhan-CoV and SARS-CoV S-proteins is 76.47%. Then according to the crystal structure of SARS-CoV S-protein RBD domain complexed with its receptor ACE2 (PDB code: 2AJF), the 3-D complex structure of the Wuhan CoV S-protein binding to human ACE2 was modeled with structural superimposition and molecular rigid docking (Li et al., 2005) (Figure 1C).

The computational model of the Wuhan CoV S-protein (using the WH-human_1 sequence as representative) showed a C_α RMSD of 1.45 Å on the RBD domain compared to the SARS-CoV S-protein structure (Figure 1C). The binding free energies for the S-protein to human ACE2 binding complexes were calculated by MOE 2019 with amber ff14SB force field parameters (Maier et al., 2015). The binding free energy between the Wuhan CoV S-protein and

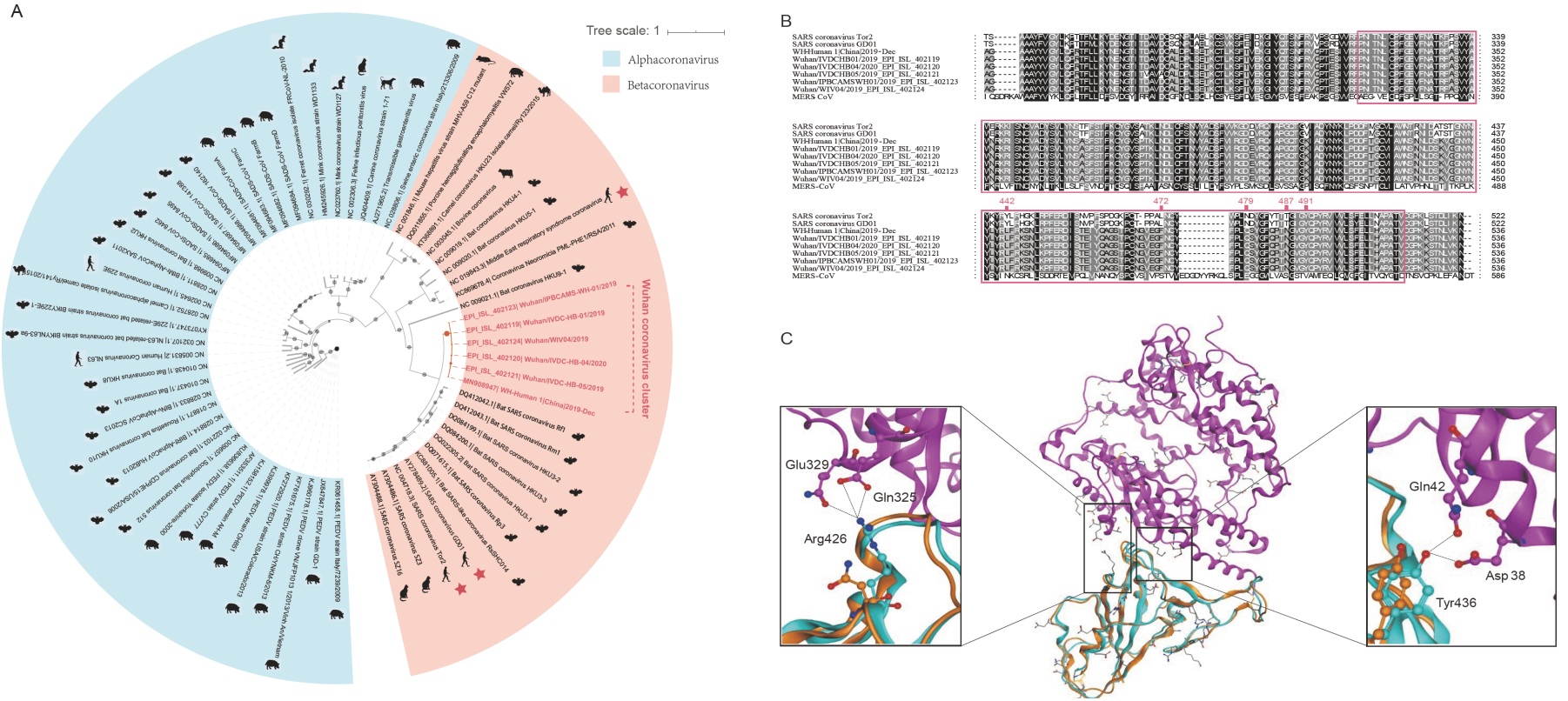

Figure 1 Evolutionary analysis of the coronaviruses and modeling of the Wuhan CoV S-protein interacting with human ACE2. A, Phylogenetic tree of coronaviruses based on full-length genome sequences. The tree was constructed with the Maximum-likelihood method using RAxML with GTRGAMMA as the nucleotide substitution model and 1,000 bootstrap replicates. Only bootstraps ≥50% values are shown as filled circles. The host for each coronavirus is marked with corresponding silhouette. Known human-infecting betacoronaviruses are indicated with a red star. B, Amino acid sequence alignment of the RBD domain of coronavirus S-protein. Residues 442, 472, 479, 487, and 491 (numbered based on SARS-CoV S-protein sequence) are important residues for interaction with human ACE2 molecule. C, Structural modeling of the Wuhan CoV (WH-human_1 as representative) S-protein complexed with human ACE2 molecule. Middle panel: The model of the Wuhan CoV S-protein (brown ribbon) is superimposed with the structural template of the SARS CoV S-protein (light blue ribbon). The protein backbone structure of human ACE2 is represented in magenta ribbon. Left panel: The region is shown for hydrogen bonding interactions between Arg426 in S-protein and Gln325/Glu329 in ACE2. The relevant residues are presented in ball and stick representations. Right panel: The region is shown for hydrogen bonding interactions between Tyr436 in S-protein and Asp38/Gln42 in ACE2.

human ACE2 was –50.6 kcal mol^{-1}, whereas that between SARS-CoV S-protein and ACE2 was –78.6 kcal mol^{-1}. A value of –10 kcal mol^{-1} is usually considered significant. Because of the loss of hydrogen bond interactions due to replacing Arg426 with Asn426 in the Wuhan CoV S-protein, the binding free energy for the Wuhan CoV S-protein increased by 28 kcal mol^{-1} when compared to the SARS-CoV S-protein binding. Although comparably weaker, the Wuhan CoV S-protein is regarded to have strong binding affinity to human ACE2. So to our surprise, despite replacing four out of five important interface amino acid residues, the Wuhan CoV S-protein was found to have a significant binding affinity to human ACE2. Looking more closely, the replacing residues at positions 442, 472, 479, and 487 in the Wuhan CoV S-protein did not alter the structural confirmation. The Wuhan CoV S-protein and SARS-CoV S-protein shared an almost identical 3-D structure in the RBD domain, thus maintaining similar van der Waals and electrostatic properties in the interaction interface.

In summary, our analysis showed that the Wuhan CoV shared with the SARS/SARS-like coronaviruses a common ancestor that resembles the bat coronavirus HKU9-1. Our work points to the important discovery that the RBD domain of the Wuhan CoV S-protein supports strong interaction with human ACE2 molecules despite its sequence diversity with SARS-CoV S-protein. Thus the Wuhan CoV poses a significant public health risk for human transmission via the S-protein–ACE2 binding pathway. People also need to be reminded that risk and dynamic of cross-species or human-to-human transmission of coronaviruses are also affected by many other factors, like the host's immune response, viral replication efficiency, or virus mutation rate.

Compliance and ethics *The author(s) declare that they have no conflict of interest.*

Acknowledgements *This work was supported in part by grants from the National Science and Technology Major Projects for "Major New Drugs Innovation and Development" (directed by Dr. Song Li) (2018ZX09711003) of China, the National Key R&D Program (2018YFC0310600) of China, the National Natural Science Foundation of China (31771412), and Special Fund for strategic bio-resources from Chinese Academy of Sciences (ZSYS-014). We also acknowledge the National Institute for Viral Disease Control and Prevention, China CDC; Wuhan Institute of Virology, Chinese Academy of Sciences; Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College; and Wuhan Jinyintan Hospital for their efforts in research and collecting the data and genome sequencing sharing. In addition, we acknowledge GISAID (https://www.gisaid.org/) for facilitating open data sharing.*

References

CDC. (2020). 2019 Novel Coronavirus (2019-nCoV), Wuhan, China. https://www.cdc.gov/coronavirus/novel-coronavirus-2019.html.

Cotten, M., Watson, S.J., Kellam, P., Al-Rabeeah, A.A., Makhdoom, H.Q., Assiri, A., Al-Tawfiq, J.A., Alhakeem, R.F., Madani, H., AlRabiah, F. A., et al. (2013). Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study. Lancet 382, 1993–2002.

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.

Li, F. (2012). Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. J Virol 86, 2856–2858.

Li, F., Li, W., Farzan, M., and Harrison, S.C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309, 1864–1868.

Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theor Comput 11, 3696–3713.

Normile, D. (2020). Mystery virus found in Wuhan resembles bat viruses but not SARS, Chinese scientist says. https://www.sciencemag.org/news/2020/01/mystery-virus-found-wuhan-resembles-bat-viruses-not-sars-chinese-scientist-says.

Raj, V.S., Mou, H., Smits, S.L., Dekkers, D.H.W., Müller, M.A., Dijkman, R., Muth, D., Demmers, J.A.A., Zaki, A., Fouchier, R.A.M., et al. (2013). Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature 495, 251–254.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003). SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 31, 3381–3385.

Shu, Y., and McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data—from vision to reality. EuroSurveillance 22, PMCID: PMC5388101.

WHO. (2020). Pneumonia of unknown cause in China. https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/.

Zhang, Y.Z. (2020). Initial genome release of novel coronavirus. http://virological.org/t/initial-genome-release-of-novel-coronavirus/319?from=groupmessage.

SUPPORTING INFORMATION

Figure S1 Sequence similarity analysis among six Wuhan CoV genomes and between Wuhan CoV and the human-infecting coronaviruses, SARS-CoV and MERS-CoV.

Table S1 Acknowledgement to the authors, originating and submitting laboratories of the sequences from GISAID

The supporting information is available online at http://life.scichina.com and https://link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.