

Predicting the angiotensin converting enzyme 2 (ACE2) utilizing capability as the receptor of SARS-CoV-2

Ye Qiu[#], Yuan-Bo Zhao[#], Qiong Wang, Jin-Yan Li, Zhi-Jian Zhou, Ce-Heng Liao, Xing-Yi Ge^{*}

Hunan Provincial Key Laboratory of Medical Virology, Institute of Pathogen Biology and Immunology, College of Biology, Hunan University, 27 Tianma Rd., Changsha, Hunan, 410012, China

[#] These authors contributed equally to this work.

^{*} Correspondence to Dr. Xing-Yi Ge

Email: xyge@hnu.edu.cn

Phone: 86-731-88822606

Address: 27 Tianma Rd, Changsha, Hunan, 410012, China

Abstract

SARS-CoV-2, the newly identified human coronavirus causing severe pneumonia epidemic, was probably originated from Chinese horseshoe bats. However, direct transmission of the virus from bats to humans is unlikely due to lack of direct contact, implying the existence of unknown intermediate hosts. Angiotensin converting enzyme 2 (ACE2) is the receptor of SARS-CoV-2, but only ACE2s of certain species can be utilized by SARS-CoV-2. Here, we evaluated and ranked the receptor-utilizing capability of ACE2s from various species by phylogenetic clustering and sequence alignment with the currently known ACE2s utilized by SARS-CoV-2, predicting potential intermediate hosts of SARS-CoV-2.

Keywords: SARS-CoV-2; coronavirus; angiotensin converting enzyme 2 (ACE2); receptor utilization; phylogenetic analysis.

1. Introduction

In December 2019, a novel human coronavirus, SARS-CoV-2, was detected in the city of Wuhan, China, and then the virus caused a severe epidemic of Coronavirus Disease 2019 (COVID-19) in China and worldwide, critically threatening the public health [1, 2]. The epidemic has caused 74,280 confirmed infections with 2,009 fatal cases in China by February 18th, 2020, surpassing any former coronavirus epidemics. The infection incubation period of SARS-CoV-2 varies from 2-14 days, and typical clinical symptoms include fever, dry cough, dyspnea, headache, and pneumonia. Finally, COVID-19 may result in progressive respiratory failure due to alveolar damage and even death [3-5].

SARS-CoV-2 is an enveloped non-segmented positive sense RNA viruses classified into the *Sarbecovirus* subgenus of the genus *Betacoronavirus* in the subfamily *Orthocoronavirinae*, together with other severe acute respiratory syndrome-related coronaviruses (SARSr-CoVs), such as SARS-CoV [6]. Occasional interspecies transmission of virus is believed to be a major cause of coronavirus epidemic. SARS-CoV-2 has a high genetic relationship with a bat coronavirus (BatCoV RaTG13) with a 96% genomic nucleotide sequence identity. The close phylogenetic relationship to Bat RaTG13 provides evidence for a bat origin of SARS-CoV-2 [6]. However, direct transmission of the virus from bats to humans is unlikely due to the lack of direct contact between bats and humans in Wuhan, China. Hence, there are probably intermediate hosts transmitting SARS-CoV-2 to humans. This speculation is supported by the reports about the intermediate hosts of other human coronaviruses. For instance, SARS-CoV, the coronavirus causing SARS epidemic in 2003, was also originated from bats [7, 8], but it transmitted to human via a variety of intermediate hosts including masked palm civet cats and raccoon dogs [9, 10]. Since native reservoir hosts of viruses usually live far away from human community, intermediate hosts play critical roles in transmitting viruses to

human and causing epidemics. A common way to determine the intermediate hosts is to collect all animals in the original epidemic area and detect the virus in them one by one, which is quite time- and labor-consuming and also is easy to be biased by passive contamination of the samples.

Receptor recognition is an important factor determining host range and cross-species infection of viruses. Angiotensin converting enzyme 2 (ACE2) has been proved to be the cellular receptor of SARS-CoV-2 [6]. ACE2 was initially identified as an exopeptidase expressed in vascular endothelial cells in the heart and the kidney that catalyses the conversion of angiotensins [11, 12]. Later, ACE2 was well known for its function as the virus receptor of SARS-CoV [13]. Utilization of ACE2 as the receptor by SARS-CoV-2 is an important rationale to classify SARS-CoV-2 to the same subgenus as SARS-CoV. ACE2 is expressed in most vertebrates, but not all ACEs can be utilized by SARS-CoV-2 as the receptor. Currently studies demonstrated that SARS-CoV-2 is able to use Chinese horseshoe bats, civet, swine but not mouse ACE2 as its entry receptor [6]. Obviously, the utilizing capability of ACE2 by SARS-CoV-2 can be used for quick screening to narrow down the range of the intermediate hosts for SARS-CoV-2. We intended to predict the utilizing capability by SARS-CoV-2 of different ACE2s via amino acid sequence comparison. Unfortunately, the amino acid identities between human ACE2 and ACE2s of Chinese horseshoe bats, civet, swine and mouse, are 80.75%, 83.48%, 81.37%, 82.11%, respectively. Mouse ACE2 which cannot be used by SARS-CoV-2 show higher similarity to human ACE2 than those of Chinese horseshoe bats and swine which can be used by SARS-CoV-2, indicating that whole amino acid sequence identities is not a good marker for predicting utilizing capability by SARS-CoV-2.

In this study, we combined the phylogenetic analysis and critical site comparison of ACE2 to predict the intermediate hosts of SARS-CoV-2. First, we collected the amino

acid sequences of ACE2s of 250 species and built a phylogenetic tree to cluster the ACE2s based on their evolutionary distance. In the tree, the ACE2s of species close to Chinese horseshoe bats, civet or swine but far from mouse were more likely to be utilized by SARS-CoV-2. Second, we aligned the amino acid sequences of ACE2s of Chinese horseshoe bats, civet, swine and mouse to identify the sites critical for SARS-CoV-2 utilization; then we ranked the ACE2s of other species according to their amino acid substitution on these sites. The top hits were selected to be the candidates potentially utilized by SARS-CoV-2. Combining the results obtained by the two methods and considering the common animals living in Wuhan, we predicted pangolin, cat, cow, buffalo, goat, sheep and pigeon as the potential intermediate hosts for SARS-CoV-2, together with civet or swine. These results may help the screening for the intermediate hosts of SARS-CoV-2, contributing to the study of the viral transmission and disease control.

2. Methods and materials

2.1 Phylogenetic analysis

The amino acid sequences of ACE2s of *Homo sapiens* (human), *Rhinolophus sinicus* (Chinese horseshoe bat), *Paguma Larvata* (civet), *Sus scrofa* (swine) and *Mus musculus* (mouse) were downloaded from UniProt Knowledgebase; the rest of 248 vertebrate species were downloaded from GenBank (**File S1**). Multiple sequence alignment was performed for the whole sequences of ACEs using MAFFT with a local alignment strategy FFT-NS-2. The phylogenetic tree was constructed using the neighbor-joining (NJ) method with 1,000 bootstrap replicates using FigTree.

2.2 Ranking of utilizing capability by SARS-CoV-2

The critical amino acid sites for the utilization by SARS-CoV-2 were selected by sequence alignment of the extracellular domain of Chinese horseshoe bats, civet, swine and mouse ACE2s using DNAMAN 8. The initial mark for every other ACE2 was 100 and the substitution on each critical sites resulted in a deduction of its mark. The utilizing capability by SARS-CoV-2 of the ACE2s were ranked by the final scores which ranged from 0 (unlikely to be utilized) to 100 (likely to be utilized).

3. Results and discussion

To evaluate the homology of ACE2s in different species, we built a phylogenetic tree harboring ACE2s of 250 vertebrate species (**Figure S1**). The tree was branched two major groups. The first group included the classes of Mammalia, Aves and Reptilia, and the second group included the classes of Amphibian, Chondrichthyes and Teleostomi. This tree indicated that besides mammalia, ACE2s of aves and reptilia were more likely to be utilized by SARS-CoV-2 than other vertebrates. As shown in **Figure 1**, among mammals, ACE2 was clustered with bat, civet and swine ACE2s but not mouse ACE2, matching the previous report about the utilizing capability by SARS-CoV-2 of these ACE2s. Intriguingly, cat, pangolin, cow and goat ACE2s were also clustered with human ACE2. Considering the widespread of stray cats, wildlife markets and stock farms in Wuhan, it was not strange that these animals could serve as potential intermediate hosts of SARS-CoV-2. Indeed, it was reported in the local news that SARS-CoV-2 was detected in pangolins, though further verification is required.

It has been proved that human, Chinese horseshoe bats, civet and swine ACE2s can be utilized by SARS-CoV-2 but mouse ACE2 cannot. We aligned the amino acid sequence of the extracellular domains of these five ACE2s (**File S2**). Ten amino acid sites distinct from the mouse but conserved in the other four species were picked as critical

sites determining the utilizing capability by SARS-CoV-2. These sites included T20, Y83, S218, A246, K353, P426, T593, N636, A714, R716 and A774 (**Figure 2**). Y83 and K353 have been reported to play critical roles in the receptor binding of SARS-CoV, and Y83F or K353H/A/D mutants would diminish the binding of SARS-CoV, while the other sites have not been reported [14, 15]. Based on these sites, we established a marking system to rank the utilizing capability by SARS-CoV-2 of different ACE2s. In the system, every one of these 10 sites was worth 10 marks and substitution on them would result in deduction of the marks (no lower than 0), except the substitution of Y83F or K353H/A/D which would be marked as 0 for the whole sequence regardless of other sites. The ACE2s with final scores higher than 50 were considered to be potentially utilized by SARS-CoV-2, while those with scores less or equal to 50 were unlikely to be utilized by SARS-CoV-2. This ranking mainly included mammalian and bird ACE2s (**File S3**), since others were quite different from human ACE2 and other animals were unlikely to mediate the transmission of viruses among mammals (from bats to humans). As shown in **Table 1**, excluding the animals unlikely contacted by people in Wuhan, pangolin, cat, cow, buffalo, goat and sheep ACE2 were ranked as the top mammals that could be potentially used by SARS-CoV-2, while murine ACE2s, such as mouse and rat ACE2s, got the lowest scores. As for birds, pigeon ACE2 reached the highest score of 60, and is the only bird with a score higher than 50 (**Table 1**). Considering pigeons are common food in Wuhan, they could also be the intermediate host of SARS-CoV-2. ACE2 of *Protobothrops mucrosquamatus* (pallas pit viper), a common snake existing in Wuhan, was also marked but got a score of 0, indicating that reptilia ACE2s were unlikely to be utilized by SARS-CoV-2. To verify the robustness of our ranking system, we labelled the utilizing capability of different ACE2s on the phylogenetic tree, and found that the ACE2s confirmed or potential to be utilized by SARS-CoV-2 (labelled with circles) were

obviously clustered from those unlikely to be utilized by SARS-CoV-2 (labelled with triangles) (**Figure 1**). The well matching with the phylogenetic tree elevated the liability of this ranking system.

In summary, we combined phylogenetic analysis and critical site marking to predict the utilizing capability of ACE2s from different animal species by SARS-CoV-2. The results from these two analysis matched each other quite well, indicating relative reliability of our prediction. We have shown that, besides currently confirmed ACE2s utilized by SARS-CoV-2, pangolin, cat, cow, buffalo, goat, sheep and pigeon ACE2s might be utilized by SARS-CoV-2, indicating potential interspecies transmission of the virus from bats to these animals and among these animals. Considering the widespread of these animals in Wuhan, some of them might serve as intermediate hosts for SARS-CoV-2, which called for the attention in disease control. However, these are still preliminary results predicted by sequence analysis, and more laboratory and epidemiological investigation are required to uncover the true intermediate hosts of SARS-CoV-2.

4. Acknowledgements

This work was jointly funded by the National Key Research and Development Program of China (grant number 2017YFD0500104), National Natural Science Foundation of China (grant number 31470260 and 81902070), the Science Fund for Distinguished Young Scholars of Hunan Province (grant number 2019JJ20004), the Provincial Natural Science Foundation of Hunan Province (grant number 2019JJ50035) and the Fundamental Research Funds for the Central Universities of China (grant number 531107051162).

5. Declaration of interest statement

The authors declare no competing interests.

6. References

- [1] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020.
- [2] Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020.
- [3] Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, 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.
- [4] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020.
- [5] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020.
- [6] 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.
- [7] Bolles M, Donaldson E, Baric R. SARS-CoV and emergent coronaviruses: viral determinants of interspecies transmission. *Curr Opin Virol* 2011;1:624-34.
- [8] 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;503:535-8.

[9] Song HD, Tu CC, Zhang GW, Wang SY, Zheng K, Lei LC, et al. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci U S A* 2005;102:2430-5.

[10] Shi Z, Hu Z. A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res* 2008;133:74-87.

[11] Ferrario CM, Trask AJ, Jessup JA. Advances in biochemical and functional roles of angiotensin-converting enzyme 2 and angiotensin-(1-7) in regulation of cardiovascular function. *Am J Physiol Heart Circ Physiol* 2005;289:H2281-90.

[12] Donoghue M, Hsieh F, Baronas E, Godbout K, Gosselin M, Stagliano N, et al. A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circ Res* 2000;87:E1-9.

[13] Kuba K, Imai Y, Rao S, Gao H, Guo F, Guan B, et al. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat Med* 2005;11:875-9.

[14] Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J* 2005;24:1634-43.

[15] 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. *J Virol* 2020.

7. Table

Table 1. Marks of ACE2s from selected species

Species	20	83	218	246	353	426	593	636	714	716	SUM										
<i>Rhinolophus sinicus</i> (Chinese horseshoe bat)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	A	0	R	0	100
<i>Viverricula indica pallida</i> (Civet)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	A	0	R	0	100
<i>Equus caballus</i> (horse)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	A	0	R	0	100
<i>Felis catus</i> (cat)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	A	0	R	0	100
<i>Homo sapiens</i> (human)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	A	0	R	0	100
<i>Lynx canadensis</i> (lynx)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	A	0	R	0	100
<i>Manis javanica</i> (pangolin)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	A	0	R	0	100
<i>Mustela erminea</i> (mustela)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	A	0	R	0	100
<i>Sus scrofa</i> (swine)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	A	0	R	0	100
<i>Capra hircus</i> (goat)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	A	0	Q	-10	90
<i>Ovis aries</i> (sheep)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	A	0	Q	-10	90
<i>Pteropus alecto</i> (flying fox)	T	0	Y	0	N	-10	A	0	K	0	P	0	T	0	N	0	A	0	R	0	90
<i>Bos taurus</i> (cattle)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	V	-10	Q	-10	80
<i>Bubalus bubalis</i> (buffalo)	T	0	Y	0	S	0	A	0	K	0	P	0	T	0	N	0	V	-10	Q	-10	80
<i>Canis lupus dingo</i> (dog)	-	-10	Y	0	S	0	T	-10	K	0	P	0	T	0	N	0	V	-10	R	0	70
<i>Columba livia</i> (pigeon)	V	-10	-	-10	S	0	H	-10	K	0	P	0	E	-10	N	0	A	0	R	0	60
<i>Nestor notabilis</i> (parrot)	-	-10	-	-10	S	0	H	-10	K	0	P	0	N	-10	D	-10	A	0	R	0	50
<i>Merops nubicus</i> (bee-eater)	V	-10	-	-10	S	0	H	-10	K	0	P	0	K	-10	D	-10	A	0	R	0	50
<i>Egretta garzetta</i> (egret)	V	-10	-	-10	S	0	H	-10	K	0	P	0	K	-10	D	-10	A	0	R	0	50
<i>Apaloderma vittatum</i> (cuckoo)	V	-10	-	-10	S	0	H	-10	K	0	P	0	N	-10	D	-10	A	0	K	-10	40
<i>Colius striatus</i> (colius)	-	-10	-	-10	S	0	H	-10	K	0	P	0	K	-10	D	-10	A	0	G	-10	40
<i>Dasyurus novemcinctus</i> (dasypus)	T	0	F	-100	N	-10	T	-10	K	0	P	0	T	0	S	-10	V	-10	G	-10	0
<i>Grammomys surdaster</i> (thicket rat)	L	-10	F	-100	N	-10	R	-10	H	-100	S	-10	D	-10	T	-10	I	-10	R	0	0
<i>Mus caroli</i> (field mouse)	L	-10	F	-100	N	-10	R	-10	H	-100	S	-10	D	-10	T	-10	I	-10	R	0	0
<i>Mus musculus</i> (house mouse)	L	-10	F	-100	N	-10	R	-10	H	-100	S	-10	D	-10	T	-10	V	-10	G	-10	0
<i>Rattus norvegicus</i> (rat)	L	-10	F	-100	N	-10	T	-10	H	-100	S	-10	V	-10	T	-10	I	-10	G	-10	0
<i>Protobothrops mucrosquamatus</i> (pallas pit viper)	V	-10	F	-100	Q	-10	R	-10	K	0	P	-10	K	-10	D	-10	A	0	K	-10	0

8. Figures

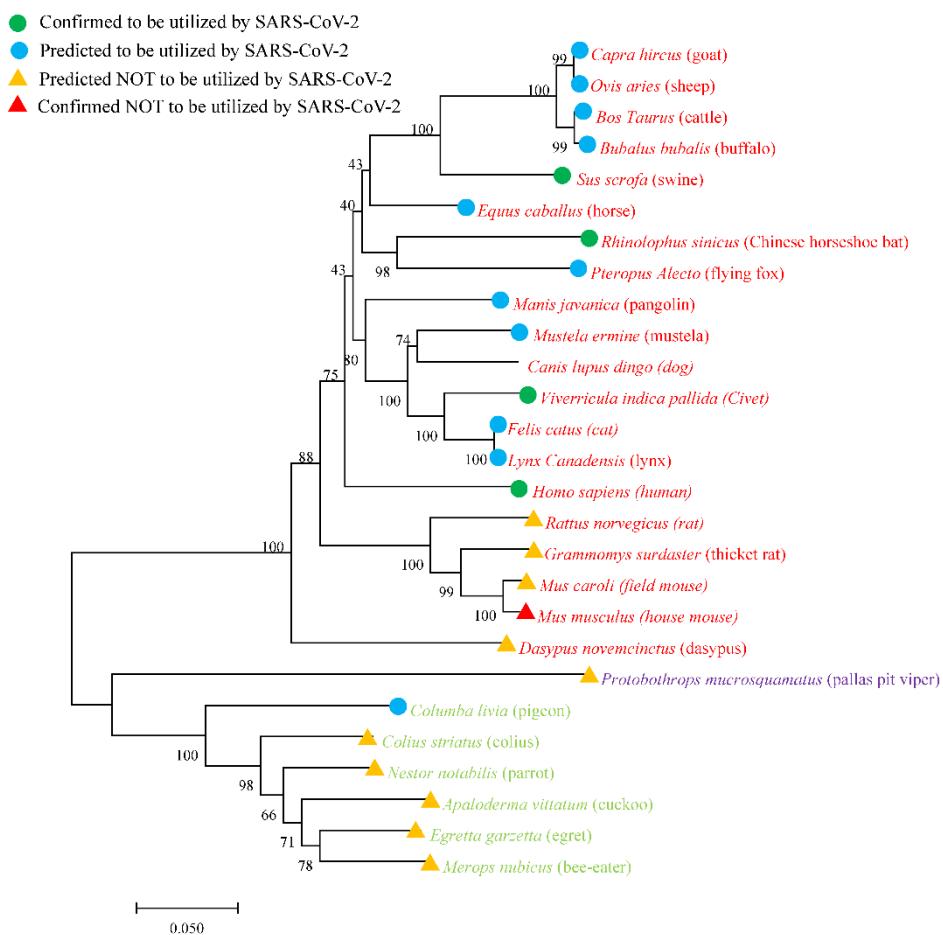


Figure 1. Part of the phylogenetic tree of ACE2s from selected species. The tree was constructed on the whole amino acid sequences of ACE2s by using NJ method with 1,000 bootstrap replicates. The utilizing capability of every ACE2 was marked with different labels as indicated. The ACE2 sequence of *Homo sapiens* (Q9BYF1), *Rhinolophus sinicus* (U5WHY8), *Paguma Larvata* (Q56NL1), *Sus scrofa* (K7GLM4) and *Mus musculus* (Q8R0I0) were downloaded from UniProt Knowledgebase. The rest were downloaded from GenBank as follows: *Capra hircus* (AHI85757.1), *Ovis aries* (XP_011961657.1), *Bos taurus* (XP_005228485.1), *Bubalus bubalis* (XP_006041602.1), *Equus caballus* (XP_001490241.1), *Pteropus alecto* (XP_006911709.1), *Manis javanica* (XP_017505752.1), *Mustela ermine* (XP_032187679.1), *Canis lupus dingo* (XP_025292934.1), *Felis catus* (AAX59005.1), *Lynx Canadensis* (XP_030160839.1),

Dasypus novemcinctus (XP_004449124.1), *Rattus norvegicus* (NP_001012006.1), *Grammomys surdaster* (XP_028617962.1), *Mus caroli* (XP_021009138.1), *Nestor notabilis* (XP_021009138.1), *Merops nubicus* (XP_010012481.1), *Egretta garzetta* (XP_009638257.1), *Apaloderma vittatum* (XP_009867056.1), *Colius striatus* (XP_009082150.1), *Columba livia* (PKK30539.1), *Rotobothrops mucrosquamatus* (XP_029140508.1).

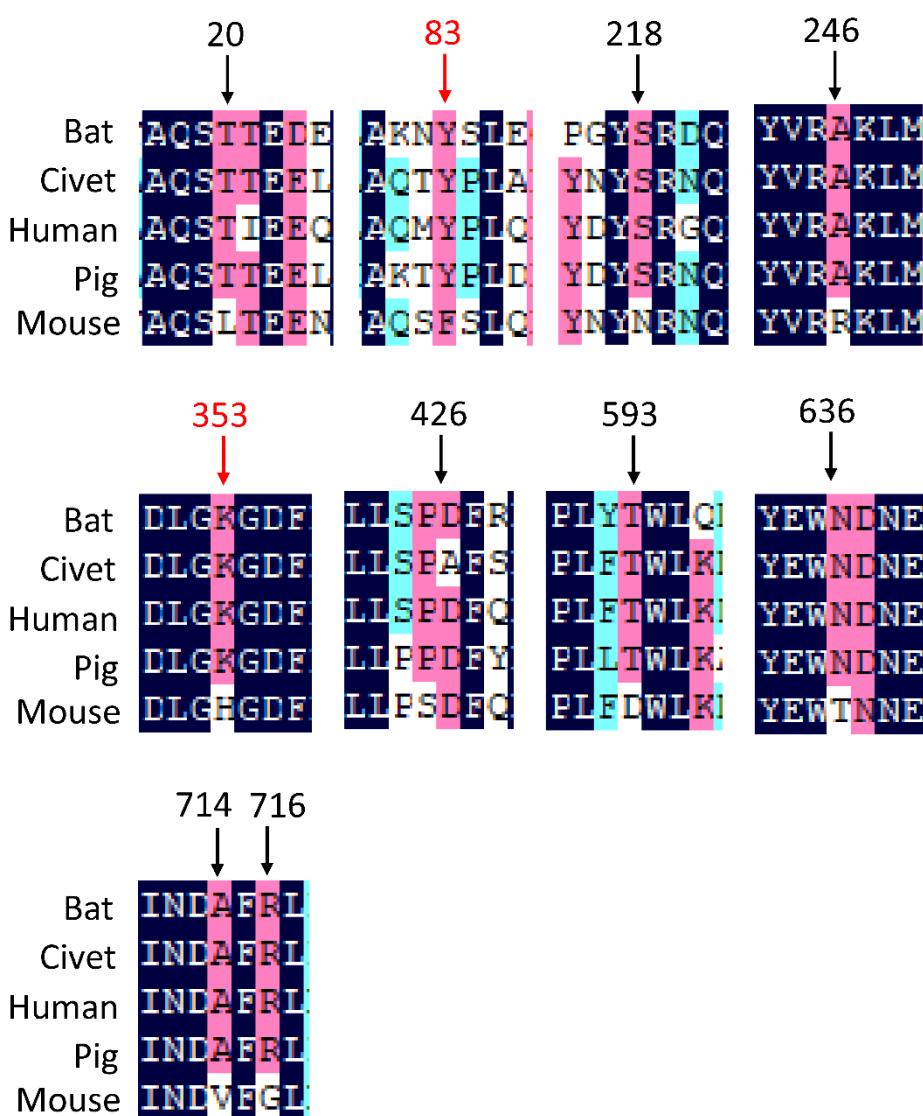


Figure 2. Critical amino acid sites predicted for the utilization by SARS-CoV-2. The extracellular domains of *Homo sapiens* (Human), *Rhinolophus sinicus* (Bat), *Paguma Larvata* (Civet), *Sus scrofa* (swine) and *Mus musculus* (mouse) ACE2s were aligned and

10 critical sites were picked. The critical sites reported in the previous articles are highlighted with red.

9. Supplementary materials

Figure S1. Phylogenetic tree of ACEs from 253 species. The tree was constructed on the whole amino acid sequences of ACEs by using NJ method with 1,000 bootstrap replicates. Different classes were labelled by different colors as indicated.

File S1. The list of all sequences analysed in this study.

File S2. Alignment of the amino acid sequences of human, mouse, swine, civet and Chinese horseshoe bat.

File S3. SARS-CoV-2 utilizing scores of all mammals and birds analysed in this study.