The Multiple Sequence Alignment and Evolutionary Analysis of Angiotensin-converting Enzyme 2 Diverse Species

Kexin Sun*
Suzhou Foreign Language School, Suzhou, Jiangsu 215000, China
*Corresponding author: kexin_sun@protonmail.com

Abstract. The sudden outbreak of new coronavirus (2019-nCoV) in Wuhan significantly influences in China, with wide attention attracted to investigate its origin, intermediate host, transmission pathway, and pathogenic mechanism. 2019-nCoV is very similar to SARS. They may use the similar receptor to recognize and attack host cell. Angiotensin-converting enzymes 2 (ACE2) is a main receptor of 2019-nCov and SARS, which interact with the surface spike glycoproteins (S proteins) of the coronaviruses. However, we still don’t know how the 2019-nCoV infected human and which species is the intermediate host. In this study we analyzed data of amino acid sequence of ACE2 and constructed phylogenetic tree. The phylogenetic tree showed ACE2 of human is highly similar to Pongo abelii and Pan troglodytes. Furthermore, our results suggest that Pongo abelii and Pan troglodytes may be another host and we must protect these animals from be infected. Our result is beneficial for a global pneumonia epidemic prevention.

Keywords: 2019-nCov, ACE2, Multiple Sequence Alignment, Phylogenetic Tree.

1. Introduction
Scientists discovered a new coronavirus in Wuhan, China in December 2019, which has caused a huge public health crisis, affecting the health of tens of thousands of people [1]. The new virus 2019-nCoV is very similar to SARS. In the clinical cases, 2019-nCoV infection will lead to fever and cough, followed by more and more serious lung infection and respiratory, thus endangering the life of the infected patients [2]. The latest research have elucidated the mechanism of virus recognition of the host cells, finding out the essential protein, angiotensin- converting enzymes 2 (ACE2), also very important in the process of SARS virus recognition [3-5]. Since the global outbreak of SARS in 2003, a large number of studies have shown the role of ACE2 on the host cell surface to be the receptor for SARS-CoV [6-9]. ACE2 which could interact with the surface spike glycoproteins (S proteins) of the coronaviruses, widely expressed in vivo, especially in kidney and testis, might play an essential role in the regulation of cardiovascular, renal function, and fertility with more than 600 complete coding sequences and about 3400 protein sequences for ACE2 searched on the NCBI [10-12]. Meanwhile, the three dimensional structure of full-length ACE2 has been elucidated recently [13, 14]. All those research will help with the foundation for further research.
What’s more, to explore the difference of ACE2 in diverse species is also an important task, which will contribute to assessing whether the other species could recognize those viruses and interact with the protein on the viruses. The most important thing is to find the intermediate host. The sequence decides the structure, which would influence the physiological function, so studying the protein sequence will provide the valuable information to better understand ACE2 [15].

In this study, I will compare amino acid sequences of ACE2 in different species and suggest possible intermediate host according to the sequence similarity and diversity. Starting from the protein sequence downloaded from NCBI, the multiple sequence alignment and phylogenetic tree building were employed to explore the conserved region and variable region, as well as the distance of protein sequence among the diverse species.

2. Method

Sequence data-set was downloaded from NCBI database searched by the keywords “angiotensin I converting enzyme 2”, including 600 reference protein sequences, the detailed species distribution showed in Table 1. In order to get the sequence similarity, ClustalW was employed for multiple sequence alignment, with all sites selected for calculation, and the other settings set to the default values [16]. Then, with the multiple sequence alignment results, MEGAX was used for phylogenetic tree building with Maximum Parsimony method, one of the criteria for choosing the best phylogenetic tree [17]. To raise the reliability of the phylogenetic tree, the resampling method was adopted with the bootstrap value set as 1000. The rough results showed more than 300 sequences redundant data and other ACE protein isoforms. Therefore, only 314 protein sequences for ACE2 were used for the final phylogenetic tree building and analysis. To present tree data clearly, the phylogenetic tree was visualized by the an online tool iTOL [18].

| Class       | Number of Species | Gene Name                          | Protein Name                          |
|-------------|-------------------|------------------------------------|---------------------------------------|
| Mammals     | 113               | angiotensin I converting enzyme 2  | angiotensin I converting enzyme 2     |
| Birds       | 54                | angiotensin I converting enzyme 2  | angiotensin I converting enzyme 2     |
| Lizards     | 7                 | angiotensin I converting enzyme 2  | angiotensin I converting enzyme 2     |
| Turtles     | 4                 | angiotensin I converting enzyme 2  | angiotensin I converting enzyme 2     |
| Insects     | 33                | angiotensin I converting enzyme 2  | angiotensin I converting enzyme 2     |
| Anthozoans  | 3                 | angiotensin I converting enzyme 2  | angiotensin I converting enzyme 2     |
| Molluscs    | 3                 | angiotensin I converting enzyme 2  | angiotensin I converting enzyme 2     |
| Bacteria    | 10                | angiotensin I converting enzyme 2  | angiotensin I converting enzyme 2     |

3. Results and Discussion

3.1. Collection of ACE2 Protein Sequences

The ACE2 protein sequence was downloaded from the NCBI database as angiotensin I converting enzyme 2 as keywords. As a result, 3472 protein sequences were obtained, of which 671 were well reviewed and formed the reference organisms. The species mainly includes mammals, birds, lizards, turtles, insects, anthozoans, molluscs, and bacteria, with the detailed information shown in Table 1.

Multiple sequence alignment of ACE2 protein. To identify the similarity and diversity of the 671 download ACE2 protein sequences, multiple sequence alignment was performed with ClustalW software. The results showed that the downloaded ACE2 protein sequences had high diversity because of existing the ACE1 proteins, ACE3 proteins, and other ACE-like isoforms. In Figure 1, the green clade represents ACE2 proteins, the others are redundant data showed in orange clade, ignored in the further analysis.
3.2. Phylogenetic Tree Analysis of ACE2 Protein

Phylogenetic tree has been widely used for evolutionary relationship analysis of biomolecules. To build a phylogenetic tree, a series of approaches have been developed, such as Neighboring Joining method, Maximum Likelihood method, Maximum Parsimony methods, Bayesian method, and so on. Here, we used the Maximum-Parsimony method to build the phylogenetic tree for the high efficiency. The partial results for primate clade showed in Figure 2, that the ACE2 of human had the highest similarity to Pongo abelii and Pan troglodytes with meanwhile Macaca nemestrina located at the furthest position with Homo sapiens. Since the protein sequence of ACE2 in Homo sapiens and Pongo abelii and Pan troglodytes are very similar, the 2019-nCov may also cause diseases in Pongo abelii and Pan troglodytes. Furthermore, these kinds of animals are most suitable for experiments because of the similar protein sequences. So the response to coronaviruses may be similar, meaning possible transmission among those species. The reliability of the branch is 0.979 which is very high and means this result is very reliable. Because the protein sequence of ACE2 in Macaca mulatta locates has higher difference from that of Homo sapiens, compared to Pongo abelii and Pan troglodytes, it is less possible that the virus would spread to human from Macaca mulatta.

![Figure 1. The Phylogenetic tree for all 671 ACE protein sequences.](image)

Furthermore, the epidemic situation caused panic, as the news reported that many families worried their pets (such as cats and dogs) may transmit the virus and killed their pets. Thus, we investigated the ACE2 similarity among Human, dog and cat. The results show that the distance of cats (Felis catus, marked in red frame) and dog (Canis lupus, marked in blue frame) is far away from Human in the tree (Figure 3). Therefore, it is a low possibility to transmit virus from pets to Human.
**Figure 2.** The Clade for ACE2 of the primates in a phylogenetic tree.

**Figure 3.** The clade for ACE2 of the other mammals far away from the primates clade.
4. Conclusion
As 2019-nCoV has spread all over the world, it is critical now to figure out all about of this coronavirus. ACE2 protein plays an essential role on virus invading the host making the analysis of ACE2 protein sequence helpful to provide information on the virus transmission among the species. In order to protect us from this virus, we must find out the potential host. In the study, we collect the data on protein sequence of ACE2 and build the phylogenetic tree to analyze the similarities and diversities among the different species which could be contributed to suggesting certain preventive measures. In particular, we found that protein sequence of ACE2 in Pongo abelii and Pan troglodytes is similar with human. So these animal may be infected by the 2019-nCoV. In the further, we should do more virus detection during them. Besides, the pets (such as cats and dogs) is a low possibility to be infected. Whereas, the ACE2 protein is common in different species and highly conservative. So 2019-nCoV may transmit between different species. Wildlife animals may be possible virus reservoir. We must protect susceptible population from these animals.

References
[1] Dong, Yuanyuan, et al. "Epidemiological characteristics of 2143 pediatric patients with 2019 coronavirus disease in China." Pediatrics (2020).
[2] World Health Organization. "Coronavirus disease 2019 (COVID-19): situation report, 72." (2020).
[3] Mehta, Puja, et al. “COVID-19: consider cytokine storm syndromes and immunosuppression.” The Lancet 395.10229 (2020): 1033-1034.
[4] Bai, Yan, et al. "Presumed asymptomatic carrier transmission of COVID-19." Jama 323.14 (2020): 1406-1407.
[5] Remuzzi, Andrea, and Giuseppe Remuzzi. "COVID-19 and Italy: what next?" The Lancet (2020).
[6] Du, Lanying, et al. "The spike protein of SARS-CoV—a target for vaccine and therapeutic development." Nature Reviews Microbiology 7.3 (2009): 226-236.
[7] Letko, Michael, Andrea Marzi, and Vincent Munster. "Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses." Nature microbiology 5.4 (2020): 562-569.
[8] Walls, Alexandra C., et al. "Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein." Cell (2020).
[9] Bartlam, Mark, Haitao Yang, and Zihe Rao. "Structural insights into SARS coronavirus proteins." Current opinion in structural biology 15.6 (2005): 664-672.
[10] Li, Rui, Songlin Qiao, and Gaiping Zhang. "Analysis of angiotensin-converting enzyme 2 (ACE2) from different species sheds some light on cross-species receptor usage of a novel coronavirus 2019-nCoV." Journal of Infection 80.4 (2020): 469-496.
[11] Ortega, Joseph Thomas, et al. "Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: An in silico analysis." EXCLI journal 19 (2020): 410.
[12] Mathewson, Alison C., et al. "Interaction of severe acute respiratory syndrome-coronavirus and NL63 coronavirus spike proteins with angiotensin converting enzyme-2." The Journal of general virology 89.Pt 11 (2008): 2741.
[13] Yan, Renhong, et al. "Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2." Science 367.6485 (2020): 1444-1448.
[14] Xiao, Fei, et al. "Evidence for gastrointestinal infection of SARS-CoV-2." Gastroenterology 158.6 (2020): 1831-1833.
[15] Cai, Guoshuai. "Bulk and single-cell transcriptomics identify tobacco-use disparity in lung gene expression of ACE2, the receptor of 2019-nCov." MedRxiv (2020).
[16] Katoh, Kazutaka, et al. "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform." Nucleic acids research 30.14 (2002): 3059-3066.
[17] Hall, Barry G. "Building phylogenetic trees from molecular data with MEGA." Molecular
biology and evolution 30.5 (2013): 1229-1235.

[18] Letunic, Ivica, and Peer Bork. "Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation." Bioinformatics 23.1 (2007): 127-128.