

NEWS AND VIEWS

SARS-like virus in the Middle East: A truly bat-related coronavirus causing human diseases

Guangwen Lu¹✉, Di Liu²

¹ CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

² Network Information Center, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

✉ Correspondence: luguangwen2001@126.com

On September 23rd, the United Kingdom Health Protection Agency (HPA) reported the diagnostic confirmation of a human infection case by a new type of coronavirus (<http://www.hpa.org.uk/NewsCentre/NationalPressReleases/2012PressReleases/120923acutererespiratoryillnessidentified/>). The patient is a 49-year-old Qatari who suffered from an acute, serious respiratory illness. The record shows that the man had been in a travel visit to the country of Saudi Arabia, where an Arabic victim of a severe acute respiratory illness caused by a coronavirus was reported several months ago (<http://www.promedmail.org/direct.php?id=20120920.1302733>). Laboratory tests showed that the identified viruses from these two patients are genetically the same (http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1317136202755).

The infection was reported to first manifest clinically with fever, cough, shortness of breath and breathing difficulties, and then combine with an acute renal failure. The Arabic patient died subsequently, and the Qatari is currently under medical treatment in a London hospital. These symptoms are very similar to those caused by the SARS coronavirus (SARS-CoV), which infected over 8000 people with more than 800 death worldwide in the 2003 pandemic (http://www.who.int/csr/sars/country/2003_08_15/en/index.html). As people are worrying about the reemergence of SARS, two cases of novel coronavirus infection instantly draw worldwide attention as to potential pandemics in future.

Coronavirus belongs to the family *Coronaviridae* within the order *Nidovirales* (Lai et al., 2007) and is named as by the featured morphological description of a wide variety of viruses with microscopic crown-like appearances. Coronaviruses usually have a circular structure of 100–160 nm in diameter with spikes of bulbous shape in the distal end deco-

rating the viral surface. Although an electronic microscopic picture of the newly identified coronavirus is not available currently, the genetic features of the new virus have been successfully characterized. The whole genome of the new coronavirus identified in the first infection case has been successfully sequenced by a group of Netherlandish researchers. The viral strain was named HCoV-EMC, and its genomic sequence was made publicly available in the GeneBank (accession code: JX869059) shortly after the confirmation of the second infection case. The determined HCoV-EMC genome is of 30118 bases, which is consistent with the normal size of a typical coronavirus genome ranging from 26 to 32 kb in size.

The genomic structure of HCoV-EMC has also been determined, which involves a short leader sequence, followed by a 5' untranslated region (5'-UTR), a protein coding region of ten putative open reading frames (ORFs), a 3'-UTR and a short poly(A) tail. A characteristic gene order of 5'-replicase ORF1ab, spike (S), envelope (E), membrane (M), and nucleocapsid (N)-3' is observed in the new virus. In coronaviruses, these genes could be further interspersed with ORFs encoding some group specific nonstructural and/or structural proteins (eg. the HE glycoprotein). In HCoV-EMC, four extra ORFs encoding nonstructural proteins NS3a-3d are observed between S and E genes. This gene organization is very similar to those reported for bat-derived coronaviruses of the 2c subgroup (e.g. the Bat-CoV HKU4 and HKU5 strains). For the replicase part, about two thirds of the genome from the 5' end encodes two polyprotein precursors 1a and 1b, which would be further processed into individual non-structural proteins (Nsps). Two putative proteases are identified in the new coronavirus, including Nsp3 and Nsp5. The former contains a putative papain-like protease domain, while the latter

resembles the chymotrypsin-like protease (or the 3C-like main protease, 3CL Mpro).

Serologically, coronaviruses can be categorized into three distinct antigenic groups (group 1, 2, and 3) (Lai et al., 2007). Several subgroups are further defined based on the phylogenetic characters of different viral genes. E.g. four subgroups 2a–2d are thus far identified in the coronavirus group 2 (Woo et al., 2007). In this classification system, SARS-CoV, which was initially believed to represent a fourth antigenic group but was later designated as a distant group 2 member, belongs to the subgroup of 2b. For the new coronavirus, a recent phylogenetic study using a partial polymerase (Nsp12) sequence from “Lonedon1_novel CoV 2012” (a tentative name for the viral strain identified in the second infection case) reveals that the new virus belongs to subgroup 2c. With the

full genomic sequence of HCoV-EMC, a more comprehensive phylogenetic analysis is now available. As shown in Fig. 1, different viral proteins, including three non-structural proteins of the putative 3CL Mpro, RNA dependent RNA polymerase (RDRP) and Helicase and two structural proteins of the spike and nucleocapsid, are involved in the *in silico* assay. Five trees are thereby constructed using the amino acid sequence. A remarkable phylogenetic “closeness” is observed between HCoV-EMC and other subgroup 2c members of the Bat-CoV HKU4 and HKU5 strains. In all five trees, the HKU4 strains are clustered with another bat-derived coronavirus strain of BtCoV/133/2005 and the HKU5 strains are clustered separately, which is consistent with a previous report (Woo et al., 2007). HCoV-EMC forms an extra sublineage and is more closely related to the HKU5 cluster than to the HKU4 cluster.

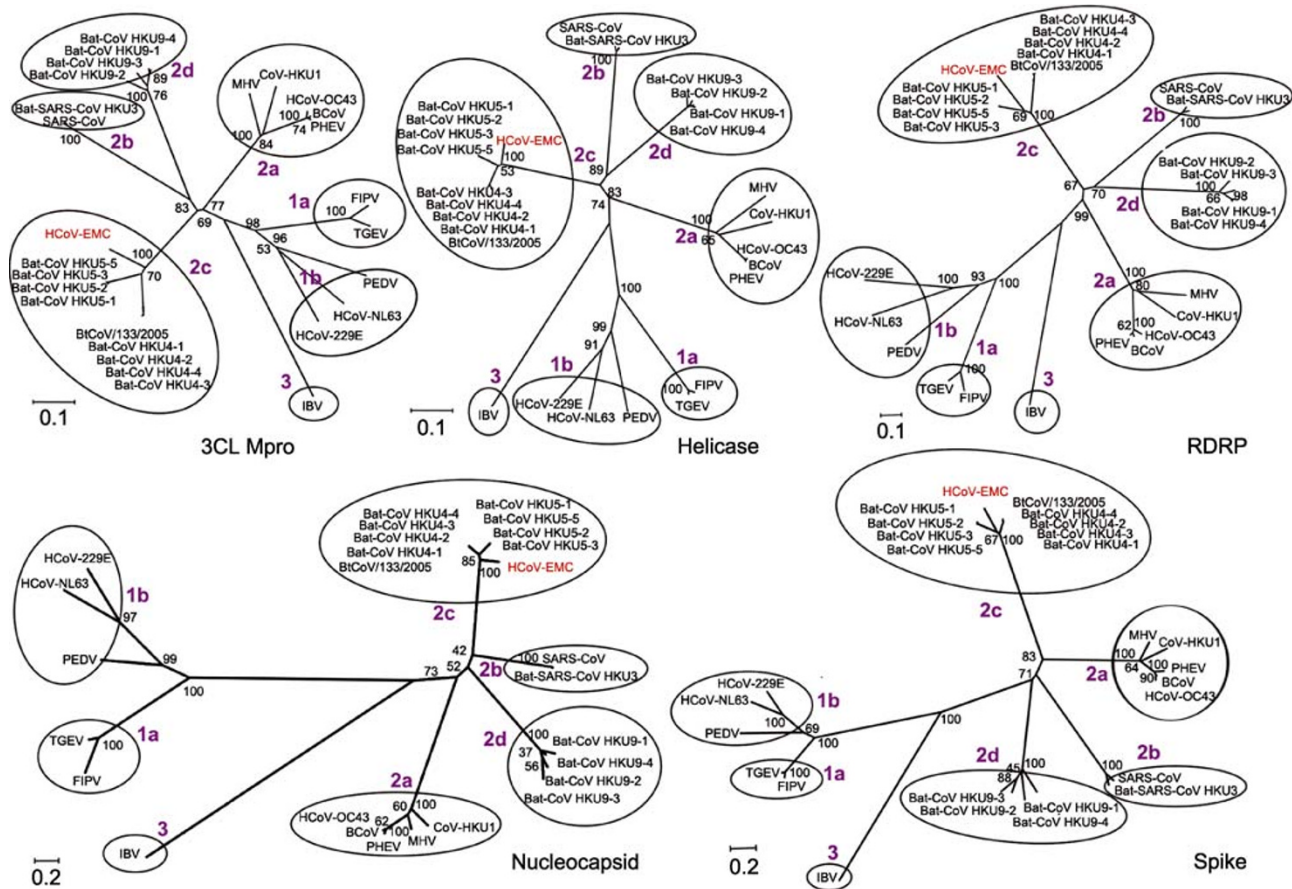


Figure 1. Phylogenetic analysis of the 3C-like main protease (3CL Mpro), Helicase, RNA-dependent RNA polymerase (RDRP), Spike and Nucleocapsid of HCoV-EMC. The trees are constructed by the neighbor-joining method using the bootstrap values. The deduced full amino acid sequence of the individual protein is used in the calculation. The subgroups are marked with magenta labels. Abbreviations and accession codes: HCoV-EMC (JX869059); Bat-CoV HKU, bat coronavirus HKU (4-1–4-4, 5-1–5-5, 9-1–9-5: EF065505 to EF065516); BtCoV/133/2005 (DQ648794); SARS-CoV (NC_004718); Bat-SARS-CoV HKU3, bat SARS coronavirus HKU3-1 (DQ022305); CoV-HKU1, human coronavirus HKU1 (NC_006577); MHV, murine hepatitis virus (NC_006852); PHEV, porcine hemagglutinating encephalomyelitis virus (NC_007732); BCoV, bovine coronavirus (NC_003045); HCoV-OC43, human coronavirus OC43 (NC_005147); IBV, avian infectious bronchitis virus (NC_001451); FIPV, feline infectious peritonitis virus (AY994055); TGEV (DQ811789); PEDV, porcine epidemic diarrhea virus (NC_003436); HCoV-229E, human coronavirus 229E (NC_002645); HCoV-NL63, human coronavirus NL63 (NC_005831).

We noted that all the subgroup 2c coronaviruses identified so far are derived from bats. This indicates that HCoV-EMC, as a 2c member, is likely of bat-origin. It is also noteworthy that the new coronavirus is only distantly related to SARS-CoV, which clearly separates the recent infection cases from SARS resurgence.

The severe respiratory illness clinically observed in the two laboratory confirmed cases makes the infection similar to SARS. But the new coronavirus does not genetically resemble SARS-CoV. Nevertheless, these two viruses, both as a group 2 coronavirus member, should share common mechanisms in the transcription and replication strategies. In this sense, it is likely that some inhibitors designed for SARS-CoV would also be applicable to the new virus. A most promising target would be the 3CL Mpro enzyme which processes the replicative pp1a and 1b precursors into functional protein subunits. The proteases from SARS-CoV and the new coronavirus exhibit over 50% identity in sequence. Similar structural and enzymatic features as observed for SARS-CoV 3CL Mpro could therefore be expected for the protease of the new virus. E.g. the protease should contain three domains; the N-terminal two domains would form a chymotrypsin-like fold; an extra C-terminal domain would facilitate the homodimerization of the protease which is indispensable for the catalytic activity, etc (Yang et al., 2003). After broad inhibitor studies worldwide targeting SARS-CoV 3CL Mpro, a large number of effective compounds have been discovered, which would benefit clinical treatment of the new coronavirus infection patients. The fusion process is also predicted to be similar between SARS-CoV and the new coronavirus. The studying strategy for the heptad repeat regions of SARS-CoV spike (Zhu et al., 2004) should easily be used in the design of short peptide fusion inhibitors targeting the new coronavirus.

There are neither prophylactic vaccines to prevent nor effective clinical methods to treat the new virus infection by far. We believe the treatment experiences with SARS patients might be referential, despite that a consensus on an optimal treatment regimen for SARS has not been reached yet. Further information are needed regarding to the immune response and the levels of proinflammatory cytokines in the patients infected by the new virus. Another public concern over the new virus is how contagious it is? There has not been a definitive answer yet. Normally, coronaviruses are aerosol transmissible. Nevertheless, coronaviruses are fairly fragile and easily destroyed by usual detergents and cleaning agents outside of the body. Transmission of the new coronavirus appears to be rather limited, as only two infection cases are currently confirmed globally. Those people having close contacts with and the medical personnel caring for these two patients are not infected by the same virus.

As to the diagnostic assays, a detection method using the real-time reverse-transcription polymerase chain reaction has

been developed (Corman et al., 2012). In addition, the World Health Organization (WHO) has already issued an interim case definition to help countries strengthen health protection measures against the new virus (http://www.who.int/csr/disease/coronavirus_infections/case_definition/en/index.htm). With global efforts and facilitated by the availability of the full genomic sequence of the new virus, more sensitive and specific diagnostic methods could be expected in the near future.

By far, little is known about the pathogenesis and virulence of the new virus. Identification of the cellular receptor for the virus would be of vital importance in understanding its pathogenesis. Although it cannot be excluded, it is unlikely that the new coronavirus also utilize human ACE2 as the entry receptor, as its spike diverges dramatically from that of SARS-CoV in sequence (less than 27% identity). Further efforts are needed to settle the issue. What is of equal importance is the mechanism underlying the viral cross-species transmission. The phylogenetic analysis reveals a possible bat origin of the new virus. If this is the truth, how the virus gains the ability to infect human is an urgent issue requiring instant studies.

Before SARS outbreak, a limited number of coronaviruses are known to be circulating in humans, causing only mild illness such as common cold. Following the 2003 SARS pandemic, people start to realize that coronaviruses, once crossing the species barrier, could be life-threatening. Further attentions are therefore needed to be paid to the new coronavirus, although it does not seem to be very contagious.

REFERENCES

- Corman, V., Eckerle, I., Bleicker, T., Zaki, A., Landt, O., Eschbach-Bludau, M., van Boheemen, S., Gopal, R., Ballhause, M., Bestebroer, T., et al. (2012). Detection of a novel human coronavirus by real-time reverse-transcription polymerase chain reaction. *Euro Surveill* 17. (In Press).
- Lai, M.M., Perlman, S., and Anderson, L.J. (2007). Coronaviridae. in *Fields Virology*. Knipe, D.M. and Howley, P.M. ed. (Philadelphia: Lippincott, Williams & Wilkins), pp. 1305–1336
- Woo, P.C., Wang, M., Lau, S.K., Xu, H., Poon, R.W., Guo, R., Wong, B.H., Gao, K., Tsoi, H.W., Huang, Y., et al. (2007). Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. *J Virol* 81, 1574–1585.
- Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., et al. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U S A* 100, 13190–13195.
- Zhu, J., Xiao, G., Xu, Y., Yuan, F., Zheng, C., Liu, Y., Yan, H., Cole, D.K., Bell, J.I., Rao, Z., et al. (2004). Following the rule: formation of the 6-helix bundle of the fusion core from severe acute respiratory syndrome coronavirus spike protein and identification of potent peptide inhibitors. *Biochem Biophys Res Commun* 319, 283–288.