Word-comparison-for-A4-print-showing-changes-to preprint-to-arrive-at-final-version.docx/pdf

This shows the deletions and additions which would need to be made to the preprint version of *The Proximal Origin of SARS-CoV-2* to arrive at final version as published in Nature Medicine on 2020-03-17:

https://www.nature.com/articles/s41591-020-0820-9

The preprint posted to virological.org is no longer there, but archive.org has it:

https://web.archive.org/web/20200217170645/https://virological.org/t/the-proximal-origin-of-sars-cov-2/398

I used Word 365 on the plain text of these two HTML web pages, making two Word files and then comparing them. I pasted the comparison text into a fresh file, saved it and here saved it as a PDF.

Robin Whittle rw@firstpr.com.au 2023-07-19

Since the first reports of ~~a~~ novel pneumonia (COVID-19) in Wuhan ~~city~~, Hubei province, ~~China~~China1,2, there has been considerable discussion ~~and uncertainty over~~on the origin of the causative virus, SARS-CoV-~~2.~~23 (also referred to as HCoV-19)4. Infections with SARS-CoV-2 are now widespread ~~in China, with cases in every province. As~~ , and as of ~~14 February~~11 March 2020, ~~64,473 such~~121,564 cases have been confirmed~~, with 1,384 deaths attributed to the virus. These official case numbers are likely an underestimate because of limited reporting of mild and asymptomatic cases, and the virus is clearly capable of efficient human-to-human transmission. Based on the possibility of spread to~~ in more than 110 countries ~~with weaker healthcare systems, the World Health Organization has declared the COVID-19 outbreak a Public Health Emergency of International Concern (PHEIC). There are currently neither vaccines nor specific treatments for this disease~~, with 4,373 deaths5.

SARS-CoV-2 is the seventh ~~member of the Coronaviridae~~ coronavirus known to infect humans~~. Three of these viruses,~~; SARS ~~-~~CoV ~~-1~~, MERS~~,~~-CoV and SARS-CoV-2~~,~~ can cause severe disease~~; four,~~, whereas HKU1, NL63, OC43 and 229E~~,~~ are associated with mild ~~respiratory symptoms. Herein,~~symptoms6. Here we review what can be deduced about the origin ~~and early evolution~~ of SARS-CoV-2 from ~~the~~ comparative analysis of ~~available genome sequence~~ genomic data. ~~In particular, we~~We offer a perspective on the notable features ~~in~~of the SARS-CoV-2 genome and discuss scenarios

by which ~~these features~~ they could have arisen. ~~Importantly, this analysis provides evidence~~Our analyses clearly show that SARS-CoV-2 is not a laboratory construct ~~nor~~or a purposefully manipulated virus.

Notable features of the SARS-CoV-2 genome

Our ~~The genomic~~ comparison of ~~both~~ alpha- and betacoronaviruses ~~(family Coronaviridae) described below~~ identifies two notable genomic features of ~~the~~ SARS-CoV-2 ~~genome~~: (i) ~~based~~ on the basis of structural ~~modelling~~studies[7,8,9] and ~~early~~ biochemical ~~experiments~~experiments[1,9,10], SARS-CoV-2 appears to be optimized for binding to the human ~~ACE2~~ receptor~~;~~ ACE2; and (ii) the ~~highly variable~~ spike ~~(S)~~ protein of SARS-CoV-2 has a functional polybasic (furin) cleavage site at the S1 ~~and~~ – S2 boundary ~~via~~through the insertion of ~~twelve nucleotides. Additionally, this event~~12 nucleotides[8], which additionally led to the predicted acquisition of three ~~predicted~~ O-linked glycans around the ~~polybasic cleavage~~ site.

1. Mutations in the receptor–binding domain of SARS-CoV-2

The receptor–binding domain (RBD) in the spike protein ~~of SARS-CoV and SARS-related coronaviruses~~ is the most variable part of the ~~virus genome.~~coronavirus genome[1,2]. Six ~~residues in the~~ RBD ~~appear~~amino acids have been shown to be critical for binding to ~~the human~~ ACE2 ~~receptor~~receptors and for determining the host ~~range[1]. Using~~range of SARS-CoV-like viruses[7]. With coordinates based on ~~the Urbani strain of~~ SARS-CoV, they are Y442, L472, N479, D480, T487~~,~~ and Y491~~. The corresponding residues in SARS-CoV-2 are~~, which correspond to L455, F486, Q493, S494, N501~~,~~ and Y505 in SARS-CoV-2[7]. Five of these six residues ~~are mutated in SARS-CoV-2 compared to its most closely related virus, RaTG13 sampled from a Rhinolophus affinis bat, to which it is ~96% identical[2] (Figure~~differ between SARS-CoV-2 and SARS-CoV (Fig. 1a). ~~Based on modeling[1]~~On the basis of structural studies[7,8,9] and biochemical ~~experiments[3,4]~~experiments[1,9,10], SARS-CoV-2 seems to have an RBD that ~~may bind~~binds with high affinity to ACE2 from ~~human, non-human primate, ferret, pig, and cat, as well as~~ humans, ferrets, cats and other species with high receptor ~~homology[1]. In contrast, SARS-CoV-2 may bind less efficiently to ACE2 in other species associated with SARS-like viruses, including rodents and civets[1]~~homology[7].

~~The phenylalanine (F) at residue 486 in the SARS-CoV-2 S protein corresponds to L472 in the SARS-CoV Urbani strain. Notably, in SARS-CoV cell culture experiments the L472 mutates to phenylalanine (L472F)[5], which is predicted to be optimal for binding of the SARS-CoV RBD to the human ACE2 receptor[6]. However, a phenylalanine in this position is also present in several SARS-like-CoVs from bats (Figure 1a). While these analyses suggest that SARS-CoV-2 may be capable of binding the human ACE2 receptor with high~~

affinity, the interaction is not predicted to be optimal1. Additionally, several of the key residues in the RBD of SARS-CoV-2 are different to those previously described as optimal for human ACE2 receptor binding6. In contrast to these computational predictions, recent binding studies indicate that SARS-CoV-2 binds with high affinity to human ACE27. Thus the SARS-CoV-2 spike appears to be the result of selection on human or human-like ACE2 permitting another optimal binding solution to arise. This is strong evidence that SARS-CoV-2 is not the product of genetic engineering.

Polybasic cleavage site and O-linked glycans

The second notable feature of SARS-CoV-2 is a predicted polybasic cleavage site (RRAR) in the spike protein at the junction of S1 and S2, the two subunits of the spike protein (Figure 1b)8,9. In addition to two basic arginines and an alanine at the cleavage site, a leading proline is also inserted; thus, the fully inserted sequence is PRRA (Figure

Fig. 1: Features of the spike protein in human SARS-CoV-2 and related coronaviruses.

1b). The strong turn created by the proline insertion is predicted to result in the addition of O-linked glycans to S673, T678, and S686 that flank the polybasic cleavage site. A polybasic cleavage site has not previously been observed in related lineage B betacoronaviruses and is a unique feature of SARS-CoV-2. Some human betacoronaviruses, including HCoV-HKU1 (lineage A), have polybasic cleavage sites, as well as predicted O-linked glycans near the S1/S2 cleavage site.

While the functional consequence of the polybasic cleavage site in SARS-CoV-2 is unknown, experiments with SARS-CoV have shown that engineering such a site at the S1/S2 junction enhances cell–cell fusion but does not affect virus entry10. Polybasic cleavage sites allow effective cleavage by furin and other proteases, and can be acquired at the junction of the two subunits of the haemagglutinin (HA) protein of avian influenza viruses in conditions that select for rapid virus replication and transmission (e.g. highly dense chicken populations). HA serves a similar function in cell-cell fusion and viral entry as the coronavirus S protein. Acquisition of a polybasic cleavage site in HA, by either insertion or recombination, converts low pathogenicity avian influenza viruses into highly pathogenic forms11-13. The acquisition of polybasic cleavage sites by the influenza virus HA has also been observed after repeated forced passage in cell culture or through animals14,15. Similarly, an avirulent isolate of Newcastle Disease virus became highly pathogenic during serial passage in chickens by incremental acquisition of a polybasic cleavage site at the junction of its fusion protein subunits16. The potential function of the three predicted O-linked glycans is less clear, but they could create a “mucin-like domain” that would shield potential epitopes or key residues on the SARS-CoV-2 spike protein. Biochemical analyses or

structural studies are required to determine whether or not the predicted O-linked glycan sites are utilized.

figure
figure2718×1487 394 KB

Figure 1. (

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 CoVscoronaviruses and SARS-CoV-1 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 therelated viruses, including SARS-CoV (Urbani strain. (). b), Acquisition of polybasic cleavage site and O-linked glycans. TheBoth the polybasic cleavage site is marked in grey withand the three adjacent predicted O-linked glycans in blue. Both the polybasic cleavage site and 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 numberscodes MN908947, MN996532, AY278741, KY417146, and MK211376. The pangolin coronavirus sequences are a consensus generated from SRR10168377 and SRR10168378 (NCBI BioProject PRJNA573298)18,1929,30. Full size image

While the analyses above suggest that SARS-CoV-2 may bind human ACE2 with high affinity, computational analyses predict that the interaction is not ideal7 and that the RBD sequence is different from those shown in SARS-CoV to be optimal for receptor binding7,11. Thus, the high-affinity binding of the SARS-CoV-2 spike protein to human ACE2 is most likely the result of natural selection on a human or human-like ACE2 that permits another optimal binding solution to arise. This is strong evidence that SARS-CoV-2 is not the product of purposeful manipulation.

2. Polybasic furin cleavage site and O-linked glycans

The second notable feature of SARS-CoV-2 is a polybasic cleavage site (RRAR) at the junction of S1 and S2, the two subunits of the spike8 (Fig. 1b). This allows effective cleavage by furin and other proteases and has a role in determining viral infectivity and host range12. In addition, a leading proline is also inserted at this site in SARS-CoV-2; thus, the inserted sequence is PRRA (Fig. 1b). The turn created by the proline is predicted to result in the addition of O-linked glycans to S673, T678 and S686, which flank the cleavage site and are unique to SARS-CoV-2 (Fig. 1b). Polybasic cleavage sites have not been observed in related 'lineage B' betacoronaviruses, although other human betacoronaviruses, including HKU1 (lineage A), have those sites and predicted O-linked glycans13. Given the level of genetic variation in the spike, it is likely that SARS-

CoV-2-like viruses with partial or full polybasic cleavage sites will be discovered in other species.

The functional consequence of the polybasic cleavage site in SARS-CoV-2 is unknown, and it will be important to determine its impact on transmissibility and pathogenesis in animal models. Experiments with SARS-CoV have shown that insertion of a furin cleavage site at the S1–S2 junction enhances cell–cell fusion without affecting viral entry14. In addition, efficient cleavage of the MERS-CoV spike enables MERS-like coronaviruses from bats to infect human cells15. In avian influenza viruses, rapid replication and transmission in highly dense chicken populations selects for the acquisition of polybasic cleavage sites in the hemagglutinin (HA) protein16, which serves a function similar to that of the coronavirus spike protein. Acquisition of polybasic cleavage sites in HA, by insertion or recombination, converts low-pathogenicity avian influenza viruses into highly pathogenic forms16. The acquisition of polybasic cleavage sites by HA has also been observed after repeated passage in cell culture or through animals17.

The function of the predicted O-linked glycans is unclear, but they could create a 'mucin-like domain' that shields epitopes or key residues on the SARS-CoV-2 spike protein18. Several viruses utilize mucin-like domains as glycan shields involved immunoevasion18. Although prediction of O-linked glycosylation is robust, experimental studies are needed to determine if these sites are used in SARS-CoV-2.

Theories of SARS-CoV-2 origins

It is ~~unlikely~~improbable that SARS-CoV-2 emerged through laboratory manipulation of ~~an existing SARS-~~a related SARS-CoV-like coronavirus. As noted above, the RBD of SARS-CoV-2 is optimized for binding to human ACE2 ~~receptor binding~~ with an efficient ~~binding~~ solution different ~~to that which would have been predicted. Further~~from those previously predicted7,11. Furthermore, if genetic manipulation had been performed, one ~~would expect that one~~ of the several reverse ~~-~~genetic systems available for betacoronaviruses would probably have been ~~used~~used19. However, ~~this is not the case as~~ the genetic data ~~shows~~irrefutably show that SARS-CoV-2 is not derived from any previously used virus ~~backbone17~~backbone20. Instead, we propose two scenarios that can plausibly explain the origin of SARS-CoV-2: (i) natural selection in ~~a non-human~~an animal host ~~prior to~~before zoonotic transfer~~,~~; and (ii) natural selection in humans following zoonotic transfer. We also discuss whether selection during passage ~~in culture~~ could have given rise to ~~the same observed features~~SARS-CoV-2.

~~Selection~~ 1. Natural selection in an animal host~~.~~ before zoonotic transfer

As many ~~of the~~ early cases of COVID-19 were linked to the Huanan ~~seafood and wildlife~~ market in ~~Wuhan~~Wuhan[1,2], it is possible that an animal source was present at this location. Given the similarity of SARS-CoV-2 to bat SARS-CoV-like ~~CoVs, particularly RaTG13,~~ coronaviruses[2], it is ~~plausible~~likely that bats serve as reservoir hosts for ~~SARS-CoV-2. It is important, however, to note that previous outbreaks of betacoronaviruses in humans involved direct exposure to animals other than bats, including civets (SARS) and camels (MERS), that carry viruses that are genetically very similar to SARS-CoV-1 or MERS-CoV, respectively. By analogy, viruses closely related to SARS-Cov-2 may be circulating in one or more animal species. Initial analyses indicate that~~ its progenitor. Although RaTG13, sampled from a Rhinolophus affinis bat[1], is ~96% identical overall to SARS-CoV-2, its spike diverges in the RBD, which suggests that it may not bind efficiently to human ACE2[7] (Fig. 1a).

Malayan pangolins (~~-~~Manis javanica~~-~~) illegally imported into Guangdong province contain ~~a CoV that is~~coronaviruses similar to SARS-CoV-~~2[18,19]~~[21]. Although the RaTG13 bat virus ~~RaTG13~~ remains the closest ~~relative~~ to SARS-CoV-2 across the ~~whole genome, the Malayan~~genome[1], some pangolin ~~CoV is identical~~ coronaviruses exhibit strong similarity to SARS-CoV-2 ~~at~~ in the RBD, including all six key RBD ~~residues (Figure~~residues[21] (Fig. 1). ~~However, no pangolin CoV has yet~~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 ~~SARS-CoV-2 across its entire genome to support direct human infection. In addition, the pangolin CoV does not carry a polybasic cleavage site insertion.~~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 ~~receptor binding~~, an animal host would ~~likely~~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 ~~orthologue. Further characterization of CoVs in pangolins and other animals that may harbour SARS-CoV-like viruses should be a public health priority~~ortholog.

~~Cryptic adaptation to humans.~~ 2. Natural selection in humans following zoonotic transfer

It is ~~also~~ possible that a progenitor ~~to~~of SARS-CoV-2 jumped ~~from a non-human animal to~~into humans, ~~with~~acquiring the genomic features described above ~~acquired~~ through adaptation during ~~subsequent~~undetected human-to-human transmission. ~~We surmise that once~~Once acquired, these adaptations ~~were acquired (either together or in series) it~~ would enable the ~~outbreak~~pandemic to take ~~-~~ off~~, producing~~ and produce a sufficiently large ~~and unusual~~ cluster of ~~pneumonia~~ cases to trigger the surveillance system that ~~ultimately~~ detected ~~it~~it[1,2].

All SARS-CoV-2 genomes sequenced so far have the ~~well adapted RBD and the polybasic cleavage site, and~~ genomic features described above and are thus derived from a common ancestor that had ~~these features.~~them too. The presence ~~of an RBD~~ in pangolins ~~that is~~of an RBD very similar to ~~the one in~~that of SARS-CoV-2 means that we can infer this was ~~likely already present~~also probably in the virus that jumped to humans~~, even if we don't yet have the exact non-human progenitor virus.~~. This leaves the insertion of polybasic cleavage site ~~insertion~~ to occur during human-to-human transmission. ~~Following the example of the influenza A virus HA gene, a specific insertion or recombination event is required to enable the emergence of SARS-CoV-2 as an epidemic pathogen.~~

Estimates of the timing of the most recent common ancestor ~~(tMRCA)~~ of SARS-CoV-2 ~~using currently available genome~~made with current sequence data point to ~~virus~~ emergence of the virus in late November 2019 to early December ~~2019~~20,~~21~~201923, compatible with the earliest retrospectively confirmed ~~cases22~~cases24. Hence, this scenario presumes a period of ~~unrecognised~~unrecognized transmission in humans between the initial zoonotic ~~transfer~~ event and the acquisition of the polybasic cleavage site. Sufficient opportunity could ~~occur~~have arisen if there had been many prior zoonotic events ~~producing~~that produced short chains of human-to-human transmission ~~(so-called 'stuttering chains')~~ over an extended period. This is essentially the situation for MERS-CoV ~~in the Arabian Peninsula where all the~~, for which all human cases are the result of repeated jumps of the virus from dromedary camels, producing single infections or short ~~chains of~~ transmission chains that eventually resolve~~. To date, after 2,499 cases over 8 years, no human adaptation has emerged that has allowed MERS-CoV to take hold in the human population~~, with no adaptation to sustained transmission[25].

~~How could we test whether cryptic spread of SARS-CoV-2 enabled human adaptation? Metagenomic studies~~Studies of banked ~~serum~~human samples could provide ~~important~~ information~~, but given the relatively short period of viremia it may be impossible to detect low-level SARS-CoV-2 circulation in historical samples.~~ on whether such cryptic spread has occurred. Retrospective serological studies

potentially could also be informative, and a few such studies have already been conducted. One found that animal importation traders had a 13% seropositivity to coronaviruses23, while another noted that 3% residents of a village in Southern China were seropositive to these viruses24. Interestingly, 200 residents of Wuhan did not show coronavirus seroreactivity. showing low-level exposures to SARS-CoV-like coronaviruses in certain areas of China26. Critically, however, these studies could not have distinguished whether positive serological responsesexposures were due to a prior infectioninfections with SARS-CoV-1, SARS-CoV-2 or -2.other SARS-CoV-like coronaviruses. Further retrospective serological studies should be conducted to determine the extent of prior human exposure to betacoronaviruses in different geographic areas, particularly using assays that can distinguish among multiple betacoronavirusesSARS-CoV-2.

3. Selection during passage.

Basic research involving passage of bat SARS-CoV-like coronaviruses in cell culture and/or animal models havehas been ongoing in BSL-2 for many years in multiplebiosafety level 2 laboratories across the world25-28. Thereworld27, and there are also documented instances of the laboratory acquisition of SARS-CoV-1 by laboratory personnel working under BSL-2 containment29,30.escapes of SARS-CoV28. We must therefore considerexamine the possibility of a deliberate or an inadvertent laboratory release of SARS-CoV-2.

In theory, it is possible that SARS-CoV-2 acquired the observed RBD mutations site(Fig. 1a) during adaptation to passage in cell culture, as has been observed in studies with SARS-CoV5 as well as MERS-CoV31. However, the acquisition ofof SARS-CoV11. 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 mutation19.

The acquisition of both the polybasic cleavage site orand predicted O-linked glycans – if functional – also argues against this scenario.culture-based scenarios. New polybasic cleavage sites have only been observed only after prolonged passagingpassage of low-pathogenicity avian influenza virus in cell culturevitro or animalsin vivo17. Furthermore, thea hypothetical generation of SARS-CoV-2 by cell culture or animal passage would have required prior isolation of a progenitor virus with a very high genetic similarity., which has not been described. Subsequent generation of a polybasic cleavage site would have then required an intense program ofrepeated passage in cell culture or animals with ACE-2 receptorACE2 receptors similar to those of humans (e.g. ferrets). It is, but such work has also questionable whethernot previously been described. Finally, the generation of the

predicted O-linked glycans ~~would~~is also unlikely to have occurred ~~on~~due to cell ~~-~~culture passage, as such ~~mutations typically~~features suggest the involvement of an immune ~~system, that is not present in vitro~~ system18.

Conclusions

In the midst of the global COVID-19 public ~~-~~health emergency, it is reasonable to wonder why the origins of the ~~epidemic~~pandemic matter. ~~A detailed~~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 ~~we are at~~there is the risk of future re-emergence events ~~even if the current epidemic is controlled.~~. In contrast, if the adaptive process ~~we describe~~ occurred in humans, then even if ~~we have~~ repeated zoonotic transfers occur, they are unlikely to take ~~-~~ off ~~unless~~without the same series of mutations ~~occurs~~. In addition, identifying the closest ~~animal~~viral relatives of SARS-CoV-2 circulating in animals will greatly assist studies of ~~virus~~viral function. Indeed, the availability of the RaTG13 bat sequence ~~facilitated the comparative genomic analysis performed here, helping to~~helped reveal ~~the~~ key RBD mutations ~~in the RBD as well as~~and the polybasic cleavage site ~~insertion~~.

The genomic features described here may ~~in part~~ explain in part the infectiousness and transmissibility of SARS-CoV-2 in humans. Although ~~genomic~~the evidence ~~does not support the idea~~shows that SARS-CoV-2 is not a ~~laboratory construct~~purposefully manipulated virus, it is currently impossible to prove or disprove the other theories of its origin described here~~, and it~~. 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 ~~unclear whether future~~ plausible.

More scientific data ~~will help resolve this issue. Identifying the immediate non-human animal source and obtaining virus~~could swing the balance of evidence to favor one hypothesis over another. Obtaining related viral sequences from ~~it~~animal sources would be the most definitive way of revealing ~~virus~~viral origins. ~~In addition, it would~~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 ~~the virus~~SARS-CoV-2, including ~~experimental~~animal studies ~~of receptor binding and the role of the polybasic cleavage site and predicted O-linked glycans.~~. The identification of a potential intermediate host of SARS-CoV-2, as well as ~~the~~ sequencing of the virus from very early cases ~~including those not connected to the Wuhan market~~, would similarly be highly informative. Irrespective of ~~how~~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.