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.

Infection, Genetics and Evolution 82 (2020) 104285

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Letter to Editor

Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species

ARTICLE INFO

Keywords:
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
Genome
Genetic evolution
Structural protein
2019 Coronavirus disease (COVID-19)

ABSTRACT

The Corona Virus Disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a Public Health Emergency of International Concern. However, so far, there are still controversies about the source of the virus and its intermediate host. Here, we found the novel coronavirus was closely related to coronaviruses derived from five wild animals, including *Paguma larvata*, *Paradoxurus hermaphroditus*, Civet, *Aselliscus stoliczkanus* and *Rhinolophus sinicus*, and was in the same branch of the phylogenetic tree. However, genome and ORF1a homology show that the virus is not the same coronavirus as the coronavirus derived from these five animals, whereas the virus has the highest homology with Bat coronavirus isolate RaTG13.

To the Editor,

The Corona Virus Disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a Public Health Emergency of International Concern, posing a serious threat to people's health as well as the medical and health system (Association, 2020; Su et al., 2020). From December 12, 2019, when the first patient was admitted to hospital to March 5, 2020, China has reported a total of 80,585 confirmed cases and 3016 deaths caused by SARS-CoV-2 infection. To date, the disease has spread worldwide and become a serious infectious disease affecting human health worldwide. There are 17,637 confirmed cases and 338 deaths in more than 78 countries and regions outside China. With the increasing number of the cases, the World Health Organization (WHO) raised the assessment of the risk of spread and impact of COVID-19 to very high at a global level on February 28, 2020 (Association, 2020; Lai et al., 2020; Su et al., 2020).

To date, a large number of studies have proved the pathogen of COVID-19 is a novel coronavirus, which belongs to the *Coronavirus* family, *Betacoronavirus* genus and *Sarbecovirus* subgenus, with a linear single-stranded positive-strand RNA genome of about 30 kb (Ceraolo and Giorgi, 2020; Jiang and Shi, 2020; Lai et al., 2020; Li et al., 2020; Lu et al., 2020; Zhou et al., 2020). However, so far, there are still controversies about the source of the virus and its intermediate host. The evolutionary analysis showed that the coronavirus was the most similar to Bat coronavirus isolate RaTG13 (GenBank No.: MN996532), with 96.2% nucleotide homology in the whole genome (Jiang and Shi, 2020; Zhou et al., 2020). Other groups suggest that pangolin, mink, snake, turtle may be potential intermediate hosts for the virus (Association, 2020; Guo et al., 2020; Ji et al., 2020; Lam et al., 2020; Li et al., 2020; Liu et al., 2020a; Liu et al., 2020b; Xiao et al., 2020; Zhang et al., 2020a; Zhang et al., 2020b), which still need more evidence to be confirmed.

Here, we compared the sequences of SARS-CoV-2 and 39 coronavirus isolates from other species (Supplementary Table 1) using DNAMAN 6. As shown in Fig. 1, 41 isolates can be grouped into 4 clades, including genus *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus*. Among which, two isolates of SARS-CoV-2 belong to *Betacoronavirus*, with a close relationship to five other isolates (GenBank Nos.: AY515512, AY613948, AY572038, KY417142, and KY417150) derived from five wild animals, *Paguma larvata*, *Paradoxurus hermaphroditus*, Civet, *Aselliscus stoliczkanus* and *Rhinolophus sinicus*, forming the same branch of the phylogenetic tree. Further analysis showed the homology of the whole genome of the seven viruses was 89.07%. Among them, SARS-CoV-2 (NC_045512) has about 73.81, 73.60, 73.58, 74.58 and 71.98% homology with other isolates (except the other SARS-CoV-2 isolate), respectively. Additionally, we previously found mink coronavirus had low similarity with SARS-CoV-2 in the whole genome, ORF1a, ORF1ab, and S proteins, with 68, 46.96, 52.44 and 47.36% identities, respectively, subgrouping into distinct clusters (Yang et al., 2020). Thus, since the homology of SARS-CoV-2 and these coronaviruses is less than 75%, it is presumed that SARS-CoV-2 is not the same virus as the coronaviruses derived from these five wild animals.

To further verify this speculation, three non-human coronaviruses and three suspected host-derived coronaviruses were selected for homology analysis. As shown in Table 1, SARS-CoV-2 and bat coronavirus RaTG13 have the highest homology in the whole genome, ORF1ab, nucleocapsid protein (N), and spike protein (S) genes. Furthermore, the amino acid homologies of ORF1ab, N, S proteins of the two viruses are 98.55, 99.05 and 97.41% respectively. These results suggest the two viruses have a high genetic relationship, which is consistent with the results reported previously (Jiang and Shi, 2020; Li et al., 2020; Paraskevis et al., 2020; Zhou et al., 2020). Moreover, the N protein of SARS-CoV-2 and bat coronavirus RaTG13 had 4 different amino acids, which were 37S/P, 215G/S, 243G/S, and 267A/Q, respectively. The S protein of the two strains has 33 different amino acids (2.59%), with larger differences are located at 439–449 and 482–505 of the S protein, respectively. Besides, it was reported that the SARS-CoV-2 virus has a unique peptide (PRRA) insertion, which may be involved in the proteolytic cleavage of the S protein by cellular proteases, and impact host range and transmissibility (Li et al., 2020). The PRRA motif is located at the 680 of the SARS-CoV-2 S protein, but not in the S protein of the bat coronavirus RaTG13. 103 different amino acid sites (1.45%) were found in the ORF1ab protein of the two strains, with

https://doi.org/10.1016/j.meegid.2020.104285
Received 6 March 2020; Received in revised form 7 March 2020; Accepted 10 March 2020
Available online 10 March 2020

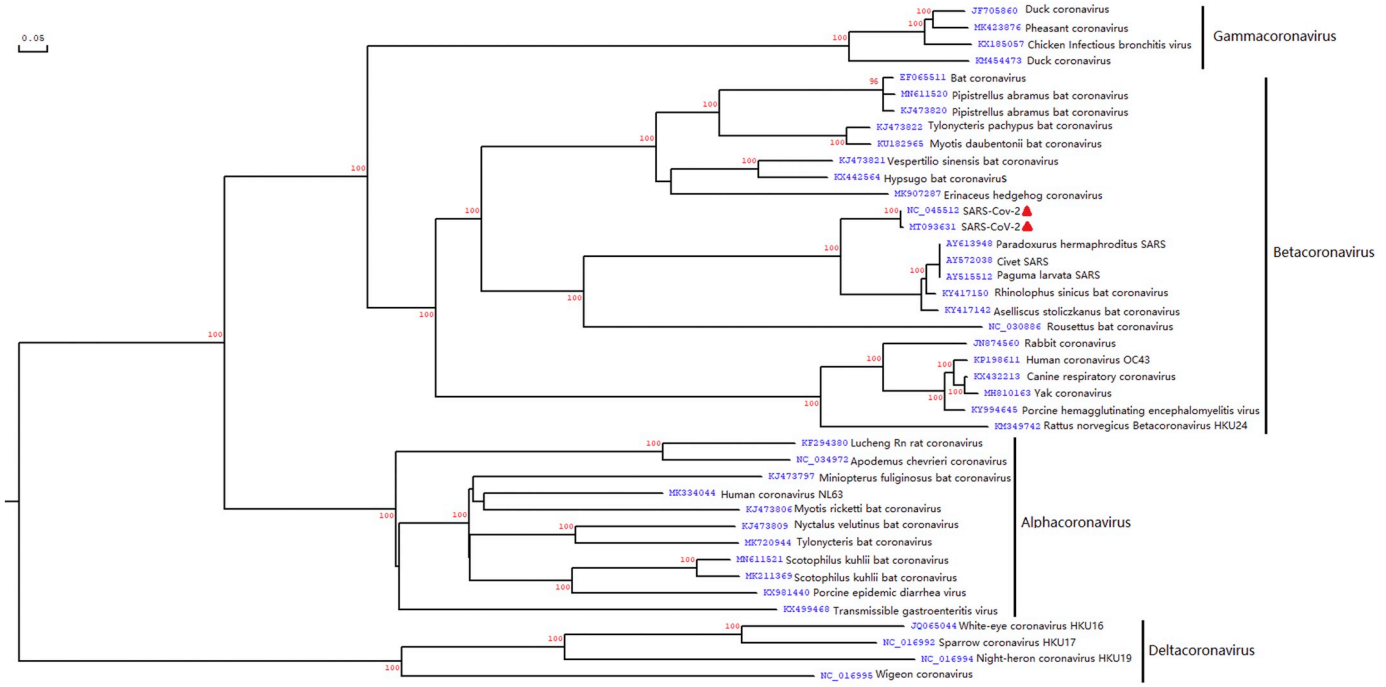

Fig. 1. Phylogenetic analysis of coronaviruses from different species and 2019 novel coronavirus in China. The 41 complete genome sequences of coronavirus were analyzed using MEGA 5.0. The phylogenetic tree was constructed using maximum-likelihood (ML) with a bootstrap of 100. Red triangle, SARS-CoV-2 isolate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Homologous analysis of SARS-CoV-2 (NC_045512) and six other Coronavirus strains isolated from different hosts in China (%).

Isolate	Host	Complete genome	ORF1ab	N	S
SARS coronavirus civet020(AY572038)	Civet	73.58	79.23	87.79	71.41
Bat SARS-like coronavirus As6526 (KY417142)	*Aselliscus stoliczkanus*	74.58	79.23	87.55	68.17
Bat SARS-like coronavirus Rs4874(KY417150)	Rhinolophus sinicus	71.98	79.18	87.94	71.29
Alphacoronavirus Mink/China/1/2016(MF113046)	Mink	34.97	38.47	33.70	30.89
Bat coronavirus isolate RaTG13(MN996532)	*Rhinolophus affinis*	93.7	96.5	96.9	92.86
Pangolin coronavirus(MT084071)	*Manis javanica*	?	?	95	90

Note: N, N protein. S, spike protein. ?, Sequence of Pangolin coronavirus (MT084071) is not completed in this part of the genome, there are many gaps needed to be filled.

larger differences are located at 919–1227. On the contrary, the homology between SARS-CoV-2 and coronaviruses from the other five species (except pangolin coronavirus) is less than 90% (Table 1). These results further indicate that the two viruses, SARS-CoV-2 and bat coronavirus RaTG13, are closely related. Interestingly, the homology of the whole genome, ORF1ab, N, and S genes of SARS-CoV-2 and another Bat SARS-like coronavirus Rs4874 isolated from *R. sinicus* were less than 90% (Table 1), indicating the coronavirus carried by different bat individuals may be different.

Moreover, SARS-CoV-2 and pangolin coronavirus is also highly related (Lam et al., 2020; Xiao et al., 2020; Zhang et al., 2020b), but due to incomplete sequence of pangolin coronavirus published in GenBank, reasonable analysis cannot be carried out in this study. Besides, it was reported that the SARS-CoV-2 virus did not come directly from pangolins (Li et al., 2020). Therefore, the relationship between SARS-CoV-2 and pangolin coronavirus and whether pangolin is the intermediate host of SARS-CoV-2 need further investigation.

Notably, although the whole genome sequence of SARS-CoV-2 is almost identical, with a similarity rate of 99.9% (Ceraolo and Giorgi, 2020; Li et al., 2020; Lu et al., 2020; Paraskevis et al., 2020), the virus begins to mutate in patients (Zhao et al., 2020). Mutations of different SARS-CoV-2 isolates mainly occurs in five genes, including S, N, ORF8, ORF3a, and ORF1ab, with about 42% of the variations are non-synonymous mutation (Zhao et al., 2020). An increased level of viral diversity was found in some SARS-CoV-2 infected patients (Shen et al., 2020), suggesting that the virus has begun to adapt to the human environment and its genomes have begun to evolve in the population. Therefore, in-depth detection and tracking of the mutation and evolution of the virus in the population, as well as a cross-species transmission mechanism, are urgent tasks to be carried out at present.

In summary, based on these results, we have some thoughts on the research of the SARS-CoV-2 in the future: (1) the SARS-CoV-2 is closely related to bat coronavirus RaTG13 but is far from coronavirus isolated from other species analyzed in this study. However, is *R. affinis* the storage host of the SARS-CoV-2? (2) If SARS-CoV-2 evolved from *R. affinis*, whether there are other host involved in the mutation of the virus; (3) If SARS-CoV-2 does not from *R. affinis*, the animal may also serve as a key intermediate host for the recombination and evolution of SARS-CoV-2; (4) due to the limitation of data, whether there are other sources of SARS-CoV-2 is worth further investigation and analysis.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2020.104285.

Author contributions

Writing-original draft, L.R; writing-review and editing, Y.Y.; figures editing, C.L.; revising and supervising and funding acquisition, L.R.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was financially supported by the National Key Research and Development Program of China (No. 2017YFD0500103), the National Natural Science Foundation of China (No. 31772747), and the Science and Technology Research Program during the 13th Five-Year Plan Period of Jilin Educational Committee (No. JJKH20190172KJ).

References

Association, S.E.G.f.C.o.t.E.o.N.C.P.o.t.C.P.M., 2020. An update on the epidemiological characteristics of novel coronavirus pneumonia(COVID-19). Zhonghua Liu Xing Bing Xue Za Zhi 41, 139–144. https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.002.

Ceraolo, C., Giorgi, F.M., 2020. Genomic variance of the 2019-nCoV coronavirus. J. Med. Virol. https://doi.org/10.1002/jmv.25700.

Guo, Q., Li, M., Wang, C., Wang, P., Fang, Z., Tan, J., Wu, S., Xiao, Y., Zhu, H., 2020. Host and infectivity prediction of Wuhan 2019 novel coronavirus using deep learning algorithm. bioRxiv. https://doi.org/10.1101/2020.01.21.914044.

Ji, W., Wang, W., Zhao, X., Zai, J., Li, X., 2020. Cross-species transmission of the newly identified coronavirus 2019-nCoV. J. Med. Virol. 92, 433–440. https://doi.org/10.1002/jmv.25682.

Jiang, S., Shi, Z.L., 2020. The first disease X is caused by a highly transmissible acute respiratory syndrome coronavirus. Virol. Sin. https://doi.org/10.1007/s12250-020-00206-5.

Lai, C.C., Shih, T.P., Ko, W.C., Tang, H.J., Hsueh, P.R., 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. Int. J. Antimicrob. Agents 105924. https://doi.org/10.1016/j.ijantimicag.2020.105924.

Lam, T.T.-Y., Shum, M.H.-H., Zhu, H.-C., Tong, Y.-G., Ni, X.-B., Liao, Y.-S., Wei, W., Cheung, W.Y.-M., Li, W.-J., Li, L.-F., Leung, G.M., Holmes, E.C., Hu, Y.-L., Guan, Y., 2020. Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. bioRxiv. https://doi.org/10.1101/2020.02.13.945485.

Li, X., Zai, J., Zhao, Q., Nie, Q., Li, Y., Foley, B.T., Chaillon, A., 2020. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. J. Med. Virol. https://doi.org/10.1002/jmv.25731.

Liu, P., Jiang, J.-Z., Wan, X.-F., Hua, Y., Wang, X., Hou, F., Chen, J., Zou, J., Chen, J., 2020a. Are pangolins the intermediate host of the 2019 novel coronavirus (2019-nCoV)? bioRxiv. https://doi.org/10.1101/2020.02.18.954628.

Liu, Z., Xiao, X., Wei, X., Li, J., Yang, J., Tan, H., Zhu, J., Zhang, Q., Wu, J., Liu, L., 2020b. Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. J. Med. Virol. https://doi.org/10.1002/jmv.25726.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J., Meng, Y., Wang, J., Lin, Y., Yuan, J., Xie, Z., Ma, J., Liu, W.J., Wang, D., Xu, W., Holmes, E.C., Gao, G.F., Wu, G., Chen, W., Shi, W., Tan, W., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395, 565–574. https://doi.org/10.1016/S0140-6736(20)30251-8.

Paraskevis, D., Kostaki, E.G., Magiorkinis, G., Panayiotakopoulos, G., Sourvinos, G., Tsiodras, S., 2020. Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. Infect. Genet. Evol. 79, 104212. https://doi.org/10.1016/j.meegid.2020.104212.

Shen, Z., Xiao, Y., Kang, L., Ma, W., Shi, L., Zhang, L., Zhou, Z., Yang, J., Zhong, J., Yang, D., Guo, L., Zhang, G., Li, H., Xu, Y., Chen, M., Gao, Z., Wang, J., Ren, L., Li, M., 2020. Genomic diversity of SARS-CoV-2 in coronavirus disease 2019 patients. Clin. Infect. Dis. https://doi.org/10.1093/cid/ciaa203.

Su, S., Li, X., Hao, H., Wang, X.-Y., Zhang, M.-M., Geng, H., Ma, M., 2020. Advances in research on SARS-CoV-2. J. Xi'an Jiaotong Univ.(Medical Sciences) 41 (preprint), 1–7. http://kns.cnki.net/kcms/detail/61.1399.R.20200224.0944.010.html.

Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X., Zou, J.-J., Li, N., Guo, Y., Li, X., Shen, X., Zhang, Z., Shu, F., Huang, W., Li, Y., Zhang, Z., Chen, R.-A., Wu, Y.-J., Peng, S.-M., Huang, M., Xie, W.-J., Cai, Q.-H., Hou, F.-H., Liu, Y., Chen, W., Xiao, L., Shen, Y., 2020. Isolation and characterization of 2019-nCoV-like coronavirus from Malayan pangolins. bioRxiv. https://doi.org/10.1101/2020.02.17.951335.

Yang, Y., Chen, S., Ren, L., 2020. Genetic evolution analysis of Marten coronavirus and 2019 novel coronavirus. Chin. J. Anim. Health Inspection 37 Accepted.

Zhang, C., Zheng, W., Huang, X., Bell, E.W., Zhou, X., Zhang, Y., 2020a. Protein structure and sequence re-analysis of 2019-nCoV genome does not indicate snakes as its intermediate host or the unique similarity between its spike protein insertions and HIV-1. bioRxiv. https://doi.org/10.1101/2020.02.04.933135.

Zhang, T., Wu, Q., Zhang, Z., 2020b. Pangolin homology associated with 2019-nCoV. bioRxiv. https://doi.org/10.1101/2020.02.19.950253.

Zhao, W., Song, S., Chen, M., Zou, D., Ma, L., Ma, Y., Li, R., Hao, L., Li, C., Tian, D., Tang, B., Wang, Y., Zhu, J., Chen, H., Zhang, Z., Xue, Y., Bao, Y., 2020. The 2019 novel coronavirus resource. Hereditas 42, 212–221. https://doi.org/10.16288/j.yczz.20-030.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., Chen, J., Luo, Y., Guo, H., Jiang, R.D., Liu, M.Q., Chen, Y., Shen, X.R., Wang, X., Zheng, X.S., Zhao, K., Chen, Q.J., Deng, F., Liu, L.L., Yan, B., Zhan, F.X., Wang, Y.Y., Xiao, G.F., Shi, Z.L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. https://doi.org/10.1038/s41586-020-2012-7.

Chun Li[c], Yanling Yang[b,*], Linzhu Ren[a,*]

[a] *College of Animal Sciences, Jilin University, Changchun 130062, China*

[b] *Institute of Special Wild Economic Animals and Plants, Chinese Academy of Agricultural Sciences, Changchun 130112, China*

[c] *The Chinese Peoples' Liberation Army 964 Hospital, Changchun, China*

E-mail addresses: m18043213639@163.com (Y. Yang), renlz@jlu.edu.cn (L. Ren).

* Corresponding authors.