2019-nCoV: A possible progenitor for SARS-CoV with bat origin?

Yongchang Xu1,2,4[#], Leyi Wang3[#], Xu Jia2, Youjun Feng1,2,4*

1 Department of Pathogen Biology & Microbiology, and Department of General

Intensive Care Unit of the Second Affiliated Hospital, Zhejiang University School of

Medicine, Hangzhou, Zhejiang 310058, China

2 Non-coding RNA and Drug Discovery Key Laboratory of Sichuan Province,

Chengdu Medical College, Chengdu, Sichuan 610500, China

3 Department of Veterinary Clinical Medicine and the Veterinary Diagnostic

Laboratory, College of Veterinary Medicine, University of Illinois, Urbana, IL, 61802,

USA

4 College of Animal Sciences, Zhejiang University, Hangzhou, Zhejiang 310058,

China

[#] these authors contribute equally to this work.

*Correspondence: Youjun Feng (fengyj@zju.edu.cn)

Abstract

A novel coronavirus (2019-nCoV) that is initially found to trigger human severe respiratory illness in Wuhan City of China, 2019, has been recognized as a public health emergency of international concern. In the past two months, this deadly agent has caused 77,785 cases with 2,666 deaths via rapid person-to-person transmission and reached at least 25 countries. However, its evolutionary origin is poorly understood. Here we show integrative evidence that 2019-nCoV is a possible progenitor for SARS-CoV with bat origin. Our finding underscores the importance of tracing origin in the efficient monitoring, and effectively preventing the interspecies transmission of such emerging/re-emerging coronaviruses.

Key words: Bat SARS-like CoV, SARS-CoV, 2019-nCoV, Phylogeny, Spike protein, Viral and host fusion

Running title: 2019-nCoV is an ancestor of SARS-CoV

Introduction

Emerging and re-emerging infectious diseases (like Severe Acute Respiratory Syndrome (SARS) [1, 2] and Ebola Virus Disease (EVD) [3]) are threatening global public health [4, 5]. As such zoonotic pathogens, coronaviruses (CoVs) result in a number of notorious infections (ranging from respiratory, enteric, to neurologic diseases) [4, 6]. Essentially, these causative agents constitute an expanding family *Coronaviridae* of enveloped viruses that harbor a single-stranded, positive-sense RNA genome (26 to 32kb) [3]. In general, four genera in family *Coronaviridae* are classified [7], namely α-CoV (such as HCoV-229E & HCoV-NL63), β-CoV exemplified with SARS-CoV and Middle Eastern Respiratory Syndrome CoV (MERS-CoV), γ-CoV (such as infectious bronchitis virus), and δ-CoV (such as porcine δ-CoV). Particularly, natural reservoirs of α-CoV and β-CoV are bats and rodents [8-11]. Prior to the outbreak of 2019 acute respiratory diseases (formerly called 2019-nCoV pneumonia in Wuhan), six known CoV species cause human diseases. In general, the four species of HCoVs (229E, NL63, OC43 and HKU1) are mild causatives of common cold-like syndrome, in that they preferentially cause self-limiting upper respiratory tract diseases. However, SARS-CoV and MERS-CoV are pandemic agents that are responsible for two large-scaled outbreaks of fatal human respiratory diseases: i) SARS originated from South China, 2002-2003 (774 deaths of 8000 cases, ~10% mortality) [2, 12], and ii) MERS occurred in Middle East, 2012 (804 deaths of 2266 cases, 35.5% mortality) [13, 14].

It seems likely that the novel CoV initially found in Wuhan City of Hubei Province, China, 2019 (initially called 2019-nCoV, and then formally named COVID-19 by WHO) represents the 7th human CoV pathogen [15, 16]. The clinical feature of those patients infected with 2019-nCoV refers to atypical viral pneumonia with abnormal findings on chest computed tomography (CT), implicating acute respiratory distress syndrome [17, 18]. Epidemiological studies reveal that certain inpatients with unknown pneumonia consistently display the history of exposure to Huanan seafood wholesale wet market in Wuhan City, China [15, 16, 19], implying the possibility of its wild animal-to-human transmission at the initial stage. Subsequently, a large-scale retrospective study of 425-confirmed human cases including familial clusters of pneumonia [18, 20], underlines that 2019-nCoV also acquires an ability of person-to-person transmission [15]. As of preparing this draft, not only has 2019-nCoV almost disseminated into the whole China, but also spread abroad to no less than 25 countries (or regions), including United States and Vietnam (https://www.cdc.gov/coronavirus/2019-ncov/locations-confirmed-cases.html).

As of February 20, 2020, 74,579 human cases are confirmed to be positive for 2019-nCoV (2130 death, 2.86% fatality), as well as 49,22 suspected cases in China alone. Given the rapid global spread and boosted number of confirmed cases, the World Health Organization (WHO) declares the outbreak of 2019-nCoV in China as a Public Health Emergency of International Concern (PHEIC). Additionally, a modelling study by Wu and coworkers [21] predicts that a basic reproductive number of 2019-

nCoV epidemic is 2.68 (95%, CtI. 2.47-2.86), suggesting a possibility of appropriate

75, 815 individuals with atypical viral pneumonia (95%, CrI 37,304–130,330) in

Wuhan City alone, as of Jan 25, 2020. Therefore, this unexpected outbreak of acute

respiratory illness is due to the emergence of a novel CoV, 2019-nCoV of global

health concern [16, 19, 22]. To prevent its further spread worldwide, this requires close

monitoring, effective interference strategies, and development of promising vaccines

(and/or anti-viral therapies) against this ongoing, and quickly-devastating challenge

with 2019-nCoV [19, 22].

Despite that clinical and genomic aspects of 2019-nCoV outbreaks are

increasingly clear, the evolutionary relationship of 2019-nCoV to SARS-CoV remains

fragmentarily understood. In this paper, we aim to close this gap of knowledge. We

report integrated evidence (ranging from viral genomics, phylogeny, to structural

biology) that 2019-nCoV acts as a possible progenitor of SARS-CoV, the causative

of 2002-03 pandemic of acute respiratory disease worldwide.

Materials and Methods

Phylogenetic Analysis and Bioinformatics

All the genome sequences of 2019-nCoVs are sampled from GISAID (https://www.gisaid.org/epiflu-applications/next-betacov-app/). Phylogenetic tree of 2019-nCoV isolates at the whole genome level was constructed with MEGA software in which the method of Maximum likelihood is used. The program of SimPlot (https://www.mybiosoftware.com/simplot-3-5-1-sequence-similarity-plotting.html) was applied in the measurement of viral genomic similarity. As for both S and N protein, sequence alignments were conducted with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/), and its resultant output of phylogeny was given with the program TreeView (https://treeview-x.en.softonic.com/). The final output of RBD (and/or ACE2) alignment was produced using the program of ESPript 3.0 (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi), along with ClustalOmega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

Structural Modeling and Analysis

In addition to cellular ACE2 protein [23], viral proteins of SARS-CoV with known structures used in this study, included S protein [23], fusion core [24], and 3CL [25, 26]. The Cryo-EM structure of the S/ACE2 complex is under accession number: 6ACG. The X-ray structure of fusion core of SARS-CoV S protein is deposited under accession number: 1WYY. The 3CL structures of SARS-CoV and 2019-nCoV 3CL are

accessed via the accession numbers of 1Z1I and 6LU7, respectively. Structural

modeling was performed

using the program of SWISS-MODEL (https://swissmodel.expasy.org/). The

structural photographs were given with PyMol (https://pymol.org/2/).

Results and Discussion

Genomic Insights into 2019-nCoV

The availability of viral genomes renders it possible to determine origin, evolution, and pathogenesis of 2019-nCoV, this deadly agent of atypical pneumonia. On January 28, 2020, a Chinese scientist, Yong-Zhen Zhang completed a first genome sequence of 2019-nCoV (29,903bp, Acc. no.: NC_045512), which refers to Wuhan-Hu-1, a patient isolate of Wuhan seafood market collected in December, 2019 [27]. Distinct from the two human pathogens (hCoVs 229E and NL63) of α-CoV (**Fig. S1**), genomic context of 2019-nCoV demonstrates that it exclusively locates within lineage B of β-CoVs (**Fig. S1**). The paradigm member of lineage B denotes SARS-CoV, rather than MERS-CoV within Lineage C (**Fig. S1**). In addition to a large polypeptide ORF1a/1b, three structural proteins are localized, namely spike (S) glycoprotein, envelope (E) protein, and matrix (M) protein (**Fig. 1A**). Among them, S protein of 2019-nCoV is a paradigmatic fusion protein responsible for viral entry into host cells (**Fig. S2**).

SimPlot-based analysis of similarity unveils that Wuhan-Hu-1, the isolate of 2019-nCoV at that time matches two bat SARS-like CoVs we have ever reported at highest level of genomic coverage (**Fig. 1B**)[10]. Namely, they give 87.8% similarity to bat-SL-CoVZC45 (29,802bp, Acc. no.: MG772933) and 87.6% similarity to bat-SL-CoVZXC21 (29,732bp, Acc. no.: MG772934)]. In contrast, it only exhibits 79.4% similarity to the paradigm strain of SARS-CoV Tor2 (29,753bp, Acc. no.: NC_004718) (**Fig. 1B**) [28], next to the aforementioned bat SARS-like CoVs. Evidently,

this supports the placement of 2019-nCoV as a new CoV, and hints its bat origin. So

far, more than 114 viral isolates with full genome sequences are been accumulated,

which include two confirmed US cases exported from Wuhan City [2019-nCoV/USA-

IL1/2020 (29,882bp, Acc. no.: MN988713) and 2019-nCoV/USA-WA1/2020

(29,882bp, Acc. no.: MN985325)] and one Shenzhen isolate from the familial cluster

of pneumonia, returned from Wuhan City (2019-nCoV_HKU-SZ-002a_2020

(29,838bp, Acc. no.: MN938384) [18]. Not only does extensive genomic analysis

reveals that all the 2019-nCoV isolates are almost identical with little change of

nucleotide acids (**Fig. 1B**), but also updates the most-homologous virus (96.0%

similarity to 2019-nCoV) from the two bat SARS-like CoVs (bat-SL-CoVZC45 & bat-

SL-CoVZXC21) [10] to another newly-released bat CoV (β-

CoV/bat/Yunnan/RaTG13/2013, Acc. no.: EPI_ISL_402131) collected from Yunnan

Province, 2013 (**Fig. 1B**) [29]. Taken together, this provides genomic evidence that

these newly-emerging viral pathogens originate from bats [27, 30].

Phylogeny of 2019-nCoV

To address its evolutionary origin or placement, we construct phylogenetic

tree of 2019-nCoV isolates using the complete genome sequences within β-CoVs.

As for 2019-nCoV, 44 representative isolates are included, 12 of which are from

sporadic cases abroad exported from Wuhan City, China, and the remaining 32

isolates are collected from patients distributed in different cities of China. Maximum

likelihood-based phylogeny reveals that i) all the isolates of 2019-nCoV outbreak are

almost identical, and clustered into the same subclade, termed Subclade I (**Fig. 2**); ii) three bat SARS-like CoVs are closely-related cousins of Subclade I, namely β-CoV/bat/Yunnan/RaTG13/2013, bat-SL-CoVZC45, and bat-SL-CoVZXC21 (**Fig. 2**); and iii) Subclade II, a neighboring cluster of Subclade I, is majorly consisted of human/civet SARS-CoVs and bat SARS-like CoVs (**Fig. 2**). Intriguingly, the afore-mentioned 3 bat SARS-like versions bridge Subclade I and Subclade II. This raises a possibility that Subclade I, is phylogenetically placed in the position of an ancestor/progenitor for Subclade II, rather than in paralleled evolution.

To further consolidate this hypothesis, we also performed phylogenetic analyses with two evolutionarily conserved proteins (S protein and nucleoprotein (N)). As expected, similar scenarios are also seen in phylogeny of both S protein (**Fig. S3**) and N protein (**Fig. S4**). Not only is 2019-nCoV closely adjacent to the two bat-SARS-like CoVs (bat SL-CoVZC45, and bat-SL-CoVZXC21), but also it is evolutionarily placed as an ancestor/progenitor for SARS-CoV-forming subclade (**Figs S3-S4**). Although it leads to the outbreak of 2019(20) atypical pneumonia, in parallels to SARS pandemic in 2002-2003, we favor believing 2009-nCoV as an ancestry version, rather than a domesticated version of SARS-CoV. This is because the latter's survival presumably requires its already being adapted gradually to human hosts. As a result, it evolves to only give a mild illness of cold-like syndrome to some extent. On the contrary, the fact that the onset of severe respiratory diseases is linked to its global spread via rapid person-to-person transmission, extensively argues the statue of 2019-nCoV as an ancestry or fresh invader.

Entry of 2019-nCoV by S Protein

The surface S glycoprotein of 2019-nCoV plays critical roles in successful entry into the infected host cells. This is dependent on efficient binding of 2019-nCoV S to angiotensin converting enzyme 2 (ACE2) receptor, followed by the formation of membrane fusion between CoV and host (**Fig. S5**). In general, the S protein of 2019-nCoV consists of two distinct domains (S1 and S2, **Fig. S2**). As predicted, a putative receptor-binding domain (RBD, 331-583aa) is also detected in C-terminus of S1 protein (its counterpart of SARS-CoV locates at the position 318-569aa). Furthermore, overall architecture of 2019-nCoV RBD (**Fig. S5C**) is in appreciable similarity to the counterparts of SARS-CoV (**Fig. S5B**) and other 3 bat SARS-like CoVs (**Figs S5D-F**). In spite that functional motifs' organization of S protein is almost identical, receptor-binding motifs (RBM) is highly variable (**Fig. S5A**). When compared with those of bat-SL-CoVZC45 and bat-SL-CoVZXC21, two discontinuous sequence inserts consistently appear in 2019-nCoV, SARS-CoV, and β-CoV/bat/Yunnan/RaTG13/2013 (**Fig. S5A**). In brief, the two inserts correspond to the two regions (i.e., 434-438aa and 473-486aa for 2019-nCoV, equivalent to 431-435aa, 460-472aa in SARS-CoV, in **Fig. S5A**). Very recently, a modeling study by Wan *et al*. [31] indicates that a putative RBM of 2019-nCoV possesses five distinct residues (L445, F486, Q493, S494, N501) from those of SARS-CoV (Y442, L472, N479, D480 & T487, **Fig. S5A**) and is implicated into crosstalk with four putative

important amino acids (K31, E35, D38, and K353) of human ACE2 receptor (**Fig. S6**)
[32]. This might suggest the possible diversity of CoV/ACE2 receptor.

Re-analysis of cryo-EM structure of SARS-CoV RBD and its ACE2 receptor
(PDB: 6ACG) allows us to illustrate a binding interface comprising two motifs (motif I
& II, **Fig. 3A**). Obviously, binding residues of SARS-CoV RBM we propose here
(**Figs 3A-C**) are quite different from the prediction by Wan *et al*. [31] on the basis of
SARS-CoV RBD domain (Acc. no.: 2AJF). Indeed, N473 of RBM interacts with Q24
and Y83 of ACE2 in motif I (**Fig. 3B**); motif II involves at least eight pairs of hydrogen
bonds [i.e., the six residues (Y408, R426, G482, Y484, T486 & Q492) provided by
SARS-CoV RBM cross-talk with the six residues (D38, Y41, Q42, Q325, N330, &
K353) donated by human ACE2, **Fig. 3C**]. Structural modeling of 2019-nCoV
RBD/ACE2 complex also gives two unique binding motifs (**Fig. 3D**): i) N487 of RBM
interacts with Q24 and Y83 of ACE2 in motif I (**Fig. 3E**); ii) though that the contact
residues of ACE2 remains intact, the residues of RBM are changed into Y421, G496,
T500 & Q506 (**Fig. 3F**). Of note, the two binding-residues (D38 and K353) of ACE2
are consistently revealed in our analysis along with other reports [31, 32]. The variable
snapshots on RBM/ACE2 are in part (if not all) due to the different structural
templates we utilized [23, 31, 32]. However, it might constitute a relatively comprehensive
picture on receptor recognition by 2019-nCoV S protein.

In addition, S2 region of S protein contains two typical motifs of heptad repeat
region: HR1 (920-970aa), and HR2 (1163-1202aa) in front of its transmembrane
region (**Fig. S2**). Obviously, such organization of motifs are quite conserved across

bat SARS-like CoVs (**Figs 4A** and **S2**). The two HR motifs of CoVs participate in formation of membrane fusion core with coiled-coil structure (**Figs 4B-F**). Not surprisingly, fusion core of 2019-nCoV (**Fig. 4C**) is structurally similar to those of SARS-CoV (**Fig. 4B**) and the other bat SARS-like CoV (**Figs 4D-F**). Thus, S protein of 2019-nCoV is a paradigmatic type I viral fusion protein, assuring its efficient entry into host cells.

Structural Similarity of SARS-CoV and 2019-nCoV 3CL

The chymotrypsin-like protease (3CL^{pro}), also called main protease (M^{pro}), is a prevalent nonstructural protein that mediates the proteolytic processing of viral replicase polypeptides, pp1a (486 kDa) and pp1ab (790 kDa) across different species of CoV. Unlike that papain-like protease 2 (PL2^{pro}) cleaves three sites [33-37], the paradigm 3CL (~34kDa) of SARS-CoV recognizes 11 cleavage sites, giving numbers of functional enzymes, like RNA-dependent RNA polymerase [25, 26]. The maturation of 3CL is dependent on its auto-processing at both N-terminal and C-terminal sites [25]. Like that of SARS-CoV (**Fig. 5A**), a 3CL^{Pro} homologue is predicted within the polypeptide ORF1a/b of 2019-nCoV (**Fig. S2**), and supposed to be excised by its own auto-cleavage ability [38]. Sequence alignment indicates that the 3CL enzymes have 98%~99% similarity each other, with the limited substitution of only ~10 residues (**Fig. 5A**). In relative to the prototypical version (H41 and C145) of SARS-CoV [26], an evolutionarily-conserved catalytic dyad (H41 and C145) is also examined in 2019-nCoV 3CL (**Fig. 5A**). X-ray crystal structure (PDB: 1Z1I) allows us

to define that SARS-CoV 3CL is composed of three unique domains [24]. Among them, the former two domains (Domain I & II) are structurally featuring with anti-parallel β-barrels and represent the reminiscent of trypsin-like serine proteases (**Fig. 5B**). By contrast, the last one (Domain III) that is linked to Domain II, seems to display a globular structure comprising five α-helices (**Fig. 5B**). A similar scenario is also seen in the overall structure of 2019-nCoV 3CL (PDB: 6LU7). More importantly, the fusion inhibitors targeting HR1 domain is a promising therapy against the 2019-nCoV infections [39]. Such almost-identical structural snapshots of 3CL proteases validate their functional identity (**Fig. 5C**), highlighting the bat origin shared by SARS-CoV and 2019-nCoV in the context of viral evolution.

Conclusion

As of preparing this manuscript, WHO(https://www.who.int/) renamed the severe respiratory illness caused by 2019-nCoV epidemic, as Corona Virus Infectious Disease (COVID-19). Meanwhile, the International Committee on Taxonomy of Viruses has replaced 2019-nCoV with SARS-CoV-2 (https://talk.ictvonline.org/) on February 11, 2020. Obviously, such declaration aims to formally recognize it as a cousin of SARS-CoV. Given that the clinical consequence of 2019-nCoV is paralogous to that of SARS, we agree with this announcement of renaming it to some extents. It rules out the possibility that 2019-nCoV is regarded a variant of SARS-CoV domesticated in an ongoing evolution. However, our phylogeny of 2019-nCoV combined with series of structural analysis also proposes an alternative explanation for the emergence of 2019-nCoV. This denotes that 2019-nCoV can be a progenitor for SARS-CoV with bat origin. However, it requires more experimental evidence, and heightens the importance of tracing the origin of 2019-nCoV epidemic in the efficient monitoring and effective preventing such emerging/re-emerging CoV.

Acknowledgements

This work was financially supported by National Natural Science Foundation of China (31830001, YF).

Declaration of interest statement

The authors declare that they have no competing interests.

Author contributions

YF and LW designed and supervised this project; YF, YX and LW performed experiments; YF, YX, LW and XJ analyzed the data and prepared the figures; YF drafted this manuscript.

References

1.   Lee, N. *et al.* A major outbreak of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* **348**, 1986-1994 (2003).

2.   Ksiazek, T.G. *et al.* A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* **348**, 1953-1966 (2003).

3.   Holmes, E.C., Dudas, G., Rambaut, A. & Andersen, K.G. The evolution of Ebola virus: Insights from the 2013-2016 epidemic. *Nature* **538**, 193-200 (2016).

4.   Cui, J., Li, F. & Shi, Z.L. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* **17**, 181-192 (2019).

5.   Gao, G.F. From "A"IV to "Z"IKV: Attacks from Emerging and Re-emerging Pathogens. *Cell* **172**, 1157-1159 (2018).

6.   Perlman, S. & Netland, J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol* **7**, 439-450 (2009).

7.   Zumla, A., Chan, J.F., Azhar, E.I., Hui, D.S. & Yuen, K.Y. Coronaviruses - drug discovery and therapeutic options. *Nat Rev Drug Discov* **15**, 327-347 (2016).

8.   Poon, L.L. *et al.* Identification of a novel coronavirus in bats. *J Virol* **79**, 2001-2009 (2005).

9.   Lau, S.K. *et al.* Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A* **102**, 14040-14045 (2005).

10.   Hu, D. *et al.* Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg Microbes Infect* **7**, 154 (2018).

11.   Feng, Y. & Gao, G.F. Towards our understanding of SARS-CoV, an emerging and devastating but quickly conquered virus. *Comp Immunol Microbiol Infect Dis* **30**, 309-327 (2007).

12.   Rota, P.A. *et al.* Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394-1399 (2003).

13.   Assiri, A. *et al.* Epidemiological, demographic, and clinical characteristics of 47 cases of Middle East respiratory syndrome coronavirus disease from Saudi Arabia: a descriptive study. *Lancet Infect Dis* **13**, 752-761 (2013).

14.   de Groot, R.J. *et al.* Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J Virol* **87**, 7790-7792 (2013).

15.   Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* (2020).

16.   Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* (2020).

17.   Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* (2020).

18.   Chan, J.F. *et al.* A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* (2020).

19.   Wang, C., Horby, P.W., Hayden, F.G. & Gao, G.F. A novel coronavirus outbreak of global health concern. *Lancet* (2020).

20.   Phan, L.T. *et al.* Importation and Human-to-Human Transmission of a Novel Coronavirus in Vietnam. *N Engl J Med* (2020).

21.   Joseph T Wu, K.L., Gabriel M Leung Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*, https://doi.org/10.1016/S0140-6736(1020)30260-30269 (2020).

22.   The, L. Emerging understandings of 2019-nCoV. *Lancet* (2020).

23.   Song, W., Gui, M., Wang, X. & Xiang, Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog* **14**, e1007236 (2018).

24.   Xu, Y. *et al.* Crystal structure of severe acute respiratory syndrome coronavirus spike protein fusion core. *J Biol Chem* **279**, 49414-49419 (2004).

25.   Muramatsu, T. *et al.* SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc Natl Acad Sci U S A* **113**, 12997-13002 (2016).

26.   Yang, H. *et al.* The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U S A* **100**, 13190-13195 (2003).

27.   Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* (2020).

28.   Marra, M.A. *et al.* The Genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399-1404 (2003).

29.   Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* (2020).

30.   Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* (2020).

31.   Wan, Y., Shang, J., Graham, R., Baric, R.S. & Li, F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *J Virol* (2020).

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

33.   Sulea, T., Lindner, H.A., Purisima, E.O. & Menard, R. Deubiquitination, a new function of the severe acute respiratory syndrome coronavirus papain-like protease? *J Virol* **79**, 4550-4551 (2005).

34.   Lindner, H.A. *et al.* The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme. *J Virol* **79**, 15199-15208 (2005).

35.   Harcourt, B.H. *et al.* Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. *J Virol* **78**, 13600-13612 (2004).

36.   Gao, F., Ou, H.Y., Chen, L.L., Zheng, W.X. & Zhang, C.T. Prediction of proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing SARS-CoV genomes. *FEBS Lett* **553**, 451-456 (2003).

37.   Barretto, N. *et al.* The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. *J Virol* **79**, 15189-15198 (2005).

38.   Muramatsu, T. *et al.* Autoprocessing mechanism of severe acute respiratory syndrome coronavirus 3C-like protease (SARS-CoV 3CLpro) from its polyproteins. *FEBS J* **280**, 2002-2013 (2013).

39.   Xia, S. *et al.* Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell Mol Immunol* (2020).
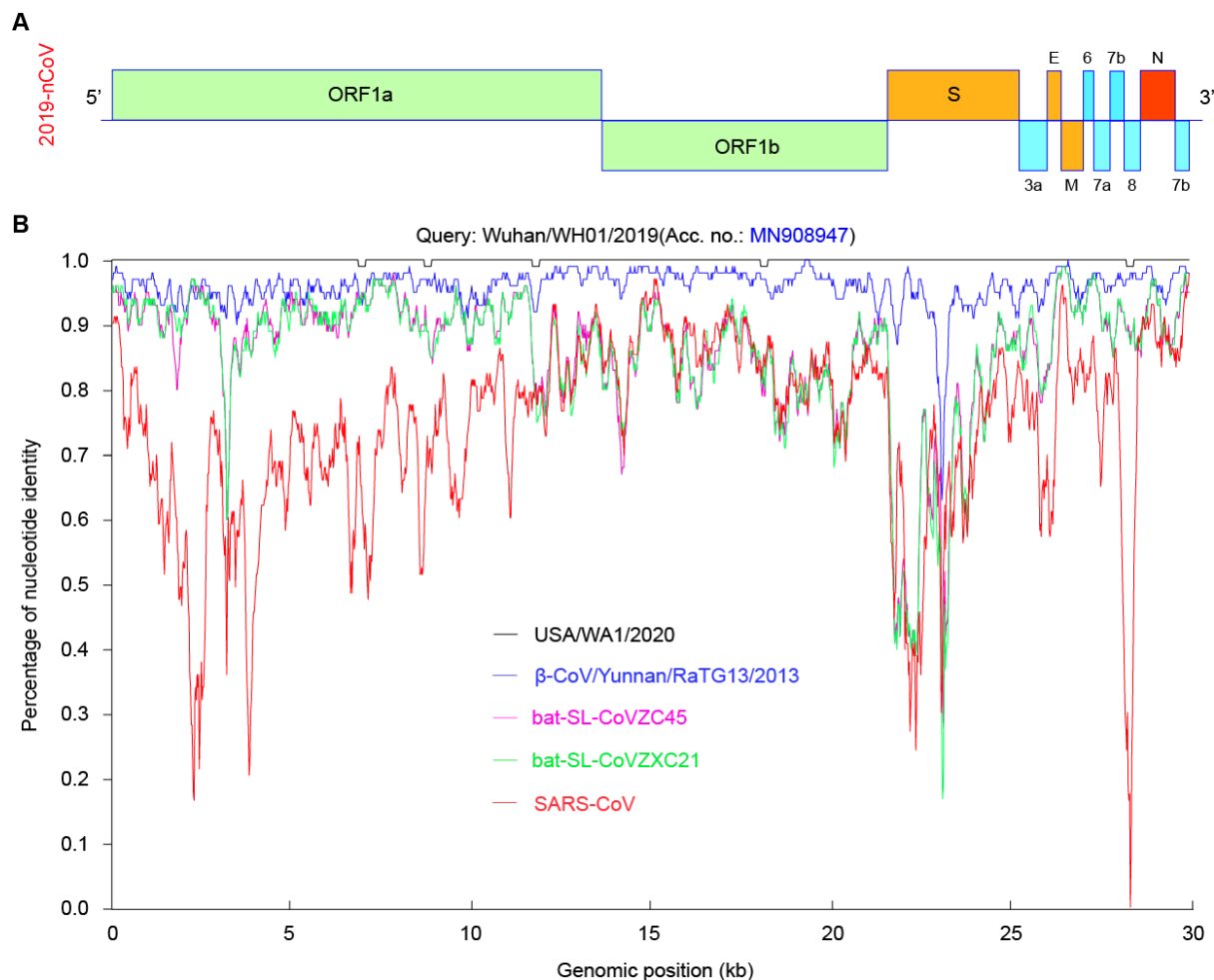
Figure legends

Figure 1 Genomic organization and similarity analysis of 2019-nCoV

A. Scheme for genomic organization of 2019-nCoV

B. Genome-wide similarity analyses of 2019-nCoV with five closely-related coronaviruses

The software of SimPlot3.5.1 was applied in viral similarity plotting, the query of which refers to Wuhan/WH01/2019 (Acc. no.: MN908947), a prototypical isolate of 2019-nCoV pneumonia outbreak in Wuhan City, 2019. The five isolates aligned correspond to USA/WA1/2020, β-CoV/Yunnan/RaTG/2013, bat-SL-CoVZC45, bat-SL-CoVZXC21, and SAARS-CoV, respectively.
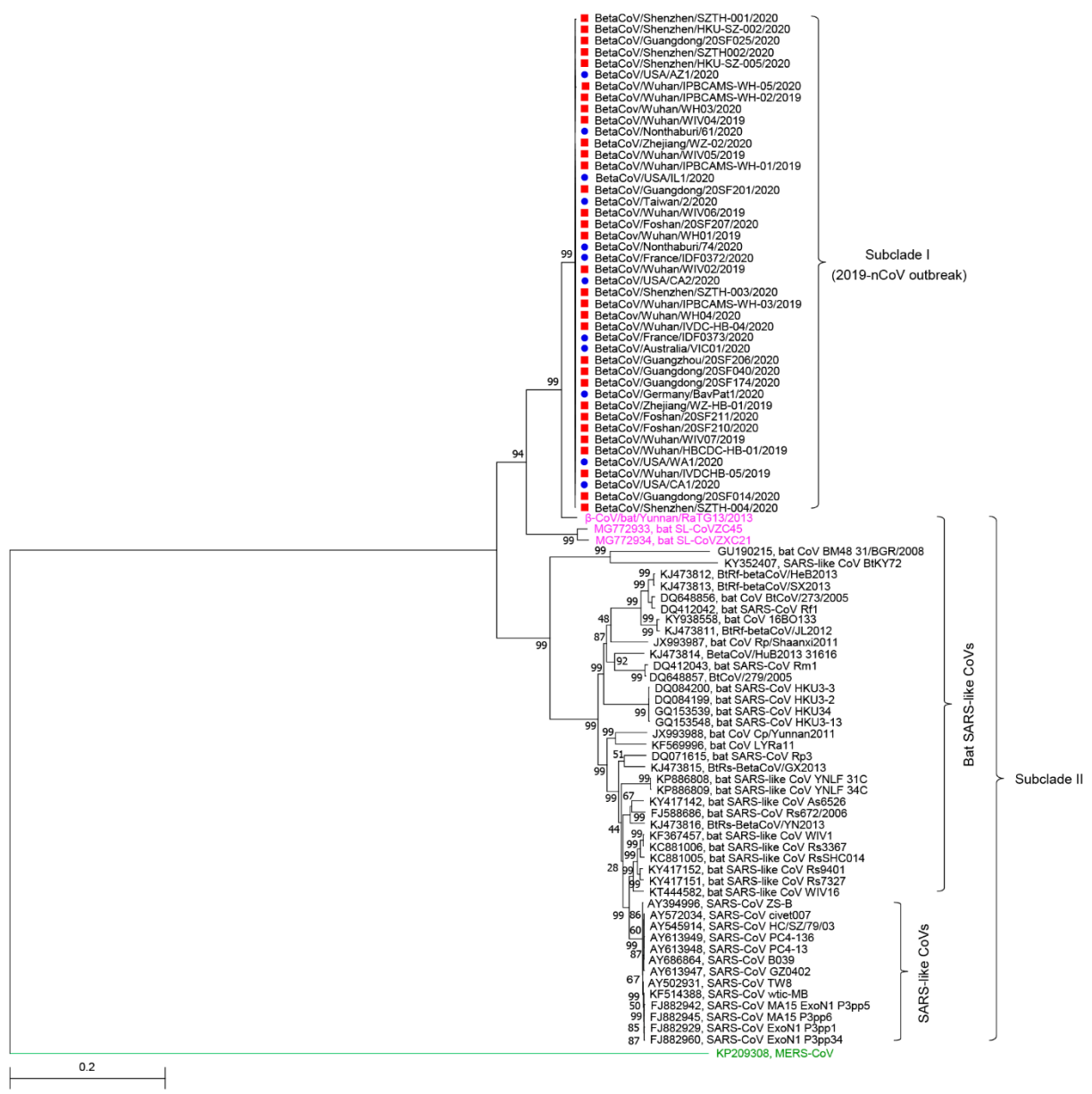
Figure 2 Phylogenetic analysis of 2019-nCoV in the context of β-coronaviruses

Maximum likelihood-based phylogeny of full genomes of 2019-nCoV isolates was

generated with the MEGA software. MERS-CoV acts as an outgroup here. It infers

that i) all the 2019-nCoV isolates form a unique subclade, termed Subclade I; ii) the

representative isolates of SARS-CoV and bat SARS-like CoV can be organized in a

neighboring subclade, named Subclade II; and three bat SARS-like CoVs (β-

CoV/Yunnan/RaTG/2013, bat-SL-CoVZC45, bat-SL-CoVZXC21) are closely-related

to 2019-nCoV (87.6%-96% similarity), and bridge Subclade I and Subclade II.

20

Evidently, we speculate that 2019-nCoV is bat origin, and behaves as possible
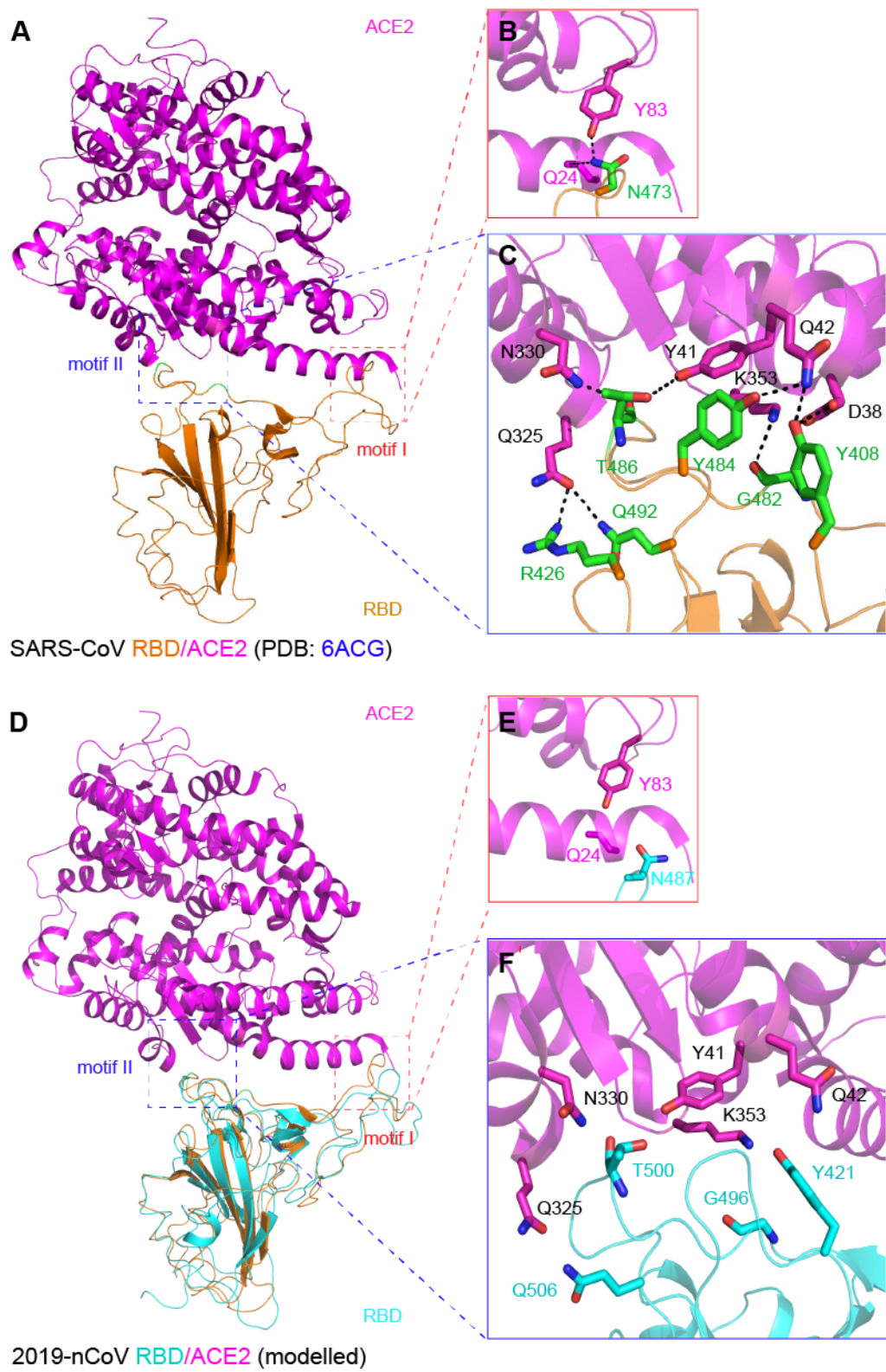
ancestor or progenitor for SARS-CoV.

Figure 3 Structural insights in interplay between 2019-nCoV RBD and its putative

receptor ACE2

A. Complex structure of SARS-CoV RBD and its receptor ACE2

RBD of SARS-CoV spike protein is in orange, and ACE2 is given in magenta. Two

critical binding motifs (motif I and motif II) are squared with dashed lines.

Enlarged views of the interface between SARS-CoV RBD and ACE2, consisting of

motif I (**B**) and motif II (**C**)

The binding of SARS-CoV spike protein its cellular receptor ACE2 is revealed by

Song *et al*. [23] with Cryo-EM structure deposited into NCBI under accession number:

6ACG.

D. Modeled structure of 2019-nCoV RBD complexed with its putative receptor ACE2

Unlike that ACE2 is in purple, RBD of 2019-nCoV S protein is in cyan. The two motifs

(motif I and motif II) essential for RBD/ACE2 binding are squared with dashed lines.

Structural snapshots of the binding of 2019-nCoV RBD to ACE2, which are featuring

with motif I (**E**) and motif II (**F**)

The important residues from ACE2 are labeled with dark letters, and those of SARS-

CoV RBD are indicated with green letters. In contrast, the residues of 2019-nCoV

RBD implicated into its contact with ACE2 are highlighted in cyan.

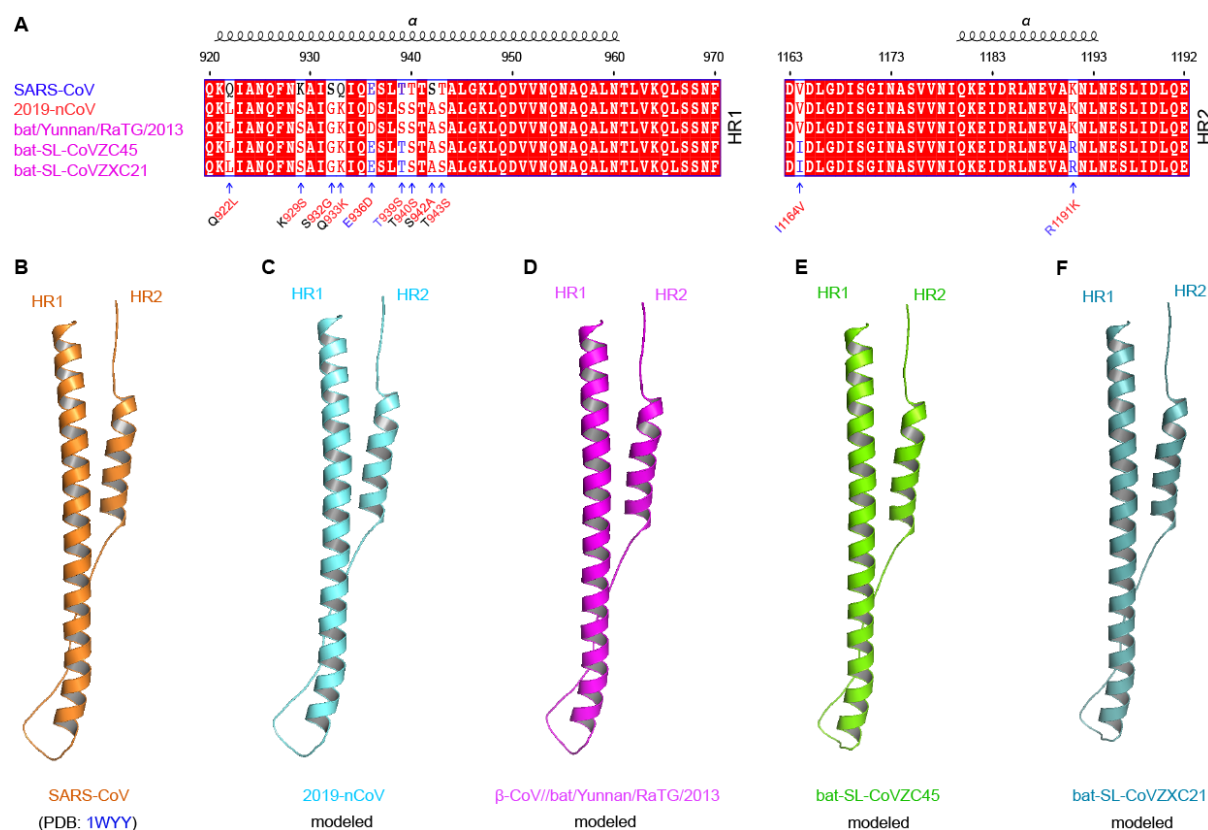Abbreviations: RBD, receptor binding domain; ACE2, angiotensin-converting enzyme

2.

Figure 4 Sequence and structural analyses of membrane protein fusion cores of 2019-nCoV

A. Sequence comparison of two heptad regions (HR), HR1 and HR2, from 2019-nCoV and its related four coronaviruses

The detailed position of HR1 (HR2) in spike proteins are depicted in Fig. S2.

B. Ribbon presentation of SARS-CoV membrane protein fusion core, a heterodimer bundle of HR1 and HR2

It is derived from the S protein structure of SARS-CoV (PDB: 1WYY).

C. Modeled structure of 2019-nCoV membrane protein fusion core

Ribbon structures of membrane protein fusion cores of β-CoV/bat/Yunnan/RaTG/2013 (**D**), bat-SL-CoVZC45 (**E**), and bat-SL-CoVZXC21 (**F**)

A flexible linker (SGGRGG) is designed to connect HR1 with HR2. Structural

24

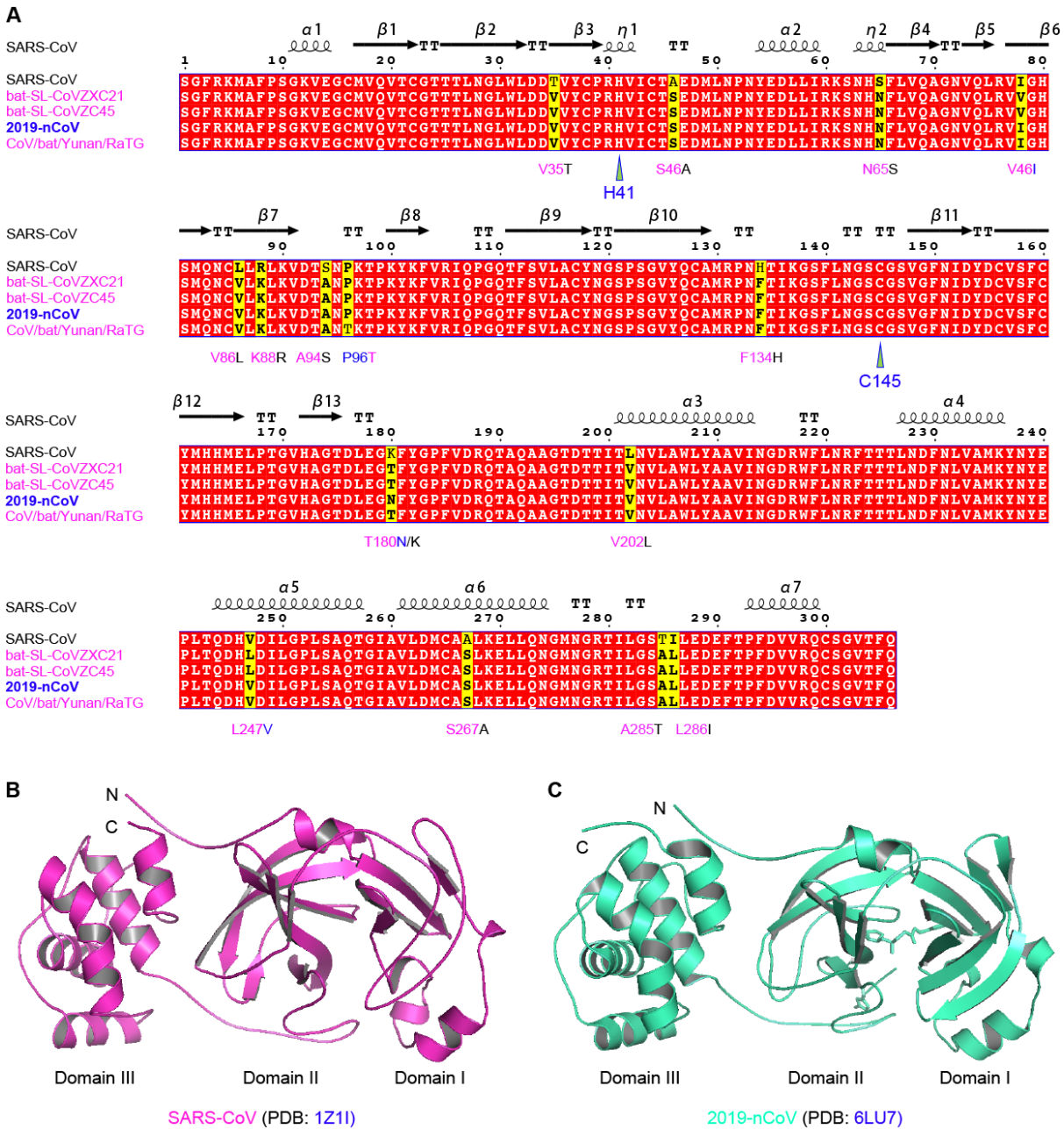modelling proceeds using the server of SWISS-MODEL

(https://swissmodel.expasy.org/).

Figure 5 Sequence and structural similarity of SARS-CoV and 2019-nCoV 3CL enzymes

A. Sequence alignment of 3CL proteases

The catalytic dyad (H41 and C145) is highlighted with triangle. The residue substitution is indicated.

In relative to that of SARS-CoV, the substitution of 12 amino acids occurs in the

2019-nCoV 3CL protease. However, only three residues alteration is present in 3CL

protease of 2019-nCoV when compared with bat-SL-CoVZXC21 (and/or bat-SL-

CoVZC45). Two residues of 3CL enzyme are different between 2019-nCoV and β-

CoV/bat/Yunan/RaTG/2013.

B. Crystal structure of SARS-CoV 3CL protease in monomeric form (PDB: 1Z1I)

C. Ribbon structure of 2019-nCoV 3CL enzyme (PDB: 6LU7)

It indicates that 3CL proteins of both SARS-CoV and 2019-nCoV are evolutionarily

conserved, and structurally similar.

Supplementary Materials for

Figure S1 Genomic context of 2019-nCoV and other viral relatives

Figure S2 Linear scheme for distinct domains/motifs in spike protein of different

coronaviruses

Figure S3 Phylogeny of 2019-nCoV surface spike protein

Multiple sequence alignment of spike protein was conducted with Clustal Omega

(https://www.ebi.ac.uk/Tools/msa/clustalo/), and the phylogenic analysis was given

with TreeView.

Figure S4 Phylogeny of 2019-nCoV nucleoprotein

Phylogenetic tree of nucleoproteins was generated with TreeView, which follows the

multiple sequence alignment with Clustal Omega

(https://www.ebi.ac.uk/Tools/msa/clustalo/).

Figure S5 Structural analyses of RBD from 2019-nCoV and four closely-related

CoVs

A. Sequence analyses of RBD from 2019-nCoV and four closely related CoVs

Putative residues of RBD interacting with ACE2 are highlighted with letters coloring

blue for SARS-CoV, and red for 2019-nCoV.

B. Ribbon structure of RBD from SARS-CoV

C. Modeled structure of 2019-nCoV RBD

Structural architectures of RBD from β-CoV/bat/Yunnan/RaTG/2013 (**D**), bat-SL-CoVZC45 (**E**), and bat-SL-CoVZC45 (**F**)

Figure S6 Sequence analysis of the ACE2 receptor

The putative residues of ACE2 indicated with blue letters, are implicated into its RBD binding. Three ACE2 proteins used here are sampled from bats, civets, and humans (homo sapiens).