

# **DIGGING THROUGH THE SCIENCE—AND THE NOISE—ON WHAT IS KNOWN ABOUT THE ORIGIN OF SARS COV-2**

*Update: In a new post we find that Shi Zhingli of Wuhan Institute of Virology has provided convincing evidence to Scientific American that SARS CoV-2 is the result of a natural jump to humans from an animal host and was not accidentally released from her lab, which had no isolates of any viruses that match closely enough to be the outbreak virus.*

Although it seems that all of this has been going on forever at this point, it's important to realize that the COVID-19 pandemic outbreak probably began less than six months ago. In the context of how we develop an understanding of a disease like this one, and the virus that causes it, SARS CoV-2, that means that we really have only just begun our analysis. Nevertheless, because of the ongoing disastrous impact on global public health as well as the global economy, it is imperative that we learn as much as we can as fast as we can.

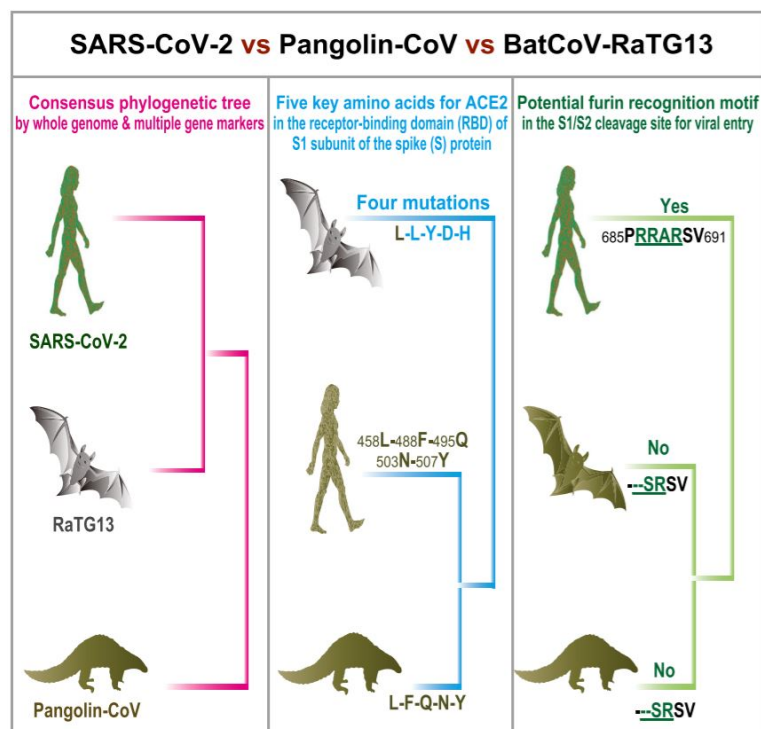
In this post, I want to take a deep dive into what virologists and epidemiologists have pieced together on the emergence of SARS CoV-2. The problem is that what might initially appear to be straightforward scientific and public health questions eventually get muddled by accusations of disinformation, accusations of hiding data and offerings of potential leaks of intelligence that also have a chance to be disinformation. These noisy battles relate to basic facts that have a direct bearing on our understanding of the virus' origin.

As a result, it needs to be stated from the outset that because some of the needed basic

information may be hidden or some of what we think we know might be wrong. Therefore, this analysis will be unable to come to a definite conclusion. With any luck, the discussion will help us to have a framework within which we can proceed as more facts become verified.

### Overview Derived From SARS CoV-2 Genetic Sequence

I want to start with the science. The very helpful graphic below is lifted from this paper in Current Biology. It is in three sections. The section on the left illustrates what we know from the genetic sequence of the virus when that is compared to other known viruses. What it shows is that the closest overall relative to SARS CoV-2, with a sequence identity of 96%, is RaTG13, another coronavirus isolated from a bat:



Let's move to this Nature Medicine article from March 17 and this Cell article from April 16 for the narrative on diving into the distinguishing features of SARS CoV-2 from its genetic sequence.

From the Nature Medicine article, we get a description of the features of SARS CoV-2 that distinguish it from other known viruses (these features are what the center and right panels of

the graphic address):

Our comparison of alpha- and betacoronaviruses identifies two notable genomic features of SARS-CoV-2: (i) on the basis of structural studies and biochemical experiments, SARS-CoV-2 appears to be optimized for binding to the human receptor ACE2; and (ii) the spike protein of SARS-CoV-2 has a functional polybasic (furin) cleavage site at the S1-S2 boundary through the insertion of 12 nucleotides, which additionally led to the predicted acquisition of three O-linked glycans around the site.

To translate some of the terms and clarify a bit, there are four genera of coronaviruses, with alpha and beta infecting mammals and delta and gamma infecting birds. The genome is the genetic sequence of the virus. I would usually say the DNA sequence, but coronaviruses are RNA viruses. There has been much discussion of ACE2 on this blog in the comments, so for now let's just say ACE stands for angiotensin converting enzyme and ACE2 is present on the surface of many cell types found in many different tissues within the body. So what stands out here is that the structure of the virus spike protein, as determined from its genetic sequence and tests in the lab, allows it to bind exceptionally well to ACE2 when compared to other coronaviruses.

The middle panel of the graphic shows us that although the overall sequence of SARS CoV-2 is very closely aligned to the bat virus, when we narrow it down to only compare the region where the spike protein binds to ACE2, it is a perfect match of that part of a pangolin virus, while it is very different from the bat virus. For the important stretch of the spike protein (these amino acids are not next to each other when the gene sequence is read from start to finish, but once the protein is assembled from amino acids, the amino acids are close to each other from the way the protein assumes its three dimensional

structure), the gene encodes a string of five amino acids in the protein that matches exactly with the pangolin virus sequence but in only the first of the five positions on the bat virus sequence.

But that final panel and the second half of the Nature Medicine snippet goes further in what is different about this virus. The gene for the spike protein encodes two subunits, S1 and S2. Remarkably, SARS CoV-2 has acquired a site where the two subunits can be separated using an enzyme called furin that is found in mammalian cells. The right panel shows us that neither the bat sequence nor the pangolin sequence has a furin cleavage site.

The Cell paper tells us that a furin cleavage site has not been seen in the betacoronaviruses closely related to SARS CoV-2. It has been seen in other human coronaviruses, though. Of further significance is that a furin cleavage site also appears in the more pathogenic bird flu viruses.

### **Not A Lab Construct**

From the Nature Medicine article, we get one of the most convincing arguments I've seen against the virus being created in a lab:

While the analyses above suggest that SARS-CoV-2 may bind human ACE2 with high affinity, computational analyses predict that the interaction is not ideal and that the RBD sequence is different from those shown in SARS-CoV to be optimal for receptor binding. 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.

So, in other words, if someone in the lab wanted to set out to make a virus with the best

possible ACE2 binding site, this is not the sequence the computer or the literature would have given them. That suggests that this very good binding sequence is a product of natural evolution instead. The Nature Medicine article also further noted that the genetic sequence of SARS CoV-2 differs too much from that of any other known coronavirus sequence for one of the known viruses to have been used as a starting point in engineering this stronger pathogen.

### **The Species Jump**

Perhaps the most important step in the emergence of SARS CoV-2 is the jump from its initial host species to humans. This could have happened directly, or as in the case of MERS CoV, which went from bats to camels to humans, with an intermediate host. Note that MERS still has not adapted to efficient human to human transmission, and so when we see it, it's usually from multiple camel to human events.

The problem here is that we don't have proof of the host from which humans were first infected with SARS CoV-2. In other words, no virus isolated from an animal so far is related closely enough at the sequence level to SARS CoV-2 that we can say this is where humans were first infected, as we can tell from the MERS jumps from camels to humans. As we will discuss below, and as you are well aware, early suspicion on the origin of human infection centered on the wet market in Wuhan. Remarkably, authors of the Cell paper visited the market and took these pictures in October 2014 because they were concerned that wet markets in general, and this one in particular, represent a particularly large risk for bringing humans into contact with less commonly encountered hosts of potentially deadly viruses:



[Download : Download high-res image \(2MB\)](#)

[Download : Download full-size image](#)

Figure 1. The Huanan Seafood and Wildlife Market in Wuhan, China

The photographs (credit: E.C.H.) were taken when both authors visited the market together in October 2014 and highlight some of the wide variety of wildlife on sale, providing a potent mechanism for zoonotic transmission. Importantly, although many of the early COVID-19 cases were linked to this market, its role in the initial emergence of SARS-CoV-2 remains uncertain.

The caption properly notes that many early cases are linked to the market, but we don't yet have proof of where and how the first human infection(s) took place. In discussing the jump and subsequent outbreak, the Cell authors continue:

The emergence and rapid spread of COVID-19 signifies a perfect epidemiological storm. A respiratory pathogen of relatively high virulence from a virus family that has an unusual knack of jumping species boundaries, that emerged in a major population center and travel hub shortly before the biggest travel period of the year: the Chinese Spring Festival.

/snip/

While our past experience with coronaviruses suggests that evolution in animal hosts, both reservoirs and intermediates, is needed to explain the emergence of SARS-CoV-2 in humans, it cannot be excluded that the virus acquired some of its key mutations during a period of "cryptic" spread in humans prior to its first detection in December 2019. Specifically, it is possible that the virus emerged earlier

in human populations than envisaged (perhaps not even in Wuhan) but was not detected because asymptomatic infections, those with mild respiratory symptoms, and even sporadic cases of pneumonia were not visible to the standard systems used for surveillance and pathogen identification. During this period of cryptic transmission, the virus could have gradually acquired the key mutations, perhaps including the RBD and furin cleavage site insertions, that enabled it to adapt fully to humans. It wasn't until a cluster of pneumonia cases occurred that we were able to detect COVID-19 via the routine surveillance system. Obviously, retrospective serological or metagenomic studies of respiratory infection will go a long way to determining whether this scenario is correct, although such early cases may never be detected.

So, the sequence information comes to a dead end here until the details of the epidemiology are reconstructed. As the authors note, it likely will prove impossible to sample many of the most important animals and humans that would clarify the route and timing. It is further worth noting that the bat from which the RaTG13 sequence is derived was found in Yunnan province, a very long way from Wuhan.

### **Epidemiology**

It appears that as of this writing, the earliest known infection may have been a shrimp seller in the wet market who first developed symptoms on November 17. Also, this Lancet article provides further details on some of the early studies showing a high concentration of cases affiliated with the market in December. The Lancet graphic suggests a case on December 1 not affiliated with the market and the start of the market cluster on the tenth, with 27 of the 41 early patients considered here being associated with the wet market. If that were indeed the earliest

case, we might think we've seen the index case. But if the South China Post article is to be believed, the shrimp seller fell ill on November 17 and, according to the article, one to five people a day from that day forward had the disease. If we believe that information, then the virus appears to have already been circulating before the middle of November.

It is when we start getting into this information that accusations of hiding information are thrown about. Were there earlier cases that China suppressed or that simply went undetected? We have no way of knowing at this point.

A further point that comes from the Cell paper is that SARS CoV-2 has been circulating long enough that minor variations in the gene sequence are arising that don't affect pathogenicity but allow for tracing of various lineages of the virus in its spread around the globe. They also note that the lineages allow them to go back in time over the evolution of those sequences and the diversity diminishes a lot as they get back to the early isolates from Wuhan. This is further confirmation for Wuhan being essential in the earliest part of the outbreak.

### **Accidental Release**

It is here that the noise gets really loud. If we accept the really strong evidence that SARS CoV-2 was not deliberately made in a laboratory, there remains the possibility that the virus could have escaped from a laboratory that studies potential pandemic agents.

As long ago as 2004, Rutgers scientist Richard Ebright spoke out against the massive amount of funding that was funneled into research on bioweapons after the 2001 anthrax attacks. From the New York Times:

Dr. Ebright disagrees with much of the security community about how best to protect the nation from attacks with biological weapons.



The government and many security experts say one crucial step is to build more high-security laboratories, where scientists can explore the threats posed not only by deadly natural germs, but also by designer pathogens – genetically modified superbugs that could outdo natural viruses and bacteria in their killing power. To this end, the Bush administration has earmarked hundreds of millions of dollars to erect such laboratories in Boston; Galveston, Tex.; and Frederick, Md., among other places, increasing eightfold the overall space devoted to the high-technology buildings.

Dr. Ebright, on the other hand, views the plans as a recipe for catastrophe. The laboratories, called biosafety level 4, or BSL-4, are costly, unnecessary and dangerous, he says.

"I'm concerned about them from the standpoint of science, safety, security, public health and economics," he added in an interview. "They lose on all counts."

Ebright continues:

The labs, Dr. Ebright says, are a perilous overreaction to an inflated threat and will do more harm than good.

Although the threat of biological warfare is real, the weapons used by terrorists are unlikely to be the next-generation agents that the high-security labs are intended to study, he says. Yet by increasing the availability of such pathogens, Dr. Ebright argues, the labs will "bring that threat to fruition."

"It's arming our opponents," he said.

In addition, he says, the laboratories could leak. They could put deadly

pathogens into irresponsible hands and they will divert money from other worthy endeavors like public health and the frontiers of biology. Moreover, their many hundreds of new employees would become a pool of deadly expertise that could turn malevolent, unleashing lethal germs on an unsuspecting public.

Note the “leak” bit. The article goes on:

But Dr. Ebright noted that the deadly SARS virus recently escaped from BSL-4 and BSL-3 labs in Taiwan, Singapore and Beijing, in each case setting off minor epidemics that killed or sickened people.

This 2014 paper from the Center for Arms Control goes into detail on two separate escapes of SARS from the same laboratory in Beijing, along with four other documented cases of releases of possibly pandemic pathogens if you care to read further. Suffice it to say that Ebright was right that with the proliferation of these new labs, there would be leaks. So far, they’ve all been accidental instead of the type feared by Ebright where someone from inside a laboratory deliberately releases a pathogen.

With regard to the SARS CoV-2 outbreak, rumors from nearly the very beginning swirled about a lab in Wuhan. There is in fact a level 4 containment lab in Wuhan and there is also a level 2 lab as well, that I believe is very close to the wet market.

Should there have been an accidental release from either of these labs, at this point we would have to postulate that China has specifically quashed all information relating to this event and kept the laboratory personnel and any close family or other contacts who may have been infected out of the databases of patients.

But that hasn’t stopped the noise. Some aspects of the noise even begin to look to me like an

information operation of sorts. Of course, since we don't know the originator of the operation, we don't know if it is actual intelligence being leaked or if it is disinformation being sown to add to the chaos.

At any rate, this April 2 column from David Ignatius put the idea of an accidental leak from a Wuhan lab into the Washington Post. Those who follow intelligence community news know that Ignatius is often thought of as a mouthpiece for information the CIA wants disseminated. Are they his source here? Was some other information operative his source?

Then things really heated up on April 15. Here is John Roberts of Fox News asking Trump a question during the April 15 "press conference":

Wow. That's an incredibly specific question. It assumes a female intern at the lab who infected a boyfriend and then she (or did he, not clear to me from Roberts' phrasing) went to the market. Even though this was April 15, I've seen no further pushing of this specific version of the story.

But Trump's response is a bit concerning. Note that he says they're "hearing that story a lot", but then makes a really big deal of the word "sources". Given Trump's history of spilling classified intelligence, and the constant warnings to him about such leaks compromising "sources and methods", I almost wonder if that's a genuine response of his lizard brain to all those warnings. We simply have no way of knowing that or knowing if perhaps those "sources" happen to lie outside the intelligence community and among circle of wingnuts who have the ears of Trump and Fox News and he's really proud of them but doesn't want to divulge them.

That same day, Josh Rogin put out a Washington Post column pushing the leak from a lab story, this time tying it directly to the State Department cables in 2018 about lax biosecurity protocols at the level 4 containment lab in Wuhan that Roberts mentioned. But Rogin didn't

include the specifics about the intern.

I've heard nothing further on the intern question, but the general idea of an escape from a Wuhan lab still gets tossed around. Ignatius returned to the idea of an accidental release on April 23. He even talked to Ebright:

“Science is not going to shift this from a ‘could have been’ to a ‘probably was,’ ” messaged Richard H. Ebright, a leading biosafety expert at Rutgers. “The question whether the outbreak virus entered humans through an accidental infection of a lab worker . . . can be answered only through a forensic investigation, not through scientific speculation.” Ebright told me the Chinese government should launch a forensic investigation by reviewing “facilities, samples, records, and personnel.”

Given Ebright's history of predicting just such an accidental release, I find it very reassuring that he isn't ready to say that's what happened. As he rightfully points out, we can only know what happened when detailed information is assembled on the epidemiology of the earliest cases. Only Chinese medical investigators can know whether any laboratory personnel, and especially whether any family or other close contacts of them appear on the timeline of the early infections. It is also crucial to know where any such infections, if they exist, fall on the timeline in relation to cases affiliated with the wet market.

My gut feeling is that the evidence still very strongly points to the virus originating through the wet market, but I also think the index case there likely goes back even earlier than the November 17 case discussed above, since there are suggestions there were other cases appearing daily by then. Also, it's hard to imagine that if the official intelligence community had a story as specific as the intern story and had

evidence to back it up, that Trump wouldn't be trumpeting it on a daily basis to deflect the criticism being heaped on his response to the outbreak.

Stay tuned. I suspect the story will take several more turns before we ever reach any level of certainty.