Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China

A novel coronavirus (CoV) named “2019 novel coronavirus” or “2019-nCoV” by the World Health Organization (WHO) is responsible for the recent pneumonia outbreak that started in early December, 2019 in Wuhan City, Hubei Province, China (Huang et al., 2020; Zhou et al., 2019). The first ORF representing approximately 67% of the entire genome encodes 16 non-structural proteins (nsps), while the remaining ORFs encode accessory proteins and structural proteins (Cui et al., 2019). The four major structural proteins are the spike surface glycoprotein (S), small envelope protein (E), matrix protein (M), and nucleocapsid protein (N). The spike surface glycoprotein plays an essential role in binding to receptors on the host cell and determines host tropism (Li, 2016; Zhu et al., 2019). The genome of coronaviruses, whose size ranges between approximately 26,000 and 32,000 bases, includes a variable number (from 6 to 11) of open reading frames (ORFs) (Song et al., 2019). The genome of 2019-nCoV consists of more than 13 ORFs with 29,923 nucleotides.

A comparison of genomes of these three coronaviruses, namely Wuhan/IVDC-HB-01/2019 (GISAID accession ID: EPI_ISL_402119) (HB01), Wuhan/IVDC-HB-04/2019 (EPI_ISL_402120) (HB04), and Wuhan/IVDC-HB-05/2019 (EPI_ISL_402121) (HB05), an in-depth genome annotation of this virus was performed with a comparison to related coronaviruses, including 1,008 human SARS-CoV, 338 bat SARS-like CoV, and 3,131 human MERS-CoV, whose genomes were published before January 12, 2020 (release date: September 12, 2019) from Virus Pathogen Database and Analysis Resource (ViPR) (http://www.viprbrc.org/) and NCBI.

Comparison of genomes of these three strains showed that they are almost
identical, with only five nucleotide differences in the genome of ~29.8 kb nucleotides (Figure S1). The 2019-nCoV genome was annotated to possess 14 ORFs encoding 27 proteins (Figure 1A and Tables S1A and S1B). The orf1ab and orf1a genes located at the 5' terminus of the genome respectively encode the pp1ab and pp1a proteins, respectively. They together comprise 15 nsps including nsp1 to nsp10 and nsp12 to nsp16 (Figure 1A and Table S1B). The 3' terminus of the genome contains four structural proteins (S, E, M, and N) and eight accessory proteins (3a, 3b, p6, 7a, 7b, 8b, 9b, and orf14). At the amino acid level, the 2019-nCoV is quite similar to that of SARS-CoV, but there are some notable differences. For example, the 8a protein is present in SARS-CoV and absent in 2019-nCoV; the 8b protein is 84 amino acids in SARS-CoV, but longer in 2019-nCoV, with 121 amino acids; the 3b protein is 154 amino acids in SARS-CoV, but shorter in 2019-nCoV, with only 22 amino acids (Table S1A). Further studies are needed to characterize how these differences affect the functionality and pathogenesis of 2019-nCoV.

As shown in a phylogenetic tree based on whole genomes (Figures 1B and S2) with the Molecular Evolutionary Genetics Analysis (MEGA) (version 7.0), the 2019-nCoV is in the same Betacoronavirus clade as MERS-CoV, SARS-like bat CoV, and SARS-CoV. The phylogenetic tree falls into two clades. The Betacoronavirus genus constitutes one clade, while the Alphacoronavirus, Gammacoronavirus, and Deltacoronavirus genera constitute the other clade. The 2019-nCoV is parallel to the SARS-like bat CoVs, while the SARS-CoVs are descended from the SARS-like bat CoVs, indicating that 2019-nCoV is closer to the SARS-like bat CoVs than the SARS-CoVs in terms of the whole genome sequence. Tables S1C and S1D also show that the genome
of 2019-nCoV has the highest similarity with that of a SARS-like bat CoV (MG772933). In comparison, 2019-nCoV is distant from and less related to the MERS-CoVs. In terms of the encoded proteins of pp1ab, pp1a, envelope, matrix, accessory protein 7a, and nucleocapsid genes, phylogenetic analyses showed that the 2019-nCoV is closest to the SARS-like bat CoVs (Figure 1C and Table S1D). Regarding the spike gene, the 2019-nCoV is closest to the bat CoVs, while the 3a and 8b accessory genes are both closest to the SARS-CoVs. Although phylogenetic analyses for the whole genome and individual genes clearly show that the 2019-nCoV is most closely related to SARS-like bat CoVs, while the 3a and 8b accessory genes are both closest to the SARS-CoVs.
viruses (Figures 1B and 1C), we did not find a single strain of a SARS-like bat virus that harbors all proteins with the most similarity to counterparts of the 2019-nCoV (Figures 1B and 1C).

Given the close relationship between 2019-nCoV and SARS-CoVs or SARS-like bat CoVs (Figures 1B and 1C), an examination of the amino acid substitutions in different proteins could shed light into how 2019-nCoV differs structurally and functionally from SARS-CoVs. In total, there were 380 amino acid substitutions between the amino acid sequences of 2019-nCoV (HB01) and the corresponding consensus sequences of SARS and SARS-like viruses (Figure 2 and Tables S1E and S1F). No amino acid substitutions occurred in nonstructural protein 7 (nsp7), nsp13, envelope, matrix, or accessory proteins p6 and 8b (Table S1F). Respectively, 102 and 61 amino acid substitutions are located in nsp3 and nsp2. In addition, 27 amino acid substitutions were found in the spike protein with a length of 1,273 amino acids, including six substitutions in the RBD at amino acid region 569-655. Moreover, four substitutions (Q560L, S570A, P663S, and R694C) in the C-terminal of the receptor-binding subunit S1 domain (Figure 2) are situated in two peptides previously reported to be antigens for SARS-CoV (Guo et al., 2004).

Due to very limited knowledge of this novel virus, we are unable to give reasonable explanations for the significant number of amino acid substitutions between the 2019-nCoV and SARS or SARS-like CoVs. For example, no amino acid substitutions were present in the receptor-binding motifs that directly interact with human receptor ACE2 protein in SARS-CoV (Ge et al., 2013), but six mutations occurred in the other region of the RBD. Whether these differences could affect the host tropism and transmission property of the 2019-nCoV compared to SARS-CoV is worthy of future investigation.

SUPPLEMENTAL INFORMATION
Supplemental Information can be found online at https://doi.org/10.1016/j.chom.2020.02.001.

ACKNOWLEDGMENTS
This work was supported by the National Key Plan for Scientific Research and Development of China (2016YFD0500301 and 2016YFC1200202), CAMS Initiative for Innovative Medicine (CAMS-I2M and 2016-I2M-1-005), the National Natural Science Foundation of China (U1603126), the Central Public-Interest Scientific Institution Basal Research Fund (2016ZX10105, 2017PT31026, and 2018PT30116), and NIH R01AI089120 (United States).

REFERENCES
China CDC (2020). Tracking the Epidemic. http://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm?from=timeline#Beijing%20Municipality%20Update.
Cui, J., Li, F., and Shi, Z.L. (2019). Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. 17, 181–192.
Drosten, C., Kellam, P., and Memish, Z.A. (2014). Evidence for camel-to-human transmission of MERS coronavirus. N. Engl. J. Med. 371, 1359–1360.
Ge, X.Y., Li, J.J., Yang, X.L., Chmura, A.A., Zhu, G., Epstein, J.H., Mazet, J.K., Hu, B., Zhang, W., Peng, C., et al. (2013). Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503, 535–538.
Guan, Y., Zheng, B.J., He, Y.Q., Liu, X.L., Zhuang, Z.X., Cheung, C.L., Luo, S.W., Li, P.H., Zhang, L.J., Guan, Y.J., et al. (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302, 276–278.
Guo, J.P., Petric, M., Campbell, W., and McGeer, P.L. (2004). SARS corona virus peptides recognized by antibodies in the sera of convalescent viruses. Virology 324, 251–256.
Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet, https://doi.org/10.1016/s0140-6736(20)30183-5.
Jeffers, S.A., Tussell, S.M., Gillim-Ross, L., Hemmila, E.M., Achenbach, J.E., Babcock, G.J., Thomas, W.D., Jr., Thackray, L.B., Young, M.D., Mason, R.J., et al. (2004). CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. Proc. Nati. Acad. Sci. USA 101, 15748–15753.
Li, F. (2016). Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu. Rev. Virol. 3, 237–261.
Song, Z., Xu, Y., Bao, L., Zhang, L., Yu, P., Qu, Y., Zhu, H., Zhao, W., Han, Y., and Qin, C. (2019). From SARS to MERS, Thrusting Coronaviruses into the Spotlight. Viruses 11, E59.
Tang, Q., Song, Y., Shi, M., Cheng, Y., Zhang, W., and Xia, X.Q. (2015). Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition. Sci. Rep. 5, 17155.
Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., et al. (2020). Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. bioRxiv, https://doi.org/10.1101/2020.01.22.914952.
Zhu, Z., Zhang, Z., Chen, W., Cai, Z., Ge, X., Zhu, H., Jiang, T., Tan, W., and Peng, Y. (2018). Predicting the receptor-binding domain usage of the coronavirus based on kmer frequency on spike protein. Infect. Genet. evol. 67, 183–184.
Zhu, N., Zhang, D., Wang, L., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al.; China Novel Coronavirus Investigating and Research Team (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. N. Engl. J. Med. https://doi.org/10.1056/NEJMoa2001017.