## LETTER

# A Unique Protease Cleavage Site Predicted in the Spike Protein of the Novel Pneumonia Coronavirus (2019-nCoV) Potentially Related to Viral Transmissibility

Qiong Wang[1] · Ye Qiu[1] · Jin-Yan Li[1] · Zhi-Jian Zhou[1] · Ce-Heng Liao[1] · Xing-Yi Ge[1]

**Dear Editor,**

In December 2019, a novel human coronavirus caused an epidemic of severe pneumonia (Coronavirus Disease 2019, COVID-19) in Wuhan, Hubei, China (Wu *et al.* 2020; Zhu *et al.* 2020). So far, this virus has spread to all areas of China and even to other countries. The epidemic has caused 67,102 confirmed infections with 1526 fatal cases worldwide by February 14th, 2020. The viral infection incubation period varies from 2 to 14 days and typical clinical symptoms are fever, dry cough, dyspnea, headache, and pneumonia. Disease onset may result in progressive respiratory failure due to alveolar damage and even death (Chan *et al.* 2020; Chen *et al.* 2020; Huang *et al.* 2020).

The nomenclature for this coronavirus is still controversial. It was initially named as 2019-nCoV, indicating that it is a novel coronavirus identified in the year of 2019. Recently, the International Committee on Taxonomy of Viruses (ICTV) has suggested to name this virus as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) based on its phylogenetic relationship with SARS-CoV. However, many researchers point out that the name SARS-CoV-2 implies this virus as a pathogen of SARS, but actually, COVID-19 is a disease different from SARS, and the name SARS-CoV-2 may mislead the treatment and prevention of COVID-19. Hence, a new name, human coronavirus 2019 (HCoV-19), has been suggested, which is more appropriate and consistent with the disease name of COVID-19 (Jiang *et al.* 2020). Nevertheless, to avoid

potential confusions, we use 2019-nCoV to indicate the new coronavirus in this letter, mimicking most articles published recently.

Coronaviruses are enveloped non-segmented positive sense RNA viruses belonging to the genus *Betacoronavirus* of the subfamily *Orthocoronavirinae* in the family *Coronaviridae*. Based on the phylogenetic tree, 2019-nCoV is clustered into the *Sarbecovirus* subgenus with other severe acute respiratory syndrome-related coronaviruses (SARSr-CoVs), such as SARS-CoV and bat SARSr-CoV (Zhou *et al.* 2020). Compared to other human coronaviruses, 2019-nCoV emerges several unique features. First, the mortality of 2019-nCoV infection is estimated to be 3.06% which is much lower than that of SARS-CoV (10%) and MERS-CoV (37%). Second, SARS-CoV-2 has a basic reproduction rate ($R_0$) of 3.3–5.5, higher than those of SARS-CoV (2–5) and MERS-CoV (2.7–3.9), indicating a higher transmissibility of SARS-CoV-2 than other human coronaviruses (Lipsitch *et al.* 2003; Wallinga 2004; Lin *et al.* 2018). The underlying mechanism determining these features is still unclear and more relevant studies are urged considering the serious 2019-nCoV epidemic recently.

Receptor recognition is an important factor determining host range and cross-species infection of viruses. For coronaviruses, their receptor binding and subsequent internalization mainly depend on the spike protein (S protein) anchored in the viral envelope. All coronavirus S proteins consist of three domains: an extracellular domain (EC), a transmembrane anchor domain and a short intracellular tail. EC contains two functional subunits, a receptor-binding subunit (S1) and a membrane-fusion subunit (S2). S1 contains two independent domains, an N-terminal domain (S1-NTD) and receptor binding domain (RBD), which plays a key role in receptor recognition and binding (Heald-Sargent and Gallagher 2012; Li 2012). During host-virus membrane fusion, spikes protein is cleaved at the S1/S2 boundary by host proteases, releasing

---

Qiong Wang and Ye Qiu have authors contributed equally to this work.

✉ Xing-Yi Ge
xyge@hnu.edu.cn

[1]  Hunan Provincial Key Laboratory of Medical Virology, Institute of Pathogen Biology and Immunology, College of Biology, Hunan University, Changsha 410082, Hunan, China
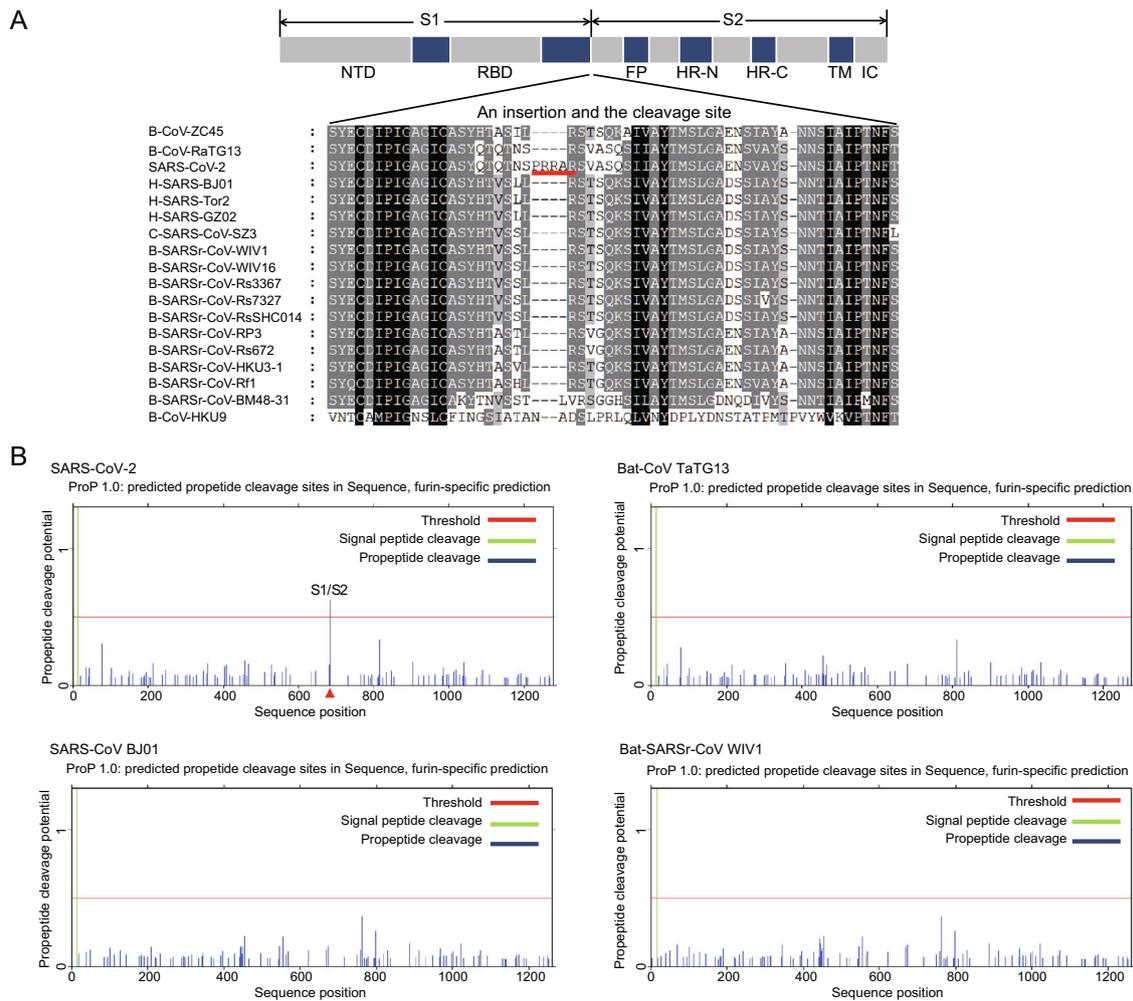
**Fig. 1** The predicted cleavage site between S1/S2 in the spike protein of 2019-nCoV. **A** The PRRA insertion (underlined) in the S of 2019-nCoV. **B** Prediction of a furin-specific cleavage site (indicated by a red arrow) in the S protein of 2019-nCoV.

the spike fusion peptide, which is necessary for virus entry. The host proteases for S protein cleavage vary among different coronaviruses, which is a key factor determining the epidemiological and pathological features of virus, including host range, tissue tropism, transmissibility and mortality. For instance, a variety of human proteases, such as trypsin, tryptase Clara, human airway trypsin-like protease (HAT) and transmembrane protease serine 2 (TMPRSS2), are known to cleave and activate the S protein of SARS-CoV (Bosch *et al.* 2008; Bertram *et al.* 2011). These proteases are widely expressed in many important organs, which is critical reason for the systematic infection, serious pathogenicity and high mortality of SARS-CoV. Laboratory studies also demonstrated that the IBV Beaudette strain, compared to other IBV strains, have an extended tropism, which infects not only primary chicken cells but also many other cell lines, due to containing an additional consensus furin cleavage site on its S protein (Yamada and Liu 2009).

In this study, we discovered one deletion and three insertions in 2019-nCoV S protein by amino acid sequence alignment. Notably, four additional amino acid residues (–PRRA–) were inserted between S1 and S2 subunits, potentially affecting the cleavage of S protein as our hypothesis (Fig. 1A). To verify our speculation, we comprehensively predicted the protease cleavage sites on different coronavirus S protein by using the ProP 1.0 server (www.cbs.dtu.dk/services/ProP/). This server is designed to predict arginine and lysine propeptide cleavage sites in eukaryotic protein sequences by using an ensemble of neural networks. As a result, 2019-nCoV S protein showed a unique furin cleavage (–RRAR–) within the S1/S2 domain which was overlapped with insertion described above (Fig. 1B). This furin cleavage site was located between the residues 682 and 685, distinct from SARS-CoV and all other SARS-like coronaviruses which only contain a trypsin or TMPRSS2 cleavage site at R667 (corresponding to residues 685 in 2019-nCoV S) (Fig. 1B).

Furin is a protease ubiquitously expressed in a variety of organs and tissues, including brain, lung, gastrointestinal tract, liver, pancreas and reproductive tissues. With the furin cleavage site on the S protein, 2019-nCoV probably gains ability to infect organs or tissues insensitive to other coronaviruses, leading to systematic infection of 2019-nCoV in the body. Even worse, the wide distribution of 2019-nCoV in a patient body may release the virus into the environment via more diverse ways, severely enhancing the transmission of 2019-nCoV. This hypothesis is supported by the current reports about the trace of 2019-nCoV in some place distinct from other coronaviruses, such as feces and eyes. However, these speculations are mainly based on our sequence studies, and further functional studies are required to characterize how these differences affect the functionality and pathogenesis of 2019-nCoV.

In summary, our sequence analysis on the S protein of 2019-nCoV has predicted a novel furin cleavage site at S1/S2 linkage. The ubiquitous expression of furin in different organs and tissues may be a reason for the high transmissibility and pathogenicity of 2019-nCoV observed in the current epidemic. However, since our findings were mainly based on bioinformatic analysis, more laboratory studies on 2019-nCoV in cell and animal models are required to verify our speculations and to avoid any bias.

## Compliance with Ethical Standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Animal and Human Rights Statement** This article does not contain any studies with human or animal subjects performed by any of the authors.

## References

Bertram S, Glowacka I, Muller MA, Lavender H, Gnirss K, Nehlmeier I, Niemeyer D, He Y, Simmons G, Drosten C, Soilleux EJ, Jahn O, Steffen I, Pohlmann S (2011) Cleavage and activation of the severe acute respiratory syndrome coronavirus spike protein by human airway trypsin-like protease. J Virol 85:13363–13372

Bosch BJ, Bartelink W, Rottier PJ (2008) Cathepsin L functionally cleaves the severe acute respiratory syndrome coronavirus class I fusion protein upstream of rather than adjacent to the fusion peptide. J Virol 82:8887–8890

Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, Xing F, Liu J, Yip CC, Poon RW, Tsoi HW, Lo SK, Chan KH, Poon VK, Chan WM, Ip JD, Cai JP, Cheng VC, Chen H, Hui CK, Yuen KY (2020) A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet. https://doi.org/10.1016/S0140-6736(20)30154-9

Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. https://doi.org/10.1016/S0140-6736(20)30211-7

Heald-Sargent T, Gallagher T (2012) Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. Viruses 4:557–580

Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. https://doi.org/10.1016/S0140-6736(20)30183-5

Jiang S, Shi Z, Shu Y, Song J, Gao GF, Tan W, Guo D (2020) A distinct name is needed for the new coronavirus. Lancet. https://doi.org/10.1016/S0140-6736(20)30419-0

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

Lin QCA, Zhao S, He D (2018) Modeling the spread of Middle East respiratory syndrome coronavirus in Saudi Arabia. Stat Methods Med Res 27:1968–1978

Lipsitch MCT, Cooper B, Robins JM, Ma S, James L et al (2003) Transmission dynamics and control of severe acute respiratory syndrome. Science 300:1966–1970

Wallinga JTP (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol 160:509–516

Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ (2020) A new coronavirus associated with human respiratory disease in China. Nature. https://doi.org/10.1038/s41586-020-2008-3

Yamada Y, Liu DX (2009) Proteolytic activation of the spike protein at a novel RRRR/S motif is implicated in furin-dependent entry, syncytium formation, and infectivity of coronavirus infectious bronchitis virus in cultured cells. J Virol 83:8744–8758

Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. https://doi.org/10.1038/s41586-020-2012-7

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus I, Research T (2020) A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. https://doi.org/10.1056/NEJMoa2001017