The proximal origin of SARS-CoV-2
Kristian G. Anderson

Thursday, March 19, 2020   10:41 AM

Introduction
This review uses a number of studies to trace the origin of the SARS-CoV-2 coronavirus.

Conclusions
The section of the virus (Receptor Binding Domain) responsible for binding a receptor on human cell surface is characterized as "not ideal".

There is great similarity in the structure of the bat coronavirus and the pangolin coronavirus' receptor binding domain, giving circumstantial evidence it crossed and mutated both of these viruses.

It is also possible the virus escaped a lower security laboratory, as this has happened in the past.

Amendments

3. CoV-2 does bind the ACE tightly/strongly, but it doesn't do it optimally. This non-ideal binding of the virus spike protein to the human ACE and the difference from the original CoV implies it has mutated to a human-like ACE that allows an actual optimal binding, but not binding to a human ACE, itself. It just so happens that it binds to the human version of ACE tightly, but since it is not optimal, it implies it is better suited for some other specie's ACE.

1. CoV-2 (also known as CoVID-19) virus binds (via a spike protein - a protrusion common to viruses) the Angiotensin Converting Enzyme that is located on the surface of particular cells of our body. Once bound, it is endocytosed (meaning, it is "absorbed") by the cell.

2. The spike protein on the virus that allows the virus to bind the ACE receptor on human cells is made up of variable/different amino acids - this is a source of identification on how the virus interacts with human cells (we can cross reference the 6 amino acids that make up the spike protein to the human genome and see where they might match together). Although both CoV viruses (the first one and now this second one) are the same in their mechanism of entry to the cell, 5 of the 6 amino acids that make up the spike protein are different. This spike protein binds the ACE receptor of humans, cats, and other species with high affinity (strongly).

4. CoV-2 could have emerged from an animal and then infected a human, because a bat specific version of a similar virus (RaTG13) is 96% identical in genome (DNA) to this CoV-2 in humans, except for its spike proteins (as explained earlier, these bind the ACE receptor on human cells). RaTG13 cannot bind human cells well, so it does not infect humans like it does bats due to this difference in spike proteins. However, pangolins also have a Coronavirus that does share the same receptor binding domain (RBD) as this CoV-2 in humans. It is then feasible that these two versions of the virus kept coming into contact in one of these animals and eventually formed a new virus with a cross breed of both by natural selection.
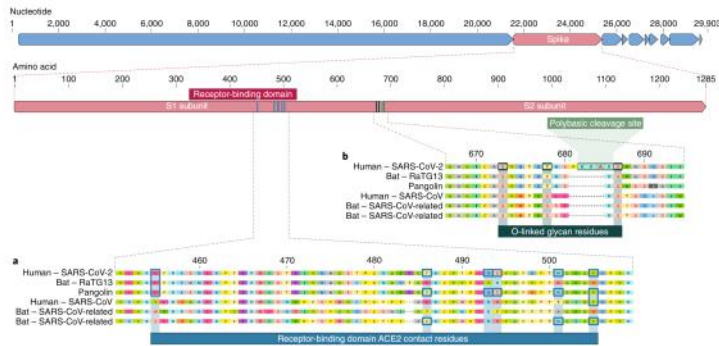
Fig. 1 | Features of the spike protein in human SARS-CoV-2 and related coronaviruses. a, Mutations in contact residues of the SARS-CoV-2 spike protein. The spike protein of SARS-CoV-2 (red bar at top) was aligned against the most closely related SARS-CoV-like coronaviruses and SARS-CoV itself. Key residues in the spike protein that make contact to the ACE2 receptor are marked with blue boxes in both SARS-CoV-2 and related viruses, including SARS-CoV (Urbani strain). **b**, Acquisition of polybasic cleavage site and O-linked glycans. Both the polybasic cleavage site and the three adjacent predicted O-linked glycans are unique to SARS-CoV-2 and were not previously seen in lineage B betacoronaviruses. Sequences shown are from NCBI GenBank, accession codes MN908947, MN996532, AY278741, KY417146 and MK211376. The pangolin coronavirus sequences are a consensus generated from SRR10168377 and SRR10168378 (NCBI BioProject PRJNA573298)[29,30].

genome[1], some pangolin coronaviruses exhibit strong similarity to SARS-CoV-2 in the RBD, including all six key RBD residues[21] (Fig. 1). This clearly shows that the SARS-CoV-2 spike protein optimized for binding to human-like ACE2 is the result of natural selection.

Neither the bat betacoronaviruses nor the pangolin betacoronaviruses sampled thus far have polybasic cleavage sites. Although no animal coronavirus has been identified that is sufficiently similar to have served as the direct progenitor of SARS-CoV-2, the diversity of coronaviruses in bats and other species is massively undersampled. Mutations, insertions and deletions can occur near the S1–S2 junction of coronaviruses[22], which shows that the polybasic cleavage site can arise by a natural evolutionary process. For a precursor virus to acquire both the polybasic cleavage site and mutations in the spike protein suitable for binding to human ACE2, an animal host would probably have to have a high population density (to allow natural selection to proceed efficiently) and an ACE2-encoding gene that is similar to the human ortholog.

2. Natural selection in humans following zoonotic transfer. It is possible that a progenitor of SARS-CoV-2 jumped into humans, acquiring the genomic features described above through adaptation during undetected human-to-human transmission. Once acquired, these adaptations would enable the pandemic to take off and produce a sufficiently large cluster of cases to trigger the surveillance system that detected it[1,2].

All SARS-CoV-2 genomes sequenced so far have the genomic features described above and are thus derived from a common ancestor that had them too. The presence in pangolins of an RBD very similar to that of SARS-CoV-2 means that we can infer this was also probably in the virus that jumped to humans. This leaves the insertion of polybasic cleavage site to occur during human-to-human transmission.

Estimates of the timing of the most recent common ancestor of SARS-CoV-2 made with current sequence data point to emergence of the virus in late November 2019 to early December 2019[23], compatible with the earliest retrospectively confirmed cases[24]. Hence, this scenario presumes a period of unrecognized transmission in humans between the initial zoonotic event and the acquisition of the polybasic cleavage site. Sufficient opportunity could have arisen if there had been many prior zoonotic events that produced short chains of human-to-

human transmission over an extended period. This is essentially the situation for MERS-CoV, for which all human cases are the result of repeated jumps of the virus from dromedary camels, producing single infections or short transmission chains that eventually resolve, with no adaptation to sustained transmission[25].

Studies of banked human samples could provide information on whether such cryptic spread has occurred. Retrospective serological studies could also be informative, and a few such studies have been conducted showing low-level exposures to SARS-CoV-like coronaviruses in certain areas of China[26]. Critically, however, these studies could not have distinguished whether exposures were due to prior infections with SARS-CoV, SARS-CoV-2 or other SARS-CoV-like coronaviruses. Further serological studies should be conducted to determine the extent of prior human exposure to SARS-CoV-2.

3. Selection during passage. Basic research involving passage of bat SARS-CoV-like coronaviruses in cell culture and/or animal models has been ongoing for many years in biosafety level 2 laboratories across the world[27], and there are documented instances

of laboratory escapes of SARS-CoV[28]. We must therefore examine the possibility of an inadvertent laboratory release of SARS-CoV-2.

In theory, it is possible that SARS-CoV-2 acquired RBD mutations (Fig. 1a) during adaptation to passage in cell culture, as has been observed in studies of SARS-CoV[17]. The finding of SARS-CoV-like coronaviruses from pangolins with nearly identical RBDs, however, provides a much stronger and more parsimonious explanation of how SARS-CoV-2 acquired these via recombination or mutation[19].

The acquisition of both the polybasic cleavage site and predicted O-linked glycans also argues against culture-based scenarios. New polybasic cleavage sites have been observed only after prolonged passage of low-pathogenicity avian influenza virus in vitro or in vivo[17]. Furthermore, a hypothetical generation of SARS-CoV-2 by cell culture or animal passage would have required prior isolation of a progenitor virus with very high genetic similarity, which has not been described. Subsequent generation of a polybasic cleavage site would have then required repeated passage in cell culture or animals with ACE2 receptors similar to those of humans, but such work has also not previously been described. Finally, the generation of the predicted O-linked glycans is also unlikely to have occurred due to cell-culture passage, as such features suggest the involvement of an immune system[19].

Conclusions

In the midst of the global COVID-19 public-health emergency, it is reasonable to wonder why the origins of the pandemic matter. Detailed understanding of how an animal virus jumped species boundaries to infect humans so productively will help in the prevention of future zoonotic events. For example, if SARS-CoV-2 pre-adapted in another animal species, then there is the risk of future re-emergence events. In contrast, if the adaptive process occurred in humans, then even if repeated zoonotic transfers occur, they are unlikely to take off without the same series of mutations. In addition, identifying the closest viral relatives of SARS-CoV-2 circulating in animals will greatly assist studies of viral function. Indeed, the availability of the RaTG13 bat

sequence helped reveal key RBD mutations and the polybasic cleavage site.

The genomic features described here may explain in part the infectiousness and transmissibility of SARS-CoV-2 in humans. Although the evidence shows that SARS-CoV-2 is not a purposefully manipulated virus, it is currently impossible to prove or disprove the other theories of its origin described here. However, since we observed all notable SARS-CoV-2 features, including the optimized RBD and polybasic cleavage site, in related coronaviruses in nature, we do not believe that any type of laboratory-based scenario is plausible.

More scientific data could swing the balance of evidence to favor one hypothesis over another. Obtaining related viral sequences from animal sources would be the most definitive way of revealing viral origins. For example, a future observation of an intermediate or fully formed polybasic cleavage site in a SARS-CoV-2-like virus from animals would lend even further support to the natural-selection hypotheses. It would also be helpful to obtain more genetic and functional data about SARS-CoV-2, including animal studies. The identification of a potential intermediate host of SARS-CoV-2, as well as sequencing of the virus from very early cases, would similarly be highly informative. Irrespective of the exact mechanisms by which SARS-CoV-2 originated via natural selection, the ongoing surveillance of pneumonia in humans and other animals is clearly of utmost importance. □

Kristian G. Andersen[1,2]✉,
Andrew Rambaut[3], W. Ian Lipkin[4],
Edward C. Holmes[5] and Robert F. Garry[6,7]

[1]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, USA. [2]Scripps Research Translational Institute, La Jolla, CA, USA. [3]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. [4]Center for Infection and Immunity, Mailman School of Public Health of Columbia University, New York, NY, USA. [5]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, Australia. [6]Tulane University, School of Medicine, Department of Microbiology and Immunology, New Orleans, LA, USA. [7]Zalgen Labs, Germantown, MD, USA.
✉e-mail: andersen@scripps.edu

References

1. Zhou, P. et al. Nature https://doi.org/10.1038/s41586-020-2012-7 (2020).
2. Wu, F. et al. Nature https://doi.org/10.1038/s41586-020-2008-3 (2020).
3. Gorbalenya, A. E. et al. bioRxiv https://doi.org/10.1101/2020.02.07.937862 (2020).
4. Jiang, S. et al. Lancet https://doi.org/10.1016/S0140-6736(20)30419-0 (2020).
5. Dong, E., Du, H. & Gardner, L. Lancet Infect. Dis. https://doi.org/10.1016/S1473-3099(20)30120-1 (2020).
6. Corman, V. M., Muth, D., Niemeyer, D. & Drosten, C. Adv. Virus Res. 100, 163–188 (2018).
7. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. J. Virol. https://doi.org/10.1128/JVI.00127-20 (2020).
8. Walls, A. C. et al. bioRxiv https://doi.org/10.1101/2020.02.19.956581 (2020).
9. Wrapp, D. et al. Science https://doi.org/10.1126/science.abb2507 (2020).
10. Letko, M., Marzi, A. & Munster, V. Nat. Microbiol. https://doi.org/10.1038/s41564-020-0688-y (2020).
11. Sheahan, T. et al. J. Virol. 82, 2274–2285 (2008).
12. Nao, N. et al. MBio 8, e02298-16 (2017).
13. Chan, C.-M. et al. Exp. Biol. Med. 233, 1527–1536 (2008).
14. Follis, K. E., York, J. & Nunberg, J. H. Virology 350, 358–369 (2006).
15. Menachery, V. D. et al. J. Virol. https://doi.org/10.1128/JVI.01774-19 (2019).
16. Alexander, D. J. & Brown, I. H. Rev. Sci. Tech. 28, 19–38 (2009).
17. Ito, T. et al. J. Virol. 75, 4439–4443 (2001).
18. Bagdonaite, I. & Wandall, H. H. Glycobiology 28, 443–467 (2018).
19. Cui, J., Li, F. & Shi, Z.-L. Nat. Rev. Microbiol. 17, 181–192 (2019).
20. Alexander, F. et al. Virus Res. 189, 262–270 (2014).
21. Zhang, T., Wu, Q. & Zhang, Z. bioRxiv https://doi.org/10.1101/2020.02.19.950253 (2020).
22. Yamada, Y. & Liu, D. X. J. Virol. 83, 8744–8758 (2009).
23. Rambaut, A. Virological.org http://virological.org/t/356 (2020).
24. Huang, C. et al. Lancet https://doi.org/10.1016/S0140-6736(20)30183-5 (2020).
25. Dudas, G., Carvalho, L. M., Rambaut, A. & Bedford, T. eLife 7, e31257 (2018).
26. Wang, N. et al. Virol. Sin. 33, 104–107 (2018).
27. Ge, X.-Y. et al. Nature 503, 535–538 (2013).
28. Lim, P. L. et al. N. Engl. J. Med. 350, 1740–1745 (2004).
29. Wong, M. C., Javornik Cregeen, S. J., Ajami, N. J. & Petrosino, J. F. bioRxiv https://doi.org/10.1101/2020.02.07.939207 (2020).
30. Liu, P., Chen, W. & Chen, J.-P. Viruses 11, 979 (2019).

New Section 4 Page 3