

Origins of the SARS Corona Virus 2

Ariel Fernández

Daruma Institute for Applied Intelligence, AF Innovation GmbH,
1005 Oakhurst Ave, High Point, NC 27262
CONICET–Argentine National Research Council, Buenos Aires, Argentina
Contact: ariel@afinnovation.com

L. Ridgway Scott

University of Chicago, Chicago, Illinois 60637
Contact: ridg@uchicago.edu

September 28, 2021

Abstract

SARS-CoV-2 exhibits genomic features that are unusual for a natural virus. The U. S. federal government has funded research related to virus mutation in similar viruses in a laboratory in Wuhan, China. This raises questions of interest to physicians and policy makers. We explain how the Delta variant of SARS-CoV-2 adds weight to the probability of a laboratory leak. We make suggestions for policy changes in the future and for strategies to investigate further the possibility of a lab leak.

The origins of the SARS-CoV-2 virus has received significant attention [1, 2]. The focus of a possible laboratory leak has been Wuhan, China, especially the Wuhan Institute of Virology (WIV). A comparison has been made of the evidence for a natural origin versus a lab leak [3]. It has been lamented that information from China is not forthcoming, but we observe here that it may yet be possible to gain definitive information regarding a possible lab origin from the public record in the U. S., provided the right questions are asked.

1 Gain of function

It has been documented that the U. S. government has funded research regarding corona viruses at the WIV lab [4]. In a recent exchange [5], Dr. Fauci denied that the NIH funding was for gain of function. However, it has been reported that such research was funded by DoD [6], in part from the Defense Threat Reduction Agency. The precise definition of gain-of-function may be debatable, but since such research was prohibited under the Obama administration [7], it would be useful to provide more clarity.

Research that reduces threats is clearly in the interest of humanity. One example is developing an anti-missile missile. But to do so, you need to have an idea of what missiles you need to be defending against. If you make a mistake in building such target missiles, bad things can happen. But these can typically be localized to remote sites.

In a biological context, this means developing a suitable target, e.g., a virus, and thus gain-of-function research is one possible approach. But the issue in this case is: what happens if something goes wrong? Blame is not the issue. Rather we need to ask what are the policy implications for the future. We also need to ask how it might be possible to determine whether a lab leak occurred.

2 Sequence-based analysis

A recent paper [2] presented a sequence-based analysis of the SARS-CoV-2 virus that indicates how the genomic sequence differs from what would be expected for a naturally occurring genome in that lineage. One observation is the occurrence of a critical pair of arginines for which the nucleotide coding [8, Figure 8.6] is not one of the natural codings primarily used in corona viruses of the same lineage [9, Figure 1c]. Two consecutive arginines in SARS-CoV-2, namely R682 and R683, are encoded by CGGCGG, an extremely rare event in natural CoVs of the same lineage [3, 9]. The CGG codon is almost never used to encode for arginine in natural coronavirus, and two consecutive codons have an exceedingly low probability of occurring (estimated to be $< 0.25\%$ [2]).

These arginines are in a part of the virus (corresponding to the amino acid sequence PRRA, at positions 681–684) that is present in SARS-CoV-2 [10] but not in the purported [11] precursor bat virus CoV RaTG13 [2], or other known viruses (see e.g. [12, Table 1] and [13, Figure 3B]), in the region of focus regarding putative gain of function [14]. In addition, the mutation P681R occurs in the Delta variant [15] of SARS-CoV-2, also known as lineage B.1.617.2.

The Alpha variant (B.1.1.7) has the mutation P681H [16]. Like R, histidine (H) can be positively charged, making the mutated site more basic as in the Delta variant, although less basic than Delta (see [8, Table 4.2] and the discussion in [8, Section 4.5]).

The fact that the most infective variants have focused on the PRRA sequence suggests that natural selection is paying inordinate attention to the fragment and thus implies it is a major contributor to virulence. Thus sequence analysis raises important questions that should be addressed, in particular, where did the PRRA sequence in SARS-CoV-2 come from?

Of course, anything can occur in nature, and one single area of a virus sequence may not be the only critical one. So it is not possible to say for sure that this unusual signature is a smoking gun. Nature can make an astronomical number of mutations, so even very rare ones may occur. Estimating the product of a very small number and a very large number is not easy.

3 Attempts at suppression

The journal that published [2] received several requests to retract the letter, including from Chinese nationals, using specious and illogical arguments. Based on this pressure, the journal is planning to issue an Expression of Concern in an attempt to discredit the letter. It is understandable that China and those involved with the WIV funding do not want adverse press, and there is nothing wrong with individuals sending such requests to journals. Our journals must be sufficiently robust to withstand and process such comments. But such attempts at a coverup just elevate concerns over the virus origins.

4 Why do we care?

Does it really matter where a virus comes from? The animal impact of a virus may be no different whatever the source. But there are two reasons for wanting to know.

Some physicians feel that it is important to tell patients and their families as much as possible about what has happened to them. Thus they argue the origin matters for this discussion.

In addition, the origin of a virus is important to know for future planning regarding strategic policies. One possible conclusion may be that it is not a good idea to fund research on gain-of-function in countries which lack high safety standards [3] and open discourse.

If a lab leak occurred at Fort Detrick, it might have the same impact as a leak from anywhere else. But what is different is that a sargent could contact the Wall Street Journal to blow a whistle.

5 Investigational techniques

How would one be sure that a virus mutation was designed in a laboratory? The most obvious one is that lab members had documented, say in emails, the particular sequence of the mutated virus. Given the current government in China, it is not likely that such correspondence would ever be found.

However, if the research was externally funded (in the U. S.), then it would be possible that correspondence with the funding parties might indicate key features of the sequence that were designed to fingerprint the genomic sequence, or simply utilized due to experimental constraints. For example, someone might have related that a particular coding of a particular sidechain was chosen for this purpose. Thus it might be possible to shed some light on the issue by examining the emails of the funding parties. In crime drama, they refer to this as following the money.

6 Policy implications

We have already noted that the challenges of determining virus origin suggests limiting certain types of research funded by the U. S. government. But in addition, it may be appropriate to expand the expertise of the U. S. national security apparatus of the government to include genomic biology, so that we can better assess certain threats. SARS-CoV-2 is certainly a weapon of mass destruction. Whether it arose naturally or in a lab, we need to know better how to combat such agents.

The paper [3] does a great job of comparing the lab-leak and natural-cause theories for the origins of SARS-CoV-2. At this point, it is best to assume that either could be correct, or that there is some third explanation so far not articulated. But whatever the origins, we need to make policy changes now. In a sense, it would be better if the lab leak theory were correct. Because the usual natural-origin explanations (e.g., for SARS and MERS) have not been able to explain SARS-CoV-2 [3]. If it was not a lab leak, then there is some natural process, yet to be understood, which was able to create an extremely deadly virus. So we should be focusing on what that process could be, and it may be very difficult to discover.

References

- [1] Xiaolu Tang, Changcheng Wu, Xiang Li, Yuhe Song, Xinmin Yao, Xinkai Wu, Yuange Duan, Hong Zhang, Yirong Wang, Zhaohui Qian, et al. On the origin and continuing

- evolution of SARS-CoV-2. *National Science Review*, 7(6):1012–1023, 2020.
- [2] Ariel Fernández. Molecular biology clues portray SARS-CoV-2 as a gain-of-function laboratory manipulation of bat CoV RaTG13. *ACS Medicinal Chemistry Letters*, 12(6):941–942, 2021.
- [3] Nicholas Wade. The origin of COVID: Did people or nature open Pandora’s box at Wuhan. *Bulletin of the Atomic Scientists*, May 5, 2021.
- [4] Katherine Eban. The lab-leak theory: Inside the fight to uncover COVID-19’s origins. *Vanity Fair*, June 23, 2021.
- [5] Chris Sims. What is gain-of-function research? Here’s what Rand Paul and Dr. Fauci are fighting over. *Louisville Courier Journal*, July 23, 2021.
- [6] Pentagon gave USD 39 million to Peter Daszak’s NGO that funded coronavirus research at Wuhan lab. *Times Now Digital*, June 6, 2021.
- [7] Lisa Schnirring. Feds lift gain-of-function research pause, offer guidance. *CIDRAP News*, December 19, 2017.
- [8] L. Ridgway Scott and Ariel Fernández. *A Mathematical Approach to Protein Biophysics*. Springer-Verlag, 2017.
- [9] Wei Hou. Characterization of codon usage pattern in SARS-CoV-2. *Virology Journal*, 17(1):1–10, 2020.
- [10] Bryan A Johnson, Xuping Xie, Birte Kalveram, Kumari G Lokugamage, Antonio Muruato, Jing Zou, Xianwen Zhang, Terry Juelich, Jennifer K Smith, Lihong Zhang, et al. Furin cleavage site is key to SARS-CoV-2 pathogenesis. *BioRxiv*, 2020.
- [11] Yuri Deigin and Rossana Segreto. SARS-CoV-2’s claimed natural origin is undermined by issues with genome sequences of its relative strains: Coronavirus sequences RaTG13, MP789 and RmYN02 raise multiple questions to be critically addressed by the scientific community. *BioEssays*, page 2100015, 2021.
- [12] Laura E Frazier, Bailey Lubinski, Tiffany Tang, Susan Daniel, Javier A Jaimes, and Gary R Whittaker. Spike protein cleavage-activation mediated by the SARS-CoV-2 P681R mutation: a case-study from its first appearance in variant of interest (VOI) A.23.1 identified in Uganda. *bioRxiv*, 2021. <http://dx.doi.org/10.1101/2021.06.30.450632> doi:10.1101/2021.06.30.450632.
- [13] Rossana Segreto, Yuri Deigin, Kevin McCairn, Alejandro Sousa, Dan Sirotkin, Karl Sirotkin, Jonathan J Couey, Adrian Jones, and Daoyu Zhang. Should we discount the laboratory origin of covid-19? *Environ Chem Lett*, 19:2743–2757, 2021.
- [14] Shuai Xia, Qiaoshuai Lan, Shan Su, Xinling Wang, Wei Xu, Zezhong Liu, Yun Zhu, Qian Wang, Lu Lu, and Shibo Jiang. The role of furin cleavage site in SARS-CoV-2 spike protein-mediated membrane fusion in the presence or absence of trypsin. *Signal transduction and targeted therapy*, 5(1):1–3, 2020.

- [15] Saathvik R Kannan, Austin N Spratt, Alisha R Cohen, S Hasan Naqvi, Hitendra S Chand, Thomas P Quinn, Christian L Lorson, Siddappa N Byrareddy, and Kamal Singh. Evolutionary analysis of the Delta and Delta Plus variants of the SARS-CoV-2 viruses. *Journal of Autoimmunity*, 124:102715, 2021.
- [16] Suman Pokhrel, Benjamin R Kraemer, Scott Burkholz, and Daria Mochly-Rosen. Natural variants in SARS-CoV-2 Spike protein pinpoint structural and functional hotspots with implications for prophylaxis and therapeutic strategies. *Scientific Reports*, 11(1):1–10, 2021.