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.

Microbes and Infection 22 (2020) 221–225

Contents lists available at ScienceDirect

Microbes and Infection

journal homepage: www.elsevier.com/locate/micinf

Short communication

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

Ye Qiu [1], Yuan-Bo Zhao [1], 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

ARTICLE INFO

Article history:
Received 21 February 2020
Accepted 10 March 2020
Available online 19 March 2020

Keywords:
SARS-CoV-2
Coronavirus
Angiotensin converting enzyme 2 (ACE2)
Receptor utilization
Phylogenetic analysis

ABSTRACT

SARS-CoV-2, the newly identified human coronavirus causing severe pneumonia pandemic, 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. As a result, we predicted that SARS-CoV-2 tends to utilize ACE2s of various mammals, except murines, and some birds, such as pigeon. This prediction may help to screen the intermediate hosts of SARS-CoV-2.

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 pandemic of Coronavirus Disease 2019 (COVID-19) in China and worldwide, critically threatening the public health [1,2]. The pandemic has caused 80,813 confirmed infections with 3073 fatal cases in China by March 7th, 2020, surpassing any former coronavirus epidemics. Even worse, this virus emerges a worldwide spread currently and COVID-19 cases increased more rapidly than China in many countries around the world, including South Korean, Iran, Japan, Italy, Germany, France and United States. The interquartile deviation of the infection incubation period of SARS-CoV-2 is 2–7 days, but the maximum can be as long as 24 days according to the current reports, which is longer than other human coronaviruses and exacerbates the epidemic. The typical clinical symptoms of COVID-19 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. 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 labour-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

* Corresponding author.
E-mail address: xyge@hnu.edu.cn (X.-Y. Ge).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.micinf.2020.03.003
1286-4579/

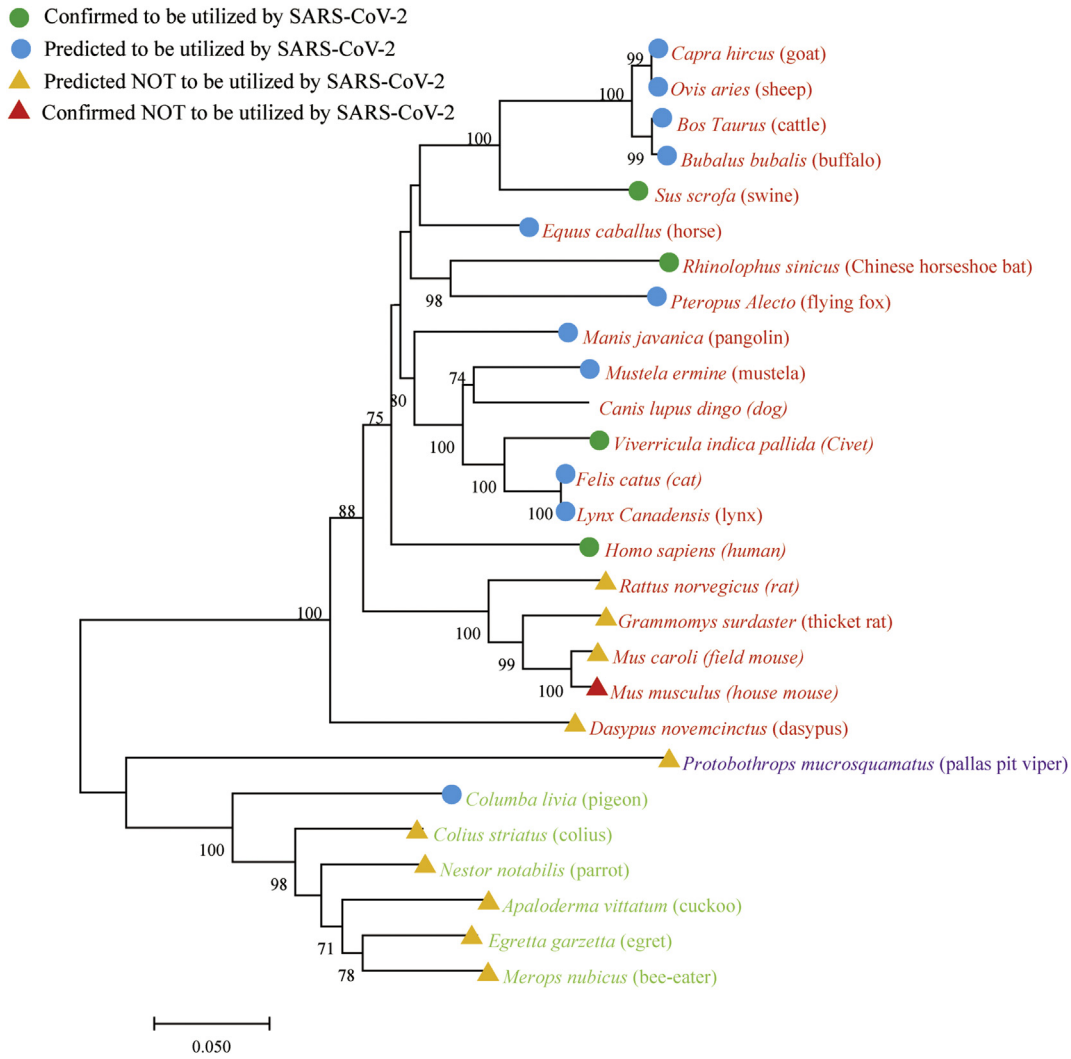

Fig. 1. Part of the phylogenetic tree of ACE2s from selected species. The tree was constructed on the whole aa sequences of ACE2s using NJ method by MEGA7 with 1000 bootstrap replicates. The utilizing capability of every ACE2 was marked with different labelled as indicated. The ACE2 sequence of *Homo 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).

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 (aa) sequence comparison. Unfortunately, the aa 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 aa 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 aa sequences of ACE2s of 253 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 aa 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 aa 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 living animals, we predicted pangolin, cat, cow, buffalo, goat, sheep and pigeon as the potential intermediate hosts for SARS-CoV-2, together with

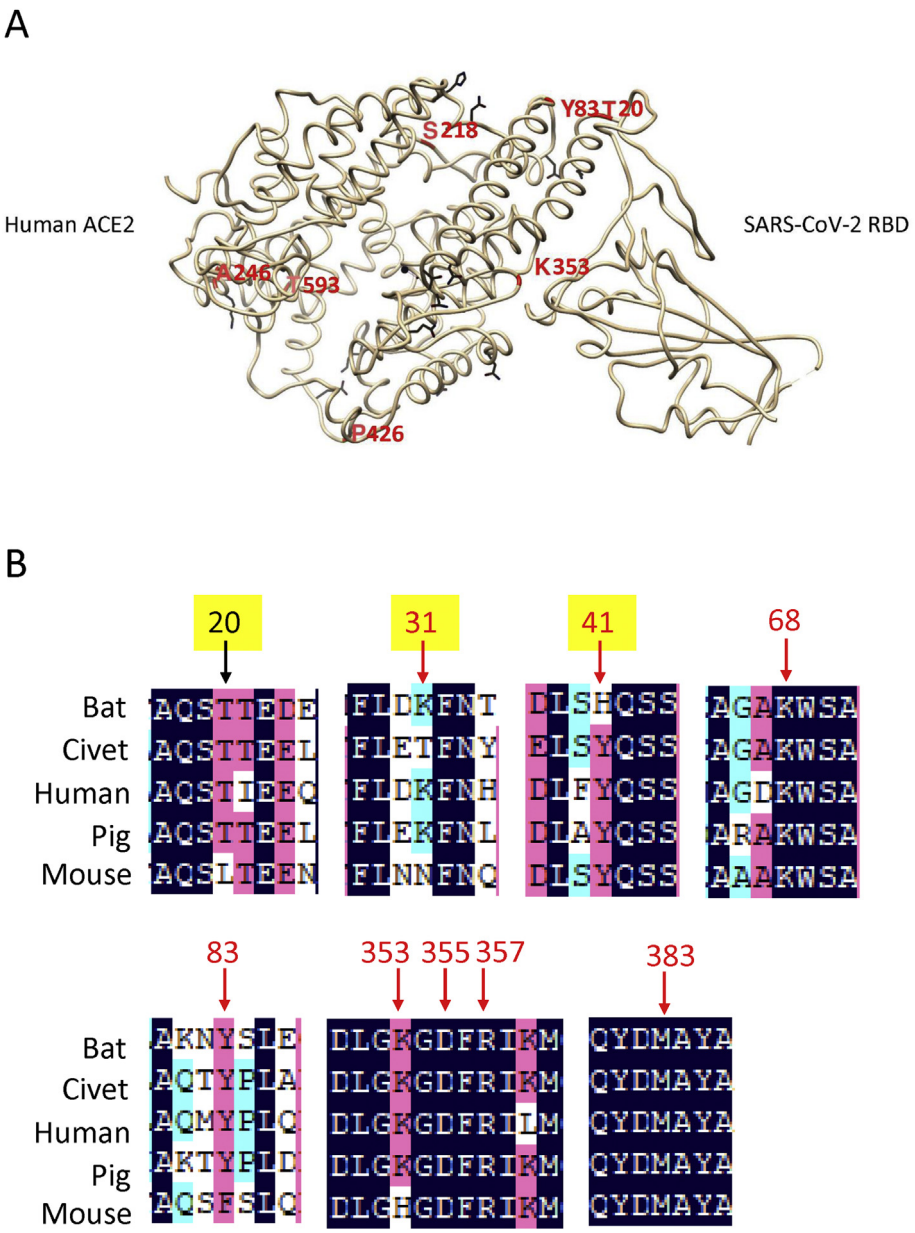

Fig. 2. Critical amino acid sites predicted for the utilization of SARS-CoV-2. (A) The structure of the binding complex of human ACE2 and SARS-CoV-2 RBD. The structure was adapted from Protein Data Bank (PDB ID: 6VW1) and the aa residues distinct in mouse ACE2 (T20, Y83, S218, A246, K353, P426 and T593) are labelled with red. N636, A714, R716 and A774 are not included in the structure, since they are too far away from the interface and were not crystallized in the structure analysis. (B) The aa sequences of *Homo sapiens* (Human), *Rhinolophus sinicus* (Bat), *Paguma Larvata* (Civet), *Sus scrofa* (swine) and *Mus musculus* (mouse) ACE2s were aligned. The critical sites identified by the alignment were highlighted with yellow background. The critical sites reported in SARS-CoV binding are highlighted with red font.

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.

1. Methods and materials

1.1. Phylogenetic analysis

The aa 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 aa sequences of ACEs using MAFFT with a local alignment strategy FFT-NS-2. The phylogenetic tree was constructed by MEGA7 using the neighbour-joining (NJ) method with 1000 bootstrap replicates and visualized using FigTree.

1.2. Marking of utilizing capability by SARS-CoV-2

The critical aa sites for the utilization by SARS-CoV-2 were selected by picking the unique aa residues in mouse ACE2 which distinct from human ACE2 and the sites reported to be critical for the receptor binding of SARS-CoV in the previous study [14]. The initial mark for every ACE2 was 100 and the substitution on each critical sites resulted in a reduction of its mark of 50 (for substitutions reported to abolish the receptor binding of SARS-CoV) or

Table 1
Marks for ACE2s of selected animals.

Species	20		31		41		68		83		353		355		357		383		SUM
Viverricula_indica_pallida (Civet)	T	0	T	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Felis_catus (cat)	T	0	K	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Homo_sapiens (human)	T	0	K	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Lynx_canadensis (lynx)	T	0	K	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Manis_javanica (pangolin)	T	0	K	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Mustela_erminea (mustela)	T	0	K	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Sus_scrofa (swine)	T	0	K	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Capra_hircus (goat)	T	0	K	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Ovis_aries (sheep)	T	0	K	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Pteropus_alecto (flying fox)	T	0	K	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Bos_taurus (cattle)	T	0	K	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Bubalus_bubalis (buffalo)	T	0	K	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Rhinolophus_sinicus (Chinese horseshoe bat)	T	0	K	0	H	0	K	0	Y	0	K	0	D	0	R	0	M	0	100
Canis_lupus_dingo (dog)	-	-10	K	0	Y	0	K	0	Y	0	K	0	D	0	R	0	M	0	90
Equus_caballus (horse)	T	0	K	0	H	0	R	-10	Y	0	K	0	D	0	R	0	M	0	90
Columba_livia (pigeon)	V	-10	E	-10	Y	0	K	0	-	-10	K	0	D	0	R	0	M	0	70
Merops_nubicus (bee-eater)	V	-10	E	-10	Y	0	K	0	-	-10	K	0	D	0	R	0	M	0	70
Egretta_garzetta (egret)	V	-10	E	-10	Y	0	K	0	-	-10	K	0	D	0	R	0	M	0	70
Apaloderma_vittatum (cuckoo)	V	-10	E	-10	Y	0	K	0	-	-10	K	0	D	0	R	0	M	0	70
Nestor_notabilis (parrot)	-	-10	-	-10	-	-10	-	-10	-	-10	K	0	D	0	R	0	M	0	50
Colius_striatus (colius)	-	-10	-	-10	-	-10	-	-10	-	-10	K	0	D	0	R	0	M	0	50
Dasypus_novemcinctus (dasypus)	-	-10	T	0	H	0	K	0	F	-50	K	0	D	0	R	0	M	0	40
Protobothrops_mucrosquamatus (pallas pit viper)	V	-10	Q	-10	Y	0	I	-10	F	-50	K	0	D	0	R	0	M	0	20
Grammomys_surdaster (thicket rat)	L	-10	K	0	Y	0	K	0	F	-50	H	-50	D	0	R	0	M	0	0
Mus_caroli (field mouse)	L	-10	K	0	Y	0	K	0	F	-50	H	-50	D	0	R	0	M	0	0
Rattus_norvegicus (rat)	L	-10	K	0	Y	0	K	0	F	-50	H	-50	D	0	R	0	M	0	0
Mus_musculus (house mouse)	L	-10	N	-10	Y	0	K	0	F	-50	H	-50	D	0	R	0	M	0	0

10 (for other substitutions). The utilizing capability by SARS-CoV-2 of the ACE2s were ranked by the final scores and ACE2s with scores >50 were considered to be potentially utilized by SARS-CoV-2.

2. Results and discussion

To evaluate the homology of ACE2s in different species, we built a phylogenetic tree haboring ACE2s of 253 vertebrate species (Fig. S1). The tree was branched to two major groups. The first group included the classes of Mammalia, Aves and Reprilia, and the second group included the classes of Amphibian, Chondrichthyes and Teleostomi. This tree indicated that besides mammalia, ACE2s of aves and reprilia were more likely to be utilized by SARS-CoV-2 than other vertebrates. As shown in Fig. 1, among mammals, human 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, 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.

We further tried to predict the receptor utilizing capability by SARS-CoV-2 via looking into the critical single aa polymorphism in different ACE2s. We screened the aa sites critical for SARS-CoV-2 utilization by two approaches. First, it has been proved that human, Chinese horseshoe bats, civet and swine ACE2s can be utilized by SARS-CoV-2 but mouse ACE2 cannot [6]. We aligned the aa sequence of these five ACE2s (File S2). Eleven aa residues distinct in mouse ACE2 but conserved in other ACE2s were identified, which were T20, Y83, S218, A246, K353, P426, T593, N636, A714, R716 and A774. However, according to the complex structure of human ACE2 and the receptor binding domain (RBD) in SARS-CoV-2 S protein (PDB ID: 6VW1), only three of the eleven sites are located near the interface and potentially mediate the binding, which were T20, Y83 and K353 (Fig. 2A). Second, SARS-CoV-2 shares an identity of 74.6% with SARS-CoV in the aa sequences of their RBDs, and thus the aa residues critical for the receptor binding of SARS-CoV may be also critical for SARS-CoV-2. According to the previous study, such critical sites include K31, Y41, K68, Y83, K353, D355, R357 and M383 [14]. Among these sites, K31D, Y41A, Y83F, K353 H/A/D, D355A and R357A abolish or strongly inhibit SARS-CoV binding, while the substitutions on K68 and M383 slightly inhibit SARS-CoV binding [14,15]. Combining the two screening, we got nine critical aa sites as shown in Fig. 2B. Based on these sites, we established a marking system to evaluate the utilizing capability by SARS-CoV-2 of different ACE2s. In the system, the initial score for every ACE2 was 100 and substitution on the critical sites would result in reduction of 50 marks (for K31D, Y41A, Y83F, K353 H/A/D, D355A and R357A) or 10 marks (for other substitutions). K31T and Y41H would not reduce the scores since these two substitutions exist in civet and Chinese horseshoe bat ACE2s, respectively (Fig. 2B). A final score less than 0 was adjusted to 0. As the result, a final score equal to or less than 50 means that the ACE2 contains either at least one substitution that will abolish the receptor binding or at least five other substitutions that will make it quite different from human ACE2. Therefore, 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. The final scores for all ACE2s we analysed were listed in Table S1. As shown in Table 1, excluding the animals unlikely contacted by people, 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 70 (Table 1). Though birds are not natural reservoirs for currently know species of betacoronavirus, the classification of coronaviruses is based on the genome feature not the host range, and thus we could not exclude the possibility that new betacoronavirus may infect birds. Considering pigeons are food in many places around the world, more attention should be paid on the non-mammalian hosts of SARS-CoV-2. ACE2 of *Protobothrops mucrosquamatus* (pallas pit viper), a common snake, got a score of 20, indicating that reprilia 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) (Fig. 1). The well matching with the phylogenetic tree elevated the liability of this marking 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 existing of these animals, 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 which cannot accurately reflect the infection in the animals, and more laboratory and epidemiological investigation are required to uncover the true intermediate hosts of SARS-CoV-2.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

This work was jointly funded by the National Natural Science Foundation of China (grant number 31470260 and 81902070), National Key Research and Development Program of China (grant number 2017YFD0500104), 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 2020SK3001) and the Fundamental Research Funds for the Central Universities in China (grant number 531107051162).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.micinf.2020.03.003.

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;382:727–33.
[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. https://doi.org/10.1038/s41586-020-2008-3 [Epub ahead of print].
[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;395:514–23.
[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;395:507–13.
[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;395:497–506.
[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 Mar;579(7798):270–3. https://doi.org/10.1038/s41586-020-2012-7.
[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. https://doi.org/10.1128/JVI.00127-20. pii: JVI.00127-20. [Epub ahead of print].